University of Kent

Kent Academic Repository

Paberzyte, Darija (2021) *Exploring the Significance of the APOBEC3B Polymorphism on Cancer Risk.* Master of Science by Research (MScRes) thesis, University of Kent,.

Downloaded from
https://kar.kent.ac.uk/90083/ The University of Kent's Academic Repository KAR

The version of record is available from
https://doi.org/10.22024/UniKent/01.02.90083

This document version
UNSPECIFIED

DOI for this version

Licence for this version
UNSPECIFIED

Additional information

Versions of research works

Versions of Record
If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts
If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries
If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

Exploring the Significance of the APOBEC3B Polymorphism on Cancer Risk

A thesis submitted to the University of Kent for the degree of

Master of Science in Cell Biology

2020

Darija Paberzyte

School of Biosciences

I Declaration

No part of this thesis has been submitted in support of an application for any degree or qualification of the University of Kent or any other University or institute of learning.

Darija Paberzyte

December 2020

II Acknowledgements

First of all, I would like to express my sincere gratitude to my research supervisor, Dr Tim Fenton, who made my dream getting into cancer research possible: he always kept the research interest alive and taught me to be curious and excited, even about the smallest results. Without his guidance and support this work would not have been possible.

I would like to thank all the members of the Fenton lab for creating a very warm and enjoyable work environment. I am extremely grateful to Nicola Smith for spending her invaluable time on guiding me through this project and teaching me everything she knows herself, even the special love song to the cell cultures. Continuous encouragement to never give up and persistence during research are only a couple of the indispensable skills that I learned during our time together. Additionally, I am thankful to Maxmilan Jeyakumar for his patience in answering the same questions over and over again, and surely for making my time in the laboratory more enjoyable, especially throughout the night and weekend shifts. I would also like to thank Dr Nerissa Kirkwood for her support and advice through my PCR experiments. A very special thank you goes to Ian Reddin, who, despite my endless questions and long emails throughout the COVID-19 lockdown, was always very patient and kindly helped me tackle the computational part of the research.

Additionally, I would also like to thank all the fellow Research Master's students and my closest laboratory friends James, Tanaka, Aroub and Emmanuel for their constant support throughout the most confusing and stressful research times and a good time we had together.

Furthermore, I thank Dr Rokas Juodeikis and Dr Linas Tamosaitis, the only Lithuanian friends I made in Canterbury (who, oddly enough, both happened to share their wonderful PhD in Biochemistry journey with me), for their friendship, guidance, trust and everlasting availability round-the-clock to discuss not only scientific topics.

Finally, I am extremely grateful to my family, who supported me from the very beginning of this difficult, yet rewarding, journey and never ceased to believe in me. I would also like to express my deep gratitude to my partner Gediminas, who did everything in his power to help me fulfil my dream and encouraged my passion for the research.

III Abstract

Activity of the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) deaminases on ssDNA have been ascribed to innate and adaptive immunity, restricting replication of viruses, retrotransposons, and triggering antibody diversification in B-cells. However, emerging studies have also shown implication of deregulated APOBEC3 subgroup members, in particular APOBEC3A and APOBEC3B, in cancer mutagenesis with ~15% of all sequenced tumours displaying C>T or C>G substitutions within TCW (where W = T or A) trinucleotide motif, generating distinctive mutational signatures 2 and 13, respectively. Mechanisms, inducing characteristic off-target activity are yet to be elucidated. Nevertheless, APOBEC3A_B deletion polymorphism, resulting in a loss of entire APOBEC3B coding region, is naturally present in ~22% of global population and breast cancers from APOBEC3A_B carriers harbour a higher mutational load of signatures 2 and 13 than non-carriers. Additionally, recent work has showed that APOBEC3A_B deletion is associated with higher risk of breast cancer. In this study, we aimed to investigate the significance of APOBEC3B polymorphism on cancer risk for different regions by conducting an extensive meta-analysis of all published association studies. The results showed a significant association with breast and ovarian cancers, particularly in Asian cohorts. Moreover, single-cell cloning of novel isogenic model for APOBEC3A_B deletion was performed to identify successful clones, representing homozygous or heterozygous deletion for future studies. Lastly, Western blotting and immunochemical staining revealed APOBEC3B expression and subcellular localisation in the G_2 and mitotic phases of the cell cycle, suggesting a potential novel function during cell division. This work will aid future APOBEC studies for understanding mechanisms driving carcinogenesis in deletion carriers, by firstly elucidating functions of APOBEC3B in healthy versus cancerous cells.

IV Table of Contents

V List of Figures

VI Abbreviations

AAV	Adeno-Associated Viruses
AID	Activation-Induced Cytidine Deaminase
APOBEC	Apolipoprotein B mRNA Editing Catalytic Polypeptide-Like
bp	Base Pair
BRCA	Breast Cancer
CDA	Cytidine Deaminase
CDK	Cyclin-Dependent Kinase
DAPI	4',6-diamidino-2-phenylindole hydrochloride
DBS	Doublet Base Substitution
DSB	Double-Strand Break
DMEM	Dulbecco’s Modified Eagle Medium
Eif4g2	Eukaryotic translation initiation factor 4 gamma 2
ER	Estrogen Receptor
HA-A3B	NIKS with Endogenously Expressed HA-Tag on A3B
HIV	Human Immunodeficiency Virus
IFN	Interferon
LB	Lysogeny Broth
LINE	Long Interspersed Nuclear Element
LTRs	Long Terminal Repeats
miRNA	microRNA
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
NIKS	Normal Immortalized Keratinocytes from Skin
nuDNA	Nuclear DNA
ORF	Open Reading Frame

PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PTEN	Phosphatase and Tensin Homolog
ROS	Reactive Oxygen Species
SBS	Single Base Substitution
SDS	Sodium Dodecyl Sulphate
SINE	Short Interspersed Nuclear Element
SOC	Super Optimal broth with Catabolite repression
ssDNA	single-stranded DNA
TAE	Tris-acetate-EDTA
TBS	Tris-Buffered Saline
TCGA	The Cancer Genome Atlas
TEMED	Tetramethylethylenediamine
TRIS	Tris(hydroxymethyl)aminomethane
UNG	Uracil DNA Glycosylase 2
UTR	Untranslated Region
UV	Ultraviolet
WT-NIKS	Wild-type NIKS

1 Introduction

1.1 The APOBEC family

The apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) superfamily consists of cytidine deaminases, which have diverse functions in adaptive and innate immunity, metabolism, epigenetics, evolution and cancer (Knisbacher *et al*., 2016). APOBEC enzymes highly vary in expression within tissues and their functions, however most of them elicit a biochemical activity in deaminating cytosine bases to uracil in single-stranded DNA (ssDNA) and RNA, respectively (Salter *et al*., 2016). The APOBEC family consists of activation-induced cytidine deaminase (AID) and five APOBECs (A1-5), with human genome encoding the largest APOBEC repertoire of all organisms – eleven members of the family: AID, A1, A2, seven A3 subfamily proteins (A3A-H) and A4 (Salter *et al*., 2016). All these proteins share an evolutionary conserved zinc-dependent deaminase sequence motif, which forms cytidine deaminase (CDA) catalytic domain core. However, cytidine deamination activity for both A2 and A4 proteins still have not been shown experimentally, these remain the least studied proteins of the APOBEC family (Lada *et al*., 2011).

AID is the most ancestral member of the AID/APOBEC family, and possibly led the evolutionary process to the rest of the family through gene duplication, complex fusions and positive selection (Conticello, 2008). It is situated on chromosome 12 and is crucial in humoral immune response for antibody diversification (Muramatsu *et al*., 2000). AID targets ssDNA and deaminates deoxycytidine (dC) into deoxyuridine (dU), driving through this process gene class-switch recombination (CSR) and somatic hypermutation (SHM), targeting variable V(D)J region within

immunoglobulin genes in B lymphocytes (Henderson and Fenton, 2015; Muramatsu *et al*., 2000). The importance of AID for immune system is further highlighted by autosomal recessive AID deficiency syndrome – hyper-IgM syndrome (HIGM2), causing IgM accumulation in the body, which is driven by absence of immunoglobulin CSR and SHM (Revy *et al*., 2000).

The appearance of APOBEC1 is a recent evolutionary event, however, it was the first family member to be cloned, in 1993 (Teng *et al*., 1993). A1 is situated on the same chromosome as AID, in the range of 40 kb – 1 Mb away from the *Aid* gene in primates and is thought to result from an inverted duplication of *Aid* (Conticello, 2008). In humans, the enzyme is primarily expressed in the small intestine and some species (rodents, horses, dogs) show expression in the liver (Salter *et al*., 2016). A1 carries out deamination of cytidine 6666 to uridine in the ApoB mRNA sequence, leading to incorporation of premature stop codon, and consequently truncated ApoB protein – ApoB-48, important for dietary lipid uptake and lipid homeostasis (Daniels *et al*., 2009). It was later discovered through bacterial assays that A1 exhibits ability to also act as a DNA mutator through dC deamination (Harris *et al*., 2002). Together with earlier discovered link to cancer (through transgenic rabbits and mice, overexpressing APOBEC1 in liver and consequently developing liver dysplasia or hepatocellular carcinomas), more studies focused on elucidating mutagenic activity, regulation, function and role of A1 in cancer development and progression (Yamanaka *et al*., 1995; Saraconi *et al.*, 2014).

The second characterised member of the AID/APOBEC family was A2 (situated on chromosome 6), which is thought to be an ancestral member of the APOBEC family and is expressed only in skeletal and cardiac muscles (Conticello *et al*., 2004; Sato

et al., 2010). It has been shown that A2 is involved in normal muscle development, since the protein is associated with slow-twitch fibers and A2-deficient mice develop mild myopathy and atrophy (Sato *et al*., 2010; Sato *et al*., 2018). Recent studies have also confirmed importance of A2 in regeneration and development during embryogenesis in vertebrates (Vonica *et al*., 2011) and although it was thought that the protein is nonmutagenic (Lada *et al*., 2011), emerging evidence showed aberrant A2 expression in hepatocytes of transgenic mouse model , in which it was proposed to cause nucleotide alterations in *Eif4g2* and *PTEN* RNA sequences (Okuyama *et al.*, 2012).

A4 is located on chromosome 1 and was only recently identified as a member of APOBEC protein family (Rogozin *et al*., 2005). Comparably to AID and A2, A4 is found in all jawed vertebrates and is suggested to be one of the first APOBECs to have evolved due to its low sequence similarity at its cytidine deaminase site compared with the other APOBEC enzymes (Conticello *et al*., 2007; Conticello, 2008). A4 is primarily expressed in testes, suggesting its possible role in spermatogenesis (Rogozin *et al*., 2005). It still has not been found to carry out deamination activity in RNA or ssDNA in bacterial or yeast assays (Lada *et al*., 2011), possibly due to weak interaction with ssDNA *in vitro* (Marino *et al*., 2016). Further studies are needed to elucidate biochemical, structural and functional information about A4.

1.1.1 The APOBEC3 subfamily

Seven homologs A3A-A3D, A3F, A3G and A3H of the APOBEC3 (A3) subfamily are localised on chromosome 22 within a 200 kb region in tandem array and are considered to have originated through a series of duplication and divergence events

(Münk *et al*., 2012; Swanton *et al*., 2015). This, alongside evolutionary pressures, resulted in a wide selection of the possible substrate, while conserving zinc-dependent cytidine deaminase domains (ZD-CDAs), which function in DNA editing and viral restriction (Knisbacher *et al*., 2016). The A3 subfamily can be further classified into two subgroups, depending on the number of ZD-CDAs present in the enzyme; A3A, A3C and A3H have a single ZD-CDA, whereas A3B, A3D, A3G and A3F express two ZD-CDAs (Rebhandl *et al*., 2015). A3s are mostly known for their function in innate immunity as potent restrictors of mobile retroelements and both DNA and RNA viruses by DNA editing through deamination (Harris and Dudley, 2015; Knisbacher *et al*., 2016). The studies of A3s role in viral restriction commenced with findings of A3G restriction mechanism on HIV-1 (Sheehy *et al*., 2002), which was later shown to act in viral reverse transcription through deaminating dC➔dU on the negative-strand of cDNA, resulting in dG➔dA hypermutations on the positive cDNA strand, thus leading to excessive mutations in the viral genome. This results in viral cDNA degradation, or in the case of DNA integration in the host genome – highly defective virions (Vieira, 2013). Other A3 enzymes, in particular A3B, A3D and A3F also have been shown to exhibit anti-HIV activity in human primary CD4+ T-cells and macrophages, albeit with decreased potency (Chiu and Greene, 2008; Chaipan *et al*., 2012). A3 enzymes have been seen to exhibit antiviral role against a variety of human disease-inducing viruses, such as human papillomavirus (HPV), Epstein-Barr virus (EBV), hepatitis B virus (HBV), hepatitis C virus (HCV), human T-cell lymphotropic virus (HTLV) and Herpes simplex virus-1 (HSV-1), either *in vivo* or *in vitro* (Vieira, 2013). Interestingly, in addition to exerting deaminase activity on ssDNA, A3A, similarly to A1, displays C>U RNA editing activity in macrophages and monocytes (Sharma *et al*., 2015). As

mentioned above, members of the A3 family exhibit several inhibition mechanisms on retrotransposable elements, that are classified into long terminal repeats (LTRs) containing retroviruses and non-LTR retroelements, consisting of short (SINEs) and long (LINEs) interspersed nuclear elements (Chiu and Greene, 2008). Protection of the human genome from genotoxicity and prevention of diseases caused by activity of retroelements can also occur through deamination-independent mechanism. For instance, A3G inhibits Alu RNA element retrotransposition in SINE through physically preventing it to interfere with reverse transcriptase machinery L1 (Chiu *et al.*, 2006; Kinomoto *et al.*, 2007; Anwar *et al*., 2013). In addition, A3G and A3F directly hypermutate retroelement DNA during reverse transcription, targeting C at CC and TC sites on the negative-strand, leading to generation of (guanines (G) in bold) **G**G and **G**A motifs on the positive strand, respectively (Anwar *et al.*, 2013).

Cellular localisation and trafficking of A3 deaminases has to be tightly regulated due to the nature of their functions in the cell – the ability to edit the genome by mutating ssDNA, and to prevent genotoxicity. Single deaminase domain proteins A3A, A3C and A3H are smaller in size (~23 kDa) and exhibit even cytoplasmic/nuclear distribution, due to ability to passively diffuse into nucleus (Lackey *et al*., 2013; Henderson and Fenton, 2015). Interestingly, seven haplotypes with diverse antiviral activities of A3H have been identified in the human population and HapII, so far, is the only haplotype with predominant cytoplasmic localisation, which was shown to correlate with the RNA-binding activity of the haplotype (Zhen *et al*., 2012; Mitra *et al*., 2015). Dual deaminase domain A3s (A3B, A3D, A3F, A3G) are larger in size (~45-47 kDa) and are predominantly cytoplasmic throughout interphase, with the exception being A3B, which is localised in the nucleus (Lackey *et al*., 2013). A3G is restricted in cytoplasm through a cytoplasmic retention signal (CRS) not only during

interphase, but also is excluded from DNA in mitosis, and such localisation is also important for A3G ability to bind RNA. Nevertheless, studies showed that RNA binding is not required for retaining enzyme in the cytoplasm, suggesting a presence of additional factors involved in controlling such localisation (Bennett *et al.*, 2008; Smith *et al.*, 2012). A later, yet conflicting study, revealed that A3G is being recruited to the nucleus upon DNA damage for promoting repair of double-strand breaks (DSB) by non-homologous end joining (Nowarski *et al*., 2012; Rebhandl *et al.*, 2015).

1.2 APOBEC contribution to cancer

Genome instability and mutation is one of the underlying hallmarks of cancer (Hanahan and Weinberg, 2011). Although a vast variety of other factors, discussed by Hanahan and Weinberg, are needed for cancer formation, one of the main regulators is defective DNA repair, leading to build-up of errors in the genome and consequently, survival and progression of tumours. The emergence of publicly available data for large-scale next generation sequencing of more than 5000 tumours by The Cancer Genome Atlas (TCGA) aided the identification of mutational signatures present in cancers (Alexandrov *et al*., 2013). Such signatures are characteristic “footprints” of distinct somatic mutations in different cancers and caused by both exogenous (carcinogens in tobacco smoke and UV light) and endogenous (abnormal DNA mismatch repair, upregulation of DNA editing enzymes, e.g. APOBEC) mutation-inducing factors (Alexandrov *et al.*, 2013; Helleday, *et al*., 2014; McGranahan *et al*., 2015).

Beside protective role of APOBEC enzymes described above, loss of DNA editing regulation in the cell can become a drawback, causing enzymes to aberrantly act

on host and leading potentially to the introduction of oncogenic mutations. In fact, analysis of whole-exome sequencing data from TCGA suggested that APOBEC induced mutations contribute to up to 68% of the total SNV burden in many cancer types, such as bladder, breast, cervical, head and neck and lung cancers (Burns *et al*., 2013; Roberts *et al*., 2013). Possible APOBEC involvement in carcinogenesis was suggested upon discovery of first members of the family A1 and later AID, with the link of overexpression of A1 and consequent development of hepatocellular carcinomas, as mentioned above (Yamanaka *et al*., 1995; Smith and Fenton, 2019). AID-overexpressing transgenic mice develop lung adenocarcinomas and T-cell lymphomas due to high levels of C>T and G>A point mutations, driven by AID, observed in *c-myc* and T-cell receptor (*TCR*) genes (Okazaki *et al*., 2003). Later studies with transgenic mice have found *K-ras* being another target for mutagenic activity of AID, causing development of gastric, liver and lung cancers (Morisawa *et al*., 2008).

Contribution to mutagenesis of A2 and A4 proteins is still not fully elucidated. However, A2 deficient mice have not been found to be affected by any abnormalities during development, in comparison to A2 transgenic mice, overexpressing the gene, who have shown to have an increased number of RNA nucleotide alterations of *Eif4g2* and *PTEN* genes in hepatocytes, contributing to lung and liver tumourigenesis (Mikl *et al*., 2005; Okuyama *et al*., 2012). A recent study revealed importance of A2 in sustaining normal function of mitochondria in skeletal muscles. A2 knockout mice had dysmorphic mitochondria, inducing generation of reactive oxygen species (ROS) and activating mitophagy pathways, which is indirectly implicated in cancers by reducing nuclear stability (Sato *et al*., 2018). Although no published studies could be found on A4 relation to cancer, data from The Human

Protein Atlas revealed that high RNA expression of A4 is observed in endometrial cancer (data used from TCGA) and high/medium protein expression is observed in liver, stomach and pancreatic cancers in the decreasing order.

The A3 group of enzymes, in particular A3A and A3B, are mostly known for their contribution to cancer mutagenesis through induction of mutations in both nuclear and mitochondrial DNA (Suspene *et al*., 2011). Mainly, single base substitution (SBS) mutations are observed in primary cancers and A3 leaves a couple of distinctive mutational signatures in a broad array of cancers, named SBS-2 and SBS-13 (Nik-Zainal *et al.*, 2012; Roberts *et al*., 2013; Alexandrov *et al.*, 2019). Whole-genome sequencing of 21 breast cancers revealed small genomic regions of base substitution hypermutations at the cytosine (in bold) in the context of T**C**W (where W=A/T, C>G or C>T), named “kataegis” (translated from ancient Greek as “thunderstorm”) (Nik-Zainal *et al.*, 2012; Roberts *et al*., 2013). SBS-2 is characterised by C>T transitions with less C>G transversions at T**C**W trinucleotides, while SBS-13 is conversely enriched with C>G transversions (Alexandrov *et al.*, 2013). Interestingly, DBS11– a newly found doublet base substitution (DBS) with yet unknown aetiology, is composed mainly with CC>TT mutations and was associated with SBS-2, suggesting another possible mutagenic effect of A3 enzymes (Alexandrov *et al.*, 2019). It has also been noted that such mutational patterns often occur in close proximity to chromosomal rearrangement breakpoints, where abnormal repair of DSBs and stalled replication forks result in accessible substrate for APOBEC editing–ssDNA (Roberts *et al*., 2012; Roberts *et al*., 2013; Henderson and Fenton, 2015). Initially, studies on A3G suggested its preferential deamination of C>U in 3’→5’ orientation on a ssDNA (Chelico *et al.*, 2006), making it a suggested model of tumorigenic activity for other A3s. Upon discovery of strand-

coordinated mutation clusters, it was further elucidated that A3A and A3B favour T towards the 5' of the strand within 5'-T**C**W-3' motif (where W=A/T, mutated base in bold) (Burns *et al*., 2013). Studies with yeast, grown with DNA alkylation-inducing agent, revealed clusters of simultaneous (within the same cell cycle) point mutations of C and G, that occur during DSB repair. Mutations were observed on 5' cytosines and 3' guanines after bi-directional resection of DSB, suggesting A3 activity on opposing DNA strands following resection either side of the DSB (Roberts *et al*., 2012). Another yeast study suggested uracil-DNA glycosylase (UNG), which removes deoxyuridine (dU) prior DNA replication in base excision repair (BER), to be important constituent of kataegis. This is due to generation of DSBs, caused by cuts of endonucleases at UNG-created abasic sites (Taylor *et al*., 2013; Petljak *et al*., 2019). The phenomenon of kataegis was also observed in many other cancer types, including lung, liver, pancreatic, prostate, head and neck, leukemias (chronic lymphocytic and acute lymphoblastic) and B-cell lymphomas (Alexandrov *et al*., 2013; Roberts *et al*., 2013). Furthermore, analysis of TCGA data revealed APOBEC-mediated mutagenetic activity even in the oncogenes, such as *EGFR*, *PTEN*, *TP53* and *PIK3CA* (Alexandrov *et al*., 2013; McGranahan *et al*., 2015).

1.2.1 APOBEC3A and APOBEC3B in cancer

APOBEC3A and APOBEC3B are the key APOBEC3 family enzymes, implicated in tumourigenesis, tumour evolution, cancer progression and metastasis and/or therapy resistance in many cancers, thus emphasising the importance of understanding regulation, role in cell-cycle, mechanism of action and processes of mutagenesis of these deaminases (Caval *et al*., 2014; Chan *et al*., 2015; Henderson and Fenton, 2015).

A large proportion of head and neck squamous cell carcinomas (HNSCC), cervical squamous cell carcinomas and endocervical adenocarcinomas (CESC) are caused by infection of high-risk HPV types (Henderson *et al*., 2014). A positive correlation was established between APOBEC signature mutations, displayed in these cancers, APOBEC expression levels and activation of immune signaling upon viral infection (Cannataro *et al.*, 2019). Sequencing of ~300 HPV^+ and HPV^- HNSCC tumours revealed increased A3B expression and number of TCW point mutations in HPV^+ tumours, whereas HPV^- tumours had a higher burden of tobacco-associated signature mutations (Henderson *et al*., 2014). Amongst those mutations, observed in HPV^+ HNSCC, characteristic G>A substitutions (E542K and E545K), situated within helical domain of *PIK3CA*, were suggested to be mediated by A3 deamination on the negative strand (Henderson *et al*., 2014). This finding was later supported by a quantitative assessment of A3B deamination of *PIK3CA* hotspots *in vitro* (Cannataro *et al.*, 2019). Sequencing of various regions of seven non–small cell lung cancer (NSCLC) tumours for analysis of driver mutations in clonal evolution also revealed a presence of *PIK3CA* hotspot mutations in an APOBEC context, implicating abberant activity of A3B on driver genes in various cancers, with HPV acting as a selective pressure on certain mutations, although it is yet to be elucidated which enzyme A3A, A3B, or synergistic activity of both is responsible for the observed mutations (de Bruin *et al*., 2014; Henderson and Fenton, 2015).

1.2.1.1 APOBEC3A activity in the cell and role in cancer

Localisation of A3A in the cell is equally distributed in both cytoplasm and nucleus, as mentioned previously, however, forced overexpression (such as by transfection) of this deaminase is genotoxic and cytotoxic to the cells (Landry *et al*., 2011; Mussil *et al*., 2013; Caval *et al*., 2014; Suspène *et al*., 2017). A3A has been shown to induce

C>T hypermutations in DNA and cause DSBs, which, if unrepaired lead to apoptosis or cell cycle arrest in G_1 phase. Deleterious A3A activity is proposed to cause stalled replication forks in S phase and, consequently, induce cell-cycle arrest in early S phase, which is too late for G_1 checkpoint (Landry *et al*., 2011; Mussil *et al*., 2013). Before discovery of APOBEC signature mutations in cancers, mutagenic activity of A3A has been investigated in mitochondrial (mtDNA) and nuclear (nuDNA) DNA, with findings showing editing activity in exons of the *MYC* and *TP53* genes. Mutated proto-oncogene *MYC* is found in Burkitt lymphoma and multiple myeloma patients, and the tumour-suppressor *TP53* is the most frequently mutated gene in more than half of all human cancers, proposing contribution of A3A hyperediting in cancer development (Landry *et al*., 2011; Suspène *et al*., 2011; Dang, 2012*;* Surget *et al.*, 2013). However, a checkpoint protein human Tribbles Homolog 3 (*TRIB3*), has been shown to have a protective function in genome stability against malicious A3A editing. In particular, physical interaction of TRIB3 with A3A reduces nuDNA editing by preventing A3A translocation to the nucleus, which was additionally observed through dC deamination in *CMYC* site. *TRIB3* had restricting effect on A3A hyperediting when expression of TRIB3 in the cell was exceeding that of A3A. Additionally, it also maintains integrity of DNA, minimising occurrence of γH2AX-activating DSBs, induced by extensive deamination of A3A (Aynaud *et al*., 2012).

The discovery of kataegis, and attribution of novel signature mutations to A3A and A3B activity, firstly suggested A3B (as a nuclear protein) to be the leading enzyme responsible for clusters of mutations in multiple cancers (**Fig.1)**, in particular breast cancer (Burns *et al*., 2013a; Burns *et al*., 2013b). However, more recent studies have shown A3A to have higher potency than A3B to predominantly drive carcinogenesis, through stronger induction of DNA damage and, consequently,

generating more point mutations in the genome (Chan *et al*., 2015; Cortez *et al*., 2019). This was surprising, since A3B was considered to be critically involved in mutagenesis driving breast cancer (BRCA), due to significantly upregulated mRNA expression levels and correlating increased mutational loads (signature base substitutions), observed in more than half of all sequenced breast tumours and in about three quarters of cancer cell lines (Burns *et al*., 2013b; Kanu *et al*., 2016). Interestingly, A3A is now considered to be the primary source of APOBEC-induced mutations, even in breast cancer (Cortez *et al.,* 2019). Comparative analysis of 28 whole-exome sequenced BRCA cell lines revealed that knockdown of A3A but not A3B expression in BT474, CAMA-1 and MDA-MB-453 cancer cell lines had a dramatic effect on deaminase activity, despite A3B mRNA expression levels being ~240-fold higher than those of A3A (Cortez *et al*., 2019). Additionally, APOBEC signatures SBS-2 and SBS-13 were found to be predominant out of all observed mutation signatures in Taiwanese oral squamous cell carcinoma (OSCC) patients, with observed stronger A3A expression and contribution to mutational burden in tumours. The study proposed that increased interferon (IFN) signaling, that induces A3A and A3B gene expression and is prominent in oral cancers, combined with the reduced activity of A3B 3'UTR due to involvement of miRNA, could drive the upregulation of A3A in OSCC (Chen *et al*., 2017). Further studies are needed to elucidate whether A3A and A3B expression levels and activity in the cell are co-dependent, however, in the study conducted by Cortez and colleagues, induced mutagenesis of A3A in TCGA cancer genomes was proposed to be independent from A3B in the cell (Cortez *et al*., 2019). The C-terminal deaminase domain of A3B shares 91% similarity with the A3A catalytic domain, which could be a potential reason for mismatching observed RNA expressions with the appropriate deaminases from RNA-seq analyses (Caval *et al*., 2014; Starrett *et al*., 2016).

