University of Kent

Generating an isogenic model of the APOBEC3A_B deletion polymorphism using CRISPR/Cas9.

Supervisor: Dr Tim Fenton

A thesis submitted to the University of Kent for the degree of Master of Science in Genetics.

Rosalba Biondo

August, 2019

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.

Rosalba Biondo

August 2019

II Acknowledgements

My gratitude goes my supervisor, Dr Tim Fenton, who made this project possible and who patiently answered my endless questions. I would like to thank him for his support and encouragement while I was working with the flow cytometer to carry out my FACS experiments. My gratitude also goes to the other members of the lab, for also making this project possible and for their constant support and help throughout my experiments.

I would also like to thank Dr Emmanuela Cuomo, from Astra Zeneca, for kindly providing some of the cell line used in this project.

Additionally, I want to thank my closest friends Timo, Cindy, Kulveer, Anastasia for never giving up on me and always believing in me. Thank you for your support and for all the laughs. A special thanks goes to Laurens for the constant encouragement and all the food provided, and to Roberta and Adriana for always being there, despite living miles away.

Lastly, I want to thank my family for daily phone calls of encouragement and eternal support, I would have not been able to do this without them.

III Abstract

Deamination of cytosine, leading to C>T transitions, are one of the most common mutation leading to carcinogenesis. Recently, APOBEC3 (A3) enzymes have been accounted as the main cause. Under normal conditions, these deaminase enzymes have antiviral activities, which mutate viral genome preventing infections. Deregulation of these enzymes has been linked with C>T mutation in host genome, within the specific TpCpW (where W = A or T) nucleotide sequence, termed APOBEC-signature mutations. These are very common in several tumours, such as breast, cervical, head and neck and lung cancer. A3B has predominantly been associated with high APOBEC-signature mutations, especially in breast cancer. However, a naturally occurring ~29.5kb germline deletion polymorphism, spanning from intron 4 of A3A to intron 7 of A3B, generates the hybrid A3A_B, which causes the loss of almost the entire A3B. However, A3B-null tumours still present high APOBEC-signature mutations and have a high cancer risk. To investigate this paradox, we generated an isogenic model, using CRISPR/Cas9, which replicates the ~29.5kb deletion in a non-cancerous cell line. The genotyping assay created is very sensitive as it was possible to detect as low as 1% of cells with the deletion in a heterogeneous population. Furthermore, we analysed the use of the dual puro-ΔTK selection cassette in NIKS, to enrich for cells with the deletion and laid the ground for future studies. This is a novel and important system, which can help understand how A3B deletion might affect the other A3 enzymes and cell proliferation, what drives carcinogenesis in these cells and what are the impacts on treatment development and advances. Overall the results obtained from this work are important for further analysis on APOBEC3 and its implication in cancer formation, prompting further studies and analysis.

VI Contents

V List of Figures

VI Abbreviations

APOBEC	A3
APOBEC3A	A3A
APOBEC3B	A3B
APOBEC3A_B	A3A_B
HER2	Human epidermal growth factor receptor 2
ER	Estrogen receptor
PBS	Phosphate-Buffered Saline
FACS	Fluorescence-Activates Cell Sorting
sgRNA	single guide RNA
GCV	Ganciclovir
DMSO	Dimethyl sulfoxide
NIKS	Normal immortalized keratinocytes
IC_{50}	Drug concentration that inhibits growth of 50% of cells
IC_{90}	Drug concentration that inhibits growth of 90% of cells
FIAU	5-iodo-2’-fluoro-2’-deoxy-1-ß-D-arabino-furonosyluracil
TK	Thymidine kinase

1 Introduction

1.1 APOBEC gene family

The human apolipoprotein B mRNA-editing enzyme (APOBEC), catalytic polypeptide-like gene family comprises 11 genes that code for proteins with tissue-specific function and expression (Salter *et al.*, 2016). The gene family is thought to originate from the ancestral AID gene, and some members of the family share similar characteristics (Salter *et al.*, 2016) They bind to both RNA and single stranded DNA (ssDNA) and play important roles in immunity and carcinogenesis (Revathidevi *et al.*, 2016). The encoded proteins are: activation-induced cytidine deaminase (AICDA or AID); APOBEC1 (both AICDA and APOBEC1 genes are on chromosome 12), APOBEC2 (gene found on chromosome 6), seven APOBE3 proteins (genes found on chromosome 22) and APOBEC4 (gene located on chromosome 1) (Salter *et al.*, 2016). These proteins show cytidine deaminase activity at their catalytic site, except for APOBEC2 and APOBEC4, which are the least studied proteins and whose detailed functions have not been elucidated yet (Henderson and Fenton, 2015).

The first protein to be discovered and the family founder was APOBEC1, which gave the name to most of the proteins in the family, expect AID. APOBEC1 (also known as A1) carries out cytidine deamination at position 6666 in the mRNA of ApoB protein, in the small intestine of mammals where ApoB mediates the uptake of dietary lipids (Mehta *et al.*, 2000). The deamination converts cytidine to uridine (C →U), creating a stop codon and therefore a shorter version of ApoB (Mehta *et al.*, 2000). Additionally, A1 is also able to deaminate ssDNA (Salter *et al.*, 2016) (Henderson and Fenton, 2015). With a similar structure and function as A1, AID

carries out deoxycytidine (dC) deamination in ssDNA, which generates deoxyuridine (dU). This activity is performed by its activation-induced deaminase (AID) motif, responsible for somatic hypermutation and class switching recombination in B cells which generates antibodies variation with specific antigen affinity (Henderson and Fenton, 2015) (Muramatsu *et al.*, 2000). AID is an important enzyme that has influenced the immune system throughout evolution; it is found in bony and cartilaginous fish, the first organisms to have developed an adaptive immune system, and the enzyme has been conserved throughout evolution (Salter *et al.*, 2016). Additionally, dysfunctional AID enzyme gives rise to a type of autosomal recessive immunodeficiency disease, further highlighting its importance in immunity (Salter *et al.*, 2016) (Rebhandl *et al.*, 2015).

1.1.1 APOBEC3 subfamily

The APOBEC3 (A3) genes in humans, are located in tandem in a 100kb region on chromosome 22 and encode seven proteins: APOBEC3A (A3A), APOBEC3B (A3B), APOBEC3C (A3C), APOBEC3D (A3D), APOBEC3F (A3F), APOBEC3G (A3G) and APOBEC3H (A3H) (Swanton *et al.*, 2015). These are divided in two groups, based on the number of zinc-finger-containing catalytic sites they possess: A3A, A3C, A3H have a single deaminase domain, whereas A3B, A3D, A3G and A3F have two deaminase domains (Rebhandl *et al.*, 2015) (Swanton *et al.*, 2015).

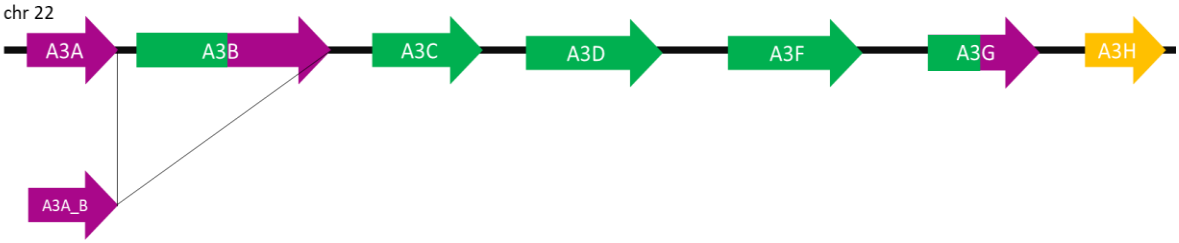


Figure 1. APOBEC3 genes on chromosome 22

Schematic representation of the APOBEC3 gene on chromosome 22. The A3A_B deletion polymorphism is also shown, caused as a result of A3B deletion and the fusion of A3A OFR to A3B 3' UTR (detailed explanation of the A3A_B deletion polymorphism in section 0).

A3 proteins localize in different cell compartments, enabling protection against nuclear or cytoplasmic replicating pathogens, and explaining their role in DNA modification and implications in human health and diseases. A3G and A3F are restricted to the cytoplasm (Lackey *et al.*, 2013; Henderson and Fenton, 2015) A3A is highly cytotoxic if overexpressed as it is found both in the cytoplasm and the nucleus and it is able to induce double strand DNA (dsDNA) breaks (Caval *et al.*, 2014; Suspène *et al.*, 2016; Gansmo *et al.*, 2018). A3B is constitutively found in the nucleus, due to specific nuclear localization sequence which allows its active transport in the nucleus after cell division (Bogerd *et al.*, 2006; Lackey *et al.*, 2012). Here, A3B is able to inhibit transcription of retrotransposons such as LINE-1 and Alu (Bogerd *et al.*, 2006). A3B nuclear localization and role, might be the result of conservation with ancestral AID, however the other functions of the two enzymes, diverted over time functions (Lackey *et al.*, 2012). A3H is polymorphic and four major haplotypes (I-IV) have been discovered in humans (OhAinle *et al.*, 2008; Harari *et al.*, 2009): haplotype I localises in the nucleus and is less active than haplotype II, which localizes in the cytoplasm and it's the most active form; haplotypes III and IV are inactive (Li and Emerman, 2011).

A3 enzymes carry out deamination in ssDNA, and are involved in the defence against retroviruses, retrotransposons and viruses such as HPV (human papilloma virus), HBV (hepatitis B virus) and HIV; for these reasons, APOBEC3 enzymes constitute an important part of the innate immune response (Burns *et al.*, 2015; Henderson and Fenton, 2015). Indeed, it was shown that A3G mediates HIV restriction, by becoming incorporated into HIV viral particles in the cytoplasm; A3G

causes deamination of HIV cDNA during reverse transcription, causing inactivation of the virus (Sheehy *et al.*, 2002; Koning *et al.*, 2009). However, the viral protein Vif overcomes the host antiviral activity, by depleting cellular levels of A3G in two ways: by preventing translation of A3G mRNA and by targeting A3G for proteasomal degradation (Mangeat *et al.*, 2003; Stopak *et al.*, 2003). A3B and A3F have also been shown to have antiviral activity against HIV, with a similar mechanism as A3G, however, A3B does not appear to be regulated by Vif action (Koning *et al.*, 2009; Vieira and Soares, 2013). A3A and A3B provide defence against viruses such as HPV and HBV (Vartanian *et al.*, 2008; Vartanian *et al.*, 2010a); indeed, A3A and A3B have been seen to be upregulated upon HPV infection and it is highly expressed in HPV-associated tumours (Henderson *et al.*, 2014; Smith and Fenton, 2019). A3A and A3B are also powerful inhibitors of retrotransposon elements, such as LINE-1 (Bogerd *et al.*, 2006); A3F and, to a certain extent, the other A3 enzymes have also been shown to stop retrotransposons activity (Muckenfuss *et al.*, 2006; Stenglein and Harris, 2006; Lovin and Peterlin, 2009; Pak *et al.*, 2011; Richardson *et al.*, 2014; Henderson and Fenton, 2015).

It is easy to understand that with nuclear localization of these highly active enzymes, there is the risk of off-target activities directed to the host genome, if regulation is lost. One possible detrimental consequence of regulation loss is the formation of genomic mutations which might lead to cancer development. This would suggest the importance of unravelling the role APOBE3 proteins in carcinogenesis, as well as understanding whether these proteins might be a target for novel therapeutics.

