Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Juliane Hafermann, M. Sc. Molecular Biotechnology

Born in Regensburg

Oral examination: 25.1.2018

The effect of APOBEC3A-mediated mutagenesis on tumour growth of immortalised, transformed cells and on survival of head and neck squamous cell carcinoma patients

Referees: Prof. Dr. Peter Angel

Prof. Dr. Martin Löchelt

Acknowledgements

I would like to express my gratitude to my advisor Prof. Dr. Martin Löchelt for the support of my PhD studies and all related research. I would especially like to thank him for giving me the chance to work in his laboratory, his continued interest in my project and his constant willingness for discussion.

My sincere thanks also goes to Dr. Axel Szabowski for his patience, encouragement, and the many valuable insights into science, management, presentation and writing he shared with me. He challenged me to go outside my comfort zone and so helped me grow as a scientist and as a person. I am also grateful for his help with editing my thesis.

I would also like to thank Prof. Dr. Michael Pawlita for his valuable support and the many helpful discussions and comments.

I am indebted to PD Dr. Jochen Heß for providing me with the samples and data of the HIPO-POP019 study, and for his enthusiasm to discuss ideas and try out new analyses.

My gratitude is also due to Dr. Marc Zapatka for valuable input, and for analysing the whole genome sequences and performing the mutational signature analysis together with Dr. Susanne Gröbner, Dr. Mario Hlevnjak, and Dr. Agnes Hotz-Wagenblatt.

I would further like to thank PD Dr. Karin Müller-Decker and Brigitte Steinbauer for their support in planning the *in vivo* study and handling the animal experiments.

My thanks goes to the members of my Thesis Advisory Committee, Prof. Dr. Peter Angel, Prof. Dr. Ralf Bartenschlager and Prof. Dr. Peter Lichter, for their support and constructive criticism of my research project.

Support by the DKFZ Light Microscope Facility, in particular by Dr. Damir Krunic and Dr. Felix Bestvater, in microscope handling and image processing is gratefully acknowledged.

I thank the Sample Processing Laboratory and High Throughput Sequencing units of the Genomics & Proteomics Core Facility, German Cancer Research Center (DKFZ), for providing excellent sample quality control, sequencing library preparation and whole genome sequencing services.

I am grateful to Dr. Lea Schröder and Dr. Dana Holzinger for determining the HPV status of the tumours in the HIPO-POP019 study.

I gratefully acknowledge the funding from the Helmholtz International Graduate School that made my research possible. In addition, I would like to thank all members of the Graduate Office for their help and support, and for the chance to get involved at the DKFZ beyond the confines of my PhD research.

I am grateful to all my colleagues present and past in F020 who contributed to my time at the DKFZ for the many helpful discussions, the good advice and the many friendships.

My sincere thanks also goes to Annica Flemming, Felix Behr, Alexander Hogrebe, Taga Lerner, Samuel Bousseau, and Isabell Hafermann, who proof-read my thesis.

Last but by no means least, I would like to thank my friends and in particular my family for far too many things to be listed here.

Table of contents

Abstract

Inflammation and mutagenesis contribute to cancer progression. The cytidine deaminase APOBEC3A might link these two processes. On the one hand, the expression of APOBEC3A is regulated by infections and pro-inflammatory stimuli. On the other hand, APOBEC3A causes mutational signatures that have been found in the genomes of various cancer types. However, the major contribution of APOBEC3A during the process from tumour initiation to a fully formed disease is still under debate. APOBEC3A-mediated mutagenesis could (a) affect early stages of tumour development by mutating cancer driver and/or tumour suppressor genes, (b) have an influence by increasing the total mutational load, or (c) shape the tumour by increasing genetic heterogeneity, in particular during late stages of cancer progression. In addition, it is unclear whether parameters capturing the process of APOBEC3A-mediated mutagenesis during cancer progression have a prognostic value. Here, we address (I) whether APOBEC3A-mediated mutations alter tumour growth of already immortalised, transformed cells and (II) whether a combination of parameters describing past and ongoing APOBEC3-mediated mutagenesis have a prognostic value in a cohort of head and neck squamous cell carcinoma (HNSCC) patients. In this study, we established HEK293 cell populations carrying the APOBEC3-mediated mutational signatures. As ongoing APOBEC3A-driven mutagenesis causes a genetically heterogeneous cell population, single cell clones were isolated and analysed. APOBEC3A activity was lost in these clones by various mechanisms after creating the APOBEC3 mutational signature. The clones were characterised by quantifying proliferation, migration, cisplatin resistance and *in vivo* tumour growth in a xenograft mouse model. No phenotypic difference was observed between APOBEC3A-mutagenised cells and controls regarding any of the studied phenotypes either *in vitro* or *in vivo*. This suggests that APOBEC3A-mediated mutagenesis has no effect on the tumour growth of immortalised and transformed HEK293 cells. These results hint that additional APOBEC3A-mediated mutagenesis in cells with a large number of pre-existing tumourigenic alterations cannot further affect cellular growth. Analysing material from tumour patients with HNSCC revealed that tumours can show APOBEC3A expression and/or APOBEC3B expression and/or the APOBEC3 mutational signature. None of these parameters by themselves succeeded in stratifying an HNSCC cohort. In contrast, a principal component analysis combining transcriptional and genomic data lead to a statistically significant stratification regarding progression-free survival. The approach developed here might be a valuable additional tool in personalised medicine.

Zusammenfassung

Entzündung und Mutagenese tragen zur Tumorprogression bei. Die Cytidin-Deaminase APOBEC3A ist eine potentielle Verbindung zwischen diesen Prozessen. Einerseits wird die Expression von APOBEC3A durch Infektionen und pro-inflammatorische Stimuli reguliert. Andererseits verursacht APOBEC3A Mutationssignaturen, die in den Genomen verschiedener Krebsarten gefunden wurden. Der wesentliche Beitrag von APOBEC3A zu dem Prozess von der Tumorinitiierung bis hin zur vollendeten Krankheit wird jedoch immer noch diskutiert. APOBEC3A-vermittelte Mutagenese könnte (a) frühe Stadien der Tumorentstehung durch Mutation von Tumortreiber- und/oder Tumorsuppressorgenen beeinflussen, (b) die Gesamtmutationslast erhöhen oder (c) den Tumor durch Erhöhung der genetischen Heterogenität, insbesondere in späten Stadien der Tumorprogression, verändern. Darüber hinaus ist unklar, ob Parameter, die den Prozess der APOBEC3A-vermittelten Mutagenese während der Krebsentstehung festhalten, zur Prognose verwendet werden können. Hier wird untersucht, (I) ob APOBEC3A-vermittelte Mutationen das Tumorwachstum bereits immortalisierter, transformierter Zellen verändern kann und (II) ob eine Kombination von Parametern, die vergangene und gegenwärtige APOBEC3-vermittelte Mutagenese beschreiben, zur Prognose einer Patientenkohorte mit Kopf- und Halskrebs (HNSCC) beitragen kann. In dieser Studie wurden HEK293-Zellpopulationen etabliert, die APOBEC3-vermittelte Mutationssignaturen tragen. Da die APOBEC3A-getriebene Mutagenese eine genetisch heterogene Zellpopulation erzeugt, wurden Einzelzellklone isoliert und analysiert. Die APOBEC3A-Aktivität ging in diesen Klonen nach dem Entstehen der APOBEC3 Mutationssignaturen durch verschiedene Mechanismen verloren. Die Klone wurden bezüglich Proliferation, Migration und Cisplatinresistenz sowie *in vivo* Tumorwachstum in einem Xenograft-Mausmodell charakterisiert. Es wurde kein Unterschied zwischen APOBEC3A-mutagenisierten Zellen und Kontrollen für die *in vitro* oder *in vivo* untersuchten Phänotypen beobachtet. Dies deutet darauf hin, dass die APOBEC3A-vermittelte Mutagenese keinen Einfluss auf das Tumorwachstum von immortalisierten und transformierten HEK293-Zellen hat. Diese Ergebnisse deuten an, dass eine zusätzliche APOBEC3A-vermittelte Mutagenese das Wachstum von Zellen, die bereits eine große Anzahl tumorerzeugender Veränderungen enthalten, nicht weiter verändern kann. Bei der Analyse von HNSCC-Tumorpatienten zeigte sich, dass Tumore APOBEC3A-Expression und/oder APOBEC3B-Expression und/oder die APOBEC3-Mutationssignatur aufweisen können. Mit keinem dieser Parameter allein kann die HNSCC-Kohorte stratifiziert werden. Im Gegensatz dazu führt eine Principal Component Analyse, die transkriptionelle und genomische Daten kombiniert, zu einer statistisch signifikanten

Stratifizierung für progressionsfreies Überleben. Der hier entwickelte Ansatz könnte ein wertvolles zusätzliches Hilfsmittel in der personalisierten Medizin sein.

List of abbreviations

A	adenine
A3A	APOBEC3A
A3B	APOBEC3B
ACTB	β-actin
AID	activation-induced deaminase
apoB	apolipoprotein B
APOBEC	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
APS	ammonium persulphate
ATM	ataxia telangiectasia mutated
ATR	ataxia telangiectasia and Rad3-related protein
BGH	bovine growth hormone
bp	base pair
BSA	bovine serum albumin
C	cytidine
cDNA	copy DNA
CMV	cytomegalie virus
COSMIC	Catalogue of Somatic Mutations in Cancer
Cys	cysteine
DAB	3,3'-diaminobenzidine
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dox	doxycycline
dsDNA	double-stranded DNA
DTT	1,4-dithiothreitol
ECL	enhanced chemiluminescence
ECS	extracapsular spread
EDTA	ethylenediaminetetraacetic acid
Fam	6-carboxyfluorescein
FCS	foetal calf serum
FFPE	formalin-fixed paraffin-embedded
fwd	forward
G	guanine
Glu	glutamic acid
gDNA	genomic DNA
HBV	hepatitis B virus
HE	haematoxylin and eosin
HEK293	human embryonic kidney cells
Hex	6-hexachlorofluorescein
HF	high fidelity
His	histidine
HIV	human immunodeficiency virus
HNSCC	head and neck squamous cell carcinoma

HPV	human papillomavirus
HRP	horse radish peroxidase
IC_{50}	half-maximal inhibitory concentration
IFN	interferon
IFN-R	interferon receptor
JAK	Janus kinase
LINE	long interspersed nuclear elements
LPS	lipopolysaccharides
LTR	long terminal repeat
mRNA	messenger RNA
NF-κB	nuclear factor kappa-light-chain-enhancer of B cells
NOD	non-obese diabetic
NRT	no reverse transcriptase
NSG	NOD scid gamma mice
nt	nucleotides
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PD-1	programmed death-1
PD-L1	programmed death ligand-1
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PKC	protein kinase C
PMA	phorbol 12-myristate-1-acetate
pop.	population
Pro	proline
qPCR	quantitative polymerase chain reaction
R	purine base
rcf	relative centrifugal force
rev	reverse
RNA	ribonucleic acid
RT	reverse transcriptase
scid	severe combined immunodeficiency
SD	standard deviation
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SINE	short interspersed nuclear elements
ssDNA	single-stranded DNA
STAT	signal transducer and activator of transcripton
STING	stimulator of interferon genes
T	thymine
TAE buffer	buffer containing tris base, acetic acid and EDTA
TBE buffer	Tris borate EDTA buffer
TCGA	The Cancer Genome Atlas
TEMED	N,N,N',N'-tetramethylethane-1,2-diamine
TLR	Toll-like receptor

TPA	12-O-tetradecanoylphorbol 13-acetate
UPL	universal probe library
UTR	untranslated region
Vif	virus infectivity factor
Xaa	any amino acid
Y	pyrimidine base

List of figures

List of tables

1. Introduction

Cancer is often linked closely with inflammation and infection (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011; Hoste et al., 2015). Mutagenesis and genetic instability have been linked to an inflammatory tumour environment (Colotta et al., 2009). Mutations are essentially the driving force of cancer (Stratton et al., 2009; Stratton, 2011). They start accumulating well before cancer initiation and shape the behaviour and characteristics of a tumour throughout its promotion and progression. The mutations a tumour accumulates during its development can determine the prognosis and survival of patients. While it is clear that infection and inflammation are linked to mutagenesis in cancer, the causal relation still remains unclear. One potential link between infection, inflammation and cancer is the family of APOBEC3 cytidine deaminases, which can act as mutagens under inflammatory conditions.

1.1. APOBEC3A is part of the innate immune system

The apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) proteins are a family of cytidine deaminases (reviewed in Harris and Liddament, 2004; Cullen, 2006; Conticello et al., 2007; Aguiar and Peterlin, 2008; Chiu and Greene, 2008; Conticello, 2008; Smith et al., 2012; Refsland and Harris, 2013; Knisbacher et al., 2015; Swanton et al., 2015). The APOBEC family members exhibit the ability to deaminate cytidine to uracil in RNA (Teng et al., 1993) and/or DNA substrates (Harris et al., 2002a). They are defined by cytidine deaminase active sites with the consensus sequence His-Xaa-Glu-Xaa_{23-28}-Pro-Cys-Xaa_{2-4}-Cys (with Xaa denoting any amino acid). This conserved zinc-coordinating motif is present in all members of the APOBEC family and is required for deaminase activity (Wedekind et al., 2003). APOBEC1 was the first of this enzyme family to be discovered. It specifically deaminates C_{6666} of the apolipoprotein B (apoB) mRNA, thus creating a truncated version of the apoB protein by creating a premature stop codon (Teng et al., 1993). Expression of APOBEC1 in *E. coli* furthermore showed its potential to deaminate cytidine in DNA (Harris et al., 2002a). Another family member, activation-induced deaminase (AID), is essential for somatic hypermutation and class switch recombination during antibody maturation by editing the immunoglobulin-encoding genes (Muramatsu et al., 2000; Harris et al., 2002b). So far, the physiological functions of APOBEC2 (Liao et al., 1999; Jarmuz et al., 2002) remain unknown, and it does not appear to share the cytidine deaminase activity of the other family members (Harris et al., 2002a). Hardly anything is known about APOBEC4. APOBEC4 contains the zinc-coordinating domain characteristic of the APOBEC family, suggesting that it may possess cytidine deaminase activity (Rogozin et al., 2005). However, so far, no deamination of single stranded DNA substrates has been observed (Ito et al., 2017).

1.1.1. APOBEC3 cytidine deaminase family

There are in total seven human APOBEC3 isoforms, namely APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3DE, APOBEC3F, APOBEC3G and APOBEC3H (Figure 1), encoded in tandem on chromosome 22 (Jarmuz et al., 2002). While APOBEC3A, C and H each contain only one active site, APOBEC3B, DE, F and G each contain two cytidine deaminase domains (Jarmuz et al., 2002; Wedekind et al., 2003; Conticello et al., 2005).

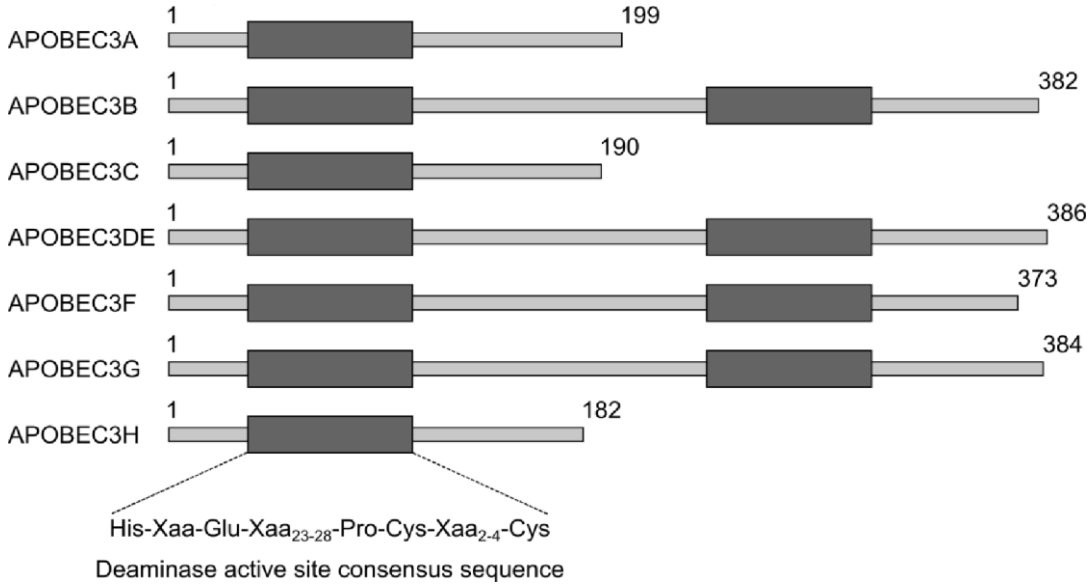

Figure 1 The human APOBEC3 family of cytidine deaminases. APOBEC3A, C and H contain one deaminase domain, whereas APOBEC3B, DE, F and G contain two deaminase domains. The dark grey boxes represent the cytidine deaminase active sites, whose consensus sequence is given below. The numbers represent the number of amino acids for each protein. Adapted and modified from Conticello et al. (2005), Holmes et al. (2007), Chiu and Greene (2008), Smith et al. (2012) and Wang et al. (2015).

The APOBEC3 family members are expressed in different tissues and cell types. APOBEC3C, F and G are ubiquitously expressed (Refsland et al., 2010 and reviewed by Conticello, 2008 and Aguiar and Peterlin, 2008). The expression profile of the other APOBEC3 family members is more restricted. APOBEC3A is expressed in monocytes and keratinocytes, and has also been found to be upregulated in lung tissue (Madsen et al., 1999; Jarmuz et al., 2002; Refsland et al., 2010; Land et al., 2013; Wang et al., 2014; Yang et al., 2016). APOBEC3B expression has been described in the intestines, mammary glands, uterus, and liver, as well as in keratinocytes (Madsen et al., 1999; Jarmuz et al., 2002; Bonvin et al., 2006; Wang et al., 2014 and reviewed by Conticello, 2008). APOBEC3DE and H are restricted to expression in immune cells and related organs such as the thymus and spleen as well as the thyroid (reviewed by Conticello, 2008). APOBEC3A, C, DE, F and G have also been found in primary fallopian epithelial tissues (Brachova et al., 2017).

1.1.2. APOBEC3 as restriction factors of viruses, endogenous retroelements and foreign DNA

The family of APOBEC3 deaminases are antiviral restriction factors and as such part of the innate and intrinsic immune defense (reviewed by Harris and Liddament, 2004; Aguiar and Peterlin, 2008). APOBEC3G was first identified as an HIV restriction factor when Sheehy et al. (2002) noticed that virus infectivity factor (Vif)-deficient HIV-1 was able to replicate in some permissive cell lines, but not in other (non-permissive) cells. The presence or absence of APOBEC3G expression was found to be the distinguishing factor between a non-permissive and a permissive cell line, respectively. APOBEC3G restricts retroviruses based on its ability to cause in the the viral genome after reverse transcription (Harris et al., 2002a; Harris et al., 2003). APOBEC3G deaminates cytidines to uracils in the first single-stranded cDNA during the reverse transcription step of retroviral genome replication (Suspène et al., 2004). The uracils then serve as templates for the incorporation of adenines in the complementary strand, thereby introducing a large number of G-to-A mutations into the viral genome and affecting the viability of the virus (reviewed by Harris and Liddament, 2004 and Chiu and Greene, 2008). Since then, various members of the APOBEC3 family have been found to restrict various retroviruses (summarised by Chiu and Greene, 2008). In addition, some APOBEC3 family members are able to restrict DNA viruses. For instance, APOBEC3A has been found to restrict adeno-associated virus (Chen et al., 2006) and human papillomavirus (HPV) (Warren et al., 2015b). APOBEC3B is able to restrict hepatitis B virus (HBV) (Lucifora et al., 2014). A review by Warren et al. (2017) gives an overview of the DNA viruses restricted by the different APOBEC3 family members.

The APOBEC3 cytidine deaminases are furthermore involved in the restriction of endogenous retroelements (Schumacher et al., 2008, reviewed by Aguiar and Peterlin, 2008). Retroelements are mobile DNA sequences that propagate by coupling transcription with reverse transcription. The reverse transcribed retroelement DNA is then inserted into the genome in a process called retrotransposition. Endogenous retroelements comprise retrovirus-like elements containing long terminal repeats (LTRs), as well as non-LTR retrotransposons such as human long interspersed nuclear elements (LINEs, e.g. LINE-1) and short interspersed nuclear elements (SINEs, such as Alu) (reviewed by Kazazian, 2004). All APOEBC3 family members are able to inhibit retrotransposition of LINE-1 to different degrees (Chen et al., 2006; Muckenfuss et al., 2006; Kinomoto et al., 2007; Bulliard et al., 2011). Retrotransposition of murine intracisternal A-particle (IAP), as an example of an LTR-retrotransposon, is strongly inhibited by APOBEC3A and APOBEC3B, and to a lesser degree by APOBEC3G (Bogerd et al., 2006; Chen et al., 2006; Knisbacher and Levanon, 2015). The reviews by Chiu and

Greene (2008) and Smith et al. (2012) provide an overview about the spectrum of viruses and retroelements restricted by the different APOBEC3 cytidine deaminases.

In addition, APOBEC3A was found to cause the degradation of foreign DNA such as transfected plasmids both in cell culture (Stenglein et al., 2010) and in mice (Kostrzak et al., 2015), and was found to edit mRNA in human cells (Niavarani et al., 2015; Sharma et al., 2015; Sharma et al., 2017).

1.1.3. Regulation of APOBEC3 expression by inflammatory stimuli

Inflammation and cancer are intricately linked (reviewed by Colotta et al., 2009). As restriction factors and as part of the innate immune system, APOBEC3 expression can be upregulated by various inflammatory stimuli in a range of cell types and tissues (Figure 2). APOBEC3A and APOBEC3B (formerly called phorbolin-1 and phorbolin-1-related protein/phorbolin-2, respectively) were first identified in primary human keratinocytes after treatment with the tumour promoter phorbol 12-myristate-1-acetate (PMA), also known as 12-O-tetradecanoylphorbol 13-acetate (TPA) (Madsen et al., 1999). PMA/TPA treatment has also been found to induce APOBEC3B expression in a range of cell lines through protein kinase C (PKC) and nuclear factor kappa-light-chain-enhancer of B cells (NF-κB) signalling (Leonard et al., 2015; Maruyama et al., 2016). All APOBEC3 family members with the exception of APOBEC3DE are induced by interferon-α (IFN-α) in macrophages and monocytes (Peng et al., 2006; Refsland et al., 2010; Stenglein et al., 2010; Thielen et al., 2010; Aynaud et al., 2012; Carpenter et al., 2012; Mehta et al., 2012; Suspene et al., 2017). IFN-α was also shown to stimulate the expression of all APOBEC3 members to different degrees in a range of cancer cell lines (Menendez et al., 2017). Similarly, lipopolysaccharides (LPS) can also induce expression of all APOBEC3 family members with the exception of APOBEC3B in human monocytes (Mehta et al., 2012). APOBEC3A expression is potently upregulated by IFN-α in $CD4^+$ T-cells, while APOBEC3G, F and H are only moderately induced (Koning et al., 2009). Treatment with IFN-α also increases expression of APOBEC3A, B, F and G in hepatocytes (Bonvin et al., 2006; Lucifora et al., 2014) as well as expression of APOBEC3A and G in both chimpanzee and human livers (Lucifora et al., 2014). It furthermore induces expression of APOBEC3A, G and F in myeloid dendritic cells (Mohanram et al., 2013).

In addition, APOBEC3A can also be induced by IFN-β in normal immortalised skin keratinocytes (Wang et al., 2014; Warren et al., 2015b), by cytoplasmic double-stranded DNA (such as viral DNA genomes or transfected plasmids) in a human monocytic cell line (Suspene et al., 2017), by follicular fluid in fallopian cell lines (Brachova et al., 2017) and by viral infection in breast and bladder cancer cell lines (Middlebrooks et al., 2016).

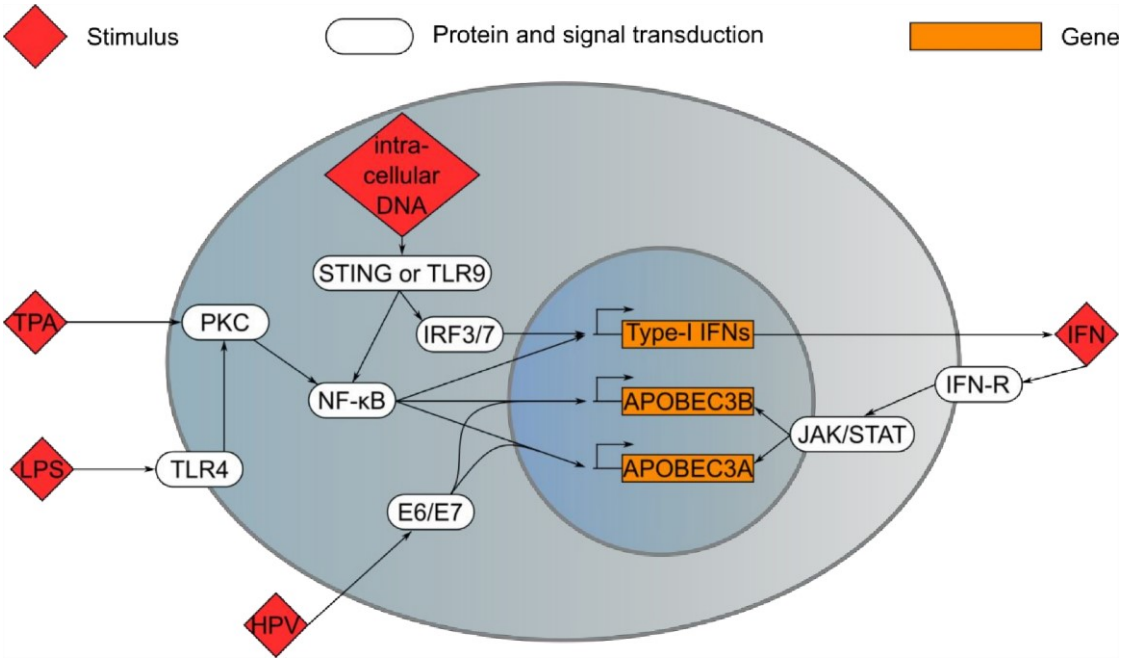

Figure 2 Inflammatory stimuli induce expression of APOBEC3A and B. Lipopolysaccharides (LPS) via Toll-like receptor 4 (TLR4) and 12-O-tetradecanoylphorbol 13-acetate (TPA) induce expression of APOBEC3A and B as well as type-I interferon (type-I IFN) expression through activation of protein kinase C (PKC) and nuclear factor kappa-light-chain-enhancer of B cells (NF-κB) signalling. Intracellular DNA (such as viral DNA genomes or transfected plasmids) is recognised mainly by stimulator of interferon genes (STING) or toll-like receptor 9 (TLR9), which activate expression of APOBEC3A and B and type-I interferons via NF-κB and interferon regulatory factors 3 and 7 (IRF3 and 7). Type-I interferons (either autocrine or paracrine) are recognised by the interferon receptor (IFN-R) and activate expression of APOBEC3A and B through the Janus kinase/signal transducer and activator of transcripton (JAK/STAT) pathway. Human papilloma virus E6 and E7 oncoproteins can also cause APOBEC3A and B expression. Figure based on Madsen et al., 1999; Bonvin et al., 2006; Peng et al., 2006; Schoggins et al., 2011; Mehta et al., 2012; Vieira et al., 2014; Wang et al., 2014; Dempsey and Bowie, 2015; Leonard et al., 2015; Warren et al., 2015b; Maruyama et al., 2016; Raftery and Stevenson, 2017; Suspene et al., 2017.

Drug-induced replication stress has recently been linked to increased APOBEC3B expression (Kanu et al., 2016). A role of p53 in the transcriptional regulation of APOBEC3 was recently reported by Menendez et al. (2017). Activation of p53 in response to genotoxic stress leads to an increase in expression of APOBEC3A, C, DE, H and in some cases also APOBEC3G, but interestingly also causes repression of APOBEC3B expression in human cancer cell lines. APOBEC3B repression is prevented in cells where p53 is either silenced or contains tumour-associated mutations. Furthermore, p53 can enhance IFN-induced APOBEC3 expression (Menendez et al., 2017). The HPV oncoproteins E6 and E7 can also induce APOBEC3A and APOBEC3B expression in human keratinocytes (Vieira et al., 2014; Warren et al., 2015b). In summary, APOBEC3 expression can be induced by inflammatory stimuli and danger signals such as cytoplasmic DNA in a range of cells and tissue types.

1.2. APOBEC3A and APOBEC3B mutate genomic DNA

The different APOBEC3 family members can deaminate DNA viruses to different degrees. Thus, if they have access to the nucleus, they may also induce mutations in genomic DNA. The different APOBEC3 family members are located either in the nucleus, the cytoplasm or both. Overexpression of transgenic APOBEC3 proteins showed that APOBEC3B is the only variant predominantly located in the nucleus due to an active nuclear import mechanism (Lackey et al., 2012). APOBEC3DE, G and F are exclusively found in the cytoplasm, whereas APOBEC3A, C and H can be found cell-wide (Burns et al., 2013b; Burns et al., 2013a). Localisation of APOBEC3 proteins seem to be independent of the cell type (summarised by Land et al., 2013). In total, four of the seven APOBEC3 proteins (A, B, C and H) have access to the nucleus. This means that these enzymes can potentially deaminate the cytidine residues in genomic DNA. Out of the APOBEC3 family members that locate to the nucleus, overexpression of transgenic APOBEC3A has been shown to edit both mitochondrial (Suspène et al., 2011) and genomic DNA (Suspène et al., 2011; Aynaud et al., 2012; Shinohara et al., 2012; Caval et al., 2014b). Endogenous APOBEC3A expression in $CD4^+$ T lymphocytes has been shown to edit nuclear DNA (Mussil et al., 2013). APOBEC3B can also edit nuclear DNA (Caval et al., 2014a).

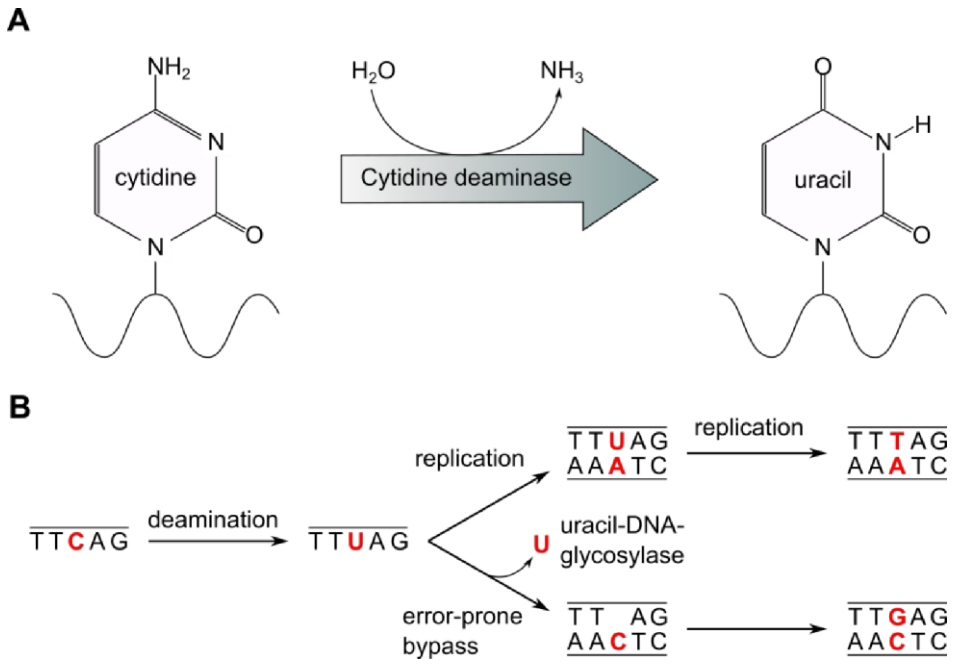

Figure 3 APOBEC3 cytidine deaminases cause C-to-T and C-to-G mutations in genomic DNA. **A** Cytidine deaminases catalyse deamination of cytidine to uracil. **B** Uracil can pair with adenine during DNA replication, resulting in a C-to-T substitution (top). Uracil-DNA-glycosylase remove uracil from DNA, and a C is inserted during error-prone bypass in translesion DNA synthesis, resulting in a C-to-G substitution (bottom). Adapted and modified from Harris and Liddament (2004), Conticello et al. (2007), Refsland and Harris (2013), and Helleday et al. (2014).

The activity of both APOBEC3A and APOBEC3B on chromosomal DNA results in the deamination of cytidine to uracil (Harris et al., 2002a) (Figure 3A). If this uracil is paired with adenine during DNA replication, it results in a permanent C-to-T substitution. However, uracil-DNA-glycosylases (UNGs) remove uracil resulting from deamination of cytidine in DNA, thus creating an abasic site (Lindahl, 1974). Translesion DNA synthesis then inserts adenine or cytidine opposite the abasic site during error-prone bypass of the abasic sites (Chan et al., 2013) (Figure 3B). Deamination of methyl-cytidine directly results in a C-to-T substitution (Carpenter et al., 2012). Thus, deamination of cytidine and methyl-cytidine in genomic DNA by APOBEC3A and APOBEC3B creates C-to-T and C-to-G mutations.

1.2.1. APOBEC3 deaminases edit ssDNA

APOBEC3A binds to and deaminates single-stranded DNA (ssDNA), but it is unable to deaminate double-stranded DNA (dsDNA) or DNA:RNA hybrids (Chen et al., 2006; Carpenter et al., 2012; Bohn et al., 2015; Sharma et al., 2015; Harjes et al., 2017; Kouno et al., 2017). Processive APOBEC3-mediated deamination of ssDNA (Chelico et al., 2006) results in the generation of strand-coordinated mutation clusters (Chan et al., 2012). In the nucleus, ssDNA is present either during transcription, at double-strand breaks or during DNA synthesis. No transcriptional strand bias has been observed for C-to-T or C-to-G mutations (Chan et al., 2012; Nik-Zainal et al., 2012a; Alexandrov et al., 2013b; Kazanov et al., 2015), suggesting that APOBEC3 deaminases do not edit genomic DNA during transcription. It has also been discussed that break-induced replication at dsDNA breaks may be a substrate for APOBEC3-mediated mutagenesis (Roberts et al., 2012; Sakofsky et al., 2014).

The majority of APOBEC3-mediated mutations, however, are enriched in late-replicating regions of the lagging strand during DNA replication (Kazanov et al., 2015; Haradhvala et al., 2016; Hoopes et al., 2016; Morganella et al., 2016; Seplyarskiy et al., 2016b; Seplyarskiy et al., 2016a). Furthermore, APOBEC3A-mediated deamination is more prevalent in replicating than in resting cells (Green et al., 2016). This suggests that APOBEC3-driven mutagenesis occurs primarily on the ssDNA of the lagging-strand template that is exposed between Okazaki fragments during DNA replication (discussed in Haradhvala et al., 2016).

1.2.2. Context of APOBEC3-mediated mutations

APOBEC3 family members show a preference regarding the nucleotide immediately 5’ of their target cytidine (reviewed by Refsland and Harris, 2013). Both APOBEC3A and APOBEC3B prefer a thymine in the 5’-position of their target cytidine, so deamination

occurs in a 5'-TC context (the edited base is underlined) (Bishop et al., 2004; Chen et al., 2006; Thielen et al., 2010; Love et al., 2012; Shinohara et al., 2012; Burns et al., 2013a; Byeon et al., 2013; Chan et al., 2015; Shi et al., 2016; Harjes et al., 2017).

In summary, APOBEC3A and B edit genomic DNA, which is accessible in a single-stranded form as the lagging strand during DNA replication, thus ultimately creating strand-coordinated C-to-T and C-to-G mutations in a 5'-TC context.

1.3. APOBEC3 deaminases create mutational signatures

APOBEC3 cytidine deaminases with a 5'-TC target specificity have been attributed as the source for the mutational signatures 2 and 13 described by Alexandrov et al. (2013b; 2013a). A mutational signature is a characteristic pattern of mutations. All single nucleotide exchanges can be classified into six subgroups if they are referred to by the pyrimidine of the mutated Watson-Crick base pair: C-to-A, C-to-G, C-to-T, T-to-A, T-to-C and T-to-G. These can be directly flanked by either A, C, T or G on both sides. Thus, including information on the bases immediately 5' and 3' of the mutated position results in a total of 96 unique combinations. These combinations can be used to distinguish the same base substitutions occurring in different sequence contexts (Alexandrov et al., 2013b; Alexandrov et al., 2013a). Mutational signatures are characterised by the frequency of all possible 96 combinations of single nucleotide exchanges in their immediate sequence contexts. Different mutational processes cause different mutational patterns (reviewed by Alexandrov and Stratton, 2014; Helleday et al., 2014; Roberts and Gordenin, 2014b; Hollstein et al., 2016; Petljak and Alexandrov, 2016).

1.3.1. Mutational signatures are the product of mutagenesis and repair

Generally, mutational signatures are thought to be the result of both mutagenesis and DNA repair processes. In combination, mutational and repair processes generate the characteristic patterns of the different mutational signatures (reviewed by Helleday et al., 2014). For instance, mutational signature 2 is characterised by C-to-T mutations, while signature 13 consists of mostly C-to-G mutations in a 5'-TC context. APOBEC3-mediated deamination of cytidine to uracil in genomic DNA is likely the mutagenic process behind both these signatures (Nik-Zainal et al., 2012a; Roberts et al., 2012; Alexandrov et al., 2013b; Alexandrov et al., 2013a; Roberts et al., 2013). The C-to-T mutations found in signature 2 are created by uracil serving as a template in DNA replication. The C-to-G transversions making up signature 13 are caused by the error-prone bypass (Chan et al., 2013) of abasic sites created through the excision of uracil by uracil-DNA-glycosylase (Lindahl, 1974). This signature shows a bias for C-to-G over C-to-A mutations, which could be caused by REV1 inserting a C opposite the abasic site during

replication (Nik-Zainal et al., 2012a; Taylor et al., 2013). Thus, the underlying mutagenic process is mediated by APOBEC3 deaminase activity in both cases, and the differences in the mutational signatures are likely caused by different involvement of repair processes (Alexandrov et al., 2013a; Taylor et al., 2013; Helleday et al., 2014).

1.3.2. APOBEC3-mediated mutagenesis creates a mutational signature

Based on their nuclear localisation and deaminase activity in a 5'-TC sequence context leading to C-to-T or C-to-G mutations in single-stranded genomic DNA (Nik-Zainal et al., 2014, Morganella et al., 2016 and reviewed by Roberts and Gordenin, 2014b), APOBEC3A and APOBEC3B have been discussed as the most likely causes of mutational signatures 2 and 13 (Figure 4) (Nik-Zainal et al., 2012a; Alexandrov et al., 2013a; Burns et al., 2013b; Burns et al., 2013a; Roberts et al., 2013; Taylor et al., 2013; Alexandrov and Stratton, 2014; Caval et al., 2014b; Caval et al., 2014a; Chan et al., 2015; Hedegaard et al., 2016; Rogozin et al., 2017; Warren et al., 2017).

APOBEC3-mediated mutations have furthermore been associated with regional clusters of hypermutation called *kataegis* (Greek for thunderstorm) originally found in breast cancer genomes (Lada et al., 2012; Nik-Zainal et al., 2012a). The mutations within regions of *kataegis* match the APOBEC3 mutational signatures and are found on the same chromosomal strand over long genomic distances, implying that they were introduced either simultaneously or in a processive manner over a short period of time (Nik-Zainal et al., 2012a; Alexandrov et al., 2013a). They are furthermore associated with chromosomal rearrangements and with chromothripsis (Nik-Zainal et al., 2012b), a single event of catastrophic shattering and subsequent complex rearrangement of chromosomal regions first described by Stephens et al. (2011). The induction of these *kataegis* mutation showers have been linked to APOBEC3A and B, which act upon ssDNA exposed during repair of DNA double-strand breaks (Roberts et al., 2012; Taylor et al., 2013). *Kataegis* has been observed in bladder, lung, head and neck and breast cancer, which all also show a prominent presence of the APOBEC3 mutational signatures (Burns et al., 2013b; Secrier et al., 2016), as well as a range of other cancer types (reviewed in Roberts and Gordenin, 2014a).

Overall, the APOBEC3-mediated mutational signatures are found in 16 out of 30 cancer types and contribute more than a quarter of the total mutations in 15% of all examined cancer samples (Alexandrov et al., 2013a). APOBEC3-mediated mutagenesis can thus be considered one of the most important human carcinogens, whose prevalence surpasses even that of tobacco smoke and exposure to UV light (Alexandrov and Stratton, 2014).