Additionally, yeast studies have showed ability of both deaminases to act on each other's preferred trinucleotide (T**C**W) sequences, which could be another caveat in distinguishing unique contributions of A3A and A3B in cancer (Starrett *et al*., 2016). Nevertheless, current distinction of A3A caused mutations in the genome against A3B is alleviated by their preferred motifs in the ssDNA, revealed by yeast studies, with shown preference of A3A for YT**C**A, and A3B for RT**C**A (where Y stands for pyrimidine, R stands for purine and cytosine in bold is a targeted nucleotide for deamination) (Chan *et al*., 2015; Chen *et al*., 2017). Interestingly, a study by Chan and colleagues also observed enrichment of analogous YT**C**A mutations in human cancer genomes, in particular bladder, breast, cervical, head and neck, lung squamous cell carcinoma and lung adenocarcinoma; suggesting A3A to be a more potent mutator than A3B, as discussed earlier (Chan *et al*., 2015).

Whilst considering the question of which member of A3 family plays the key role in carcinogenesis, it is worth noting that observed elevated levels of A3 mRNA in tumours might not necessarily represent the exact cause of cancer development, since expression pattern might vary throughout the process of tumourigenesis. Studies, investigating APOBEC-driven genome editing upon HPV infection in cervical cancers, have been using HPV+ keratinocytes and pre-cancerous cervical lesions as an appropriate model to further understand A3 expression throughout different stages of tumour development (Warren *et al*., 2014; Henderson and Fenton, 2015).

1.2.1.2 APOBEC3B activity in the cell and role in cancer

Analyses of emerging cancer genome and whole exome datasets implicated A3B to be the leading candidate in APOBEC-induced mutations for almost a decade

(Figure 1) (Burns *et al*., 2013a; Burns *et al*., 2013b; Roberts *et al.*, 2013; Harris, 2015). In particular, research of breast cancer brought attention to significance of A3B and raised further questions whether APOBEC is a potential cancer driver gene or a downstream effector of aberrant activity of other driving forces. Links of A3B with various oncogenes were observed in several studies: Burns and colleagues observed elevated mRNA levels of A3B in majority of breast cancer cell lines having positive correlation with C>T mutation burden in *TP53*, leading to inactivation of this tumor suppressor gene (Burns *et al*., 2013b). Later studies showed *TP53* has a regulatory role in A3 activation: p53 indirectly represses the expression of A3B through involvement of p21, which by inhibiting cyclin-dependent kinases 4 and 6 (CDK4/6) negates hyperphosphorylation of p107/p130 retinoblastoma proteins. This in turn allows recruitment of p107/p130 and E2F4 to the DREAM repressor complex, which is directed to the A3B gene promoter site and consequently, inhibits A3B expression (Menendez *et al*., 2017; Periyasamy *et al.*, 2017). Interestingly, activated p53 during chromosomal stress conditions (agents Nutlin, doxorubicin) has been proposed to bind not only to A3B, but also to A3A/C/H transcriptional regulatory regions, showing a novel contribution to A3/*TP53* interaction as a transcriptional regulator (Menendez *et al*., 2017). A weak association of A3B with the *PIK3CA* proto-oncogene was also observed in several studies on HPV-associated cancers, as mentioned previously in section **1.2.1** (Henderson *et al*., 2014; Cannataro *et al.*, 2019).

Clinical studies focusing on breast cancer tumour evolution associated with APOBEC dysregulation, have suggested A3B as a possible biomarker in estrogen receptor positive (ER^+) cancers. Increased levels of A3B mRNA, observed in primary breast tumours, are associated with worse prognosis in ER^+ versus ER^-

breast cancer patients, as measured by both shorter overall and disease-free survival (Sieuwerts *et al*., 2014; Harris, 2015). Conversely to higher A3B expression levels observed in ER^- primary tumours, A3B is considered to be prognostic for ER^+ cancers. Particularly, it is thought to drive tumour evolution and, consequently, resistance to anti-cancer drugs and relapse, as emerging resistance to tamoxifen in response to overexpression of A3B was observed in clinical results and mice models (Law *et al*., 2016). Later studies have compared A3B expression levels in matched metastatic tumours to primary breast tumours; the results revealed further increase in enzyme levels within metastases (Sieuwerts *et al*., 2017). Additionally, independent from ER expression status in tumours, distant metastases had higher A3B expression than local metastases in lymph nodes, implying that A3B has a key role throughout cancer progression and is one of the elements, strongly fueling clonal evolution (Law *et al*., 2016; Sieuwerts *et al*., 2017).

Investigation of A3B activity and expression levels in cancer cell lines and clinical samples, such as tumours, can aid understanding associations with certain clinicopathological features and outcomes of the disease, including disease-free survival and overall survival of the patients. Endogenous A3B has been found to drive C>U deamination in serous ovarian cancer, which is the most aggressive and common ovarian cancer type (Leonard *et al.*, 2013). The study showed a significant connection between A3B expression levels and preferred C>A and C>G transversions within ssDNA; as well as visualised localisation of A3B in the nucleus. This was supported by later studies, suggesting A3B being exclusively nuclear due to specific amino acid sequence on the N-terminal region, crucial for nuclear import of the protein (Caval *et al*., 2014; Suspène *et al*., 2017). On contrary, immunohistochemistry analysis of cancerous ovarian tissues revealed that A3B

localisation was predominant in the cytoplasm, with a significant correlation between over-expression of A3B and higher FIGO (International Federation of Gynecology and Obstetrics) stage of the sample. Additionally, knockdown of A3B in human ovarian cancer cell lines resulted in decline of cell viability, prompting A3B being one of key factors for viability maintenance (Du *et al*., 2018). Interestingly, high A3B expression was also seen in >40% of nasopharyngeal carcinoma patients, with expression being significantly higher in those, who had recurrence of metastasis within 5 years (Feng *et al*., 2020). Evaluation of over 200 gastric cancer tissues in Chinese population further confirmed worse prognosis for cancer patients, which correlated with elevated A3B expression, possibly due to extensive editing and interference with cell death regulating protein PDCD2 (Zhang *et al.*, 2015).

Together, these findings in a vast array of cancers create a strong ground for creating a link between over-expression of A3B and poor prognosis for cancer patients, suggestively by enriching proto-oncogenes with somatic mutations, exacerbating DNA replication stress, interfering with transcription of cell-survival regulating proteins and affecting viability of cancerous cells (Henderson and Fenton, 2015; Zhang *et al*., 2015 Kim *et al*., 2020). Interestingly, A3B has been also proposed to have a role in ER-stimulated tumour growth in BRCA *in vitro* and *in vivo*, that is completely independent from its mutagenic effects (Periyasami *et al*., 2015). A3B knockdown in ER^+ cancer cell lines MCF7 and T47D has shown to inhibit tumour growth together with reduced expression of ER-responsive genes. C-to-U deamination by A3B at ER binding regions causes generation of DSBs through base excision repair and non-homologous end-joining pathways, consequently stimulating chromatin remodeling and RNA polymerase II recruitment to the ER binding region to promote gene expression (Periyasami *et al*., 2015). Thus, A3B has

been considered as a potential marker for multiple human cancers for a while already, in particular for prediction of disease development, risks of metastasis and relapse. Additionally, A3 subfamily members, which have been found to be active mutators in cancers (A3A, A3B and A3H), are also viewed at as a promising diagnostic and therapeutic targets, that could be combined with existing therapy strategies (Olson *et al.*, 2018). However, extensive research of A3A and A3B regulation and roles in the cell, their interaction with proto-oncogenes, as well as further understanding of factors, driving deregulation of these enzymes is essential to successfully contribute to the development of new treatment strategies.

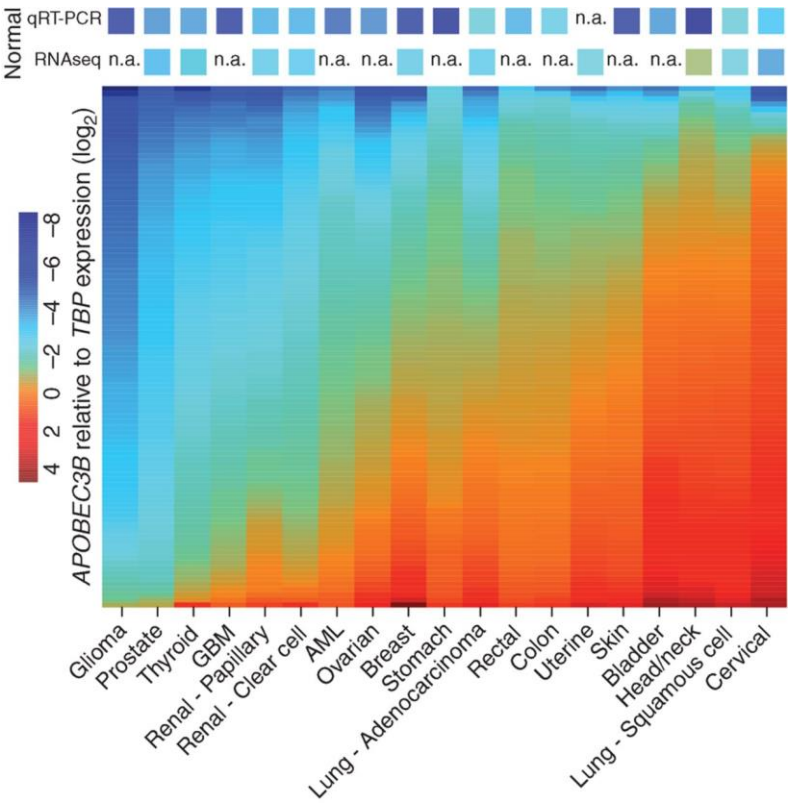


Figure 1. APOBEC3B expression in nineteen different cancer types.

APOBEC3B mRNA expression values were normalised to TATA-binding protein (TBP) to allow quantitative comparison between qRT-PCR and RNAseq results. The expression levels were visualised using heatmap, where red and orange colours indicate elevated A3B expression, green and blue represent values from normal tissue. AML – acute myeloid leukemia; GBM – glioblastoma multiforme; n.a – not available. *Image adapted from Swanton et al., 2015.*

1.3 APOBEC3A_B deletion polymorphism and link to cancer

Regardless of suggested A3B importance in cancer development, it seems that A3B is not a critical protein in humans, due to a widely spread A3B deletion polymorphism phenotype in the human population. Additionally, this germline deletion is estimated to be present in ~40% of all humans, carrying at least 1-copy number deletion (otherwise known as heterozygous deletion carriers), and has also been associated with increased risk of breast cancer and elevated levels of APOBEC-associated signature mutations found in tumours of cancer patients (Chan *et al*., 2015; Cortez *et al*., 2019; Kidd *et al*., 2007; Long *et al*., 2013; Nik-Zainal *et al*., 2014).

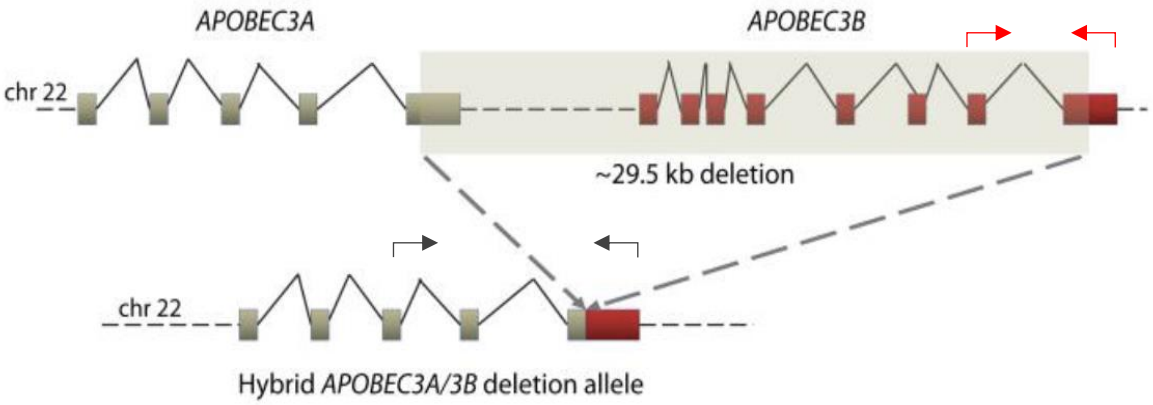


Figure 2. **Formation of APOBEC3A_B deletion allele.**
The 29.5 kb deletion spans from APOBEC3A (shown in green) exon 5 up to APOBEC3B (shown in red) exon 8, resulting in a loss of APOBEC3B coding region. The truncated deletion allele is a product of fused APOBEC3A open reading frame (ORF) with APOBEC3B 3'-untranslated region (UTR). Red arrows indicate insertion and grey arrows indicate deletion PCR primers, created by *Kidd et al., 2007. Image adapted from Nik-Zainal et al., 2014.*

This common ~29,500-bp deletion fragment (although, a more recent analysis by 1000 Genomes Project, revised the number up to 29,936-bp) spans from the fifth exon of the A3A gene up to the eighth exon of the A3B gene on chromosome 22

(A3A exon 5 is 100% identical to A3B exon 8), resulting in full deletion of A3B coding region and fusing completely unaltered and functional A3A gene with 3' UTR of A3B **(Figure 2)** (Kidd *et al*., 2007).

In literature, the fusion gene is referred to as hybrid A3A, A3A-A3B fusion, or A3A_B, and the resulting mRNA has been shown to be up to 20-fold more stable than the wild-type A3A when transfected into HEK 293T cells (Caval *et al*., 2014; Klonowska *et al*., 2017). It has been shown experimentally, that elevated levels of polymorphic mRNA, therefore upregulated A3A_B protein, lead to DSB formation and DNA damage in SKBR3 cells (Caval *et al*., 2014). Although the exact cause of altered stability has not yet been elucidated, involvement of the chimaeric A3B 3'-UTR regulation (instead of normal A3A 3'-UTR) has been proposed. Several cancer case-control studies in Asia have observed correlation between downregulation of certain miRNAs and increased expression and stability of A3A_B gene: miR-34b-3p is reported to be downregulated in various cancers (such as endometrial serous adenocarcinoma, cervical and NSCLC), as it has tumour suppressor function, as well as being an A3B gene repressor (Wang *et al*., 2013; Revathidevi *et al*., 2016). A study conducted in South Indian population proposed its regulatory role for A3A_B transcript by targeting A3B 3'-UTR (miR-34b-3p regulates A3B through repressing) and saw a significant downregulation of miR-34b-3p in cervical cancer samples, leading to reduction of repression by A3B 3'-UTR, overexpression of A3A_B and APOBEC-associated somatic mutations in tissues (Revathidevi *et al*., 2016). Additionally, miR-409 (miRNA, which targets and reduces activity of A3B 3'-UTR, but not A3A 3'-UTR) was found to be downregulated in Taiwanese OSCC samples, proposing its regulatory role in enhanced stability of A3A-A3B fusion protein. Furthermore, A3A expression levels were found to be significantly elevated in

tumours of the homozygous and heterozygous individuals for the A3B deletion, compared to those with the functional A3B gene, however the regulatory link between A3A and A3A_B deletion is still unknown and more research is needed (Chen *et al*., 2017).

The deletion is particularly of interest for the cancer research due to its high prevalence rates across the globe, with <1% in African populations, 6% in European, ~37% in East Asian, ~58% in Amerindian, and astounding 93% seen in Oceanic populations (Kidd *et al*., 2007). Emerging evidence has shown that individuals with this deletion are also carrying a higher incidence of SBS-2 and SBS-13 mutational signatures observed in cancers, in comparison to non-carriers, despite the complete loss of A3B protein (Caval *et al*., 2014; Nik-Zainal *et al*., 2014). In particular, the earlier genome-wide association studies have linked this deletion polymorphism with approximately two-fold elevated risk of breast cancer in Chinese, Japanese and European populations. Beyond this, copy number of the gene has shown to be significantly associated with the cancer risk – individuals, carrying homozygous deletion are predisposed to higher risk in comparison to those with heterozygous deletion (Komatsu *et al.*, 2008; Long *et al*., 2013; Xuan *et al*., 2013). Another genome-wide association study in a small cohort of Chinese women have also ascribed such association of increased cancer risk to copy number variation in epithelial ovarian cancer, suggesting a functional role of A3A_B deletion polymorphism in various cancers, with no restriction to a particular one (Qi *et al*., 2014; Cortez *et al*., 2019). Intriguingly, the results and overall consensus on the deletion polymorphism role started to diverge with the increasing number of population or hospital-based association studies. In contrast to previous findings, the case-control study conducted in Swedish (Caucasian) population has found no

risk or survival association for breast cancer (Göhler *et al*., 2016). This was later supported by the study conducted in another Scandinavian country, Norway, showing no risk association found with any of the deletion genotype not only for breast, but for other inspected cancers (lung, prostate and colon) as well (Gansmo *et a*l., 2018). Outside European region, similar results for breast cancer were observed in Moroccan, Brazilian and Indian populations (Marouf *et al*., 2016; Vitiello *et al*., 2020; Revathidevi *et al*., 2016).

It has been hypothesised that such variation in risk association could be related to stratification across the human population, with increased risk observed in cohorts where deletion is more prevalent. However, as briefly mentioned above, significant discrepancies are found in regions outside Africa and Europe, where deletion is commonly found within the population. It is still unclear what other factors drive such variation in the results, hence all the up to date published association studies of A3A_B deletion and various cancers will be collected and overviewed in Results section **3.1** to obtain a clearer picture of the polymorphism's association with cancer type-specific risk.

1.4 APOBEC – a novel target for cancer therapy

Throughout the past years the emerging studies focused primarily on A3A and A3B out of A3 group due to their implication in cancer mutagenesis. These enzymes are of particular importance due to their driving roles in tumour evolution and intra-tumour heterogeneity in the case of deregulation, which is essential in the context of cancer development and therapeutics. Additionally, it is worth noting that germline A3B deletion allele, commonly found in certain populations, has also presented higher breast and ovarian cancer risk. This makes A3 group of enzymes an

attractive novel therapeutic target in cancers, where APOBECs are found to be overexpressed and consecutively, lead to somatic hypermutation.

Despite the knowledge gained throughout the past decade on behaviour of A3A and A3B in various cancers, it still remains unclear what drives such common deregulation characterised by recurring mutational signatures. Currently, there are no APOBEC inhibitors available for use. However, there are several therapeutic strategies being proposed in studies, such as therapy by hypomutation, therapy by hypermutation, immune checkpoint blockade and combination of several therapies. First, and the simplest strategy – therapy by hypomutation, otherwise known as development of inhibitors for A3 directly, or targeting inhibition of signaling pathways, leading to A3 gene expression, was suggested by many studies (Leonard *et a*l., 2015; Kanu *et al*., 2016; Olson *et al.*, 2018). In this approach, the objective would be to suppress the accumulation of somatic mutations and hence limit tumour evolution driven by APOBEC, which eventually would limit adverse outcomes of the illness, including metastasis, drug resistance and decreased survival. The proof of principle has been shown experimentally in breast cancer cell lines by targeting ATR/Chk1 (which is activated by replication stress) pathway, which consequently activates transcription of A3B. Results showed consistency of two Chk1 inhibitors potently inhibiting increase of A3B transcription, thus decreasing its mutagenic activity in cancer cells (Kanu *et al*., 2016; Nikkilä *et al*., 2017). Additionally, another signaling pathway has been previously found to be implicated in A3 upregulation: PKC/NF-κB induces elevated expression for both A3A and A3B, however other A3 family members have not been found to be affected by this signal transduction pathway (Leonard *et al*., 2015). Importantly, the study showed that PKC inhibitor successfully reduced A3B mRNA levels by more than half in multiple cancer cell

lines, including breast, ovarian and head and neck, where endogenous A3B is constitutively elevated (Leonard *et al*., 2015). Besides focusing on the main A3 members, commonly detected in cancers, other deaminases of the APOBEC family have been also looked at as potential targets, albeit, to a lesser extent. For instance, A3G is found to be overexpressed in mesenchymal glioblastomas (the most aggressive type) and correlate with poor survival prognosis (shorter life-span), as well as to be involved in radioresistance (due to promotion of DSB repair through Chk2 pathway) and migration of the cells, which infiltrate the brain rendering surgery impossible. Instead, cells with the A3G knockout, hence attenuated DNA repair mechanism, showed increased sensitivity to irradiation and successfully underwent apoptosis (Wang *et al*., 2017). Although successful small molecule inhibitors for A3G were developed almost a decade ago due to emerging interest in its potency against Vif-deficient HIV-1, efforts to target A3B have lagged behind but the co-structure of the protein bound to a small molecule was recently revealed (Li *et al*., 2012; Shi *et al*., 2020). These findings are a leap forward, aiding development and optimisation of A3B ligands and consequently, small molecule inhibitors, which eventually could be particularly beneficial in some cancers, combined with other treatments.

Second strategy, and more challenging than the first, is therapy by hypermutation, which was initially based on inhibiting Vif, as this protein causes A3G to be degraded by the proteasome, however, adaptation of this therapy to A3A and A3B in particular, is also suggested to be effective in cancers (Olson *et al.*, 2018). The idea behind this therapy lies in increasing the mutational load within an APOBEC-high cancerous cell to the cytotoxic levels with small molecule A3 agonists, leading to targeted and specific cell death. Recent studies have shown ATR inhibitors

successfully sensitise cells to accumulation of ssDNA (substrate for A3A and A3B enzymes), causing apoptosis in A3A and A3B-upregulated various cancer cell lines (including AML, breast, ovarian, head and neck and others) (Green *et al.*, 2017; Buisson *et al*., 2017). Interestingly, lethality of the cells overexpressing A3B, can also be achieved by inhibition of uracil DNA glycosylase 2 (UNG), which activates base excision repair pathway for repairing genomic DNA, with the condition that mismatch repair proteins and p53 remain functional (or can become with pharmacological intervention) for uracil processing (Serebrenik *et al*., 2019).

Besides these intra-cellular therapeutic approaches, immune checkpoint blockade is gaining more interest as a complementary therapy. As it was established, that APOBECs drive tumour evolution and heterogeneity, and have distinctive mutational footprints in the genome, the next would be incorporation of immune system to recognise neoantigens from APOBEC-upregulated tumours (Swanton *et al.*, 2015). Previous genome studies revealed that SBS-2 signature mutations, attributable to A3A, are both clonal and subclonal (i.e. they occur throughout tumour development), although more commonly in the subclonal, or in later stages of tumour evolution. On contrary, SBS-13 signature was more prevalent in the clonal, or earlier part of the tumour evolution (McGranahan *et al*., 2015). Such therapy might therefore be expected to be predominantly beneficial for cancers displaying SBS-13, since high-clonal neoantigen burden containing tumours respond better to immune checkpoint blockade and importantly, have been shown to result in longer progression-free survival (McGranahan *et al*., 2016).

The emerging data of APOBEC3-driven drug resistance in cancer also suggests the need of combination therapy, to attenuate disease progression. In treated tumours,

APOBEC may act in a feed-forward loop, as DNA-damaging drug induces DNA breaks and causes accumulation of ssDNA, which in turn is a substrate for A3A and A3B, leading to mutational load increase and introduction of novel mutations in oncogenes (Kanu *et al*., 2016; Law *et al*., 2016). Such interaction of A3 deaminase activity and cytotoxic chemotherapy drugs would explain therapy resistance obtained from rapid evolution and high heterogeneity, commonly seen in cancer with upregulated A3B activity. In breast cancer cell lines, common cytotoxic drugs inducing replication stress such as hydroxyurea, gemcitabine, camptothecin, aphidicolin and a selective ER modulator tamoxifen increase mRNA expression levels and activity of A3B (Kanu *et al*., 2016; Law *et al*., 2016). Additionally, as elevated expression of A3B in ER+ breast cancer has been previously associated with poor patient survival, a study by Law *et al*. linked tamoxifen resistance in primary ER+ breast cancer tumours with overexpression of A3B mRNA (Sieuwerts *et al*., 2014; Periyasami *et al*., 2015; Law *et al*., 2016). Further, murine xenograft ER+ model studies have shown that genetic knockdown of A3B improves response of tamoxifen therapy (Law *et al*., 2016). A correlation between A3 upregulation, the A3 mutational signature and resistance to chemotherapy has also been observed in bladder cancer (Middlebrooks *et al.,* 2016; Faltas *et al*., 2016). Interestingly, all studied bladder cancer cell lines treated with PKC inhibitor (refer to first therapeutic strategy, described above) achieved a very small reduction of A3B mRNA levels, suggesting that alternative signaling pathway exists for induction of A3B (Leonard *et a*l., 2015).

To summarise, it is clear that more studies are needed to address the knowledge gap of A3A and A3B induction mechanisms and differentiate between early and late stage induced mutations in tumour microenvironment. This would aid development

of small molecule inhibitors, to limit A3-driven cancer evolution and consequently therapy evasion and recurrence. Additionally, understanding occurrence time of signature mutations in each cancer individually, where APOBECS are found to be upregulated, would help administer when suitable and limit escape from immune checkpoint blockade therapy. At the moment, combination therapy seems to be the most effective method to inhibit A3-driven resistance in tumours and increase overall survival in patients.

1.5 Aims of the project

The focus of this project fell on both sides of APOBEC3B implication in cancer – absence (germline deletion of APOBEC3B) and presence (functional APOBEC3B protein) of the protein in NIKS cells, a non-transformed keratinocyte cell line. In fact, it aimed to continue two studies, previously carried out in the laboratory, hence the aims and objectives are further divided into two sections for clarity:

Investigation of APOBEC3A_B deletion polymorphism part of the work aimed to isolate and characterise single-cell clones (SCCs) that had germline A3A_B deletion. The previous MSc-R student Rosalba Biondo has successfully generated a novel isogenic model for A3A_B deletion in NIKS cell line using CRISPR/Cas9 technique and created a collection of SCCs, in which heterozygous or homozygous deletion event has occurred. After the screening stage and successful identification of clones with the deletion, by genotyping assay, the initial plan was to demonstrate the loss of A3B in the cell population and observe expression of mRNA levels of remaining APOBEC3 genes and whether/how A3B deletion affects their expression using qPCR. Additionally, it was of our interest to compare the findings between clones with homozygous deletion versus ones with the heterozygous deletion.

Unfortunately, due to unexpected laboratory shutdown in March, caused by COVID-19, this work could not be continued beyond the screening phase.

The second part of the project focused on understanding localisation and a putative cell-cycle related function of APOBEC3B in the NIKS, to elucidate which normal functions of A3B are absent in deletion carriers. This work was based on results, recently obtained by the PhD student Nicola Smith in the laboratory, where A3B protein levels were found to be the highest in the G_2/M phase of the cell cycle **(Figure 3)**. Immuno-fluorescent staining of A3B was implemented to observe localisation of the protein in mitotic NIKS cells, as well as any association with the cellular structures. Initially, investigation of the cell-cycle related functions of A3B included looking at the solubility and stability of the protein, since there is no evidence for significant mRNA increase during mitosis despite protein accumulation, hence the question rises whether this is due to post-translational modification leading to increased stability or due to increased translation. Albeit, we were not able to pursue the later part of the work fully due to the laboratory shutdown.