1.2 APOBEC3 and cancer

Cancer arises as a result of an accumulation of mutations in the DNA, coupled with the failure of repair mechanisms to resolve the lesion before the next round of DNA replication. Characterized by high levels of heterogeneity due to mutations that can arise at the start of the tumour, or that can evolve as the tumour progresses and eventually metastasises. Several mutational processes have been proposed to shape tumour progression, both from exogenous sources, such as tobacco smoke or UV light, to endogenous factors such as mismatch repairs or upregulation of certain enzymes (Alexandrov *et al.*, 2013) (McGranahan *et al.*, 2015a). The C-to-T transition is one of the most common mutation found in cancers, largely caused by UV light lesions within di-pyrimidine context, seen in melanomas and head and neck squamous carcinomas (Alexandrov *et al.*, 2013) (McGranahan *et al.*, 2015a). These can also occur as a spontaneous water-mediated mutation, in a CpG context. Interestingly, it has recently been found that APOBEC3 enzymes are predominantly responsible for the high volumes of C-to-T mutation in cancers, however this occurs in a different nucleoside sequence (Nik-Zainal *et al.*, 2012; Burns *et al.*, 2013a; Alexandrov *et al.*, 2013; McGranahan *et al.*, 2015a; Law *et al.*, 2016a). A3 are involved in mutations seen in cancer driver genes such as *PTEN*, *EGFR*, and *TP53* (Alexandrov *et al.*, 2013; McGranahan *et al.*, 2015a). Moreover, strikingly high APOBEC mutations are also see in the helical domain of the oncogene *PIK3CA* across several cancers (Henderson *et al.*, 2014; McGranahan *et al.*, 2015b). APOBEC3 enzymes are able to generate somatic hypermutations which frequently occur in clusters and are characterized by a specific pattern (TpCpW, where W = A or T); these are referred to as kataegis (from Greek for rain shower), originally identified in breast cancer with *BRCA* mutation (Nik-Zainal *et al.*, 2012), and later

also seen other cancers (Alexandrov *et al.*, 2013) . Upon ssDNA exposure, A3 convert cytosine bases (C) to uracil (U), resulting in C to T transition, if DNA replication occurs before repair has been competed; C to G transversions can also be seen, depending on the DNA polymerase involved in the translesion synthesis (Burns *et al.*, 2013c; Roberts *et al.*, 2013). Mutation clusters (kataegis) that occur on stretches of cytosines within the 5'-TCA-3' or 5'-TCT-3' motif have been attributed to A3 activity on ssDNA and are referred to as APOBEC-signature mutations, shown to be characteristic of certain cancers (Burns *et al.*, 2013c; Chan *et al.*, 2015).

1.2.1 APOBEC3 access to ssDNA

As previously mentioned, APOBEC3 target ssDNA, therefore, A3-mediated mutagenesis in cancer cells depends on substrate availability. Exposure of ssDNA can occur during resection of double strand break (DSB) repair, and A3-signature mutations are often seen in rearrangement breakpoints in tumours (Roberts *et al.*, 2012; Henderson and Fenton, 2015). ssDNA is also exposed at replication forks, especially if these are stalled or collapsed, a characteristic of tumour cells due to forced S-phase entry during cell cycle. Another source of long stretches of ssDNA is seen at transcription bubbles, where helicase unzips the DNA and exposes ssDNA (Henderson and Fenton, 2015). In fact, A3-mutations are present in highly transcribed gens in bladder cancer (Nordentoft *et al.*, 2014). Interestingly, recent studies show that A3-mutations are seen in the lagging strand during DNA replication, and these mutations are attributed to A3A and A3B (Hoopes *et al.*, 2016).

Analysis of subcellular localization of A3 enzymes, has showed that they are kept away from chromosomes during the different stages of mitosis, but they are shuttled in different cell compartment during the interphase stage of cell cycle (Lackey *et al.*, 2013). Their deaminase activity is not impaired during the different stages of cell cycle and A3 enzymes also have the capacity to arrest cell cycle. A3A, A3C and A3H gain access to DNA during interphase and telophase (Lackey *et al.*, 2013); A3B, A3D, A3F and A3G do not come in contact with DNA during the different phases of mitosis, however, during interphase, A3B is nuclear as it is transported in the nucleus after mitosis (Lackey *et al.*, 2012; Lackey *et al.*, 2013), while A3D, F and G are cytoplasmic (Lackey *et al.*, 2013). Despite A3 are excluded from mitotic chromosomes, it is possible that the nuclear envelope loses its integrity during interphase, altering genome stability and halting cellular processes such as DNA transcription and replication prematurely, causing DNA damage such as DSB. DNA damage seen as a result of nuclear envelope rupture, is associated with kataegis and chromothripsis (Hatch, 2018). These two detrimental phenomena work together. Chromothripsis is a not-well-characterized mutational process where regions of a chromosome undergo shattering; these chromosomal pieces are randomly rearranged together, originating several mutations (Maciejowski *et al.*, 2015). It is not surprising to see chromothipsis across multiple cancers. Chromothipsis can arise as a result of telomere crisis, where the ends of chromosomes are joined together, forming long chromatin bridges that link the two nuclei while a cell undergoes cell division (Maciejowski *et al.*, 2015). When the bridges are broken, long strands of ssDNA are formed and might be subjected to A3 deaminase action. In fact, kataegis attributed to the action of APOBEC3A/B are seen at chromothipsis break points (Maciejowski *et al.*, 2015). Rupture of nuclear

envelope and shuttling of A3 enzymes during the cell cycle, might enable several A3 enzymes to cause DNA mutations. However, they act on different nucleotides motifs, distinguishable from the APOBEC-signature mutation (Faltas *et al.*, 2016a).

A3B is thought to be the major source of mutations in several cancers; as mentioned already, it is the only A3 protein that constitutively localises in the nucleus and has remarkable high levels of mRNA expression, especially compared to A3A (Burns *et al.*, 2013c). Interestingly, the post-translation regulation of A3B activity and the molecular mechanism of its transport in the nucleus, remain unclear; however, it has recently been shown that A3B interacts with some of the proteins involved in the cell cycle. The study from McCann *et. Al*, showed that A3B interacts with CDK4 (cyclin-dependent kinase 4), preventing cyclin D1 import in the nucleus (McCann *et al.*, 2019) . The complex CDK4-Cyclin D1 is essential for the regulation of the G1/S phase transition of the cell cycle (Day *et al.*, 2009), and it is the most common check point to be altered in many cancers (Barretina *et al.*, 2010; Baker and Reddy, 2012). Enforced over-expression of A3B in cancer cells, causes cell cycle arrest at the G1/S phase, due to the A3B-CDK4 interaction, leading to increased exposure of ssDNA substrates for A3B. (McCann *et al.*, 2019). As a result, this could result in a surge of DNA deamination, mutations and high genomic instability.

1.3 Role of APOBEC3A and APOBEC3B in cancer development

The main A3 proteins implicated in cancer development and prognosis are A3A and A3B (Shinohara *et al.*, 2012; Caval *et al.*, 2014; Chan *et al.*, 2015).

A3B was first implicated in causing mutations in breast cancer, due to analysis of breast cancer cell lines which showed an upregulation of A3B (Burns *et al.*, 2013a). Both A3A and A3B can cause deamination, however, A3A is prevalently found in

myeloid cells and therefore less detectable in breast cancer cells. It was observed that *A3B* mRNA was increased in the majority of primary breast cancers and high expression of A3B is likely to cause mutations in *TP53*, leading to its inactivation (Burns *et al.*, 2013b). High levels of *A3B* mRNA, in primary breast tumour compared to the corresponding normal tissue, was matched to increased mutational load and poor disease prognosis in estrogen receptor positive (ER+) breast cancer, but not in ER- tumours, suggesting a link between oestrogen receptor and A3B (Sieuwerts *et al.*, 2014). It was also seen that patients with high levels of A3B in primary breast tumour, are more likely to have relapse, suggesting that A3 also contribute to disease progression and ongoing mutations (Sieuwerts *et al.*, 2014).

Studies which looked at range of different types of breast cancers, showed that human epidermal growth factor receptor 2 (HER2+) subtype had exceptionally high levels of A3B as well as signature-mutations, suggesting that A3 mutagenesis is linked with cancer development (Roberts *et al.*, 2013). Independent of HER and ER expression, A3B was also high in distant metastasis of breast cancer, especially in tissues from liver, brain, bones and lungs (Sieuwerts *et al.*, 2017a). This further suggests that A3B is not only a major contributor of primary cancer development, but it also shapes and dictates disease evolution and progression.

Subsequent analysis showed that bladder, cervix, lung adenocarcinoma, lung squamous cell carcinoma, head and neck, and breast, show an elevated A3B expression, as well as increased APOBEC-mutation signature and kataegis (Burns *et al.*, 2013c).

Given the role of A3 enzymes against viral infections, it is not surprising to see that A3B implication in HPV related cancers (HPV+), such as head and neck squamous cell carcinoma (HNSCC) and cervical cancers (Henderson *et al.*, 2014). Comparison

of HPV+ and HPV- HNSCC, showed that HPV- tumours are more likely to develop as a result of tobacco smoke, whereas mutations seen in HPV+ HNSCC are attributed to A3B mutations, and occur in non-smokers (Henderson *et al.*, 2014). A3B signature mutations are also seen in *PIK3CA* gene of some HPV- HNSCC, suggesting that A3B action can be seen in many cancers and its activity might be exacerbated by HPV presence (Henderson and Fenton, 2015).

Ovarian cancers present a lower expression of A3B compared to other types of cancer; however, it was shown that deamination activities due A3B action, have been detected in serous ovarian cancer, which is the most common and severe form of ovarian cancer (Leonard *et al.*, 2013). High A3B expression and its activity has also been linked with gastric cancers, which results in poor prognosis (Zhang *et al.*, 2015).

APOBEC-signature mutation affects the outcome of myelomas too. The cytogenetic myeloma subgroups t(14;16) and t(14;20), show A3A and A3B overexpression which corresponds in elevated expression of a maf transcription factor (Walker *et al.*, 2015). In particular this results in *MAF* upregulation in the t(14;20) subtype, and *MAFB* upregulation in the t(14;20), both linked with poor prognosis. It is not clear how the APOBEC-signature mutations arise in these myeloma subgroups, however it was seen that loci translocation between the chromosomes involved, exposes ssDNA so APOBEC3 enzyme can use it as a substrate (Walker *et al.*, 2015). APOBEC-signature mutations were also seen in both genes when translocation occurred between *MYC* and *IG loci,* responsible for another type of myeloma (Walker *et al.*, 2015). Interestingly, the study by *Walker et al.* did not distinguish if the mutations were caused by A3A or A3B. Urothelial carcinoma also present

APOBE3-signature mutations, seen both in early and late disease stages. These were attributed to A3B, as A3A expression in urothelial tissue and cell lines is very low, although this remains controversial (Vasudevan *et al.*, 2018).

A3B impact on cancer formation, evolution and responsiveness to treatment is astonishing. Therefore, it is not surprising to see that the number of studies carried out in the last decade have portrayed A3B as the principal source of mutations. However, a deeper analysis of these studies reveals no clear discrimination between A3B and A3A (or other A3). Therefore, other members of the A3 family, such as A3A, may also play an important role in tumorigenesis, making the story more complex (Starrett *et al.*, 2016).

1.3.1 The controversial role of APOBEC3A in carcinogenesis

Alongside A3B, A3A has also been thought to contribute to some of the signature-mutations seen in cancers. Interestingly, opposing to what was originally thought, it has been shown that A3A is more potent and stable than A3B, and it is considered to be the predominant deaminase that causes mutations in cancers. This finding was very surprising, considering the significantly higher mRNA level of *A3B* in cancer cells, compared to A3A (Caval *et al.*, 2014; Chan *et al.*, 2015; Kanu *et al.*, 2016) Due to more than 90% nucleotide similarities between A3A and A3B genes (Burns *et al.*, 2015), it has been difficult to finely determine their protein expression as not specific enough antibodies are yet available (Starrett *et al.*, 2016). However, by using a yeast model, *Chan et al. (2015)* were able to show that A3A and A3B signature mutations are statistically distinguishable (Chan *et al.*, 2015). The study showed that A3A and A3B prefer different motifs within the ssDNA: A3A favoured

YTCA and A3B favoured RTCA, where Y is a pyrimidine and R is a purine. Although the study was conducted in yeast, the A3A and A3B motif preference in yeast can be a suitable model for deamination in human cancers (Chan *et al.*, 2015). The assumption was made because their results were in accordance with the data generated from *Taylor et al. (2013).* A comparison between cancer kataegis and yeast kataegis triggered by A3A and A3B, shows that the flanking regions of the target cytosine and the length of the kataegic stretch are similar for both species. However, cancer kataegis showed two to fivefold higher mutation density, indicating that there may be differences in deaminase activity between the two organisms (Law *et al.*, 2016b). The study by *Chan et al. (2015)* then compared A3A-mutations vs A3B-mutations in several different cancer samples. The results indicated that at least 6 cancer types, such as bladder (BLCA),breast (BRCA), head and neck (HNSC), lung adenocarcinoma (LUAD), lung squamous carcinoma (LUSC) and cervical (CESC) cancer had higher A3A-mutations, supporting A3A to be a more efficient mutator than A3B (Chan *et al.*, 2015). The different signature mutations of A3A and A3B were exploited in a study on oral cancer in a Taiwanese population; the results pointed A3A to cause the most APOBEC-associated mutations, and that higher A3A expression is seen in cancerous tissues, compared to neighbouring corresponding healthy tissues, due to different 3'UTR regulation and greater interferon signalling (Chen *et al.*, 2017). A3A was also seen to be the dominant A3 to cause mutations both in nuclear DNA, especially in *TP53* and *MYC* genes, and in mitochondrial DNA (Suspene *et al.*, 2011).