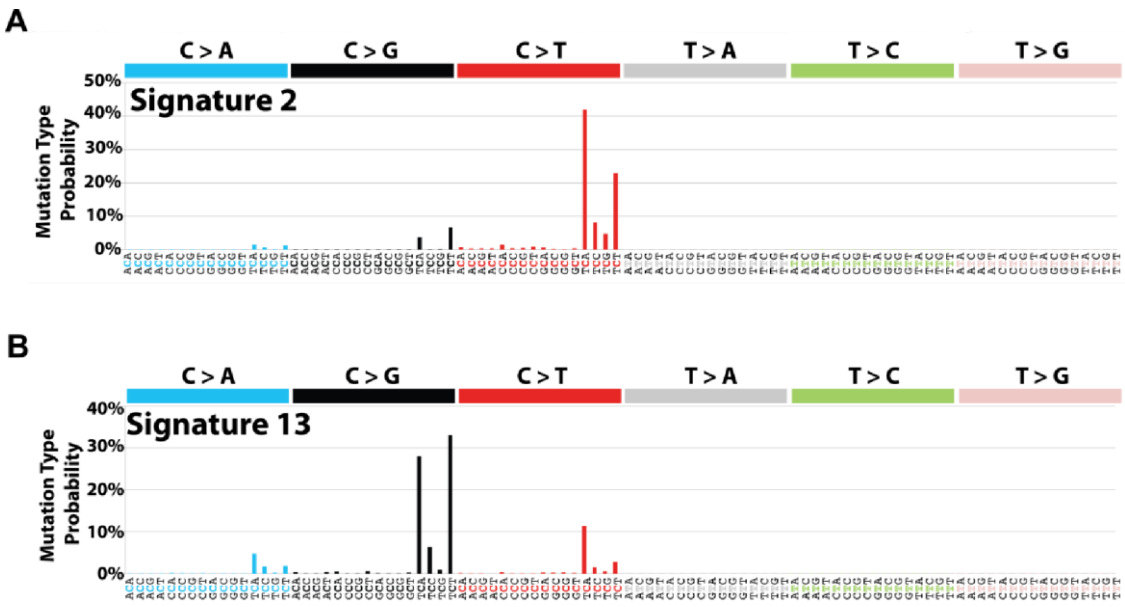

Figure 4 Mutational signatures attributed to APOBEC3-mediated mutagenesis. Each signature consists of the 96 categories classified by the single nucleotide exchange and their immediate sequence context as depicted on the horizontal axis. The bars on the vertical axis represent the probability of specific mutations found within a mutational pattern. **A** Mutational signature 2 is characterised by C-to-T mutations in a 5'-TC context. **B** Mutational signature 13 is characterised by C-to-G mutations in a 5'-TC context. Adapted from http://cancer.sanger.ac.uk/cosmic/signatures (Forbes et al., 2017) accessed on 27.9.2017.

1.3.3. Cancer types carrying APOBEC3-mediated mutations

The APOBEC3 mutational signatures 2 and 13 were originally found in 16 cancer types (Alexandrov et al., 2013a). This has since been expanded to 22 cancer types according to the Catalogue Of Somatic Mutations In Cancer (COSMIC) and Alexandrov (2015). The APOBEC3 mutational signatures are particularly prominent in bladder cancer, cervical cancer, head and neck squamous cell carcinoma, breast cancer and lung cancer (Roberts et al., 2012; Alexandrov et al., 2013a; Roberts et al., 2013; Jia et al., 2014; Lin et al., 2014a; Nordentoft et al., 2014; Roberts and Gordenin, 2014a; Zhang et al., 2015a; Hedegaard et al., 2016; Zehir et al., 2017; Zhou et al., 2017).

Cervical and head and neck cancer are of particular interest due to the interactions between human papillomavirus (HPV) and APOBEC3. APOBEC3A can restrict HPV infectivity (Vartanian et al., 2008; Warren et al., 2015b), edit HPV genomes (Wang et al., 2014; Wakae et al., 2015) and may even trigger integration of the HPV virus into the host genome (Herdman et al., 2006; Lace et al., 2015; Kondo et al., 2017). Warren et al. (2017) discussed the possibility that APOBEC3A may mediate the clearance of HPV DNA during persistent infection.

Almost all cervical cancers (Walboomers et al., 1999) as well as 45-65% of oropharyngeal cancer and 5-25% of other head and neck cancers (Stransky et al., 2011; Ndiaye et al.,

2014; The Cancer Genome Atlas Network, 2015) are caused by high-risk HPV infection. While the E6 and E7 oncoproteins of high-risk HPVs cause immortalisation of cells, they are not sufficient for transformation: HPV-driven tumourigenesis requires the acquisition of additional mutations, which accounts for the latency period between HPV infection and development of cancer (reviewed by McLaughlin-Drubin and Münger, 2009, discussed by Henderson et al., 2014). Thus, it is interesting to note that mutations matching the APOBEC3 mutational signatures are found in many head and neck cancers (Lin et al., 2014a; Zhang et al., 2015a; Fanourakis et al., 2016; Sawada et al., 2016; Secrier et al., 2016; Chang et al., 2017) as well as more than 50% of cervical cancers (Ojesina et al., 2014), and are particularly frequent in HPV-driven cases (Lawrence et al., 2013; Henderson et al., 2014, The Cancer Genome Atlas Network, 2015, The Cancer Genome Atlas Research Network, 2017 and reviewed by Litwin et al., 2017).

All APOBEC3 family members except APOBEC3B are more highly expressed in mucosal tissue, which is more vulnerable to virus infection in comparison to cutaneous skin (Warren et al., 2015a). Normal skin keratinocytes already show expression of APOBEC3A (Yang et al., 2016). An antiviral response can upregulate APOBEC3A expression in both skin keratinocytes (Wang et al., 2014; Warren et al., 2015b) and cancer cell lines (Middlebrooks et al., 2016). Expression of both APOBEC3A and APOBEC3B is furthermore induced by HPV E6 and E7, and it is higher in HPV-positive cancers than in HPV-negative ones (Vieira et al., 2014; Warren et al., 2015b; Kondo et al., 2017). Additionally, Warren et al. (2015b) found that expression of both APOBEC3A and B is upregulated in early stages of HPV-driven cervical neoplasia. This upregulation is no longer significant in fully formed cervical cancers.

1.4. APOBEC3A and APOBEC3B can shape the genomic landscape of cancer

APOBEC3A and APOBEC3B share the characteristics that allow them to create the mutational signatures attributed to APOBEC3-mediated deamination. Both have access to the nucleus and thus to genomic DNA (Chen et al., 2006; Muckenfuss et al., 2006; Kinomoto et al., 2007; Bulliard et al., 2011; Landry et al., 2011; Aynaud et al., 2012; Lackey et al., 2012; Burns et al., 2013a; Lackey et al., 2013; Mussil et al., 2013; Lucifora et al., 2014). Both can introduce C-to-T mutations in genomic DNA (Shinohara et al., 2012; Burns et al., 2013a). Both prefer to deaminate cytidine in a 5'-TC context (Bishop et al., 2004; Chen et al., 2006; Thielen et al., 2010; Burns et al., 2013a; Byeon et al., 2013; Taylor et al., 2013; Mitra et al., 2014; Shi et al., 2016; Kouno et al., 2017). They show different expression patterns in various tissues, but their expression is generally low (Koning et al., 2009; Refsland et al., 2010; Refsland and Harris, 2013), yet overexpression

has been observed in cancer (Lin et al., 2014a; Boichard et al., 2017; Chen et al., 2017). Both may contribute to the deaminase activity observed in cancer cells (Buisson et al., 2017). Opinions are divided about whether APOBEC3A (Roberts et al., 2013; Caval et al., 2014a, 2014a; Chan et al., 2015; Faltas et al., 2016) or APOBEC3B (Burns et al., 2013b; Burns et al., 2013a; Nikkilä et al., 2017) is the main source of mutations forming the APOBEC3 mutational signatures.

APOBEC3B alone is not sufficient to explain the APOBEC3 mutational signatures (discussed by Nik-Zainal et al., 2014, Akre et al., 2016 and Radmanesh et al., 2017). Similar to APOBEC3B, APOBEC3A expression is induced by HPV E7 (Warren et al., 2015b) and found to be overexpressed in various cancers and cancer cell lines (Hedegaard et al., 2016; Boichard et al., 2017; Buisson et al., 2017; Chang et al., 2017; Chen et al., 2017; Green et al., 2017). APOBEC3A expression has also been correlated with the presence of the APOBEC3 mutational signatures (Jia et al., 2014; Hedegaard et al., 2016). In contrast to the findings arguing for APOBEC3B as the cause of the APOBEC3 mutational signatures, the absence of APOBEC3B in a deletion polymorphism has been associated with an increased risk of breast cancer (Komatsu et al., 2008). A similar association has been found for liver cancer (Zhang et al., 2013) and oesophageal cancer (Chen et al., 2017). Further studies had some conflicting results: while this association with an increased breast cancer risk has been confirmed in a Chinese and European population (Long et al., 2013; Xuan et al., 2013), others did not find an association in a Swedish, South Indian and an Eastern European population (Göhler et al., 2015; Revathidevi et al., 2016; Klonowska et al., 2017). Meta-analyses of several studies confirmed the overall association with breast cancer in Asian, but not Caucasian populations (Han et al., 2016; Klonowska et al., 2017). It has been suggested that the discordances between the studies may be due to differences in genetic background (Klonowska et al., 2017).

The APOBEC3 polymorphism deletes the sequence between the fifth exon of APOBEC3A and the eighth exon of APOBEC3B. This results in a complete loss of the coding sequence for APOBEC3B and a chimeric construct consisting of the APOBEC3A coding sequence with the 3'-untranslated region (UTR) of APOBEC3B (Kidd et al., 2007). The protein sequence encoded by this chimeric construct is identical to APOBEC3A. Heterozygous carriers of the deletion polymorphism show reduced APOBEC3B expression, and the APOBEC3B transcript cannot be detected in carriers of homozygous deletions (Komatsu et al., 2008; Zhang et al., 2013; Nik-Zainal et al., 2014; Cescon et al., 2015; Zhang et al., 2015b; Chen et al., 2017; Klonowska et al., 2017). While some studies found an increase in APOBEC3A mRNA levels in carriers of the polymorphism (Caval et al., 2014a; Chen et al., 2017), others found the APOBEC3A transcript levels unchanged (Cescon et al., 2015). As the 3'-UTR of APOBEC3A is replaced with the 3'-UTR of APOBEC3B in the

polymorphism, the change may be in protein rather than transcript levels due to differential regulation of translation (reviewed by Matoulkova et al., 2012). Caval et al. (2014a) showed that the chimeric construct resulted in an increase of APOBEC3A protein expression, as well as an increase in DNA damage.

No difference is apparent in the frequency of C-to-T or C-to-G mutations in patients with two wild-type APOBEC3B genes and heterozygous or homozygous deletion of APOBEC3B (Zhang et al., 2015b). In fact, the number of C-to-T and C-to-G mutations in the 5'-TC context preferred by APOBEC3 is increased in heterozygous carriers and even higher in homozygous carriers of the APOBEC3B deletion polymorphism (Zhang et al., 2015b). Furthermore, the APOBEC3B deletion allele was associated with a higher mutational burden of APOBEC3-attributed mutational signatures 2 and 13, and with a subset of hypermutator breast cancers (Nik-Zainal et al., 2014). Despite the inconsistencies between the studies regarding an increase in cancer risk, the fact that the mutational signatures 2 and 13 are not just present but increased in carriers of homozygous deletions of APOBEC3B shows that APOBEC3B cannot be solely responsible for the mutations attributed to APOBEC3 (discussed by Caval et al., 2014a; Nik-Zainal et al., 2014). Instead, the results point to the possibility that APOBEC3A may be a major mutagenic force shaping cancer genomes.

There are additional findings which hint that APOBEC3A may have greater relevance than APOBEC3B in cancer. The sequence context of single nucleotide exchanges in the mutational signature analysis only considers the nucleotides directly adjacent to the mutated base. Extending the sequence context of APOBEC3-mediated mutations shows that APOBEC3A prefers a YTCA sequence context, whereas APOBEC3B prefers RTCA sites (with Y being a pyrimidine base and R being a purine base) (Chan et al., 2015). This means that APOBEC3A-like and APOBEC3B-like mutations can be distinguished. Cancers with a high contribution of APOBEC3-type mutations (bladder, breast, lung and head and neck cancer) were subcategorised into APOBEC3A- and APOBEC3B-like subgroups depending on the enrichment of YTCA or RTCA sites, respectively. Cancers with an APOBEC3A-like mutational pattern were found to carry more than tenfold the number of mutations found in APOBEC3B-like cancers. Furthermore, cancers with the APOBEC3B deletion polymorphism also show an APOBEC3A-like mutational pattern. The APOBEC3A-like mutational pattern was furthermore found to contribute more than APOBEC3B-like mutations to the mutational landscape of urothelial carcinoma (Faltas et al., 2016; Lamy et al., 2016) and oesophageal squamous cell carcinoma (Chen et al., 2017). Chan et al. (2015) suggest that in APOBEC3B-like cancers, the low mutational activity of APOBEC3B and low or no activity of APOBEC3A results in a low APOBEC3 mutational load with an APOBEC3B-like pattern. In contrast, the background APOBEC3B-like mutations are

dwarfed by the high mutagenic activity of APOBEC3A in APOBEC3A-like cancers, resulting in a high mutational load of APOBEC3A-like mutations. This is consistent with the observation that APOBEC3A deaminates cytidine more efficiently than APOBEC3B. APOBEC3A outperforms both APOBEC3B (Caval et al., 2014a; Chan et al., 2015; Lamy et al., 2016; King and Larijani, 2017) and APOBEC3G (Carpenter et al., 2012; Shlyakhtenko et al., 2016), and shows the highest deaminase activity for both cytidine and methyl-cytidine (Ito et al., 2017).

Finally, Akre et al. (2016) found that overexpression of transgenic APOBEC3B in HEK293 cells created mutations in the genomic DNA and increased the mutational load, but did not observe the mutational signatures 2 and 13 attributed to APOBEC3 in these cells. In summary, APOBEC3A, possibly supported by APOBEC3B, can contribute a substantial amount of mutations to the total mutational load of various cancer types. This raises the question whether APOBEC3-mediated mutations functionally affect the tumours that carry them.

1.5. APOBEC3A – a double-edged sword

APOBEC3A can be considered a double-edged sword. On the one hand, it is involved in innate immunity to viral infection. On the other hand, it can contribute to cancer mutagenesis. These contrasting effects of APOBEC3A also extend within the context of cancer: APOBEC3A expression has deleterious acute effects, but the potential long-term effects may cause or influence cancer progression.

1.5.1. Acute effects of APOBEC3A expression

Overexpression of transgenic APOBEC3A has a deleterious and genotoxic acute effect on cells. It causes cell cycle arrest, apoptosis and dsDNA breaks (Landry et al., 2011; Burns et al., 2013a; Lackey et al., 2013; Mussil et al., 2013; Green et al., 2016). The deleterious acute effect of APOBEC3A expression has even been shown in mouse tumours, where transgenic expression of APOBEC3A reduces tumour volume and causes necrosis (Kostrzak et al., 2016). These negative effects may build selection pressure on cells with high APOBEC3A activity.

1.5.2. Potential long-term effects of APOBEC3A-mediated

APOBEC3-mediated mutagenesis may also have a more long-term effect in cells that survive the acute deleterious effects. Other members of the AID/APOBEC family have been shown to be directly involved in carcinogenesis. Transgenic mice which overexpress either APOBEC1 or AID were found to develop tumours (Yamanaka et al., 1995; Okazaki et al., 2003), with the mutations found in genes commonly mutated in

cancer matching their typical deaminase activity and target sequence (Beale et al., 2004). APOBEC3 signature mutations have been observed to affect cancer driver genes and tumour suppressor genes (Table 1) (Roberts et al., 2013; Tornesello et al., 2014; Li et al., 2017; Litwin et al., 2017), and a high APOBEC3 mutational load is associated with mutations in various cancer driver and tumour suppressor genes (Kanu et al., 2016).

Table 1 Selection of cancer driver and tumour suppressor genes carrying APOBEC3-mediated mutations

Mutated gene	Function	Reference
PIK3CA	Cancer driver	Kang et al. (2005), Bader et al. (2006), Nichols et al. (2013), Roberts et al. (2013), Henderson et al. (2014), Tornesello et al. (2014), Zhang et al. (2015a), McGranahan and Swanton (2015), Litwin et al. (2017)
TP53	Tumour supressor	Roberts et al. (2013), Tornesello et al. (2014), McGranahan and Swanton (2015), Kanu et al. (2016)
PTEN	Cancer driver	Roberts et al. (2013), Kanu et al. (2016), McGranahan and Swanton (2015), Litwin et al. (2017)
BRAF	Cancer driver	Roberts et al. (2013)
ATM	Tumour suppressor	Roberts et al. (2013)
APC	Cancer driver	Roberts et al. (2013)
NF1	Tumour suppressor	Roberts et al. (2013), Kanu et al. (2016)
KRAS	Cancer driver	Litwin et al. (2017)

The majority of mutations in the most frequently mutated cancer drivers in cervical cancer were found to correspond to the APOBEC3 mutational signatures (analysis by Litwin et al., 2017 based on data from The Cancer Genome Atlas Research Network, 2017). In particular, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA) is the most frequently mutated oncogene in HPV-driven cancers (reviewed by Litwin et al., 2017). PIK3CA contains mutation hotspots with the ability to induce oncogenic transformation in its helical and its kinase domain (Kang et al., 2005; Bader et al., 2006). Helical domain mutations are particularly frequent in HPV-positive cervical (Ojesina et al., 2014; The Cancer Genome Atlas Research Network, 2017) and head and neck cancers (Nichols et al., 2013; The Cancer Genome Atlas Network, 2015). Mutations in the helical domain have been linked to APOBEC3-mediated mutagenesis (Henderson et al., 2014). They occur more frequently in cancers with a high number of mutations matching the APOBEC3 mutational signature (Henderson et al., 2014; Zhang et al., 2015a; Kanu et al., 2016; Sawada et al., 2016) and are more frequent in HPV-positive than HPV-negative cancers (Nichols et al., 2013; Henderson et al., 2014; Zhang et al., 2016, reviewed by Litwin et al., 2017).

APOBEC3-mediated mutations of cancer driver and tumour suppressor genes may give the mutagenised cells an advantage and help them survive the negative acute effects of

APOBEC3 expression. Ultimately, APOBEC3-mediated hypermutation of genomic DNA may thus contribute to carcinogenesis, especially if it is transient. The acute negative effects disappear with APOBEC3 expression, but the mutations it generated remain (reviewed in Roberts and Gordenin, 2014a, discussed by Middlebrooks et al., 2016).

2. Aims

We hypothesized that cells which received mutations conferring a selective advantage such as increased proliferation could survive under the selective pressure created by the negative effects of APOBEC3A, while cells with neutral or detrimental mutations would be selected against. Cancer development is caused by mutagenesis and selection in an evolutionary process (reviewed by Stratton et al., 2009), so the combination of mutagenesis and selection caused by APOBEC3A could influence tumour growth. Thus, we asked whether APOBEC3A-mediated mutations can alter tumour growth of already immortalised, transformed cells.

The aim of this thesis was to study the effect of APOBEC3A-mediated mutagenesis on tumour growth of already immortalised, transformed cells. This required genetically defined cell lines that carry the APOBEC3 mutational signatures, but were stable concerning further APOBEC3-mediated mutagenesis to prevent moving targets. In order to find out whether APOBEC3A-mediated mutagenesis had an effect on tumour growth, cancer-relevant phenotypes of the APOBEC3A-mutagenised cells were assessed both *in vitro* and *in vivo*. Specifically, we asked whether the behaviour of APOBEC3A-mutagenised cells was altered concerning proliferation, migration and cisplatin sensitivity in cell culture, and concerning proliferation, necrosis, tumour growth and survival time in a xenograft mouse model. Finally, it was addressed whether an integration of genomic data and gene expression levels reflecting past and present APOBEC3-mediated mutagenesis in head and neck squamous cell carcinomas can be applied as a novel way to identify patient subgroups with prognostic value.

3. Material & Methods

3.1. Material

All material used in this study is listed alphabetically under the relevant heading, together with the supplier it was acquired from. In addition, the recipes of all buffers used here are listed.

3.1.1. Equipment

Bio Gard Hood	The Baker Company
Bioruptor Sonicator	Diagenode
Bolt™ Mini Gel Tank	Novex
Bolt™ Mini Blot Module	Novex
Centrifuge 5417R	Eppendorf
Electrophoresis chamber large	Peqlab
Electrophoresis chamber small	Peqlab
Electrophoresis power supply	Bio-Rad
Electrophoresis power supply PEQPower	Peqlab
Gel Doc EZ Imager	Bio-Rad
Heater/magnetic stirrer	Fisher Scientific
Intas ECL chemical imager	Intas
Lightcycler 480	Roche
Microcentrifuge	Eppendorf
Microplate reader Multiskan Go	Thermo Scientific
Mini Spin plus centrifuge	Eppendorf
Minifuge RF	Heraeus SEPATECH
Nanodrop Spectrophotometer	Peqlab
Nikon Eclipse Ti fluorescence microscope	Nikon
Precellys 24 tissue homogeniser	Bertin
Rotary microtome Microm HM 355 S	Thermo Scientific
Scales	Kern & Sohn GmbH
Steam cooker Vitacuisine	Tefal
Thermo Forma series II water jacketed CO_2 incubator	Thermo Scientific
Thermomixer 5436	Eppendorf
Thermomixer 5437	Eppendorf
UItra Cruz Mini centrifuge	Santa Cruz Biotechnology
Vortex-Genie 2	Scientific Industries, lnc.
Water bath	GFL

Widefield Olympus IX81 Cell^R	Olympus
Widefield motorized inverted Cell Observer.Z1	Zeiss
Xcell sure lock electrophoresis cell	Invitrogen

3.1.2. Kits, reagents and consumables

Commercially available kits

AllPrep DNA/RNA/Protein Mini Kit	QIAgen
Cell proliferation reagent WST -1	Roche
DC protein assay kit	Bio-Rad
ImmPRESS™ HRP Anti-Rabbit IgG (Peroxidase) Polymer Detection Kit	Vector Laboratories Inc.
LightCycler ® 480 Probes Master Kit	Roche
QIAquick Gel Extraction Kit	QIAgen
QIAshredder columns	QIAgen
QuantiTect Reverse Transcription Kit	QIAgen
RNeasy Mini Kit	QIAgen

Chemicals

1,4-Dithiothreitol (DTT)	Roth
2 - Mercaptoethanol	Sigma-Aldrich
Acetic acid	Roth
Agarose basic	AppliChem GmbH
Amersham ECL Select Western Blotting Detection Reagent	GE Healthcare
Amersham ECL Western Blotting Detection Reagents	GE Healthcare
Ammonium persulphate (APS)	Sigma-Aldrich
Boric acid	J.T. Baker
BSA - Type H1	Gerbu
Cisplatin	Sigma-Aldrich
Coomassie GelCode™ Blue Stain Reagent	Thermo Scientific
DAB substrate kit for peroxidase	Vector Laboratories Inc.
Dako fluorescent mounting medium	Dako
Dimethyl sulphoxide (DMSO)	Roth
Eosin	Roth
Ethanol	Sigma-Aldrich
Ethidium bromide	Sigma-Aldrich

Ethylenediaminetetraacetic acid (EDTA)	Roth
Formaldehyde solution (37%)	Merck
Formamide	Fluka
Glycerol	Roth
HEPES	Eurobio
Hoechst 33342	Sigma-Aldrich
Hygromycin B	Roche
Isopropanol	Fisher Scientific
Mayer's Haemalaun solution	AppliChem GmbH
Methanol	Sigma-Aldrich
Methylene blue·HCl	Serva
Milk powder	Roth
Mitomycin C	Sigma-Aldrich
N,N,N',N'-Tetramethylethylenediamine (TEMED)	Sigma-Aldrich
NaOH	VWR
Paraformaldehyde	Roth
poly-L-lysine (0.01%)	Sigma-Aldrich
Rotiphorese Gel 30 (37.5:1) acrylamide/bis-acrylamide	Roth
Sodium dodecyl sulphate (SDS)	Gerbu
Tris base	Sigma-Aldrich
Tri-sodium citrate 2-hydrate ($C_6H_5Na_3O_72H_20$)	Merck
Triton X-100	Roth
Trypan blue	Merck
Tween-20	Gerbu
Urea	Roth
Xylene cyanol	Sigma-Aldrich
Xylene	Fischer Scientific

Consumables

Amersham Hyperfilm ECL	GE Healthcare
Coverslips	Menzel
DAKO pen	Dako
Distilled water, DNase/RNase free	Gibco
Dulbecco's Modified Eagle's Medium	Sigma-Aldrich
Eukitt	Kindler GmbH
Foetal calf serum (FCS)	Biochrom AG

Gel cassettes	Novex
LightCycler ® 480 Multiwell plate 96 well, white	Roche
Nalgene Cryoware cryogenic vials	Thermo Scientific
Nitrocellulose membrane	Thermo Scientific
NuPAGE 4-12% Bis-Tris Gel 12 well	Novex
NuPAGE 4-12% Bis-Tris Gel 15 well	Novex
Penicillin Streptomycin solution	Gibco
Precellys Soft Tissue Homogenizing Ceramic Beads	Bertin Technologies
Protease Inhibitor Cocktail	Roche
Superfrost Plus microscope slides	Thermo Scientific
Tissue culture dish 100	Sarstedt
Tissue culture flask 25	TPP
Tissue culture flask 75	TPP
Tissue culture flask 100	TPP
Tissue culture plate 6 wells	Corning Incorporated
Tissue culture plate 12 wells	Corning incorporated
Tissue culture plate 24 wells	Falcon
Tissue culture plate 96 wells	Falcon

Buffers and solutions

1x PBS	137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4 in mqH_2O, pH 7.4
1x TAE buffer	40 mM Tris (pH 7.6), 20 mM acetic acid, 1 mM EDTA in mqH_2O
6x DNA loading dye	10 mM Tris-HCl (pH 7.6), 0.03% bromophenol blue, 0.03% xylene cyanol, 60% glycerol, 60 mM EDTA
Trypan blue	0.5 g trypan blue, 0.9 g NaCl in 100 ml mqH_2O

Buffers for immunofluorescence:

Fixation solution	4% (w/v) paraformaldehyde in PBS
Permeabilisation solution	0.5% (v/v) Triton X-100 in PBS
Blocking buffer	1% (w/v) BSA in PBS

Buffers for immunohistochemistry:

Antigen retrieval buffer	10 mM tri-sodium citrate 2-hydrate in mqH_2O, adjusted to pH 6.0 with citric acid

Buffers for deamination assay:

1x HED buffer	25 mM HEPES, 5 mM EDTA, 10% (v/v) glycerol, 1 mM DTT, 4% (v/v) protease inhibitor cocktail, pH 7.8
10x TBE buffer	89 mM Tris, 89 mM boric acid, 2 mM EDTA
2x formamide buffer	80% (v/v) formamide, 0.05% (w/v) bromophenol blue, 0.01% (w/v) xylene blue in 1x TBE buffer

Buffers for SDS-PAGE and Western Blot:

Lysis buffer	40 µl protease inhibitor cocktail in 960 µl 1% (v/v) SDS
5x lane marker reducing sample buffer	Ready to use by Thermo Scientific
Running buffer	1x NuPAGE™ MOPS SDS running in mqH_2O
Transfer buffer	1x NuPAGE™ transfer buffer, 20% (v/v) methanol in mqH_2O
Coomassie GelCode™ Blue Stain Reagent	Ready to use (Thermo Scientific)
Blocking buffer	4% (w/v) skim milk powder, 0.05% Tween-20 in PBS
PBS-T	0.05% (v/v) Tween-20 in PBS
Stripping buffer	67.5 mM Tris-HCl, 2% (w/v) SDS, 100 mM β-mercaptoethanol in PBS, pH 6.8

DNA and protein size markers

GeneRuler 1 kb DNA Ladder	Thermo Scientific
GeneRuler 100 bp DNA Ladder	Fermentas
Page Ruler Prestained Protein Ladder	Thermo Scientific

Antibodies

The anti-APOBEC3G antibody listed below is cross-reactive with both APOBEC3A and APOBEC3B.

Rabbit anti-human APOBEC3G (polyclonal)	Sigma-Aldrich
Rabbit anti-Ki-67 (monoclonal)	Thermo Scientific
Mouse anti-ACTB, HRP-conjugated (monoclonal)	Abcam
Goat anti-rabbit Alexa Fluor 594	Thermo Scientific
Goat anti-rabbit, HRP-conjugated	Sigma-Aldrich

Enzymes

Accutase	Sigma-Aldrich
Phusion High Fidelity DNA Polymerase	New England Biolabs

RNase A	QIAgen
Trypsin-EDTA	Sigma-Aldrich
Uracil-DNA glycocylase	New England Biolabs

Oligonucleotides

The standard primers CMV forward and BGH reverse are used for amplification of the APOBEC3A and APOBEC3A$_{E72A}$ transgenes in polymerase chain reactions (PCR). The same primers are used to sequence the resulting PCR products.

CMV forward	5'-CGC AAA TGG GCG GTA GGC GTG-3'
BGH reverse	5'-TAG AAG GCA CAG TCG AGG-3'

The gene-specific qPCR primers listed below are used to specifically amplify the target genes in order to quantify their transcript levels in a qPCR. The separate primer pairs for each target gene consist of one left (L) or forward primer and one right (R) or reverse primer.

APOBEC3A-L3	5'-GAG AAG GGA CAA GCA C-3'
APOBEC3A-R3	5'-GTG TGG ATC CAT CAA AAG TG-3'
APOBEC3B-L6	5'-GAC CCT TTG GTC CTT CGA C-3'
APOBEC3B-R6	5'-GCA CAG CCC AGG AGA AG-3'
ACTB-L1	5'-ATT GGC AAT GAG CGG TTC-3'
ACTB-R1	5'-CGT GGA TGC CAC AGG ACT-3'

The following probes are used to detect the target transcripts in qPCR. The APOBEC3A and APOBEC3B probes are labelled with Hex, the ACTB probe is labelled with Fam.

APOBEC3A-Hex probe	5'-HEX-GAA GCC AGC CCA GCA TCC-BBQ-3'
APOBEC3B-Fam probe	Universal Probe Library #39 (Roche)
ACTB-Fam probe	Universal Probe Library #11 (Roche)

3.1.3. Software and internet resources

The software programs and internet resources listed in Table 2 are used for data processing, statistical analyses, control of equipment and image acquisition as well as image processing.

Table 2 Overview of software programs and internet resources used

Software or webtool	Source
Citavi for Windows, version 5.7.0.0	Swiss Academic Software
Clustal Omega	http://www.ebi.ac.uk/Tools/msa/clustalo/
ClustVis webtool	http://biit.cs.ut.ee/clustvis/ (Metsalu and Vilo, 2015)
GIMP version 2.8.16	https://www.gimp.org/
GraphPad Prism for Windows, version 5.04	GraphPad Software, San Diego California USA, https://www.graphpad.com/
ImageJ version 1.50b	http://rsbweb.nih.gov/ij/ (Schneider et al., 2012)
Inkscape version 0.92	https://inkscape.org
Intas ChemoStar Software	Intas
LightCycler Software 480 release 1.5.0.SP1	Roche Diagnostics
NanoDrop version 3.6.0	NanoDrop Technologies
NIS-Elements AR version 4.40.00	Nikon
SkanIt Software 3.2 for Multiskan GO	Thermo Fisher Scientific
Xcellence rt version 2.0	Olympus
Zeiss ZEN blue	Zeiss

3.2. Methods

3.2.1. Cell Culture

All cells used in this study are Flp-In™ T-Rex™ HEK293 cells (Life Technologies). They carry a stably integrated FRT site followed by a lacZ/zeocin resistance fusion gene as well as a tetracyclin-repressor gene. All cell lines used here were originally created and characterised by Ann-Mareen Franke. She created cell lines with stable integrations of APOBEC3 transgenes under transcription control of a tet-on regulator. The cell population containing the APOBEC3A transgene already contains the APOBEC3 mutational signatures.

All cell lines used in this study are frequently confirmed to be free of contaminations including mycoplasma infection, and their identity is authenticated by Multiplexion GmbH (Heidelberg).

3.2.1.1. Culturing cells

All cells in this study are cultured in Dulbecco's Modified Eagle's Medium (DMEM) with low glucose at a concentration of 1000 mg/l supplemented with L-glutamine, sodium pyruvate and sodium bicarbonate containing 10% (v/v) foetal calf serum (FCS) and 1% (v/v) Penicillin-Streptomycin solution. As transgene expression in the HEK293 cells is doxycycline-regulated, certified tetracycline-free FCS is used for their culture. The transgene-containing HEK293 cells furthermore carry a gene for hygromycin B resistance, so they can also be cultivated in standard medium containing 200 µg/ml hygromycin B. All cells are incubated in cell culture incubators at 37°C in a humidified atmosphere containing 5% CO_2. An exponentially growing asynchronous cell population is passaged by removing the medium, washing the cells with PBS and detaching them from the cell culture vessel by incubation with accutase or trypsin/EDTA. The detached cells are washed off the cell culture vessel with an appropriate volume of pre-warmed medium, and pelleted by centrifugation at 300 rcf for 5 minutes at room temperature. The pelleted cells are then resuspended in an appropriate volume of pre-warmed medium. The cell count is determined by mixing cell suspension with trypan blue (0.5 g trypan blue and 0.9 g NaCl per 100 ml mqH_2O), and counting the unstained cells in a Neubauer chamber at 100x magnification under a light microscope. The cells are plated in a fresh cell culture vessel at a density of $1.3x10^4$ cells/cm^2 for maintenance culture, if not stated differently.

3.2.1.2. Cryo-conservation of cells

For long-term storage of cryo-conserved cells, an exponentially growing asynchronous cell population is harvested, centrifuged at 300 rcf for 5 minutes at room temperature, resuspended in freezing medium (FCS + 10% (v/v) DMSO) and transferred into cryotubes. A total of $1x10^6$ cells suspended in 1 ml freezing medium is added to each cryotube. The cells are cooled down at a rate of -1°C per minute, either by placing the cryotubes into a freezing box filled with isopropanol or by wrapping them in several layers of paper tissues and placing them in the -80°C freezer for several days. The cells are finally placed in the vapour phase of liquid nitrogen for long-term storage.

3.2.1.3. Reviving cells

Cryo-conserved cells are revived by warming them until they are almost completely thawed, and then transferring the thawed cells into 10 ml pre-warmed medium. The cells are pelleted by centrifugation at room temperature for 5 minutes at 300 rcf. The supernatant is discarded, the cells are resuspended in pre-warmed medium and plated at a cell density of $1.3x10^4$ cells/cm^2. One day after reviving the cells, the medium is replaced with fresh pre-warmed medium.

3.2.1.4. Long-term culture

The long-term culture exposes three independent cell populations called APOBEC3A pop. 01, APOBEC3A pop. 12 and APOBEC3A pop. 20 to both basal and doxycycline (dox)-induced APOBEC3A transgene expression over a period of 11 weeks. A cell population called $APOBEC3A_{E72A}$ pop. carrying a transgene of the enzymatically inactive $APOBEC3A_{E72A}$ variant as well as a transgene-free population of parental cells called HEK293 pop. are included as controls. At the start of the long-term culture, each cell population is split into one cell line that is kept without doxycycline and one cell line that is induced with doxycycline during the course of the long-term culture. The cells are plated at a density of $1.7x10^4$ cells/cm^2 in 6-well plates. On the day after plating the cells, doxycycline is added to a final concentration of 1 µg/ml to the cell lines to be kept under induction. The cells are then incubated in cell culture incubators at 37°C in a humidified atmosphere containing 5% CO_2 for 3 days. At every fourth passage, all cell lines are expanded in two 10 cm cell culture dishes each in addition to the 6-well plates. To the cell lines kept under doxycycline induction, a final concentration of 1 µg/ml doxycycline is added to the medium three days before the cells are harvested for further processing. One of these plates is harvested for extraction of total protein, genomic DNA and total RNA, the cells on the other plate are used for cryo-conservation.

3.2.1.5. Isolating and expanding single cell clones

Single cells are isolated from a cell population by plating the cells of the population in a 96-well plate at a cell density so that a single cell is on average contained in one well. This is achieved by preparing a cell suspension with a concentration of 640 cells/ml, which is then diluted in a 1:2 dilution series with the final dilution being 1:512 (or 1.25 cells/ml). Of each dilution, 100 µl/well are plated in eight wells of a 96-well plate. An average of one cell per well is expected at a dilution of 1:64 (or a cell concentration of 10 cells/ml). The cells are incubated in a cell culture incubator at 37°C in a humidified atmosphere containing 5% CO_2 for 6 days. At regular intervals starting from day 3, the plates are visually inspected under a light microscope at 100x magnification in order to identify wells containing one cell colony derived from a single cell. Three single clones are isolated from each population, with only two clones each isolated from the parental HEK293 and APOBEC3A$_{E72A}$ populations that had been treated with 1 µg/ml doxycycline, and the APOBEC3A$_{E72A}$ population kept without doxycycline over 11 weeks. The clones are expanded by passaging them into successively larger cell culture vessels, up to T100 cell culture flasks, for characterisation and cryoconservation.

3.2.2. Characterisation of HEK293 clones

The HEK293 cell populations are characterised for expression of the APOBEC3A and APOBEC3A$_{E72A}$ transgenes on a single cell level by immunofluorescence. The HEK293 cell clones are characterised regarding proliferation, migration and cisplatin sensitivity. Proliferation is determined by staining the cells for the proliferation marker Ki-67 using immunofluorescence, and determining the proliferative index. Cell migration speed is quantified in a scratch assay. Cisplatin sensitivity is determined in a viability assay of cells after exposing them to a cisplatin concentration gradient.

3.2.2.1. Immunofluorescence

For immunofluorescence stains, cells are grown on poly-lysine coated coverslips. In order to coat the coverslips, they are covered in 0.01% poly-lysine solution and incubated at room temperature for 5 minutes. Then, the liquid is removed and the coverslips are dried before they are placed in a 12-well plate with the poly-lysine coated side facing upwards. Per well, a total of $1x10^5$ cells are plated in a total volume of 2 ml medium per well (cell density $2.5x10^4$ cells/cm^2), and incubated for 24 hours in a cell culture incubator at 37°C in a humidified atmosphere containing 5% CO_2. After allowing the cells to grow for one day, the medium is aspirated, and the cells are fixed by addition of 0.5 ml 4% (w/v) paraformaldehyde in PBS per well. After 20 minutes incubation at room temperature, the fixation solution is carefully aspirated, and the coverslips are

washed in PBS for 3x5 minutes. The cells are permeabilised in 0.5 ml 0.5% (v/v) Triton X-100 in PBS for 15 minutes at room temperature, followed by 30 minutes incubation in 1 ml blocking buffer (1% (w/v) BSA in PBS) at 37°C. Following the blocking step, the cover slips are placed with the cell-bearing side facing down on 50 µl drops of primary antibody diluted in blocking buffer spotted on parafilm in a wet chamber. In order to determine the proliferative index, the anti-Ki-67 antibody is used at a 1:300 dilution. The anti-human APOBEC3G antibody is used at a 1:100 dilution to detect APOBEC3A transgene expression. The primary antibody is incubated either at room temperature for 1 hour, or overnight at 4°C. After the incubation, the coverslips are washed in PBS for 3x5 minutes at room temperature, and then placed with the cell-bearing side facing down on 50 µl drops of blocking buffer containing 1:5000 diluted Hoechst 33342 stain as well as 1:2000 diluted secondary antibody goat anti-rabbit Alexa Fluor 594. The secondary antibody and Hoechst 33342 stain are incubated for 1 hour under the exclusion of light at room temperature. After this final incubation, the cover slides are again washed in PBS for 3x5 minutes, and finally mounted on glass slides in anti-fade mounting medium. The fluorescence is recorded using the TRITC and DAPI filters of a Nikon Eclipse Ti microscope at 100x or 200x magnification. The different fluorescence channels are then combined into an overlay image using GIMP 2.8.16 (GNU Image Manipulation Program).

3.2.2.2. Ki-67 proliferative index

Ki-67 is a marker for cell proliferation. It is used to determine the fraction of proliferating cells in a population, called the proliferative index. Ki-67 positive cells are detected by a monoclonal anti-Ki-67 antibody in both immunofluorescence stains of cell culture samples and immunohistochemistry stains of tumour sections. This assay is used to determine the proliferative index of the HEK293 clones *in vitro*.