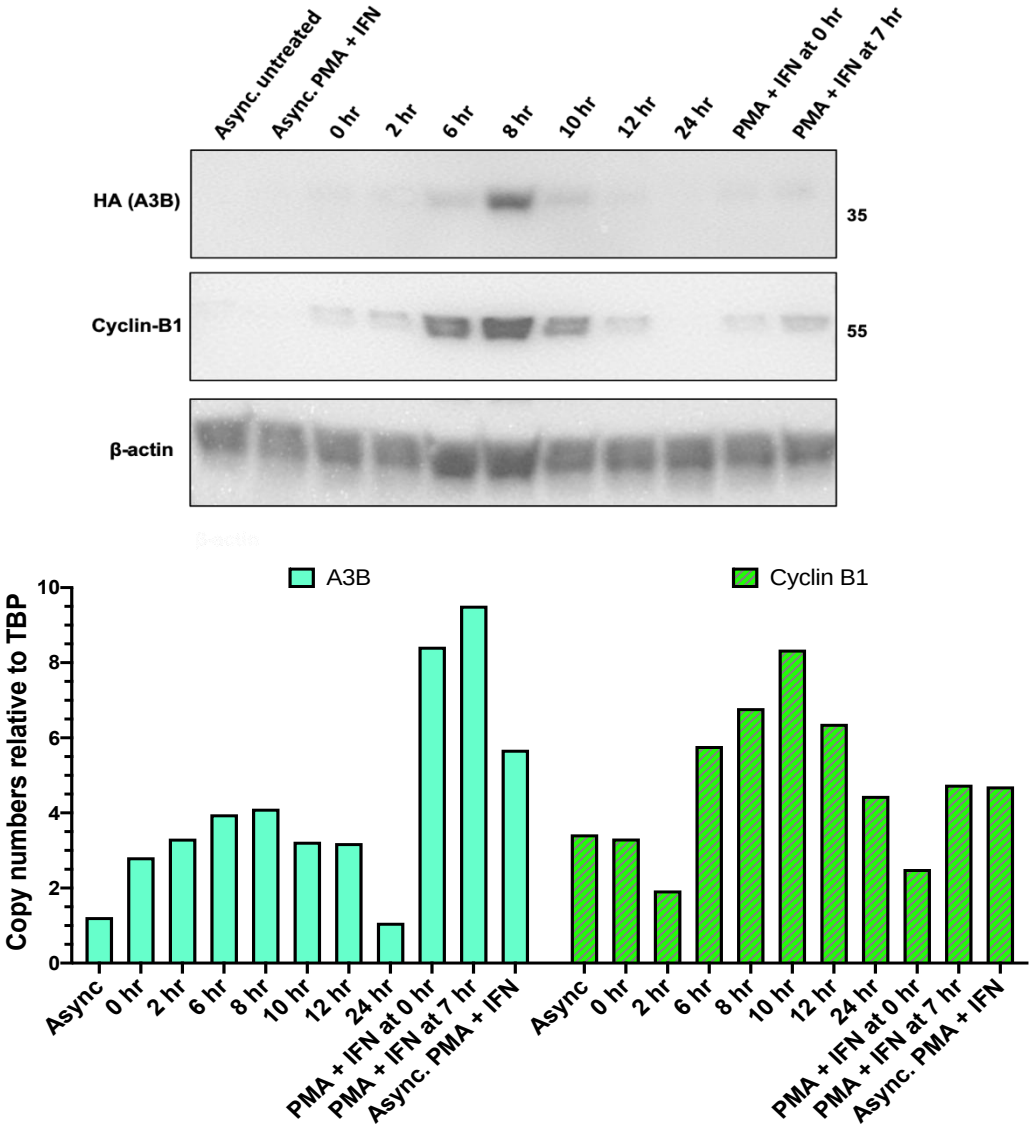


Figure 3. APOBEC3B is expressed at G_2/M phase of the cell cycle.

HA-A3B and WT-NIKS were plated at a density of 3 x 105 cells per well in 6-well plates. The next day, cell synchronisation at the G1/S boundary was initiated by incubating with 222 μl/cm^2 of fresh media containing 2 mM of thymidine for 18 hrs. In addition, asynchronous cells or those at the G1/S boundary (0 hr), or those released into fresh media for 7 hrs were treated with 100 ng/ml PMA and 1000 U/ml IFN-α for 6 hrs. Protein levels were detected by separating 20 ug of whole cell lysate and probing with anti-HA antibody (Cell Signalling, 1:1000). RNA was extracted from WT-NIKS, from which cDNA was synthesised. mRNA levels were detected by qPCR using WT-NIKS, and copy numbers generated from standard curves are shown relative to TBP*. Image obtained from Nicola Smith; unpublished.*

2 Materials and Methods

2.1 Tissue culture

2.1.1 Culture media preparation

FC medium was prepared by mixing 1 part of Dulbecco's modified Eagle's medium with 1% L-Glutamine (DMEM, Invitrogen 11965-084) with 3 parts of Ham's F-12 medium (Invitrogen 11765-054). It was further supplemented with 5% Foetal Bovine Serum (FBS, Pan Biotech P30-3031), 24 µg/mL Adenine (Sigma A2786-5G), 8.3 ng/mL Cholera Toxin (EMD Biosciences 227036), 5 µg/mL Insulin (Sigma I6634-50MG), 0.4 µg/mL Hydrocortisone (Sigma H0888-1G), 10 ng/mL Epidermal Growth Factor (EGF, Sigma E1257-0.1MG) and 1% (v/v) Penicilin-Streptomycin solution – 10,000 U/mL Penicillin, 10 mg/mL Streptomycin (Pan Biotech P06-07100). Media was then filter sterilised and stored at 4°C.

FC medium

Chemical	Volume
DMEM with 1% L-Glutamine	110 mL
Ham's F-12 medium	335 mL
Fetal Bovine Serum	25 mL
100X Hydrocortisone	5 mL
100X Cholera Toxin	5mL
100X Insulin	5mL
100X Adenine	5 mL
100X Epidermal Growth Factor	5mL
Penicilin-Streptomycin	5mL

DMEM medium was supplemented with 10% (v/v) FBS and 1% (v/v) Penicilin-Streptomycin (10,000 U/mL Penicillin, 10 mg/mL Streptomycin) prior to filter sterilisation. The medium was stored at 4°C.

2.1.2 J2-3T3 Mouse Fibroblasts (“feeder” cells)

Feeder culture was maintained in DMEM medium in humidified atmosphere at 37°C / 5% CO_2, with media change every 2-3 days. Upon reaching ~80% confluency, the culture was treated with 4 µg/mL Mitomycin-C (Sigma M4287-2MG) in DMEM medium and incubated in humidified atmosphere at 37°C / 5% CO_2 for 4 hours in order to arrest the growth of cells. The culture was neutralised with 1X Phosphate Buffered Saline (PBS, Oxoid), incubated with 0.25% trypsin for 2-5 minutes at 37°C and re-suspended in DMEM medium to neutralise trypsin prior pelleting the cells at 300 x g for 6 minutes. Treated and washed feeder culture was then stored in DMEM medium with 10% (v/v) dimethyl sulfoxide (DMSO, Fisher Scientific D/4121) freezing mix at -80°C for future use in NIKS co-cultures.

2.1.3 Normal Immortalised Keratinocytes from Skin (NIKS)

Prior to plating NIKS cell suspension, mitomycin C treated J2-3T3 mouse fibroblasts were left to attach for at least 2 and up to 24 hours to the base of T25, T75 or T175 cell culture flask. NIKS were co-cultured on a layer of growth-arrested fibroblasts at a ratio 3:1 respectively, in FC medium and humidified atmosphere at 37°C / 5% CO_2. FC medium was changed every 2 days if cells were prepared for downstream use in experiments and every 3 days if the culture was maintained for passaging. Expanded culture was stored in freezing mix, described above, and stored at -80°C for future use. NIKS cells above passage 6 were not used for experimental purposes.

2.2 Single-cell cloning

2.2.1 Generation of Single-Cell Clones

Generation of single-cell clones was performed by low-density seeding method. Transfected and sorted polyclonal NIKS pools, generated by R. Biondo, were grown in T25 cell culture flasks for expansion and upon reaching 50-60% confluency, the cells were harvested by trypsinisation and deactivated in FC medium. Concentration of cells in homogenised suspension was counted and serial dilutions were carried out to obtain 10 cells/mL final concentration. 50 µL of cell suspension was transferred into each well of a 96 well-plate, to obtain cell density of 0.5 cells per well. Mitotically inactivated feeder cells were aliquoted at 50 µL to the 96 well-plate at a concentration of 20,000 cells/mL at least two hours prior to plating NIKS. Plates were initially screened after 4-5 days, and after that every day to visually identify wells, containing no more than one colony. Monoclonal lines, that were viable and successfully reached confluency, were trypsinised and passaged into a 24 well-plate, and subsequently to a 6 well-plate, after which were harvested for freeze-downs and DNA extraction.

2.2.2 Purification and quantification of genomic DNA

QIAamp DNA mini kit (QIAGEN™, 51306) was used for genomic DNA purification from cultured cells, following the standard protocol with the recommended step. Alternatively, later in the project, KAPA Express Extract Kit (Sigma Aldrich, KK7100) was employed for DNA extraction, following steps 1 and 2 in the KAPA Express Extract Protocol. Additionally, ethanol precipitation method was used to purify the DNA afterwards. DNA samples were mixed by vortexing with 0.1 v/v 3M sodium acetate (Sigma S2889) and 3 v/v ice cold 100% ethanol (Fisher), after which were

placed in the -80°C freezer for 1 hour. Later, samples were centrifuged at 13000 rpm for 30 minutes at 4°C. Pellets were washed with 0.5 mL of ice cold 75% ethanol twice and centrifuged at 13000 rpm for 10 minutes at 4°C each time. Supernatant was carefully removed without disturbing the pellet and samples were centrifuged again at 13000 rpm for 10 seconds, trying to remove ethanol traces. Lastly, the pellets were left to air-dry for 30-60 minutes and resuspended in 200 µL of nuclease-free water (New England Biolabs). DNA concentration in samples was quantified in triplicate using the NanodropOne© (ThermoFisher).

2.2.3 PCR genotyping assay

List of Primers

All primers were purified by standard desalting method. Supplied by Integrated DNA Technologies.

Name	Sequence	Description
TRF635	TAGGTGCCACCCCGAT	Kidd *et al*., 2007 deletion forward
TRF636	TTGAGCATAATCTTACTCTTGTAC	Kidd *et al*., 2007 deletion reverse
TRF639	TGTCCCTTTTCAGAGTTTGAGTA	Kidd *et al*., 2007 insertion 2 forward
TRF640	TGGAGCCAATTAATCACTTCAT	Kidd *et al*., 2007 insertion 2 reverse

PCR genotyping assay was carried out to distinguish the A3B deletion and insertion genotypes of obtained single-cell clones. Each sample was genotyped separately with deletion and insertion primers, which were originally developed by *Kidd et al.* Following primer pairs were used: TRF635 (forward primer): TAGGTGCCACCCCGAT and TRF636 (reverse primer):

TTGAGCATAATCTTACTCTTGTAC for generating an amplified 700-base pairs deletion sequence product, and TRF639 (forward primer): TGTCCCTTTTCAGAGTTTGAGTA and TRF640 (reverse primer): TGGAGCCAATTAATCACTTCAT for generating a 705-base pairs (bp) insertion PCR product.

PCR was performed following the HotStarTaq *Plus* Master Mix Kit protocol without the addition of 10x CoralLoad Concentrate. Reactions were carried out in 20 µL total volume, mixing reagents displayed above with 100 ng of DNA template (however, in cases of low DNA concentration, 70 ng was used for the assay). Cycling conditions used were following: 5 mins at 95°C, 30 cycles of 0.5 min at 94°C, 0.5 min at 54°C, 1 min at 72°C, and a single cycle of 10 min at 72°C.

Reagent	**Volume**	**Final Concentration**
HotStarTaq Plus Master Mix, 2x	10 µL	1U DNA polymerase, 1x PCR Buffer, 200 µM of each dNTP
Forward Primer	1 µL	0.5 µM
Reverse Primer	1 µL	0.5 µM
RNase-Free Water	Variable	-
Template DNA	Variable	100 ng
Final volume	**20 µL**	**-**

2.2.4 DNA Electrophoresis

DNA fragments were separated and identified on a 1% agarose gel in 1x TAE buffer with ethidium bromide (Sigma Aldrich, E1510) at a final concentration of 0.5 µg/mL. Before loading, 10 µL of each DNA sample was mixed with either 2 µL 6x Blue/Orange Loading Dye (Promega, G190A) or alternatively, 9 µL of sample was mixed with 1 µL 10x BlueJuice™ Gel Loading Buffer (Invitrogen, 10816015). DNA samples were loaded directly into wells alongside 5 µL of Quick-Load® Purple 1 kb

Plus DNA Ladder (New England Biolabs, N0550S), or 1 kb Plus DNA Ladder (Invitrogen, 10787018). The gel was run at 100V for 40-50 minutes and visualised under a UV transilluminator or Syngene G:Box imager.

2.3 Cell synchronisation experiments

2.3.1 Cell Synchronisation at G_2/M Phase

Cell synchronisation was achieved by chemical blockade method with either DNA synthesis inhibitor thymidine (Sigma Aldrich, T1895-1G) or inhibitor of microtubule formation nocodazole (Sigma Aldrich, M1404-2MG). Prior to treatment, NIKS were passaged 2-3 times and grown under the conditions, mentioned in the section **2.1.3**. WT-NIKS and HA-A3B NIKS were plated at density of 2.5 x 10^5 cells per well on a monolayer of 1 x 10^5 feeder cells in 6-well plate or scaled up, if required, retaining the same ratio of cells per area. Single thymidine block was carried out when NIKS reached 40-50% confluency, medium was replaced with FC medium, containing 2mM thymidine. 18 hours later, cells were washed twice with pre-warmed PBS and further incubated in FC medium for either: 1) 8 hours to release from thymidine block and allow cells to enter G_2/M cell cycle phase; or 2) for 2 hours prior to nocodazole treatment. Media was replaced with FC media containing 100 ng/mL nocodazole and the cultures were incubated for further 10 hours, to arrest at G_2/M phase. Cell synchronisation was confirmed by microscopical examination of the cultures, as NIKS change morphologically and become more rounded-up before undergoing mitosis.

2.3.2 SDS-PAGE and Western Blotting

Prior to analysis, cells were lysed by incubating scraped cells on ice for 15-30 minutes upon addition of either MPER or RIPA lysis buffers, supplemented with 1x protease inhibitor (cOmplete tablets EASYpack, Roche, #04693116001), 1x phosphatase inhibitor (PhosSTOP, Roche, #04906837001) and 0.2U/µL benzonase (Sigma, E1014-25KU). Lysates were further centrifuged at 13,000 x g for 20 mins at 4°C. The proteins were denatured by mixing the samples with 4x Laemmli buffer and incubating on a heating block at 70°C for 10 minutes, followed by SDS-PAGE for 20 µg of protein, using Criterion™TGX™Precast Gel (Bio-Rad, #5671125) and running buffer containing 25mM Tris Base, 192mM Glycine, 0.1% SDS. PageRuler Plus Prestained Protein Ladder (Thermo Scientific, #26619) used for quantitating observed protein sizes. Electrophoresis was performed at 150V for 55 minutes, after which proteins were transferred onto an activated in ethanol Polyvinylidene Difluoride (PVDF) membrane. PVDF membrane was blocked in Tris-buffered saline (TBS) with 0.1% Tween-20 (Sigma, P1379) (this further will be referred as TBST) and 5% (w/v) skimmed milk powder (Oxoid, LP003) for 1 hour at room temperature with gentle agitation, followed by three washes for 5 minutes. The proteins were then incubated overnight at 4°C in primary antibodies: rabbit anti-HA (1:1000, Cell signalling #3724S), rabbit anti-pH3 (1:1000, Cell signalling #2947S), mouse anti-Cyclin B1 (1:500, Santa-Cruz #sc-245). The next morning, the membrane was washed in TBST three times for 10 minutes and incubated on a rocker for 1h in room temperature in following HRP-conjugated secondary antibodies: Donkey anti-rabbit HRP-conjugated (1:10,000; Invitrogen #A16023), or Donkey anti-mouse HRP-conjugated (1:10,000; Invitrogen #A16011). Finally, the membrane was washed in TBST three times for 10 minutes before being exposed on film or imaging with

Syngene G:Box-fluorescence imager, with the addition of Enhanced Chemiluminescence (ECL) reagent (BioRad #1705061).

4x Laemmli Buffer

Chemical	Concentration in dH_2O
Tris base	0.25M
Sodium dodecyl sulfate	0.08 g/L
Glycerol	0.4 mL/mL
2-Mercapto-ethanol	0.2 mL/mL
Bromphenol blue	0.8 mg/mL

2.4 Immunofluorescent microscopy

2.4.1 Sample preparation

Prior to seeding the cells, coverslips were sterilized by autoclaving, inserted in wells of 6-well plates, washed once with 70% ethanol, rinsed twice with sterile dH2O and left to air-dry under UV light for 10-15 minutes. WT-NIKS and HA-A3B NIKS were seeded at density of 3 x 10^5 cells per well on a monolayer of 1 x 10^5 feeder cells. Cells were allowed to settle and continue cycling for 20-24 hours to reach 40-50% confluency, prior to synchronization by single-thymidine block (with 2mM thymidine) for 18 hours. In order to arrest the cells in G_2/M cell cycle phase, cells were released in fresh FC media for 8-hours.

2.4.2 Fixing, Staining and Mounting

Upon reaching a fixing time point, the coverslips were washed twice with cold PBS and fixed in 1 mL of 4% paraformaldehyde (PFA) for 15 minutes at room temperature. The coverslips were washed with PBS three times, permeabilised with 1 mL 0.3% Triton X-100 (Sigma, T8787) in TBS and washed with TBS twice. The cells were subsequently blocked in filtered 1 mL of 2% (w/v) BSA in TBS for 1 hour

at room temperature. Without additional washes, the coverslips were inverted and incubated in dark with primary antibody (all antibodies used listed in the table below) in diluent (filtered TBS with 1% (w/v) BSA and 0.05% Tween-20 (Sigma, P1379)) for 2 hours at room temperature. Coverslips were then washed three times with filtered TBS with 0.1% (w/v) BSA and 0.05% Tween-20 for 10 minutes on a rocker and incubated in dark in secondary antibody, diluted in the same diluent as primary antibody, for 1 hour at room temperature. Lastly, the cells were washed with filtered TBS with 0.1% (w/v) BSA and 0.05% Tween-20 twice for 10 minutes and afterwards with TBS for 10 minutes, stained with NucBlue™ Live ReadyProbes™ nuclear stain (Invitrogen, R37606) in TBS for additional 5 minutes and mounted on a glass microscope slide, using ProLong™ Gold antifade (Invitrogen, P10144) without DAPI. The coverslips were stored overnight in the dark at room temperature to set before storing at 4°C.

Name	Species	Dilution	Supplier
	Primary antibody		
HA-Tag	Rabbit mAb	1:500	Cell Signalling #3724S
Phospho-Histone H3	Mouse mAb	1:500	Cell Signalling #9706
	Secondary antibody		
Alexa Fluor 488	Donkey anti-Mouse	1:1000	Invitrogen #A21202
Alexa Fluor 594	Donkey anti-Rabbit	1:1000	Invitrogen #A21207

2.4.3 Optimisation of immunofluorescence for A3B

A panel of six different methanol and PFA fixation methods was set with HA-A3B NIKS arrested in G_2/M cell cycle phase to optimize immunofluorescence of A3B by minimising background staining. Sample preparation was the same as described in section **2.3.1**, however only a single NIKS treatment with thymidine and subsequent 8-hours post release was chosen as a single condition for visualisation.

Methanol fixation method was carried out according to Immunofluorescence Protocol with Methanol Fixation from Cell Signaling with two different conditions: either 1) 5% Donkey Serum (Sigma, D9663) or 2) 2% BSA was used for preparation of Blocking Buffer. 4% PFA fixation method incorporated four conditions: 1) 5% Donkey Serum or 2) 2% BSA in Blocking Buffer, and performing subcellular fractionation by extracting 3) cytoplasmic and loosely held nuclear proteins or 4) tightly held nuclear fraction with cytoskeleton (CSK) buffer. CSK buffer was prepared fresh and filtered on the day of fixing the coverslips, consisting of 10mM PIPES, 300mM Sucrose, 100mM NaCl, 3mM MgCl2 and 1mM EDTA. Coverslips were firstly washed with ice cold PBS twice and incubated with 200µL of ice-cold CSK buffer, containing 0.1% (v/v) Triton X-100, for 1 minute, to extract loosely held proteins. Two wash steps with ice-cold PBS followed and either the coverslips were fixed with 4% PFA at 4°C for 30 minutes or were further incubated on ice in 200µL CSK buffer with 0.5% (v/v) Triton X-100, for 20 minutes, prior to fixation step.

2.4.4 Visualisation

The samples were visualized using a Widefield CytoVision Olympus BX-61 epifluorescence microscope with a cooled Hamamatsu Photonics Digital CCD camera, using DAPI, FITC and Texas Red filters at the magnifications of x20,000, x60,000 and x100,000.

2.5 Meta-Analysis

2.5.1 Identification of association studies for the meta-analysis

Relevant association studies were searched in the PubMed database up to June 2020. Keywords including “APOBEC3 OR APOBEC3B” AND “deletion OR

polymorphism OR polymorphisms OR variant OR SNP“ AND “cancer OR cancers OR tumour OR tumours OR malignancy” were used for search purposes. Additionally, the references from each eligible article were screened manually to identify more potential studies that could be included. Case-control and genome-wide association studies were included in the meta-analysis if: 1) the study looked specifically for association between A3A_B deletion polymorphism and any type of cancer risk and 2) there was enough data provided to calculate insertion (I) and deletion (D) alleles present in controls and cases, and to calculate pooled ORs with corresponding 95% CIs for different inheritance models.

2.5.2 Calculations and statistical analysis

For each study the following data was collected for the meta-analysis: authors of the study, year published, overall population size and individual numbers for cases and controls, ethnicity of the population, cancer type, and where possible, numbers of I/I, I/D and D/D carriers in cases and controls and percentage of A3B deletion alleles within cases and controls. If latter data elements were not publicly available, they were calculated manually (Equation 1, Equation 2).

Equation 1. Calculation for I/D and D/D individuals in case-control groups.

$$ID\ alleles + DD\ alleles = Del\ alleleles\ (\%)\ /\ (\frac{100\%}{total\ number\ of\ alleles})$$

Equation 2. Calculation for A3A_B allele deletion frequency in case-control groups.

$$Del\ alleles\ (\%) = (\frac{100\%}{total\ number\ of\ alleles}) \times (ID\ alleles + DD\ alleles)$$

Statistical analysis was performed using MedCalc for Windows, version 19.6 (MedCalc Software, Ostend, Belgium). MedCalc odds ratio (OR) calculator was used for computation of the OR, standard error (SE), 95% confidence interval (95% CI) and standard normal deviate (z-value). P-value was calculated according to Sheskin, 2004 (p. 542). The formulas for OR, 95% CI, SE and z-value are available online at https://www.medcalc.org/calc/odds_ratio.php. The data collected is presented in **Supplementary Table 3**.

3 Results

3.1 Association of APOBEC3A_B deletion genotype with increased risk of cancer

There is no clear consensus to what extent A3A_B deletion contributes to increased risk of cancer due to numerous contradicting studies and the data obtained. This overview of existing association studies on A3A_B deletion polymorphism will aim to reflect the ones carried out for the last decade and gain a clearer picture to what degree A3B is responsible for cancer development and progression.

After initial search in the PubMed database with the keywords specified in Chapter 2.5.1, a total of 100 records were identified that matched the criteria, and consequently screened (**Figure 4**). The complete list of all the publications screened for the meta-analysis is presented in the **Supplementary Table 1.** During the screening process, shown in **Figure 4**, 87 records were excluded, due to not studies being not entirely related to the topic (n=27), were reviews or previously conducted meta-analyses (n=11), or were not association or case-control studies (n=49). Thus, for this meta-analysis, 20 association studies (including genome-wide association studies and case-control studies, which identified A3A_B genotypes by PCR) from 13 publications (published 2013-2020) were selected, which included 9 distinct cancer types: breast, ovarian, cervical, prostate, bladder, colon, lung, oral and liver (hepatocellular carcinoma, HCC). Predominantly, breast cancer association studies comprise this review, due to increased public interest and focus on this specific cancer type. In total, 26 516 cases and 33 724 controls were subjected in the analysis. The characteristics of each association study, such as sizes and ethnicity of case-control groups, frequency of the A3A_B deletion allele, cancer type and odds ratios (OR) with 95% confidence intervals (95% CI) for effect of one or two-

copy number deletions compared to subjects with no deletion, are summarised in **Table 1**. Original statistical values from each study and extended data of statistical results obtained for this analysis and can be found in **Supplementary Table 2** and **Supplementary Table 3,** respectively. It was decided to focus on two genetic inheritance models – heterozygous codominant (I/D vs I/I, where I – insertion allele and D – deletion allele; otherwise known as 1-copy deletion vs no deletion) and homozygous codominant (D/D vs I/I, where I – insertion allele and D – deletion allele; also known as 2-copy deletion vs no deletion) for cancer risk association with A3A_B deletion investigation.

The results for heterozygous codominant model are shown in **Figure 5**. The findings show association between heterozygous deletion carriers in the general population and increased risk of breast cancer [OR(95%CI)=1.21(1.15–1.27), p<0.0001], ovarian cancer [OR(95%CI)=1.42(1.21–1.65), p<0.0001] and HCC [OR(95%CI)=1.78(1.47–2.16), p<0.0001]. Conversely, there is no evidence for significant effect of the A3A_B heterozygous deletion on cancer risk in bladder [OR(95%CI)=1.03(0.91–1.17), p=0.5953], cervical [OR(95%CI)=1.17(0.69–1.99), p=0.5628], prostate [OR(95%CI)=0.91(0.78–1.07), p=0.2654], colon [OR(95%CI)=0.94(0.80–1.10), p=0.4286], lung [OR(95%CI)=1.08(0.92–1.27), p=0.3528] and oral [OR(95%CI)=0.96(0.57–1.60), p=0.8655] cancers. However, it is noteworthy, that only one association study for each of these cancer types was carried out so far.

Sub-analysis of the same cancer-type risk in heterozygous individuals across different geographic regions or ethnicities was carried out for breast and ovarian cancers. The findings for breast cancer were accordant across the studies, showing

significant effect of A3A_B deletion in Asian region/populations [OR(95%CI)=1.31(1.22–1.40), p<0.0001], however in European region the association was lower and not statistically significant [OR(95%CI)=1.06(0.97–1.16), p=0.1641]. Single studies from Northern Africa [OR(95%CI)=0.64(0.34–1.21), p=0.1685] and Latin America [OR(95%CI)=0.76(0.51–1.12), p=0.1622] regions also showed no statistically significant association of A3A_B deletion with breast cancer.

The findings for homozygous codominant model are shown in **Figure 6**, with data pointing to a significantly increased risk shift for individuals harbouring two-copy A3B deletion. Overall, risk of breast cancer [OR(95%CI)=1.57(1.44–1.72), p<0.0001] and ovarian cancer [OR(95%CI)=2.99(1.53–5.87), p=0.0014] considerably grew for homozygous deletion carriers. In addition, single study on HCC by *Zhang et al., 2013* has also indicated significant association between A3B 2-copy deletion and risk of HCC development [OR(95%CI)=2.31(1.69–3.17), p<0.0001]. Results for bladder [OR(95%CI)=1.51(1.12–2.05), p=0.0071], cervical [OR(95%CI)=0.70(0.23–2.11), p= 0.5318], prostate [OR(95%CI)=0.82(0.47–1.43), p= 0.4913], colon [OR(95%CI)=0.70(0.38–1.31), p= 0.2679], lung [OR(95%CI)=1.23(0.72–2.12), p= 0.4462] and oral [OR(95%CI)=0.18(0.02–1.34), p= 0.0941] cancers were in agreement with those, seen in heterozygous codominant model (**Figure 5**), as the evidence is not strong enough to state that A3A_B had an effect.

Sub-analysis of region/population significance on risk of developing certain cancer with 2-copy deletion of A3B gene has shown even higher risk for breast cancer in Asia [OR(95%CI)=1.65(1.49–1.83), p<0.0001], although for European [OR(95%CI)=1.17(0.78–1.73), p=0.4491] and Latin American [OR(95%CI)=0.49(0.15–1.60), p=0.2368] regions the effect was not significant. The

African study by *Marouf et al., 2016* could not be included in this comparison, as the study did not have any individuals with 2-copy deletion genotype.