However, due to the similarities between the preferred motifs and the protein sequences, it is difficult to distinguish differences between A3A and A3B and which one is more involved in mutagenesis and cancer development.

Therefore, an intricate interplay between A3A and A3B might be the result of kataegis seen in the genome, however, this area warrants further studies.

1.3.2 APOBEC3 activity and cancer therapy

Adding to the tumorigenic role of A3, there is evidence showing that A3B can also contribute to drug resistance by continuous generation of mutations throughout the treatment. This is shown by the recurrence of cancers, which occurs when the patient is undergoing therapy with curative intent (Davies *et al.*, 2013; Law *et al.*, 2016a) Although difficult to discern whether a mutation was present in a minor subclone if the primary tumour, or it was a result of ongoing genomic changes, it has been shown that A3B contributes to accumulation of mutations as well as drug resistance in oestrogen receptor (ER^+) breast tumours (Law *et al.*, 2016a). The study by *Law et al. (2017)* showed that in cells derived from primary ER^+ breast tumour that never received hormonal treatment, high levels of *A3B* mRNA decreases benefits of tamoxifen treatment in metastatic ER^+ breast tumours. Additionally, knockdown of A3B in ER^+ breast cancer cells showed tamoxifen response in murine xenograft. Conversely, A3B overexpression caused tamoxifen resistance in murine xenografts. The study shows that high *A3B* results in tamoxifen resistance in ER+ breast cancer, although no clear evidence linking A3B-generated mutations to the resistance was presented (Law *et al.*, 2016a).

The interplay between A3 and cytotoxic drugs to treat tumours, is rather interesting. A3B has been shown to contribute to on-going mutations throughout the cancer lifetime; cytotoxic drugs, used to treat cancer, increase DNA replication stress and genomic instability, by creating DNA double strand breaks and therefore increasing

ssDNA, a substrate for A3B. This can create mutations in oncogenes and/or tumour suppressor genes, further feeding into cancer therapy resistance, in a feedback loop.

Experimental models have shown that treating breast cancer cells and normal cell with drugs such as hydroxyurea, aphidicolin, gemcitabine and camptothecin, cause DNA replication stress (Kanu *et al.*, 2016). It is speculated that this leads to an upregulation of the ATR/Chk1 pathway and the subsequent increased activity of A3B, responsible for cytidine-deamination, typical of APOBEC3-signature mutations (Kanu *et al.*, 2016).

Additionally, analysis of metastatic urothelial carcinoma, relapsed following treatment with a combination of platinum-based drugs and gemcitabine, also showed an increased enrichment for APOBEC3-signature mutations (Faltas *et al.*, 2016b). Both A3A and A3B-related mutations seemed to be enhanced in the post-chemotherapy tumour; interestingly, however, the study indicated A3A as is the main responsible, due to higher mutations in the A3A-preferred motif (YTCA) (Faltas *et al.*, 2016b). Moreover it was shown that mutations in chemotherapy-treated cancers significantly diverge from primary, chemotherapy-naïve cancers, attributing the clonal tumour evolution to APOBEC-mutations, which occur throughout tumour development as well as during therapy (Faltas *et al.*, 2016a). Furthermore, treatment of some bladder and breast cancer cell lines with the drug bleomycin (which causes DNA damage and induces DNA breaks), showed an increased induction of both A3A and A3B, with A3B levels in bladder cancer cell lines (Middlebrooks *et al.*, 2016)

Although a clear distinction between A3A and A3B activity in cancer progression has not been made yet, it is certain that these enzymes do play a role. Developing

A3B (and/or A3A) inhibitors, to be administered together with known chemotherapies, might reduce APOBEC-signature mutations, decrease tumour drug resistance and increases survival. This seemingly straightforward approach still requires more research to further our understanding of the APOBEC3 enzymes and their specific role in carcinogenesis.

1.4 APOBEC3B deletion polymorphism

In addition to being the cause of signature mutations in certain cancers, A3B also presents an interesting phenotype. This resulted in extensive research to unravel the intricate details of its behaviour. High levels of A3B might be beneficial to fight off viral infections, however, it can also cause cytotoxicity and increase signature mutations (Starrett *et al.*, 2016). Interestingly, it has been shown that a naturally occurring deletion of A3B, present in a large proportion of the population, appears to result in high levels of signature mutations which might be a contributor to elevated cancer risk (Caval *et al.*, 2014) (Kidd *et al.*, 2007) (Chan *et al.*, 2015) (Middlebrooks *et al.*, 2016).

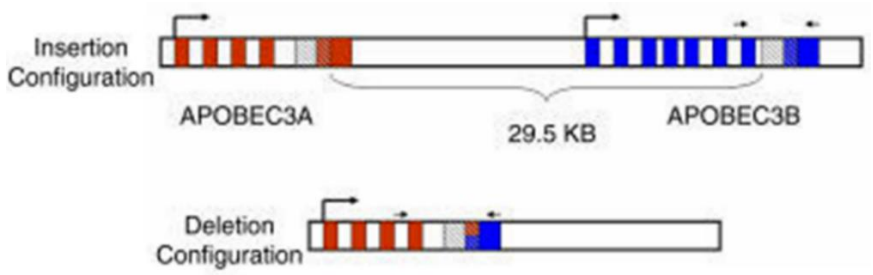


Figure 2. Insertion and deletion configuration of A3A and A3B

The insertion configuration corresponds to the wild type human genome. The ~29.5 deletion spanning from intron 4 of A3A (in red) and intron 7 of A3B is shown, which forms a shorrter, hybrid A3A_B gene. The black arrows on top of the genes, indicate PCR primers. The primers above APOBEC3B in the insertion configuration are referred to as the insertion primers; the primers on top of the deletion configuration are referred to as deletion primers. *Image adapted from Kidd et al* (Kidd *et al.*, 2007).

A naturally occurring ~29.5kb deletion between the 5th exons of A3A and the 8th exon of A3B, results in the formation of a hybrid gene (A3A_B, also referred to as hybrid A3A), where the coding sequence and the regulatory elements are identical to A3A, but the 3' UTR is derived from the A3B gene. As a result, there is a complete loss of the coding region of A3B gene (Kidd *et al.*, 2007). The hybrid gene behaves virtually identically to A3A, which is much more potent and more stable than A3B. In fact, despite the higher cellular level of A3B, A3A mutations are much more prominent than A3B (Caval *et al.*, 2014) (Chan *et al.*, 2015) . Moreover, when the deletion is present and the A3A_B gene is formed, a higher level of A3A signature-mutations are seen, which could indicate that the A3A_B gene presents an enhanced stability compared to both A3A and A3B (Klonowska *et al.*, 2017). One possible explanation for this increased stability, could be the loss of regulatory miRNAs which mediated repression of the A3A 3' UTR. This is consistent with the observation in certain cervical cancers, where the downregulation of miR-34b-3p (one of the miRNAs downregulated in many tumours) corresponds to high expression levels of A3A_B gene and increased somatic mutations. Hence, it is clear that a comprehensive analysis of the miRNAs, may help to further understand A3A_B regulation and its implications in carcinogenesis (Revathidevi *et al.*, 2016). Increased hybrid A3A expression was also seen in oral cancers in a Taiwanese population with A3B homozygous deletion, compared to individuals with heterozygous deletion and individuals with A3B present (Chen *et al.*, 2017). The same study, also indicated that miRNA downregulation might be involved in the increased stability of hybrid A3A; they found that miR-409, which decreases A3B 3'UTR activity but not that A3A 3'UTR, was indeed downregulated in oral cavity cancer (Chen *et al.*, 2017).

1.5 APOBEC3B deletion frequency, population variation and cancer risk

The signature mutations were firstly identified in breast cancer screens; interestingly, the A3B deletion polymorphism is also associated with such mutations and the C-to-T and C-to-G mutations are present despite the A3B deletion. The signature mutations were later also found in other cancer types, suggesting that A3s action was not only restricted to breast tissues (Nik-Zainal *et al.*, 2014). However, several studies show different association between the deletion polymorphism and cancer risk. Suggestion of A3B deletion and increased breast cancer risk, was showed in a Chinese population (Long *et al.*, 2013); the same association was showed in small studies on Iranian (Rezaei *et al.*, 2015), Malaysian (Wen *et al.*, 2016) and European-American (Xuan *et al.*, 2013) populations. However, association was not seen in studies done in Swedish (Göhler *et al.*, 2016), Indian (Revathidevi *et al.*, 2016) and Moroccan (Marouf *et al.*, 2016) populations. A study conducted on Norwegian population did not show association with A3B deletion and increased risk of lung, prostate, breast and colon cancer (Gansmo *et al.*, 2018). However, a stratification of the population by age showed an increased risk of lung cancer in individuals younger than 50 years of age, which progressively decreased with age (Gansmo *et al.*, 2018). A similar pattern was seen for prostate cancer, however, for both cancer types diagnosis occurred at a young age. The results obtained for breast cancer are in agreement with the Swedish study (Göhler *et al.*, 2016) previously mentioned, showing no association for A3B deletion and increased cancer risk (Gansmo *et al.*, 2018). Furthermore, a study on a South Indian population showed no association between breast, cervical and oral cancer and A3B deletion. Interestingly, the same study indicated a higher incidence of the deletion

polymorphism in women than in men both in oral cancer samples and controls, however, it is unclear as why that might be the case (Revathidevi *et al.*, 2016). The discrepancies in these findings can be explained by the varying frequencies of A3B deletion polymorphism across the global population. It is found in ~6% of Europeans, ~37% of East Asians, ~58% of Amerindians and a very high fixed frequency of ~93% in Oceanic population (Kidd *et al.*, 2007) (Xuan *et al.*, 2013). The association studies, indicate conflicting results about the effects of A3B deletion on Caucasian population, mirrored by the frequency distribution of the deletion.

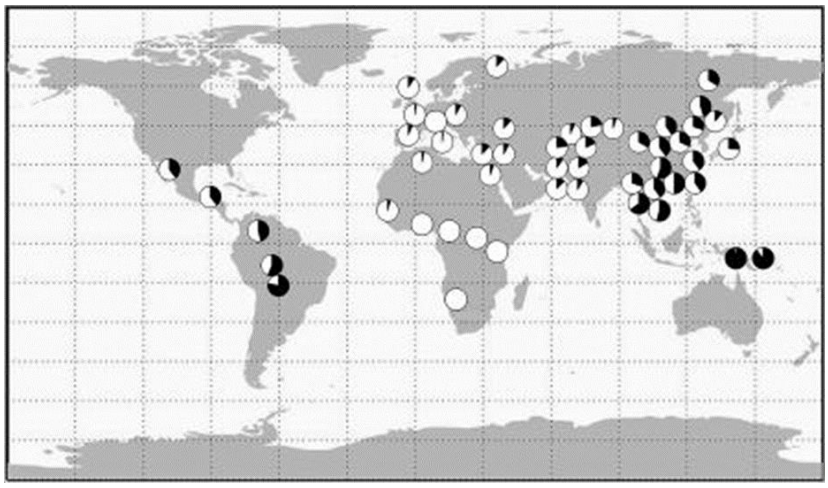


Figure 3. Distribution of the A3A_B deletion frequency.

The map shows the deletion A3A_3B frequency worldwide. The black portion of the pie charts indicate the deletion, the white sections indicate insertion frequency. (Kidd *et al.*, 2007)

Thinking of A3B deletion as another factor that might increase cancer risk is very simplistic; the different ethnic background as well as the expression of A3B in different tissues must be taken into consideration when making such a claim. More lager scale studies to gain a better understanding of the deletion polymorphism and why it is so different across worldwide populations, are essential to further elucidate APOBEC3 role in carcinogenesis and therapeutics. Additionally, the role of this interesting phenotype needs to be evaluated in light of the specific genetic

background of specific individuals, as A3A and A3B are expressed differently, depending on the presence of A3B deletion (Chen *et al.*, 2017).