In cell culture, the proliferative index is determined using immunofluorescence (as described in chapter 3.2.2.1). The cells are stained with monoclonal rabbit anti-Ki-67 IgG as primary and anti-rabbit IgG AlexaFluor 594 conjugated as secondary antibody, as well as Hoechst 33342 fluorescent DNA stains as a counterstain. The results are imaged using the DAPI and TRITC filters of the Nikon Eclipse Ti microscope at 100x magnification. At least three different fields of vision at 100x magnification are analysed per sample. The proliferative index is determined by counting the number of Ki-67 positive cells and dividing it by the total number of living cells for each field of vision.

3.2.2.3. Scratch migration assay

The migration speed of the HEK293 clones is determined in a live-cell imaging assay that tracks the closure of a scratch created in an otherwise confluent layer of cells. Two variations of this assay are used. One is employed to characterise all HEK293 clones for migration only, and an optimised version is used to determine both migration speed and the scratch closure speed with the combination of migration and proliferation in the six selected HEK293 clones. For the characterisation of migration speed in all HEK293 clones, a total of $1x10^6$ cells per well (cell density $8.9x10^5$ cells/cm^2) are plated in a 12-well plate in a minimum of 3 ml medium. The cells are allowed to grow confluent for 1 day in a cell culture incubator at 37°C in a humidified atmosphere containing 5% CO_2. The method is optimised for the quantification of scratch closure speed using migration alone or a combination of migration and proliferation in the six selected HEK293 clones. In this case, a total of $4x10^5$ cells per well (cell density $3.6x10^5$ cells/cm^2) are plated in a 12-well plate in a minimum of 3 ml medium. The cells are allowed to grow confluent for 3 days in a cell culture incubator at 37°C in a humidified atmosphere containing 5% CO_2.

Four hours before inflicting the scratch and starting the life-cell imaging, the supernatant is aspirated and gently replaced with fresh medium either with or without a final concentration of 5 µg/ml mitomycin C. Mitomycin C inhibits proliferation, and is added in order to exclude the effect of cell proliferation on scratch closure and measure only the effect of migration. In cases where the combined effect of proliferation and migration on scratch closure is determined, the cells are cultivated in mitomycin C-free medium. After the medium change, the cells are incubated in a cell culture incubator at 37°C in a humidified atmosphere containing 5% CO_2 for another four hours. Then, a scratch is created in the confluent cell layer using a p10 pipette tip. Scratch closure is imaged in time-lapse microscopy of live cells using the widefield Olympus IX81 Cell^R microscope. At least three positions along the scratch are imaged per sample, and a picture at 100x magnification is taken every 10 minutes over a period of 22.5 hours. Care is taken that the imaged positions are not in the immediate neighbourhood of the end of a scratch. During the time-lapse imaging, the cells are kept at 37°C in a humidified atmosphere containing 5% CO_2. The area of the scratch is quantified using ImageJ 1.50b (Schneider et al., 2012). The difference between the scratch area A_0 at the beginning of the time-lapse and the scratch area A_t at time point t describes the area ΔA_t covered by cells within t minutes:

$$\Delta A_t = A_0 - A_t$$

The area ΔA_t covered by cell migration or cell migration and proliferation is plotted against time. A line is fit through the linear part of this graph between time point

t=100 min and t=1000 min. The slope of this linear fit represents the scratch closing speed of the different cell clones.

3.2.2.4. Cell viability measurement after cisplatin treatment

The cisplatin sensitivity of HEK293 cells is determined by measuring the cell viability in a WST-1 test according to the manufacturer's instructions after the cells had been exposed to a cisplatin dilution series for 2 days. The half-maximal inhibitory concentration (IC_{50}) is used as a measure for cisplatin sensitivity. It describes the cisplatin concentration that causes a cell viability halfway between the maximum (100% cell viability untreated with cisplatin) and baseline (at the highest cisplatin concentration). HEK293 cells are plated at 3,000 cells in 50 µl medium per well in 96-well plates (at a cell density of $9.4x10^3$ cells per cm^2). Each cell line is plated into a total of 3x11 wells, so that the ten cisplatin dilutions and a background control without the addition of WST-1 is measured in triplicate each. In addition, wells containing only medium but no cells are included in triplicate as blank controls. After incubating the cells for 24 hours in cell culture incubators at 37°C in a humidified atmosphere containing 5% CO_2, 50 µl 2x cisplatin diluted in medium is added to the 50 µl medium per well to reach the final cisplatin concentrations of 0 µM, 0.05 µM, 0.2 µM, 0.8 µM, 2 µM, 3.2 µM, 5 µM, 15 µM, 50 µM and 150 µM in 3 wells per cell line each. After 2 days of cisplatin treatment, 10 µl WST-1 is added per well to determine the cell viability. After incubating the cells with WST-1 for 4 hours in a cell culture incubator at 37°C in a humidified atmosphere containing 5% CO_2, the absorption of each well is measured at a wavelength of 440 nm. The absorption values are blank-corrected by subtracting the mean value of the blank wells containing no cells but treated with WST-1. The cell viability is calculated by normalising all values to the mean absorption of cells cultivated without cisplatin, which is set as 100%. The normalised cell viability in percent is plotted against the log_{10} values of the cisplatin concentrations in a dose-response curve, and the IC_{50} is calculated by nonlinear regression using the FindECanything function of GraphPad Prism 5 for Windows.

3.2.3. DNA methods

Genomic DNA is extracted from HEK293 cells and mouse xenografts for whole genome sequencing as well as PCR amplification and sequencing of specific target genes. The Sanger sequencing service provided by the company GATC (Konstanz) is used to sequence DNA fragments such as PCR products.

3.2.3.1. Total DNA extraction

Total genomic DNA is isolated from cells using the DNeasy Blood & Tissue Kit according to manufacturer's instructions. Up to $5x10^6$ cells are collected by centrifugation for 5 minutes at 300 rcf. The cell pellet is resuspended in 200 µl PBS, and 20 µl proteinase K are added to the cell suspension. The cells are lysed by addition of 200 µl buffer AL and thoroughly vortexed. After incubation at 56°C for 10 minutes, 200 µl ethanol (96-100%) are added, and the samples are mixed thoroughly by vortexing. The cell lysate is then pipetted into a DNeasy Mini spin column and centrifuged for 1 minute at 20,000 rcf. The flow-through and the collection tube are discarded, and the column is placed in a new collection tube. The column is washed with 500 µl buffer AW1, then centrifuged for 1 minute at 20,000 rcf and placed in a new collection tube. The old collection tube and the flow-through are discarded. The column is washed again with 500 µl buffer AW2, then centrifuged for 3 minutes at 20,000 rcf to dry the membrane of the column. The flow-through and collection tube are discarded, and the column is placed in a new 1.5 ml collection tube. Finally, 200 µl buffer AE are pipetted directly onto the membrane of the column, incubated for 1 minute at room temperature and centrifuged for 1 minute at 20,000 rcf to elute the genomic DNA. Finally, the DNA concentration and the ratio of absorption at 260 nm and 280 nm wavelength as a measure of DNA purity are determined using the NanoDrop spectrophotometer. A 260 nm/280 nm ratio of 1.8 is generally considered ideal for DNA, and DNA extracts with a ratio between 1.8 and 2.0 are accepted as possessing a sufficient quality for further processing. If DNA, RNA and protein extracts are required of the same sample, the AllPrep DNA/RNA/Protein Mini Kit is used instead (see chapter 3.2.6).

3.2.3.2. Polymerase chain reaction PCR

The polymerase chain reaction is used to amplify DNA segments from genomic DNA templates. The amplified segments are then visualised by gel electrophoreses to check whether the target sequence is present in the genome, or used for sequencing of the target sequence. Amplicons used for verifying the target sequence by Sanger sequencing are produced in 20 µl reactions. The components and concentrations of a PCR reaction are detailed in Table 3.

Table 3 Components and concentrations of a PCR reaction

Component	Stock concentration	Final concentration
Nuclease-free H_2O	n/a	n/a
5x Phusion HF buffer	5x	1x
dNTPs	2.5 mM	0.2 mM
Forward primer	100 µM	0.5 µM
Reverse Primer	100 µM	0.5 µM
Phusion HF DNA Polymerase	2 units/µl	0.02 units/µl
Template genomic DNA	variable	250 ng/reaction

The thermocycling conditions (Table 4) are modified depending on the primers used. The annealing temperature is chosen as 3°C above the lower melting temperature of the primer pair.

Table 4 Thermocycling conditions for PCR

Step	Temperature	Time	Cycles
Initial denaturation	98°C	3 min	1x
Denaturation	98°C	15 sec	35x
Annealing	Primer dependant	30 sec	
Elongation	72°C	60 sec per kb	
Final elongation	72°C	10 min	1x
Hold	4°C	indefinite	1x

3.2.3.3. Agarose gel electrophoresis

DNA fragments such as PCR products are separated according to their size by agarose gel electrophoresis. The gels are created by melting 1% (w/v) agarose in an appropriate volume of 1x TAE electrophoresis buffer (40 mM Tris (pH 7.6), 20 mM acetic acid, 1 mM EDTA in mqH_2O), pouring the melted agarose into a gel chamber of the desired size and adding a comb to create the desired number of wells. Ethidium bromide is added to the gel to a final concentration of 0.5 µg/ml in order to visualise the DNA fragments. One volume 6x DNA loading dye (10 mM Tris-HCl (pH 7.6), 0.03% bromophenol blue, 0.03% xylene cyanol, 60% glycerol, 60 mM EDTA) is added to five volumes of DNA sample, and 10 µl DNA sample are loaded into each well. On each gel, 10 µl of a DNA molecular-weight size marker with a size range appropriate for the expected DNA fragments is included. The gel is run in 1x TAE buffer at 120 V and 300 mA for 1 hour. The DNA fragments are visualised and recorded in a Gel Doc EZ Imager under UV light, which causes the ethidium bromide that intercalated into the DNA to emit light at a wavelength of 590 nm.

3.2.3.4. DNA extraction from agarose gels

DNA fragments are extracted from agarose gels using the QIAquick Gel Extraction Kit according to manufacturer's instructions. The DNA fragment with the desired size is visualised with UV light, excised from the agarose gel with a clean scalpel and transferred to a 2 ml tube. The weight of the gel slice is determined. 100 mg gel approximately correspond to a volume of 100 µl. Three volumes of buffer QG are added to each volume of gel, and the mixture is incubated at 50°C for 10 minutes (or until the gel slice is completely dissolved) with intermittent vortexing. All following steps are performed at room temperature. One gel volume of isopropanol is added, the samples are mixed by vortexing and then transferred to a QIAquick spin column. The columns are centrifuged for 1 minute at 20,000 rcf, and the flow-through is discarded. The columns are first washed with 500 µl QG buffer and then with 750 µl PE buffer; after the addition of each buffer the columns are centrifuged for 1 minute at 20,000 rcf and the flow-through is discarded. In order to remove residual wash buffer and dry the membrane, the columns are centrifuged again for 1 minute at 20,000 rcf, before they are placed in fresh 1.5 ml collection tubes. To elute the DNA, 50 µl buffer EB are applied to the centre of the membrane, incubated for 1 minute at room temperature and centrifuged for 1 minute at 20,000 rcf. Finally, the DNA concentration and the ratio of absorption at 260 nm and 280 nm wavelength as a measure of DNA purity are determined using the NanoDrop spectrophotometer. A 260 nm/280 nm ratio of 1.8 is generally considered ideal for DNA, and DNA extracts with a ratio between 1.8 and 2.0 are accepted as possessing a sufficient quality for further processing.

3.2.3.5. Whole genome sequencing

Whole genome sequencing is performed by the DKFZ Genomics & Proteomics Core Facility. Total genomic DNA is submitted to the core facility for sample quality control and library preparation using the Illumina® TruSeq® Nano DNA Library Prep kit, followed by HiSeq X (150 bp Paired End) whole genome sequencing using the Illumina HiSeq X sequencing platform.

3.2.4. RNA methods

Total RNA is extracted from HEK293 cells and xenograft tumours in order to quantify APOBEC3A, APOBEC3B and ACTB transcript levels. Furthermore, mRNA extracted from head and neck squamous cell carcinoma samples is obtained for quantification of APOBEC3A, APOBEC3B and ACTB mRNA expression. Transcript levels are quantified by a gene-specific quantitative PCR (qPCR) after reverse transcription of the mRNA into cDNA.

3.2.4.1. Total RNA extraction

Total RNA is isolated from cultured cells using the RNeasy Mini Kit according to manufacturer's instructions. Up to $5x10^6$ cells are collected by centrifugation at 300 rcf for 5 minutes. The supernatant is discarded, and the pelleted cells are lysed in 350 µl buffer RLT. The cell lysate is pipetted directly into a QIAshredder spin column and centrifuged for 2 minutes at 20,000 rcf. One volume (350 µl) 70% ethanol is added to the homogenised lysate before immediately transferring it to an RNeasy spin column in a 2 ml collection tube, followed by centrifugation for 15 seconds at 20,000 rcf. The flow-through is discarded. The column is washed with 700 µl buffer RW1 and centrifuged for 15 seconds at 20,000 rcf before the flow-through is discarded. Afterwards, the column is washed twice with 500 µl buffer RPE, centrifuged for 15 seconds at 20,000 rcf and the flow-through is discarded. To remove any remaining wash buffer, the column is centrifuged again for 15 seconds at 15,000 rpm. Finally, the column is placed in a 1.5 ml collection tube, and 50 µl RNase-free water is pipetted directly onto the centre of the membrane. The RNA is eluted into the collection tube by a 1 minute centrifugation at 20,000 rcf. Finally, the RNA concentration and the ratio of absorption at 260 nm and 280 nm wavelength as a measure of RNA purity are determined using a NanoDrop spectrophotometer. A 260 nm/280 nm ratio of 2.0 is generally considered ideal for RNA, and RNA extracts with a ratio between 2.0 and 2.2 are accepted as possessing a sufficient quality for further processing. If DNA, RNA and protein extracts are required of the same sample, the AllPrep DNA/RNA/Protein Mini Kit is used instead (see chapter 3.2.6).

3.2.4.2. Reverse Transcription

Total mRNA extracted from cells is reverse transcribed into cDNA using the QuantiTect Reverse Transcription Kit according to the manufacturer's instructions. The total mRNA extract is thawed on ice. Per sample, two genomic DNA elimination reactions are performed to provide enough DNA-free RNA for one cDNA synthesis reaction containing reverse transcriptase (RT) and one without reverse transcriptase (NRT). The NRT samples serve as a control to exclude a contamination of the samples with remaining gDNA, which may lead to a false positive signal during qPCR. The components and final concentrations per reaction are given in Table 5.

Table 5 Components and concentrations of a standard genomic DNA elimination reaction

Component	Stock concentration	Final concentration	Volume per reaction
RNase-free H_2O	n/a	n/a	Fill up to 14 µl
7x gDNA wipeout buffer	7x	1x	2 µl
Template RNA	variable	1 µg/reaction	variable

The genomic DNA elimination reaction is incubated for ten minutes at 42°C and then immediately placed on ice. For each RNA extract, both RT and NRT reactions are prepared on ice according to Table 6.

Table 6 Components and concentrations of a standard RT and NRT cDNA synthesis reaction

Component	Stock concentration	Final concentration	Volume per RT reaction [µl]	Volume per NRT reaction [µl]
5x RT buffer	5x	1x	4	4
RT primer mix			1	1
Reverse transcriptase (RT)			1	-
RNase-free H_2O	n/a	n/a	-	1
gDNA elimination reaction			14	14

For cDNA synthesis, the reactions are incubated at 42°C for 30 minutes. The reverse transcriptase is then inactivated by an incubation at 95°C for three minutes. Finally, the concentration of the cDNA is adjusted to 10 ng/µl by diluting the sample 1:4 with RNase-free H_2O. The samples are either stored on ice and used directly for qPCR, or stored at -20°C until further processing.

3.2.4.3. Quantitative polymerase chain reaction qPCR

Quantitative real-time PCR (qPCR) is employed to quantify the expression of APOBEC3A and APOBEC3B on mRNA level. The number of APOBEC3A and APOBEC3B transcripts is determined relative to the transcripts of the household gene ACTB (β-actin). The reactions are set up by first preparing a primer mix according to Table 7.

Since the APOBEC3A probe is labelled with Hex and the ACTB probe UPL 11 is labelled with Fam, both signals can be measured in the same reaction. As the APOBEC3B probe UPL 39 also carries a Fam fluorophore, it has to be quantified separately from ACTB. To create the qPCR master mix, 1 µl of the primer mix is added to 5 µl 2x LightCycler 480® Probes Master and 3 µl PCR-grade H_2O per reaction. Then, 1 ng cDNA sample is added to 9 µl master mix per well in a 96-well qPCR plate. The plate is covered in sealing foil and briefly centrifuged before the transcript levels are determined in a Lightcycler. All samples are measured in duplicates.

Table 7 Primer mix: concentrations of primers and probes per reaction used for transcript detection in qPCR

Transcript	Component		Stock concentration [µM]	Final concentration [µM]
APOBEC3A	Left Primer	APOBEC3A-L3	10	0.75
	Right Primer	APOBEC3A-R3	10	0.75
	Probe	APOBEC3A-Hex	1	0.1
ACTB	Left Primer	ACTB-L1	10	0.5
	Right Primer	ACTB-R1	10	0.5
	Probe	UPL 11	1	0.1
APOBEC3B	Left Primer	APOBEC3B-L6	10	0.25
	Right Primer	APOBEC3B-R3	10	0.25
	Probe	UPL 39	1	0.1
RNase-free H_2O			n/a	Fill up to 1 µl

3.2.5. Protein methods

Whole cell lysates of HEK293 cells are produced by different methods depending on whether a native protein is required or not for downstream processes. If native protein is required (e.g. for a deamination assay, which requires the deaminases to be intact and enzymatically active), cells were lysed in 1x HED buffer using sonication. If the protein extract is solely used for Western Blotting, the cells can be lysed using SDS.

3.2.5.1. Whole cell lysate in SDS

Total protein is extracted from cultured cells by producing whole cell lysates. Cell lysis using sodium dodecyl sulphate (SDS) denatures all proteins, so the resulting whole cell lysate can be used for SDS-PAGE, but is unsuitable for the deamination assay. Whole cell lysate in SDS is produced by centrifuging cultured cells at room temperature for 5 minutes at 300 rcf, discarding the supernatant, resuspending the cells in 20 µl PBS and adding 200 µl lysis buffer. The lysis buffer consists of 40 µl protease inhibitor cocktail in 960 µl 1% (v/v) SDS. The cell lysate is vortexed thoroughly. In order to homogenise the lysate, it is applied to a QIAshredder spin column and centrifuged at room temperature for 5 minutes at 20,000 rcf. The flow-through is re-applied to the column and centrifuged again. The homogenised lysate is then stored at -20°C until its total protein concentration is determined and it is further processed for SDS-PAGE followed by Western Blot. If DNA, RNA and protein extracts are required of the same sample, the AllPrep DNA/RNA/Protein Mini Kit is used instead (see chapter 3.2.6).

3.2.5.2. Whole cell lysate (native)

Whole cell lysates containing native proteins are used for deamination assays requiring enzymatic function, as well as for SDS-PAGE followed by Western Blot. As the addition of SDS denatures all proteins, the cells are instead disrupted by sonication. Native whole cell lysate is produced by centrifuging cultured cells for 5 minutes at 300 rcf, discarding the supernatant and resuspending the cells in 200 µl 1x HED buffer (25 mM HEPES, 5 mM EDTA, 10% (v/v) glycerol, 1 mM DTT, 4% (v/v) protease inhibitor cocktail, pH 7.8). The cells are disrupted in a bath sonicator, which cycles between 30 seconds sonication and 30 seconds inactivity for 10 minutes. The samples are cooled to 4°C during sonication by adding ice to the sonicator bath. The lysed cells are finally homogenised by applying them to QIAshredder spin columns and centrifuging them at room temperature for 5 minutes at 20,000 rcf. The homogenised lysate is then stored at -20°C until its total protein concentration is determined and it is further processed for a deamination assay or SDS-PAGE followed by Western Blot.

3.2.5.3. Protein concentration assay

The total protein concentration of all whole cell lysates is determined using the DC protein assay according to the manufacturer's instructions. The assay is a colorimetric determination of protein concentration after solubilisation in detergent; essentially it is a slightly improved version of the assay described by Lowry et al. (1951) (reviewed by Peterson, 1979). A 1:2 dilution series of bovine serum albumin (BSA) from 8 mg/ml to 0.125 mg/ml in lysis buffer as well as a negative control containing no protein serves as a protein standard. This standard curve is included each time the assay is performed. If the protein samples contain detergents, a working reagent consisting of 20 µl reagent S per 1 ml reagent A is required. A volume of 10 µl per protein sample and protein standard are pipetted into a 1.5 ml tube. First, 50 µl working reagent are added to each tube, followed by 400 µl reagent B. After 15 minutes incubation at room temperature, 200 µl of the mixture is transferred into two wells of a clear microtiter plate each. Finally, the plate is mixed for 5 seconds in the microplate reader Multiskan Go and the absorbance is read at 750 nm. The absorbance values are blank-corrected by subtracting the reading of the wells containing no protein. Finally, the protein concentration of the samples is calculated using the standard curve as a reference. The slope of the standard curve multiplied by the absorbance of a sample equals the protein concentration.

3.2.5.4. SDS-PAGE and Western Blot

The proteins of whole cell lysates are separated according to their size using a discontinuous sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

A total of 20 µg protein per lane are applied to the SDS-PAGE. The total protein concentration of all whole cell lysates used for SDS-PAGE is adjusted to 2.5 µg/µl by dilution with 1% (v/v) SDS. Four volumes of the concentration-adjusted whole cell lysate are mixed with one volume 5x lane marker reducing sample buffer to a final concentration of 2 µg/µl total protein, incubated at 95°C for 5 minutes and briefly centrifuged. The protein size marker Page Ruler^{TM} Prestained Protein Ladder and 20 µg total protein per well are loaded onto an SDS gel submerged in 1x NuPAGE MOPS SDS running buffer. The gel is run for 3 minutes at 200 V followed by 1.5 hours at 120 V.

After separation, the proteins are transferred from the SDS-PAGE to a carrier material. Wet (tank) electroblotting is employed to transfer the proteins from the SDS-PAGE to a nitrocellulose membrane. Per SDS-PAGE, two sponges, two pieces of filter paper and one piece of nitrocellulose membrane are briefly submerged in transfer buffer (1x NuPAGE^{TM} transfer buffer, 20% (v/v) methanol in mqH_2O). In the wet tank transfer assembly, the nitrocellulose membrane is located on the anode-facing side of the SDS-PAGE gel. They are sandwiched between a filter paper and a sponge on either side, with the sponges being in direct contact with the electrodes. The assembly is placed in a tank, submerged in transfer buffer and exposed to a voltage of 1.5 mA x cm^2 for 1 hour.

After blotting, the gel is stained in Coomassie GelCode™ Blue Stain Reagent for 1 hour to detect untransferred protein. The result is recorded using the Gel Doc EZ Imager.

The protein-bearing nitrocellulose membrane is used to detect specific target proteins by probing with antibodies. All following incubation steps are performed while gently agitating the membrane on a shaker. Unless otherwise noted, incubation is done at room temperature. Firstly, the nitrocellulose membrane is incubated in blocking buffer (4% (w/v) skim milk powder, 0.05% Tween-20 in PBS) for one hour in order to block all unbound membrane sites and prevent unspecific binding of probes to the membrane. Secondly, the membrane is washed three times in PBS-T (0.05% (v/v) Tween-20 in PBS) for five minutes each. The primary anti-human APOBEC3G antibody used to probe for the APOBEC3A and APOBEC3B expression is diluted 1:2500 in blocking buffer and added to the membrane overnight at 4°C. Following three washing steps of five minutes in PBS-T, the membrane is incubated with the goat anti-rabbit secondary antibody diluted 1:3000 in blocking buffer for 1 hour at room temperature. The secondary antibody recognises the constant region of the primary antibody and is conjugated to horse radish peroxidase (HRP). After another three washing steps in PBS-T for five minutes each and an additional five minutes washing in PBS, a 1:1 mixture of the enhanced chemiluminescence (ECL) detection reagents A and B is added to the membrane. The

resulting chemiluminescence signal is either detected using Amersham Hyperfilm ECL X-ray film or the Intas ECL chemical imager at suitable exposure times.

In order to allow probing the membrane with additional primary antibodies, the membrane is stripped of the original probe and detection antibodies. This is achieved by incubating the membrane in 2 ml stripping buffer (67.5 mM Tris-HCl, 2% (w/v) SDS, 100 mM β-mercaptoethanol in PBS, pH 6.8) for 20 minutes at 55°C. Afterwards, the membrane is first washed in PBS-T for five minutes and then incubated in blocking buffer for one hour at room temperature. Following three washing steps of five minutes each in PBS-T, the membrane is incubated with anti-ACTB IgG-HRP 1:5000 diluted in blocking buffer for one hour at room temperature. After the final three washing steps of 5 minutes PBS-T followed by 5 minutes PBS, a 1:1 mixture of the ECL detection reagents A and B is added to the membrane and the chemiluminescence is detected as described above.

3.2.5.5. Deamination assay

The deamination assay is a method to determine the enzymatic activity of cytidine deaminases such as APOBEC3A in a native whole cell lysate adapted from McDougle et al. (2013). The deaminase assay is an oligonucleotide cleavage assay of an oligo that carries only a single C nucleotide (5'-ATT ATT ATT ATT CTA ATG GAT TTA TTT ATT TAT TTA TTT ATT T-AlexaFluor488-3', Eurofin). The sequence context of this C is in the TC target sequence of APOBEC3A. Enzymatically active deaminases can deaminate this cytidine to uracil. Addition of uracil-DNA glycosylase cleaves the glycoside bond between the uracil base and the deoxyribose, creating an abasic site that can be cleaved by addition of NaOH. Thus, in samples containing enzymatically active deaminases, the shorter cleavage product can be observed when separating the oligonucleotides according to their size on a gelelectrophoresis in TBE buffer (8.9 mM Tris, 8.9 mM boric acid, 0.2 mM EDTA). The intact and cleaved oligo are made visible by methylene blue staining. Alternatively, it can be detected using the green fluorescence of the Alexa-488 tag attached to the 3'-end of the oligonucleotide for visualisation.

First, the total protein concentration of all samples for the deamination assay is adjusted to 2 µg/µl by diluting 10 µg of the homogenised native whole cell lysates in nuclease-free H_2O to a final volume of 5 µl. To each sample, 5 µl of deamination master mix is added, resulting in a final protein concentration of 1 µg/µl. The concentrations and volumes of the components per reaction used in the deamination master mix is detailed in Table 8.

Table 8 Deamination reaction master mix: components, concentrations and volumes used per reaction

Component	Stock concentration	Final concentration in 10 µl reaction	Volume per 10 µl reaction
Nuclease-free H_2O	n/a	n/a	3 µl
10x UDG buffer (New England Biolabs)	10x	1x	1 µl
UDG (New England Biolabs)	5,000 units/ml	125 units/ml	0.25 µl
RNase A (QIAgen)	4 mg/ml	0.1 mg/ml	0.25 µl
Alexa-488 labelled oligonucleotide (Eurofin)	1,000 ng/µl	50 ng/µl	0.5 l

The deamination reactions are incubated at 37°C for 2 hours. Afterwards, 2 µl 1M NaOH are added to each 10 µl reaction, and the samples are incubated for 10 minutes at 95°C before they are immediately cooled on ice. Finally, the samples are mixed with an equal volume (12 µl) of 2x formamide buffer (80% (v/v) formamide, 0.05% (w/v) bromophenol blue, 0.01% (w/v) xylene blue in 1x TBE buffer), which serves as loading buffer for the gel electrophorese.

During the incubation time of the deamination reaction, the 15% acrylamide/urea gel for the gel electrophoresis of the samples is prepared. This gel is capable of resolving 18-30 nucleotide long oligonucleotides. For one gel measuring 7 cm x 10 cm x 0.1 cm, 7.5 ml Rotiphorese Gel 30 (37.5:1) acrylamide/bis-acrylamide solution is mixed with 7.2 g urea and 1.5 ml 1x TBE buffer. This mixture is incubated at 37°C while shaking for approximately 10-15 minutes, or until the urea is completely dissolved. After the solution has cooled down to room temperature, 75 µl 10% ammonium persulphate solution (APS) and 7.5 µl N,N,N',N'-tetramethylethane-1,2-diamine (TEMED) are added to start polymerisation of the gel. After briefly mixing the solution by swirling, the gel is immediately poured into an empty standard gel cassette, and a comb is placed on the top of the cassette. Best results are generally accomplished with a 12-well gel. The gel should solidify within 10-15 minutes at room temperature.

Once the gel is completely solidified, it is placed in a gel running chamber filled with 1x TBE buffer. After a pre-run of the gel without the deamination assay samples for 30 minutes at 200 V, the wells of the gel are rinsed to remove unpolymerised material by pipetting the contents of each well up and down a few times and then aspirating the buffer in the wells until no more unpolymerised material is visible. The gel is loaded with 10 µl of each deamination reaction before it is run for 1.5 hours at 200 V.

For visualisation of the substrate and product oligonucleotides, the gel is stained with methylene blue (0.02 % (w/v) methylene blue in 1x TBE buffer) for 5 minutes, followed by several washing steps in water until non-specific staining is removed. An image of the

methylene blue stained gel is recorded using the Gel Doc EZ Imager with the settings for Coomassie blue stained gels.

3.2.6. Simultaneous isolation of total genomic DNA, RNA and protein from the same sample

Simultaneous extraction of total genomic DNA, total RNA and total protein from the same sample cell or tissue sample is done with the QIAgen AllPrep DNA/RNA/Protein Mini Kit according to manufacturer's instructions. The disruption and homogenisation steps depend on the starting material.

If cells serve as starting material, up to $5x10^6$ cells are collected by centrifugation at 300 rcf for 5 minutes at room temperature. All subsequent centrifugation steps are done at 20,000 rcf. The supernatant is discarded, and the pelleted cells are disrupted by the addition of 350 µl buffer RLT. The cell lysate is homogenised by loading it onto a QIAshredder spin column and centrifuging it for 2 minutes. The homogenised lysate is then used for AllPrep extraction.

If tissue serves as starting material, a 30 mg sample is cut off a frozen tumour using a clean scalpel. During the handling and weighing, the tumour is kept frozen by keeping it on dry ice whenever possible. The 30 mg tumour sample is placed in 600 µl buffer RLT in a Precellys tissue lysis tube containing ceramic beads, and disrupted and homogenised using a Precellys tissue lyser for twice 30 seconds at 5,000 Hz with a 5 second break in between. After disruption, the lysate is centrifuged for 3 minutes, and the entire supernatant is then used for AllPrep extraction.

The disrupted and homogenised lysate of cells or tissue is transferred to an AllPrep DNA spin column and centrifuged for 30 seconds. The AllPrep DNA spin column is placed in a fresh 2ml collection tube and stored at 4°C for later DNA purification, while the flow-through is used for RNA and protein purification. Depending on the amount of buffer RLT used for the starting material, 250 µl (if 350 µl buffer RLT is used) or 400 µl (if 600 µl buffer RLT is used) 96-100% ethanol is added and mixed by pipetting. Up to 700 µl of the sample is transferred to an RNeasy spin column and centrifuged for 30 seconds. The flow-through is transferred to a 2 ml tube for later protein purification, while 700 µl buffer RW1 is added to the RNeasy spin column. After centrifugation for 30 seconds, the flow-through is discarded. The column is washed twice with 500 µl RPE buffer, followed first by 30 seconds and then 2 minutes of centrifugation and disposal of the flow-through. After one more centrifugation step for 1 minute to dry the membrane, the RNeasy spin column is placed in a new 1.5 ml collection tube, and 30 µl RNase-free water is pipetted on the centre of the spin column membrane. The total RNA is eluted by

centrifugation for 1 minute. Finally, the RNA concentration and the ratio of absorption at 260 nm and 280 nm wavelength as a measure of RNA purity are determined using the NanoDrop spectrophotometer. A 260 nm/280 nm ratio of 2.0 is generally considered ideal for RNA, and RNA extracts with a ratio between 2.0 and 2.2 are accepted as possessing a sufficient quality for further processing. The sample stored at -80°C until further processing.

Protein is precipitated from the flow-through from RNA purification mentioned above. One volume (either 600 µl or 1000 µl) buffer APP are added to the flow-through, and the samples are mixed vigorously by pipetting up and down and then incubated for 10 minutes at room temperature to precipitate the total protein content. Then, the samples are centrifuged for 10 minutes, and the supernatant is decanted. The protein pellet is washed in 500 µl 70% ethanol and centrifuged again for 1 minute. The supernatant is carefully aspirated, and the protein pellet is air-dried for 5-10 minutes at room temperature. The dried pellet is dissolved in 250 µl lysis buffer (960 µl 1% (v/v) SDS and 40 µl protease inhibitor cocktail) by mixing vigorously and incubated for 5 minutes at 95°C. Then, the sample is centrifuged for 1 minute to pellet residual insoluble material, and the supernatant is stored at -20°C until it is used for protein concentration determination and SDS-PAGE followed by Western Blot.

Finally, the genomic DNA is purified from the AllPrep DNA spin column mentioned above. The column is washed with 500 µl buffer AW1, centrifuged for 30 seconds and the flow-through is discarded. Then, 500 µl buffer AW2 are added to the column, it is centrifuged for 2 minutes and the flow-through is discarded. In order to dry the membrane of the column, the centrifugation is repeated for 1 minute. The genomic DNA is eluted by adding 100 µl buffer EB pre-heated to 70°C to the centre of the column membrane, incubating it for 2 minutes at room temperature and centrifuging it for 1 minute. The elution step is then repeated with an additional 50 µl buffer EB in the same way to elute further DNA. Finally, the DNA concentration and the ratio of absorption at 260 nm and 280 nm wavelength as a measure of DNA purity are determined using the NanoDrop spectrophotometer. A 260 nm/280 nm ratio of 1.8 is generally considered ideal for DNA, and DNA extracts with a ratio between 1.8 and 2.0 are accepted as possessing a sufficient quality for further processing. The sample is stored at -20°C until further processing.

3.2.7. Mouse xenograft growth and characterisation

Six different HEK293 clones are injected into mice to characterise their tumour growth. Four of these clones had been exposed to APOBEC3A-mediated mutagenesis and two to the enzymatically inactive APOBEC3A$_{E72A}$ variant previous to the xenograft experiment.

3.2.7.1. Xenograft mouse model

All animal work in this study was done in cooperation with Karin Müller-Decker, and all animal handling was performed by Brigitte Steinbauer. A total of 52 female NOD scid gamma (NSG) mice are injected with xenografts for this study. The cells are expanded in cell culture, harvested and resuspended in a total of 300 µl PBS. Shortly before injection, the cell suspensions are mixed with an equal volume of matrigel. Each HEK293 clone is injected into eight NSG mice. Four of these mice are injected with $5x10^4$ cells, the other four with $3x10^5$ cells in a total volume of 100 µl 1:1 matrigel and PBS mixture. As an additional control, $3x10^5$ standard HEK293 Flp-In T-Rex cells in a total volume of 100 µl 1:1 matrigel and PBS mixture are injected into four mice. The size of the xenograft tumours is determined every 3-4 days with a calliper. Once the tumour reaches a diameter of 1.4 cm in any dimension, the experiment is terminated and the mice are killed by cervical dislocation. Pictures of the mice bearing the tumours are recorded. Of each mouse, the entire tumour as well as tissue samples of lung, liver and kidney are collected. The draining lymph nodes are also collected if they are found to be enlarged. The weight of the tumour is recorded. A quarter of the tumour as well as the lung, kidney and liver tissue samples are fixed in formalin and embedded in paraffin. Another quarter of the tumour is prepared for cryosections. The remaining half of the tumour is flash-frozen in liquid nitrogen, to be used for extraction of total DNA, RNA and protein.

3.2.7.2. Creating tumour sections

One half of each xenograft tumour is formalin-fixed and paraffin-embedded (FFPE). In order to assess tumour morphology and perform immunohistochemistry, the tumours are cut into sections. The paraffin blocks containing the tumour samples are cooled to 4°C on a cooling plate, and then cut into 0.5 µm thick sections using a rotary microtome. The sections are allowed to relax and straighten out on a 37°C water bath, before they are mounted on Superfrost Plus microscope slides. The slides with the tumour sections attached to them are dried at 37°C overnight.

3.2.7.3. Haematoxylin eosin stain

FFPE tumour sections are stained with haematoxylin and eosin (HE) to assess the morphology of the xenograft tumours. First, the sections are deparaffinised according to the deparaffinisation series detailed in Table 9.

Table 9 Deparaffinisation series steps and durations

Deparaffinisation steps	Duration
Xylene	5 minutes
Xylene	5 minutes
100% ethanol	3 minutes
100% ethanol	3 minutes
3% H_2O_2 in 70% ethanol	10 minutes
mqH_2O	1 minute

The slides are then submerged in haematoxylin until the tumour sections are suitably stained, followed by washing 10 minutes in running tap water to fix the stain. Then, the sections are submerged in eosin until a good counterstain is evident, and briefly washed in deionised H_2O. After staining, the sections are dehydrated in a dehydration series detailed in Table 10. The stained sections are imaged with the Cell observer microscope at 100x magnification. The overlapping images are stitched together to an image of the entire tumour section using the Zeiss Zen software.

Table 10 Dehydration series steps and duration

Dehydration steps	Duration
70% ethanol	briefly
96% ethanol	briefly
100% ethanol	briefly
100% ethanol	2 minutes
Xylene	briefly
Xylene	2 minutes

The HE-stained sections are used to determine tumour morphology and estimate how much of the tumour is necrotic. Estimation of the necrotic fraction is done by Jochen Heß.

3.2.7.4. Immunohistochemistry

Ki-67 is a marker for cell proliferation. It is used to determine the fraction of proliferating cells in a population, called the proliferative index. Ki-67 positive cells are detected by a monoclonal anti-Ki-67 antibody in both immunofluorescence stains of cell culture samples and immunohistochemistry stains of tumour sections. This assay is used to determine the proliferative index of the xenograft tumours grown in mice.

Formalin-fixed paraffin-embedded tumour sections are stained for presence of the proliferation marker Ki-67. First, the sections are deparaffinised according to the deparaffinisation series detailed in Table 9. Antigen retrieval is done by placing the slides in citrate buffer (10 mM tri-sodium citrate 2-hydrate in deionised H_2O, adjusted to

pH 6.0 with citric acid) pre-heated to 100°C and boiling them for 10-20 minutes in a steam cooker. The slides submerged in citrate buffer are allowed to cool down to room temperature for approximately 20-40 minutes. Then, the slides are washed in PBS for 5 minutes while shaking the chamber. The tumour sections are circled with a hydrophobic pen, placed in a wet chamber and incubated with normal serum from the ImmPRESS HRP anti-rabbit detection kit for 20 minutes at room temperature. After removal of the normal serum, the sections are not washed but immediately incubated with rabbit monoclonal anti-Ki-67 IgG diluted 1:300 in PBS. After incubation at room temperature for 1 hour, the primary antibody is removed and the slides are washed twice in PBS for 5 minutes at room temperature. The sections are incubated with an anti-rabbit IgG peroxidase-conjugated secondary antibody from ImmPRESS HRP anti-rabbit detection kit for 20 minutes at room temperature, before again being washed twice in PBS for 5 minutes. The 3,3'-diaminobenzidine (DAB) substrate producing a brown stain in the presence of peroxidase is prepared from the Vector Laboratories DAB substrate kit for peroxidase according to manufacturer's instructions. One drop buffer, two drops DAB and one drop H_2O_2 solution are added to 2.5 ml mqH_2O. The slides are incubated with the DAB substrate mixture until a sufficient stain develops. The stained slides are incubated for 3 minutes in 50 mM $NaHCO_3$ and briefly washed in deionised H_2O before being submerged in haematoxylin until a good counterstain is evident. Finally, the slides are washed in running tap water for 10 minutes. After staining, the sections are dehydrated in the dehydration series detailed in Table 10.

The dehydrated sections are covered with Eukitt (Kindler GmbH) and a cover slip and allowed to dry. Imaging of the stained sections is done with the Cell observer microscope at 100x magnification. The overlapping images are stitched together to an image of the entire tumour section using the Zeiss Zen software.

The proliferative index of tumours is determined by immunohistochemistry of formalin-fixed paraffin-embedded (FFPE) sections and estimating the percentage of Ki-67 positive cells in living tumour tissue. Necrotic areas are excluded from the estimation of the proliferative index. Estimation of the proliferative index of tumour sections is performed by Jochen Heß.

3.2.8. Patient sample acquisition and analysis

The tumour samples of head and neck squamous cell carcinoma (HNSCC) patients as well as their clinical data are kindly provided by Jochen Heß. They form the basis of the prospective study HIPO-POP019 to elucidate the molecular mechanisms of treatment failure in HNSCC patients.