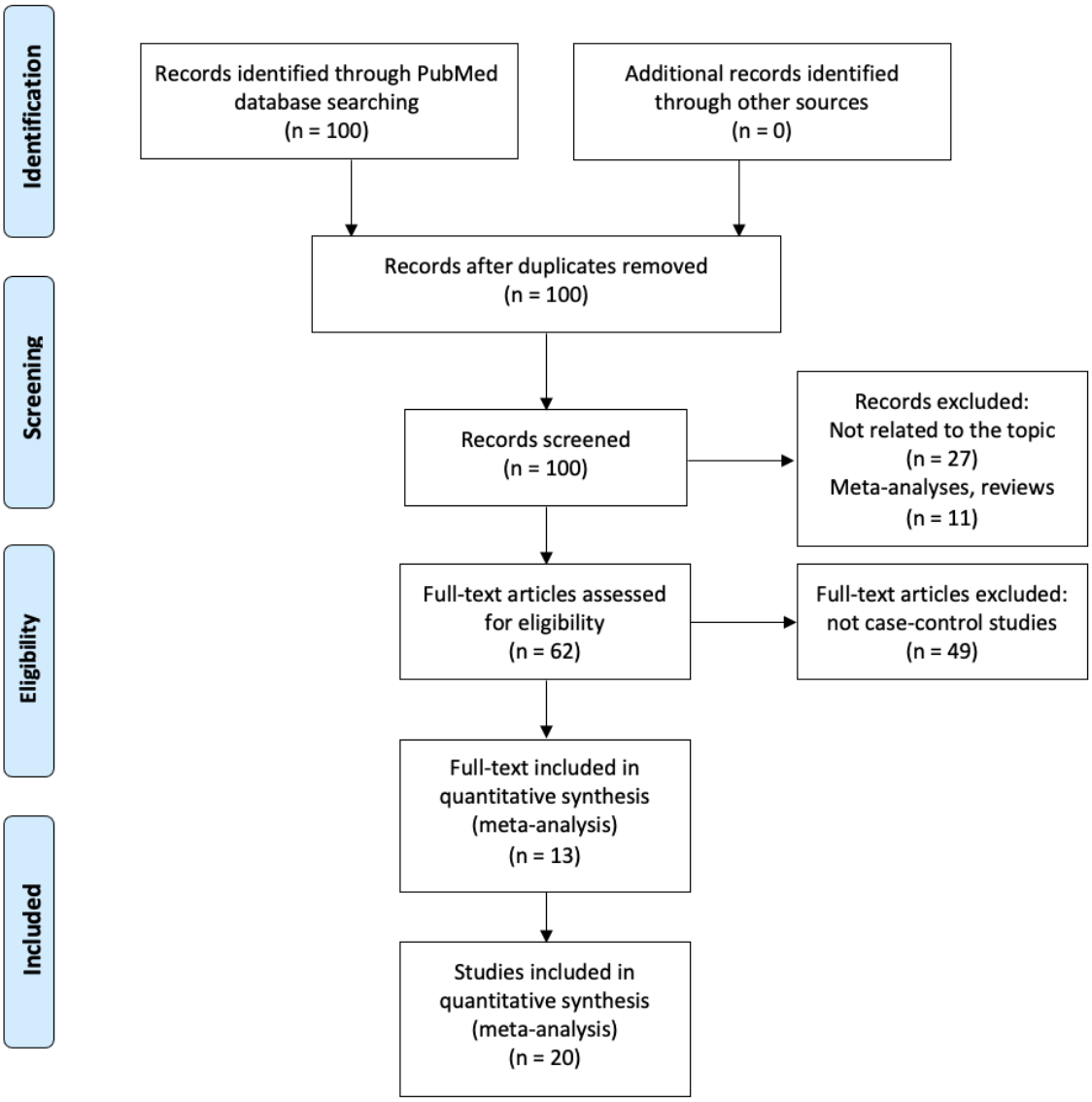


Figure 4. Flow diagram of study selection for the meta-analysis.

PRISMA flow chart showing the process of identification and selection of eligible association studies for the meta-analysis. The PRISMA diagram was accessed from *Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097.*

Table 1. **Overview of association studies of the APOBEC3A_B deletion allele and cancer risk.**
A3A_B deletion allele frequency within the studied population is displayed in percentage (ΔA3B freq.). Odds ratios (OR) are presented for I/D vs. I/I (effect for one-copy deletion) and for D/D vs. I/I (effect for two-copy deletion). BC – breast cancer; BlC – bladder cancer; CC – cervical cancer; ColC – colon cancer; HCC – hepatocellular carcinoma; LC – lung cancer; OC – oral cancer; OvC – ovarian cancer; PC – prostate cancer.

Study, Year	Population size	Cases (ΔA3B freq.)	Controls (ΔA3B freq.)	Ethnicity	Cancer type	I/D OR, (95%CI)	D/D OR, (95%CI)
Zhang, 2013	1,950	1,124 (37.9%)	826 (27.5%)	Chinese	HCC	1.78 (1.47–2.16)	2.31 (1.69–3.17)
Xuan, 2013	3,273	1,671 (12.4%)	1,602 (10.4%)	European-ancestry	BC	1.20 (1.01–1.42)	2.23 (1.01–4.91)
Long, 2013	11,622	5,792 (40.5%)	5,830 (34.0%)	Chinese	BC	1.32 (1.22–1.42)	1.76 (1.57–1.98)
Qi, 2014	2,938	1,374 (14.1%)	1,564 (10.2%)	Chinese	OvC	1.42 (1.19–1.69)	2.73 (1.29–5.79)
Rezaei, 2015	479	262 (21.6%)	217 (17.3%)	Iranian	BC	1.57 (1.07–2.31)	0.80 (0.24–2.68)
Göhler, 2016	2,341	780 (9.7%)	1,559 (8.9%)	Swedish	BC	1.16 (0.92–1.45)	0.79 (0.28–2.22)
Wen, 2016	2,893	1,451 (36.9%)	1,442 (32.8%)	Chinese, Malaysian, Indian	BC	1.23 (1.05–1.44)	1.38 (1.10–1.74)
Revathidevi, 2016	573	224 (19.0%)	349 (18.5%)	Indian	BC	1.23 (0.84–1.81)	0.77 (0.37–1.63)
	473	88 (18.2%)	349 (18.5%)		CC	1.17 (0.69–1.99)	0.70 (0.23–2.11)
	575	97 (12.9%)	478 (17.2%)		OC	0.96 (0.57–1.60)	0.18 (0.02–1.34)
Marouf, 2016	426	226 (4.2%)	200 (6.3%)	Moroccan	BC	0.64 (0.34–1.21)	-
Middlebrooks, 2016	4,285	1,719 (5.8%)	2,566 (6.9%)	European-ancestry	BlC	0.76 (0.63–0.92)	1.93 (0.81–4.60)
	2,061	1,116 (25.9%)	945 (27.6%)	Japanese	BlC	0.86 (0.72–1.03)	0.94 (0.67–1.32)
Klonowska, 2017	5,390	2,532 (6.5%)	2,858 (6.5%)	European	BC	0.96 (0.82–1.13)	1.69 (0.69–4.14)
	1,502	440 (4.5%)	1,062 (5.6%)	European	OvC	0.71 (0.47–1.05)	2.35 (0.47–11.68)
Gansmo, 2018	3,687	1,769 (8.9%)	1,918 (9.4%)	Norwegian	BC	0.98 (0.82–1.16)	0.65 (0.31–1.33)
	5,187	1,360 (10.2%)	3,827 (9.4%)		LC	1.08 (0.92–1.27)	1.23 (0.72–2.12)
	5,412	1,585 (8.7%)	3,827 (9.4%)		ColC	0.94 (0.80–1.10)	0.70 (0.38–1.31)
	4,474	2,565 (8.7%)	1,909 (9.5%)		PC	0.91 (0.78–1.07)	0.82 (0.47–1.43)
Vitiello, 2020	738	341 (8.7%)	397 (11.6%)	Brazilian	BC	0.76 (0.51–1.12)	0.49 (0.15–1.60)

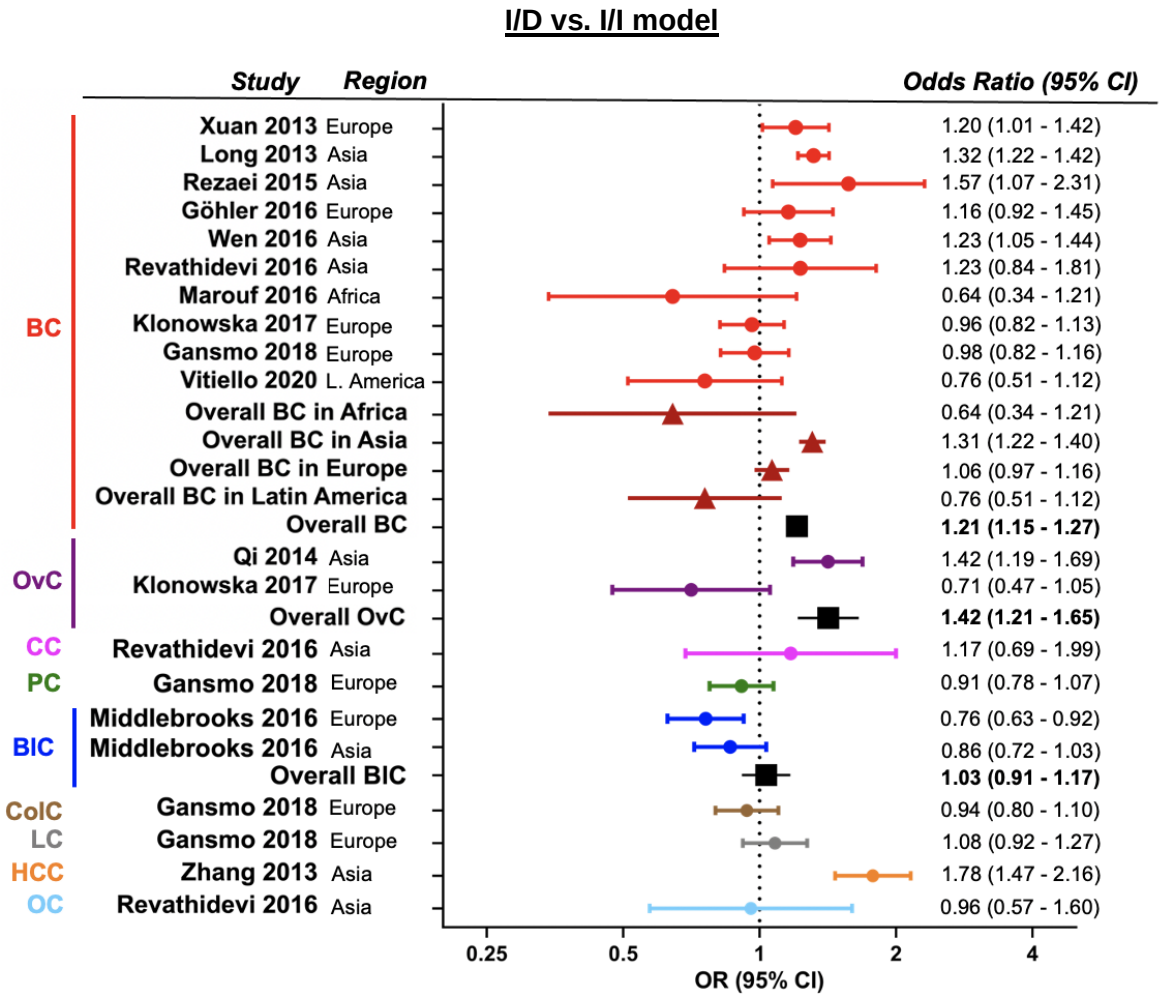


Figure 5. APOBEC3A_B deletion polymorphism and cancer risk for one-copy deletion genotype model.

Forest plot illustrating meta-analysis results of all studies on association of A3A_B deletion with various cancer types. Details of each of the study are indicated on the left. Odds ratio with corresponding 95% confidence intervals (CI), individually calculated for each study, are indicated on the right and visualised in the forest plot by different colours for each cancer type. Vertical dotted line (OR=1) represents no effect. Summary values are represented by black squares. The results are plotted on Log2 scale for visualisation purposes. BC – breast cancer; BlC – bladder cancer; CC – cervical cancer; ColC – colon cancer; HCC – hepatocellular carcinoma; LC – lung cancer; OC – oral cancer; OvC – ovarian cancer; PC – prostate cancer.

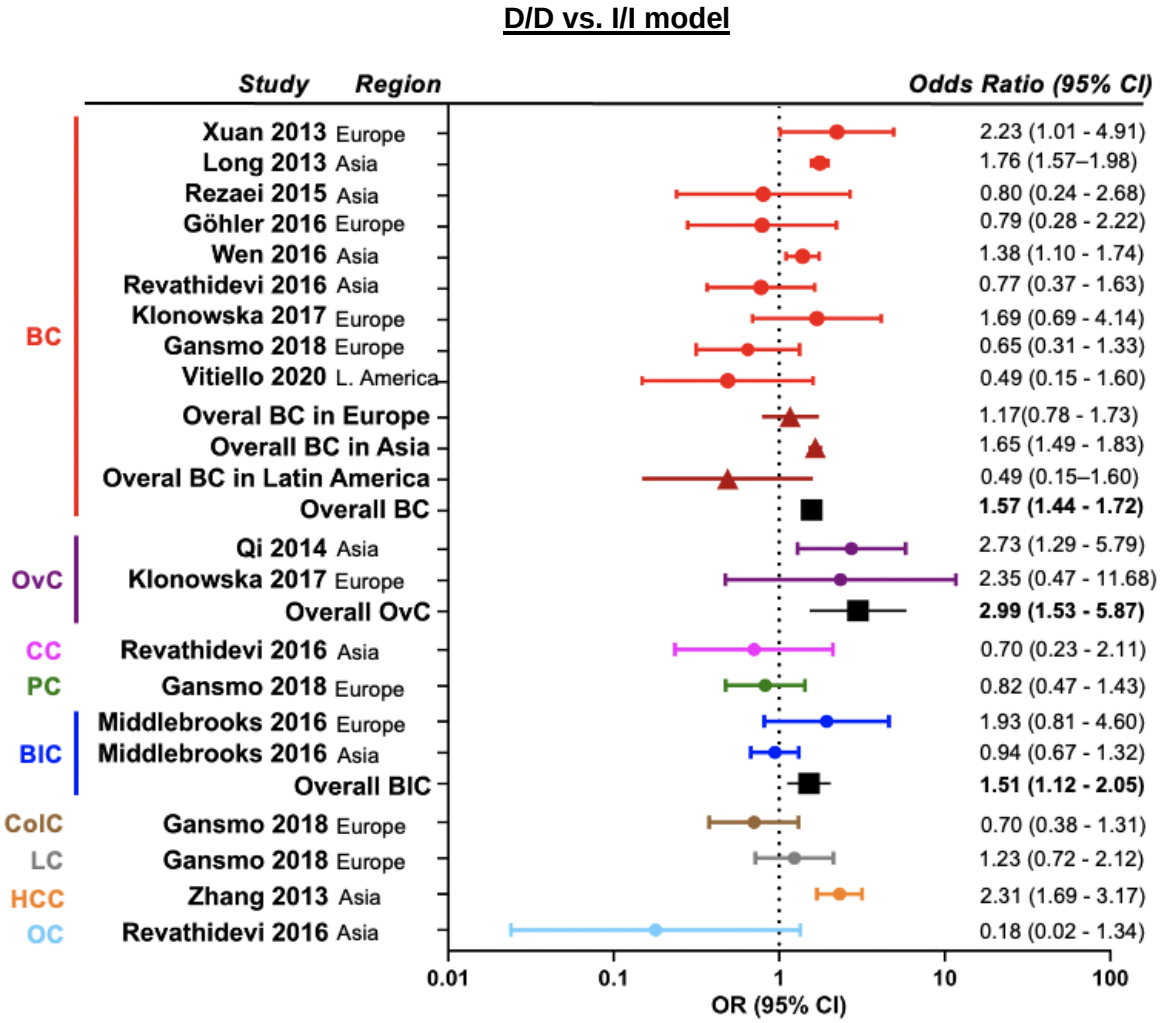


Figure 6. APOBEC3A_B deletion polymorphism and cancer risk for two-copy deletion genotype model.

Forest plot illustrating meta-analysis results of all studies on association of A3A_B deletion with various cancer types. Details of each of the study are indicated on the left. Odds ratio with corresponding 95% confidence intervals (CI), individually calculated for each study, are indicated on the right and visualised in the forest plot by different colours for each cancer type. Vertical dotted line (OR=1) represents no effect. Summary values are represented by black squares. The results are plotted on Log10 scale for visualisation purposes. BC – breast cancer; BlC – bladder cancer; CC – cervical cancer; ColC – colon cancer; HCC – hepatocellular carcinoma; LC – lung cancer; OC – oral cancer; OvC – ovarian cancer; PC – prostate cancer.

3.2 Genotyping single-cell clones and detection of A3A_B deletion

In order to generate an isogenic model of the A3A_B polymorphism, a previous MSc student (Rosalba Biondo) transfected NIKS with pX458 plasmids containing Cas9 and sequences encoding three different single guide (sg)RNAs designed to cut in both intron 4 of A3A and intron 7 of A3B (this is possible due to the extremely high sequence identity between these regions). Cells were transfected either with one of the sgRNA/Cas9 plasmids, or with a combination of all three. Rosalba demonstrated using a genotyping PCR assay that the bulk population of transfected cells contained cells in which the A3A_B deletion had been generated and generated a library of 129 single-cell clones (SCCs). My initial objective was to expand and screen this library to identify clones with either heterozygous or homozygous A3B deletion, using the previously-optimised PCR genotyping assay (Materials and Methods section 2.2.3).

SCCs were first revived and expanded in complete FC medium for 14-21 days before harvesting and further aliquoting for both DNA extraction and making freeze-down stocks for subsequent use (Materials and Methods section 2.1). Extracted genomic DNA was analysed with PCR genotyping assay twice, with insertion and deletion primers, designed by *Kidd et al*. (Materials and Methods section 2.2.3) and visualised on 1% agarose gel. Clones with successfully identified homozygous or heterozygous A3A_B deletions are shown in **Figure 7**. A clone with homozygous deletion is detected if a band of 700-bp is present, amplified by deletion primers. A clone with heterozygous deletion can be detected if a band is present in both PCR assays with deletion (700-bp) and insertion (705-bp) primers. Hence, a homozygous deletion was observed in clone number 36 in line 5 (**Figure 7A**) with a single ~700-bp amplified band by deletion primers. Two clones harbouring heterozygous

deletion (numbered 56 and 58 in lines 6 and 8) were detected by the presence of both ~700-bp bands with deletion and insertion-specific primers (**Figure 7B**). The validity of the assay is confirmed by 3 controls: negative or nuclease-free water control, eliminating the possibility of contamination; WT-NIKS DNA control, which is a negative control for A3B deletion and a positive control for A3B insertion; and lastly, SKBR3 DNA control, which is a positive control for A3B deletion and negative for insertion (as homozygous A3A_B deletion is naturally observed in the SKBR3 breast cancer cell line (Komatsu *et al.,* 2008)).

The results of the SCC library screen are indicated in **Table 2**. Out of all clones transfected with 3 different sgRNAs separately or with combination of all 3 sgRNAs (named accordingly sgRNA 74, sgRNA 75 and sgRNA 76), only those with sgRNA 75 showed presence of A3A_B deletion. Heterozygous deletion occurred in 20% of viable clones, while homozygous only in 10%. Only 51% of all screened clones were viable, hence more SCCs from sorted pools of transfected NIKS cells with gRNA 75 and gRNA 76 were generated for future screens and use (Materials and Methods section 2.2.1).

Table 2 Overview of screened single-cell clones.
Characteristics of all screened single-cell clones with APOBEC3A_B deletion (ΔA3B). -/- ΔA3B – homozygous deletion; -/+ ΔA3B – heterozygous deletion.

Clones	Overall number of clones	Survival rate, (distribution %)	-/- ΔA3B frequency, (distribution %)	-/+ ΔA3B frequency, (distribution %)
No sgRNA control	18	10 (55.6%)	-	-
sgRNA 74	33	26 (78.8%)	0	0
sgRNA 75	19	10 (52.6%)	1 (10%)	2 (20%)
sgRNA 76	31	10 (32.3%)	0	0
sgRNA 74+75+76	28	10 (35.7%)	0	0

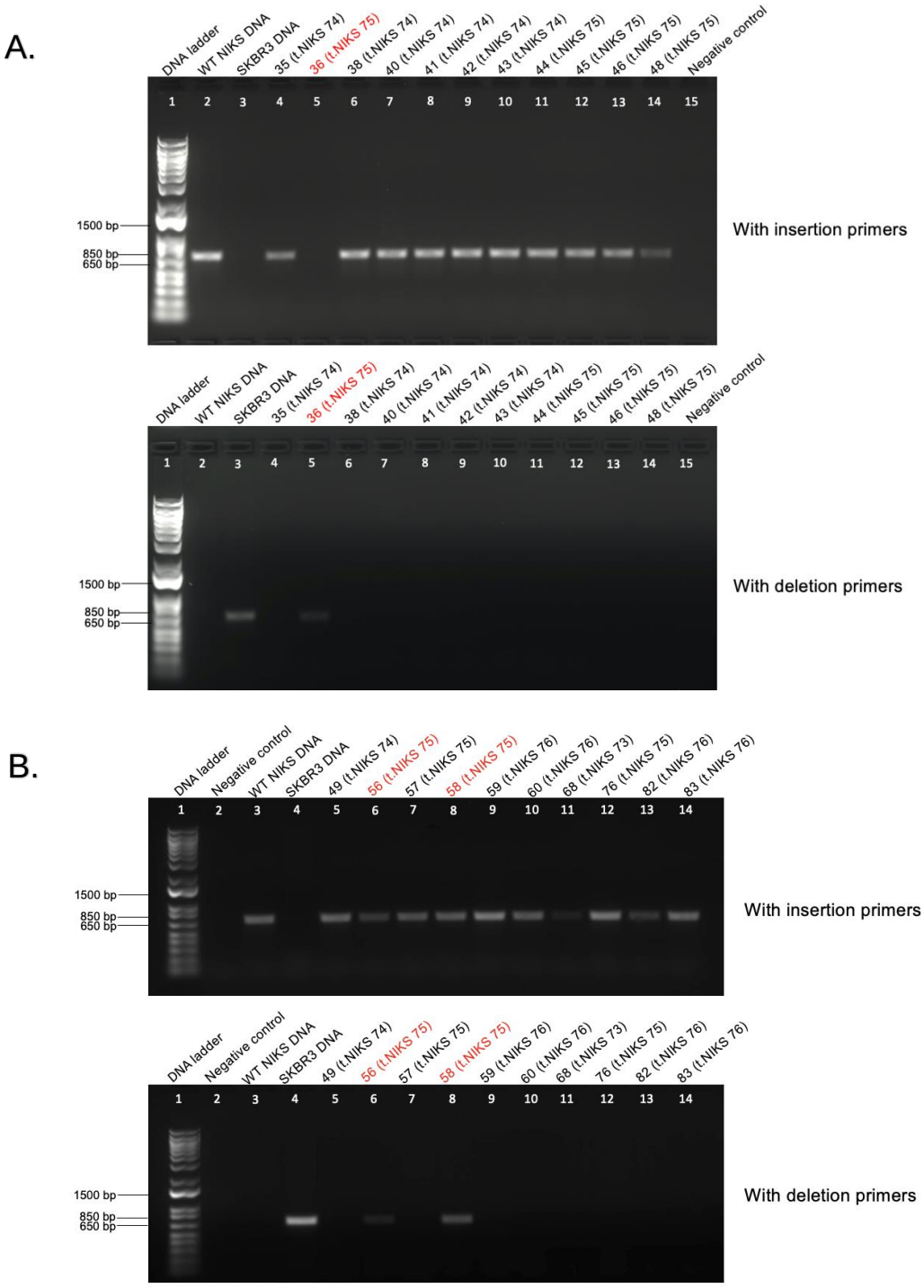


Figure 7. Successful clones with the APOBEC3A_B deletion.

PCR genotyping assay with detected clones containing homozygous (**A**) or heterozygous (**B**) deletion. Each assay was separately performed with insertion and deletion PCR primers, created by *Kidd et al., 2007.* Clone number is indicated above the lanes with corresponding sgRNA type used in brackets. Clones with the deletion are highlighted in red. Negative control corresponds to nuclease-free water.

3.3 A3B protein detection in HA-A3B NIKS cells

In order to understand the differences between cells expressing functional A3B protein and those with the deletion of entire A3B coding region, as well as to evaluate what functions are homozygous A3A_B carriers depleted of, it is firstly important to understand expression patterns and, consequently, functions of A3B in a normal cell. This part of the work was a continuation of previously made observations in the laboratory, that in a synchronised NIKS population at G_1/S boundary, A3B is mostly upregulated 8h post-release from a single thymidine block (corresponding to G_2/M stage), hence, it could be a cell cycle-dependent protein (**Figure 3**).

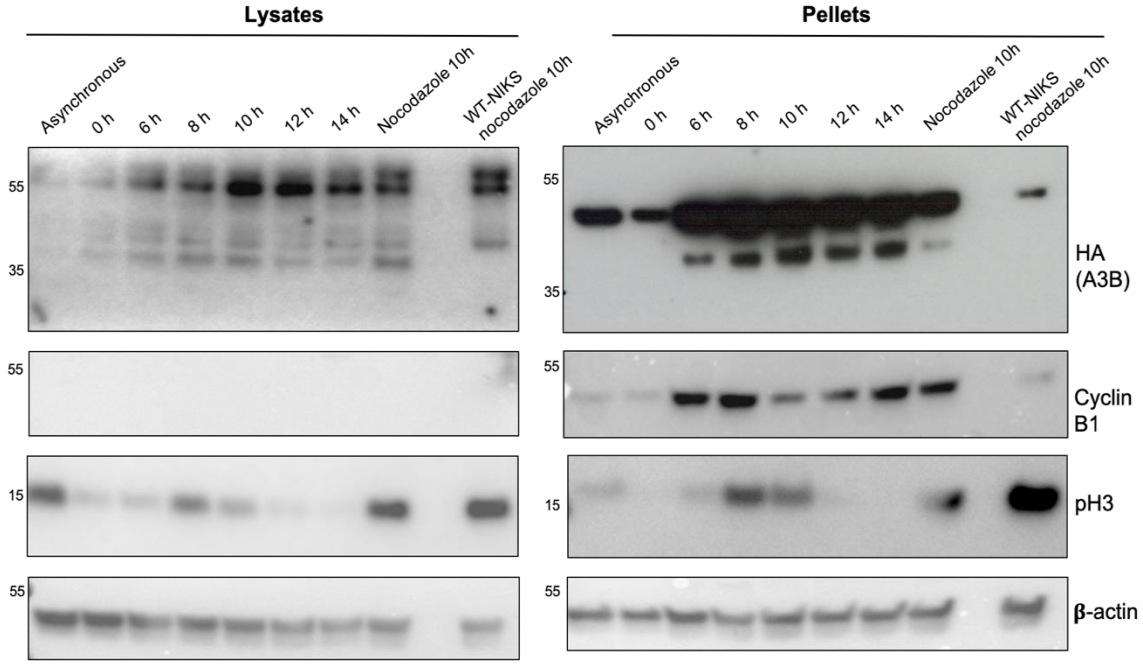


Figure 8. Expression of APOBEC3B in HA-A3B NIKS is the highest during G_2/M cell cycle phase.

Western blotting demonstrating A3B protein expression in asynchronous, thymidine-treated and nocodazole treated NIKS populations. Protein levels in 1) cell lysate were detected by lysing cells, while 2) pellets were solubilised in 4x Laemmli sample buffer and boiled for 10 minutes. A3B levels were detected by separating 20 µg of whole cell lysate by SDS-PAGE and probing with anti-HA (1:1000), anti-Cyclin B1 (1:500) and anti-pH3 (1:1000) antibodies, after which, visualised with ECL reagent. β-actin was used as a loading control.
For more details on cell synchronisation and SDS-PAGE methods, refer to Materials and Methods sections 2.3.1 and 2.3.2, respectively.

In order to confirm detection of A3B protein in G_2/M cell cycle phase, cell-synchronisation experiment was designed to observe protein expression pattern in both soluble (lysates) and insoluble (pellets) fraction of the cell (**Figure 8**). NIKS cells, in which CRISPR/Cas9 was used to tag with HA-epitope the endogenous A3B gene (further referred to as HA-A3B) were synchronised in G_1/S boundary by single-thymidine block for 18h, followed by 0-14h release. Concurrently, HA-A3B were arrested in G_2/M boundary by nocodazole (NZ, a beta-tubulin binding agent which disrupts microtubule polymerisation and therefore blocks cells in mitosis) block for 10h (Materials and Methods section 2.3.1) and were used as a positive control. Nocodazole-treated synchronised WT-NIKS population was used as negative control for anti-HA antibody specificity. Phospho-histone H3 (pH3) and Cyclin B1 were used as mitotic and G_2/M markers, respectively.