1.6 Possible causes for A3B deletion frequency and further polymorphism in the APOBE3 family

A3B plays an important role in innate immunity; it has been implicated to restrict HBV replication in hepatocytes and high expression was seen in the lungs and the spleen. It has been noted that A3B insertion is constitutively present in African populations, where malaria is endemic (Kidd *et al.*, 2007), and that there is an increased expression of the relative APOBEC1 gene in a mouse spleen infected with the parasite *P. berghei*, during early infection stages. This lead to investigate whether there is an association between A3B and malaria susceptibility in humans (Jha *et al.*, 2012), despite mice having only one A3 gene, and humans express seven. In a study on Indian populations, the deletion allele was associated with a higher probability to get infected by falciparum malaria and a low deletion frequency was seen in individuals living in endemic regions of India (Jha *et al.*, 2012). Thus, it seems safe to accept that A3B is under selective pressure and its worldwide distribution mirrors the spread of certain African pathogens such as malaria (Jha *et al.*, 2012). This, indeed, might provide an explanation for the frequency of the A3B deletion worldwide and makes it easy to understand: as populations moved out of Africa, and migrated towards Europe and then Asia, the protective action of A3B against the pathogen was no longer needed. Additionally, it has been speculated that HIV infection might also play a role in the distribution of the deletion, together with malaria (Jha *et al.*, 2012). However, population's movement might have not been the only reason influencing A3B polymorphism; other A3 proteins also show an intricate polymorphism and it has recently been hypothesized that they might

also be involved in increased cancer risk (Klonowska *et al.*, 2017). A3H is the other polymorphic gene of the APOBEC3 family, presenting several haplotypes (I-IV) with variable stability (Li and Emerman, 2011), and this has recently been exploited to explain the A3B-deletion paradox (Starrett *et al.*, 2016). There is an association between A3B deletion and increased A3H-I expression, which might be responsible for the signature mutations seen in tumours from patients lacking A3B. Despite A3H-I expression is seen almost exclusively when there is A3B deletion, further studies need to be carried out to understand the actual contribution of A3H, A3B and/or other members of the family to the mutational load observed in breast cancer, lung adenocarcinoma and other cancers (Starrett *et al.*, 2016).

1.7 Summary

A3A and A3B are important antiviral proteins, however, high cellular expression can be dangerous as it can cause mutations in the genome, leading to cancer formation. Currently, A3A and A3B mutational signatures are implicated in specific cancers, however, further studies are needed in order to make clear links and correlations. In addition, the peculiar A3B deletion phenotype adds complexity to the picture: a deletion of a highly mutagenic gene predicts a reduction in mutations and cancer incidence, however, A3B deletion and consequent A3A_B hybrid gene formation, increases cancer risk, only in specific populations.

Additionally, it has been seen that patients with A3B deletion phenotype might be more responsive to immunotherapy (Cescon *et al.*, 2015), however, further research is needed to confirm this. It was recently shown that A3A might be an important prognostic marker, albeit, only when the A3B deletion polymorphism is present. This was seen in a small Taiwanese population with oral cancer, the study can prompt

further genomic studies on different cancers nonetheless (Chen *et al.*, 2017). Further studies are hence needed to understand A3A_B gene functions within the cell and how this might affect cell cycle and proliferation, as well as how this phenotype might affect viral infection.

A3A and A3B have received most of the attention in past studies, however, it might be possible that other members of the A3 family are involved in tumorigenesis.
Due to the high mutational load caused by APOBEC3, developing inhibitors to use in conjunction with or as a substitution of chemotherapy seems a logical assumption to make. Most research has pointed to towards the development of A3B inhibitors, due to exceptionally high levels of the enzyme in cells and the numerous evidence that appoints A3B as the main mutator of genomic DNA. This idea is also supported by the germline deletion of A3B in certain patients. Interestingly, this coincides with a higher risk of tumour development, therefore it is important to understand the interplay between A3 enzymes and how this can affect disease progression, prognosis and treatment.

1.8 Project outline

The current knowledge on A3A, A3B and A3A_B, is derived from overexpressing recombinant cDNA encoding the hybrid A3A and A3B genes in plasmids and expressed in *E.coli* or yeast. Additionally, A3B knock downs have been realised using siRNA in tumour cell lines, characterized by genomic instability where other mutations might have already occurred, and which are prone to further mutations. The extent to which A3A and A3B are implicated in mutagenesis in cancer is therefore still open to debate, augmented by the fact that individuals with the A3A_B genotype also present APOBEC3-signature mutations. An isogenic model for

A3A_B does not currently exist, hence, creating an A3B-knockout cell line will be vital to help address the historical question of whether A3B is essential for mutations, and if other A3 enzymes are also involved in the process. Additionally, this might also help to understand to what extent each A3 enzyme might contribute to mutations and how they might interact with each other.

This project aims generate a model that replicates the germline APOBEC3B deletion polymorphism. This naturally occurring -29.5kb deletion, is successfully generated in NIKS cell line using CRISPR/Cas9. This cell line, derived from epithelial cells, is used as a representative of the healthy tissue from which several squamous cell carcinomas, such as breast cervical, lung, head and neck originate and present APOBEC-signature mutations. Additionally, the MCF10A cell line was used as a representative of breast epithelial, due to the high APOBEC-signature mutations seen in breast cancers.

2 Materials and methods

2.1 Tissue culture

2.1.1 Normal Immortalized Keratinocytes (NIKS)

NIKS were cultured using medium composed of Ham's F-12 medium (Invitrogen 11765-054) and Dulbecco's modified Eagle's medium (DMEM) with 1% L-Glutamine (Invitrogen 11965-084), at a ratio of 3:1 respectively, supplemented with 5% Foetal Bovine Serum (FBS) (Pan Biotech. Cat. No.P30-3031). The medium was complemented with 0.4 µg/mL Hydrocortisone (Sigma H0888-1G), 8.3 ng/mL Cholera Toxin (EMD Biosciences 227035), 24 µg/mL Adenine (Sigma A2786-5G),

10 ng/ml Epidermal Growth Factor (EGF) (Sigma E1257-0.1MG), 5 µg/mL Insulin (Sigma I6634-50MG) and 1% penicillin/streptomycin (10000 U/ml, 10 mg/ml. Cat. No. P06-07100). The complete medium is referred as FC medium. Cells were cultured on a layer of feeder cells (mouse 3T3 fibroblasts) treated with mitomycin-C. NIKS were incubated in humidified atmosphere at 37°C and 5% CO2; media was changed every 2-3 days. NIKS were grown both in T25 and T75 flasks, depending on the downstream use of the cells. For transfection, cells were grown in T75 flasks.

2.1.2 TK-NIKS cell line

NIKS were previously modified, to express the *puro∆TK* selection cassette, as shown in **Figure 4**. The cassette is formed by a combination of puromycin N-acetyltransferase (puro), which confers puromycin resistance to the cells, and a truncated version of the herpes simplex virus type 1 thymidine kinase (∆TK or HSV-TK), as shown by *Chen et al.* (Chen and Bradley, 2000). Natural substrate of HSV-TK is thymidine, however, it also phosphorylates nucleosides analogues such as FIAU and ganciclovir (GCV), which stop DNA elongation, leading to cell death (Borrelli *et al.*, 1988; Chen and Bradley, 2000). Cells that express the selection cassette are negatively selected when treated with GCV, causing cell death. The selection cassette was inserted between exon 1 and 2 of the *APOBEC3B* gene, generating the modified TK-NIKS cell line, sensitive to GCV, with no effects on the expression of the *A3B* gene.

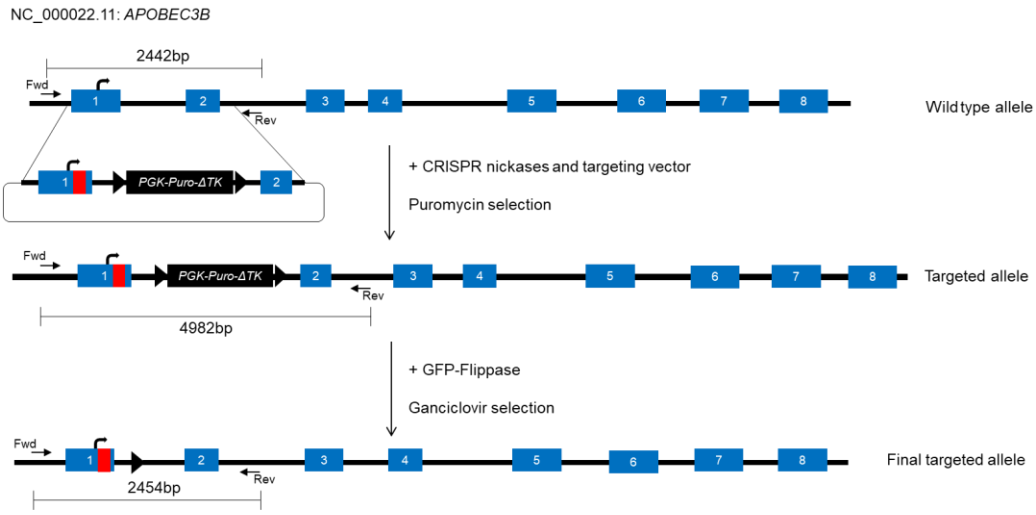


Figure 4. Gene construct for puroΔTK selection cassette insertion in A3B

puroΔTK selection cassette was inserted between exon 1 and 2 of APOBEC3B gene (intron 1). Cells that had a successful insertion of the cassette are resistant to puromycin, therefore these can be isolated from the heterogeneous population present after transfection via puromycin treatment. The presence of the inserted gene can also be showed by PCR, which should generate a product of approximately 4982-bp (shown in the *targeted allele* construct). The puroΔTK construct can be eliminated by using flippase enzymes; this generates a PCR product of approximately 2454-bp (shown in final *targeted allele* construct), which is similar the PCR product generated by the wild type allele. Additionally, cells which had the puroΔTK selection cassette removed are resistant to ganciclovir, therefore GCV-treatment eliminates cells in which flippase action was not successful.

TK-NIKS cells, were cultured the same way as NIKS (see section 2.1.1); same conditions as NIKS, also applied for TK-NIKS cell sorting.

2.1.3 3T3 mouse fibroblasts

Feeder cells were cultured in DMEM medium, supplemented with 10% FBS and 1% penicillin/streptomycin (10000U/ml and 10 mg/ml respectively), and incubated in humidified atmosphere at 37°C and 5% CO_2. When they reach 80-90% confluency, feeder cells were treated with mitomycin-C (Sigma - M4287-2MG) which arrests their growth. 2% of 50X mitomycin-C was diluted in DMEM and then added to the growing feeder cells, which were incubated 1-4 hours in humidified atmosphere at 37°C and 5% CO_2. Treated 3T3 feeder cells can be used in co-culture with NIKS or stored at -80°C, for later use. NIKS and growth arrested feeder cells are plated at a ratio of 3:1, respectively.

2.1.4 SKBR3

SKBR3, a breast cancer cell line homozygous for the A3A_B deletion allele, was cultured in DMEM medium, supplemented with 10% FBS and 1% penicillin/streptomycin (10000U/ml and 10 mg/ml respectively), and incubated in humidified atmosphere at 37°C and 5% CO_2. At 80-90% confluency, cells we were harvested for DNA extraction which was used as a positive control for deletion primers, in all the genotyping experiments.

2.1.5 MCF10A

MCF10A Odin Cas9 High, a stable cell line with tetracycline-inducible Cas9 expression was supplied by Dr Emmanuela Cuomo (Astra Zeneca). These cells were adapted to new media and cultured using 1:1 Ham's F-f12 medium and Dulbecco's modified Eagle's medium (DMEM) with 1% L-Glutamine (Pan Biotech. Cat. No. P04-41250), supplemented with 5% Horse serum. The medium was complemented with 0.4 µg/mL Hydrocortisone (Sigma H0888-1G), 8.3 ng/mL Cholera Toxin (EMD Biosciences 227035), 24 µg/mL Adenine (Sigma A2786-5G), 10 ng/ml Epidermal Growth Factor (EGF) (Sigma E1257-0.1MG), 5 µg/mL Insulin (Sigma I6634-50MG) and 1% penicillin/streptomycin (10000 U/ml, 10 mg/ml. Cat. No. P06-07100). Media was changed every 2-3 days.