3.2.8.1. Data and material available from HNSCC patients

All HNSCC samples consist of at least 50% tumour tissue. Whole exome sequences of all tumours and corresponding healthy tissue references are available. Mutations in the tumours are called by comparing the tumour whole exome sequence to that of the healthy tissue reference. The HPV status of all tumours is determined by Lea Schröder and Dana Holzinger using Multiplex Genotyping to detect the DNA of 51 different HPV types with β-globin as a control. The samples are considered HPV-positive if HPV DNA is detected, and HPV-negative if HPV DNA is absent but β-globin is detected. In addition, HPV RNA of E6*I (the spliced mRNA of HPV E6) and ubiquitin C mRNA as internal control are detected using the luminex platform. The samples are considered HPV-positive if HPV E6*I mRNA is detected, and HPV-negative if HPV E6*I mRNA is absent but ubiquitin C mRNA is detected. A tumour is considered HPV-driven if it is positive for both HPV DNA and HPV E6*I mRNA. Extraction of the mutational signatures from tumour exomes is done by non-negative matrix factorisation. The analysis of the mutational landscape in all HNSCC samples is done by Marc Zapatka and Mario Hlevnjak. Furthermore, transcriptome data is available from a microarray. RNA extracts of the tumours are also provided and are used to quantify transcript levels of APOBEC3A and APOBEC3B relative to ACTB in a gene-specific qPCR (see chapters 3.2.4.2 and 3.2.4.3 for details on reverse transcription and qPCR).

3.2.8.2. Principal component analysis

A principal component analysis of the total mutational load, the APOBEC3A and APOBEC3B expression levels and the percentage of APOBEC3-mediated mutations out of all mutations is performed and the resulting data visualised in a heatmap using the ClustVis webtool (http://biit.cs.ut.ee/clustvis/ and described by Metsalu and Vilo, 2015). The default settings are used for principal component analysis and generation of the heatmap. The parameters used for principal component analysis are first converted into z-scores by subtracting the mean μ of the parameter over all samples from the value x for a parameter in a specific sample, and dividing the difference by the standard deviation σ for this parameter.

$$z - score = \frac{x - \mu}{\sigma}$$

This is done separately for each of the parameters in question. In addition to the z-scores for the parameters mentioned, all samples are annotated with their HPV-status, lymph node involvement and extracapsular spread status where available and included in the heat map.

4. Results

The HEK293 cells carrying isogenically inserted and doxycycline (dox)-regulated transgenes used in this study had previously been established by Ann-Mareen Franke. Previous experiments by Franke et al. (unpublished) provided experimental proof that APOBEC3A causes a mutational signature in the genome of human HEK293 cells, while no signature was found in HEK293 cells exposed to the enzymatically inactive variant APOBEC3A$_{E72A}$. This APOBEC3A-mediated mutational signature is characterised by C-to-T and C-to-G mutations in a 5'-TC context, and it is a combination of the two published signatures 2 and 13 that had previously been attributed to 5'-TC-specific APOBEC3 family members in general (Alexandrov et al., 2013b; Alexandrov et al., 2013a). The HEK293 cells carrying the dox-regulated APOBEC3A transgenes were used as a model in this study to determine the effect of APOBEC3A-mediated mutagenesis on tumour growth of immortalised and transformed cells *in vitro* and *in vivo*.

4.1. APOBEC3A-mediated mutagenesis in HEK293 cells leads to a genetically mixed population

This study made use of five cell populations created by Ann-Mareen Franke. Out of these five populations, the three populations called APOBEC3A pop. 01, APOBEC3A pop. 12 and APOBEC3A pop. 20 carry a transgene encoding functional APOBEC3A. Two cell populations were included as controls. The first control was a cell population called APOBEC3A$_{E72A}$ pop. carrying a transgene that encodes an enzymatically inactive variant. The second control was a population of parental cells called HEK293 pop. All cell populations had been expanded from single cells containing the respective transgenes.

APOBEC3A pop. 01 as well as APOBEC3A$_{E72A}$ pop. have previously been used in the study by Ann-Mareen Franke. During the course of this study, the cell populations cultured in certified dox-free FCS were exposed to basal expression levels of APOBEC3A and APOBEC3A$_{E72A}$, respectively. After an additional transgene expression boost by dox treatment for three days, whole genome sequencing and signature extraction showed that APOBEC3A pop. 01 carried the APOBEC3A mutational signature, whereas it was absent in APOBEC3A$_{E72A}$ pop. before the start of this study. The other two APOBEC3A populations, APOBEC3A pop. 12 and 20, were been employed in the previous study and were derived from independent clones after transgene insertion. It was thus assumed that the time they had been exposed to basal APOBEC3A expression was shorter, and consequently that they had a lower mutational load caused by APOBEC3A-mediated mutagenesis than APOBEC3A pop. 01 at the beginning of the study presented here.

4.1.1. Prolonged mutational pressure by APOBEC3A can lead to loss of APOBEC3A activity by various mechanisms

Expression and/or activity of endogenous deaminases in HEK293 cells may create background cytidine deamination that can result in mutations. In order to exclude an influence of endogenous deaminases on the genome, the parental HEK293 cells were tested for cytidine deaminase activity as well as expression of endogenous APOBEC3A and APOBEC3B. Endogenous APOBEC3A or APOBEC3B expression was neither observed at protein level (Figure 5A) nor at mRNA level (Figure 5B) in parental HEK293 cells. An absence of background deaminase activity in the 5'-TC sequence motif by other cytidine deaminases was confirmed in the deamination assay (Figure 5C).

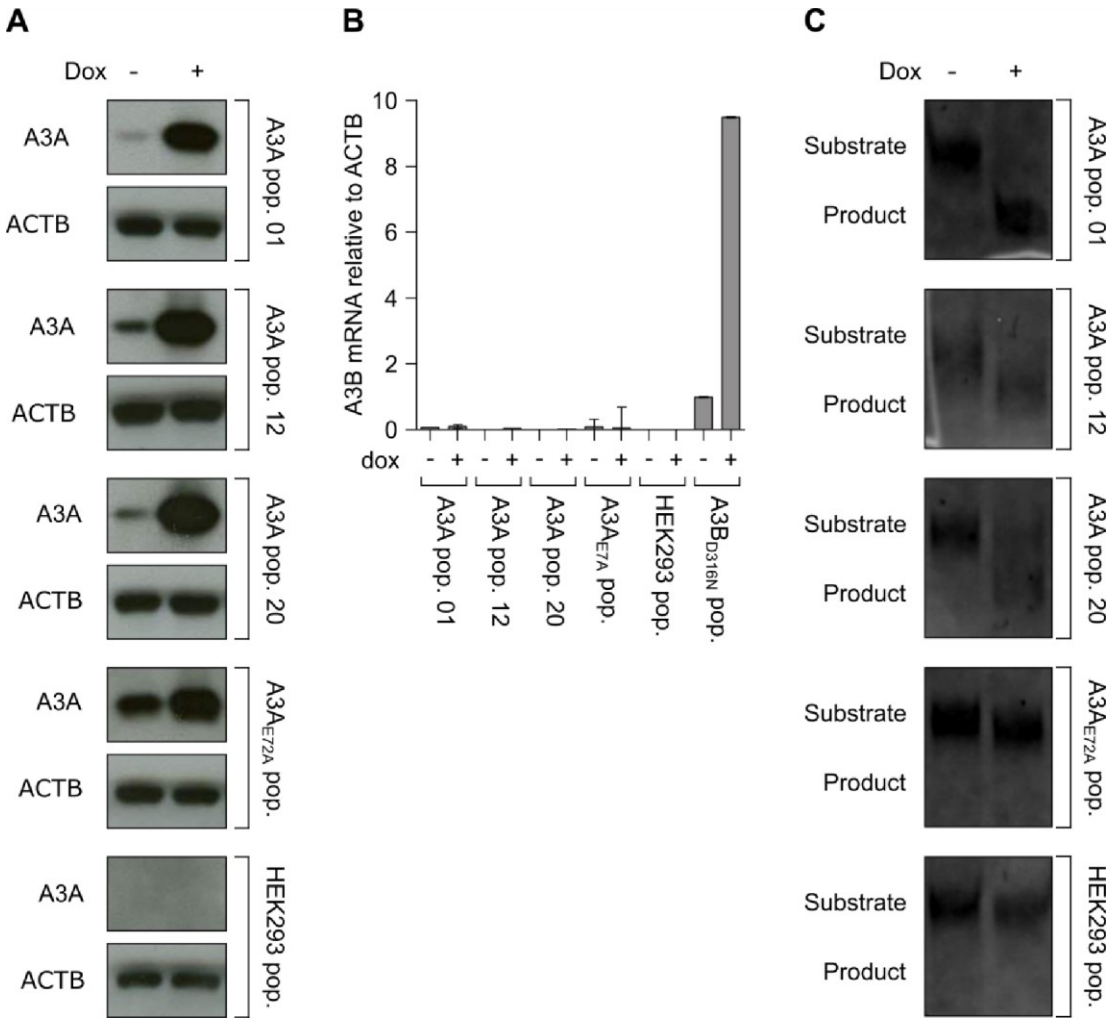

Figure 5 HEK293 cells do not show endogenous APOBEC3A (A3A) expression or activity, but the APOBEC3A transgene shows basal expression and deaminase activity that can be increased by dox treatment. The cell populations were cultured for three days either with or without 1 µg/ml dox. **A** Basal and dox-induced APOBEC3A (A3A) transgene expression in all HEK293 cell populations. Whole cell lysates in 1% SDS where separated in an SDS-PAGE and transferred to a nitrocellulose membrane in a Western blot. The membrane was probed with a polyclonal rabbit anti-APOBEC3G antibody which cross-reacts with APOBEC3A as primary and a goat anti-rabbit HRP conjugate as secondary antibody. ACTB served as a loading control. **B** Endogenous APOBEC3B (A3B) mRNA expression in all HEK293 cell populations. Total mRNA was extracted and reverse transcribed into

cDNA. APOBEC3B expression was determined in a gene-specific qPCR and normalised to ACTB mRNA expression. Expression of basal and dox-induced APOBEC3B$_{D316N}$ transgene was included as a positive control. ***C*** Basal and dox-induced deaminase activity in all HEK293 cell populations as determined in the oligonucleotide cleavage assay. Deaminase activity leads to the cleavage of the oligonucleotide substrate into a shorter product.

Based on these parental cells, Ann-Mareen Franke had established the cells carrying stably integrated APOBEC3 transgenes. In all cell populations, the transgenes showed a basal level of transgene expression in the absence of dox, which could be enhanced by dox treatment (Figure 5A). Dox-induced expression of the functional APOBEC3A transgene caused a strong induction of deaminase activity. No deaminase activity was observed in cells expressing the enzymatically inactive APOBEC3A$_{E72A}$ (Figure 5C). The constant basal expression implies that constant APOBEC3A-mediated mutagenesis occurs in the cells carrying the enzymatically active transgene. This basal expression likely also causes cytidine deamination that is below the detection threshold of the deamination assay, but still able to mutagenise the genome at a low rate over a prolonged period of time.

In order to test whether all cells in each of the populations expressed the transgene, APOBEC3A and APOBEC3A$_{E72A}$ expression after dox induction was analysed at single cell level. After two days of culturing the cells in medium containing 1µg/ml dox, APOBEC3A expression was detected by immunofluorescence (Figure 6A). The enzymatically inactive variant APOBEC3A$_{E72A}$ was expressed in all cells, whereas on average only 70% of the cells expressed the enzymatically active APOBEC3A transgene (Figure 6B).

There are two possible explanations why the APOBEC3A cell populations were heterogeneous for transgene expression: either the expression levels were below the detection limit, or the cells lost transgene expression entirely. The loss of transgene expression in a fraction of the cells was observed in all three APOBEC3A populations, but not in the APOBEC3A$_{E72A}$ population. This hints at a gradual loss of APOBEC3A expression in the population, which may be caused by sustained APOBEC3A-mediated mutagenesis, the cyto- and genotoxic effects is causes and the resulting selection processes.

In order to test whether the APOBEC3A-mediated mutational signatures 2 and 13 can still be observed in a cell population that is heterogeneous for transgene expression, total genomic DNA was extracted and the whole genome sequenced. The APOBEC3A mutational signatures were detected in the whole genome sequence of the cell population APOBEC3A pop. 01, although 20% of the cells no longer expressed APOBEC3A even after dox induction (Figure 6C).

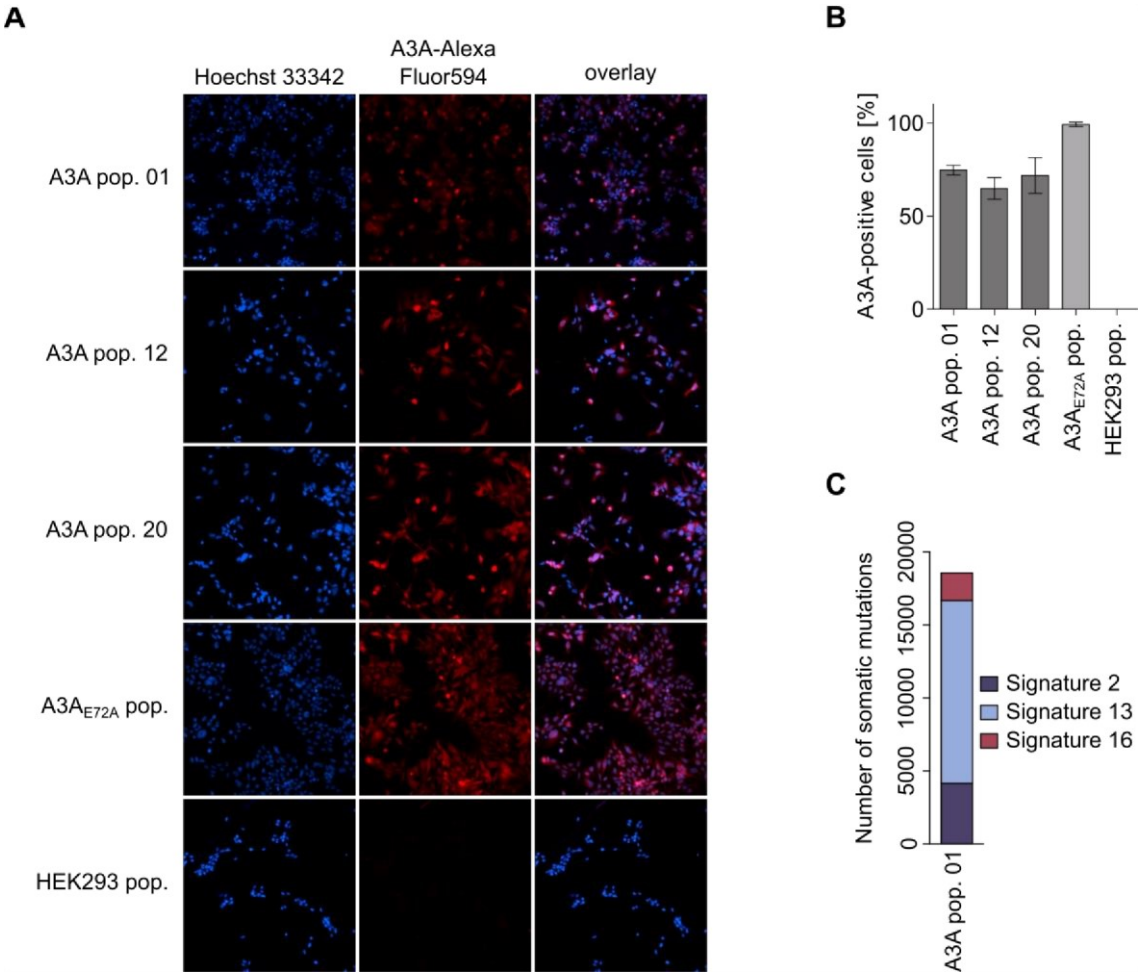

Figure 6 APOBEC3A transgene expression is absent in a fraction of the APOBEC3A-mutagenised HEK293 cell populations. **A** APOBEC3A (A3A) transgene expression at single cell level after dox induction. The cell populations had been cultured on poly-Lys coated cover slips for two days with 1 µg/ml dox. APOBEC3A expression was detected in an immunofluorescence stain with a polyclonal rabbit anti-APOBEC3G antibody which cross-reacts with APOBEC3A as primary and a goat anti-rabbit AlexaFluor 594 conjugate as secondary antibody. Hoechst 33342 served as counterstain. **B** Quantification of APOBEC3A-expressing cells (mean ± standard deviation) in immunofluorescence images. For each cell population, the fraction of APOBEC3A-positive cells out of the total number of cells in percent was determined. Three images of each cell line were quantified. **C** APOBEC3A mutational signatures 2 and 13 can be detected in a cell population which is heterogeneous for APOBEC3A expression. Total DNA extracted from APOBEC3A pop. 01 was whole genome sequenced, and an extraction of mutational signatures was performed. The stacked bars represent the number of somatic mutations ascribed to each of the mutational signatures detected in the cell population.

If the loss of APOBEC3A expression is indeed gradual, the fraction of cells which lost the transgene expression should increase over time. If it is caused by sustained APOBEC3A-mediated mutagenesis and selection, an increased rate of mutagenesis as well as an increase in selection pressure should speed up the loss of transgene expression. Both mutagenesis and selection can be increased by increasing APOBEC3A expression. Thus, the three APOBEC3A populations as well as one APOBEC3A$_{E72A}$ and one parental HEK293

population were exposed to basal or dox-induced transgene expression over 11 weeks (22 passages).

Cells that had been subjected to high mutagenic pressure caused by dox-induced APOBEC3A expression over a course of 11 weeks completely lost all transgene expression at the end of this experiment. Cells that had been subjected to lower mutagenic pressure caused by basal expression of APOBEC3A showed a loss of basal transgene expression and greatly reduced transgene expression after dox induction. The expression levels of the enzymatically inactive APOBEC3A$_{E72A}$ mutant remained unchanged over time (Figure 7A). As expected, cell populations which lost the expression of the enzymatically active APOBEC3A transgene also lost deaminase activity (Figure 7B). These observations suggest that the loss of APOBEC3A expression is indeed gradual, and as it depends on the APOBEC3A dose, it was most likely caused by APOBEC3A-mediated mutagenesis and selection processes.

One potential reason for the loss of APOBEC3A expression may be either the loss or mutagenesis of the APOBEC3A transgene inserted into the HEK293 genome. Thus, a transgene-specific PCR was employed to determine the presence and sequence integrity of the APOBEC3A transgene. It showed that the transgene itself was absent in the APOBEC3A pop. 12 population (Figure 7C). Sequencing of the PCR products revealed that the other two APOBEC3A populations, pop. 01 and pop. 20, both carried a different premature stop codon in the transgene sequence (Figure 7D). Both stop codons were caused by C→A mutations in a 5‘-T<u>C</u> context and thus match the APOBEC3 mutational signature 13. The loss of APOBEC3A expression in the cell populations with the premature stop codon may be due to nonsense-mediated decay of the mRNA. Alternatively, this mutation might also result in loss of the epitope, so that the protein can no longer be detected in Western Blot. Overall, the activity of the APOBEC3A transgene was lost due to various mechanisms, while the enzymatically inactive mutant remained unchanged.

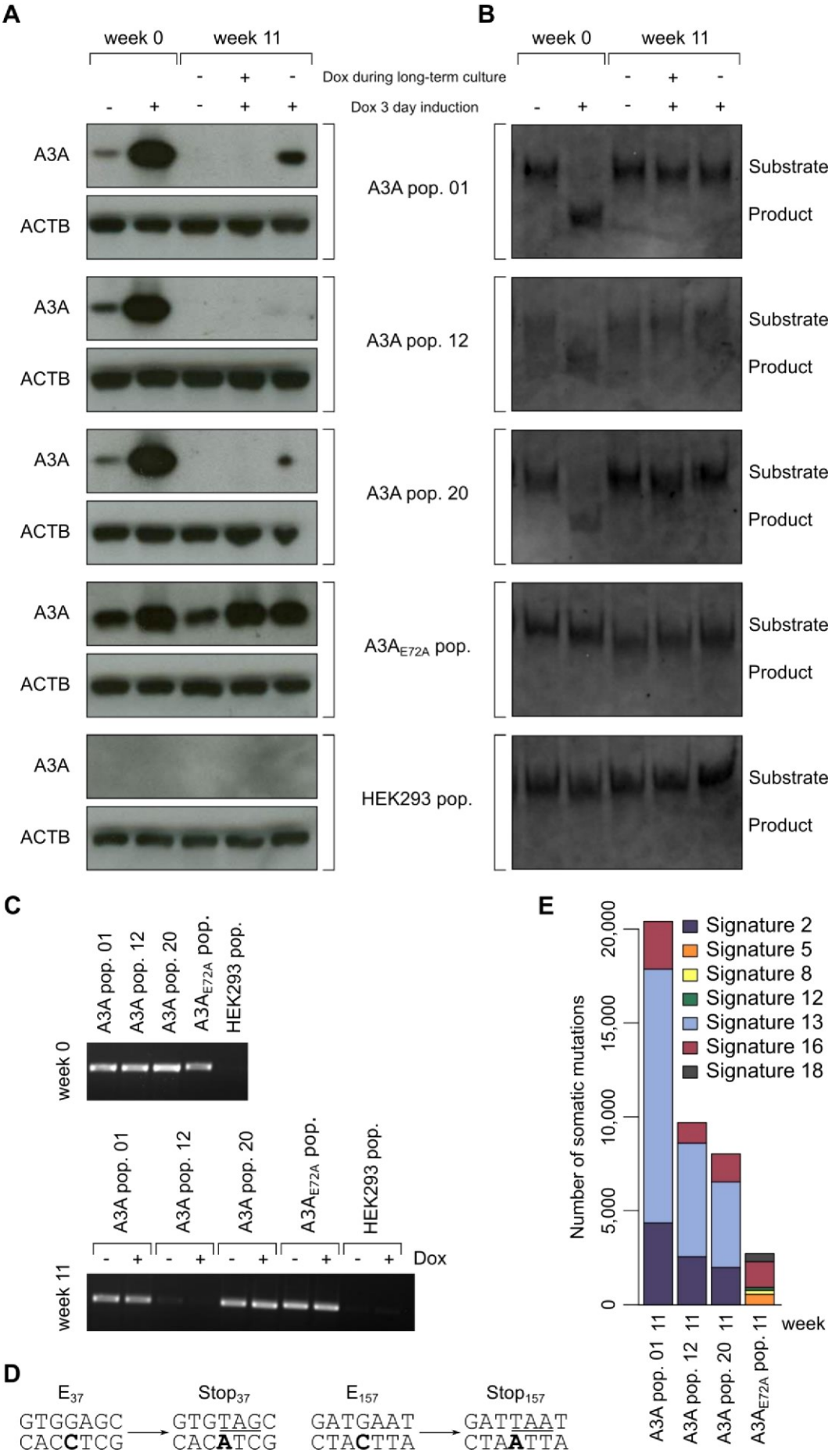
A
week 0
week 11
Dox during long-term culture
Dox 3 day induction
A3A
ACTB
A3A pop. 01
A3A pop. 12
A3A pop. 20
A3A$_{E72A}$ pop.
HEK293 pop.
B
Substrate
Product
C
week 0
week 11
Dox
D
E$_{37}$
Stop$_{37}$
E$_{157}$
Stop$_{157}$
GTGGAGC
CACCTCG
GTGTAGC
CACATCG
GATGAAT
CTACTTA
GATTAAT
CTAATTA
E
Number of somatic mutations
0
5,000
10,000
15,000
20,000
Signature 2
Signature 5
Signature 8
Signature 12
Signature 13
Signature 16
Signature 18
week
A3A pop. 01 11
A3A pop. 12 11
A3A pop. 20 11
A3A$_{E72A}$ pop. 11

Figure 7 APOBEC3A (A3A) activity in HEK293 cell culture is lost over time due to different mechanisms after inducing a mutational signature. For 11 weeks, the cells were either exposed to basal transgene expression (without dox during long-term culture) or to enhanced transgene expression (by adding 1 µg/ml dox to the medium during long-term culture). The cell populations were additionally cultured for three days either with or without 1 µg/ml dox before extraction of protein samples. **A** APOBEC3A expression at the start (week 0) and the end (week 11) of the long-term culture. Whole cell lysates in 1% SDS where separated in an SDS-PAGE and transferred to a nitrocellulose membrane in a Western blot. The membrane was probed with a polyclonal rabbit anti-APOBEC3G antibody which cross-reacts with APOBEC3A as primary and a goat anti-rabbit HRP conjugate as secondary antibody. ACTB served as a loading control. **B** Basal and dox-induced deaminase activity in all HEK293 cell populations as determined in the oligonucleotide cleavage assay. Deaminase activity leads to the cleavage of the oligonucleotide into a shorter product. **C** All cell populations at the start (week 0) and end (week 11) of the long-term culture were tested for the presence of the APOBEC3A transgene. A transgene-specific PCR was performed on an extract of total genomic DNA of all cell populations at the beginning and end of the long-term culture. The PCR products were separated and visualised by gel electrophoresis. **D** Potentially APOBEC3A-mediated mutations (in bold) causing premature stop codons (underlined) in the transgene-specific PCR products of APOBEC3A pop. 01 (left) and APOBEC3A pop. 20 (right). The mutations were identified by Sanger sequencing of the transgene-specific PCR products. **E** APOBEC3 mutational signatures 2 and 13 can be detected in all APOBEC3A cell populations despite the absence of APOBEC3A activity. Total DNA extracted from APOBEC3A pop. 01, 12 and 20 as well as the $APOBEC3A_{E72A}$ population as control were whole genome sequenced. All cell populations used for whole genome sequencing had been kept under dox-induced transgene expression during the long-term culture. An extraction of mutational signatures was performed. The stacked bars represent the number of somatic mutations ascribed to each of the mutational signatures detected in the cell population.

Two processes could have led to the absence of functionally active APOBEC3A transgene. The first possibility is that there were a few cells in the original population that lost functional APOBEC3A shortly after it was inserted. These cells would have an advantage over all cells expressing functional APOBEC3A, as they would not experience the acute negative effects caused by APOBEC3A activity, such as increased cell cycle arrest and apoptosis. Due to this selection advantage, they could have overgrown all other cells. In this case, as the cells had been exposed to APOBEC3A-mediated mutagenesis only for a very short time, detection of the APOBEC3A mutational signature should not be possible.

The other possibility is that the cells were subjected to APOBEC3A-mediated mutagenesis and carry the resulting mutational signature. In this case, the APOBEC3A activity was lost over time because it was selected against due to its negative acute effects. Despite the absence of continuing APOBEC3A activity, a large number of somatic mutations were found in the APOBEC3A-mutagenised populations, and most of these mutations are due to the APOBEC3A mutational signatures 2 and 13. These mutational signatures are entirely absent in the $APOBEC3A_{E72A}$ control population, and the total mutational load is also 7-fold lower than in APOBEC3A population 01 and 3-fold lower than APOBEC3A population 20 (Figure 7E). Due to the loss of functional transgenic

APOBEC3A, the cells were no longer subjected to APOBEC3A-mediated deamination and can thus be considered genetically stable, at least concerning APOBEC3A-driven mutagenesis.

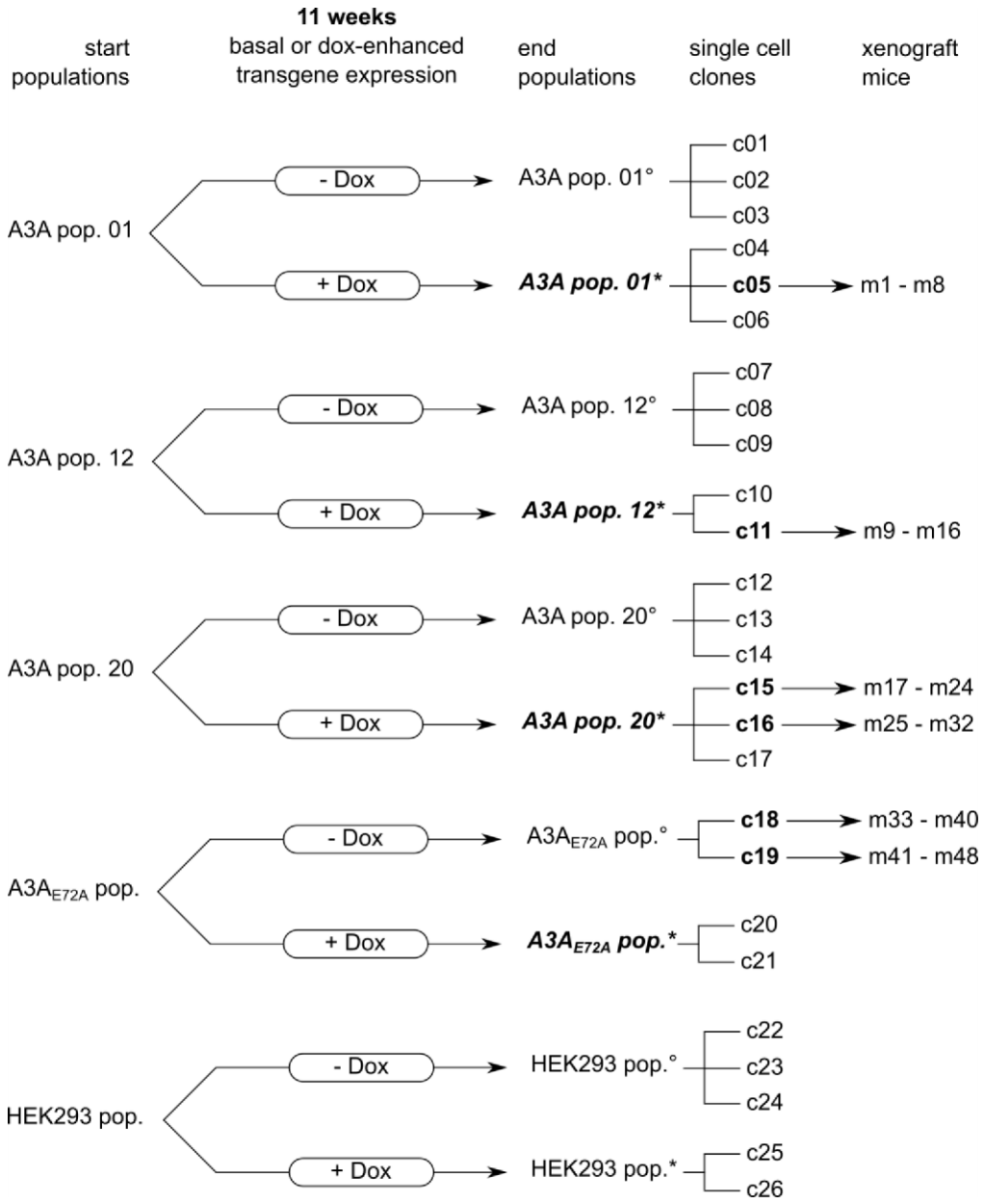

Figure 8 Dynastic tree illustrating the relationships between the original HEK293 cell populations containing the APOBEC3A transgene, the cell populations resulting from the different treatment during long-term culture, the single cell clones and the mouse xenografts. Three independent APOBEC3A (A3A) populations as well as one APOBEC3A$_{E72A}$ (A3A$_{E72A}$) population and one population of parental HEK293 cells were cultured either with 1 µg/ml dox (denoted by *) or without 1 µg/ml dox (denoted by °) over a period of 11 weeks. The cell populations printed in bold italics were whole genome sequences, and the mutational signatures were extracted (results see Figure 7E). A total of 26 single cells were isolated and expanded from the populations after the long-term culture. The six clones c05, c11, c15, c16, c18 and c19 printed in bold were whole genome sequenced. They were furthermore used to verify the stability and reproducibility of the proliferation and migration phenotypes that had been tested in all clones, and injected into eight mice each as xenografts.

The resulting cell populations consisted of varying numbers of clones, each carrying a distinct set of mutations and therefore showing differences in phenotype. The phenotype of the population was thus defined by its clonal composition. The clonal composition of a cell population depends on selection processes, and may change under subtly different culture conditions. To prevent effects on the phenotype by changes in the clonal composition of the culture, single cell clones from each population were isolated and expanded. A total of 26 clones were isolated. The exact relationships between the different populations and clones are depicted in Figure 8.

All clones were tested for transgene expression after dox treatment. None of the clones derived from the cell populations that had been subjected to APOBEC3A-driven mutagenesis showed any expression of the transgene even after dox induction. In contrast, the clones derived from the APOBEC3A$_{E72A}$ cultures all retained the ability to express the transgene after dox induction (Figure 9A). This confirms that no more APOBEC3A-mediated mutagenesis occurs in these clones, and they are genetically stable in this respect. The absence of APOBEC3A expression is caused by a loss of the transgene in 16 out of the 17 APOBEC3A clones (Figure 9B). The sequence of the transgene was found to carry a premature stop codon in c16, which still contained the transgene sequence. As in the underlying cell population, the mutation causing the premature stop codon matches the APOBEC3 mutational signature 13 (Figure 9C).

The clones were compared to the populations they were derived from. This confirmed that APOBEC3A activity was lost by different mechanisms within the same population. For instance, the population APOBEC3A pop. 20 contained transgenes carrying a premature stop codon (Figure 7D). Two out of three clones derived from this population (c14 and c17) lost the transgene, while the remaining clone (c16) shows the same premature stop codon mutation found in the population (Figure 9B and C). This implies that a part of the population lost APOBEC3A activity due to loss of the transgene, while another part received a mutation resulting in the premature stop codon.

As none of the clones still contained functional APOBEC3A, proof was needed that they had been exposed to enzymatically active APOBEC3A in the past and contained the APOBEC3A mutational signature. Out of the 26 clones, a total of 6 were whole genome sequenced. Clones c5, c11, c15 and c16 were derived from populations that had been exposed to APOBEC3A and that carry the APOBEC3 mutational signatures. The clones should therefore also carry the APOBEC3 mutational signatures. Clones c18 and c19 were included as controls, as they had only been subjected to the enzymatically inactive APOBEC3A$_{E72A}$ and should therefore have a much lower total mutational load and not contain the APOBEC3A mutational signatures.

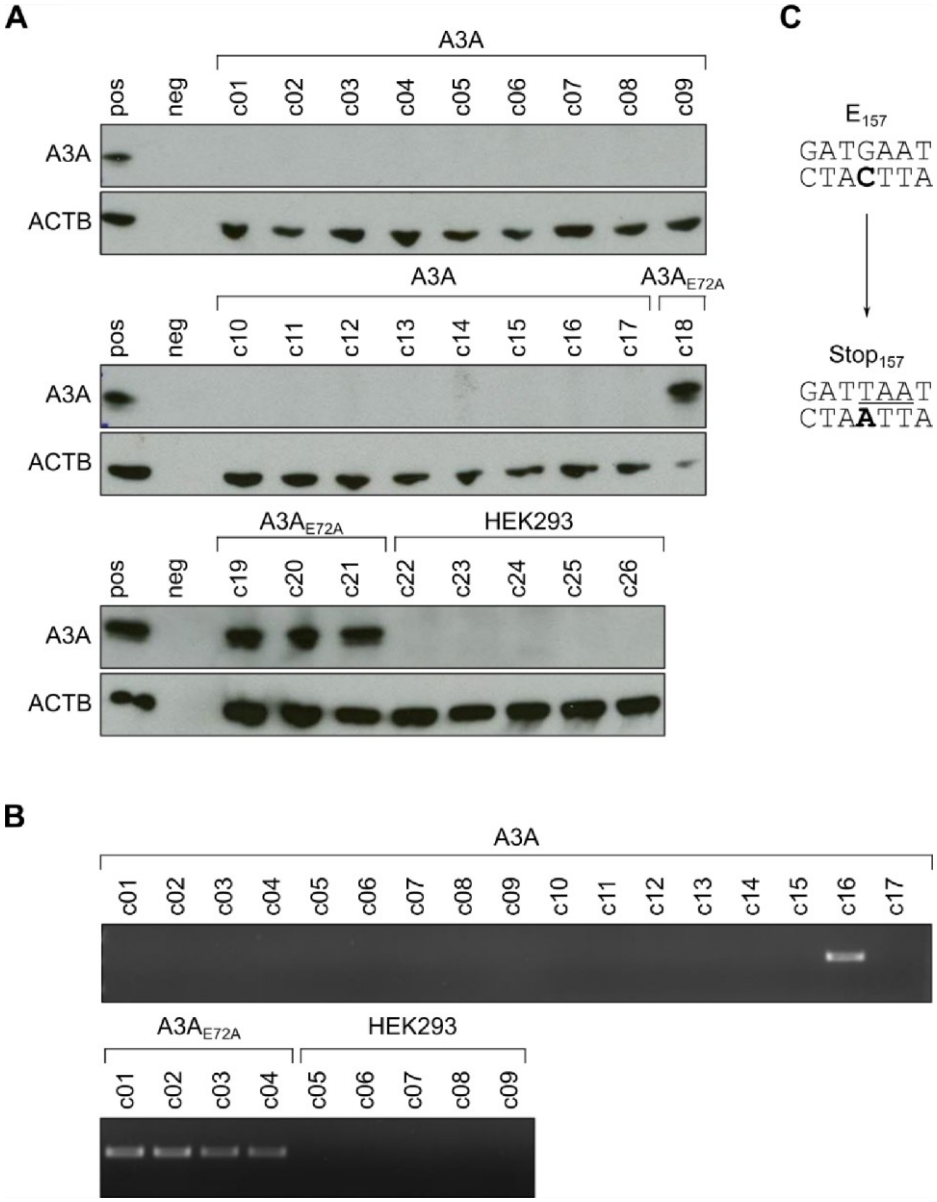

Figure 9 HEK293 clones derived from APOBEC3A-mutagenised cell populations are genetically stable regarding further APOBEC3A-mediated mutagenesis after they acquired the APOBEC3 mutational signature. **A** Dox-induced APOBEC3A (A3A) expression of all single clones. The clones were cultured with 1 µg/ml dox for three days. Whole cell lysates in 1% SDS where separated in an SDS-PAGE and transferred to a nitrocellulose membrane in a Western blot. The membrane was probed with a polyclonal rabbit anti-APOBEC3G antibody which cross-reacts with APOBEC3A as primary and a goat anti-rabbit HRP conjugate as secondary antibody. ACTB served as a loading control. Dox-induced APOBEC3A$_{E72A}$ expression and a sample containing no cell lysate served as positive (pos) and negative (neg) controls. **B** All single clones were tested for the presence of the APOBEC3A transgene. A transgene-specific PCR was performed on an extract of total genomic DNA of all cell populations at the beginning and end of the long-term culture. The PCR products were separated and visualised by gel electrophoresis. **C** Potentially APOBEC3A-mediated mutations (in bold) causing a premature stop codon (underlined) in the transgene-specific PCR product c16. The mutation was identified by Sanger sequencing of the transgene-specific PCR product.

However, the extensive number of mutations found in the HEK293 clones and the large number of structural changes such as loss of heterozygosity made a comparison

between the whole genome sequences of the clones and the whole genome of the original parental population impossible. This was further complicated by the differences in sequencing depth: the original parental HEK293 cell population was sequenced at a 10x coverage, whereas the clones were sequenced at 30x coverage. As a result, mutations could not be confidently called, and an extraction of mutational signatures was not possible.

4.1.2. APOBEC3A-mediated mutagenesis does not affect mean cellular proliferation and migration in HEK293 cell populations

Following the confirmation that the clones were genetically stable concerning additional APOBEC3A-mediated mutagenesis, they were characterised concerning their proliferation and migration phenotypes. The proliferation phenotype was characterised as the proliferative index, i.e. the fraction of Ki-67 positive cells out of all living cells. The migration phenotype was described by the speed at which the cells closed a scratch in an otherwise confluent cell layer. In order to exclude the effect of cell proliferation on scratch closure, cell division was inhibited by the addition of mitomycin C to the cell culture medium during the scratch assay.

Differences in proliferation and migration phenotype were observed between clones derived from the same underlying population. These phenotypic differences between clones suggest an underlying phenotypic variability within each cell population. Only the APOBEC3A-mutagenised clone c15 showed a significantly higher proliferative index than was observed in any of the clones derived from the $APOBEC3A_{E72A}$ or parental HEK293 control populations. The proliferation phenotypes of all other clones that had been exposed to APOBEC3A were consistent with the phenotypic variability present within the controls (Figure 10A). None of the APOBEC3A-mutagenised clones displayed a migration speed outside the range observed in the controls (Figure 10B).