Strong non-specific bands were observed in both Western blots for lysates and pellets, however, the absence of lower bands in negative controls at ~46 kDa suggested detection of A3B. Thus, the previous findings were confirmed as the strongest induction of A3B protein was seen at 6-10h post-release and in nocodazole treated HA-A3B in lysates; and in 8-14h post-release in pellets. In both blots (whole cell lysate and cell pellets) A3B expression could be detected from 6h up to 14h post-release. The same pattern of pH3 detection was observed in both lysates and pellets, indicating the period of 8-10h post-release from single-thymidine block to be the stage of mitosis. Although levels of Cyclin B1 in soluble fraction were undetectable under different exposure times, in addition to stronger bands observed for A3B in insoluble fraction of the cells, expression pattern of the protein coincides with the expected one and suggests an issue with solubilisation.

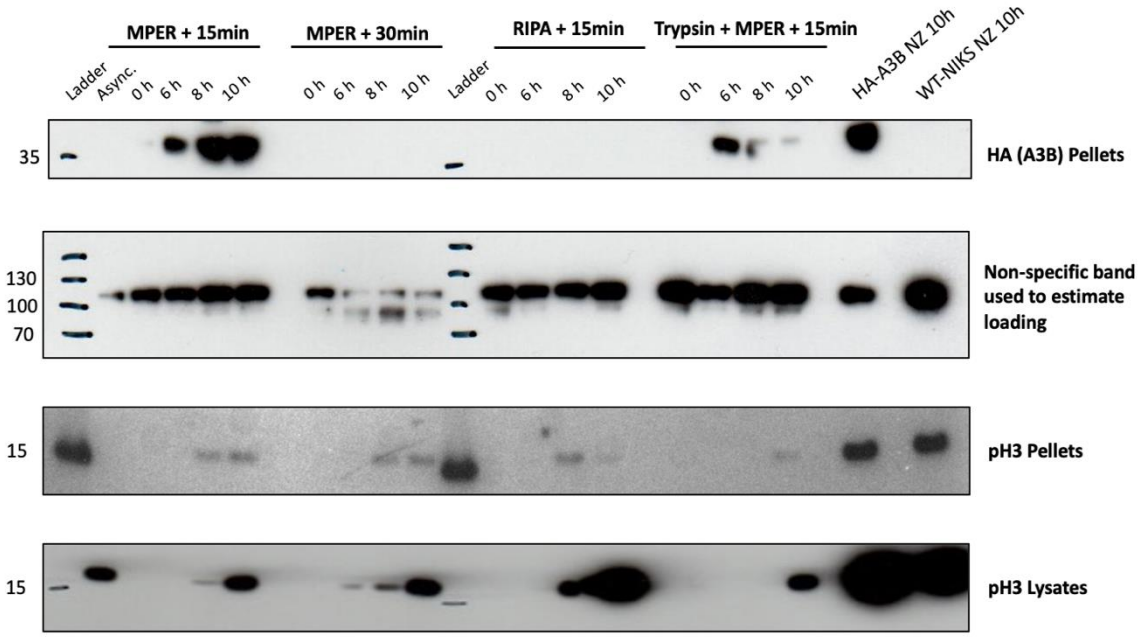


Figure 9. Optimisation of APOBEC3B solubilisation.

Western blotting demonstrating a panel of A3B protein expression in asynchronous, thymidine-treated and nocodazole treated NIKS populations using different protein extraction reagents (MPER, RIPA) with varying lysis times on ice. Protein levels in 1) cell lysate were detected by lysing cells, while 2) pellets were solubilised in 4x Laemmli sample buffer and boiled for 10 minutes. For A3B detection, 20 µg of whole cell lysate was separated by SDS-PAGE and probed with anti-HA (1:1000) and anti-pH3 (1:1000) antibodies, after which, visualised with ECL reagent. Note, the blot is lacking β-actin as a loading control, but a non-specific band observed in cell lysates included to show consistent loading.
For more details on cell synchronisation and SDS-PAGE methods, refer to Materials and Methods sections 2.3.1 and 2.3.2, respectively.

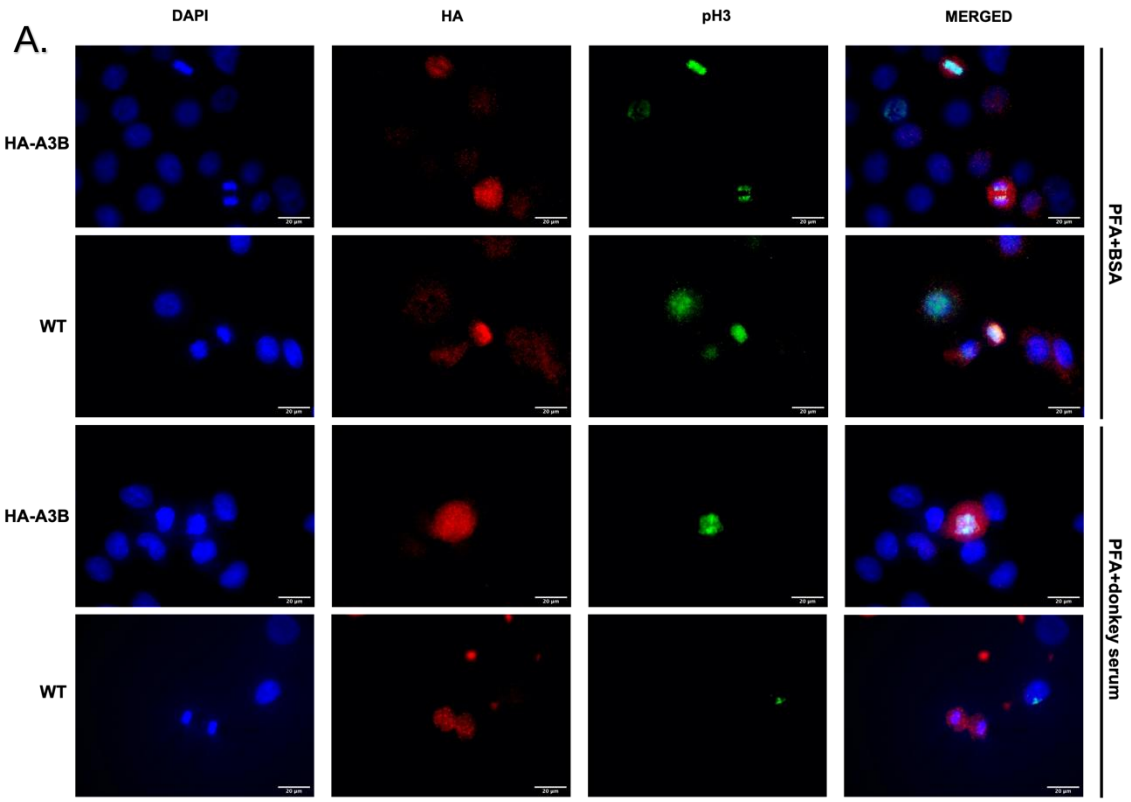
To further address the solubilisation issue, an optimisation panel was designed to identify the conditions in which A3B expression could be seen predominantly in the lysates, opposed to the cell pellets. Four different conditions were chosen for cell lysis to extract A3B, using either Mammalian Protein Extraction Reagent (MPER) or radioimmunoprecipitation assay (RIPA) buffer with varying time of 0.2U/µL benzonase nuclease incubation on ice (Materials and Methods section 2.3.2). One of the conditions included trypsinisation of cells prior to treatment with MPER lysis buffer, to dissociate adherent cell links. The results shown in **Figure 9** indicated, that A3B solubilisation panel was not successful, as A3B was detected in insoluble

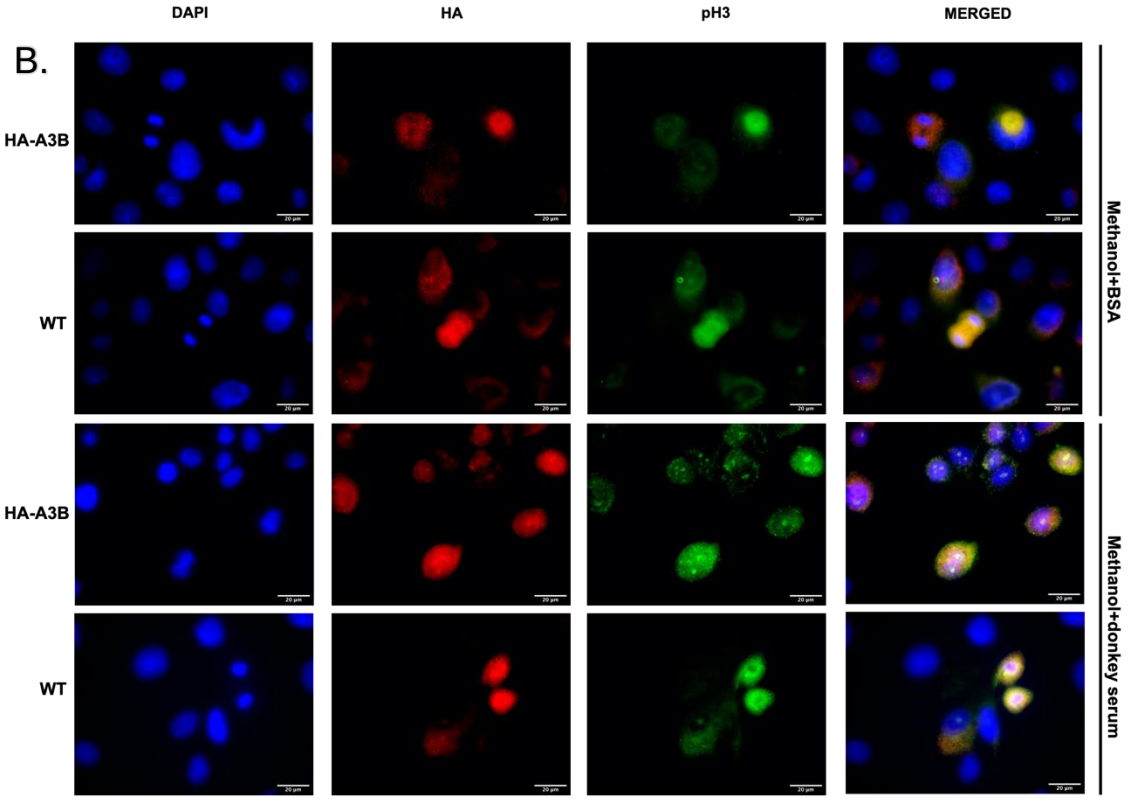
fraction at expected time points, however A3B-specific size bands were not observed in the soluble fraction (lysates). Instead, only strong non-specific bands above 70 kDa were seen in the whole cell lysate. A3B expression observed in cell pellet fraction corroborated that the protein is abundant during mitosis stage of the cell cycle (8-10h post-release from thymidine block or in nocodazole treated HA-A3B NIKS). The expression pattern also corresponded to phosphorylation on serine-10 of pH3 marker, which was consistent on both blots in lysates and pellets. Two of the conditions (lysis of cells with MPER, followed 30 min incubation on ice with benzonase and lysis with RIPA, followed by standard 15 min incubation interval) did not show any bands for A3B even in insoluble fraction of the cells. This was not surprising due to once again observable A3B solubilisation issue, the reasons for this will be discussed in the later section. Unfortunately, due to the shutdown of the laboratory caused by COVID-19 in March, the experiment could not be repeated. In addition, there was no possibility to re-probe this Western blot for β-actin control. However, as non-specific high-molecular weight bands look fairly even across all time points in “Non-specific band used to estimate loading” section of the blot, a deduction of consistent protein loading on the gel could be drawn.

3.4 Visualisation of A3B in G_2/M cell cycle phase by Immunofluorescence Microscopy

To further investigate A3B activity in NIKS cells expressing functional protein, visualisation of A3B localisation by fluorescence microscopy was performed. Based on previously obtained data from Western blot analyses (**Figure 3, Figure 8, Figure 9**), it was decided to visualise only a single condition of HA-A3B NIKS: 8h post-release from single-thymidine block, when cells are in G_2/M (entering or undergoing mitosis) phase and the expression of the protein is the highest. WT-NIKS were also synchronised with a single-thymidine block and released for 8h to represent negative control for HA-specific staining (since A3B is untagged in these cells, Materials and Methods section 2.4.1). Additionally, the experiment aimed to visualise co-localisation of A3B and pH3 mitotic marker, to mark and examine both cells in mitosis phase and mitotic chromosome condensation.

Firstly, it was decided to set up an optimisation panel, using 6 different fixation methods with methanol, 4% paraformaldehyde (PFA) or cytoskeleton (CSK) buffer, in order to minimise anti-HA background staining in HA-A3B NIKS and non-specific signal in WT-NIKS populations (Materials and Methods section 2.4.3). The results are shown in **Figure 10**, where co-staining of HA and pH3 was successful in all conditions, however 4% PFA fixation with BSA or donkey serum (same species as the host for secondary antibody) blocking buffers (**Figure 10A**) and methanol fixation with the same blocking buffers (**Figure 10B**) showed HA staining in negative controls (WT-NIKS), thus, were further disregarded. Fixation with 4% PFA and further subcellular fractionation method using CSK buffer with either one (0.1%) or two


A.
DAPI
HA
pH3
MERGED
HA-A3B
WT
HA-A3B
WT
PFA+BSA
PFA+donkey serum


B.
DAPI
HA
pH3
MERGED
HA-A3B
WT
HA-A3B
WT
Methanol+BSA
Methanol+donkey serum

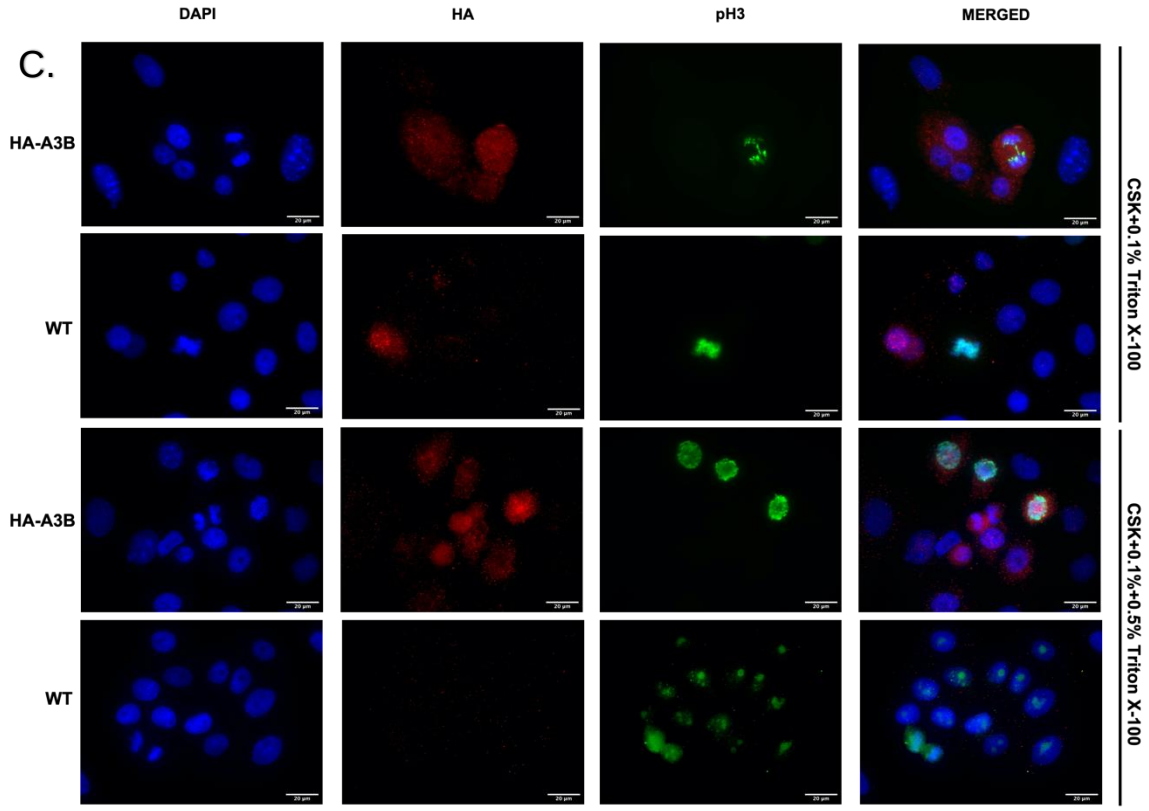


Figure 10. Sample fixing optimisation panel for visualisation APOBEC3B by immunofluorescence microscopy.

Optimisation of A3B staining in G_2/M phase (8 hours post-release from single thymidine block) with various sample fixing methods: **A.** 4% PFA with 5% Donkey Serum or 2% BSA; **B.** Methanol with 5% Donkey Serum or 2% BSA; **C**. 4% PFA with CSK buffer+0.1% (v/v) Triton X-100 or with CSK buffer+0.1%+0.5% (v/v) Triton X-100 (two incubation steps of Triton X-100). The samples were co-stained for anti-HA (1:500, rabbit monoclonal), anti-pH3 (1:400, mouse monoclonal) and nuclei of the cells were counterstained with DAPI. AlexaFluor594 (red) was used as a secondary antibody for anti-HA (1:1000, anti-rabbit), AlexaFluor488 (green) was used for anti-pH3 (1:1000, anti-mouse). The images were taken at x60,000 magnification. Background staining was normalised to synchronised WT-NIKS population in G_2/M phase (negative control). Scale bars of 20μm are shown on the images. For more details on sample preparation and fixation differences, refer to Materials and Methods section 2.4.3.

(0.1%, followed by 0.5%) Triton X-100 incubations showed little to no non-specific background staining in negative controls, respectively (**Figure 10C**). As a result, 4% PFA fixation method with CSK buffer and two 0.1%+0.5% Triton X-100 incubations was seen to be specific for immunostaining HA-A3B with complete elimination of non-specific background and was carried further.

After establishing the right conditions for A3B visualisation by immunofluorescence, a biological repeat was conducted to confirm consistent results and staining specificity for HA-A3B, but not WT-NIKS (**Figure 11**). Additionally, with the confidence that observed signal of the staining is true (**Figure 11, row 5**), co-localisation of A3B and pH3 was further inspected.

It was not surprising, that A3B was co-stained with pH3, hence, detectable in cells undergoing mitosis, however localisation of A3B changed, depending on the stage of mitosis the cell was undergoing. In the early stages of mitosis, such as prophase (**Figure 11, row 3**) and prometaphase (**Figure 11, row 2**), A3B seems to localise around condensed chromatin. In early anaphase (**Figure 11, rows 1; 4**) and late anaphase (**Figure 11, row 3**) A3B is seen around sister chromatids and anaphase bridges and is concentrated more towards the equatorial region of the cell. However, in telophase (**Figure 11, rows 1; 2**) the localisation shifts presumably to the opposite poles of mitotic spindles. Additionally, A3B staining was observed in cells without phosphorylated pH3 (**Figure 11, rows 1; 2; Figure 10C**), consistent with our western blotting data showing A3B expression over a longer time period than that seen for H3 phosphorylation. A further interesting phenotype was observed in WT-NIKS, as the nucleoli were co-stained with DAPI and anti-pH3 (**Figure 11, row 5**).

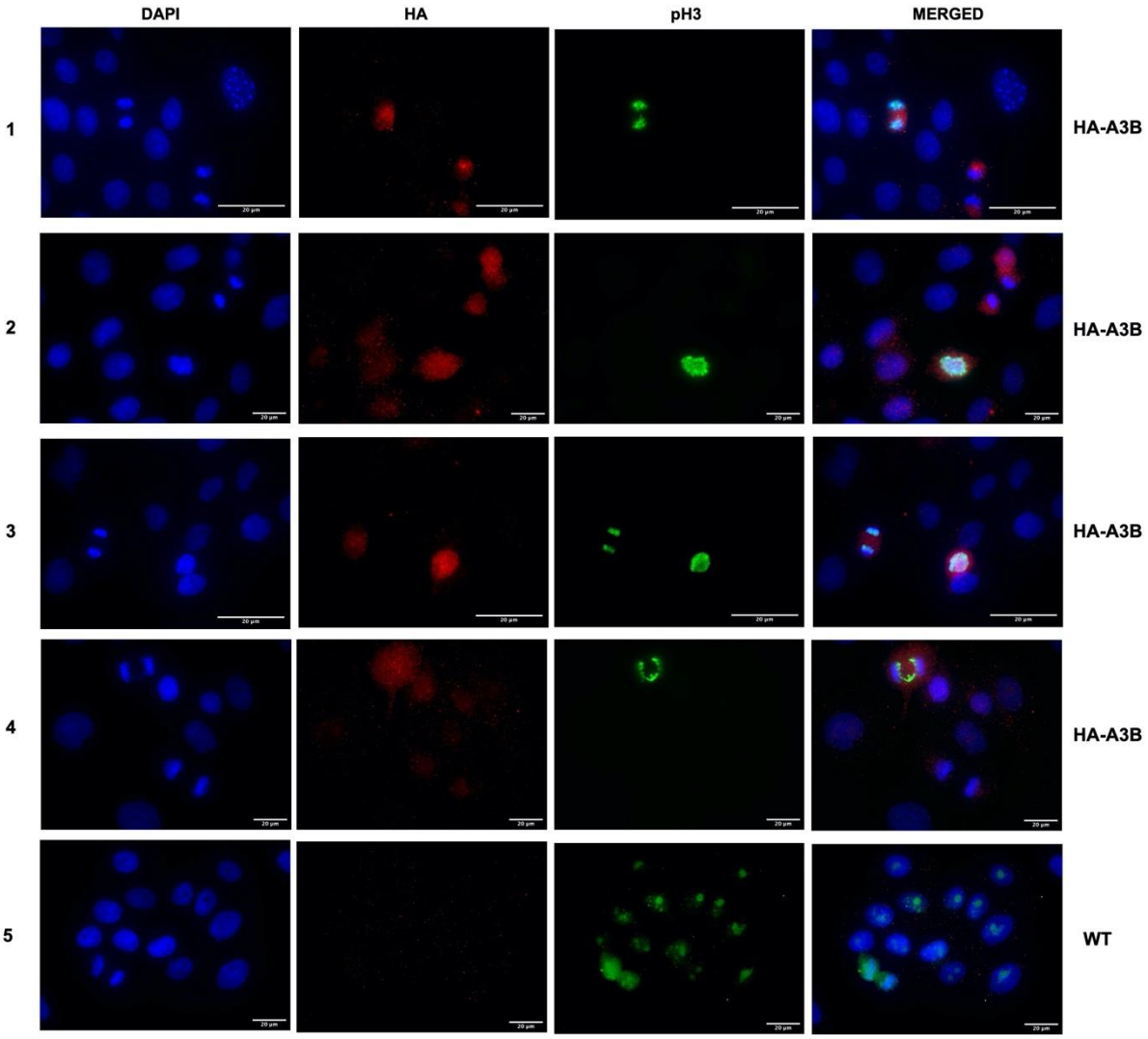


Figure 11. APOBEC3B is detected around condensed chromatin and chromosomes in prophase, anaphase and telophase stages of mitosis.

Immunofluorescence images of synchronised HA-A3B NIKS in G_2/M phase (8 hours post-release from single thymidine block), fixed with 4% PFA with CSK buffer+0.1%+0.5% Triton X-100 (double incubation). The samples were co-stained for anti-HA (1:500, rabbit monoclonal), anti-pH3 (1:400, mouse monoclonal) and nuclei of the cells were counterstained with DAPI. AlexaFluor594 (red) was used as a secondary antibody for anti-HA (1:1000, anti-rabbit), AlexaFluor488 (green) was used for anti-pH3 (1:1000, anti-mouse). The images were taken at x60,000 magnification. Background staining was normalised to synchronised WT-NIKS population in G_2/M phase (negative control). Scale bars of 20μm are shown on the images. For more details on sample preparation, refer to Materials and Methods sections 2.4.

4 Discussion

4.1 Meta-analysis of association studies on APOBEC3B deletion and cancer susceptibility

Over the recent years APOBEC3 deaminases have been in the spotlight of research due to their broad roles in innate immunity, as potent restrictors of transposable retroelements, retroviruses (HTLV-1, HIV-1, HIV-2, HBV) and DNA viruses (HPV, AAV) (Chen *et al*., 2006; Vieira, 2013; Harris and Dudley, 2015; Knisbacher *et al*., 2016). However, with the emerging large-scale studies and incorporation of high-throughput and fast next-generation sequencing an additional outcome of A3 activity has been observed, which is not so protective for the host. On contrary, overexpression and deregulation of A3 members, A3A and A3B, have been associated with characteristic mutational signatures (C>T transitions and C>G transversions at TpCpN trinucleotide motifs), contributing to cancer mutagenesis (Alexandrov *et al*., 2013; Roberts *et al*., 2013; Burns *et al*., 2013a). Furthermore, with the numerous studies debating which A3 deaminase is critical in carcinogenesis, linking overexpression of both A3A and A3B with tumourigenesis, cancer evolution, progression and worse clinical outcomes, the importance of A3B has been questioned since discovery of a commonly seen A3B deletion polymorphism, stratified across the global population (Burns *et al*., 2013b; Nik-Zainal *et al*., 2014; Kidd *et al*., 2007; Chen *et al*., 2017). Notably, A3A_B hybrid gene presents almost 2-fold increased breast cancer risk and, surprisingly, high levels of APOBEC signature mutations in sequenced tumours, suggesting both – A3B being a non-essential gene and a definite involvement of other mechanisms for cancer development (Long *et al*., 2013; Caval *et al*., 2014; Nik-Zainal *et al*., 2014). Absence of one of the main enzymes, causing *kataegis* in the genome, yet presence of distinguishable APOBEC signatures, could be explained by activity of other A3

members of the family. So far, apart from the several studies pointing observed APOBEC somatic mutagenesis to the elevated activity of A3A, A3H haplotype I (A3H-I) has been also shown to localise in the nucleus and mutagenise genomic DNA in lung and breast cancer within tumours harbouring A3A_B deletion (Starrett *et al*., 2016). Implication of A3H was also supported by a case-control study, where a missense SNP of A3H gene was associated with decreased risk of lung cancer, suggesting A3H contribution to genomic deregulation (Zhu *et al*., 2015). Additionally, recruitment of A3G should not be excluded from considerations due to its DSB repair role in the nucleus and experimentally shown enhanced radiotherapy resistance association in lymphomas (Nowarski *et al*., 2012). Furthermore, immune response may be involved in breast cancer progression in A3A_B carriers, as a strong correlation of increased expression of immune response-related genes (arising from tumour-infiltrating immune cells) and A3B deletion in hypermutated tumours was observed; conversely causing increase, rather than decrease of the mutational load (Cescon *et al*., 2015; Wen *et al*., 2016).

Inconclusive and sometimes conflicting data of association of common A3A_B deletion polymorphism and cancer risk, led us to conduct a meta-analysis of all published case-control studies on A3B deletion so far. In this computational part of the project, we set to derive a more uniform conclusion to the main question – whether deletion genotype increases or, conversely, does not exhibit an effect on developing certain cancer in each of the analysed regions.