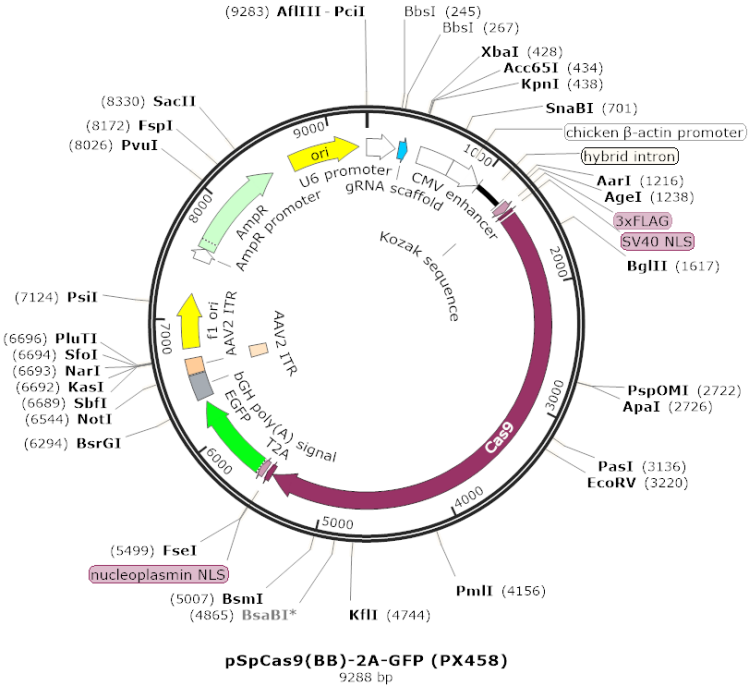
MCF10A cells were grown in T25 flasks and passaged every 5 days.

For transfection, MCF10A cells were plated in 6 well plates.

2.2 Bacteria culture

Previously transformed *E.coli* cells were streaked on LB agar plates, in the presence of ampicillin (stock concentration 100 mg/ml, 1:1000 dilution to obtain a final concentration of 100 µg/ml in the plates). Four different plates were streaked, each

with bacteria containing one of 4 different plasmids. Three of these encoded Cas9 protein and sgRNA targeting intron 4 of APOBEC3A and intron 7 of APOBEC3B at different locations in proximity. The sgRNA were classified as sgRNA1, sgRNA2 and sgRNA3 (for sgRNA sequences see **Table 1**). Additionally, an empty plasmid without sgRNA, was used as a negative control. All plasmids encoded GFP protein, which allows to determine transfection efficiency via a fluorescent microscope and allows for cell sorting (as elucidated later. A diagram of the plasmid construct and of the sgRNA targets on A3A and A3B is shown in **Figure 5** and **Figure 6**, respectively. The sgRNA oligonucleotides were previously cloned in a pX458 plasmid, following the protocol from *Zhang at al.* (Cong and Zhang, 2015). The agar plates were incubated overnight at 37◦C. The following day, a single colony from each plate was taken to carry out an overnight culture, in 250 ml of LB broth containing 100 µg/ml of ampicillin (as explained above). The bacterial cells were harvested the next day, for plasmid DNA extraction.



pSpCas9(BB)-2A-GFP (PX458)
9288 bp

Figure 5. pX458 plasmid construct to express CRISPR/Cas9

The plasmid contains the gRNA scaffold, which encodes for the specific sgRNA designed, Cas9 protein and EGFP. Additionally, the ampicillin resistant gene (AmpR) is also encoded by the plasmid. These are all under the same promoter, so they all get synthesized upon transfection into cells. Image obtained from SnapGene.

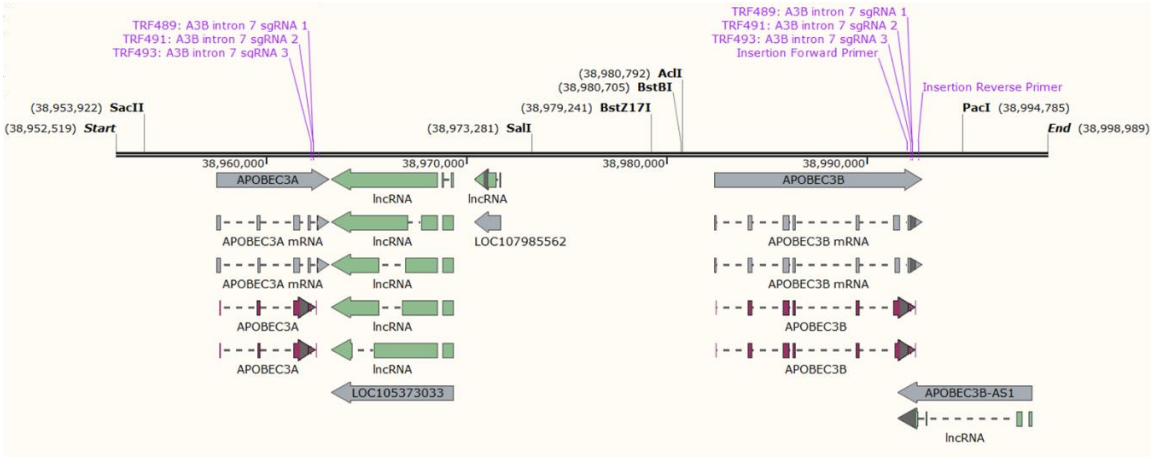


Figure 6. Schematic representation of A3A and A3B genes and sgRNA cut points

Representation of the sgRNA cut points in A3A and A3B. Gene sequence of A3A and A3B is almost identical, including the regions of interest, intron 4 of A3A and intron 7 of A3B. This means that the same sgRNA can be used to target both introns, as shown in the picture above. Intron 4 and intron 7 are found where the three sgRNA cluster; sgRNAs have been designed to perfectly match the targeted sequence. The PCR insertion primers are also shown on the gene map. Because only the wild type version (also called insertion configuration) of the genes A3A and A3B is shown here, it was not possible to add the PCR deletion primers. These, however, are shown in ***Figure 2***. When deletion occurs, the binding site for the insertion primers on A3B is no longer available and the deletion configuration, allows binding of the deletion primers. size of extected PCR product from wild type and deletion configuration are listed in **Error! Reference source not found.** Image obtained from SnapGene.

2.2.1 Plasmid DNA extraction

The plasmid containing the CRISPR/Cas9 system was extracted from bacteria cells following the QIAGEN Plasmid Plus Midi Kit (Cat. No. 12943) protocol.

2.3 NIKS Cell transfection

CRISPR/Cas9 containing plasmids were delivered to wild type NIKS and TK- NIKS. Cells were transfected according to the Technical Manual Fugene HD, following the instruction for T75 flasks. A 3:1 ratio Transfection Reagent:DNA was used for the experiments, which required 18 µg DNA, mixed with Fugene HD (Promega. Cat No. E2311) and Opti-MEM I. This formed the transfection complex that was then added

to the cells, which were incubated for 48 hours in humidified atmosphere at 37°C and 5% CO_2. Transfection was carried out when cells were at 80% confluency. Pictures of cells were taken 48h post transfection, using a Lumascope with 100x magnification.

2.4 MCF10A transfection

In order to allow the expression of Cas9, MCF10A cells were treated with different concentration of doxycycline (Doxycycline Hydrochloride ready-made solution, Sigma-Aldrich Cat. No. D3070). The concentration used were 0 ng/mL, 10 ng/mL, 50 ng/mL, 100 ng/mL, 250 ng/mL, 500 ng/mL and 100ng/mL. Doxycycline was added to the medium 2 days before transfection and was kept till two days post transfection, when the cells were harvested for analysis. MCF10A were transfected as described by the Dharmacon Edit-R CRISPR-Cas9 gene engineering with lentiviral Cas9 and synthetic guide RNAs Technical Manual, following the protocol for Transfection of synthetic guide RNA. The protocol was adapted to this specific experiment, as the transfection reagent used was Lipofectamie RNAiMAX (Invitrogen. Cat. No. 13778030). Additionally, throughout transfection, doxycycline (at the same concentrations mentioned above) was added to the complete growth medium to allow the expression of Cas9. Briefly, 10μM crRNA and tracrRNA were mixed with Tris buffer, to make a final 2μM solution. This was mixed with serum-free media, to form the gRNA complex, to which was added the transfection reagent Lipofectamie RNAiMAX, after it was diluted at 1/50. The transfection solution (tracrRNA:crRNA and the transfection reagent) were incubated at room temperature for 20 minutes. This was then added to the cells and incubated for 48 hours in humidified atmosphere at 37°C and 5% CO_2, before proceeding with further analysis.

2.5 Genomic DNA extraction and quantification

DNA was extracted from all the samples by following the protocol from QIAGEN QIAamp DNA blood mini kit (Cat. No. 51306), as described in appendix B for cultured cells.

Briefly, cells were cultured in T25 flasks, following the appropriate procedure, and when confluent, they were harvested and suspended in phosphate buffered saline (PBS), to which was added proteinase K. To this, Buffer AL is added and the sample in incubated at 56°C for 10 minutes, followed by the addition of ethanol. The mixture was then applied to a DNA QIAmp mini spin column and centrifuged. The supernatant was discarded, Buffer AW1 was added to the column and the sample was centrifuged. Again, the supernatant was discarded, Buffer AW2 was added to the column and the sample was centrifuged. The supernatant was discarded, and the DNA was eluted by adding Buffer AE to the column and collecting the flow through. DNA was quantified using the NanodropOne^{C} (ThermoFisher).

2.6 PCR genotyping assay

The primers used were the same as the ones developed by *Kidd et al.* (Kidd *et al.*, 2007), designed to distinguish insertion and deletion genotypes for A3B. Deletion primers generate a 700-bp PCR product (Forward primer: TAGGTGCCACCCCGAT; Reverse primer: TTGAGCATAATCTTACTCTTGTAC). Insertion 2 (from herein simply termed insertion) primers generate a 705-bp PCR product (Forward primer: TGTCCCTTTTCAGAGTTTGAGTA; Reverse primer: TGGAGCCAATTAATCACTTCAT). PCR was carried out according to the *HotStarTaq Plus PCR* handbook, following the *HotStarTaq Plus PCR Master Mix* protocol. Each reaction was made up to a total of 10ul, composed of 5ul of *HotStarTaq Plus PCR Master Mix,* 0.5 µM Forward Primer, 0.5 µM Reverse Prime,

100ng of DNA template, variable volume of RNase-Free Water and 2ul of the provided CoralLoad dye. The cycling conditions were the following: 5 mins at 95°C, 0.5 min at 94°C, 0.5 min at 54°C, 1 min at 72°C, plus a final extension time of 10 min at 72°C; the number of cycles used for each PCR was 30.

2.6.1 PCR optimization and sensitivity

PCR conditions and primers used were the same as described in section 2.6. DNA from wt NIKS and SKBR3 was mixed, in different proportion, in the same PCR reaction A total of 100 ng of DNA per reaction was used and DNA was mixed as follows: 50ng wt NIKS DNA and 50ng SKBR3 DNA; 70ng wt NIKS DNA and 30ng SKBR3 DNA; 90ng wt NIKS DNA and 10ng SKBR3 DNA; 99ng wt DNA and 1ng SKBR3 DNA. Additionally, two reactions with 100ng wt NIKS DNA and 100ng SKBR3 DNA respectively were also added.

2.6.2 Agarose gel

PCR products were analysed on a 1% agarose gel, using the 1kb DNA ladder from Fisher-Invitrogen.

2.7 Summary of oligonucleotides used for cell transfection and PCR primers

Oligo sequence	Function	Material and Method section
sgRNA sequence: GACGGAGCTGGGCGCCCTGTGGG	sgRNA 1 encoded from plasmid, used for NIKS transfection	Section: 2.2, Figure 6
sgRNA sequence: ACGGAGCTGGGCGCCCTGTG	sgRNA 2 encoded from plasmid, used for NIKS transfection	Section: 2.2, Figure 6
sgRNA sequence: GGGGAGGCCGATGGGGCACC	sgRNA 3 encoded from plasmid, used for NIKS transfection	Section: 2.2, Figure 6
crRNA sequence: GACGGAGCTGGGCGCCCTGTGGG	crRNA. RNA complementary to the targeted DNA section. It forms a complex with tracrRNA, which acts as a scaffold. The complex formed is gRNA. The gRNA binds with the targeted DNA and causes cuts.	Section: 2.4
Forward primer: TAGGTGCCACCCCGAT; Reverse primer: TTGAGCATAATCTTACTCTTGTAC	Deletion PCR primers, which generate a 700-bp PCR product	Section: 2.6
Forward primer: TGTCCCTTTTCAGAGTTTGAGTA; Reverse primer: TGGAGCCAATTAATCACTTCAT	Insertion primers, which generate a 705-bp PCR product	Section: 2.6

Table 1. List of oligonucleotides used and their function

The table lists all the oligonucleotides used for cell transfection as well as PCR primers. Gene sequence of A3A and A3B is almost identical, including the regions of interest, intron 4 of A3A and intron 7 of A3B. This means that the same sgRNA can be used to target both introns, as shown in **Figure 6**. The table also summarizes the expected size of the wild type product, detected by the insertion primers, and the deletion product, detected with the deletion primers. The deletion configuration is obtained as a result of the sgRNAs cut points from intron 4 of A3A, to intron 7 of A3B. This causes a genomic deletion of ~29.5kb (see section 1.4), which produces PCR products of ~700-bp

2.8 Cell sorting via FACS

After transfection, cells were harvested from the flasks, resuspended in PBS and analysed on BD FACS*JAZZ* cell sorter (BD Biosciences), with 10,000 events per sample. Cells were sorted according to their GFP signal. Data was acquired and

analysed using the BD FACS Sortware software. Non-transfected NIKS and TK-NIKS, were used to establish the background level of cells autofluorescence.