In order to determine the stability and reproducibility of the phenotypes of different clones as well as the phenotypic differences between clones, the proliferation and migration of the same six clones c5, c11, c15, c16, c18 and c19 that had been selected for whole genome sequencing were repeatedly characterised in three additional independent experiments. In contrast to previous experiments, cell migration was determined with a modified experimental setup. Regarding the proliferative index, the results in both the original characterisation and the verification experiments do not differ for the selected clones (Figure 10C), i.e. a comparison of the proliferative index of these six clones between the original characterisation and the verification experiments showed no significant difference.

Comparing the three independent migration experiments, the observed phenotypes and differences between clones were found to be stable and reproducible (Figure 10D). Due to the modified experimental setup, a direct quantitative comparison with the previous results is not valid.

In addition, two-dimensional growth was determined as a further verification. Two-dimensional growth is essentially a combination of migration and proliferation. It was determined in a scratch assay without the addition of proliferation-inhibiting mitomycin C, with the speed of scratch closure as a measure for the speed of two-dimensional growth (Figure 10D). As expected, the scratch closed faster if the cells were proliferating as well as migrating than it did if it was closed only by cell migration. The increase in scratch closing speed caused by proliferation mostly matched the proliferative index of the clones: clones with a higher proliferative index generally showed a larger increase in scratch closing speed, with c11 showing a stronger increase than would be expected from its proliferative index (Figure 10A, C and D). This result obtained with an independent method confirmed the proliferation phenotypes determined with Ki-67 staining.

Finally, the mean phenotype of each cell population was calculated as the mean phenotype over all clones derived from the population in question. This is possible under the assumption that the average phenotypes of the clones derived from one population reflects the average phenotype of that population. Neither the mean proliferation (Figure 10E) nor the mean migration (Figure 10F) were found to be altered in APOBEC3A-mutagenised populations in comparison to the $APOBEC3A_{E72A}$ and HEK293 control populations. This suggests that APOBEC3A-mediated mutagenesis does not influence proliferation and migration in HEK293 cells.

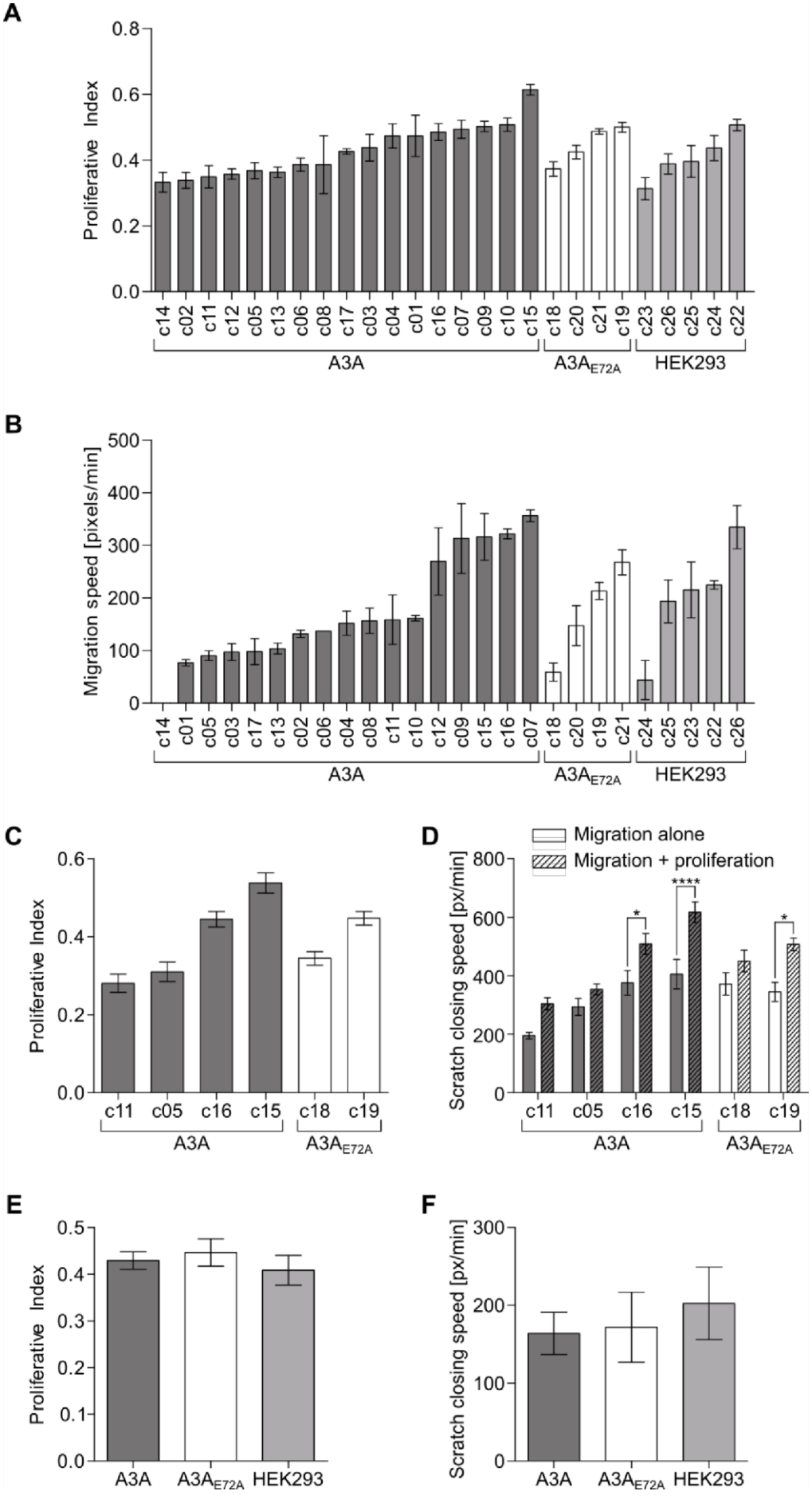
A
Proliferative Index
0.0
0.2
0.4
0.6
0.8
c14 c02 c11 c12 c05 c13 c06 c08 c17 c03 c04 c01 c16 c07 c09 c10 c15 c18 c20 c21 c19 c23 c26 c25 c24 c22
A3A
A3A$_{E72A}$
HEK293
B
Migration speed [pixels/min]
0
100
200
300
400
500
c14 c01 c05 c03 c17 c13 c02 c06 c04 c08 c11 c10 c12 c09 c15 c16 c07 c18 c20 c19 c21 c24 c25 c23 c22 c26
A3A
A3A$_{E72A}$
HEK293
C
Proliferative Index
0.0
0.2
0.4
0.6
c11 c05 c16 c15 c18 c19
A3A
A3A$_{E72A}$
D
Migration alone
Migration + proliferation
Scratch closing speed [px/min]
0
200
400
600
800
*
****
*
c11 c05 c16 c15 c18 c19
A3A
A3A$_{E72A}$
E
Proliferative Index
0.0
0.1
0.2
0.3
0.4
0.5
A3A
A3A$_{E72A}$
HEK293
F
Scratch closing speed [px/min]
0
100
200
300
A3A
A3A$_{E72A}$
HEK293

Figure 10 Characterisation of HEK293 single cell clones derived from APOBEC3A-mutagenised cell populations regarding their proliferation and migration phenotype. The clones had been exposed to APOBEC3A (A3A) or the enzymatically inactive APOBEC3A$_{E72A}$ (A3A$_{E72A}$) during long-term culture, or belonged to the parental HEK293 population. **A** Proliferative index of all single cell clones as determined by Ki-67 immunofluorescence stain (mean ± standard error). The proliferative index is a measure of the number of dividing cells in a sample, and it is calculated as the quotient of proliferating cells to total cells as determined in a Ki-67 immunofluorescence assay. Per sample, at least three independent visual fields at 100x magnification were counted to determine the proliferative index. **B** Migration speed in pixels per minute of all single clones as determined in a scratch assay (mean ± standard error). Proliferation of these cells was inhibited by 5 µg/ml mitomycin C in the medium. Per sample, at least three scratch positions were recorded at 100x magnification and analysed. **C** Proliferative index of selected single clones (mean ± standard error), determined in three independent experiments. The proliferative index was determined as described in A. **D** Scratch closure speed of selected single clones (mean ± standard error) with and without the added influence of proliferation, determined in three independent experiments. Scratch closure speed was compared between cells treated with mitomycin C (migration alone) and untreated cells (migration + proliferation) to determine the combined effect of migration and proliferation in selected clones with low, medium and high proliferative index. The migration speed was determined as described in B. Significance analysed using 2-way ANOVA followed by Bonferroni multiple comparison (* P<0.05, **** P<0.0001) **E** Mean proliferative index (± standard error) of all clones derived from APOBEC3A (A3A), APOBEC3A$_{E72A}$ (A3A$_{E72A}$) or parental HEK293 cell populations. The mean for each population was calculated as the average proliferative index of all clones derived from the respective population as shown in Figure 10A. **F** Mean migration speed (± standard error) of all clones derived from APOBEC3A, APOBEC3A$_{E72A}$ or parental HEK293 cell populations. The mean for each population was calculated as described in E.

4.1.3. APOBEC3A-mediated mutagenesis affects chemotherapy tolerance in HEK293 cells

Proliferation and migration are phenotypes that can directly influence the growth of a tumour. However, APOBEC3A-mediated mutagenesis may also influence other aspects of the mutagenised cells which only become important in patients. Chemotherapy is an important factor that puts selective pressure on fully formed tumours in patients. APOBEC3A-mediated mutagenesis may enable some cells to become more tolerant to chemotherapy. Thus, the HEK293 cell clones were tested for their tolerance towards cisplatin treatment. Cisplatin is a commonly used agent in chemotherapy of various cancer types, among them head and neck squamous cell carcinomas. It causes DNA damage by cross-linking guanidines adjacent on the same DNA strand or located on different DNA strands, which ultimately leads to apoptosis, as well as oxidative stress (reviewed by Dilruba and Kalayda, 2016).

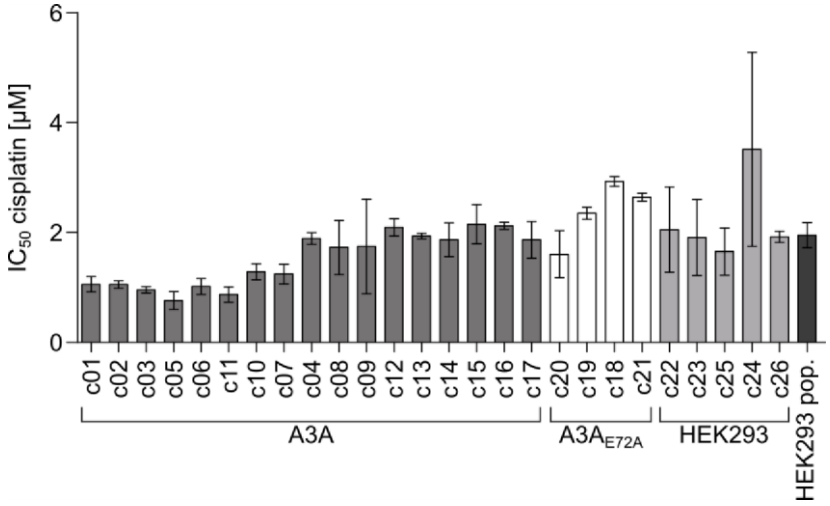

Figure 11 Cisplatin sensitivity of HEK293 single cell clones derived from APOBEC3A-mutagenised populations. The clones had been exposed to APOBEC3A (A3A) or the enzymatically inactive APOBEC3A$_{E72A}$ (A3A$_{E72A}$) during long-term culture, or belonged to the parental HEK293 population. The cisplatin sensitivity is expressed as the half-maximal inhibitory concentration (IC_{50}). The IC_{50} value is determined in a WST-1 viability assay of cells treated with a cisplatin gradient for 2 days. It was calculated from the resulting dose-response curve using non-linear regression. The cisplatin sensitivity was determined in at least two independent experiments, with all clones measured in triplicates in each experiment.

The cisplatin sensitivity of all HEK293 single cell clones was determined, including a population of parental HEK293 cells as a control (Figure 11). The half-maximal inhibitory concentration (IC_{50}) was used as a measure for cisplatin sensitivity. It describes the cisplatin concentration that caused a cell viability halfway between the untreated sample with 100% viability and the sample with the highest cisplatin concentration, where all cells were dead. Thus, a lower IC_{50} indicates a higher sensitivity towards cisplatin.

Without APOBEC3A-mediated mutations, both the clones from the APOBEC3A$_{E72A}$ and the parental populations as well as the population of parental HEK293 cells have an IC_{50} between 1.8 µM and 3.75 µM, with the majority between 2 µM and 3 µM. Out of the 17 APOBEC3A clones, nine clones also show an IC_{50} of approximately 2 µM, while the IC_{50} of the remaining eight APOBEC3A clones is at approximately half that value, between 0.8 µM and 1.3 µM. This indicates that some of the HEK293 clones that had been exposed to APOBEC3A have become more sensitive towards cisplatin.

4.2. Tumour growth of immortalised and transformed HEK293 cells is not altered by APOBEC3A-mediated mutagenesis

Tumour growth *in vivo* is influenced by many different factors, and the conditions are quite different from cell culture. In order to find out whether APOBEC3A-mediated mutagenesis has an effect on tumour growth of immortalised, transformed HEK293

cells, six clones were injected into NOD scid gamma mice as xenografts. As the parental HEK293 cells are already tumorigenic, any effect by APOBEC3A-driven mutations would be modulating. Four mice per clone were injected with $5x10^4$ cells, the other four with $3x10^5$ cells.

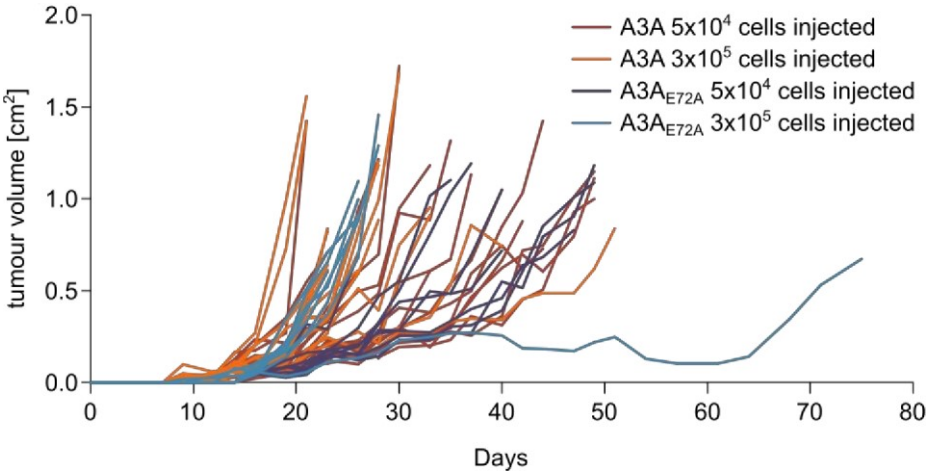

Figure 12 Tumour volume of xenografts formed by APOBEC3A-mutagenised HEK293 clones over time. A total of 24 NSG mice each were injected either with a low cell number of APOBEC3A (A3A) or APOBEC3A$_{E72A}$ clones ($5x10^4$ cells, red and dark blue lines, respectively) or a high cell number of APOBEC3A and APOBEC3A$_{E72A}$ clones ($3x10^5$ cells, orange and light blue lines, respectively) in a 1:1 mixture of PBS and matrigel into the flank. Tumour size was monitored with a calliper, and tumour volume was calculated as $½hd^2$, with d being the tumour diameter and h being its height in centimetres. The experiment was terminated once the tumour reached a diameter of 1.4 cm in any dimension.

All mice developed a visible tumour within approximately one to two weeks after injection of the cells (Figure 12). Once a tumour reached a diameter of 1.4 cm in any dimension, the experiment was terminated. Within seven weeks after injection of the cells, all mice but one had been sacrificed (Figure 12).

4.2.1. Xenograft tumours did not acquire additional APOBEC3A-mediated mutations during tumour growth

In order to allow a reasonable comparison between the results of the tumour characterisation and the results observed in cell culture, it was confirmed that the cells had not acquired additional APOBEC3-mediated mutations during tumour growth. One representative tumour of each clone was tested for expression of endogenous APOBEC3A and APOBEC3B. Neither expression of endogenous APOBEC3A (Figure 13A and B) nor endogenous APOBEC3B (Figure 13B) was observed in any of the tumours. The tumours formed by clones c18 and c19 still retained the basal expression of the enzymatically inactive APOBEC3A$_{E72A}$ transgene (Figure 13A and B). The lack of

functional APOBEC3A and APOBEC3B implies that in the fully formed tumours, no additional APOBEC3-mediated mutations could arise.

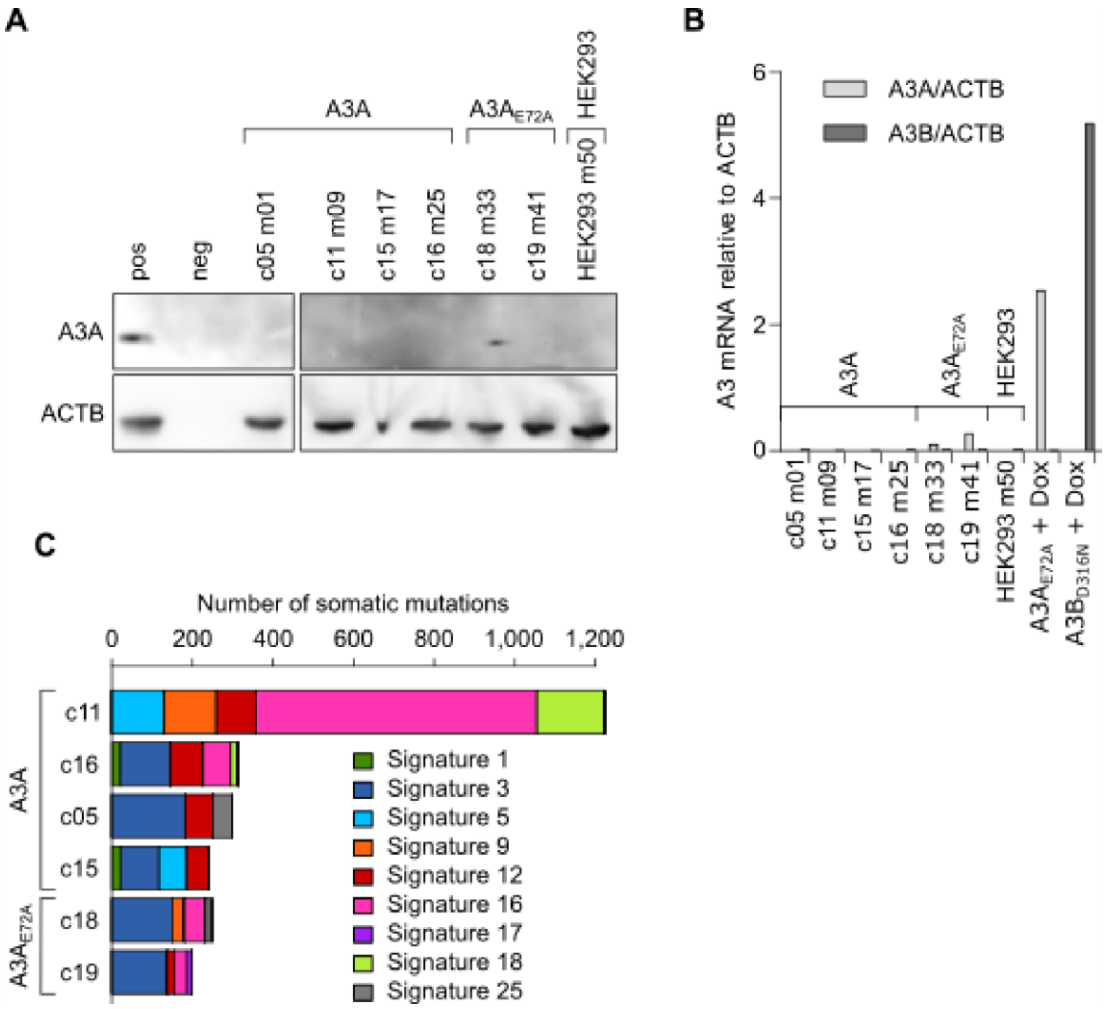

Figure 13 Xenograft tumours did not acquire additional APOBEC3A (A3A)- or APOBEC3B (A3B)-mediated mutations in comparison to clones in cell culture. One representative xenograft tumour per clone is shown. **A** Endogenous APOBEC3A protein expression in xenograft tumours. Tissue lysates in 1% SDS where separated in an SDS-PAGE and transferred to a nitrocellulose membrane in a Western blot. The membrane was probed with a polyclonal rabbit anti-APOBEC3G antibody (Sigma) which cross-reacts with APOBEC3A as primary and a goat anti-rabbit HRP conjugate as secondary antibody. ACTB served as a loading control. **B** Endogenous APOBEC3A and APOBEC3B mRNA expression in xenograft tumours. Total mRNA was extracted and reverse transcribed into cDNA. APOBEC3A and APOBEC3B expression was determined in a gene-specific qPCR and normalised to ACTB mRNA expression. Dox-induced expression of the APOBEC3A$_{E72A}$ and APOBEC3B$_{D316N}$ transgenes was included as positive controls. **C** Comparison of whole genome sequences between cell culture clones and respective mouse xenografts. Total DNA extracted from clones c05, c11, c15, c16, c18 and c19 in cell culture as well as one representative xenograft per clone were whole genome sequenced. Single nucleotide variants in the xenografts were called using the cell culture clones as reference. An extraction of mutational signatures was performed based on the single nucleotide variants found in the xenografts. The stacked bars represent the number of somatic mutations ascribed to each of the mutational signatures detected in the xenografts.

However, the absence of endogenous APOBEC3A and APOBEC3B expression in the fully formed tumour does not exclude previous activity during tumour growth, and previous results in this study suggest that loss of APOBEC3A expression and/or activity is a frequent consequence of APOBEC3A-mediated mutagenesis. In order to find out whether additional mutagenesis occurred in the HEK293 cell clones during tumour growth in mice, the whole genome of one representative tumour of each clone was sequenced. The resulting sequences of the tumours were compared to the whole genome sequences of the HEK293 clones in cell culture to find out if any additional APOBEC3-mediated mutations arose during tumour growth. None of the clones acquired additional APOBEC3-mediated mutations during tumour growth, as shown by the absence of mutational signatures 2 and 13 (Figure 13C). However, all clones acquired additional mutations by other mutagenic processes. This indicates that the HEK293 clones are genetically unstable, and subject to background mutagenic processes independent of editing by APOBEC3A. With the exception of clone c11, the APOBEC3A clones (c05, c15 and c16) as well as the APOBEC3A$_{E72A}$ clones (c18 and c19) acquired between 200 and 300 additional mutations. The majority of these mutations correspond to mutational signatures 3, 12, 16 and 25. Clone c11 acquired four to six times the number of additional mutations present in the other clones. This suggests that an additional mutagenic process was switched on in this clone. While the mutational signatures present in clone c11 are the same as the ones detected in the other clones, they are responsible for a larger number of mutations. Signature 16 in particular caused roughly 700 mutations, more than half the mutational load acquired by clone c11 during tumour growth in mice.

4.2.2. Tumour growth of APOBEC3A mutagenised cells is unchanged *in vivo*

The xenograft tumours derived from APOBEC3A-mutagenised HEK293 clones were compared to the controls that had been exposed to the enzymatically inactive APOBEC3A$_{E72A}$ variant concerning parameters that could describe or influence tumour growth. This includes tumour morphology, proliferation, necrosis, and survival time.

The tumours were xenografts of HEK293 clones that already showed interclonal phenotypic differences *in vitro*, even though the means of the populations they were derived from showed no difference. Furthermore, as four out of the six clones had undergone APOBEC3A-mediated mutagenesis, the clones presumably all carried their own distinct set of mutations. Thus, the morphological heterogeneity observed in the HE stains of the xenograft tumours was not unexpected (Figure 14). For instance, some of the tumours had very defined borders, while others had more jagged outlines; some tumour tissues looked very homogeneous and consisted solely of HEK293 cells, whereas

others also contained mouse cells and were thus much more heterogeneous. It was, however, also observed that the morphology of tumours derived from the same clone varied greatly, so that it was not possible to assign a clear morphological phenotype to each clone.

Figure 14 Morphological heterogeneity both between and within the xenograft tumours formed by the different HEK293 clones visualised by HE-staining of tissue sections at 100x magnification. Each of the six clones was injected into eight mice. Each row depicts the tumours derived from the same clone, labelled with the identity of the mouse it was grown in.

Proliferation, necrosis and tumour growth of the xenografts were characterised in more detail. Proliferation was determined by immunohistochemistry for the proliferation marker Ki-67 and expressed as the proliferative index, i.e. the percentage of Ki-67 positive cells out of all living cells. The necrotic area was estimated from the HE stains, and expressed as a percentage of the whole tumour area. Survival time served as a measure for tumour growth. The experiment had to be terminated once the tumour reached a diameter of 1.4 cm in any dimension. Thus, the survival time represents the time it took the tumour to grow to this size and so indirectly reflects tumour growth.

The differences in proliferative index between tumours derived from different clones are not significant (Figure 15A). Due to the small sample sizes *in vivo*, the standard deviations were very high. The necrotic area shows a very high variability even between tumours formed by the same clone (Figure 15B). The necrotic area of APOBEC3A-mutagenised clones and the APOBEC3A$_{E72A}$ control clones did not show a significant difference. Overall, this suggests that neither cell proliferation nor necrosis of

immortalised, transformed HEK293 cells is affected by APOBEC3A-mediated mutagenesis.

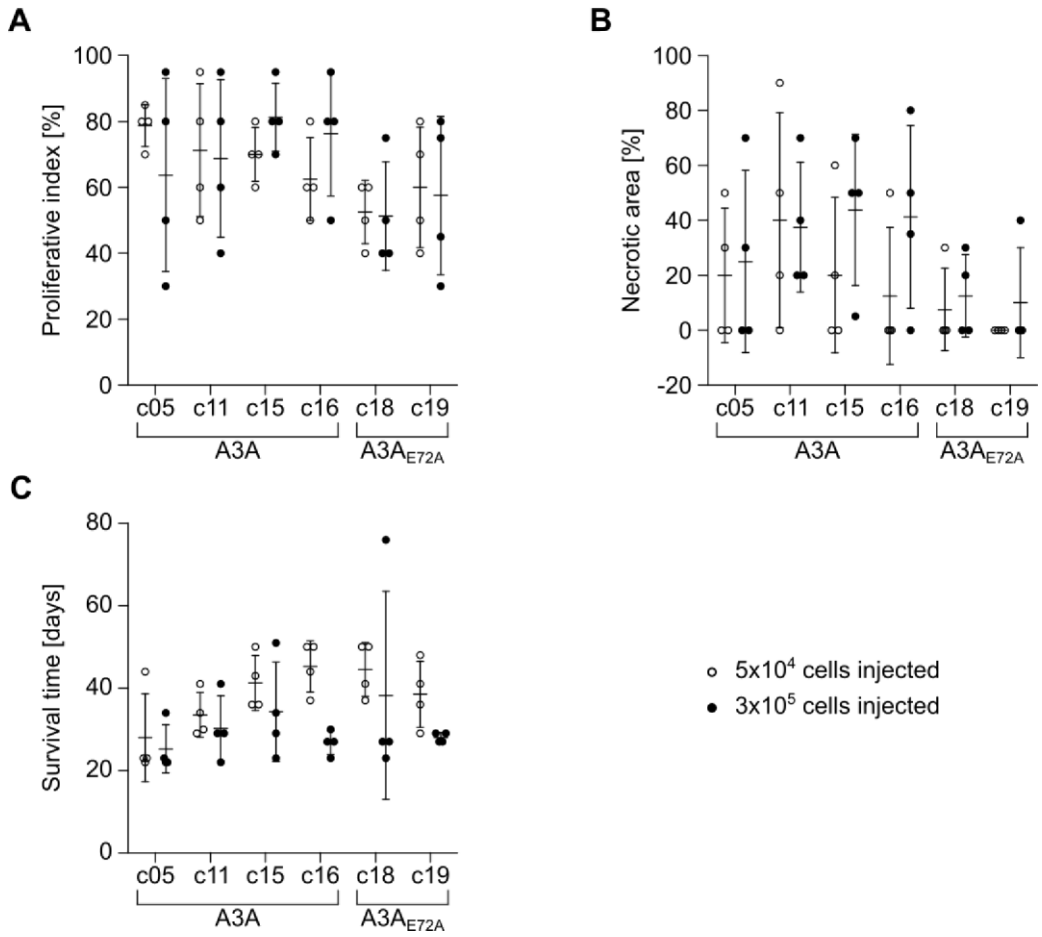

Figure 15 APOBEC3A (A3A)-mediated mutagenesis does not affect proliferation, necrosis or tumour growth of HEK293 xenografts. **A** The proliferative index (mean ± standard deviation) of all xenograft tumours was determined by performing an immunohistochemistry stain of formalin-fixed paraffin-embedded tissue sections for the proliferation marker Ki-67, and counterstained with haematoxylin and eosin. The proliferative index is the percentage of Ki-67 positive cells out of all living cells. It was estimated by Jochen Heß. **B** The necrotic area (mean ± standard deviation) as a percentage of the entire tumour area was estimated in haematoxylin and eosin-stained tumour sections of all xenografts by Jochen Heß. **C** The survival time (mean ± standard deviation) of each mouse separated by number of injected cells. It measures the time until the tumour reached a diameter of 1.4 cm in any dimension, at which point the experiment was terminated. Thus, survival time reflects the time it took the tumour to reach a certain size and can thus be used as a measure of tumour growth.

No difference was observed between the survival times of clones derived from populations with and without the APOBEC3A mutational signature (Figure 15C). Generally, the mice which were injected with a smaller number of cells survived longer than the mice which received the higher cell number, as it took some more cell cycles to form a tumour of the same size. This is why the tumours grown from a lower and a higher number of injected cells are depicted separately, as the starting cell number may

influence the final survival time, whereas an influence on other parameters such as proliferation or necrosis in the tumour at the time of harvest are not influenced by starting cell number. The difference in survival time between mice injected with a high or low number of cells was not found to be statistically significant. A higher proliferative tendency did not correspond with a decrease in survival time.

Overall, no difference was observed between APOBEC3A-mutagenised and control xenografts in either proliferative index (Figure 15A), necrosis (Figure 15B) or survival time (Figure 15C). All of the clones showed a high variability in their phenotypes *in vivo*. In summary, these results suggest that APOBEC3A-mediated mutagenesis does not have an effect on tumour growth of already immortalised and transformed HEK293 cells. This implies that APOBEC3A-driven mutagenesis does not play a role in the growth of fully formed cancer.

4.3. A combination of APOBEC3A and APOBEC3B expression and APOBEC3 mutational signature can be used to stratify an HNSCC patient cohort

APOBEC3-mediated mutations can make up a large fraction of the total mutational load in various cancer types. Furthermore, both APOBEC3A and APOBEC3B expression has been observed in some of the tumours containing the APOBEC3 mutational signatures. One cancer type where this is the case is head and neck squamous cell carcinoma (HNSCC). It is, however, not clear which, if any, of these parameters influence the outcome for the patients. Thus, expression of APOBEC3A and APOBEC3B as well as total mutational load and contribution of APOBEC3-mediated mutations to the cancer genome were determined in a HNSCC patient cohort, and their influence on outcome was assessed.

The patient samples in this work were kindly provided by Jochen Heß. They were collected in the prospective study HIPO-POP019 to elucidate the molecular mechanisms of treatment failure in HNSCC patients. This study includes a total of 90 HNSCC samples consisting of 88 primary tumours and 2 metastases along with healthy tissue used as reference from 88 HNSCC patients. All tumour samples contain a minimum of 50% tumour tissue. A total of 11 samples were excluded from the following analyses due to either lack of sequencing data, low mRNA extract quality, accidental sample swapping during whole exome sequencing or different enrichment kits being used for a sample and its cognate control. The two metastases were also excluded. The remaining 79 samples were included for all following analyses.

4.3.1. A subgroup of HNSCC patients carries the APOBEC3 mutational signatures in the tumour genome

Whole exome sequencing of all the tumours in this study was performed by the DKFZ Genomics & Proteomics Core Facility. The whole exome sequences were compared to the whole exome of the corresponding healthy tissue sample to identify the mutations present in the tumour. The mutations were called and the extraction of mutational signatures was done by Marc Zapatka and Mario Hlevnjak. The mutational signature extraction showed that the APOBEC3 mutational signatures 2 and 13 are present in a subset of patients, and together contribute up to 76% of the total mutational load in this cohort (Figure 16). Signature 4, which is associated with smoking, was also found to a lesser degree and contributes up to 43% of the total mutational load, despite most of the patients being heavy smokers. Signature 29, which is associated with tobacco chewing rather than smoking, was also observed in a number of samples and contributed up to nearly 22% of the total mutational load. Signature 16 was found to be very prominent in this patient cohort, contributing up to 75% of the total mutations. It was originally found in liver cancer, and its aetiology remains unknown.

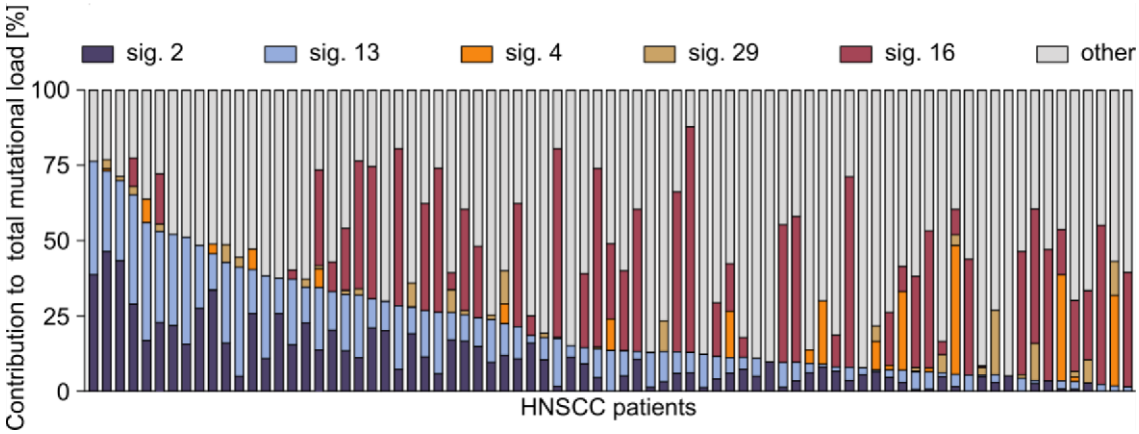

Figure 16 Mutational signatures in HNSCC patients. The bars represent the contribution of mutational signatures associated with APOBEC3 (signatures 2 and 13) and tobacco (signatures 4 and 29) as well as signature 16 of unknown aetiology to the total mutational load of HNSCC patients in percent. All remaining mutational signatures not separately included were combined into the "other" category. All signatures were extracted from whole exome sequencing of DNA extracts from samples consisting of at least 50% tumour cells. The samples were ordered by total contribution of the APOBEC3 mutational signatures.

4.3.2. APOBEC3A and APOBEC3B mRNA expression in a cohort of HNSCC patients

The expression of both APOBEC3A and APOBEC3B mRNA was determined in all available patient samples using a gene-specific qPCR. Both APOBEC3A (Figure 17A) and APOBEC3B (Figure 17B) were found to be expressed to different degrees in a subset of HNSCC patients.

The presence of the APOBEC3 mutational signature and a simultaneous absence of APOBEC3A and APOBEC3B expression was observed in some of the HNSCC tumours. This reflects the situation in the HEK293 cells, which lost APOBEC3A activity by different mechanisms after acquiring the mutational signature.

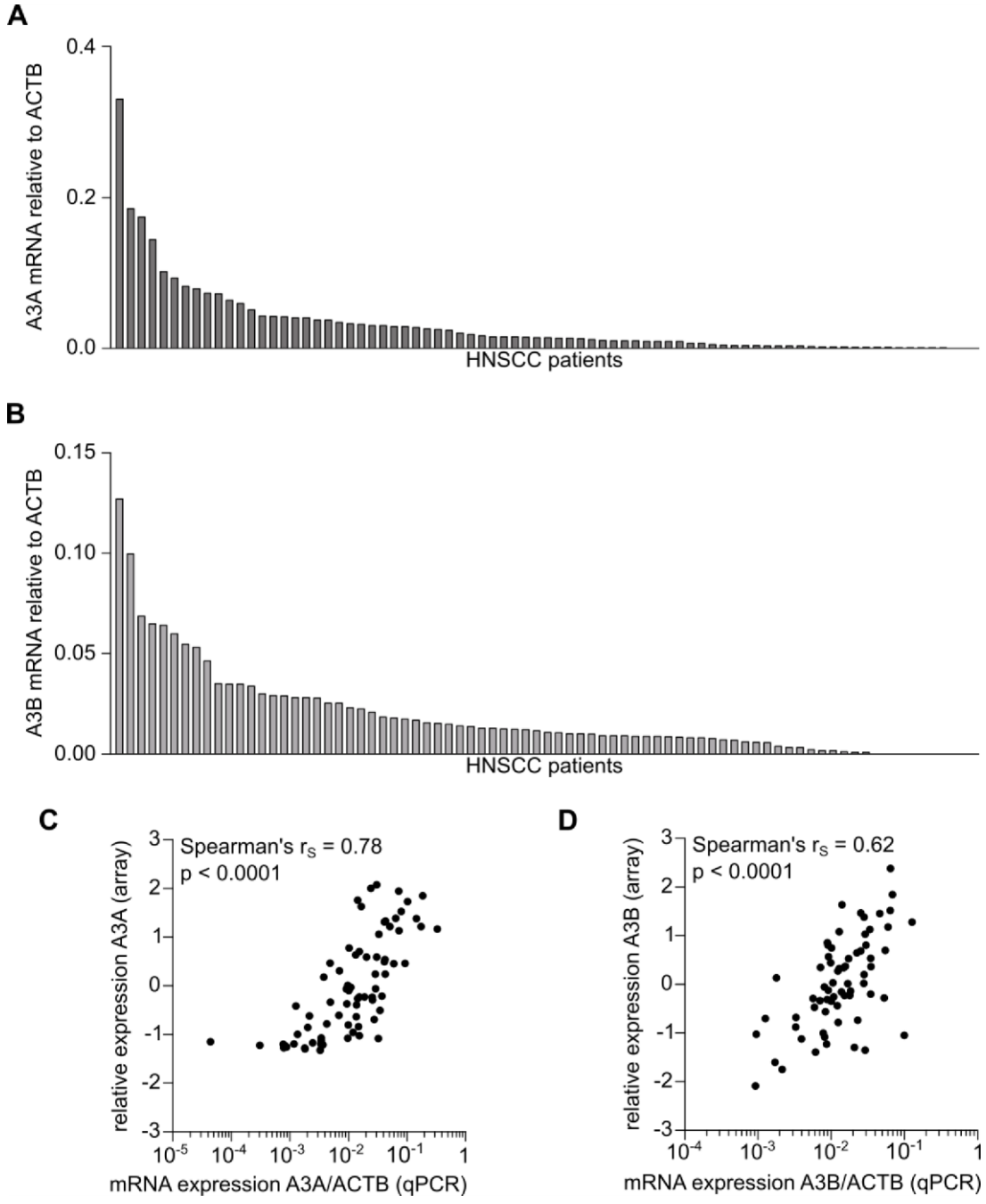

Figure 17 APOBEC3A and APOBEC3B expression in HNSCC patients. **A** APOBEC3A (A3A) and **B** APOBEC3B (A3B) mRNA expression ordered by expression level in HNSCC patients. Total mRNA was extracted from tissue samples consisting of at least 50% tumour cells and reverse transcribed into cDNA. APOBEC3A and APOBEC3B expression was determined in a gene-specific qPCR and normalised to ACTB mRNA expression. **C and D** Correlation of **C** APOBEC3A and **D** APOBEC3B expression as measured in the same samples by gene-specific qPCR relative to ACTB and by microarray carried out in the context of the HIPO-POP 019 study. Correlation between qPCR and microarray results was determined using Spearman's rank correlation.

A microarray expression profile was also available for all HNSCC samples. The expression levels of APOBEC3A and APOBEC3B obtained by qPCR were compared to the results obtained in the microarray expression analysis (Figure 17C and D). The samples were ranked according to their respective expression levels as determined by each method, and the ranks of the two methods were compared for each sample in a Spearman's rank correlation. The correlation between the qPCR results and the microarray results is better for APOBEC3A than for APOBEC3B, but it was highly significant for both APOBEC3A and APOBEC3B expression. This validates the quantification of transcript levels using an independent method.