Herein, the data from eligible large-scale case-control studies was analysed separately for each cancer type irrespective of region, as well as in each region where possible, under two main inheritance models – heterozygous (**Figure 5**) and

homozygous (**Figure 6**) codominant. The results for breast cancer were consistent with the previous extensive meta-analyses by Han *et al*. and Klonowska *et al*., with statistically significant and strong association observed for homozygous [OR(95%CI)=1.57(1.44–1.72), p<0.0001] and heterozygous [OR(95%CI)=1.21(1.15–1.27), p<0.0001] A3B deletion genotype and cancer risk (Han *et al*., 2016; Klonowska *et al*., 2017). However, it is worth noting, that study by Marouf, looking at association in North African (Moroccan) population was excluded from the homozygous codominant genetic model (**Figure 6**) due to the absence of individuals with the 2-copy deletion of A3B gene (in both control and cases groups) (Marouf et al., 2016). This is not surprising as firstly, the population size of this study is quite small, with only 426 individuals assessed, and secondly, the frequency of the deletion allele in the study was very low (**Table 1**), which was expected since A3B deletion is observed in <1% of African population (Kidd *et al.,* 2007). Interestingly, low allelic frequency, similar to European region, was seen in the newest case-control association study in Brazilian population, on contrary to previously observed ~58% for Amerindian region (Kidd *et al.,* 2007; Vitiello *et al*., 2020). Although patient data is confidential and no information on their deep-ethnic background and what parts of the country the individuals came from, all tumour tissue and blood samples were collected in Londrina Cancer Hospital, which is situated in the Southern part of Brazil. Southeast and South of Brazil population genomes display >70% European ancestry, which would allow us to deduce that individuals in this study also have European ancestry, hence low frequency of deletion genotypes (Kehdy *et al*., 2015).

As anticipated, due to low population size and A3B deletion allele frequency, the study did not show any statistically significant association between individuals

harbouring A3A_B deletion and increased breast cancer risk. Generally, results obtained for studies looking at European cohorts were in agreement with previous meta-analyses mentioned, as the observed effect of A3A_B deletion variant was low and statistically insignificant for heterozygous ([OR(95%CI)=1.06(0.97–1.16), p=0.164] for homozygous and [OR(95%CI)=1.17(0.78–1.73), p=0.449] A3A_B deletion) (Göhler *et al*., 2016; Klonowska *et al*., 2017; Gansmo *et al*., 2018). The exception was the study by Xuan *et al*., so far being the only association study of European-ancestry women, revealing a highly significant association of breast cancer risk with the deletion variant (Xuan *et al*., 2013). Moreover, the association of risk was higher in individuals with homozygous A3A_B deletion, than in those with 1 A3B allele (**Table 1**). The reasons for discrepancies in European association studies are unclear, however, a slightly higher allelic frequency in both cases and controls representing individuals in the study conducted by Xuan *et al.* are seen (**Table 1**). It is doubtful whether that would contribute to higher effect of the deletion genotype, however, additional observation while considering this paradox, is that the study was conducted in Nashville, USA, and with the ~58% of Amerindian region presenting A3B gene deletion, hence, it would be useful to perform whole-genome sequencing or genotype SNPs, to confirm that all subjects have European-ancestry (Kidd *et al.,* 2007). A consistent with strong statistical significance breast cancer risk association has been observed in Asian populations, with ([OR(95%CI)=1.31(1.22–1.40), p<0.0001] for heterozygous and [OR(95%CI)=1.65(1.49–1.83), p<0.0001] for homozygous A3A_B deletion (**Figure 5**, **Figure 6**) (Long *et al.*, 2013; Rezaei *et al*., 2015; Wen *et al*., 2016; Revathidevi *et al*., 2016).

Results for ovarian cancer were analogous to those observed in breast cancer. To elaborate, association for the increase of ovarian cancer risk significantly correlated

with the loss of 1 and 2-copy A3B alleles in the general population ([OR(95%CI)=1.42(1.21–1.65), p<0.0001] for heterozygous and [OR(95%CI)=2.99(1.53–5.87), p=0.0014] for homozygous A3A_B deletion). However, association varied for different ethnic groups: strong association was observed in Asian population in both homozygous and heterozygous deletion carriers compared to subjects with no deletion, but European cohort showed lack of significant association (**Table 1**) (Qi *et al*., 2014; Klonowska *et al*., 2017). The study by Qi *et al*. in Asian region had a larger screened population size than European by almost 2-times, hence would explain shorter CIs seen in **Figure 5** and **Figure 6**. It is also worth noting, that study by Klonowska *et al*. had a relatively small sample size for cases, compared to controls. Furthermore, deletion alleles of A3B are more frequent in Asian cohort than seen in European population (**Table 1**), showing highly heterogenous results, observed in this analysis. Although it is unclear why similar association pattern for both breast and ovarian cancers is observed, it could be due to similar hormonal or/and genetic background, or to other overlapping risk factors usually seen as a driving tumourigenesis (Holschneider and Berek, 2000).

Obtained results for bladder cancer looked particularly interesting, with decreased cancer risk and A3A_B deletion association, also known as protective association, observed in heterozygous codominant inheritance model (**Figure 5**) and increased bladder cancer risk association in homozygous inheritance model (**Figure 6**) for the Asian and European populations combined. However, the study by Middlebrooks *et al*. have shown that SNP rs1014971 (located 20 kb upstream of A3A) but not A3B deletion, is driving the association with cancer risk, as adjustment of OR and CIs without this variant did not reveal any significant association (Middlebrooks *et al*., 2016). A weaker association of SNP rs1014971 was also seen in ER+ breast

cancers (Middlebrooks *et al*., 2016). In this regard, such example leads to consideration that effect of SNPs could influence the association results of A3A_B deletion with cancer risk in other cancer types, hence more studies are needed to investigate correlation between known high risk SNP variants and A3A_B deletion effect.

The remaining association studies, included in this meta-analysis, had been carried out only in one region for each cancer investigated, hence overall association for cancer risk, irrespective of region, could not be carried out. The study by Zhang *et al.* looking at HBV-related HCC association with genomic A3A_B deletion in Chinese cohort, has shown a significantly elevated cancer risk depending on individual's genotype ([OR(95%CI)=1.78(1.47–2.16), p<0.0001] for heterozygous and [OR(95%CI)=2.31(1.69–3.17), p<0.0001] for homozygous A3A_B deletion) (Zhang *et al*., 2013). These results support a conclusion that absence of A3B protein, normally expressed in primary human hepatocytes and inhibiting HBV replication upon induction of IFN-α, impairs aforementioned functions, attenuating clearance of HBV and consequently leading to liver carcinogenesis (Bonvin *et al*., 2006; Lucifora *et al*., 2014). A study by Revathidevi *et al.* looked at association of A3A_B deletion with oral and cervical cancer risk in Indian cohort, however the results did not show significant evidence for association (**Supplementary Table 3**), which could be due to the low sample size of cases (**Table 1**) (Revathidevi *et al*., 2016). Associations for prostate, lung and colon cancers with deletion polymorphism in European population also did not reach statistical significance in both inheritance models, despite the large population size investigated (Gansmo *et al*., 2018).

There are several limitations to this meta-analysis that need to be addressed. First of all, this was a preliminary meta-analysis of all eligible studies to summarise impact of A3A_B deletion on risk of various types of cancer, as comprehensive statistical analysis was not carried out due to time and resources constraints. Investigation of heterogeneity among the studies for each genetic inheritance model or publication bias was not carried out. Additionally, predominant unavailability of patient data, which are covariates of the study (e.g. smoking status, number of individuals over the age of 60-65), left us to work with the only accessible information, such as cancer type, ethnicity, population size and APOBEC genotype. In this regard, it seemed clear that ethnicity is predominant factor for positive association, with Asian populations showing stronger and more significant correlation than European-descent populations.

4.2 Screening the SCC library for homozygous and heterozygous A3A_B deletion identification

The second part of the work aimed to PCR genotype a collection of previously generated SCCs, representing an isogenic model for A3A_B deletion in NIKS human epithelial cell line, and isolate clones, containing 1-copy or 2-copy deletion of the gene. The NIKS cell line is a spontaneously immortalised human keratinocyte cell line, representing a versatile study model for healthy stratified squamous epithelia, from which many squamous cell carcinomas develop (such as lung, skin, bladder, breast, head and neck) (Allen-Hoffmann *et al*., 2000). Additionally, NIKS supports a life cycle of human papillomaviruses (e.g. HPV-16 and HPV-18), which are strongly associated with cervical cancer development, hence could also represent a model healthy tissue for HPV-dependent cancers (Genther *et al.*, 2003). High APOBEC3 expression in healthy keratinocytes, as well as extensive genome

editing observed in HPV-infected cells make NIKS a decent candidate for isogenic model with ~29.5kb deletion, representing individuals with identical germline A3B deletion (Warren *et al*., 2014).

Genotyping results of previously generated 129 single-cell clones revealed only 3 clones, containing the desired deletion: 1 clone with homozygous A3A_B deletion, also known as 1-copy deletion (**Figure 7A**) and 2 clones with heterozygous or 2-copy deletion (**Figure 7B**). All three successful clones originally came from the same pool of FACS-sorted cells, which were previously transfected with the same sgRNA (in this project named sgRNA75). This was not surprising due to several reasons. Firstly, overall survival of the clones was rather low, with sgRNA76 transfected clones surviving the least (**Table 2**). Declining survivability numbers next to each group of clones correlate with the failure of our liquid nitrogen storage over the Christmas period, as only a fraction of clones was viable after culturing. Thus, it was impossible to calculate the distribution frequency (%) of homozygous and heterozygous A3B deletion, as possibly more successful clones were lost due to the accident. However, exclusively from the data obtained, out of 10 viable clones from the pool of 19, 10% of clones had homozygous and 20% had heterozygous genotypes (**Table 2**). Another possible reason for low deletion frequency within the sorted populations could be low transfection efficiency and non-stringent gating setting on FACS for sorting transfected cells, which would in turn cause majority of cells within the population to have uncut A3B gene, as was previously proposed by Rosalba Biondo.

Interestingly, slightly altered phenotype of the successful clones was observed by the bright-field microscopy. Despite the differences in genotypes, it took 19 days for

all three clones to revive (with comparison to ~14 days for clones without the deletion), start proliferating and form colonies. However, in comparison with control WT-NIKS and clones where deletion has not occurred, the cells were forming colonies slower and intriguingly, majority of observed keratinocytes were differentiating faster (differentiation is normally observed if NIKS reach ~70-80% confluency in the culture). This coincides with the recent study in mouse keratinocytes, showing a novel regulatory role of A3 enzyme (note: mice express only one A3 deaminase) in keratinocyte differentiation – A3 negatively regulates differentiation by inhibiting Notch3 (Dainichi *et al.*, 2019). The differentiation was not favourable in this study, as it deregulates cell cycle and triggers mitosis checkpoints, thus would negate investigation of all A3 expression levels throughout the cycle (Freije *et al*., 2012; Gandarillas, 2012).

Unfortunately, as a result of the liquid nitrogen storage accident mentioned above, a single clone with homozygous deletion was lost and could not be revived. That being said, both heterozygous clones were expanded and frozen down for future investigation; due to the shutdown of the laboratories, caused by COVID-19, characterisation work could not be carried forward.

4.3 A3B protein detection in the cell cycle

The last part of the project focused on investigation of A3B protein expression and functions at the cellular level, which still remain largely unknown despite strong A3B overexpression and characteristic mutational signature observed in cancer samples and genomes. Thus, in combination with the previous observations (*Nicola Smith*, unpublished) that A3B is only detected in G_2/M phase of the cell cycle without

a significant change of mRNA levels (**Figure 3**), it was decided to look into expression pattern and stability of the protein.

WT NIKS, HA-NIKS (NIKS cell line, in which the endogenous A3B gene has been tagged with haemagglutinin (HA)-derived epitope) and anti-HA antibody were used for protein detection and expression experiments, to attenuate a common problem stumbled upon in the studies of A3 deaminases – currently, no antibodies with high-affinity and specificity are available for A3B, due to high structural homology between A3B and A3A (and to lesser extent, structural similarity is shared with other A3 family members) (Salter *et al*., 2016). Intriguingly, the data obtained was conflicting: even though expression of A3B was the strongest in G_2/M phase (8-10h post-release from single thymidine block), as expected and shown in previous work, the protein was strongly associated with insoluble fraction of the cell, which was contradictory to previous findings (**Figure 8, Figure 9**). The findings suggested an issue with protein solubilisation, as not only A3B, but also cyclin B1 were observed in the insoluble fraction (which should not be the case), however even the optimisation panel with various cell lysis buffers and prolonged incubation time did not yield expected results (**Figure 9**). This was surprising and never observed before in the laboratory, pointing to a possible issue with Benzonase nuclease, which is commonly used for DNA digestion and consequently easier extraction of closely associated nuclear proteins, such as A3B. Additionally, the results were exacerbated by possible issues with transfer, as seen from complete lack of signal in **Figure 8** for Cyclin B1 and in **Figure 9** for A3B in the insoluble fraction. However, it is also possible that lack of signal could be attributable to part of the PVDF membrane drying.

Nevertheless, expression of A3B, was observed to last longer than previously seen and the induction pattern coincides with Cyclin B1 expression. Cyclin B1 is a key G_2/M cell-cycle phase regulator and is expressed predominantly in the end of G_2 phase up until anaphase in mitosis, after which it is rapidly degraded (Morgan, 2007). Although biological repeats are needed to confirm A3B expression pattern, in addition to further optimisation of protein solubilisation, these are preliminary additional findings for current understanding of A3B biology.

4.4 Visualisation of A3B expression localisation in G_2/M cell cycle phase

Further exploration of A3B expression pattern and localisation was carried out by double labeling WT and HA-A3B NIKS, arrested in G_2/M cell cycle phase, for visualisation by immunofluorescence microscopy. The results obtained in this (**Figure 8**), as well as during previous project in the laboratory (**Figure 3**), indicated that induction of A3B is seen specifically during G_2/M phase. Additionally, A3B is known to be the only A3 deaminase, which is constitutively nuclear (Pak *et al.,* 2011; Lackey *et al.*, 2013). Hence, in addition to staining with anti-HA to visualise the localisation of A3B in HA-A3B NIKS, the cells were additionally co-stained with DAPI, to determine the substage of mitosis by observable morphological changes of chromatin, and with anti-pH3 to confirm G_2/M phase by phosphorylation of Serine-10 on histone H3.

Immunofluorescence was firstly performed on optimisation panel with different fixation methods, commonly used for immunostaining to find optimal conditions for clear and true signal A3B visualisation (**Figure 10**). Whereas successful co-staining of mitotic cells was observed in all conditions tested, images obtained from methanol fixation with either BSA or donkey serum were particularly blurry and the localisation

of all labels was not very clear. This was surprising, as methanol technique has been shown to preserve and visualise nuclear content better that PFA, in addition to high-specificity immunofluorescence and low background-noise (Levitt and King, 1987). Conversely, loss of integrity of cytoplasmic organelles and damage to both, nuclear envelope and contents, are also attributable to methanol fixation, thus it is not recommended for detailed immunolocalisation (Hoetelmans *et al*., 2001). Blurry images could also be attributable to insufficient PBS washes after fixation, leaving methanol residue on cells and disrupting further immunostaining steps. All PFA fixation methods showed non-specific HA staining in controls (WT NIKS), apart from combination of 4% PFA with CSK and two Triton X-100 incubations for subcellular fractionation. *In situ* fractionation technique allows extraction of the loosely and (with the second Triton X-100 incubation) tightly held nuclear proteins, to reveal stable localisation of protein of interest, which in our study is A3B (Sawasdichai *et al.,* 2010). Due to the absence of non-specific HA signal in WT NIKS population, this method was carried forward, as an optimised immunostaining technique, which is specific for HA-A3B investigation. It is worth noting, that the method, although specific to visualisation of A3B, might not be for pH3. As can be seen in **Figure 10C**, pH3 is localised in the nucleoli of WT NIKS during interphase and a blurry staining is seen around condensed chromatin during late anaphase (cell at the bottom left). Such staining pattern was not seen elsewhere, only in WT NIKS, suggesting either sub-optimal immunostaining conditions for pH3, or non-specific background signal seen around the cell in anaphase. Dotted localisation of pH3 during interphase could be due to chromatin condensation in late S or G_2 phase (Bruno *et al*., 1991).

Results obtained from optimised immunostaining conditions for visualisation of A3B subcellular localisation, have shown that cells undergoing mitosis phase co-express

both pH3 and A3B signals. Intriguingly, this co-staining was one-sided, since mitotic cells with visible staining for pH3 always were seen expressing A3B, however A3B expression was also detectable in the cells without phosphorylated H3 (**Figure 11**, **rows 1-4**). Phosphorylation of Ser-10 of histone H3 is observed from late G_2 phase, through mitotic stages and up until late anaphase, when dephosphorylation is initiated (Hendzel *et al*., 1997). This can be observed by the presence of weak pH3 staining signal in the late anaphase stage (**Figure 11, row 3**) and no signal during telophase (**Figure 11, rows 1-2**). Since phosphorylation is a rapid and dynamic cellular process, in comparison to slower gene expression mechanisms, this might explain observable A3B expression seen in cells, before or after histone H3 is dephosphorylated. Additionally, these results correlate with the previous Western blot analyses (**Figure 3**, **Figure 8**, **Figure 9**), where expression of pH3 is the strongest at 8-10h post-release from single-thymidine block (corresponds to G_2/M phase), whereas expression signal of A3B is detected from 6 to 14h post-release (albeit, in the insoluble fraction of the cell) (**Figure 8**). With the support of obtained results, we propose a schematic model for A3B expression pattern in the NIKS cell (**Figure 12**). In our model, expression peak of A3B coincides with the phosphorylation of histone H3 in late G_2 and M phases, however the protein is expressed and detectable from G_2 phase until the beginning of cytokinesis.

Subcellular localisation of endogenous A3B in the nucleus and around condensed chromatin has been confirmed, as previous studies observed this localisation with transfected A3B-GFP, hence in more artificial conditions (Lackey *et al.*, 2013; Salamango *et al*., 2018). However, the project left unanswered question on why localisation of A3B around the chromatin changes in telophase and is observed on the opposite sides of the condensed chromatin. It is probably due to yet

uninvestigated function of A3B, or unknown cellular mechanism that A3B is additionally implemented in, such as regulation of cell-cycle.

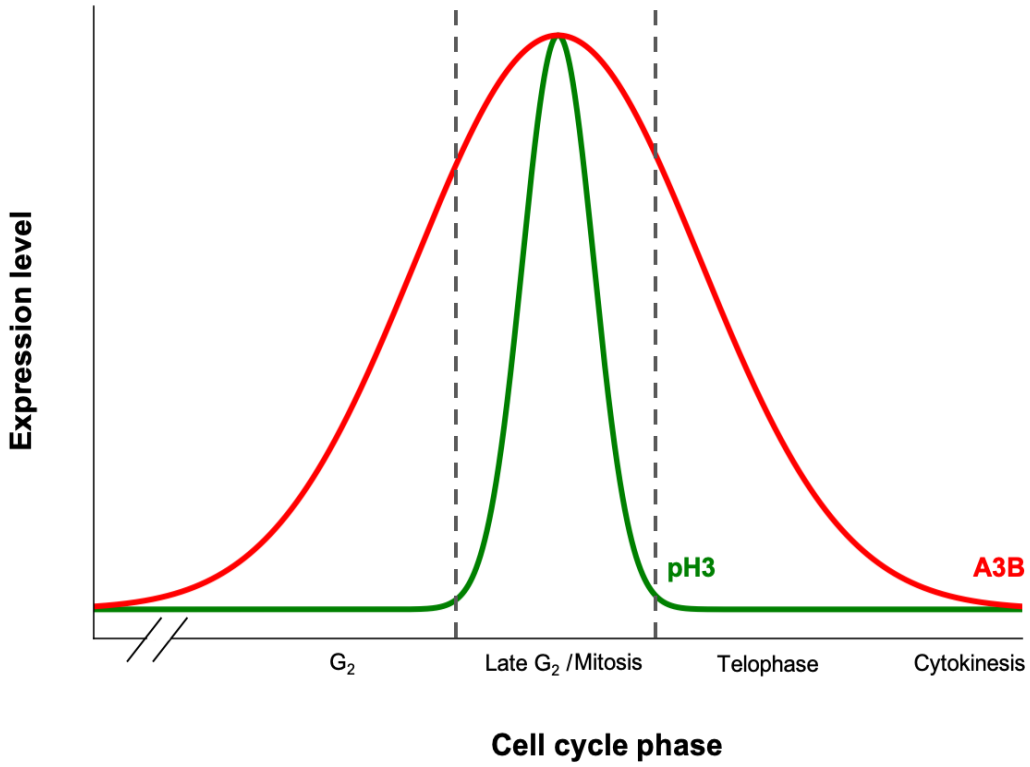


Figure 12. Schematic representation of APOBEC3B expression pattern.

A proposed model for APOBEC3B expression pattern in the cell cycle. Expression of phosphorylated histone H3 on Ser-10 residue (pH3), a mitotic marker, is visualised in green and APOBEC3B (A3B) in red. A peak for expression for both pH3 and A3B is observed at the same cell cycle phases – G_2/M transition and all mitosis stages up until telophase. A3B is considered to be induced in the cell from late G_2 cell cycle phase until cytokinesis, during which A3B is not detectable anymore.

4.5 Future directions

The main limitation of this study was time constraints due to the global COVID-19 pandemic and a consequent laboratory shut-down. The project consisted of three distinct parts, with each investigating implication of A3B in tumourigenesis from different perspectives, including application of computational and experimental techniques, and while comprehensive results were obtained in this project, there are still many outstanding questions left. Here, we suggest directions for the future studies continuing this work.

Firstly, to address the meta-analysis performed, it could be beneficial to explore association of A3B deletion with breast, ovarian, cervical, prostate, bladder, colon, lung, liver and oral cancers under other genetic inheritance models, such as dominant, recessive, overdominant, allelic and, possibly, additive (Horita and Kaneko, 2015). In addition, calculation of overall OR(95% CI) for every region separately, irrespective of cancer type, would be beneficial for comparisons of association between regions. However, this might not be possible or would not provide a clearer insight due to the lack of association studies for other cancer types rather than breast cancer, in regions such as Africa and Latin America. Having access of clinical patient data from each of the association studies, collected in this meta-analysis, and investigation of covariates that are associated with higher risk of developing cancer (smoking history, age, hormonal status), in addition to performing an extensive statistical analysis with Cochran's Q-test to test for heterogeneity and assessment for publication bias would greatly improve significance of the meta-analysis. It is worth noting, that all the case-control studies in this meta-analysis are retrospective, hence it is difficult to postulate which covariates also have a significant role in association, due to the differences of collected patient data in the past, which

is currently available for analysis. In this regard, it would be worth considering setting more prospective case-control association studies to further explore influence of ethnicity on A3B (and other members of A3 family) role in carcinogenesis. A multivariate analysis carried out to design the study upfront should include the major known factors to induce cancer, such as: age, body mass index (BMI), smoking history, infectious agents (HPV, EBV infections etc.), hormone replacement history, reproductive history, breastfeeding history and others (published by the National Cancer Institute). It would also be of particular interest to see association studies carried out Oceanic populations, where A3B deletion is particularly common and present in ~93% of inhabitants (Kidd *et al.,* 2007). Lastly, as previously mentioned (Discussion section 4.1), a statistically significant association was observed between APOBEC signatures 2 and 13 and A3H-I variant in breast and lung cancers harbouring A3A_B deletion (Starrett *et al*., 2016). The study has also shown that breast tumours with A3A_B deletion and other haplotypes of A3H do not display APOBEC signatures. This was additionally supported by extensive analysis of all TCGA tumour datasets, confirming presence of at least one copy of A3B or A3H-I with APOBEC signatures. Thus, further research on A3H-I variant distribution within A3A_B deletion carriers in different geographic areas, as well as exploring significance of ethnicity on cancer risk through APOBEC3 deregulation, are essential to understand mechanism through which APOBEC enzymes contribute to mutagenesis.

For the second part of the work with predominant focus on finding homozygous and heterozygous clones of the previously generated A3A_B isogenic model, the aim was achieved, and the library of the SCCs was screened. However, the loss of the single homozygous clone identified, led us to perform additional single-cell cloning

from transfected and sorted pools of NIKS with sgRNA75 or sgRNA76 at the last stages of work, which have not been genotyped yet. This should enhance the number of successful clones obtained in the future, after screening and genotyping steps, and allow further investigation of expression patterns of A3 deaminases, when A3B deletion is present. Comparison of expression of all A3 enzymes within the same and different genotypes (homozygous vs. heterozygous) should help get a clearer view on misregulation of A3 functions, in addition to what drives cancer development observed in deletion carriers.

Finally, investigation of A3B expression, regulation and functions within healthy and cancer cells are essential for understanding its role in tumourigenesis. Experiments for looking at stability of A3B protein with the treatment of protein synthesis inhibitor cycloheximide (CHX) could help answer why increase of A3B mRNA levels is not observed when the protein is abundant in G_2/M phase. Whether the observed effect of protein accumulation is a consequence of post-translational modification or increased translation in G_2/M is important, have not been yet explored, but would give us more insight on A3B biology within the cell and what mediates aberrant A3B activity causing genome instability. Moreover, now with the set optimised conditions for visualisation of A3B localisation, it would be particularly interesting to look at A3B co-localisation with other mitotic cell cycle markers, such as Cyclin A2 and Cyclin B1. Additional immunostaining work for investigation of A3B association with the cellular structures, in particular mitotic machinery, would aid further understanding of A3B localisation changes.

4.6 Concluding Remarks

In summary, the study has explored the importance of germline APOBEC3A_B deletion, commonly observed certain regions of global population, in the context of cancer from different perspectives. Herein, an extensive meta-analysis was conducted, combining all the up-to-date published association studies on A3A_B polymorphism and cancer risk. Obtained results proposed a significant association between A3B deletion and breast, as well as ovarian cancer, for both homozygous and heterozygous deletion carriers. Additionally, ethnicity has shown to be one of the main factors influencing the outcome of the disease, with Asian populations exhibiting a stronger association with cancer than Caucasians, suggesting deletion genotype as a possible biomarker for breast and ovarian cancer prevention and susceptibility. Although numerous case-control studies have reported association of A3B deletion with increased cancer risk, in addition to observed APOBEC-signature mutations within A3B-deficient tumours, mechanisms driving aberrant A3 expression and activity are yet to be fully explored. Hence, two heterozygous clones for isogenic A3A_B deletion model within non-cancerous NIKS line were successfully identified during this project, which will aid to provide insight on A3 activity and regulation, for future comparative studies and, possibly, shaping future targets for cancer therapy in A3B deletion carriers. Lastly, the most intriguing results of this project were obtained through immunochemical staining, where A3B was found to be expressed from late G_2 cell cycle phase until cytokinesis, with the observable localisation change from telophase subphase of mitosis. Intriguingly, such localisation shift might further explain A3B role in genome editing and its implication in cancer, however, it warrants further investigation.

5 References

Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Ng, A.W.T., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., Islam, S.M.A., Lopez-Bigas, N., Klimczak, L.J., McPherson, J.R., Morganella, S., Sabarinathan, R., Wheeler, D.A., Mustonen, V., Getz, G., Rozen, S.G. and Stratton, M.R. 2019. The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv*., p.322859.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A. V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., Boyault, S., Burkhardt, B., Butler, A.P., Caldas, C., Davies, H.R., Desmedt, C., Eils, R., Eyfjörd, J.E., Foekens, J.A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinsk, M., Jäger, N., Jones, D.T.W., Jonas, D., Knappskog, S., Koo, M., Lakhani, S.R., López-Otín, C., Martin, S., Munshi, N.C., Nakamura, H., Northcott, P.A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X.S., Raine, K., Ramakrishna, M., Richardson, A.L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T.N., Span, P.N., Teague, J.W., Totoki, Y., Tutt, A.N.J., Valdés-Mas, R., Van Buuren, M.M., Van 'T Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L.R., Zucman-Rossi, J., Andrew Futreal, P., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S.M., Siebert, R., Campo, E., Shibata, T., Pfister, S.M., Campbell, P.J. and Stratton, M.R. 2013. Signatures of mutational processes in human cancer. *Nature*. 500(7463), pp.415–421.

Allen-Hoffmann, B., Schlosser, S., Ivarie, C., Meisner, L., O'Connor, S. and Sattler, C., 2000. Normal Growth and Differentiation in a Spontaneously Immortalized Near-Diploid Human Keratinocyte Cell Line, NIKS. Journal of Investigative Dermatology, 114(3), pp.444-455. https://doi.org/10.1046/j.1523-1747.2000.00869.x.