2.9 Single-cell cloning

Single cell cloning was performed following the low-density seeding method. NIKS that had been transfected and sorted, were diluted to a density as low as 0.5 cell per well of a 96-well plate. A cell suspension of 10 cells/mL in complete media was made and 50 µl of this was transferred to each well of a 96-well plate. Four hours prior to seeding the NIKS, feeder cells were added to the plate. These were diluted in a suspension of 1000 cells per 50 µl; 50 µl aliquots were added to each well of a 96-well plate. Two 96-well plates were set up for each plasmid sample. Plates were checked after 2-3 days and after that every day for two weeks, to identify wells that contained only one cell which would give raise to a colony with genetically identical cells. Wells that contained more than one cell, and therefore generating more than one colony, were disregarded.

2.10 Cell viability assay

96 well plates were seeded with 3T3 feeder cells, wt NIKS and A3B A5 NIKS. The 3T3 feeder cells were seeded at a density of $1x10^3$ per well, and both types of NIKS were seeded at a density of $3x10^3$ per well. A3B A5 NIKS and wt NIKS were plated in different 96 well plates, at the same time as the feeder cells, with FC media as described in the Tissue Culture section for NIKS. Cells were treated for 5 days with different concentrations of ganciclovir (8 serial dilution were done with 1:3 dilution. The starting concentration was 200 µM) and incubated in humidified atmosphere at 37°C and 5% CO_2. On the 5^{th} day, cell proliferation was measured using CellTiter 96 AQueous One Solution Cell Proliferation Assay system (Promega), according to

manufacturer's instructions. 20µl of CellTiter 96® AQueous One Solution reagent was added per well, the plates were incubated in humidified atmosphere at 37°C and 5% CO_2. After 3 hours, absorbance was measured at 490nm using a 96-well plate reader. A dose-response curve was generated using GraphPad Prism 8.0.2. Three replicates well per each concentration were used to measure cell proliferation. The optimal ganciclovir concentration was then added to cells in culture.

2.11 Electroporation

Electroporation was carried out by using the Gene Pulser Xcell Electroporation System. MFCF10 and NIKS were harvested from the flasks and, after a cell count, $1x10^6$ cells were resuspended in 400 µl OptiMEM. Previously purified Cas9 protein is added to the cell suspension (Cas9 stock concentration is 6 µg/ml). 1mL of this mx was then added to a 0.2 cm electroporation cuvette, which was then added to the Gene Pulser chamber. The exponential protocol was selected from the machine, with one pulse at 300V and capacitance of 300 µF. After electroporation, cells are left to rest for 5 mins and complete medium was added in a 1:1 ration with Opti-MEM the cell are transferred into wells of 24-well plate. A coverslip coated with poly-D-lysine (Sigma P7280) had been previously inserted in the wells which was then needed for fluorescence microscopy. The plate was incubated overnight in humidified atmosphere at 37°C and 5% CO_2.

2.12 Cell fixation

Cells that adhered to the cover slip inserted in the 24-well plate, were washed twice with cold PBS and fixed in 4% paraformaldehyde for 10 mins. Cells were then washed three times with PBS, incubated for 15 mins with 50 mM NH_4Cl and washed again three times with PBS. The cells were then permeabilized and blocked for 15

mins with 2% (w/v) BSA in PBS, supplemented with 0.1% (v/v) Triton X-100. Without being washed, the cells were incubated for 1 hour with nucleus stai3
n Hoechst (Invitrogen. Cat. No. H3570), stock concentration of 20mM. A dilution of 1:12 was made with PBS and the diluted dye was place on parafilm. The coverslips were then placed inverted on the dye and covered as the dye is light sensitive. The cells were then washed three time with PBS and one time with dH_2O.The cover slips were then mounted on microscope slide with mowiol, supplemented with 2.5% (w/v) DABACO (Sigma. Cat. No. D-2522). The coverslips were stored overnight in the dark at room temperature, until the mowiol set, then they were stored at 4◦C. the slides were then analysed with fluorescence microscopy.

2.13 Fluorescence microscopy

Imaging was conducted with Widefield CytoVision Olympus BX61 microscope with Olympus UPlanFl 100X 1.30 Oil Japan numerical aperture and Hamamatsu Photonics Digital CCD camera ORCA-R2 C10600-10B-H, using DAPI and Cy5 filters.

3 Results

3.1 Optimizing a PCR genotyping assay to detect A3A_B deletion

PCR primers previously designed by *Kidd et al.* (Kidd *et al.*, 2007) were used to carry out all genotyping assays. These were insertion and deletion-specific primers (see **Figure 2**). The genetic insertion configuration corresponds to the wild type cell make up; the same labelling as used in the study by *Kidd et al.* is used in this project. The PCR primers were tested against wt NIKS and SKBR3 DNA. Additionally, the DNA of the two cell lines was mixed with different ratios, as shown in **Figure 7**, in order to determine the sensitivity of the assay.

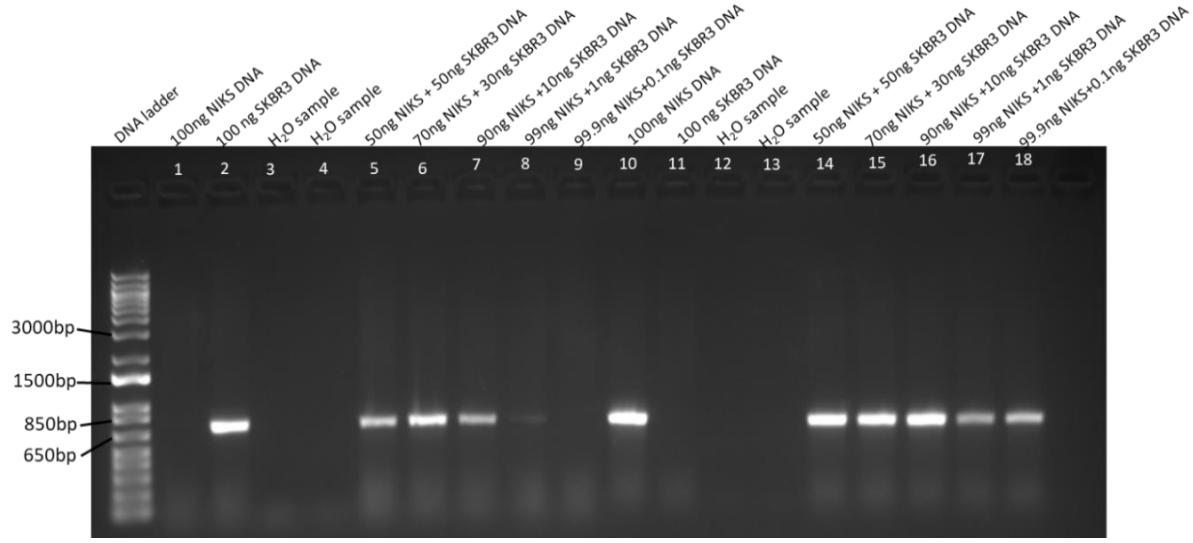


Figure 7. **Agarose gel showing PCR results for the insertion-specific and deletion-specific genotyping primers.**

Deletion-specific PCR reactions generated products of the expected size of 700 bp (lines 1-9). Insertion-specific PCR reactions generated products of the expected size of 705-bp (lines 10-18).

DNA of wt NIKS was used as a negative control for A3B deletion (line 1. Line 10 indicates insertion for NIKS), as the cells present a normal genomic make up. Previous sequencing and gene expression analysis conducted in the laboratory, has indeed demonstrated that the NIKS cell line is homozygous for the insertion of A3B allele (wild type). DNA from SKBR3 cells was used as the positive control for the

deletion (line 2. Line 11 shows that this cell line differs from the insertion and hence wild type genotype), as SKBR3 cell line derives from mammary breast metastatic tumour and is homozygous for the A3B deletion. Line 8 shows that, even when there is as little as 1ng of DNA presenting the deletion in a heterogeneous sample, it can still be detected by PCR. Therefore, in a mixed population sample, a low number of cells with the deletion can be detected, making this s a very useful tool for testing the pool of mixed cell population after transfection and A3B gene knockout.

3.2 Establishment of GCV sensitivity in NIKS harbouring a puro-ΔTK dual selection cassette in A3B intron 1

In order to enrich for cells with the deletion, and increase the purity of the population, the TK-NIKS cell line was used. As shown in **Figure 4**, TK-NIKS cells were previously modified to express the *puro-ΔTK* selection cassette, which makes them sensitive to GCV. However, the optimal concentration needed to eliminate all TK-NIKS has not been established. Different serial dilutions were used to carry out the MTS assay, as shown in **Figure 8**. As GCV is dissolved in DMSO, it is difficult to say whether the effect seen in cells is due to GCV or DMSO. It appears that GCV has an effect on cell viability at concentrations between 3.125 μM and 100 μM. At concentrations higher than 100 μM, DMSO also seems to play a role in cell viability, however, it is difficult to establish for sure. At concentrations lower than 3.125 μM, there is not much difference between GCV and DMSO and GCV does not seem to have an effect as almost 100% of the cells is alive.

The cassette seems to confer different susceptibility to GCV, depending on the cell line used, with different IC_{50} values. This is further analysed in the discussion.

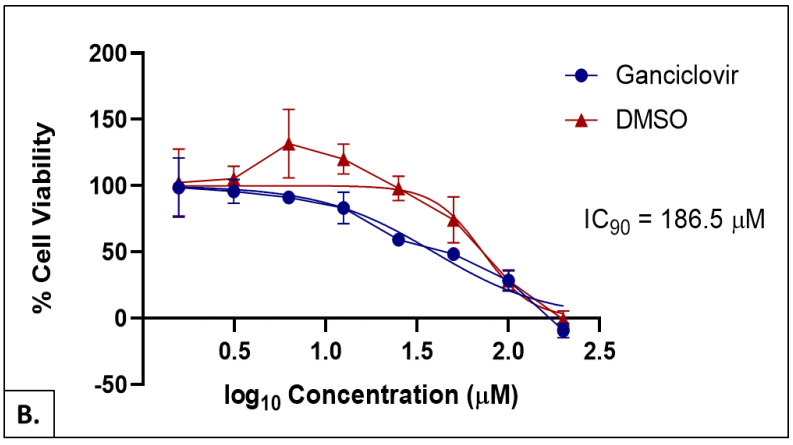


Figure 8. The effect of ganciclovir on TK- NIKS

Non-transfected A3B A5 NIKS were treated with different concentrations of GCV. DMSO was used as control and treated the same as GCV. For both experiments, cells were plated in 96-well plate together with either GCV or DMSO and after 5 days, MTS reagent was added, and OD was read at 490nm. Cell viability was calculated as a percentage of untreated cells and each data point was an average of triplicate repeats**.** The highest drug concentration used was 200 μM, and 1:2 serial dilutions were carried out. IC_{90} for GCV is shown in the graph.

3.3 Cell transfection

3.3.1 Transfection of NIKS

To generate the APOBEC3B deletion, NIKS were selected as the isogenic model. NIKS were transfected with plasmids containing the CRISPR/Cas9 system, sgRNA targeting intron 4 of A3A and intron 7 of A3B, as described in Material and methods, section 2.3. **Figure 9** shows the cells 48h post transfection, indicating that transfection efficiency is very low; despite being ~80% confluent, not many fluorescent cells are seen. The GFP reporter gene inserted in the plasmid used for transfection (**Figure 5**), allows visualization of successful transfection.