4.3.3. APOBEC3 expression correlates with clinical parameters

For correlation of APOBEC3A and APOBEC3B expression levels with clinical parameters, the HNSCC samples were stratified into two groups according to APOBEC3A or APOBEC3B expression. The groups of interest were delimited at the median: samples with an mRNA expression value greater than the median ($APOBEC3A_{high}$ and $APOBEC3B_{high}$) were compared with samples with expression levels equal to or below the median ($APOBEC3A_{low}$ and $APOBEC3B_{low}$).

In a contingency analysis (Table 11), APOBEC3A expression was found to be associated with the pathological grade of the tumours and with extracapsular spread (ECS). Extracapsular spread is the infiltration of cancer cells beyond the nodal capsule of a metastatic lymph node into the perinodal fatty tissue (Lagarde et al., 2006; Metzger et al., 2009) and was thus only assessed in tumours with lymph node involvement. However, the mean APOBEC3A expression was the same in tumours both with and without extracapsular spread (Figure 18C). Furthermore, APOBEC3A expression levels did not differ between HPV-driven and non-HPV driven tumours (Figure 18A). The contingency table shows that APOBEC3B expression correlates with the HPV status of the tumours as well as with tumour location and gender of the patient (Table 11). The mean APOBEC3B expression level was also found to be significantly higher in HPV-driven tumours than in non-HPV driven ones (Figure 18B). Tumours with lymph node involvement showing extracapsular spread also showed higher APOBEC3B expression levels than tumours without extracapsular spread (Figure 18D). This association was, however, not found in the contingency analysis (Table 11).

Table 11 Association between APOBEC3A and APOBEC3B mRNA expression levels and clinical parameters. APOBEC3A and APOBEC3B expression levels were each delimited into interest groups at the median. Statistical significance was determined by Fisher's exact test for all parameters except tumour location, where it was determined with the chi-square test.

Feature	**Category**	**APOBEC3A$_{low}$**	**APOBEC3A$_{high}$**	**P-value**	**APOBEC3B$_{low}$**	**APOBEC3B$_{high}$**	**P-value**
Gender	Female	8	9	0.7909	13	4	**0.0278**
	Male	34	32		29	37	
Age [years]	≤61.44	22	16	0.2730	17	21	0.3817
	>61.44	20	25		25	20	
Tumour size	T1-2	20	20	1.0000	16	24	0.0802
	T3-4	22	21		26	17	
Lymph node	N0	17	18	0.8257	21	14	0.1839
	N1-3	25	23		21	27	
Pathological grading	G1-2	22	31	**0.0396**	30	23	0.1745
	G3	20	10		12	18	
Tumour location	Larynx and hypopharynx	8	10	0.061	12	6	**0.0274**
	Nasopharynx	9	1		6	4	
	Oral cavity	9	13		14	8	
	Oropharynx	16	17		10	23	
Alcohol	No	23	18	0.3828	12	12	1.0000
	Yes	19	23		30	29	
Tobacco	No	16	8	0.0898	12	12	1.0000
	Yes	26	33		30	29	
HPV status	Non-HPV driven	29	30	0.8095	35	24	**0.0161**
	HPV-driven	13	11		7	17	
Extracapsular spread	Negative	13	4	**0.0168**	9	8	0.3775
	Positive	12	19		12	19	

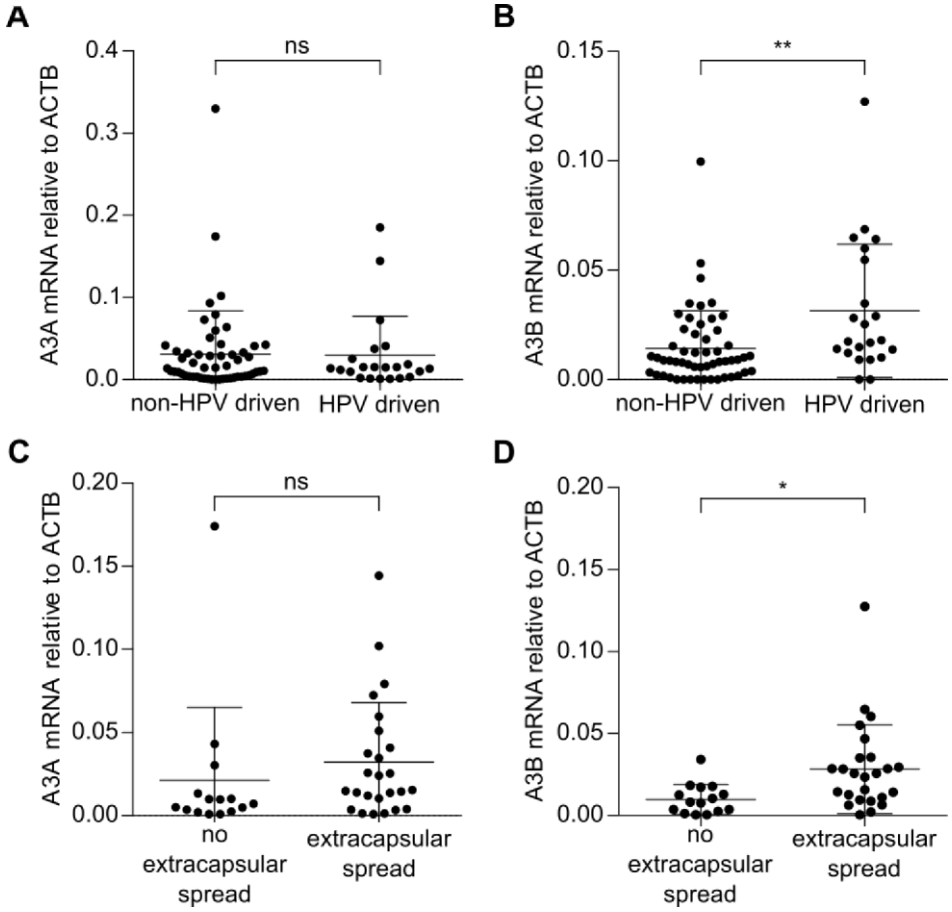

Figure 18 Association between APOBEC3A (A3A) and APOBEC3B (A3B) mRNA expression levels, HPV status and extracapsular spread in head and neck squamous cell carcinomas. **A and B** Expression levels of APOBEC3A (**A**) and APOBEC3B (**B**) in non-HPV-driven and HPV-driven tumours. **C and D** Expression levels of APOBEC3A (**C**) and APOBEC3B (**D**) in tumours with lymph node involvement without or with extracapsular spread. Significance was determined in a t-test (* P<0.05; ** P<0.01; ns not significant).

4.3.4. Enhanced APOBEC3A expression is associated with an increased presence of APOBEC3 mutational signature

The HNSCC tumours carry the APOBEC3 mutational signatures (Figure 16) and show expression of both APOBEC3A and APOBEC3B to different extents (Figure 17A and B). In order to find out whether the expression of APOBEC3A and/or APOBEC3B is associated with the mutations found in head and neck tumours, the mutational load was compared between patients with high and low APOBEC3A and APOBEC3B expression levels.

The total number of mutations was found to be significantly increased in the patient group with high APOBEC3A expression (Figure 19A). While the contribution of APOBEC3-mediated mutations to the total mutational load in percent is not associated with APOBEC3A expression (Figure 19C), samples with a high APOBEC3A expression were associated with a higher number of mutations matching the APOBEC3 mutational signatures (Figure 19E).

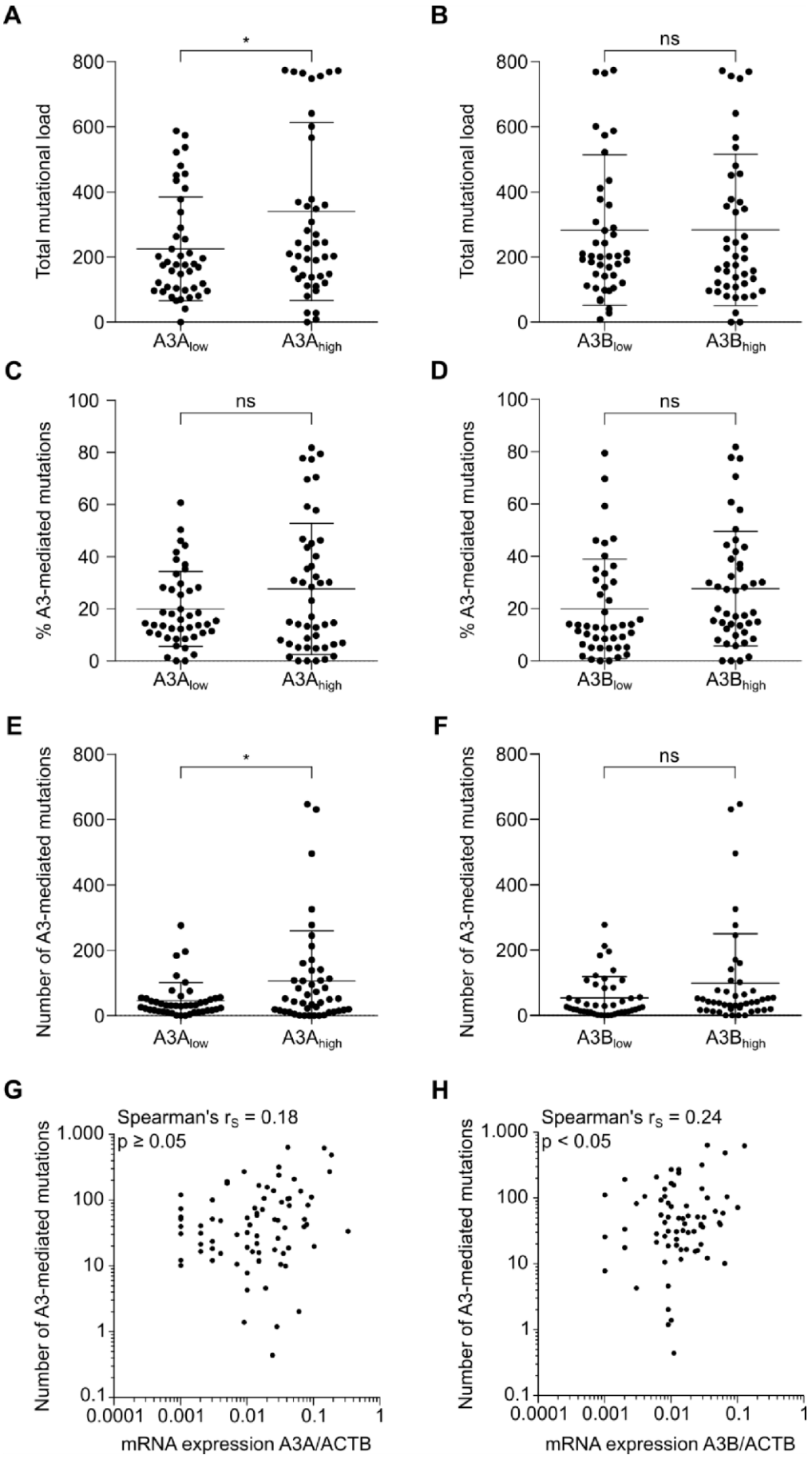
A
B
C
D
E
F
G
H
Total mutational load
% A3-mediated mutations
Number of A3-mediated mutations
*
ns
A3A$_{low}$
A3A$_{high}$
A3B$_{low}$
A3B$_{high}$
Spearman's r_S = 0.18
p ≥ 0.05
Spearman's r_S = 0.24
p < 0.05
mRNA expression A3A/ACTB
mRNA expression A3B/ACTB

Figure 19 Correlation of APOBEC3A (A3A) and APOBEC3B (A3B) expression with the mutational landscape in HNSCC patients. The mutational landscape of the HNSCC patient samples was extracted from whole exome sequencing data. APOBEC3A and APOBEC3B mRNA expression levels were quantified using gene-specific qPCR and normalised to ACTB mRNA expression. **A - F** Patients were stratified into groups with APOBEC3A or APOBEC3B expression greater than the median value ($APOBEC3A_{high}$, $APOBEC3B_{high}$) and equal to or lower than the median value ($APOBEC3A_{low}$, $APOBEC3B_{low}$). **A and B** Total number of mutations found in patients with high and low expression levels of either APOBEC3A (**A**) or APOBEC3B (**B**). **C and D** Percentage of mutations matching the APOBEC3 (A3) mutational signatures out of all mutations in the patients with high or low expression levels of either APOBEC3A (**C**) or APOBEC3B (**D**). **E and F** Number of single nucleotide exchanges matching the APOBEC3 mutational signatures in the patients with high or low expression levels of either APOBEC3A (**E**) or APOBEC3B (**F**). Statistical significance was determined in a t-test for figures A-F (* $P<0.05$, ns not significant). **G and H** Scatter plot of the number of mutations matching the APOBEC3 mutational signatures found in the patients plotted against the mRNA expression levels of APOBEC3A (**G**) or APOBEC3B (**H**) normalised to ACTB. Correlation between APOBEC3 mRNA expression and the number of APOBEC3 signature mutations was determined using Spearman's rank correlation.

The level of APOBEC3B expression was not associated with any of these parameters (Figure 19B, D and F). A scatter plot of the APOBEC3A expression against the number of APOBEC3-mediated mutations for each individual sample showed, however, that the correlation is in this case not significant (Figure 19G), whereas the same plot with APOBEC3B expression showed a weak, but significant correlation (Figure 19H).

4.3.5. APOBEC3 expression in combination with APOBEC3 mutational signature and total mutational load correlates with survival

APOBEC3-mediated mutagenesis plays a large role in shaping the genomes of some HNSCC patients. Yet, no difference in overall progression-free survival was observed between patients with high or low APOBEC3A (Figure 20A) or APOBEC3B expression levels (Figure 20B). Additional stratifications into patients with a high and low total mutational load (Figure 20C) or a high and low contribution of APOBEC3-mediated mutations to the cancer genome (Figure 20D) also did not result in a difference in progression-free survival. Neither expression of APOBEC3A or APOBEC3B, nor the total number of mutations or the contribution of APOBEC3-mediated mutagenesis to the final landscape of the cancer genome as single parameters resulted in any difference in progression-free survival for HNSCC patients.

This approach, however, does not take time into account. It does not consider whether APOBEC3-mediated mutagenesis occurred in the past, was ongoing at the time of tumour resection or started only a short time previously. Thus, an approach was chosen that combines the APOBEC3A and APOBEC3B expression levels with the contribution of APOBEC3-mediated mutations to the total mutational load and the total number of

mutations for each patient. The aim was to identify patterns in the parameter values, and form groups of patients with similar patterns.

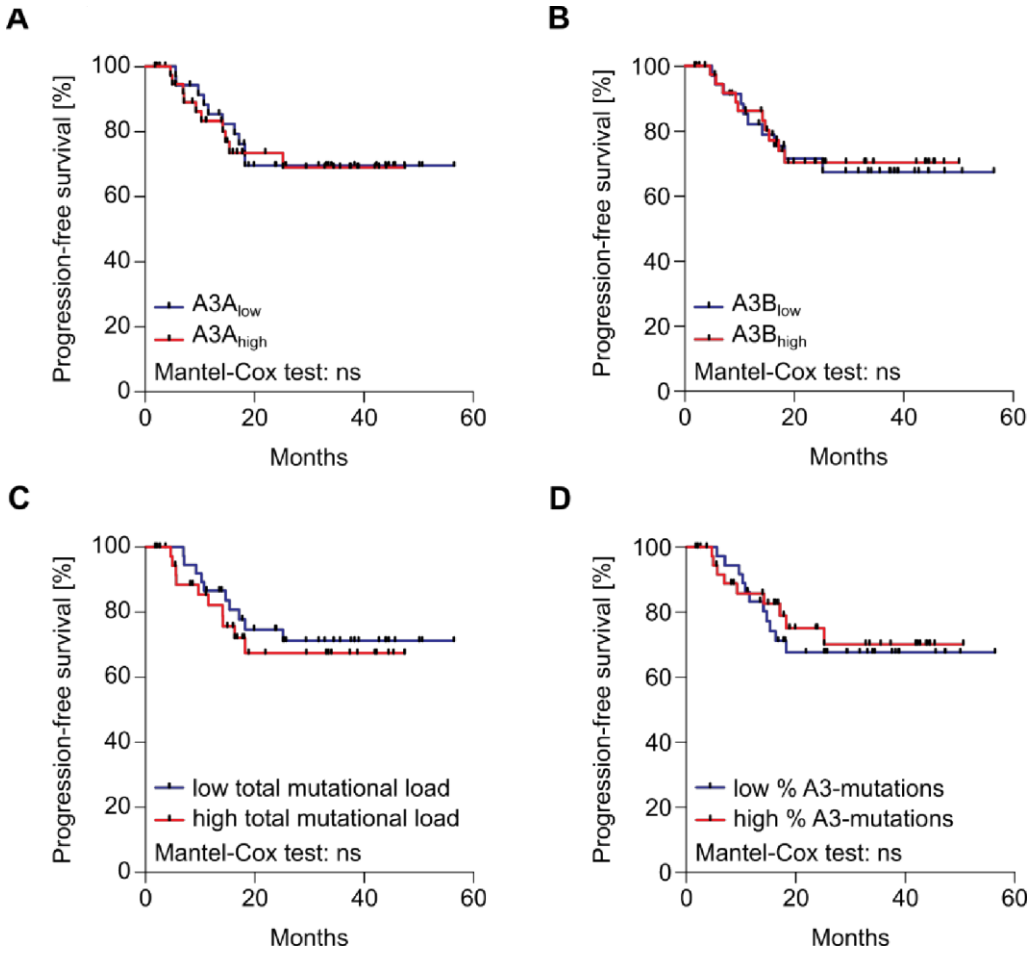

Figure 20 Progression-free survival shows no difference between patients with high or low expression of APOBEC3A (A3A) and APOBEC3B (A3B), total mutational load and fraction of mutations due to APOBEC3-mediated mutagenesis. Patients were stratified into groups with values greater than the median and values equal to or lower than the median. Differences in survival were assessed using the Mantel-Cox (log-rank) test (ns not significant). **A and B** APOBEC3A (**A**) and APOBEC3B (**B**) mRNA expression levels were quantified using gene-specific qPCR and normalised to ACTB mRNA expression. **C and D** Total mutational load (**C**) and the percentage of APOBEC3-mediated mutations (**D**) were determined from whole exome sequences of tumours in comparison to matched healthy tissue.

A principal component analysis of the four parameters mentioned was performed, and the results were visualised and clustered in groups in a heatmap (Figure 21A). Out of the 16 possible patterns that can be formed by combining four parameters, nine different patterns were identified in the HIPO-POP019 study. This suggests that the groups identified by the principal component analysis are not just random patterns. If this holds true, identical patterns should be identified in an independent dataset. Indeed, the same analysis applied to 500 HNSCC samples extracted from The Cancer Genome Atlas (TCGA) found the same patterns (Figure 22).

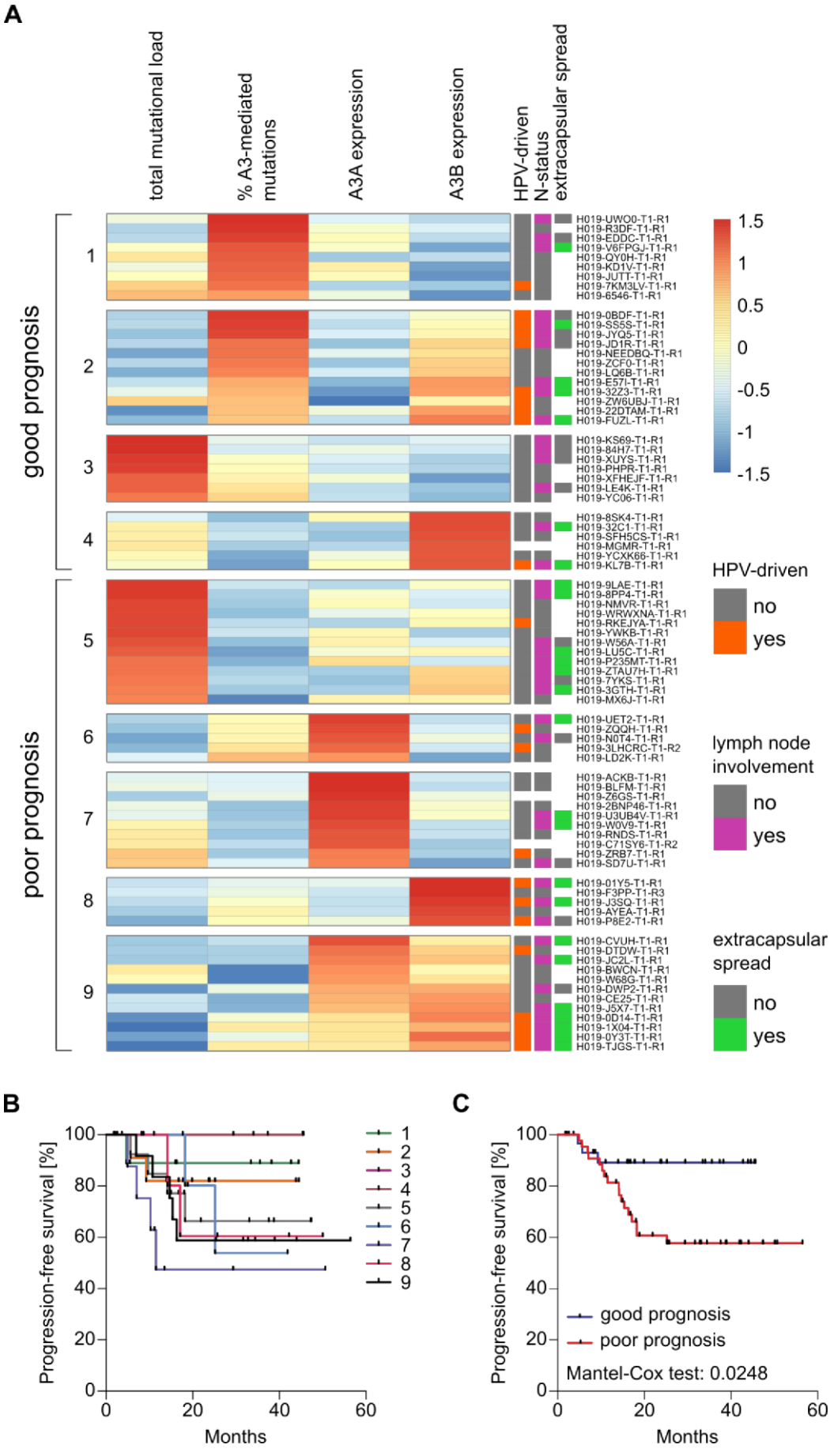
A
total mutational load
% A3-mediated mutations
A3A expression
A3B expression
HPV-driven
N-status
extracapsular spread
good prognosis
poor prognosis
1
2
3
4
5
6
7
8
9
1.5
1
0.5
0
-0.5
-1
-1.5
HPV-driven
no
yes
lymph node involvement
no
yes
extracapsular spread
no
yes
B
Progression-free survival [%]
Months
C
Progression-free survival [%]
Months
good prognosis
poor prognosis
Mantel-Cox test: 0.0248

Figure 21 Subgroups of HNSCC patients in the HIPO-POP019 study based on a combination of genomic landscape and APOBEC3 expression data show a difference in progression-free survival. Stratification and survival of patients in the HIPO-POP019 study with good and poor prognosis based on APOBEC3A (A3A) and APOBEC3B (A3B) expression, total mutational load and contribution of APOBEC3-mediated mutagenesis to total mutational load. **A** Heatmap of principal component analysis of patient samples according to total mutational load, the contribution of APOBEC3A-mediated mutations to the total number of mutations in percent identified in the whole exome sequence, and standardised APOBEC3A and APOBEC3B mRNA expression levels relative to ACTB expression as determined by gene-specific qPCR. All parameters were standardised by subtracting the mean value from the value of each sample and dividing the difference by the standard deviation, thus converteing the values into z-scores. These z-scores were used as the basis for a principal component analysis. The colour range visualises the values of the principal component analysis assigned to each sample and parameter. The samples were grouped in a heatmap according to similarity using unsupervised clustering. HPV-status, lymph node involvement and extracapsular spread status of each patient were additionally included. Rows are centred; unit variance scaling is applied to rows. Rows are clustered using correlation distance and average linkage. 79 rows, 4 columns. **B** Progression-free survival of patients assigned to each of the different groups defined in A. **C** Progression-free survival of the groups defined in A with good or poor prognosis. Differences in survival were assessed using the Mantel-Cox (log-rank) test.

As the heatmap represents the principal components rather than the absolute values for all the parameters, the colours represent the values each in relation to the others. The different groups identified in the principal component analysis represent two things. The genomic data represents the different ways the tumours acquired their mutational landscapes during tumour development. The expression data represents the different states of gene expression at the time of tumour resection. While the mutational signatures reflect the mutagenic forces that were active in the past, the APOBEC3 expression status reflects the APOBEC3 mutagenic activity in the tumour in the tumour at time of resection

The genetic landscapes of the tumours in group 1 were shaped by APOBEC3-mediated mutagenesis as the major mutagenic force in the past, as suggested by the high percentage of APOBEC3 signature mutations. Other mutagenic processes independent of APOBEC3 hardly contributed to the total mutational load. The low to moderate expression levels of APOBEC3A and B indicate that little to no additional APOBEC3-mediated mutagenesis is active in these tumours. The situation in group 2 is very similar, but the moderate APOBEC3B expression suggests that a moderate level of APOBEC3B-mediated mutagenesis is ongoing in these tumours. The cancer genomes in group 3 carry a large number of mutations that were caused by both APOBEC3-mediated mutagenesis as well as other mutagenic processes, as suggested by the moderate contribution of APOBEC3 signature mutations to the total mutational load. The low expression of both APOBEC3A and B indicates that no additional APOBEC3-mediated mutagenesis is happening in the final tumour. In group 4, other mutagenic processes independent of APOBEC3 shaped the cancer genome, as indicated by the low

contribution of APOBECC3 signature mutations to the total mutational load. The high APOBEC3B expression levels further suggest that APOBEC3B-mediated mutagenesis is ongoing. Group 5 is characterised by the high total mutational load, which is entirely due to other mutagenic processes independent of APOBEC3, as indicated by the low contribution on APOBEC3-mediated mutations to the total mutations. Moderate expression levels of either APOBEC3A or B suggest ongoing APOBEC3-mediated mutagenesis in the tumour at time of resection. The moderate percentage of APOBEC3 signature mutations in group 6 suggests that both APOBEC3-mediated mutagenesis as well as other mutagenic processes shaped the genomes of these cancers. In contrast, most of the mutations in group 7 were caused by non-APOBEC3 mutagenic processes. The high APOBEC3A expression levels in groups 6 and 7 indicate that APOBEC3-mediated mutagenesis is ongoing in these tumours at the time of resection. Only a low to moderate percentage of the total mutations in groups 8 and 9 are due to APOBEC3-mediated mutagenesis, which indicates that other processes were the main mutagenic force in these tumours in the past. The high APOBEC3B expression levels in group 8 and the moderate to high expression of both APOBEC3A and APOBEC3B in group 9 indicate ongoing APOBEC3-mediated mutagenesis in these tumours. The low contribution of APOBEC3-mediated mutations to the total mutational load in groups 6, 7, 8 and 9 suggests that APOBEC3A and/or APOBEC3B expression has not been going on long enough to create a mutational signature in these tumours.

The groups identified in the principal component analysis show differences in progression-free survival (Figure 21B). Combining groups 1 to 4 and groups 5 to 9 into two categories with good and poor prognosis, respectively, results in a statistically significant difference in progression-free survival between these two categories (Figure 21C). In addition, the HPV-status, lymph node involvement (N-status) and extracapsular spread status is indicated for each patient (Figure 21A). The frequency of HPV-driven tumours, tumours with lymph node involvement and tumours with extracapsular spread are the same in the categories with good and poor prognosis (Fisher's exact test, HPV-status $P = 1.0$, N-status $P = 0.64$, extracapsular spread $P = 0.09$ and Figure 21A), implying that this stratification is independent of these parameters.

In summary, the principal component analysis using a combination of genomic data and APOBEC3 expression levels at time of resection allows a clear stratification of patients into groups with good and poor prognosis (Figure 21C) when the single parameters failed to do so (Figure 20).

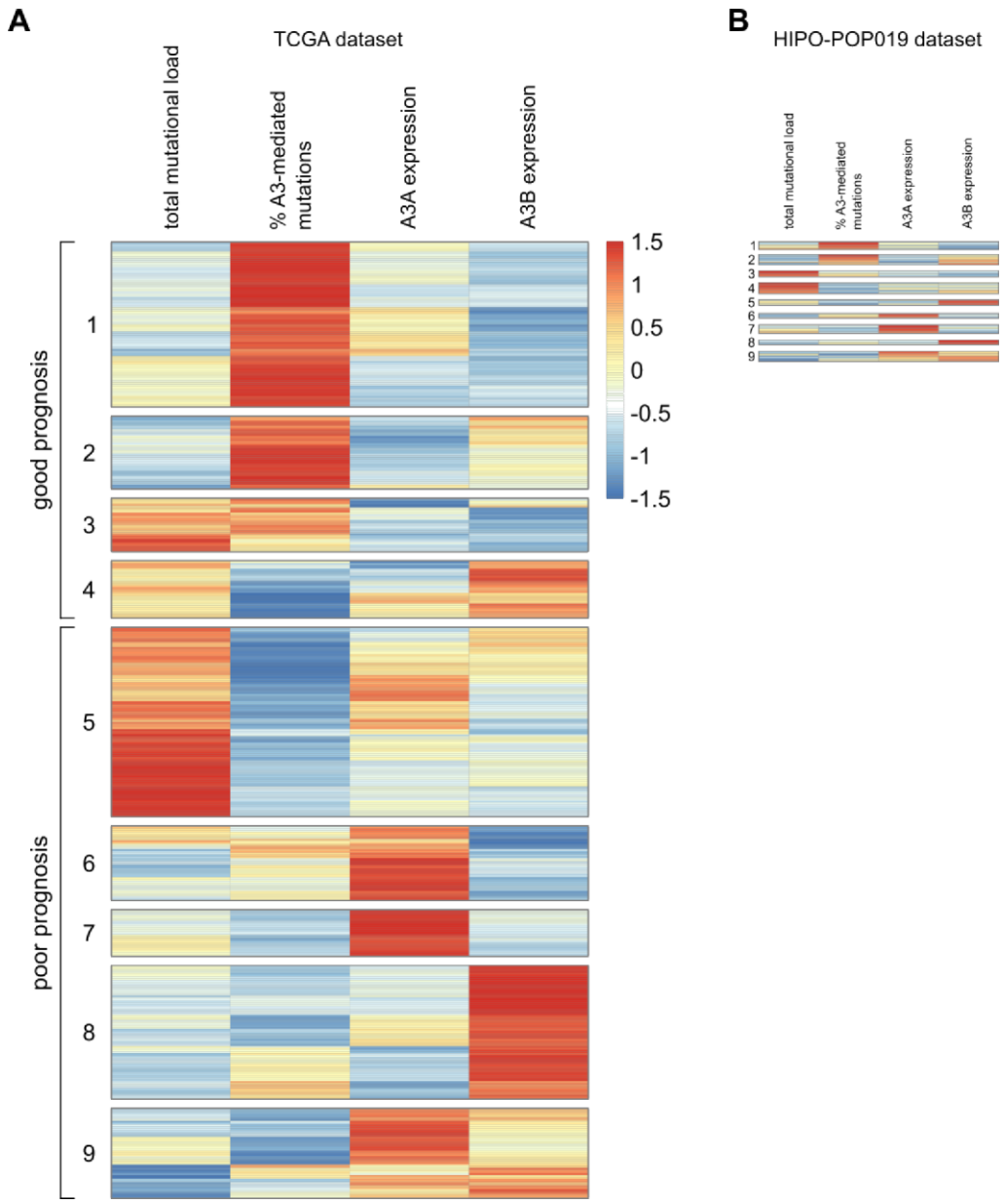

Figure 22 The same subgroups of HNSCC patients identified in the HIPO-POP019 study are also observed in the TCGA database. Stratification of 500 HNSCC patients in the TCGA database into groups with good and poor prognosis based on APOBEC3A (A3A) and APOBEC3B (A3B) expression, total mutational load and contribution of APOBEC3-mediated mutagenesis to total mutational load. **A** Heatmap of principal component analysis of TCGA patient data according to total mutational load, the contribution of APOBEC3-mediated mutations to the total number of mutations in percent identified in the whole exome sequence, and relative APOBEC3A and APOBEC3B mRNA expression levels as determined by RNAseq. All parameters were standardised by subtracting the mean value from the value of each sample and dividing the difference by the standard deviation, thus converted the values into z-scores. These z-scores were used as the basis for a principal component analysis. The colour range visualises the values of the principal component analysis assigned to each sample and parameter. The samples were grouped in a heatmap according to similarity using unsupervised clustering. Rows are centred; unit variance scaling is applied to rows. Rows are clustered using correlation distance and average linkage. There are 500 rows and 4 columns in total. **B** Heatmap of HIPO-POP019 dataset as shown in Figure 21 to scale.

5. Discussion

This study showed that a cohort of HNSCC patients that could not be stratified using single APOBEC3-related parameters was divided into subgroups with good and poor prognosis using a principal component analysis of a combination of APOBEC3-related genomic data and transcriptional data. The parameters in question reflected past and present APOBEC3-mediated mutagenesis, suggesting that APOBEC3-mediated mutagenesis has an impact on these cancers. However, APOBEC3A-mediated mutagenesis had no detectable phenotypic effect on the tumour growth of immortalised, transformed HEK293 cells either *in vitro* or *in vivo*. This implies that APOBEC3-mediated mutagenesis at a late stage during tumour development does not influence tumour growth. It was also observed that prolonged mutational pressure by APOBEC3A in HEK293 cells lead to a loss of APOBEC3 activity by various mechanisms. The premature stop codons found in some of the APOBEC3A transgenes match the APOBEC3 mutational signature, suggesting that APOBEC3A activity may be self-limiting. Finally, the HEK293 cells were genetically unstable even in the absence of cytidine deaminase activity.

5.1. Limitations of the HEK293 cell model

The cell culture model used in this study is based on HEK293 cells (ATCC accession number CRL-1573; www.atcc.org). HEK293 cells are primary human embryonic kidney cells transformed by sheared DNA of human adenovirus type 5 (Graham et al., 1977). The origin of HEK293 cells is still debated, with different studies suggesting kidney epithelial cells or fibroblasts, while others propose them to be derived from a neuronal lineage (reviewed by Stepanenko and Dmitrenko, 2015). Their expression profile suggests that they are most likely derived from an adrenal lineage of ectodermal origin (Lin et al., 2014b). They resemble dedifferentiated cells (Shaw et al., 2002), and could thus be considered a valid model for dedifferentiated tumour cells of ectodermal origin in a late stage of tumour development. However, as the HEK293 cells are already immortalised and transformed, they do not allow to study the effects of APOBEC3A-mediated mutagenesis during tumour promotion and/or progression.

HEK293 cells are known to be pseudotriploid (Bylund et al., 2004; Lin et al., 2014b), meaning that they are triploid but with chromosomal translocations (Battaglia, 1956). Repeated whole genome sequencing of the same HEK293 clones during this study revealed that the HEK293 cells are genetically unstable and acquire single nucleotide variants over time (Figure 13C). The majority of mutations acquired by the HEK293 cell clones were attributed to mutational signatures 3, 12, 16 and 25. Signature 3 has been

attributed to failure of DNA double-strand break repair by homologous recombination and BRCA1/2 mutations; the aetiology of signatures 12, 16 and 25 remains unknown (Alexandrov et al., 2013a). In addition, deletions and loss of heterozygosity were observed for large genomic regions. These results are consistent with prior studies that found a large number of mutations and a wide variety of karyotypes including loss of heterozygosity in different HEK293-derived cell lines (Lin et al., 2014b) and HEK293 cells from different sources (reviewed in Stepanenko and Dmitrenko, 2015). A large number of single nucleotide variants and copy number alterations were also found in the same Flp-In T-Rex™ HEK293 cell system used here (Akre et al., 2016). Akre et al. (2016) suggest that the mutator phenotype may be caused by faulty replicative DNA polymerase proofreading and/or defective mismatch repair. Stable transfection of an empty vector or replication stress is sufficient to increase chromosomal instability and genetic heterogeneity in HEK293 cells (Stepanenko et al., 2015, reviewed by Stepanenko and Dmitrenko, 2015). Both APOBEC3A and APOBEC3B can induce replication stress by deamination of cytidine at DNA replication forks (Kanu et al., 2016; Buisson et al., 2017), thus possibly increasing chromosomal instability beyond the normal effect of APOBEC3-mediated mutagenesis. In addition, the increase in chromosomal instability has even been shown to influence phenotype, including but not limited to proliferation, migration and apoptosis (reviewed by Stepanenko and Dmitrenko, 2015). Bylund et al. (2004), Lin et al. (2014b) and Stepanenko and Dmitrenko (2015) further discuss that it is impossible to discriminate whether the expression or activity of the introduced transgene or the stress-induced chromosomal changes are the cause of any observed phenotype. Thus, any phenotypes observed in the APOBEC3A-mutagenised clones cannot be unequivocally attributed to APOBEC3A.

Furthermore, as HEK293 cells are highly genetically unstable (Lin et al., 2014b; Stepanenko and Dmitrenko, 2015; Akre et al., 2016) and have been around since 1977 (Graham et al., 1977), they have very likely acquired a high load of pre-existing mutations since then. As no healthy tissue reference is available for HEK293 cells, it is impossible to determine how large the contribution of APOBEC3 signature mutations in the APOBEC3A-mutagenised cells actually is. Gene expression patterns of HEK293 cells suggest that they have already been selected for rapid growth by extensive *in vitro* cultivation (Lin et al., 2014b), so that an additional increase in proliferation by APOBEC3A-mediated mutagenesis may not be possible.

Finally, the genetic background can have an influence on the effect of certain mutations on the phenotype (discussed by Klonowska et al., 2017). The HEK293 cells used in this study went through several genetic bottlenecks during the introduction of transgenes as well as during the expansion of single cell clones following the APOBEC3A-mediated

mutagenesis. Due to the genetic instability of the HEK293 cells, it has to be assumed that the genetic background is different for each of the single cell clones. In order to find out whether and how the genetic background has an influence on the phenotypic effect of APOBEC3-mediated mutagenesis, it would be necessary to study the effects of APOBEC3-mediated mutagenesis in different cell types.

In summary, the HEK293 cells can serve as a valid model for dedifferentiated tumour cells of ectodermal origin in a late stage of tumour development. However, cells at such late stages already have many mutations that have been selected for increased proliferation and tumourigenicity. The genetic instability of HEK293 cells, which was also observed in the HEK293 cells used in this study, contributes to this by creating further mutations that selection can act on. Thus, it is likely that additional mutations introduced by APOBEC3A in HEK293 cells do not result in additional selection advantages that are large enough to produce a phenotypic difference. This implies that APOBEC3-mediated mutagenesis at a late stage of tumour development may not influence tumour growth due to previous extensive mutagenesis and selection processes.

5.2. The contribution of APOBEC3A and APOBEC3B to the APOBEC3 mutational signature

It is an ongoing debate whether APOBEC3A or APOBEC3B is the main source of the APOBEC3 mutational signatures. APOBEC3B was found to be overexpressed in cancer cell lines (Zhang et al., 2015b; Kanu et al., 2016; Buisson et al., 2017) breast cancers (Burns et al., 2013a), head and neck squamous cell carcinomas (Lin et al., 2014a; Fanourakis et al., 2016; Kosumi et al., 2016) and other tumour types (Burns et al., 2013b; Hedegaard et al., 2016). Its expression is upregulated by HPV oncoproteins E6 (Vieira et al., 2014) and E7 (Warren et al., 2015b). These findings suggest that APOBEC3B is involved in APOBEC3-mediated mutagenesis of cancer genomes.