Anwar, F., Davenport, M. and Ebrahimi, D., 2013. Footprint of APOBEC3 on the Genome of Human Retroelements. *Journal of Virology*, 87(14), pp.8195-8204.

Aynaud, M., Suspène, R., Vidalain, P., Mussil, B., Guétard, D., Tangy, F., Wain-Hobson, S. and Vartanian, J., 2012. Human Tribbles 3 Protects Nuclear DNA from Cytidine Deamination by APOBEC3A. *Journal of Biological Chemistry*, 287(46), pp.39182-39192.

Bennett, R.P., Presnyak, V., Wedekind, J.E. and Smith, H.C. 2008. Nuclear exclusion of the HIV-1 host defense factor APOBEC3G requires a novel cytoplasmic retention signal and is not dependent on RNA binding. *Journal of Biological Chemistry*. 283(12), pp.7320–7327.

Bonvin, M., Achermann, F., Greeve, I., Stroka, D., Keogh, A., Inderbitzin, D., Candinas, D.,

Sommer, P., Wain-Hobson, S., Vartanian, J. and Greeve, J., 2006. Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. *Hepatology*, 43(6), pp.1364-1374.

Bruno, S., Crissman, H.A., Bauer K.D., Darzynkiewicz, Z., Changes in cell nuclei during S Phase: Progressive chromatin condensation and altered expression of the proliferation-associated nuclear proteins Ki-67, cyclin (PCNA), p105, and p34, *Experimental Cell Research*, Volume 196, Issue 1, 1991, Pages 99-106, ISSN 0014-4827, https://doi.org/10.1016/0014-4827(91)90460-C.

Buisson, R., Lawrence, M., Benes, C. and Zou, L., 2017. APOBEC3A and APOBEC3B Activities Render Cancer Cells Susceptible to ATR Inhibition. *Cancer Research*, 77(17), pp.4567-4578.

Burns, M., Lackey, L., Carpenter, M., Rathore, A., Land, A., Leonard, B., Refsland, E., Kotandeniya, D., Tretyakova, N., Nikas, J., Yee, D., Temiz, N., Donohue, D., McDougle, R., Brown, W., Law, E. and Harris, R., 2013b. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*, 494(7437), pp.366-370.

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

Cannataro, V.L., Gaffney, S.G., Sasaki, T., Issaeva, N., Grewal, N.K.S., Grandis, J.R., Yarbrough, W.G., Burtness, B., Anderson, K.S. and Townsend, J.P. 2019. APOBEC-induced mutations and their cancer effect size in head and neck squamous cell carcinoma. *Oncogene*. 38(18), pp.3475–3487.

Caval, V., Suspène, R., Shapira, M., Vartanian, J. and Wain-Hobson, S., 2014. A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3′UTR enhances chromosomal DNA damage. *Nature Communications*, 5(1).

Cescon, D., Haibe-Kains, B. and Mak, T., 2015. APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation. *Proceedings of the National Academy of Sciences*, 112(9), pp.2841-2846.

Chaipan, C., Smith, J., Hu, W. and Pathak, V., 2012. APOBEC3G Restricts HIV-1 to a Greater Extent than APOBEC3F and APOBEC3DE in Human Primary CD4+ T Cells and Macrophages. *Journal of Virology*, 87(1), pp.444-453.

Chan, K., Roberts, S., Klimczak, L., Sterling, J., Saini, N., Malc, E., Kim, J., Kwiatkowski, D., Fargo, D., Mieczkowski, P., Getz, G. and Gordenin, D., 2015. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nature Genetics*, 47(9), pp.1067-

1072.

Chelico, L., Pham, P., Calabrese, P. and Goodman, M., 2006. APOBEC3G DNA deaminase acts processively 3′ → 5′ on single-stranded DNA. *Nature Structural & Molecular Biology*, 13(5), pp.392-399.

Chen, H., Lilley, C., Yu, Q., Lee, D., Chou, J., Narvaiza, I., Landau, N. and Weitzman, M., 2006. APOBEC3A Is a Potent Inhibitor of Adeno-Associated Virus and Retrotransposons. *Current Biology*, 16(5), pp.480-485.

Chen, T., Lee, C., Liu, H., Wu, C., Pickering, C., Huang, P., Wang, J., Chang, I., Yeh, Y., Chen, C., Li, H., Luo, J., Tan, B., Chan, T., Hsueh, C., Chu, L., Chen, Y., Zhang, B., Yang, C., Wu, C., Hsu, C., See, L., Tang, P., Yu, J., Liao, W., Chiang, W., Rodriguez, H., Myers, J., Chang, K. and Chang, Y., 2017. APOBEC3A is an oral cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion polymorphism. *Nature Communications*, 8(1).

Chiu, Y. and Greene, W., 2008. The APOBEC3 Cytidine Deaminases: An Innate Defensive Network Opposing Exogenous Retroviruses and Endogenous Retroelements. *Annual Review of Immunology*, 26(1), pp.317-353.

Chiu, Y., Witkowska, H., Hall, S., Santiago, M., Soros, V., Esnault, C., Heidmann, T. and Greene, W., 2006. High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proceedings of the National Academy of Sciences*, 103(42), pp.15588-15593.

Conticello, S. 2008. The AID/APOBEC family of nucleic acid mutators. *Genome Biology*, 9(6), p.229.

Conticello, S., Langlois, M., Yang, Z., Neuberger, M. (2007). DNA deamination in immunity: AID in the context of its APOBEC relatives. *Advances in Immunology*, 94, 37–73. https://doi.org/10.1016/S0065-2776(06)94002-4

Cortez, L., Brown, A., Dennis, M., Collins, C., Brown, A., Mitchell, D., Mertz, T. and Roberts, S., 2019. APOBEC3A is a prominent cytidine deaminase in breast cancer. *PLOS Genetics*, 15(12), p.e1008545.

Dainichi, T., Nakano, Y., Wakae, K., Otsuka, M., Muramatsu, M. and Kabashima, K. 2019. APOBEC3 regulates keratinocyte differentiation and expression of Notch3. *Experimental Dermatology*. 28(11), pp.1341–1347.

Dang, C., 2012. MYC on the Path to Cancer. *Cell*, 149(1), pp.22-35.

Daniels, T., Killinger, K., Michal, J., Wright Jr., R. and Jiang, Z., 2009. Lipoproteins,

cholesterol homeostasis and cardiac health. *International Journal of Biological Sciences*, pp.474-488.

de Bruin, E., McGranahan, N., Mitter, R., Salm, M., Wedge, D., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A., Gronroos, E., Muhammad, M., Horswell, S., Gerlinger, M., Varela, I., Jones, D., Marshall, J., Voet, T., Van Loo, P., Rassl, D., Rintoul, R., Janes, S., Lee, S., Forster, M., Ahmad, T., Lawrence, D., Falzon, M., Capitanio, A., Harkins, T., Lee, C., Tom, W., Teefe, E., Chen, S., Begum, S., Rabinowitz, A., Phillimore, B., Spencer-Dene, B., Stamp, G., Szallasi, Z., Matthews, N., Stewart, A., Campbell, P. and Swanton, C., 2014. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*, 346(6206), pp.251-256.

Du, Y., Tao, X., Wu, J., Yu, H., Yu, Y. and Zhao, H., 2018. APOBEC3B up-regulation independently predicts ovarian cancer prognosis: a cohort study. *Cancer Cell International*, 18(1).

Faltas, B., Prandi, D., Tagawa, S., Molina, A., Nanus, D., Sternberg, C., Rosenberg, J., Mosquera, J., Robinson, B., Elemento, O., Sboner, A., Beltran, H., Demichelis, F. and Rubin, M., 2016. Clonal evolution of chemotherapy-resistant urothelial carcinoma. *Nature Genetics*, 48(12), pp.1490-1499.

Feng, C., Zhang, Y., Huang, J., Zheng, Q., Yang, Y. and Xu, B., 2020. The Prognostic Significance of APOBEC3B and PD-L1/PD-1 in Nasopharyngeal Carcinoma. *Applied Immunohistochemistry & Molecular Morphology*, p.1.

Freije, A., Ceballos, L., Coisy, M., Barnes, L., Rosa, M., De Diego, E., Blanchard, J. and Gandarillas, A., 2012. Cyclin E drives human keratinocyte growth into differentiation. *Oncogene*, 31(50), pp.5180-5192.

Gandarillas, A., 2012. The mysterious human epidermal cell cycle, or an oncogene-induced differentiation checkpoint. *Cell Cycle*, 11(24), pp.4507-4516.

Gansmo, L.B., Romundstad, P., Hveem, K., Vatten, L., Nik-Zainal, S., Lønning, P.E. and Knappskog, S. 2018. APOBEC3A/B deletion polymorphism and cancer risk. *Carcinogenesis*. 39(2), pp.118–124.

Genther, S.M., Sterling, S., Duensing, S., Münger, K., Sattler, C. and Lambert, P.F. 2003. Quantitative Role of the Human Papillomavirus Type 16 E5 Gene during the Productive Stage of the Viral Life Cycle. *Journal of Virology*. 77(5), 2832 LP – 2842.

Göhler, S., Da Silva Filho, M., Johansson, R., Enquist-Olsson, K., Henriksson, R., Hemminki, K., Lenner, P. and Försti, A., 2016. 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*, 142(1), pp.273-276.

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*, 7(46), pp.74979-74986.

Hanahan, D. and Weinberg, R., 2011. Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), pp.646-674.

Harris, R. and Dudley, J., 2015. APOBECs and virus restriction. *Virology*, 479-480, pp.131-145.

Harris, R., 2015. Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. *Breast Cancer Research*, 17(1).

Harris, R., Petersen-Mahrt, S. and Neuberger, M., 2002. RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators. *Molecular Cell*, 10(5), pp.1247-1253.

Helleday, T., Eshtad, S. and Nik-Zainal, S., 2014. Mechanisms underlying mutational signatures in human cancers. *Nature Reviews Genetics*, 15(9), pp.585-598.

Henderson, S. and Fenton, T. 2015. APOBEC3 genes: Retroviral restriction factors to cancer drivers. *Trends in Molecular Medicine*. 21(5), pp.274–284.

Henderson, S., Chakravarthy, A., Su, X., Boshoff, C. and Fenton, T., 2014. APOBEC-Mediated Cytosine Deamination Links PIK3CA Helical Domain Mutations to Human Papillomavirus-Driven Tumor Development. *Cell Reports*, 7(6), pp.1833-1841.

Hendzel, M., Wei, Y., Mancini, M., Van Hooser, A., Ranalli, T., Brinkley, B., Bazett-Jones, D. and Allis, C., 1997. Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma*, 106(6), pp.348-360.

Hoetelmans, R.W.M., Prins, F.A., Velde, I.C., Van der Meer, J., Van de Velde, C.J.H. & Van Dierendonck, J.H. 2001. Effects of acetone, methanol, or paraformaldehyde on cellular structure, visualized by reflection contrast microscopy and transmission and scanning electron microscopy. *Applied Immunohistochemistry and Molecular Morphology*, vol. 9, no. 4, pp. 346-351.

Holschneider, C. and Berek, J., 2000. Ovarian cancer: Epidemiology, biology, and prognostic factors. *Seminars in Surgical Oncology*, 19(1), pp.3-10.

Horita, N. and Kaneko, T., 2015. Genetic model selection for a case–control study and a meta-analysis. *Meta Gene*, 5, pp.1-8.

Kanu, N., Cerone, M., Goh, G., Zalmas, L., Bartkova, J., Dietzen, M., McGranahan, N., Rogers, R., Law, E., Gromova, I., Kschischo, M., Walton, M., Rossanese, O., Bartek, J., Harris, R., Venkatesan, S. and Swanton, C., 2016. DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. *Genome Biology*, 17(1).

Kehdy, F., Gouveia, M., Machado, M., Magalhães, W., Horimoto, A., Horta, B., Moreira, R., Leal, T., Scliar, M., Soares-Souza, G., Rodrigues-Soares, F., Araújo, G., Zamudio, R., Sant Anna, H., Santos, H., Duarte, N., Fiaccone, R., Figueiredo, C., Silva, T., Costa, G., Beleza, S., Berg, D., Cabrera, L., Debortoli, G., Duarte, D., Ghirotto, S., Gilman, R., Gonçalves, V., Marrero, A., Muniz, Y., Weissensteiner, H., Yeager, M., Rodrigues, L., Barreto, M., Lima-Costa, M., Pereira, A., Rodrigues, M. and Tarazona-Santos, E., 2015. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proceedings of the National Academy of Sciences*, 112(28), pp.8696-8701.

Kidd, J., Newman, T., Tuzun, E., Kaul, R. and Eichler, E., 2007. Population Stratification of a Common APOBEC Gene Deletion Polymorphism. *PLOS Genetics*, 3(4), p.e63.

Kim, Y., Sun, D., Yoon, J., Ko, Y., Won, H. and Kim, J., 2020. Clinical implications of APOBEC3A and 3B expression in patients with breast cancer. *PLOS ONE*, 15(3), p.e0230261.

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(9), pp.2955–2964.

Klonowska, K., Kluzniak, W., Rusak, B., Jakubowska, A., Ratajska, M., Krawczynska, N., Vasilevska, D., Czubak, K., Wojciechowska, M., Cybulski, C., Lubinski, J. and Kozlowski, P., 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*, 8(44), pp.76357-76374.

Knisbacher, B.A., Gerber, D. and Levanon, E.Y. 2016. DNA Editing by APOBECs: A Genomic Preserver and Transformer. *Trends in Genetics*. 32(1), pp.16–28.

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, pp.261–270.

Komatsu, A., Nagasaki, K., Fujimori, M., Amano, J., & Miki, Y., 2008. Identification of novel deletion polymorphisms in breast cancer. *International Journal of Oncology*, 33, 261-270. https://doi.org/10.3892/ijo_00000005

Lackey, L., Law, E., Brown, W. and Harris, R., 2013. Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. *Cell Cycle*, 12(5), pp.762-772.

Lada, A.G., Krick, C.F., Kozmin, S.G., Mayorov, V.I., Karpova, T.S., Rogozin, I.B., Pavlov, Y.I. Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. *Biochemistry*. 2011; 76: 131-146.

Landry, S., Narvaiza, I., Linfesty, D. and Weitzman, M., 2011. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO reports*, 12(5), pp.444-450.

Law, E., Sieuwerts, A., LaPara, K., Leonard, B., Starrett, G., Molan, A., Temiz, N., Vogel, R., Meijer-van Gelder, M., Sweep, F., Span, P., Foekens, J., Martens, J., Yee, D. and Harris, R., 2016. The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Science Advances*, 2(10), p.e1601737.

Leonard, B., Hart, S.N., Burns, M.B., Carpenter, M.A., Temiz, A., Rathore, A., Vogel, R.I., Nikas, J.B., Law, E.K., Brown, W.L., Li, Y., Zhang, Y., Maurer, M.J., Oberg, A.L., Julie, M., Shridhar, V., Bell, D.A., April, C., Bentley, D., Cheetham, R.K., Fan, J., Grocock, R. and Humphray, S. 2013. APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer Research*, 73(24).

Leonard, B., McCann, J., Starrett, G., Kosyakovsky, L., Luengas, E., Molan, A., Burns, M., McDougle, R., Parker, P., Brown, W. and Harris, R., 2015. The PKC/NF-κB Signaling Pathway Induces APOBEC3B Expression in Multiple Human Cancers. *Cancer Research*, 75(21), pp.4538-4547.

Levitt, D. and King, M., 1987. Methanol fixation permits flow cytometric analysis of immunofluorescent stained intracellular antigens. *Journal of Immunological Methods*, 96(2), pp.233-237.

Li, M., Shandilya, S.M.D., Carpenter, M.A., Rathore, A., Brown, W.L., Perkins, A.L., Harki, D.A., Solberg, J., Hook, D.J., Pandey, K.K., Parniak, M.A., Johnson, J.R., Krogan, N.J., Somasundaran, M., Ali, A., Schiffer, C.A. and Harris, R.S. 2012. First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. *ACS Chemical Biology*. 7(3), pp.506–517.

Long, J., Delahanty, R., Li, G., Gao, Y., Lu, W., Cai, Q., Xiang, Y., Li, C., Ji, B., Zheng, Y., Ali, S., Shu, X. and Zheng, W., 2013. A Common Deletion in the APOBEC3 Genes

and Breast Cancer Risk. *Journal of the National Cancer Institute*, 105(8), pp.573-579.

Lucifora, J., Xia, Y., Reisinger, F., Zhang, K., Stadler, D., Cheng, X., Sprinzl, M.F., Koppensteiner, H., Makowska, Z., Volz, T., Remouchamps, C., Chou, W.-M., Thasler, W.E., Hüser, N., Durantel, D., Liang, T.J., Münk, C., Heim, M.H., Browning, J.L., Dejardin, E., Dandri, M., Schindler, M., Heikenwalder, M. and Protzer, U. 2014. Specific and Nonhepatotoxic Degradation of Nuclear Hepatitis B Virus cccDNA. *Science*. 343(6176), 1221 LP – 1228.

Marino, D., Perković, M., Hain, A., Jaguva Vasudevan, A., Hofmann, H., Hanschmann, K., Mühlebach, M., Schumann, G., König, R., Cichutek, K., Häussinger, D. and Münk, C., 2016. APOBEC4 Enhances the Replication of HIV-1. *PLOS ONE*, 11(6), p.e0155422.

Marouf, C., Göhler, S., Filho, M., Hajji, O., Hemminki, K., Nadifi, S. and Försti, A., 2016. Analysis of functional germline variants in APOBEC3 and driver genes on breast cancer risk in Moroccan study population. *BMC Cancer*, 16(1).

McGranahan, N., Favero, F., de Bruin, E. C., Birkbak, N. J., Szallasi, Z., & Swanton, C., 2015. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Science Translational Medicine*, 7(283), 283ra54. https://doi.org/10.1126/scitranslmed.aaa1408

McGranahan, N., Furness, A. J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S. K., Jamal-Hanjani, M., Wilson, G. A., Birkbak, N. J., Hiley, C. T., Watkins, T. B., Shafi, S., Murugaesu, N., Mitter, R., Akarca, A. U., Linares, J., Marafioti, T., Henry, J. Y., Van Allen, E. M., Miao, D., Swanton, C., 2016. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*, 351(6280), 1463–1469. https://doi.org/10.1126/science.aaf1490

Menendez, D., Nguyen, T., Snipe, J. and Resnick, M., 2017. The Cytidine Deaminase APOBEC3 Family Is Subject to Transcriptional Regulation by p53. *Molecular Cancer Research*, 15(6), pp.735-743.

Middlebrooks, C., Banday, A., Matsuda, K., Udquim, K., Onabajo, O., Paquin, A., Figueroa, J., Zhu, B., Koutros, S., Kubo, M., Shuin, T., Freedman, N., Kogevinas, M., Malats, N., Chanock, S., Garcia-Closas, M., Silverman, D., Rothman, N. and Prokunina-Olsson, L., 2016. Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. *Nature Genetics*, 48(11), pp.1330-1338.

Mikl, M., Watt, I., Lu, M., Reik, W., Davies, S., Neuberger, M. and Rada, C., 2005. Mice Deficient in APOBEC2 and APOBEC3. Molecular and Cellular Biology, 25(16),

pp.7270-7277.

Mitra, M., Singer, D., Mano, Y., Hritz, J., Nam, G., Gorelick, R., Byeon, I., Gronenborn, A., Iwatani, Y. and Levin, J., 2015. Sequence and structural determinants of human APOBEC3H deaminase and anti-HIV-1 activities. *Retrovirology*, 12(1), p.3.

Morgan, D., 2007. The Cell Cycle. London: New Science Press.

Morisawa, T., Marusawa, H., Ueda, Y., Iwai, A., Okazaki, I., Honjo, T. and Chiba, T., 2008. Organ-specific profiles of genetic changes in cancers caused by activation-induced cytidine deaminase expression. *International Journal of Cancer*, 123(12), pp.2735-2740.

Münk, C., Willemsen, A. and Bravo, I., 2012. An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC Evolutionary Biology*, 12(1), p.71.

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(5), pp.553-563.

Mussil, B., Suspène, R., Aynaud, M., Gauvrit, A., Vartanian, J. 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(8), p.e73641.

Nikkilä, J., Kumar, R., Campbell, J., Brandsma, I., Pemberton, H., Wallberg, F., Nagy, K., Scheer, I., Vertessy, B., Serebrenik, A., Monni, V., Harris, R., Pettitt, S., Ashworth, A. and Lord, C., 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(1), pp.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., Stebbings, L.A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K.W., Mudie, L.J., Varela, I., McBride, D.J., Bignell, G.R., Cooke, S.L., Shlien, A., Gamble, J., Whitmore, I., Maddison, M., Tarpey, P.S., Davies, H.R., Papaemmanuil, E., Stephens, P.J., McLaren, S., Butler, A.P., Teague, J.W., Jönsson, G., Garber, J.E., Silver, D., Miron, P., Fatima, A., Boyault, S., Langerod, A., Tutt, A., Martens, J.W.M., Aparicio, S.A.J.R., Borg, Å., Salomon, A.V., Thomas, G., Borresen-Dale, A.L., Richardson, A.L., Neuberger, M.S., Futreal, P.A., Campbell, P.J. and Stratton, M.R. 2012. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 149(5), pp.979–993.

Nik-Zainal, S., Wedge, D., Alexandrov, L., Petljak, M., Butler, A., Bolli, N., Davies, H.,

Knappskog, S., Martin, S., Papaemmanuil, E., Ramakrishna, M., Shlien, A., Simonic, I., Xue, Y., Tyler-Smith, C., Campbell, P. and Stratton, M., 2014. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nature Genetics*, 46(5), pp.487-491.

Nowarski, R., Wilner, O., Cheshin, O., Shahar, O., Kenig, E., Baraz, L., Britan-Rosich, E., Nagler, A., Harris, R., Goldberg, M., Willner, I. and Kotler, M., 2012. APOBEC3G enhances lymphoma cell radioresistance by promoting cytidine deaminase-dependent DNA repair. *Blood*, 120(2), pp.366-375.

Okazaki, I., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K. and Honjo, T., 2003. Constitutive Expression of AID Leads to Tumorigenesis. *Journal of Experimental Medicine*, 197(9), pp.1173-1181.

Okuyama, S., Marusawa, H., Matsumoto, T., Ueda, Y., Matsumoto, Y., Endo, Y., Takai, A. and Chiba, T. 2012. Excessive activity of apolipoprotein B mRNA editing enzyme catalytic polypeptide 2 (APOBEC2) contributes to liver and lung tumorigenesis. *International Journal of Cancer*. 130(6), pp.1294–1301.

Olson, M.E., Harris, R.S. and Harki, D.A. 2018. APOBEC Enzymes as Targets for Virus and Cancer Therapy. *Cell Chemical Biology*. 25(1), pp.36–49.

Pak, V., Heidecker, G., Pathak, V. and Derse, D., 2011. The Role of Amino-Terminal Sequences in Cellular Localization and Antiviral Activity of APOBEC3B. *Journal of Virology*, 85(17), pp.8538-8547.

Pan, J., Ahmad Zabidi, M.M., Chong, B., Meng, M., Ng, P., Hasan, S.N., Sandey, B., Bahnu, S., Rajadurai, P., Yip, C., Rueda, O.M., Caldas, C., Chin, S., Teo, S. *bioRxiv* 2020.06.04.135251; doi: https://doi.org/10.1101/2020.06.04.135251

Periyasamy, M., Patel, H., Lai, C., Nguyen, V., Nevedomskaya, E., Harrod, A., Russell, R., Remenyi, J., Ochocka, A., Thomas, R., Fuller-Pace, F., Győrffy, B., Caldas, C., Navaratnam, N., Carroll, J., Zwart, W., Coombes, R., Magnani, L., Buluwela, L. and Ali, S., 2015. APOBEC3B-Mediated Cytidine Deamination Is Required for Estrogen Receptor Action in Breast Cancer. *Cell Reports*, 13(1), pp.108-121.

Periyasamy, M., Singh, A.K., Gemma, C., Kranjec, C., Farzan, R., Leach, D.A., Navaratnam, N., Pálinkás, H.L., Vertessy, B.G., Fenton, T.R., Doorbar, J., Fuller-Pace, F., Meek, D.W., Coombes, R.C., Buluwela, L. and Ali, S. 2017. P53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells. *Nucleic Acids Research*. 45(19), pp.11056–11069.

Petljak, M., Alexandrov, L., Brammeld, J., Price, S., Wedge, D., Grossmann, S., Dawson, K., Ju, Y., Iorio, F., Tubio, J., Koh, C., Georgakopoulos-Soares, I., Rodríguez–Martín, B., Otlu, B., O'Meara, S., Butler, A., Menzies, A., Bhosle, S., Raine, K., Jones, D., Teague, J., Beal, K., Latimer, C., O'Neill, L., Zamora, J., Anderson, E., Patel, N., Maddison, M., Ng, B., Graham, J., Garnett, M., McDermott, U., Nik-Zainal, S., Campbell, P. and Stratton, M., 2019. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell*, 176(6), pp.1282-1294.e20.

Qi, G., Xiong, H. and Zhou, C., 2014. APOBEC3 deletion polymorphism is associated with epithelial ovarian cancer risk among Chinese women. *Tumor Biology*, 35(6), pp.5723-5726.

Radmanesh, H., Spethmann, T., Enßen, J., Schürmann, P., Bhuju, S., Geffers, R., Antonenkova, N., Khusnutdinova, E., Sadr-Nabavi, A., Shandiz, F., Park-Simon, T., Hillemanns, P., Christiansen, H., Bogdanova, N. and Dörk, T., 2017. Assessment of an APOBEC3B truncating mutation, c.783delG, in patients with breast cancer. *Breast Cancer Research and Treatment*, 162(1), pp.31-37.

Rebhandl, S., Huemer, M., Greil, R. and Geisberger, R. 2015. AID/APOBEC deaminases and cancer. *Oncoscience*. 2(4), pp.320–333.

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

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A., Brousse, N., Muramatsu, M., Notarangelo, L., Kinoshita, K., Honjo, T., Fischer, A. and Durandy, A., 2000. Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2). *Cell*, 102(5), pp.565-575.

Rezaei, M., Hashemi, M., Hashemi, S.M., Mashhadi, M.A., Taheri, M. APOBEC3 deletion is associated with breast cancer risk in a sample of southeast Iranian population. Int J Mol *Cell Med*. 2015;4:103–108.

Roberts, S., Lawrence, M., Klimczak, L., Grimm, S., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G., Carter, S., Saksena, G., Harris, S., Shah, R., Resnick, M., Getz, G. and Gordenin, D., 2013. An APOBEC cytidine deaminase mutagenesis pattern is

widespread in human cancers. *Nature Genetics*, 45(9), pp.970-976.

Roberts, S., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L., Kryukov, G., Malc, E., Mieczkowski, P., Resnick, M. and Gordenin, D., 2012. Clustered Mutations in Yeast and in Human Cancers Can Arise from Damaged Long Single-Strand DNA Regions. *Molecular Cell*, 46(4), pp.424-435.

Rogozin, I., Basu, M., Jordan, I., Pavlov, Y. and Koonin, E., 2005. APOBEC4, a New Member of the AID/APOBEC Family of Polynucleotide (Deoxy)Cytidine Deaminases Predicted by Computational Analysis. *Cell Cycle*, 4(9), pp.1281-1285.

Salamango, D., McCann, J., Demir, Ö., Brown, W., Amaro, R. and Harris, R., 2018. APOBEC3B Nuclear Localization Requires Two Distinct N-Terminal Domain Surfaces. *Journal of Molecular Biology*, 430(17), pp.2695-2708.

Salter, J.D., Bennett, R.P. and Smith, H.C. 2016. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends in Biochemical Sciences*. 41(7), pp.578–594.

Saraconi, G., Severi, F., Sala, C., Mattiuz, G. and Conticello, S.G. 2014. The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome Biology*. 15(7), pp.1–10.

Sato, Y., Ohtsubo, H., Nihei, N., Kaneko, T., Sato, Y., Adachi, S., Kondo, S., Nakamura, M., Mizunoya, W., Iida, H., Tatsumi, R., Rada, C. and Yoshizawa, F., 2018. Apobec2 deficiency causes mitochondrial defects and mitophagy in skeletal muscle. *The FASEB Journal*, 32(3), pp.1428-1439.