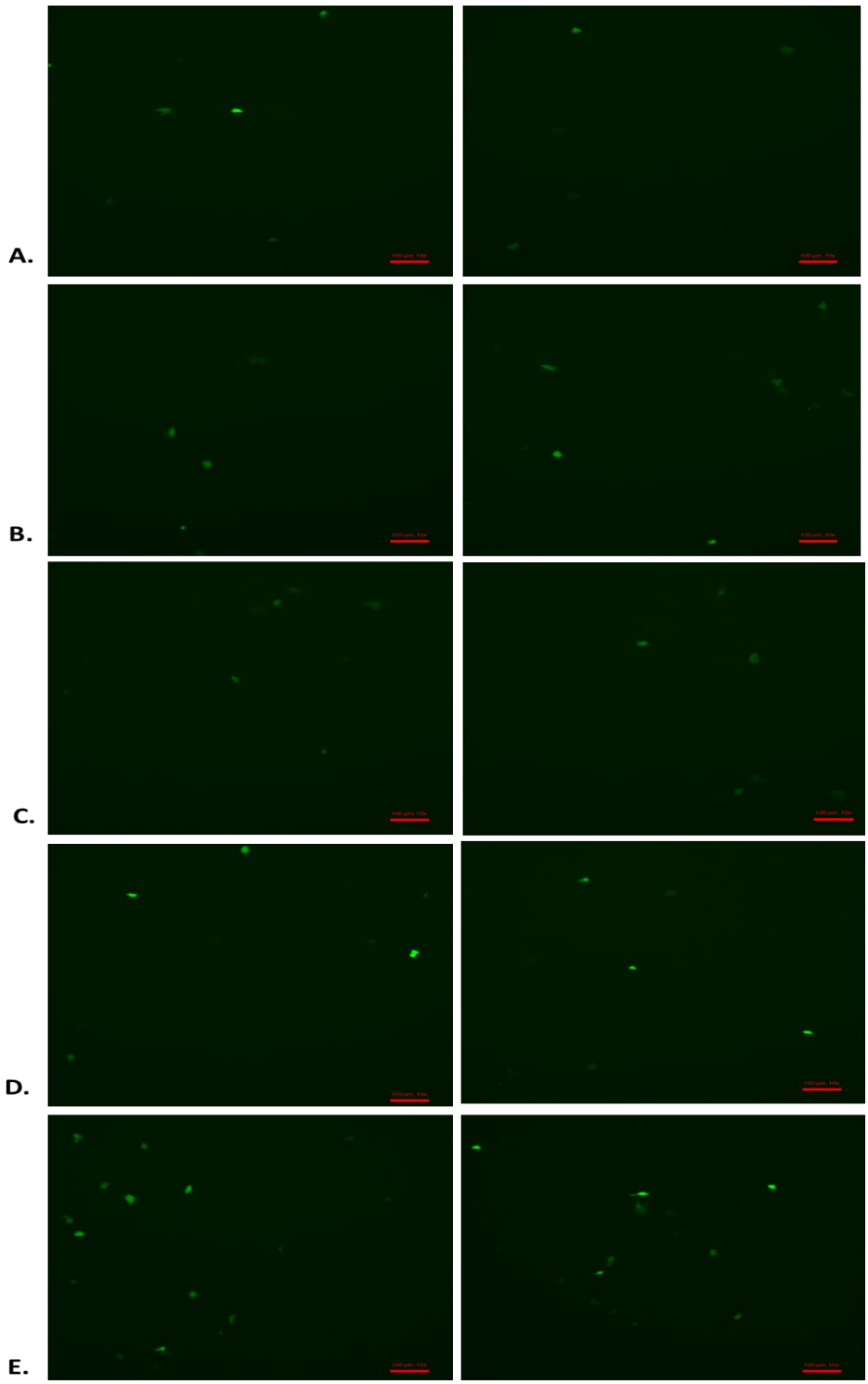


Figure 9. **Transfected NIKS.**

At ~80% confluency, wt NIKS were transfected with empty plasmid. The pictures show transfection efficiency for each condition used chosen for each condition used. **(A.)**, NIKS transfected with sgRNA1 **(B.)**, NIKS transfected with sgRNA2 **(C.)**, NIKS transfected with sgRNA3 **(D.)** NIKS transfected with sgRNA1+2+3 **(E.)**. Imaging was done 48h post transfection, at 100x magnification.

3.3.2 Transfection of TK-NIKS

TK-NIKS cell line was also transfected in the same way as NIKS cell line (Material and Methods section 2.3). Cells were examined 48h post transfection and pictures were taken, shown in **Figure 10**. This seems to show that, at ~80% cell confluency, that the transfection efficiency in TK-NIKS is higher compared to NIKS (**Figure 9**), a more of fluorescent cells are seen in **Figure 10**.

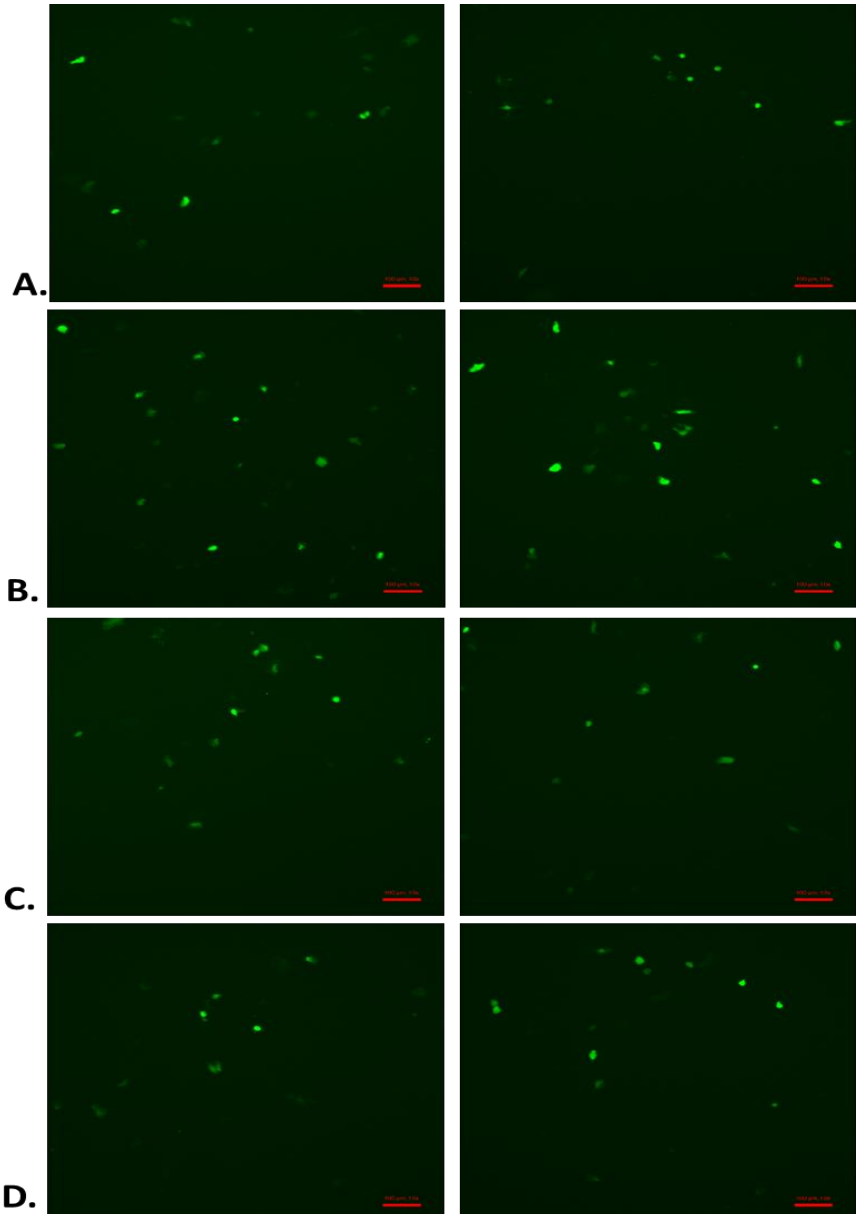


Figure 10. **Transfected TK-NIKS.**

At ~80% confluency, TK-NIKS were transfected. The pictures show transfection efficiency for each condition used. **(A.)** TK-NIKS were transfected with empty plasmid; **(B.)** TK-NIKS transfected with sgRNA1; **(C.)** TK-NIKS

transfected with sgRNA2; **(D.)** TK-NIKS transfected with sgRNA3. Imaging was done 48h post transfection, at 10x magnification.

3.4 Fluorescence Activated Cell Sorting (FACS)

3.4.1 Cell sorting of transfected NIKS

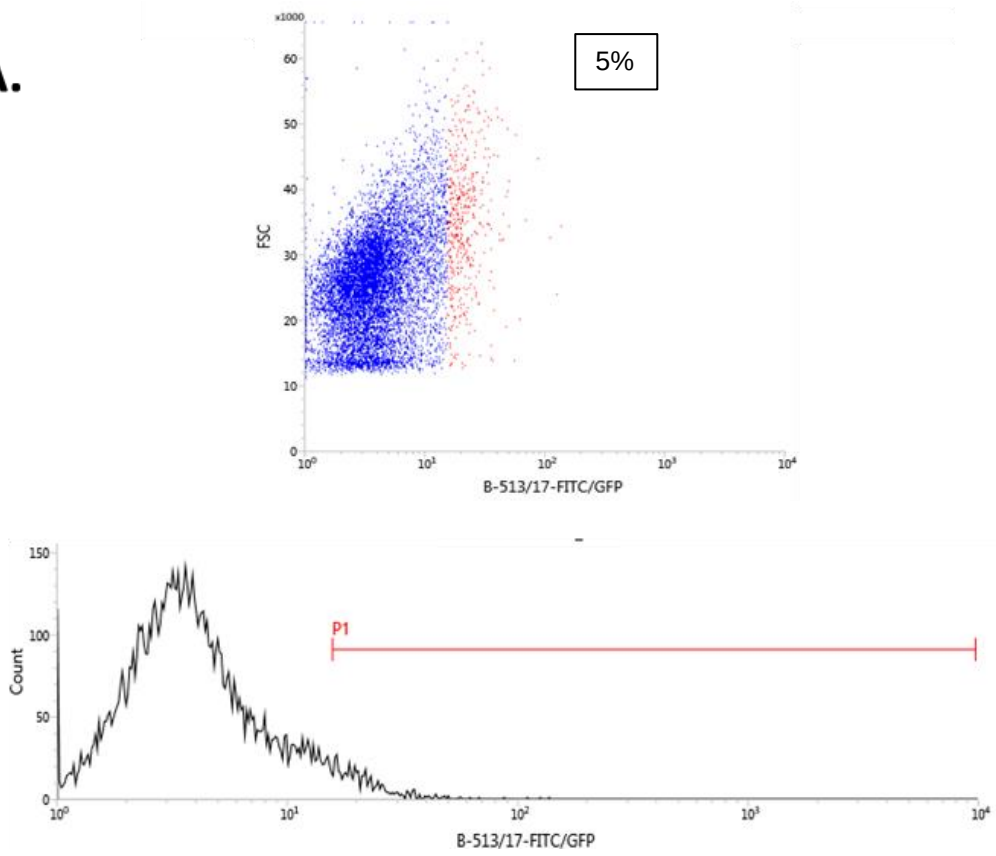
In order to enrich the population for transfected cells, FACS was carried out. This technique relies on the expression of GFP by transfected cells. The fluorescence emitted is detected by the lasers in the flow cytometer and cells are separated accordingly. **Figure 11** shows the results obtained from each sample.

Figure 11 (A) shows wt non-transfected NIKS sample, used as a control to set the population autoflourescence base line. Any variation from that, was considered to be as a result of GFP expression from the plasmid and it is represented by gate P1 (in red) in the histogram. The gating system used for sorting was not very stringent, due to the low transfection efficiency seen in **Figure 9**; the main goal was to have a population with a higher proportion of transfected cell, however, some non-transfected cells will inevitably be carried over. In fact, **Figure 11** (A), shows that a small percentage of the total cell population appears to be classified as transfected (the red cell population of the scatter plot).

The scatter plot diagrams, [**Figure 11** (B-F)] represents the entire cell population for each sample. The transfected cells, expressing GFP, which go into gate P1 gate are represented by the red population. These were separated from the whole population, collected and cultured to grow for later analysis.

It is encouraging to see that the amount of cells sorted from the transfected samples [**Figure 11** (B-F)] is higher than the control.

A.