So far, the overexpression of APOBEC3B in many different cancers has been assumed to be the cause of mutations, and in particular the cause of the APOBEC3 mutational signatures found in cancer (Burns et al., 2013b; Burns et al., 2013a). Recent studies found that there is a possibility that APOBEC3B overexpression may also be a result of mutation, inactivation or absence of p53 (Menendez et al., 2017). In breast cancer, APOBEC3B expression levels were also found to correlate with inactivation of p53 (Burns et al., 2013a). Menendez et al. (2017) showed that activation of p53 reduced expression of APOBEC3B, whereas tumour-associated mutants or absence of p53 were found to promote APOBEC3B upregulation. In contrast, activation of p53 by genotoxic stress caused an upregulation of APOBEC3 family members including APOBEC3A, and

functional p53 is required to maintain IFN-induced expression of APOBEC3A (Menendez et al., 2017). DNA replication stress has been found to increase APOBEC3-mediated mutagenesis (Green et al., 2016; Hoopes et al., 2016; Kanu et al., 2016), which potentially provides another link between p53 and APOBEC3 activity.

p53 is the most frequently mutated gene in human cancer (Zehir et al., 2017, reviewed by Kastenhuber and Lowe, 2017), which may explain why APOBEC3B overexpression is reliably found in many cancers and cancer cell lines, while the results are less clear for APOBEC3A (Burns et al., 2013b; Burns et al., 2013a; Roberts et al., 2013; Kosumi et al., 2016; Yang et al., 2016; Boichard et al., 2017; Buisson et al., 2017). APOBEC3B expression was found to be upregulated in breast cancer with p53 mutations (Cescon et al., 2015). In addition, p53 is bound and inhibited by the HPV oncoprotein E6 (Werness et al., 1990; reviewed by Freitas et al., 2014). Since p53 is already inactivated by HPV E6, an inactivating mutation of p53 would not provide a selection advantage to HPV-positive tumours. This is likely the reason why p53 is more frequently mutated in HPV-negative than HPV-positive tumours (Gaykalova et al., 2014; The Cancer Genome Atlas Network, 2015). The increased APOBEC3B expression levels in HPV-driven head and neck cancers observed in the HNSCC cohort (Figure 18B) could thus be explained by the inactivation of p53 in addition to the upregulation of APOBEC3B by HPV E6 and E7 (Vieira et al., 2014; Warren et al., 2015b). In summary, p53 loss of function provided by inactivating mutations in HPV-negative tumours and E6 in HPV-positive tumours may be one of the causes for both APOBEC3B overexpression and the low APOBEC3A expression in so many different tumours and cancer cell lines.

So far, the transcriptional regulation of APOBEC3 cytidine deaminases by p53 has only been shown in a panel of human cancer cell lines and primary human immune cells (Menendez et al., 2017). Prior studies made observations that support these findings as explained in detail above (Burns et al., 2013a; Green et al., 2016; Hoopes et al., 2016; Kanu et al., 2016). It may prove interesting to see whether a correlation between p53 loss of function and APOBEC3 expression can also be observed in the HIPO-POP019 and TCGA datasets, in particular considering HPV status.

The strongest argument brought forward in favour of APOBEC3B as the main contributor to the mutational signatures 2 and 13 is that the APOBEC3B mRNA expression levels can correlate with the presence of the APOBEC3 mutational signatures in breast cancer, bladder cancer, lung cancer, head and neck cancer, cervical cancer as well as a panel of 33 cancer types in TCGA (Burns et al., 2013b; Burns et al., 2013a; Roberts et al., 2013; Henderson et al., 2014; Seplyarskiy et al., 2016a; Chou et al., 2017). A similar weak but significant correlatin between APOBEC3B expression and the number of APOBEC3

signature mutations was found in HNSCC patients in this study (Figure 19H). C-to-T mutations in PIK3CA were found to be more frequent in oesophageal squamous cell carcinomas with higher APOBEC3B expression (Kosumi et al., 2016). Expression of transgenic APOBEC3B in HEK293 cells was also shown to cause a large number of C-to-T mutations, although neither mutational signature 2 nor signature 13 were observed (Akre et al., 2016). A recent study also suggested APOBEC3H haplotype I as an additional source of the APOBEC3 mutational signature in breast and lung cancer (Starrett et al., 2016). However, many of these arguments are based on a correlation of APOBEC3 expression with the presence of the APOBEC3 mutational signature in fully formed cancers at the time of resection. This does not necessarily reflect the situation in the cancers when the APOBEC3 signature was created (discussed by Caval et al., 2014a), especially as the exposure to APOBEC3 that leads to the mutational signature may well be transient in nature (Roberts and Gordenin, 2014a; Middlebrooks et al., 2016).

APOBEC3A activity in HEK293 cells is lost after inscribing the APOBEC3 mutational signatures. Similarly, it is possible that APOBEC3 expression is induced and afterwards lost during cancer development (discussed by Cescon et al., 2015). Such a scenario is consistent with the situation found in tumours both in this study and others, where tumours carrying the APOBEC3 mutational signature show no expression of either APOBEC3A or APOBEC3B at the time of resection (Roberts et al., 2013; Nordentoft et al., 2014; Chan et al., 2015; Lamy et al., 2016). The HIPO-POP019 dataset also contains several examples of tumours that carry the APOBEC3 mutational signature but do not express APOBEC3A and APOBEC3B. This idea is further supported by the fact that expression of both APOBEC3A and APOBEC3B is upregulated in precancerous cervical lesions, but the upregulation is no longer significant in fully formed cervical cancer (Warren et al., 2015b). Thus, the APOBEC3 transcript levels in the final tumour do not necessarily reflect the situation at the time the signatures were generated (discussed by Caval et al., 2014a; Lamy et al., 2016). It has been suggested that transient hypermutation may be more likely to cause cancer than sustained mutagenesis (Roberts and Gordenin, 2014a; Middlebrooks et al., 2016). This would be the case with a temporary induction of APOBEC3 expression during cancer development. The principal component analysis of both the HIPO-POP019 and the TCGA dataset identified groups of HNSCC patients that show signs of transient APOBEC3-mediated mutagenesis. Group 3 as well as potentially groups 1 and 2 can be explained by transient induction of APOBEC3 expression early during tumour development. After inscribing the APOBEC3 mutational signatures into the cancer genomes and so creating a large fraction of the mutations in these cancers, the APOBEC3 expression is eventually reduced or lost entirely.

5.3. Timeframe of APOBEC3-mediated mutagenesis during cancer development

It is still unclear when the APOBEC3 mutational signatures arise during tumour progression (Litwin et al., 2017). Some studies have found evidence that suggests APOBEC3-mediated mutagenesis may be a late event (Cui et al., 2009; Nik-Zainal et al., 2012b; McGranahan et al., 2015; Verlaat et al., 2015; Lamy et al., 2016), while others point towards APOBEC3-mediated mutagenesis occurring early during tumour progression (Nik-Zainal et al., 2012b; McGranahan et al., 2015). This suggests that APOBEC3-mediated mutagenesis may be active early in some cases and late in others (Yates et al., 2017, reviewed in Swanton et al., 2015), and may even be active over a longer period of time (Nordentoft et al., 2014; McGranahan et al., 2015).

The principal component analysis of the HIPO-POP019 and TCGA datasets suggests that some of the cancers were exposed to APOBEC3-mediated mutagenesis in the past, but show no signs of ongoing APOBEC3-driven mutagenesis. In contrast, APOBEC3-mediated mutagenesis is ongoing in others, but has not had a large impact on the cancer genome in the past. A strong contribution of APOBEC3-type mutations does not always co-occur with a high expression of APOBEC3A and/or APOBEC3B. This implies that all the APOBEC3 signature mutations in group 3 and the majority of the APOBEC3 signature mutations in group 1 and 2 were generated in the past at an early phase of cancer development, and APOBEC3 expression has been greatly reduced or entirely lost since then. In contrast, tumours with high expression of APOBEC3A and/or APOBEC3B in groups 4, 6, 7, 8 and 9 show a low contribution of APOBEC3 signature mutations to the total mutational load. This implies that the APOBEC3 expression observed in these groups only started a short while ago and has not yet had an impact on the genomic landscape. The ongoing APOBEC3-mediated mutagenesis at a late stage in the progression of these cancers may increase the genetic heterogeneity of the tumours.

The cancers that received the APOBEC3 mutational signatures early during tumour development can be found in group 3 as well as potentially in groups 1 and 2. In these cases, APOBEC3 was found to be the main mutagenic force shaping the cancer genome, but expression of both APOBEC3A and APOBEC3B is either absent or low. Similarly, the presence of the APOBEC3 mutational signature in the absence of APOBEC3A and/or APOBEC3B expression was also observed in some patients with various cancer types (Burns et al., 2013b; Burns et al., 2013a; Roberts et al., 2013; Henderson et al., 2014; Lamy et al., 2016; Seplyarskiy et al., 2016b; Chang et al., 2017; Chou et al., 2017). In contrast, there were also cases such as in groups 4, 7, and potentially 9, where no evidence of past APOBEC3 activity was found in the cancer genome, but either APOBEC3A, APOBEC3B or both were highly expressed. This suggests a recent

upregulation of APOBEC3 expression along with ongoing APOBEC3-mediated mutagenesis late in cancer progression. Groups 6 and 8 are examples of APOBEC3-mediated mutagenesis over a longer period of time. These tumours combine high expression of either APOBEC3A or APOBEC3B with a moderate contribution of APOBEC3 signature mutation to the total mutational load. This implies that APOBEC3-mediated mutagenesis has been active long enough to create an impact on the genomic landscape, but is also still ongoing.

In summary, the principal component analysis of patient samples found examples of early APOBEC3-mediated mutagenesis in groups 3 and potentially 1 and 2, and of late APOBEC3-mediated mutagenesis in groups 4, 7 and 9, as well as APOBEC3-mediated mutagenesis over a longer period of time in groups 6 and 8. Group 5 was found to be shaped by other mutagenic processes and thus entirely independent of APOBEC3-mediated mutagenesis.

5.3.1. APOBEC3-mediated mutagenesis as an early event

APOBEC3 activity at a very early time point during cancer development may be involved in creating the tumour. It can cause mutations in cancer drivers that are clonal in all cancer cells (McGranahan et al., 2015). Such a scenario may be particularly relevant in HPV-driven cancers. Whereas the HPV E6 and E7 oncoproteins can immortalise cells, they are not sufficient for transformation, which requires additional mutations (Münger et al., 1989, reviewed by McLaughlin-Drubin and Münger, 2009, discussed by Parfenov et al., 2014). Since viral infection (Middlebrooks et al., 2016) as well as HPV E6 and E7 can induce expression of APOBEC3A and B, and expression of both is upregulated in precancerous cervical lesions (Vieira et al., 2014; Warren et al., 2015b), it is possible to speculate that APOBEC3-mediated mutagenesis may be the factor that provides the additional mutations necessary for the transformation of HPV-immortalised cells.

As the acute effects of both APOBEC3A and APOBEC3B overexpression are toxic for the cells and cause cell death (Landry et al., 2011; Burns et al., 2013a; Lackey et al., 2013; Mussil et al., 2013; Akre et al., 2016; Kostrzak et al., 2016; Brachova et al., 2017), it is likely that only a few cells survive APOBEC3-mediated mutagenesis to become transformed cancer cells. Transient APOBEC3 expression early during cancer promotion is thus likely to result in a tumour that is homogeneous concerning the APOBEC3 mutational signatures. The genomic landscape of tumours in group 3 and potentially groups 1 and 2 suggests that APOBEC3-mediated mutagenesis was an early event in these patients. Thus, it can be speculated that these tumours are homogeneous for the APOBEC3 mutational signature, and that APOBEC3-mediated mutagenesis contributed to the transformation of these cancers.

5.3.2. APOBEC3-mediated mutagenesis as a late event

APOBEC3-mediated mutagenesis was also observed late during tumour progression of cells that were transformed by other mutagenic processes (Cui et al., 2009; Nik-Zainal et al., 2012b; de Bruin, Elza C. et al., 2014; Verlaat et al., 2015; Hao et al., 2016; Lefebvre et al., 2016). In this case, expression of APOBEC3A and/or APOBEC3B drives subclonal diversification and increases genetic heterogeneity (McGranahan et al., 2015; Lamy et al., 2016; Lefebvre et al., 2016).

APOBEC3-mediated mutagenesis late in tumour progression may be of advantage for the tumour. A high genetic diversity has been linked to higher tumorigenicity (Ye et al., 2009; Smid et al., 2016), and an increase in genetic heterogeneity has been linked to an increase in drug-resistant cells within a population (Brammeld et al., 2017). Drug resistance mutations in several different genes match the APOBEC3-type nucleotide exchange and target sequence (reviewed by Swanton et al., 2015). Expression of APOBEC3B was found to be higher in doxorubicin and etoposide-resistant cells of a breast cancer cell line (Onguru et al., 2016). Furthermore, urothelial carcinomas that had undergone chemotherapy showed an enrichment of APOBEC3 signature mutations (Faltas et al., 2016). This suggests that APOBEC3-mediated mutagenesis may be linked to the development of therapy resistance. Several resistant clones may even already be present in a lesion before start of the treatment (Bozic and Nowak, 2014). A high number of mutations also offers a larger sequence diversity on which selection can act (Stratton et al., 2009). It has furthermore been suggested that populations with greater heterogeneity may be more robust, meaning they can better maintain their function and/or performance when faced with perturbations (Kitano, 2007). Although cancer driver mutations are generally early events and thus clonal within a tumour (McGranahan et al., 2015; Yates et al., 2017), some mutations in both cancer driver genes and tumour suppressor genes were also found to be subclonal and thus happened later during cancer progression (de Bruin, Elza C. et al., 2014; McGranahan et al., 2015; Verlaat et al., 2015). In summary, APOBEC3-mediated mutagenesis late during tumour development may provide the tumour with advantages such as increased heterogeneity, subclonal diversification or resistance to therapy.

APOBEC3-mediated mutagenesis late in tumour progression may also have a deleterious effect for the tumour. More mutations increase the chance of developing neo-antigens. Neo-antigens are epitopes of mutated protein versions expressed in cancer cells that can be recognised by autologous T-cells, causing anti-tumour immunity (reviewed by Heemskerk et al., 2013; Schumacher and Schreiber, 2015; Berger and Pu, 2017). Cancer cells that express immune checkpoint molecules such as programmed death ligand-1 (PD-L1) incapacitate effector T-cells, thus evading the immune response against neo-

antigens. PD-L1 achieves this by binding to the programmed death-1 (PD-1) receptor mainly expressed on the surface of T-cells, which inhibits the ability of effective T-cells to kill the cancer cells (reviewed by Berger and Pu, 2017). This is countered in immunotherapy with checkpoint inhibitors, which block the interaction and thus activate anti-tumour immunity (reviewed by Sukari et al., 2016). Between 45-80% of head and neck cancers express PD-L1 on their surface, and chemo- and/or radiotherapy may further upregulate its expression (reviewed by Fuereder, 2016). Trials of treating HNSCC with checkpoint inhibitors have shown promising results, but predictive biomarkers to select patients who benefit from it are still missing (summarised by Fuereder, 2016). There is evidence that a high mutational load may lead to more neo-antigens and so improve the chances that a neo-antigen which can be recognised by autologous T-cells arises (Heemskerk et al., 2013; Connor et al., 2016; Madore et al., 2016, 2016; Secrier et al., 2016; Smid et al., 2016), but the total mutational load is only an imperfect marker for success of immunotherapy (Schumacher and Schreiber, 2015; Boichard et al., 2017). Interestingly, an increased expression of APOBEC3A has been found to correlate with a stronger expression of PD-L1 and an increased presence of tumour-infiltrating mononuclear cells in urothelial carcinoma (Mullane et al., 2016). In breast cancer, the presence of the APOBEC3 mutational signatures was found to be associated with tumour-infiltrating lymphocytes (Smid et al., 2016). A study on the TCGA pan-cancer dataset also found a correlation of PD-L1 expression levels with overexpression of all APOBEC3 family members, as well as the presence of APOBEC3-mediated mutations and *kataegis* (Boichard et al., 2017). Boichard et al. (2017) suggest that this correlation between APOBEC3 expression and PD-L1 expression may be caused by common underlying mechanisms. One such mechanism may be the upregulation of both by a viral infection, or a common signalling cascade such as protein kinase C. Furthermore, tumour cells with a high load of neo-antigens potentially created by APOBEC3-mediated mutagenesis derive a selection advantage from PD-L1 expression: without a mechanism to evade immunity, the cancer would be cleared by the immune response against the neo-antigens. Overall, these are hints that patients whose tumours show a high expression of APOBEC3A and/or APOBEC3B may benefit from immune checkpoint inhibitor therapy, but further studies are needed in this regard.

DNA damage and replication stress induce the DNA damage response via the ATR (ataxia telangiectasia and Rad3-related protein) and ATM (ataxia telangiectasia mutated) kinases (reviewed by Maréchal and Zou, 2013). Both APOBEC3A and APOBEC3B activity have a genotoxic effect and cause replication stress and DNA damage (Landry et al., 2011; Burns et al., 2013a; Lackey et al., 2013; Land et al., 2013; Mussil et al., 2013; Taylor et al., 2013; Akre et al., 2016; Kanu et al., 2016; Kostrzak et al., 2016; Brachova et al.,

2017; Nikkilä et al., 2017). Thus, APOBEC3A- and APOBEC3B-mediated cytidine deamination can result in activation of DNA replication checkpoints (Landry et al., 2011; Green et al., 2016), which is essentially a mechanism to protect the genome from unchecked damage and mutagenesis and to maintain genome integrity (reviewed by Maréchal and Zou, 2013, discussed by Green et al., 2017). It has been suggested that cancer cells require the DNA damage response pathways in order to be able to tolerate genomic instability, and that its inhibition in cancers with specific kinds of genomic instability may result in synthetic lethality (Maréchal and Zou, 2013). Indeed, it has been shown that cancer cells with high APOBEC3A or APOBEC3B expression are dependent on the ATR checkpoint for survival, which makes them susceptible to ATR inhibition (Buisson et al., 2017; Green et al., 2017; Nikkilä et al., 2017). Interestingly, although overexpression of APOBEC3A activates both the ATR and the ATM signalling pathways (Green et al., 2016), inhibition of ATM signalling does not result in synthetic lethality in leukaemia cells (Green et al., 2017). Activation of ATM is primarily mediated by double-stranded DNA breaks, whereas ATR is also involved in other forms of DNA repair including re-starting of stalled replication forks as well as mismatch and nucleotide excision repair (reviewed by Maréchal and Zou, 2013). One possible reason why inhibition of ATM does not lead to synthetic lethality is that what is considered high APOBEC3A expression levels in leukaemia cells may not be high enough to cause DNA double-strand breaks, while still activating the ATR DNA damage response (Buisson et al., 2017). This would result in an increased sensitivity towards inhibition of the ATR response, while sensitivity towards ATM inhibition would remain unchanged. In summary, synthetic lethality induced by ATR inhibition may also benefit tumour patients with high APOBEC3A and/or APOBEC3B expression levels.

These contrasts between the potentially negative effects of ongoing APOBEC3-mediated mutagenesis and its potentially positive effects further demonstrate its double-edged nature. As expression of APOBEC3A and/or APOBEC3B causes genetic instability and increases genetic heterogeneity by inducing mutations and can thus create neo-antigens, the patients in groups 6 to 9 with high APOBEC3A and/or APOBEC3B expression and a poor prognosis may benefit from immune checkpoint inhibition and ATR inhibition. However, these patients may also have a higher risk of developing therapy-resistant tumours, and might thus potentially benefit from a combination therapy to prevent the survival of resistant subclones.

5.4. Patient stratification and outlook

Generally, stratification attempts focus on one characteristic such as mutations in certain cancer- or therapy-relevant genes (Sawada et al., 2016; Zehir et al., 2017), expression of a single gene (Mullane et al., 2016; Chen et al., 2017) or an entire transcription profile (Keck et al., 2015; Hedegaard et al., 2016; Smid et al., 2016; Svoboda et al., 2016; Zhang et al., 2016), or the presence of mutational signatures (Connor et al., 2016; Secrier et al., 2016; Wang et al., 2017). While the subgroups that were generated using one of these characteristics were often checked for an association with other parameters (e.g. subgroups created by translational profile were checked for enrichment of mutational signatures), none of the approaches integrated both genomic data and gene expression data in creating the subgroups. In this study, stratification of head and neck squamous cell carcinoma patients in the HIPO-POP019 dataset based on a single APOBEC3-related parameter showed no difference in progression-free survival (Figure 20). In contrast, the principal component analysis performed here (Figure 21A and Figure 22) considered both the genomic data, which essentially reflects the mutagenic processes that shaped the cancer genome in the past, and the gene expression data of APOBEC3A and APOBEC3B, which indicates ongoing mutagenesis in the tumour at the time of resection. This also allowed a distinction between APOBEC3-mediated mutagenesis as early and late events during tumour development. The resulting patterns identified in the heatmap of the principal component analysis allowed to distinguish groups of patients with good and poor prognosis concerning progression-free survival (Figure 21C). The defining characteristic of a large fraction of HNSCC patients with a poor prognosis was the high APOBEC3A and/or APOBEC3B expression. As discussed above, APOBEC3-mediated mutagenesis late during tumour development may be detrimental for the patient by increasing tumour heterogeneity, or advantageous by creating neo-antigens. Thus, the patients with the high APOBEC3 expression levels may benefit from immune checkpoint inhibitor therapy. Furthermore, APOBEC3A specifically as well as APOBEC3B themselves may be potential therapeutic targets, in particular in the patients with poor prognosis.

While either high past or ongoing APOBEC3-mediated mutagenesis is the characterising feature of many of the patterns defined by the principal component analysis, group 3 and particularly group 5 show a low overall involvement of APOBEC3. This suggests that different mutagenic process shape the genomes of the cancers in these. Aside from APOBEC3-driven mutagenesis, signature 16 is prominent in HNSCC, whereas the signatures 4 and 24 attributed to tobacco only contribute a small fraction of mutations in most cases. This is consistent with observations made by Zhang et al. (2015a), Alexandrov et al. (2016) and Chang et al. (2017). An association between smoking and

drinking and mutational signature 16 has been suggested (Chang et al., 2017), possibly as an indirect effect of tobacco and alcohol.

The patterns derived from the principal component analysis were based on the z-scores of an entire dataset consisting of either 79 patients in case of the HIPO-POP019 study or 500 patients in the TCGA dataset. The z-scores are all relative and depend on the mean and standard deviation of the entire population. Thus, no cut-offs were defined for absolute values, making it impossible to assign single patients to one of the subgroups. In order to use the analysis performed in this thesis to make a prognostic statement about single patients, absolute cut-off values need to be defined, tested and verified.

Aside from prognostic value concerning progression-free survival, the patterns defined in the heatmap of the principal component analysis might also provide valuable information concerning treatment choices. Upregulation of APOBEC3A expression has been linked to higher PD-L1 expression (Mullane et al., 2016; Smid et al., 2016; Boichard et al., 2017), possibly because APOBEC3-mediated mutagenesis causes the creation of neo-antigens. This suggests that patients with ongoing APOBEC3-mediated mutagenesis in end-stage tumours may benefit from immune checkpoint inhibitor therapy. Furthermore, ongoing APOBEC3A and APOBEC3B expression activates DNA replication checkpoints (Landry et al., 2011; Green et al., 2016), thus causing synthetic lethality with ATR inhibition (Buisson et al., 2017; Green et al., 2017; Nikkilä et al., 2017). Ongoing APOBEC3-mediated mutagenesis in late stages of tumour development is also involved in subclonal diversification of tumours (Cui et al., 2009; Nik-Zainal et al., 2012b; de Bruin, Elza C. et al., 2014; Verlaat et al., 2015; Hao et al., 2016; Lefebvre et al., 2016), thus increasing their genetic heterogeneity (McGranahan et al., 2015; Lamy et al., 2016; Lefebvre et al., 2016) and potentially contributing to therapy resistance (Swanton et al., 2015; Faltas et al., 2016; Onguru et al., 2016; Brammeld et al., 2017). Thus, APOBEC3A and APOBEC3B may themselves be therapeutic targets, especially in combination with other therapeutic approaches, to reduce the chance of resistant clones arising. While certainly of interest for the groups 2 and 4 in the principal component analysis, patients in these subgroups already have a good prognosis. Thus, these treatment options would be particularly relevant for the patients in the groups 6 to 9, as they have a poor prognosis for progression-free survival and are characterised by high APOBEC3A and/or APOBEC3B expression levels. Therefore, it would be interesting to see whether the patients in these subgroups derive a benefit from any of these therapy options.

As HPV-positive head and neck cancers are associated with improved survival and better prognosis (Schwartz et al., 2001; Fakhry et al., 2008; Hay and Ganly, 2015), a de-intensification of therapy in HPV-driven tumours has been suggested (reviewed by

Benson et al., 2014). However, the principal component analysis found that approximately half the patients with HPV-driven tumours were in the subgroups with poor prognosis (Figure 21A), where a de-intensified therapy may be ill-advised. The subgroups defined by the principal component analysis may thus be helpful in further stratifying patients with HPV-driven cancers to identify the cases where therapy can be safely de-escalated.

In conclusion, the generation of the patterns based on a principal component analysis of both genomic and expression data may be a valuable addition not just for prognosis, but also an additional stratification method concerning treatment choices and personalised medicine.

6. References

Aguiar, R.S., and Peterlin, B.M. (2008). **APOBEC3 proteins and reverse transcription.** *Virus Res 134*, 74-85.

Akre, M.K., Starrett, G.J., Quist, J.S., Temiz, N.A., Carpenter, M.A., Tutt, A.N.J., Grigoriadis, A., and Harris, R.S. (2016). **Mutation Processes in 293-Based Clones Overexpressing the DNA Cytosine Deaminase APOBEC3B.** *PloS one 11*, e0155391.

Alexandrov, L.B. (2015). **Understanding the origins of human cancer. Analyzing the DNA sequences of more than 12,000 cancer patients revealed signatures of mutational processes.** *Science 350*.

Alexandrov, L.B., Ju, Y.S., Haase, K., van Loo, P., Martincorena, I., Nik-Zainal, S., Totoki, Y., Fujimoto, A., Nakagawa, H., and Shibata, T., et al. (2016). **Mutational signatures associated with tobacco smoking in human cancer.** *Science 354*, 618-622.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, Samuel A J R, Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., and Borresen-Dale, A.-L., et al. (2013a). **Signatures of mutational processes in human cancer.** *Nature 500*, 415-421.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., and Stratton, M.R. (2013b). **Deciphering signatures of mutational processes operative in human cancer.** *Cell Reports 3*, 246-259.

Alexandrov, L.B., and Stratton, M.R. (2014). **Mutational signatures: the patterns of somatic mutations hidden in cancer genomes.** *Current Opinion in Genetics & Development 24*, 52-60.

Aynaud, M.-M., Suspène, R., Vidalain, P.-O., Mussil, B., Guétard, D., Tangy, F., Wain-Hobson, S., and Vartanian, J.-P. (2012). **Human Tribbles 3 protects nuclear DNA from cytidine deamination by APOBEC3A.** *J Biol Chem 287*, 39182-39192.

Bader, A.G., Kang, S., and Vogt, P.K. (2006). **Cancer-specific mutations in PIK3CA are oncogenic in vivo.** *Proc. Natl. Acad. Sci. U.S.A. 103*, 1475-1479.

Battaglia, E. (1956). **The Concept of Pseudopolyploidy.** *Caryologia 8*, 214-220.

Beale, R.C.L., Petersen-Mahrt, S.K., Watt, I.N., Harris, R.S., Rada, C., and Neuberger, M.S. (2004). **Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes. Correlation with mutation spectra in vivo.** *Journal of molecular biology 337*, 585-596.

Benson, E., Li, R., Eisele, D., and Fakhry, C. (2014). **The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas.** *Oral oncology 50*, 565-574.

Berger, K.N., and Pu, J.J. (2017). **PD-1 pathway and its clinical application. A 20year journey after discovery of the complete human PD-1 gene.** *Gene*.

Bishop, K.N., Holmes, R.K., Sheehy, A.M., Davidson, N.O., Cho, S.-J., and Malim, M.H. (2004). **Cytidine deamination of retroviral DNA by diverse APOBEC proteins.** *Curr. Biol. 14*, 1392-1396.

Bogerd, H.P., Wiegand, H.L., Doehle, B.P., Lueders, K.K., and Cullen, B.R. (2006). **APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells.** *Nucleic acids research 34*, 89-95.

Bohn, M.-F., Shandilya, S.M.D., Silvas, T.V., Nalivaika, E.A., Kouno, T., Kelch, B.A., Ryder, S.P., Kurt-Yilmaz, N., Somasundaran, M., and Schiffer, C.A. (2015). **The ssDNA Mutator APOBEC3A Is Regulated by Cooperative Dimerization.** *Structure (London, England : 1993) 23*, 903-911.

Boichard, A., Tsigelny, I.F., and Kurzrock, R. (2017). **High expression of PD-1 ligands is associated with kataegis mutational signature and APOBEC3 alterations.** *Oncoimmunology 6*, e1284719.

Bonvin, M., Achermann, F., Greeve, I., Stroka, D., Keogh, A., Inderbitzin, D., Candinas, D., Sommer, P., Wain-Hobson, S., and Vartanian, J.-P., et al. (2006). **Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication.** *Hepatology 43*, 1364-1374.

Bozic, I., and Nowak, M.A. (2014). **Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers.** *Proc. Natl. Acad. Sci. U.S.A. 111*, 15964-15968.

Brachova, P., Alvarez, N.S., van Voorhis, B.J., and Christenson, L.K. (2017). **Cytidine deaminase Apobec3a induction in fallopian epithelium after exposure to follicular fluid.** *Gynecologic oncology*.

Brammeld, J.S., Petljak, M., Martincorena, I., Williams, S.P., Alonso, L.G., Dalmases, A., Bellosillo, B., Robles-Espinoza, C.D., Price, S., and Barthorpe, S., et al. (2017). **Genome-wide chemical mutagenesis screens allow unbiased saturation of the cancer genome and identification of drug resistance mutations.** *Genome research*.

Buisson, R., Lawrence, M.S., Benes, C.H., and Zou, L. (2017). **APOBEC3A and 3B activities render cancer cells susceptible to ATR inhibition.** *Cancer Research*.

Bulliard, Y., Narvaiza, I., Bertero, A., Peddi, S., Röhrig, U.F., Ortiz, M., Zoete, V., Castro-Díaz, N., Turelli, P., and Telenti, A., et al. (2011). **Structure-function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities.** *J Virol 85*, 1765-1776.

Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., and Nikas, J.B., et al. (2013a). **APOBEC3B is an enzymatic source of mutation in breast cancer.** *Nature 494*, 366-370.

Burns, M.B., Temiz, N.A., and Harris, R.S. (2013b). **Evidence for APOBEC3B mutagenesis in multiple human cancers.** *Nat Genet 45*, 977-983.

Byeon, I.-J.L., Ahn, J., Mitra, M., Byeon, C.-H., Hercik, K., Hritz, J., Charlton, L.M., Levin, J.G., and Gronenborn, A.M. (2013). **NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity.** *Nature communications 4*, 1890.

Bylund, L., Kytölä, S., Lui, W.-O., Larsson, C., and Weber, G. (2004). **Analysis of the cytogenetic stability of the human embryonal kidney cell line 293 by cytogenetic and STR profiling approaches.** *Cytogenetic and genome research 106*, 28-32.

Carpenter, M.A., Li, M., Rathore, A., Lackey, L., Law, E.K., Land, A.M., Leonard, B., Shandilya, S.M.D., Bohn, M.-F., and Schiffer, C.A., et al. (2012). **Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A.** *J Biol Chem 287*, 34801-34808.

Caval, V., Suspène, R., Shapira, M., Vartanian, J.-P., and Wain-Hobson, S. (2014a). **A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage.** *Nature communications 5*, 5129-5136.

Caval, V., Suspène, R., Vartanian, J.-P., and Wain-Hobson, S. (2014b). **Orthologous mammalian APOBEC3A cytidine deaminases hypermutate nuclear DNA.** *Molecular biology and evolution 31*, 330-340.

Cescon, D.W., Haibe-Kains, B., and Mak, T.W. (2015). **APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation.** *Proc. Natl. Acad. Sci. U.S.A. 112*, 2841-2846.

Chan, K., Resnick, M.A., and Gordenin, D.A. (2013). **The choice of nucleotide inserted opposite abasic sites formed within chromosomal DNA reveals the polymerase activities participating in translesion DNA synthesis.** *DNA repair 12*, 878-889.

Chan, K., Roberts, S.A., Klimczak, L.J., Sterling, J.F., Saini, N., Malc, E.P., Kim, J., Kwiatkowski, D.J., Fargo, D.C., and Mieczkowski, P.A., et al. (2015). **An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers.** *Nature genetics 47*, 1067-1072.

Chan, K., Sterling, J.F., Roberts, S.A., Bhagwat, A.S., Resnick, M.A., and Gordenin, D.A. (2012). **Base damage within single-strand DNA underlies in vivo hypermutability induced by a ubiquitous environmental agent.** *PLoS genetics 8*, e1003149.

Chang, J., Tan, W., Ling, Z., Xi, R., Shao, M., Chen, M., Luo, Y., Zhao, Y., Liu, Y., and Huang, X., et al. (2017). **Genomic analysis of oesophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and genomic alterations.** *Nature communications 8*.

Chelico, L., Pham, P., Calabrese, P., and Goodman, M.F. (2006). **APOBEC3G DNA deaminase acts processively 3' -- 5' on single-stranded DNA.** *Nature structural & molecular biology 13*, 392-399.

Chen, H., Lilley, C.E., Yu, Q., Lee, D.V., Chou, J., Narvaiza, I., Landau, N.R., and Weitzman, M.D. (2006). **APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons.** *Curr. Biol. 16*, 480-485.

Chen, T.-W., Lee, C.-C., Liu, H., Wu, C.-S., Pickering, C.R., Huang, P.-J., Wang, J., Chang, I.Y.-F., Yeh, Y.-M., and Chen, C.-D., et al. (2017). **APOBEC3A is an oral cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion polymorphism.** *Nature communications 8*, 465.

Chiu, Y.-L., and Greene, W.C. (2008). **The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements.** *Annu. Rev. Immunol. (Annual review of immunology) 26*, 317-353.

Chou, W.-C., Chen, W.-T., Hsiung, C.-N., Hu, L.-Y., Yu, J.-C., Hsu, H.-M., and Shen, C.-Y. (2017). **B-Myb Induces APOBEC3B Expression Leading to Somatic Mutation in Multiple Cancers.** *Scientific Reports 7*, 44089.

Colotta, F., Allavena, P., Sica, A., Garlanda, C., and Mantovani, A. (2009). **Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability.** *Carcinogenesis 30*, 1073-1081.

Connor, A.A., Denroche, R.E., Jang, G.H., Timms, L., Kalimuthu, S.N., Selander, I., McPherson, T., Wilson, G.W., Chan-Seng-Yue, M.A., and Borozan, I., et al. (2016). **Association of Distinct Mutational Signatures With Correlates of Increased Immune Activity in Pancreatic Ductal Adenocarcinoma.** *JAMA oncology*.

Conticello, S.G. (2008). **The AID/APOBEC family of nucleic acid mutators.** *Genome Biol 9*.

Conticello, S.G., Langlois, M.-A., Yang, Z., and Neuberger, M.S. (2007). **DNA deamination in immunity. AID in the context of its APOBEC relatives.** *Advances in immunology 94*, 37-73.

Conticello, S.G., Thomas, C.J.F., Petersen-Mahrt, S.K., and Neuberger, M.S. (2005). **Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases.** *Molecular biology and evolution 22*, 367-377.

Cui, B., Zheng, B., Zhang, X., Stendahl, U., Andersson, S., and Wallin, K.-L. (2009). **Mutation of PIK3CA. Possible risk factor for cervical carcinogenesis in older women.** *International Journal of Oncology 34*, 409-416.

Cullen, B.R. (2006). **Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors.** *J Virol 80*, 1067-1076.

de Bruin, Elza C., McGranahan, N., Mitter, R., Salm, M., Wedge, D.C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., and Rowan, A.J., et al. (2014). **Spatial and temporal diversity in genomic instability processes defines lung cancer evolution.** *Science 346*, 251-256.

Dempsey, A., and Bowie, A.G. (2015). **Innate immune recognition of DNA. A recent history.** *Virology 479-480*, 146-152.

Dilruba, S., and Kalayda, G.V. (2016). **Platinum-based drugs. Past, present and future.** *Cancer Chemotherapy and Pharmacology 77*, 1103-1124.

Fakhry, C., Westra, W.H., Li, S., Cmelak, A., Ridge, J.A., Pinto, H., Forastiere, A., and Gillison, M.L. (2008). **Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial.** *Journal of the National Cancer Institute 100*, 261-269.

Faltas, B.M., Prandi, D., Tagawa, S.T., Molina, A.M., Nanus, D.M., Sternberg, C., Rosenberg, J., Mosquera, J.M., Robinson, B., and Elemento, O., et al. (2016). **Clonal evolution of chemotherapy-resistant urothelial carcinoma.** *Nature genetics 48*, 1490-1499.

Fanourakis, G., Tosios, K., Papanikolaou, N., Chatzistamou, I., Xydous, M., Tseleni-Balafouta, S., Sklavounou, A., Voutsinas, G.E., and Vastardis, H. (2016). **Evidence for APOBEC3B mRNA and protein expression in oral squamous cell carcinomas.** *Experimental and molecular pathology 101*, 314-319.

Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., and Ponting, L., et al. (2017). **COSMIC. Somatic cancer genetics at high-resolution.** *Nucleic acids research 45*, D777-D783.

Freitas, A.C. de, Coimbra, E.C., and Leitão, M.d.C.G. (2014). **Molecular targets of HPV oncoproteins. Potential biomarkers for cervical carcinogenesis.** *Biochimica et biophysica acta 1845*, 91-103.

Fuereder, T. (2016). **Immunotherapy for head and neck squamous cell carcinoma.** *Memo 9*, 66-69.

Gaykalova, D.A., Mambo, E., Choudhary, A., Houghton, J., Buddavarapu, K., Sanford, T., Darden, W., Adai, A., Hadd, A., and Latham, G., et al. (2014). **Novel insight into mutational landscape of head and neck squamous cell carcinoma.** *PloS one 9*, e93102.

Göhler, S., Da Silva Filho, Miguel Inacio, Johansson, R., Enquist-Olsson, K., Henriksson, R., Hemminki, K., Lenner, P., and Försti, A. (2015). **Impact of functional germline variants and a deletion polymorphism in APOBEC3A and APOBEC3B on breast cancer risk and survival in a Swedish study population.** *Journal of cancer research and clinical oncology*.

Graham, F.L., Smiley, J., Russell, W.C., and Nairn, R. (1977). **Characteristics of a Human Cell Line Transformed by DNA from Human Adenoviruy Type 5.** *Journal of General Virology 36*, 59-72.

Green, A.M., Budagyan, K., Hayer, K.E., Reed, M.A., Savani, M.R., Wertheim, G.B., and Weitzman, M.D. (2017). **Cytosine deaminase APOBEC3A sensitizes leukemia cells to inhibition of the DNA replication checkpoint.** *Cancer Research*.

Green, A.M., Landry, S., Budagyan, K., Avgousti, D., Shalhout, S., Bhagwat, A.S., and Weitzman, M.D. (2016). **APOBEC3A damages the cellular genome during DNA replication.** *Cell cycle*, 0.

Han, Y., Qi, Q., He, Q., Sun, M., Wang, S., Zhou, G., and Sun, Y. (2016). **APOBEC3 deletion increases the risk of breast cancer: a meta-analysis.** *Oncotarget*.

Hanahan, D., and Weinberg, R.A. (2000). **The hallmarks of cancer.** *Cell 100*, 57-70.

Hanahan, D., and Weinberg, R.A. (2011). **Hallmarks of cancer: the next generation.** *Cell 144*, 646-674.

Hao, J.-J., Lin, D.-C., Dinh, H.Q., Mayakonda, A., Jiang, Y.-Y., Chang, C., Jiang, Y., Lu, C.-C., Shi, Z.-Z., and Xu, X., et al. (2016). **Spatial intratumoral heterogeneity and temporal clonal evolution in esophageal squamous cell carcinoma.** *Nature genetics 48*, 1500-1507.

Haradhvala, N.J., Polak, P., Stojanov, P., Covington, K.R., Shinbrot, E., Hess, J.M., Rheinbay, E., Kim, J., Maruvka, Y.E., and Braunstein, L.Z., et al. (2016). **Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair.** *Cell 164*, 538-549.

Harjes, S., Jameson, G.B., Filichev, V.V., Edwards, P.J.B., and Harjes, E. (2017). **NMR-based method of small changes reveals how DNA mutator APOBEC3A interacts with its single-stranded DNA substrate.** *Nucleic acids research*.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). **DNA deamination mediates innate immunity to retroviral infection.** *Cell 113*, 803-809.

Harris, R.S., and Liddament, M.T. (2004). **Retroviral restriction by APOBEC proteins.** *Nat. Rev. Immunol. 4*, 868-877.

Harris, R.S., Petersen-Mahrt, S.K., and Neuberger, M.S. (2002a). **RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators.** *Mol Cell 10*, 1247-1253.

Harris, R.S., Sale, J.E., Petersen-Mahrt, S.K., and Neuberger, M.S. (2002b). **AID is essential for immunoglobulin V gene conversion in a cultured B cell line.** *Curr. Biol. 12*, 435-438.

Hay, A., and Ganly, I. (2015). **Targeted Therapy in Oropharyngeal Squamous Cell Carcinoma. The Implications of HPV for Therapy.** *Rare cancers and therapy 3*, 89-117.