Sato, Y., Probst, H., Tatsumi, R., Ikeuchi, Y., Neuberger, M. and Rada, C., 2010. Deficiency in APOBEC2 Leads to a Shift in Muscle Fiber Type, Diminished Body Mass, and Myopathy. *Journal of Biological Chemistry*, 285(10), pp.7111-7118.

Sawasdichai, A., Chen, H.T., Abdul Hamid, N., Jayaraman, P.S., & Gaston, K., 2010. In situ subcellular fractionation of adherent and non-adherent mammalian cells. Journal of visualized experiments : *JoVE*, (41), 1958. https://doi.org/10.3791/1958

Serebrenik, A., Starrett, G., Leenen, S., Jarvis, M., Shaban, N., Salamango, D., Nilsen, H., Brown, W. and Harris, R., 2019. The deaminase APOBEC3B triggers the death of cells lacking uracil DNA glycosylase. *Proceedings of the National Academy of Sciences*, 116(44), pp.22158-22163.

Sharma, S., Patnaik, S., Thomas Taggart, R., Kannisto, E., Enriquez, S., Gollnick, P. and Baysal, B., 2015. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nature Communications*, 6(1).

Sheehy, A., Gaddis, N., Choi, J. and Malim, M., 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, 418(6898), pp.646-650.

Sheskin, D., 2004. Handbook of parametric and nonparametric statistical procedures. 3rd ed. *Boca Raton: Chapman & Hall /CRC*.

Shi, K., Demir, Ö., Carpenter, M., Banerjee, S., Harki, D., Amaro, R., Harris, R. and Aihara, H., 2020. Active site plasticity and possible modes of chemical inhibition of the human DNA deaminase APOBEC3B. *FASEB BioAdvances*, 2(1), pp.49-58.

Sieuwerts, A.M., Schrijver, W., Dalm, S., de Weerd, V., Moelans, C., ter Hoeve, N., van Diest, P., Martens, J. and van Deurzen, C., 2017. Progressive APOBEC3B mRNA expression in distant breast cancer metastases. *PLOS ONE*, 12(1), p.e0171343.

Sieuwerts, A.M., Willis, S., Burns, M.B., Look, M.P., Gelder, M.E.M. Van, Schlicker, A., Heideman, M.R., Jacobs, H., Wessels, L., Leyland-Jones, B., Gray, K.P., Foekens, J.A., Harris, R.S. and Martens, J.W.M. 2014. Elevated APOBEC3B Correlates with Poor Outcomes for Estrogen-Receptor-Positive Breast Cancers. *Hormones and Cancer*. 5(6), pp.405–413.

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 and Developmental Biology*. 23(3), pp.258–268.

Smith, N.J. and Fenton, T.R. 2019. The APOBEC3 genes and their role in cancer: Insights from human papillomavirus. *Journal of Molecular Endocrinology*. 62(4), pp.R269–R287.

Starrett, G., Luengas, E., McCann, J., Ebrahimi, D., Temiz, N., Love, R., Feng, Y., Adolph, M., Chelico, L., Law, E., Carpenter, M. and Harris, R., 2016. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nature Communications*, 7(1).

Surget, S., Khoury, M. and Bourdon, J., 2013. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. *OncoTargets and Therapy*, p.57.

Suspène, R., Aynaud, M., Guetard, D., Henry, M., Eckhoff, G., Marchio, A., Pineau, P., Dejean, A., Vartanian, J. and Wain-Hobson, S., 2011. Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proceedings of the National Academy of Sciences*, 108(12), pp.4858-4863.

Suspène, 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*. 45(6), pp.3231–3241.

Swanton, C., McGranahan, N., Starrett, G. and Harris, R., 2015. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discovery*, 5(7), pp.704-712.

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

Teng, B., Burant, C. and Davidson, N., 1993. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science*, 260(5115), pp.1816-1819.

Vieira, V., 2013. The Role of Cytidine Deaminases on Innate Immune Responses against Human Viral Infections. *BioMed Research International*, 2013(683095).

Vitiello, G., de Sousa Pereira, N., Amarante, M., Banin-Hirata, B., Campos, C., de Oliveira, K., Losi-Guembarovski, R. and Watanabe, M., 2020. Germline APOBEC3B deletion influences clinicopathological parameters in luminal-A breast cancer: evidences from a southern Brazilian cohort. *Journal of Cancer Research and Clinical Oncology*, 146(6), pp.1523-1532.

Vonica, A., Rosa, A., Arduini, B. and Brivanlou, A., 2011. APOBEC2, a selective inhibitor of TGFβ signaling, regulates left–right axis specification during early embryogenesis. *Developmental Biology*, 350(1), pp.13-23.

Vural, S., Simon, R. and Krushkal, J., 2018. Correlation of gene expression and associated mutation profiles of APOBEC3A, APOBEC3B, REV1, UNG, and FHIT with chemosensitivity of cancer cell lines to drug treatment. *Human Genomics*, 12(1).

Wang, L., Ni, Y., Su, B., Mu, X., Shen, H., Du, J., 2013. MicroRNA-34b functions as a tumor suppressor and acts as a nodal point in the feedback loop with Met. *International Journal of Oncology*, 42(3), pp.957-962.

Wang, Y., Wu, S., Zheng, S., Wang, S., Wali, A., Ezhilarasan, R., Sulman, E., Koul, D. and Alfred Yung, W., 2017. APOBEC3G acts as a therapeutic target in mesenchymal gliomas by sensitizing cells to radiation-induced cell death. *Oncotarget*, 8(33), pp.54285-54296.

Warren, C., Xu, T., Guo, K., Griffin, L., Westrich, J., Lee, D., Lambert, P., Santiago, M. and

Pyeon, D., 2014. APOBEC3A Functions as a Restriction Factor of Human Papillomavirus. *Journal of Virology*, 89(1), pp.688-702.

Wen, W., Soo, J., Kwan, P., Hong, E., Khang, T., Mariapun, S., Lee, C., Hasan, S., Rajadurai, P., Yip, C., Mohd Taib, N. and Teo, S., 2016. Germline APOBEC3B deletion is associated with breast cancer risk in an Asian multi-ethnic cohort and with immune cell presentation. *Breast Cancer Research*, 18(1).

Xuan, D., Li, G., Cai, Q., Deming-Halverson, S., Shrubsole, M., Shu, X., Kelley, M., Zheng, W. and Long, J., 2013. APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis*, 34(10), pp.2240-2243.

Yamanaka, S., Balestra, M., Ferrell, L., Fan, J., Arnold, K., Taylor, S., Taylor, J. and Innerarity, T., 1995. Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proceedings of the National Academy of Sciences*, 92(18), pp.8483-8487.

Zhang, J., Wei, W., Jin, H.C., Ying, R.C., Zhu, A.K. and Zhang, F.J. 2015. The roles of APOBEC3B in gastric cancer. *International Journal of Clinical and Experimental Pathology*. 8(5), pp.5089–5096.

Zhang, T., Cai, J., Chang, J., Yu, D., Wu, C., Yan, T., Zhai, K., Bi, X., Zhao, H., Xu, J., Tan, W., Qu, C. and Lin, D., 2013. Evidence of associations of APOBEC3B gene deletion with susceptibility to persistent HBV infection and hepatocellular carcinoma. *Human Molecular Genetics*, 22(6), pp.1262-1269.

Zhen, A., Du, J., Zhou, X., Xiong, Y. and Yu, X., 2012. Reduced APOBEC3H Variant Anti-Viral Activities Are Associated with Altered RNA Binding Activities. *PLOS ONE*, 7(7), p.e38771.

Zhu, M., Wang, Y., Wang, C., Shen, W., Liu, J., Geng, L., Cheng, Y., Dai, J., Jin, G., Ma, H., Hu, Z. and Shen, H., 2015. The eQTL-missense polymorphisms of APOBEC3H are associated with lung cancer risk in a Han Chinese population. *Scientific Reports*, 5(1).

6 Supplementary Data

Supplementary Table 1. List of publications screened for the meta-analysis on association of A3A_B deletion with various cancer types.
All the related articles were compiled from the PubMed database. PubMed identifier number is presented as PMID. Studies, selected as eligible for the meta-analysis, are highlighted in grey and marked as “Yes” in the “Included in the meta-analysis?” column.

PMID (PubMed)	Title	First Author	Publication Year	Included in the meta-analysis?
16729314	Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication	Bonvin	2006	No
17257057	Polymorphisms of CUL5 are associated with CD4+ T cell loss in HIV-1 infected individuals	An	2007	No
18636146	Identification of novel deletion polymorphisms in breast cancer	Komatsu	2008	No
18448976	Hepatitis B: modern concepts in pathogenesis--APOBEC3 cytidine deaminases as effectors in innate immunity against the hepatitis B virus	Bonvin	2008	No
18786991	Mouse APOBEC3 restricts friend leukemia virus infection and pathogenesis in vivo	Takeda	2008	No
19153233	Expression of murine APOBEC3 alleles in different mouse strains and their effect on mouse mammary tumor virus infection	Okeoma	2009	No
19837465	Multiple ways of targeting APOBEC3-virion infectivity factor interactions for anti-HIV-1 drug development	Smith	2009	No
19726503	The AKV murine leukemia virus is restricted and hypermutated by mouse APOBEC3	Langlois	2009	No
19698078	APOBEC3B deletion and risk of HIV-1 acquisition	An	2009	No
20617165	Adaptive evolution of Mus Apobec3 includes retroviral insertion and positive selection at two clusters of residues flanking the substrate groove	Sanville	2010	No
20980520	Persistent Friend virus replication and disease in Apobec3-deficient mice expressing functional B-cell-activating factor receptor	Santiago	2011	No
22275865	Two genetic determinants acquired late in Mus evolution regulate the inclusion of exon 5, which alters mouse APOBEC3 translation efficiency	Li	2012	No

PMID (PubMed)	Title	First Author	Publication Year	Included in the meta-analysis?
22825724	Identification of a novel population in high-grade oligodendroglial tumors not deleted on 1p/19q using array CGH	Talagas	2012	No
23411593	A common deletion in the APOBEC3 genes and breast cancer risk	Long	2013	Yes
24367644	APOBEC3G oligomerization is associated with the inhibition of both Alu and LINE-1 retrotransposition	Koyama	2013	No
23865062	The role of cytidine deaminases on innate immune responses against human viral infections	Vieira	2013	No
23870316	Gene loss and adaptation to hominids underlie the ancient origin of HIV-1	Etienne	2013	No
23880220	Identification of the feline foamy virus Bet domain essential for APOBEC3 counteraction	Lukic	2013	No
23449789	APOBEC3 inhibition of mouse mammary tumor virus infection: the role of cytidine deamination versus inhibition of reverse transcription	MacMillan	2013	No
23213177	Evidence of associations of APOBEC3B gene deletion with susceptibility to persistent HBV infection and hepatocellular carcinoma	Zhang	2013	Yes
23715497	APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry	Xuan	2013	Yes
25502777	Identification of genomic alterations in pancreatic cancer using array-based comparative genomic hybridization	Liang	2014	No
24577894	APOBEC3 deletion polymorphism is associated with epithelial ovarian cancer risk among Chinese women	Qi	2014	Yes
24728294	Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer	Nik-Zainal	2014	No
25411794	Natural polymorphisms in human APOBEC3H and HIV-1 Vif combine in primary T lymphocytes to affect viral G-to-A mutation levels and infectivity	Refsland	2014	No
25298230	A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage	Caval	2014	No
25816774	Nucleic acid recognition orchestrates the anti-viral response to retroviruses	Stavrou	2015	No
26494601	A method to avoid errors associated with the analysis of hypermutated viral sequences by alignment-based methods	Alinejad-Rokny	2015	No

PMID (PubMed)	Title	First Author	Publication Year	Included in the meta-analysis?
26682542	Integrative genomic analysis reveals functional diversification of APOBEC gene family in breast cancer	Zhang	2015	No
25401976	FHIT loss-induced DNA damage creates optimal APOBEC substrates: Insights into APOBEC-mediated mutagenesis	Waters	2015	No
27570633	Role of the host restriction factor APOBEC3 on papillomavirus evolution	Warren	2015	No
26159519	Acquired genetic alterations in tumor cells dictate the development of high-risk neuroblastoma and clinical outcomes	Khan	2015	No
25818029	APOBECs and virus restriction	Harris	2015	No
25820175	APOBEC3 genes: retroviral restriction factors to cancer drivers	Henderson	2015	No
25800874	Identification of APOBEC3B promoter elements responsible for activation by human papillomavirus type 16 E6	Mori	2015	No
25807502	APOBEC3A is implicated in a novel class of G-to-A mRNA editing in WT1 transcripts	Niavarani	2015	No
25730878	APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation	Cescon	2015	No
26459911	The eQTL-missense polymorphisms of APOBEC3H are associated with lung cancer risk in a Han Chinese population	Zhu	2015	No
26491161	A Naturally Occurring Domestic Cat APOBEC3 Variant Confers Resistance to Feline Immunodeficiency Virus Infection	Yoshikawa	2015	No
26305823	The APOBEC3B deletion polymorphism is associated with prevalence of hepatitis B virus, hepatitis C virus, Torque Teno virus, and Toxoplasma gondii co-infection among HIV-infected individuals	Prasetyo	2015	No
26394054	The Role of the Antiviral APOBEC3 Gene Family in Protecting Chimpanzees against Lentiviruses from Monkeys	Etienne	2015	No
26261799	APOBEC3 Deletion is Associated with Breast Cancer Risk in a Sample of Southeast Iranian Population	Rezaei	2015	Yes
27552096	The 29.5 kb APOBEC3B Deletion Polymorphism Is Not Associated with Clinical Outcome of Breast Cancer	Liu	2016	No
27977754	Expression of APOBEC3B mRNA in Primary Breast Cancer of Japanese Women	Tokunaga	2016	No

PMID (PubMed)	Title	First Author	Publication Year	Included in the meta-analysis?
26920143	Analysis of functional germline variants in APOBEC3 and driver genes on breast cancer risk in Moroccan study population	Marouf	2016	Yes
28199992	Fragile Genes That Are Frequently Altered in Cancer: Players Not Passengers	Karras	2016	No
26320772	Impact of functional germline variants and a deletion polymorphism in APOBEC3A and APOBEC3B on breast cancer risk and survival in a Swedish study population	Göhler	2016	Yes
27363431	The effect of HIV-1 Vif polymorphisms on A3G anti-viral activity in an in vivo mouse model	Cadena	2016	No
27163364	Mutation Processes in 293-Based Clones Overexpressing the DNA Cytosine Deaminase APOBEC3B	Akre	2016	No
27233495	Germline APOBEC3B deletion is associated with breast cancer risk in an Asian multi-ethnic cohort and with immune cell presentation	Wen	2016	Yes
26942578	Role of APOBEC3F Gene Variation in HIV-1 Disease Progression and Pneumocystis Pneumonia	An	2016	No
27643540	Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors	Middlebrooks	2016	Yes
27602762	APOBEC3 deletion increases the risk of breast cancer: a meta-analysis	Han	2016	No
27534815	Biochemical Characterization of APOBEC3H Variants: Implications for Their HIV-1 Restriction Activity and mC Modification	Gu	2016	No
27732658	A Single Nucleotide Polymorphism in Human APOBEC3C Enhances Restriction of Lentiviruses	Wittkopp	2016	No
27730215	The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer	Law	2016	No
27155849	Analysis of APOBEC3A/3B germline deletion polymorphism in breast, cervical and oral cancers from South India and its impact on miRNA regulation	Revathidevi	2016	Yes
26476745	APOBEC3B high expression status is associated with aggressive phenotype in Japanese breast cancers	Tsuboi	2016	No
27485054	Association between targeted somatic mutation (TSM) signatures and HGS-OvCa progression	Lindley	2016	No
28158858	Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization	Adolph	2017	No
28062980	Assessment of an APOBEC3B truncating mutation, c.783delG, in patients with breast cancer	Radmanesh	2017	No

PMID (PubMed)	Title	First Author	Publication Year	Included in the meta-analysis?
28127197	Association between polymorphisms of the APOBEC3G gene and chronic hepatitis B viral infection and hepatitis B virus-related hepatocellular carcinoma	He	2017	No
29100317	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	Klonowska	2017	Yes
28753428	Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes	Supek	2017	No
28575276	Structural determinants of APOBEC3B non-catalytic domain for molecular assembly and catalytic regulation	Xiao	2017	No
28475648	HIV-1 competition experiments in humanized mice show that APOBEC3H imposes selective pressure and promotes virus adaptation	Nakano	2017	No
29254236	The ubiquitous 'cancer mutational signature' 5 occurs specifically in cancers with deleted FHIT alleles	Volinia	2017	No
29116104	Germline copy number variations are associated with breast cancer risk and prognosis	Kumaran	2017	No
28878238	APOBEC3A is an oral cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion polymorphism	Chen	2017	No
29925657	Recurrent Loss of APOBEC3H Activity during Primate Evolution	Garcia	2018	No
30558640	Natural APOBEC3C variants can elicit differential HIV-1 restriction activity	Anderson	2018	No
29140415	APOBEC3A/B deletion polymorphism and cancer risk	Gansmo	2018	Yes
29025908	A germ-line deletion of APOBEC3B does not contribute to subtype-specific childhood acute lymphoblastic leukemia etiology	Wallace	2018	No
29957488	APOBEC and ADAR deaminases may cause many single nucleotide polymorphisms curated in the OMIM database	Lindley	2018	No
29783589	Re: Association of Germline Variants in the APOBEC3 Region with Cancer Risk and Enrichment with APOBEC-Signature Mutations in Tumors	Chang	2018	No
29769087	Replacement of feline foamy virus bet by feline immunodeficiency virus vif yields replicative virus with novel vaccine candidate potential	Ledesma-Feliciano	2018	No

PMID (PubMed)	Title	First Author	Publication Year	Included in the meta-analysis?
29746834	Characterization of BK Polyomaviruses from Kidney Transplant Recipients Suggests a Role for APOBEC3 in Driving In-Host Virus Evolution	Peretti	2018	No
31409681	A Protein Antagonist of Activation-Induced Cytidine Deaminase Encoded by a Complex Mouse Retrovirus	Singh	2019	No
30693645	Association of APOBEC3 deletion with cancer risk: A meta-analysis of 26 225 cases and 37 201 controls	Hashemi	2019	No
31856876	Immune gene expression profiling reveals heterogeneity in luminal breast tumors	Zhu	2019	No
30792902	Evolutionary effects of the AID/APOBEC family of mutagenic enzymes on human gamma-herpesviruses	Martinez	2019	No
30418757	Inhibiting APOBEC3 Activity with Single-Stranded DNA Containing 2'-Deoxyzebularine Analogues	Kvach	2019	No
30660178	Characterization of APOBEC3 variation in a population of HIV-1 infected individuals in northern South Africa	Matume	2019	No
30679582	A panel of eGFP reporters for single base editing by APOBEC-Cas9 editosome complexes	Martin	2019	No
30242938	Mechanisms shaping the mutational landscape of the FRA3B/FHIT-deficient cancer genome	Saldivar	2019	No
30647454	APOBEC-induced mutations and their cancer effect size in head and neck squamous cell carcinoma	Cannataro	2019	No
31073151	Endogenous APOBEC3B Overexpression Constitutively Generates DNA Substitutions and Deletions in Myeloma Cells	Yamazaki	2019	No
30578845	A lentivirus-based system for Cas9/gRNA expression and subsequent removal by Cre-mediated recombination	Carpenter	2019	No
30797859	APOBEC3 Host Restriction Factors of HIV-1 Can Change the Template Switching Frequency of Reverse Transcriptase	Adolph	2019	No
30840888	APOBEC Mutagenesis and Copy-Number Alterations Are Drivers of Proteogenomic Tumor Evolution and Heterogeneity in Metastatic Thoracic Tumors	Roper	2019	No
31603457	Selective inhibition of APOBEC3 enzymes by single-stranded DNAs containing 2'-deoxyzebularine	Barzak	2019	No
31400856	Different antiviral activities of natural APOBEC3C, APOBEC3G, and APOBEC3H variants against hepatitis B virus	Kanagaraj	2019	No

PMID (PubMed)	Title	First Author	Publication Year	Included in the meta-analysis?
31322199	Associations between the single nucleotide polymorphisms of APOBEC3A, APOBEC3B and APOBEC3H, and chronic hepatitis B progression and hepatocellular carcinoma in a Chinese population	He	2019	No
31422205	APOBEC mutagenesis is tightly linked to the immune landscape and immunotherapy biomarkers in head and neck squamous cell carcinoma	Faden	2019	No
31152021	Genetic Polymorphisms Predisposing the Interleukin 6-Induced APOBEC3B-UNG Imbalance Increase HCC Risk via Promoting the Generation of APOBEC-Signature HBV Mutations	Liu	2019	No
31533728	Integrative genomic analyses of APOBEC-mutational signature, expression and germline deletion of APOBEC3 genes, and immunogenicity in multiple cancer types	Chen	2019	No
32285256	Germline APOBEC3B deletion influences clinicopathological parameters in luminal-A breast cancer: evidences from a southern Brazilian cohort	Vitiello	2020	Yes
32176735	Clinical implications of APOBEC3A and 3B expression in patients with breast cancer	Kim	2020	No
32235597	Polymorphisms in Human APOBEC3H Differentially Regulate Ubiquitination and Antiviral Activity	Chesarino	2020	No

Supplementary Table 2. Summary of statistical results from association studies of the APOBEC3A_B deletion and cancer risk.

Odds ratios (OR) are presented for I/D vs. I/I (effect for one-copy deletion) and for D/D vs. I/I (effect for two-copy deletion). Studies, which did not include calculations for codominant gene models are marked as "no data". BC – breast cancer; BlC – bladder cancer; CC – cervical cancer; ColC – colon cancer; HCC – hepatocellular carcinoma; LC – lung cancer; OC – oral cancer; OvC – ovarian cancer; PC – prostate cancer.

Study, Year	Cancer type	I/D OR, (95%CI), *P*-value	D/D OR, (95%CI), *P*-value
Zhang, 2013	HCC	1.80 (1.48–2.19), p=1.35 × 10^{-11}	2.38 (1.73–3.29), p=1.35 × 10^{-11}
Xuan, 2013	BC	1.21 (1.02–1.43), p=0.0053	2.29 (1.04–5.06), p=0.0053
Long, 2013	BC	1.31 (1.21–1.42), p<0.005	1.77 (1.58–1.99), p<0.005
Qi, 2014	OvC	1.46 (1.13–1.87), p=0.0015	2.52 (0.93–6.87), p=0.0088
Rezaei, 2015	BC	1.57 (1.07–2.31), p=0.025	0.80 (0.24–2.66), p=0.767
Göhler, 2016	BC	1.17 (0.92–1.48), p=0.20	0.80 (0.28–2.24), p=0.67
Wen, 2016	BC	1.23 (1.05–1.44), p=0.005	1.38 (1.10–1.74), p=0.005
Revathidevi, 2016	BC	1.23 (0.84–1.81), p=0.8466	0.78 (0.37–1.64), p=0.8466
	CC	1.17 (0.69–1.99), p=0.9329	0.70 (0.24–2.11), p=0.9329
	OC	0.96 (0.57–1.60), p=0.1717	0.18 (0.02–1.34), p=0.1717
Marouf, 2016	BC	0.64 (0.34–1.21), p=0.1680	-
Middlebrooks, 2016	BlC (Eur)	no data	no data
	BlC (Asia)	no data	no data
Klonowska, 2017	BC	no data	no data
	OvC	no data	no data
Gansmo, 2018	BC	no data	no data
	LC	no data	no data
	ColC	no data	no data
	PC	no data	no data
Vitiello, 2020	BC	0.76 (0.51–1.12), p<0.05	0.49 (0.15–1.61), p<0.05

Supplementary Table 3. Summary of statistical results generated for association studies of the APOBEC3A_B deletion and cancer risk.
Odds ratios (OR) are presented for I/D vs. I/I (effect for one-copy deletion) and for D/D vs. I/I (effect for two-copy deletion). BC – breast cancer; BlC – bladder cancer; CC – cervical cancer; ColC – colon cancer; HCC – hepatocellular carcinoma; LC – lung cancer; OC – oral cancer; OvC – ovarian cancer; PC – prostate cancer.

Study, Year	Cancer type	I/D OR, (95%CI), *P*-value	*Z*-value	D/D OR, (95%CI), *P*-value	*Z*-value
Zhang, 2013	HCC	1.78 (1.47–2.16), p<0.0001	5.878	2.31 (1.69–3.17), p<0.0001	5.212
Xuan, 2013	BC	1.20 (1.01–1.42), p=0.0332	2.130	2.23 (1.01–4.91), p=0.0469	1.988
Long, 2013	BC	1.32 (1.22–1.42), p<0.0001	6.813	1.76 (1.57–1.98), p<0.0001	9.620
Qi, 2014	OvC	1.42 (1.19–1.69), p=0.0001	3.873	2.73 (1.29–5.79), p=0.0088	2.619
Rezaei, 2015	BC	1.57 (1.07–2.31), p=0.0218	2.293	0.80 (0.24–2.68), p=0.7186	0.360
Göhler, 2016	BC	1.16 (0.92–1.45), p=0.2063	1.264	0.79 (0.28–2.22), p=0.6501	0.454
Wen, 2016	BC	1.23 (1.05–1.44), p=0.0101	2.571	1.38 (1.10–1.74), p=0.0056	2.768
Revathidevi, 2016	BC	1.23 (0.84–1.81), p=0.2934	1.051	0.77 (0.37–1.63), p=0.5030	0.670
	CC	1.17 (0.69–1.99), p=0.5628	0.579	0.70 (0.23–2.11), p=0.5318	0.625
	OC	0.96 (0.57–1.60), p=0.8655	0.169	0.18 (0.02–1.34), p=0.0941	1.674
Marouf, 2016	BC	0.64 (0.34–1.21), p=0.1685	1.377	-	-
Middlebrooks, 2016	BlC (Eur)	0.76 (0.63–0.92), p=0.0054	2.779	1.93 (0.81–4.60), p=0.1357	1.492
	BlC (Asia)	0.86 (0.72–1.03), p=0.1085	1.605	0.94 (0.67–1.32), p=0.7263	0.350
Klonowska, 2017	BC	0.96 (0.82–1.13), p=0.6420	0.465	1.69 (0.69–4.14), p=0.2522	1.145
	OvC	0.71 (0.47–1.05), p=0.0890	1.701	2.35 (0.47–11.68), p=0.2973	1.042
Gansmo, 2018	BC	0.98 (0.82–1.16), p=0.7798	0.280	0.65 (0.31–1.33), p=0.2328	1.193
	LC	1.08 (0.92–1.27), p=0.3528	0.929	1.23 (0.72–2.12), p=0.4462	0.762
	ColC	0.94 (0.80–1.10), p=0.4286	0.792	0.70 (0.38–1.31), p=0.2679	1.108
	PC	0.91 (0.78–1.07), p=0.2654	1.114	0.82 (0.47–1.43), p=0.4913	0.688

Study, Year	**Cancer type**	**I/D OR, (95%CI), *P*-value**	***Z*-value**	**D/D OR, (95%CI), *P*-value**	***Z*-value**
Vitiello, 2020	BC	0.76 (0.51–1.12), p=0.1622	1.398	0.49 (0.15–1.60), p=0.2368	1.183
Overall	BC	1.21 (1.15–1.27), p<0.0001	7.585	1.57 (1.44–1.72), p<0.0001	9.882
Overall for Asia	BC	1.31 (1.22–1.40), p<0.0001	7.760	1.65 (1.49–1.83), p<0.0001	9.793
Overall for Europe	BC	1.06 (0.97–1.16), p=0.1641	1.392	1.17 (0.78–1.73), p=0.4491	0.757
Overall	OvC	1.42 (1.21–1.65), p<0.0001	4.415	2.99 (1.53–5.87), p=0.0014	3.191
Overall	BlC	1.03 (0.91–1.17), p=0.5953	0.531	1.51 (1.12–2.05), p=0.0071	2.690