Hedegaard, J., Lamy, P., Nordentoft, I., Algaba, F., Høyer, S., Ulhøi, B.P., Vang, S., Reinert, T., Hermann, G.G., and Mogensen, K., et al. (2016). **Comprehensive Transcriptional Analysis of Early-Stage Urothelial Carcinoma.** *Cancer cell 30*, 27-42.

Heemskerk, B., Kvistborg, P., and Schumacher, T.N.M. (2013). **The cancer antigenome.** *The EMBO journal 32*, 194-203.

Helleday, T., Eshtad, S., and Nik-Zainal, S. (2014). **Mechanisms underlying mutational signatures in human cancers.** *Nat Rev Genet 15*, 585-598.

Henderson, S., Chakravarthy, A., Su, X., Boshoff, C., and Fenton, T.R. (2014). **APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development.** *Cell Reports 7*, 1833-1841.

Herdman, M.T., Pett, M.R., Roberts, I., Alazawi, W.O.F., Teschendorff, A.E., Zhang, X.-Y., Stanley, M.A., and Coleman, N. (2006). **Interferon-beta treatment of cervical keratinocytes naturally infected with human papillomavirus 16 episomes promotes rapid reduction in episome numbers and emergence of latent integrants.** *Carcinogenesis 27*, 2341-2353.

Hollstein, M., Alexandrov, L.B., Wild, C.P., Ardin, M., and Zavadil, J. (2016). **Base changes in tumour DNA have the power to reveal the causes and evolution of cancer.** *Oncogene*.

Holmes, R.K., Malim, M.H., and Bishop, K.N. (2007). **APOBEC-mediated viral restriction: not simply editing?** *Trends Biochem Sci 32*, 118-128.

Hoopes, J., Cortez, L., Mertz, T., Malc, E.P., Mieczkowski, P.A., and Roberts, S.A. (2016). **APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication.** *Cell Reports 14*, 1273-1282.

Hoste, E., Arwert, E.N., Lal, R., South, A.P., Salas-Alanis, J.C., Murrell, D.F., Donati, G., and Watt, F.M. (2015). **Innate sensing of microbial products promotes wound-induced skin cancer.** *Nature communications 6*, 5932.

Ito, F., Fu, Y., Kao, S.-C.A., Yang, H., and Chen, X.S. (2017). **Family-Wide Comparative Analysis of Cytidine and Methylcytidine Deamination by Eleven Human APOBEC Proteins.** *Journal of molecular biology*.

Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., and Navaratnam, N. (2002). **An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22.** *Genomics 79*, 285-296.

Jia, P., Pao, W., and Zhao, Z. (2014). **Patterns and processes of somatic mutations in nine major cancers.** *BMC Med. Genomics (BMC medical genomics) 7*, 11.

Kang, S., Bader, A.G., and Vogt, P.K. (2005). **Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic.** *Proc. Natl. Acad. Sci. U.S.A. 102*, 802-807.

Kanu, N., Cerone, M.A., Goh, G., Zalmas, L.-P., Bartkova, J., Dietzen, M., McGranahan, N., Rogers, R., Law, E.K., and Gromova, I., et al. (2016). **DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer.** *Genome biology 17*, 185.

Kastenhuber, E.R., and Lowe, S.W. (2017). **Putting p53 in Context.** *Cell 170*, 1062-1078.

Kazanov, M.D., Roberts, S.A., Polak, P., Stamatoyannopoulos, J., Klimczak, L.J., Gordenin, D.A., and Sunyaev, S.R. (2015). **APOBEC-Induced Cancer Mutations Are Uniquely Enriched in Early-Replicating, Gene-Dense, and Active Chromatin Regions.** *Cell Reports 13*, 1103-1109.

Kazazian, H.H. (2004). **Mobile elements. Drivers of genome evolution.** *Science 303*, 1626-1632.

Keck, M.K., Zuo, Z., Khattri, A., Stricker, T.P., Brown, C.D., Imanguli, M., Rieke, D., Endhardt, K., Fang, P., and Brägelmann, J., et al. (2015). **Integrative analysis of head and neck cancer identifies two biologically distinct HPV and three non-HPV subtypes.** *Clinical cancer research : an official journal of the American Association for Cancer Research 21*, 870-881.

Kidd, J.M., Newman, T.L., Tuzun, E., Kaul, R., and Eichler, E.E. (2007). **Population stratification of a common APOBEC gene deletion polymorphism.** *PLoS genetics 3*.

King, J.J., and Larijani, M. (2017). **A Novel Regulator of Activation-Induced Cytidine Deaminase/APOBECs in Immunity and Cancer. Schrodinger's CATalytic Pocket.** *Frontiers in immunology 8*, 351.

Kinomoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., Sata, T., and Tokunaga, K. (2007). **All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition.** *Nucleic acids research 35*, 2955-2964.

Kitano, H. (2007). **Towards a theory of biological robustness.** *Molecular systems biology 3*, 137.

Klonowska, K., Kluzniak, W., Rusak, B., Jakubowska, A., Ratajska, M., Krawczynska, N., Vasilevska, D., Czubak, K., Wojciechowska, M., and Cybulski, C., et al. (2017). **The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population.** *Oncotarget*.

Knisbacher, B.A., Gerber, D., and Levanon, E.Y. (2015). **DNA Editing by APOBECs: A Genomic Preserver and Transformer.** *Trends in genetics : TIG*.

Knisbacher, B.A., and Levanon, E.Y. (2015). **DNA editing of LTR retrotransposons reveals the impact of APOBECs on vertebrate genomes.** *Molecular biology and evolution*.

Komatsu, A., Nagasaki, K., Fujimori, M., Amano, J., and Miki, Y. (2008). **Identification of novel deletion polymorphisms in breast cancer.** *International Journal of Oncology 33*, 261-270.

Kondo, S., Wakae, K., Wakisaka, N., Nakanishi, Y., Ishikawa, K., Komori, T., Moriyama-Kita, M., Endo, K., Murono, S., and Wang, Z., et al. (2017). **APOBEC3A associates with human papillomavirus genome integration in oropharyngeal cancers.** *Oncogene 36*, 1687-1697.

Koning, F.A., Newman, E.N.C., Kim, E.-Y., Kunstman, K.J., Wolinsky, S.M., and Malim, M.H. (2009). **Defining APOBEC3 Expression Patterns in Human Tissues and Hematopoietic Cell Subsets.** *J Virol 83*, 9474-9485.

Kostrzak, A., Caval, V., Escande, M., Pliquet, E., Thalmensi, J., Bestetti, T., Julithe, M., Fiette, L., Huet, T., and Wain-Hobson, S., et al. (2016). **APOBEC3A intratumoral DNA electroporation in mice.** *Gene therapy*.

Kostrzak, A., Henry, M., Demoyen, P.L., Wain-Hobson, S., and Vartanian, J.-P. (2015). **APOBEC3A catabolism of electroporated plasmid DNA in mouse muscle.** *Gene therapy 22*, 96-103.

Kosumi, K., Baba, Y., Ishimoto, T., Harada, K., Nakamura, K., Ohuchi, M., Kiyozumi, Y., Izumi, D., Tokunaga, R., and Taki, K., et al. (2016). **APOBEC3B is an enzymatic source of molecular alterations in esophageal squamous cell carcinoma.** *Medical oncology (Northwood, London, England) 33*, 26.

Kouno, T., Silvas, T.V., Hilbert, B.J., Shandilya, S.M.D., Bohn, M.F., Kelch, B.A., Royer, W.E., Somasundaran, M., Kurt Yilmaz, N., and Matsuo, H., et al. (2017). **Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity.** *Nature communications 8*, 15024.

Lace, M.J., Anson, J.R., Haugen, T.H., Dierdorff, J.M., and Turek, L.P. (2015). **Interferon treatment of human keratinocytes harboring extrachromosomal, persistent HPV-16 plasmid genomes induces de novo viral integration.** *Carcinogenesis 36*, 151-159.

Lackey, L., Demorest, Z.L., Land, A.M., Hultquist, J.F., Brown, W.L., and Harris, R.S. (2012). **APOBEC3B and AID have similar nuclear import mechanisms.** *Journal of molecular biology 419*, 301-314.

Lackey, L., Law, E.K., Brown, W.L., and Harris, R.S. (2013). **Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination.** *Cell cycle 12*, 762-772.

Lada, A.G., Dhar, A., Boissy, R.J., Hirano, M., Rubel, A.A., Rogozin, I.B., and Pavlov, Y.I. (2012). **AID/APOBEC cytosine deaminase induces genome-wide kataegis.** *Biology direct 7*, 47.

Lagarde, S.M., Kate, F.J.W. ten, Boer, D.J. de, Busch, O.R.C., Obertop, H., and van Lanschot, J.J.B. (2006). **Extracapsular lymph node involvement in node-positive patients with adenocarcinoma of the distal esophagus or gastroesophageal junction.** *The American journal of surgical pathology 30*, 171-176.

Lamy, P., Nordentoft, I., Birkenkamp-Demtröder, K., Thomsen, M.B.H., Villesen, P., Vang, S., Hedegaard, J., Borre, M., Jensen, J.B., and Høyer, S., et al. (2016). **Paired Exome Analysis Reveals Clonal Evolution and Potential Therapeutic Targets in Urothelial Carcinoma.** *Cancer Research 76*, 5894-5906.

Land, A.M., Law, E.K., Carpenter, M.A., Lackey, L., Brown, W.L., and Harris, R.S. (2013). **Endogenous APOBEC3A DNA cytosine deaminase is cytoplasmic and nongenotoxic.** *J Biol Chem 288*, 17253-17260.

Landry, S., Narvaiza, I., Linfesty, D.C., and Weitzman, M.D. (2011). **APOBEC3A can activate the DNA damage response and cause cell-cycle arrest.** *EMBO Rep. 12*, 444-450.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., and Roberts, S.A., et al. (2013). **Mutational heterogeneity in cancer and the search for new cancer-associated genes.** *Nature 499*, 214-218.

Lefebvre, C., Bachelot, T., Filleron, T., Pedrero, M., Campone, M., Soria, J.-C., Massard, C., Levy, C., Arnedos, M., and Lacroix-Triki, M., et al. (2016). **Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis.** *PLoS medicine 13*, e1002201.

Leonard, B., McCann, J.L., Starrett, G.J., Kosyakovsky, L., Luengas, E.M., Molan, A.M., Burns, M.B., McDougle, R.M., Parker, P.J., and Brown, W.L., et al. (2015). **The PKC/NF-κB Signaling Pathway Induces APOBEC3B Expression in Multiple Human Cancers.** *Cancer Research 75*, 4538-4547.

Li, Z., Abraham, B.J., Berezovskaya, A., Farah, N., Liu, Y., Leon, T., Fielding, A., Tan, S.H., Sanda, T., and Weintraub, A.S., et al. (2017). **APOBEC signature mutation generates an oncogenic enhancer that drives LMO1 expression in T-ALL.** *Leukemia*.

Liao, W., Hong, S.H., Chan, B.H.J., Rudolph, F.B., Clark, S.C., and Chan, L. (1999). **APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family (vol 260, pg 398, 1999).** *Biochemical and biophysical research communications 263*, 264.

Lin, D.-C., Hao, J.-J., Nagata, Y., Xu, L., Shang, L., Meng, X., Sato, Y., Okuno, Y., Varela, A.M., and Ding, L.-W., et al. (2014a). **Genomic and molecular characterization of esophageal squamous cell carcinoma.** *Nature genetics 46*, 467-473.

Lin, Y.-C., Boone, M., Meuris, L., Lemmens, I., van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., and Drmanac, R., et al. (2014b). **Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations.** *Nature communications 5*, 4767.

Lindahl, T. (1974). **An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues.** *Proc. Natl. Acad. Sci. U.S.A. 71*, 3649-3653.

Litwin, T.R., Clarke, M.A., Dean, M., and Wentzensen, N. (2017). **Somatic Host Cell Alterations in HPV Carcinogenesis.** *Viruses 9*.

Long, J., Delahanty, R.J., Li, G., Gao, Y.-T., Lu, W., Cai, Q., Xiang, Y.-B., Li, C., Ji, B.-T., and Zheng, Y., et al. (2013). **A common deletion in the APOBEC3 genes and breast cancer risk.** *Journal of the National Cancer Institute 105*, 573-579.

Love, R.P., Xu, H., and Chelico, L. (2012). **Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A.** *J Biol Chem 287*, 30812-30822.

Lowry, O.H., Rosebrough, N.J., Farr, A.L., and Randall, R.J. (1951). **Protein Measurement with the Folin Phenol Reagent.** *J Biol Chem 193*, 265-275.

Lucifora, J., Xia, Y., Reisinger, F., Zhang, K., Stadler, D., Cheng, X., Sprinzl, M.F., Koppensteiner, H., Makowska, Z., and Volz, T., et al. (2014). **Specific and nonhepatotoxic degradation of nuclear hepatitis B virus cccDNA.** *Science 343*, 1221-1228.

Madore, J., Strbenac, D., Vilain, R., Menzies, A.M., Yang, J.Y.H., Thompson, J.F., Long, G.V., Mann, G.J., Scolyer, R.A., and Wilmott, J.S. (2016). **PD-L1 Negative Status is Associated with Lower Mutation Burden, Differential Expression of Immune-Related Genes, and Worse Survival in Stage III Melanoma.** *Clinical cancer research : an official journal of the American Association for Cancer Research 22*, 3915-3923.

Madsen, P., Anant, S., Rasmussen, H.H., Gromov, P., Vorum, H., Dumanski, J.P., Tommerup, N., Collins, J.E., Wright, C.L., and Dunham, I., et al. (1999). **Psoriasis upregulated phorbolin-1 shares structural but not functional similarity to the mRNA-editing protein apobec-1.** *J. Invest. Dermatol. (The Journal of investigative dermatology) 113*, 162-169.

Maréchal, A., and Zou, L. (2013). **DNA damage sensing by the ATM and ATR kinases.** *Cold Spring Harbor perspectives in biology 5*.

Maruyama, W., Shirakawa, K., Matsui, H., Matsumoto, T., Yamazaki, H., Sarca, A.D., Kazuma, Y., Kobayashi, M., Shindo, K., and Takaori-Kondo, A. (2016). **Classical NF-κB pathway is responsible for APOBEC3B expression in cancer cells.** *Biochemical and biophysical research communications 478*, 1466-1471.

Matoulkova, E., Michalova, E., Vojtesek, B., and Hrstka, R. (2012). **The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells.** *RNA biology 9*, 563-576.

McDougle, R.M., Hultquist, J.F., Stabell, A.C., Sawyer, S.L., and Harris, R.S. (2013). **D316 is critical for the enzymatic activity and HIV-1 restriction potential of human and rhesus APOBEC3B.** *Virology 441*, 31-39.

McGranahan, N., Favero, F., Bruin, E.C. de, Birkbak, N.J., Szallasi, Z., and Swanton, C. (2015). **Clonal status of actionable driver events and the timing of mutational processes in cancer evolution.** *Science translational medicine 7*, 283ra54.

McGranahan, N., and Swanton, C. (2015). **Biological and therapeutic impact of intratumor heterogeneity in cancer evolution.** *Cancer cell 27*, 15-26.

McLaughlin-Drubin, M.E., and Münger, K. (2009). **Oncogenic activities of human papillomaviruses.** *Virus Res 143*, 195-208.

Mehta, H.V., Jones, P.H., Weiss, J.P., and Okeoma, C.M. (2012). **IFN-α and lipopolysaccharide upregulate APOBEC3 mRNA through different signaling pathways.** *J. Immunol. 189*, 4088-4103.

Menendez, D., Nguyen, T.-A., Snipe, J., and Resnick, M.A. (2017). **The Cytidine Deaminase APOBEC3 Family is Subject to Transcriptional Regulation by p53.** *Molecular cancer research : MCR*.

Metsalu, T., and Vilo, J. (2015). **ClustVis. A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap.** *Nucleic acids research 43*, W566-70.

Metzger, R., Drebber, U., Baldus, S.E., Mönig, S.P., Hölscher, A.H., and Bollschweiler, E. (2009). **Extracapsular lymph node involvement differs between squamous cell and adenocarcinoma of the esophagus.** *Annals of surgical oncology 16*, 447-453.

Middlebrooks, C.D., Banday, A.R., Matsuda, K., Udquim, K.-I., Onabajo, O.O., Paquin, A., Figueroa, J.D., Zhu, B., Koutros, S., and Kubo, M., et al. (2016). **Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors.** *Nature genetics 48*, 1330-1338.

Mitra, M., Hercík, K., Byeon, I.-J.L., Ahn, J., Hill, S., Hinchee-Rodriguez, K., Singer, D., Byeon, C.-H., Charlton, L.M., and Nam, G., et al. (2014). **Structural determinants of human APOBEC3A enzymatic and nucleic acid binding properties.** *Nucleic acids research 42*, 1095-1110.

Mohanram, V., Sköld, A.E., Bächle, S.M., Pathak, S.K., and Spetz, A.-L. (2013). **IFN-α induces APOBEC3G, F, and A in immature dendritic cells and limits HIV-1 spread to CD4+ T cells.** *J. Immunol. 190*, 3346-3353.

Morganella, S., Alexandrov, L.B., Glodzik, D., Zou, X., Davies, H., Staaf, J., Sieuwerts, A.M., Brinkman, A.B., Martin, S., and Ramakrishna, M., et al. (2016). **The topography of mutational processes in breast cancer genomes.** *Nature communications 7*.

Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Löwer, J., Cichutek, K., Flory, E., Schumann, G.G., and Münk, C. (2006). **APOBEC3 proteins inhibit human LINE-1 retrotransposition.** *J Biol Chem 281*, 22161-22172.

Mullane, S.A., Werner, L., Rosenberg, J., Signoretti, S., Callea, M., Choueiri, T.K., Freeman, G.J., and Bellmunt, J. (2016). **Correlation of Apobec Mrna Expression with overall Survival and pd-l1 Expression in Urothelial Carcinoma.** *Scientific Reports 6*, 27702.

Münger, K., Phelps, W.C., Bubb, V., Howley, P.M., and Schlegel, R. (1989). **The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes.** *J Virol 63*, 4417-4421.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). **Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme.** *Cell 102*, 553-563.

Mussil, B., Suspène, R., Aynaud, M.-M., Gauvrit, A., Vartanian, J.-P., and Wain-Hobson, S. (2013). **Human APOBEC3A isoforms translocate to the nucleus and induce DNA double strand breaks leading to cell stress and death.** *PloS one 8*, e73641.

Ndiaye, C., Mena, M., Alemany, L., Arbyn, M., Castellsagué, X., Laporte, L., Bosch, F.X., Sanjosé, S. de, and Trottier, H. (2014). **HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers. A systematic review and meta-analysis.** *The Lancet. Oncology 15*, 1319-1331.

Niavarani, A., Currie, E., Reyal, Y., Anjos-Afonso, F., Horswell, S., Griessinger, E., Luis Sardina, J., and Bonnet, D. (2015). **APOBEC3A Is Implicated in alpha Novel Class of G-to-A mRNA Editing in WT1 Transcripts.** *PloS one 10*.

Nichols, A.C., Palma, D.A., Chow, W., Tan, S., Rajakumar, C., Rizzo, G., Fung, K., Kwan, K., Wehrli, B., and Winquist, E., et al. (2013). **High frequency of activating PIK3CA mutations in human papillomavirus-positive oropharyngeal cancer.** *JAMA otolaryngology-- head & neck surgery 139*, 617-622.

Nikkilä, J., Kumar, R., Campbell, J., Brandsma, I., Pemberton, H.N., Wallberg, F., Nagy, K., Scheer, I., Vertessy, B.G., and Serebrenik, A.A., et al. (2017). **Elevated APOBEC3B expression drives a kataegic-like**

mutation signature and replication stress-related therapeutic vulnerabilities in p53-defective cells. *British Journal of Cancer 117*, 113-123.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., and Stebbings, L.A., et al. (2012a). **Mutational processes molding the genomes of 21 breast cancers.** *Cell 149*, 979-993.

Nik-Zainal, S., van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Raine, K., Jones, D., Marshall, J., and Ramakrishna, M., et al. (2012b). **The life history of 21 breast cancers.** *Cell 149*, 994-1007.

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

Nordentoft, I., Lamy, P., Birkenkamp-Demtröder, K., Shumansky, K., Vang, S., Hornshøj, H., Juul, M., Villesen, P., Hedegaard, J., and Roth, A., et al. (2014). **Mutational context and diverse clonal development in early and late bladder cancer.** *Cell Rep 7*, 1649-1663.

Ojesina, A.I., Lichtenstein, L., Freeman, S.S., Pedamallu, C.S., Imaz-Rosshandler, I., Pugh, T.J., Cherniack, A.D., Ambrogio, L., Cibulskis, K., and Bertelsen, B., et al. (2014). **Landscape of genomic alterations in cervical carcinomas.** *Nature 506*, 371-375.

Okazaki, I.-m., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). **Constitutive expression of AID leads to tumorigenesis.** *The Journal of experimental medicine 197*, 1173-1181.

Onguru, O., Yalcin, S., Rosemblit, C., Zhang, P.J., Kilic, S., and Gunduz, U. (2016). **APOBEC3B expression in drug resistant MCF-7 breast cancer cell lines.** *Biomedicine & pharmacotherapy = Biomédecine & pharmacothérapie 79*, 87-92.

Parfenov, M., Pedamallu, C.S., Gehlenborg, N., Freeman, S.S., Danilova, L., Bristow, C.A., Lee, S., Hadjipanayis, A.G., Ivanova, E.V., and Wilkerson, M.D., et al. (2014). **Characterization of HPV and host genome interactions in primary head and neck cancers.** *Proc. Natl. Acad. Sci. U.S.A. 111*, 15544-15549.

Peng, G., Lei, K.J., Jin, W., Greenwell-Wild, T., and Wahl, S.M. (2006). **Induction of APOBEC3 family proteins, a defensive maneuver underlying interferon-induced anti-HIV-1 activity.** *The Journal of experimental medicine 203*, 41-46.

Peterson, G.L. (1979). **Review of the Folin phenol protein quantitation method of Lowry, Rosebrough, Farr and Randall.** *Analytical biochemistry 100*, 201-220.

Petljak, M., and Alexandrov, L.B. (2016). **Understanding mutagenesis through delineation of mutational signatures in human cancer.** *Carcinogenesis 37*, 531-540.

Radmanesh, H., Spethmann, T., Enssen, J., Schuermann, P., Bhuju, S., Geffers, R., Antonenkova, N., Khusnutdinova, E., Sadr-Nabavi, A., and Shandiz, F.H., et al. (2017). **Assessment of an APOBEC3B truncating mutation, c.783delG, in patients with breast cancer.** *Breast Cancer Research and Treatment 162*, 31-37.

Raftery, N., and Stevenson, N.J. (2017). **Advances in anti-viral immune defence. Revealing the importance of the IFN JAK/STAT pathway.** *Cellular and molecular life sciences : CMLS 74*, 2525-2535.

Refsland, E.W., and Harris, R.S. (2013). **The APOBEC3 family of retroelement restriction factors.** *Curr. Top. Microbiol. Immunol. (Current topics in microbiology and immunology) 371*, 1-27.

Refsland, E.W., Stenglein, M.D., Shindo, K., Albin, J.S., Brown, W.L., and Harris, R.S. (2010). **Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction.** *Nucleic acids research 38*, 4274-4284.

Revathidevi, S., Manikandan, M., Rao, Arunagiri Kuha Deva Magendhra, Vinothkumar, V., Arunkumar, G., Rajkumar, K.S., Ramani, R., Rajaraman, R., Ajay, C., and Munirajan, A.K. (2016). **Analysis of APOBEC3A/3B germline deletion polymorphism in breast, cervical and oral cancers from South India and its impact on miRNA regulation.** *Tumor Biology 37*, 11983-11990.

Roberts, S.A., and Gordenin, D.A. (2014a). **Clustered and genome-wide transient mutagenesis in human cancers. Hypermutation without permanent mutators or loss of fitness.** *BioEssays : news and reviews in molecular, cellular and developmental biology*.

Roberts, S.A., and Gordenin, D.A. (2014b). **Hypermutation in human cancer genomes: footprints and mechanisms.** *Nat Rev Cancer 14*, 786-800.

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

Roberts, S.A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L.J., Kryukov, G.V., Malc, E., and Mieczkowski, P.A., et al. (2012). **Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions.** *Molecular cell 46*, 424-435.

Rogozin, I.B., Basu, M.K., Jordan, I.K., Pavlov, Y.I., and Koonin, E.V. (2005). **APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis.** *Cell cycle 4*, 1281-1285.

Rogozin, I.B., Pavlov, Y.I., Goncearenco, A., De, S., Lada, A.G., Poliakov, E., Panchenko, A.R., and Cooper, D.N. (2017). **Mutational signatures and mutable motifs in cancer genomes.** *Briefings in bioinformatics*.

Sakofsky, C.J., Roberts, S.A., Malc, E., Mieczkowski, P.A., Resnick, M.A., Gordenin, D.A., and Malkova, A. (2014). **Break-induced replication is a source of mutation clusters underlying kataegis.** *Cell Reports 7*, 1640-1648.

Sawada, G., Niida, A., Uchi, R., Hirata, H., Shimamura, T., Suzuki, Y., Shiraishi, Y., Chiba, K., Imoto, S., and Takahashi, Y., et al. (2016). **Genomic Landscape of Esophageal Squamous Cell Carcinoma in a Japanese Population.** *Gastroenterology 150*, 1171-1182.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). **NIH Image to ImageJ. 25 years of image analysis.** *Nat Meth 9*, 671-675.

Schoggins, J.W., Wilson, S.J., Panis, M., Murphy, M.Y., Jones, C.T., Bieniasz, P., and Rice, C.M. (2011). **A diverse range of gene products are effectors of the type I interferon antiviral response.** *Nature 472*, 481-485.

Schumacher, A.J., Haché, G., Macduff, D.A., Brown, W.L., and Harris, R.S. (2008). **The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction.** *J Virol 82*, 2652-2660.

Schumacher, T.N., and Schreiber, R.D. (2015). **Neoantigens in cancer immunotherapy.** *Science (New York, N.Y.) 348*, 69-74.

Schwartz, S.R., Yueh, B., McDougall, J.K., Daling, J.R., and Schwartz, S.M. (2001). **Human papillomavirus infection and survival in oral squamous cell cancer. A population-based study.** *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery 125*, 1-9.

Secrier, M., Li, X., Silva, N. de, Eldridge, M.D., Contino, G., Bornschein, J., MacRae, S., Grehan, N., O'Donovan, M., and Miremadi, A., et al. (2016). **Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance.** *Nature genetics*.

Seplyarskiy, V.B., Andrianova, M.A., and Bazykin, G.A. (2016a). **APOBEC3A/B-induced mutagenesis is responsible for 20% of heritable mutations in the TpCpW context.** *Genome research*.

Seplyarskiy, V.B., Soldatov, R.A., Popadin, K.Y., Antonarakis, S.E., Bazykin, G.A., and Nikolaev, S.I. (2016b). **APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication.** *Genome research*.

Sharma, S., Patnaik, S.K., Kemer, Z., and Baysal, B.E. (2017). **Transient overexpression of exogenous APOBEC3A causes C-to-U RNA editing of thousands of genes.** *RNA biology 14*, 603-610.

Sharma, S., Patnaik, S.K., Taggart, R.T., Kannisto, E.D., Enriquez, S.M., Gollnick, P., and Baysal, B.E. (2015). **APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages.** *Nature communications 6*, 6881.

Shaw, G., Morse, S., Ararat, M., and Graham, F.L. (2002). **Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells.** *FASEB journal : official publication of the Federation of American Societies for Experimental Biology 16*, 869-871.

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

Shi, K., Carpenter, M.A., Banerjee, S., Shaban, N.M., Kurahashi, K., Salamango, D.J., McCann, J.L., Starrett, G.J., Duffy, J.V., and Demir, O., et al. (2016). **Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B.** *Nature structural & molecular biology*.

Shinohara, M., Io, K., Shindo, K., Matsui, M., Sakamoto, T., Tada, K., Kobayashi, M., Kadowaki, N., and Takaori-Kondo, A. (2012). **APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells.** *Scientific Reports 2*.

Shlyakhtenko, L.S., Dutta, S., Li, M., Harris, R.S., and Lyubchenko, Y.L. (2016). **Single-Molecule Force Spectroscopy Studies of APOBEC3A-Single-Stranded DNA Complexes.** *Biochemistry 55*, 3102-3106.

Smid, M., Rodriguez-Gonzalez, F.G., Sieuwerts, A.M., Salgado, R., Prager-Van der Smissen, Wendy J C, van der Vlugt-Daane, M., van Galen, A., Nik-Zainal, S., Staaf, J., and Brinkman, A.B., et al. (2016). **Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration.** *Nature communications 7*, 12910.

Smith, H.C., Bennett, R.P., Kizilyer, A., McDougall, W.M., and Prohaska, K.M. (2012). **Functions and regulation of the APOBEC family of proteins.** *Seminars in cell & developmental biology 23*, 258-268.

Starrett, G.J., Luengas, E.M., McCann, J.L., Ebrahimi, D., Temiz, N.A., Love, R.P., Feng, Y., Adolph, M.B., Chelico, L., and Law, E.K., et al. (2016). **The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis.** *Nature communications 7*, 12918.

Stenglein, M.D., Burns, M.B., Li, M., Lengyel, J., and Harris, R.S. (2010). **APOBEC3 proteins mediate the clearance of foreign DNA from human cells.** *Nat. Struct. Mol. Biol. 17*, 222-229.

Stepanenko, A., Andreieva, S., Korets, K., Mykytenko, D., Huleyuk, N., Vassetzky, Y., and Kavsan, V. (2015). **Step-wise and punctuated genome evolution drive phenotype changes of tumor cells.** *Mutation research 771*, 56-69.

Stepanenko, A.A., and Dmitrenko, V.V. (2015). **HEK293 in cell biology and cancer research. Phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution.** *Gene 569*, 182-190.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., and Stebbings, L.A., et al. (2011). **Massive genomic rearrangement acquired in a single catastrophic event during cancer development.** *Cell 144*, 27-40.

Stransky, N., Egloff, A.M., Tward, A.D., Kostic, A.D., Cibulskis, K., Sivachenko, A., Kryukov, G.V., Lawrence, M.S., Sougnez, C., and McKenna, A., et al. (2011). **The mutational landscape of head and neck squamous cell carcinoma.** *Science 333*, 1157-1160.

Stratton, M.R. (2011). **Exploring the genomes of cancer cells: progress and promise.** *Science 331*, 1553-1558.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). **The cancer genome.** *Nature 458*, 719-724.

Sukari, A., Nagasaka, M., Al-Hadidi, A., and Lum, L.G. (2016). **Cancer Immunology and Immunotherapy.** *Anticancer research 36*, 5593-5606.

Suspene, R., Mussil, B., Laude, H., Caval, V., Berry, N., Bouzidi, M.S., Thiers, V., Wain-Hobson, S., and Vartanian, J.-P. (2017). **Self-cytoplasmic DNA upregulates the mutator enzyme APOBEC3A leading to chromosomal DNA damage.** *Nucleic acids research*.

Suspène, R., Aynaud, M.-M., Guétard, D., Henry, M., Eckhoff, G., Marchio, A., Pineau, P., Dejean, A., Vartanian, J.-P., and Wain-Hobson, S. (2011). **Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism.** *Proc. Natl. Acad. Sci. U.S.A. 108*, 4858-4863.

Suspène, R., Sommer, P., Henry, M., Ferris, S., Guétard, D., Pochet, S., Chester, A., Navaratnam, N., Wain-Hobson, S., and Vartanian, J.-P. (2004). **APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase.** *Nucleic acids research 32*, 2421-2429.

Svoboda, M., Meshcheryakova, A., Heinze, G., Jaritz, M., Pils, D., Castillo-Tong, D.C., Hager, G., Thalhammer, T., Jensen-Jarolim, E., and Birner, P., et al. (2016). **AID/APOBEC-network reconstruction identifies pathways associated with survival in ovarian cancer.** *BMC genomics 17*, 643.

Swanton, C., McGranahan, N., Starrett, G.J., and Harris, R.S. (2015). **APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity.** *Cancer discovery 5*, 704-712.

Taylor, B.J., Nik-Zainal, S., Wu, Y.L., Stebbings, L.A., Raine, K., Campbell, P.J., Rada, C., Stratton, M.R., and Neuberger, M.S. (2013). **DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis.** *Elife 2*, e00534.

Teng, B., Burant, C.F., and Davidson, N.O. (1993). **Molecular cloning of an apolipoprotein B messenger RNA editing protein.** *Science 260*, 1816-1819.

The Cancer Genome Atlas Network (2015). **Comprehensive genomic characterization of head and neck squamous cell carcinomas.** *Nature 517*, 576-582.

The Cancer Genome Atlas Research Network (2017). **Integrated genomic and molecular characterization of cervical cancer.** *Nature 543*, 378-384.

Thielen, B.K., McNevin, J.P., McElrath, M.J., Hunt, B.V.S., Klein, K.C., and Lingappa, J.R. (2010). **Innate Immune Signaling Induces High Levels of TC-specific Deaminase Activity in Primary Monocyte-derived Cells through Expression of APOBEC3A Isoforms.** *J Biol Chem 285*, 27753-27766.

Tornesello, M.L., Annunziata, C., Buonaguro, L., Losito, S., Greggi, S., and Buonaguro, F.M. (2014). **TP53 and PIK3CA gene mutations in adenocarcinoma, squamous cell carcinoma and high-grade intraepithelial neoplasia of the cervix.** *Journal of translational medicine 12*, 255.

Vartanian, J.-P., Guétard, D., Henry, M., and Wain-Hobson, S. (2008). **Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions.** *Science 320*, 230-233.

Verlaat, W., Snijders, P.J., van Moorsel, M.I., Bleeker, M., Rozendaal, L., Sie, D., Ylstra, B., Meijer, C.J., Steenbergen, R.D., and Am Heideman, D. (2015). **Somatic mutation in PIK3CA is a late event in cervical carcinogenesis.** *The journal of pathology. Clinical research 1*, 207-211.

Vieira, V.C., Leonard, B., White, E.A., Starrett, G.J., Temiz, N.A., Lorenz, L.D., Lee, D., Soares, M.A., Lambert, P.F., and Howley, P.M., et al. (2014). **Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B.** *mBio 5*.

Wakae, K., Aoyama, S., Wang, Z., Kitamura, K., Liu, G., Monjurul, A.M., Koura, M., Imayasu, M., Sakamoto, N., and Nakamura, M., et al. (2015). **Detection of hypermutated human papillomavirus type 16 genome by Next-Generation Sequencing.** *Virology 485*, 460-466.

Walboomers, J.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J., Peto, J., Meijer, C.J., and Muñoz, N. (1999). **Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.** *The Journal of pathology 189*, 12-19.

Wang, Y., Schmitt, K., Guo, K., Santiago, M.L., and Stephens, E.B. (2015). **The Role of the Single Deaminase Domain APOBEC3A in Virus Restriction, Retrotransposition, DNA Damage and Cancer.** *The Journal of general virology*.

Wang, Y.K., Bashashati, A., Anglesio, M.S., Cochrane, D.R., Grewal, D.S., Ha, G., McPherson, A., Horlings, H.M., Senz, J., and Prentice, L.M., et al. (2017). **Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes.** *Nat Genet*.

Wang, Z., Wakae, K., Kitamura, K., Aoyama, S., Liu, G., Koura, M., Monjurul, A.M., Kukimoto, I., and Muramatsu, M. (2014). **APOBEC3 deaminases induce hypermutation in human papillomavirus 16 DNA upon beta interferon stimulation.** *Journal of virology 88*, 1308-1317.

Warren, C.J., van Doorslaer, K., Pandey, A., Espinosa, J.M., and Pyeon, D. (2015a). **Role of the host restriction factor APOBEC3 on papillomavirus evolution.** *Virus evolution 1*.

Warren, C.J., Westrich, J.A., van Doorslaer, K., and Pyeon, D. (2017). **Roles of APOBEC3A and APOBEC3B in Human Papillomavirus Infection and Disease Progression.** *Viruses 9*.

Warren, C.J., Xu, T., Guo, K., Griffin, L.M., Westrich, J.A., Lee, D., Lambert, P.F., Santiago, M.L., and Pyeon, D. (2015b). **APOBEC3A functions as a restriction factor of human papillomavirus.** *Journal of virology 89*, 688-702.

Wedekind, J.E., Dance, G.S.C., Sowden, M.P., and Smith, H.C. (2003). **Messenger RNA editing in mammals. New members of the APOBEC family seeking roles in the family business.** *Trends in genetics : TIG 19*, 207-216.

Werness, B.A., Levine, A.J., and Howley, P.M. (1990). **Association of human papillomavirus types 16 and 18 E6 proteins with p53.** *Science 248*, 76-79.

Xuan, D., Li, G., Cai, Q., Deming-Halverson, S., Shrubsole, M.J., Shu, X.-O., Kelley, M.C., Zheng, W., and Long, J. (2013). **APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry.** *Carcinogenesis 34*, 2240-2243.

Yamanaka, S., Balestra, M.E., Ferrell, L.D., Fan, J., Arnold, K.S., Taylor, S., Taylor, J.M., and Innerarity, T.L. (1995). **Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals.** *Proc. Natl. Acad. Sci. U.S.A. 92*, 8483-8487.

Yang, Y., Wang, H., Zhang, X., Huo, W., Qi, R., Gao, Y., Zhang, G., Song, B., Chen, H., and Gao, X. (2016). **Heat increases the editing efficiency of human papillomavirus E2 gene by inducing upregulation of APOBEC3A and 3G.** *The Journal of investigative dermatology*.

Yates, L.R., Knappskog, S., Wedge, D., Farmery, J.H.R., Gonzalez, S., Martincorena, I., Alexandrov, L.B., van Loo, P., Haugland, H.K., and Lilleng, P.K., et al. (2017). **Genomic Evolution of Breast Cancer Metastasis and Relapse.** *Cancer cell 32*, 169-184.e7.

Ye, C.J., Stevens, J.B., Liu, G., Bremer, S.W., Jaiswal, A.S., Ye, K.J., Lin, M.-F., Lawrenson, L., Lancaster, W.D., and Kurkinen, M., et al. (2009). **Genome based cell population heterogeneity promotes tumorigenicity. The evolutionary mechanism of cancer.** *Journal of cellular physiology 219*, 288-300.

Zehir, A., Benayed, R., Shah, R.H., Syed, A., Middha, S., Kim, H.R., Srinivasan, P., Gao, J., Chakravarty, D., and Devlin, S.M., et al. (2017). **Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients.** *Nature medicine*.

Zhang, L., Zhou, Y., Cheng, C., Cui, H., Le Cheng, Kong, P., Wang, J., Li, Y., Chen, W., and Song, B., et al. (2015a). **Genomic Analyses Reveal Mutational Signatures and Frequently Altered Genes in Esophageal Squamous Cell Carcinoma.** *American Journal of Human Genetics 96*, 597-611.

Zhang, T., Cai, J., Chang, J., Yu, D., Wu, C., Yan, T., Zhai, K., Bi, X., Zhao, H., and Xu, J., et al. (2013). **Evidence of associations of APOBEC3B gene deletion with susceptibility to persistent HBV infection and hepatocellular carcinoma.** *Human molecular genetics 22*, 1262-1269.

Zhang, Y., Delahanty, R., Guo, X., Zheng, W., and Long, J. (2015b). **Integrative genomic analysis reveals functional diversification of APOBEC gene family in breast cancer.** *Human genomics 9*, 34.

Zhang, Y., Koneva, L.A., Virani, S., Arthur, A.E., Virani, A., Hall, P.B., Warden, C.D., Carey, T.E., Chepeha, D.B., and Prince, M.E., et al. (2016). **Subtypes of HPV-Positive Head and Neck Cancers Are Associated with HPV Characteristics, Copy Number Alterations, PIK3CA Mutation, and Pathway Signatures.** *Clinical cancer research : an official journal of the American Association for Cancer Research 22*, 4735-4745.

Zhou, N., Yuan, Y., Long, X., Wu, C., and Bao, J. (2017). **Mutational signatures efficiently identify different mutational processes underlying cancers with similar somatic mutation spectra.** *Mutation research 806*, 27-30.

7. Sworn affidavit

The thesis I have submitted is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given. I have not presented this thesis or parts thereof to a university as part of an examination or degree. I confirm that the declarations made above are correct. I am aware of the importance of a sworn affidavit and the criminal prosecution in case of a false or incomplete affidavit. I affirm that the above is the absolute truth to the best of my knowledge and that I have not concealed anything.

Heidelberg, 22 January 2018

_____

Juliane Hafermann