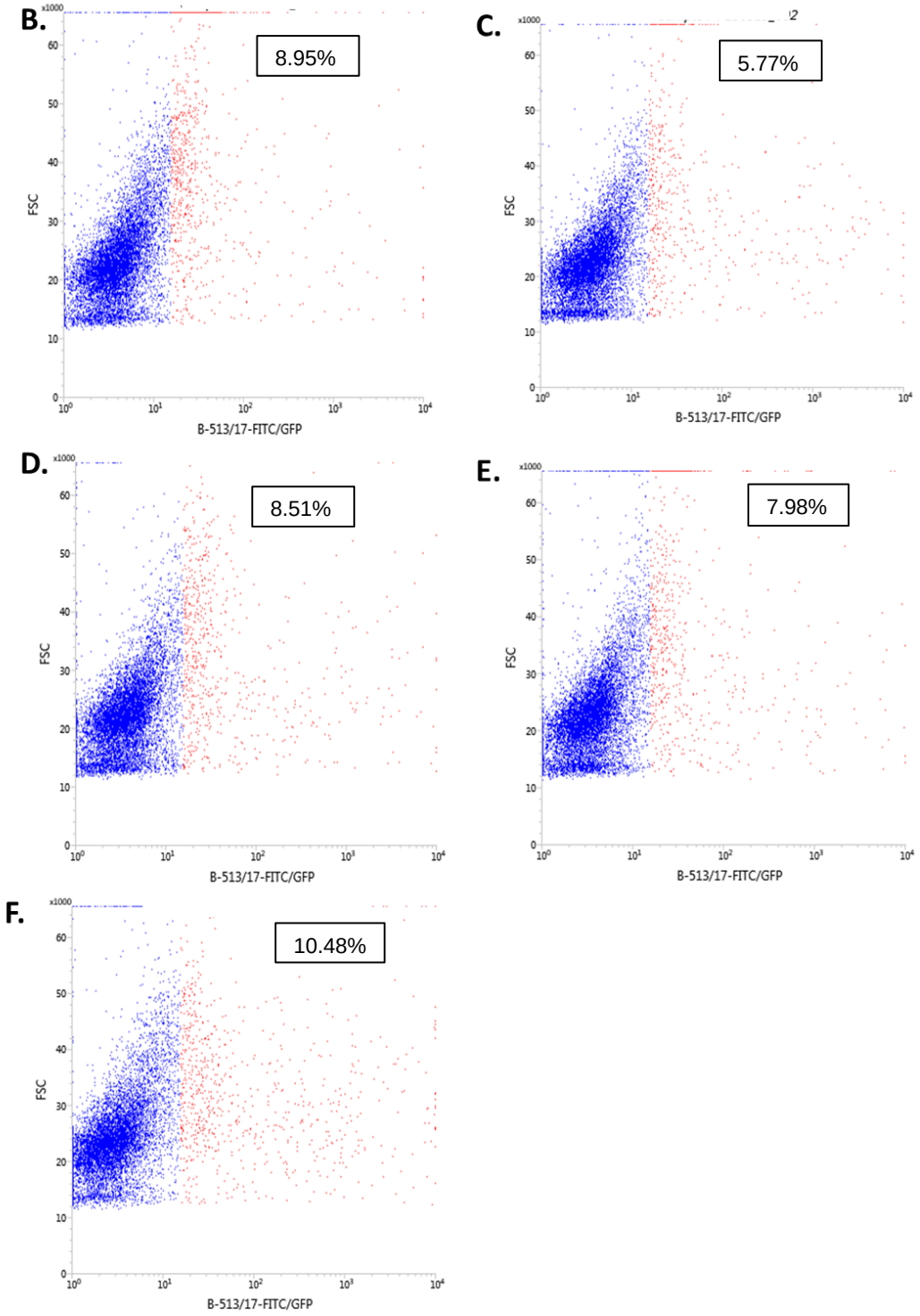


Figure 11. **Cell sorting analysis of transfected NIKS**

Post transfection, NIKS were sorted via FACS. Successfully transfected cells were selected over not transfected ones due to GFP expression. Percentage of collected cells is expressed in each diagram. Here are shown the scatter plot diagrams and histogram. The gate set as "P1" indicates cells that express GFP protein. **(A.)** Control of wt non-transfected NIKS. **(B.)** NIKS transfected with empty plasmid **(C.)** NIKS transfected with sgRNA1 **(D.)** NIK transfected with sgRNA2 **(E.)** NIKS transfected with sgRNA3 **(F.)** NIKS transfected with sgRNA1+2+3.

3.4.2 Cell sorting of transfected TK-NIKS

The same parameters using for NIKS, were used to sort transfected TK-NIKS.

Figure 12 (A) shows non-transfected TK-NIKS sample, used as a control to set the population autoflourescence base line. Any variation from this was considered to be the result of GFP expression from the plasmid, represented by gate P1 (in red) in the histogram. However, in this sample, a percentage of these cells appear to be transfected, due to the presence of a red population in the scatter plot diagram. This is as a result of low transfection efficiency (**Figure 10**), which led to use a not stringent gating system, in order to collect as many transfected cells as possible. Therefore, in the transfected samples [**Figure 12** (B- (F)], some of the non-transfected cells are carried forward and selected by the cell sorter, creating a heterogeneous population of sorted cells.

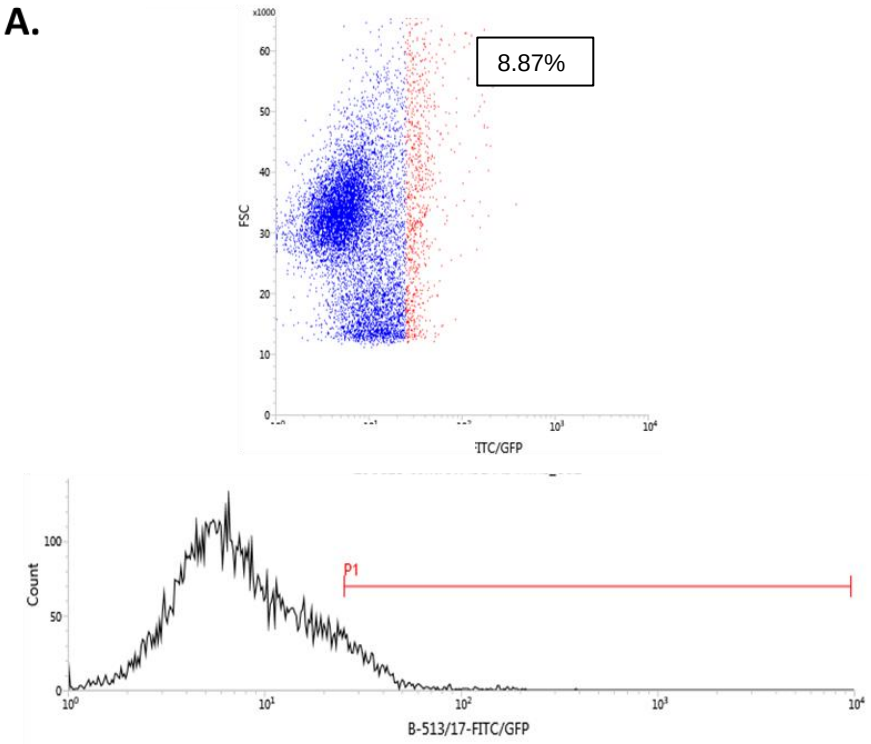


Figure 12 (B) shows TK-NIKS transfected with the empty plasmid (see Material and Methods section 2.2). This is used as a control and, together with **Figure 10** (A), indicates that cells are successfully transfected, despite the low efficiency and sort outcome.

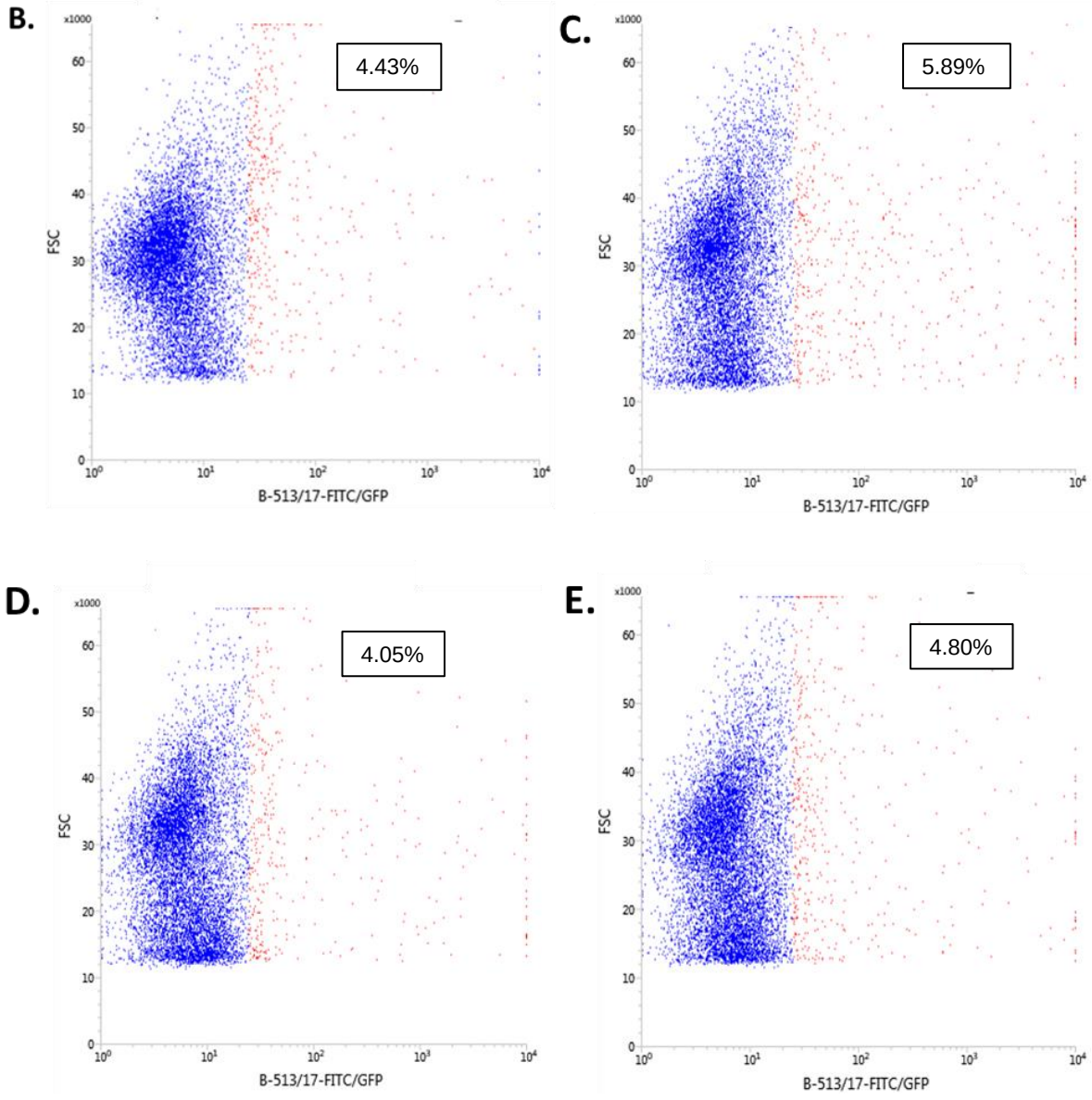


Figure 12. Cell sorting analysis of TK-NIKS

Post transfection, TK-NIKS were sorted via FACS. Successfully transfected cells were selected over not transfected ones due to GFP expression. Percentage of collected cells is expressed in each diagram. Here are shown the scatter plots and histogram. Gate P1 indicates the cell that express GFP protein. **(A.)** Control of non-transfected TK-NIKS. **(B.)** TK-NIKS transfected with empty plasmid **(C.)** TK-transfected with sgRNA1 **(D.)** TK-transfected with sgRNA2 **(E.)** TK-transfected with sgRNA3.

3.5 Generation of the ~29.5kb deletion in NIKS using CRISPR/Cas9

After NIKS were transfected and sorted, they were left to recover and grow in complete medium for ~2 weeks. At this point, a section of the cells was harvested for genomic DNA extraction, the rest were kept growing for subsequent analysis (see later sections). The extracted DNA was used to carry out PCR, using the primers developed by *Kidd et al.* (Kidd *et al.*, 2007), as described in the Material and Methods section 2.6. PCR products were then analysed on agarose gel, as shown in **Figure 13**. A band of approximately 700-bp is seen in lines 2-5, suggesting that a successful deletion, spanning from intron 4 of A3A to intro 7 of A3B, was obtained. Similar sized bands, however, were seen in lines 11-14, suggesting that most cells are not transfected the deletion might have not occurred in all the cells in the sample (compare this with line 16, containing SKBR3 DNA). SKBR3 DNA is used as a positive control, as this cell line presents the A3B deletion; wt NIKS DNA is used as a negative control. In lines 8 and 17, 1ng of SKBR3 DNA is used to understand the sensitivity of the assay, as explained in section 3.1; a band of approximately 700-bp is seen in line 8 with the deletion primers so, even when A3B deletion is present in a small proportion of cells, it can be detected by PCR.

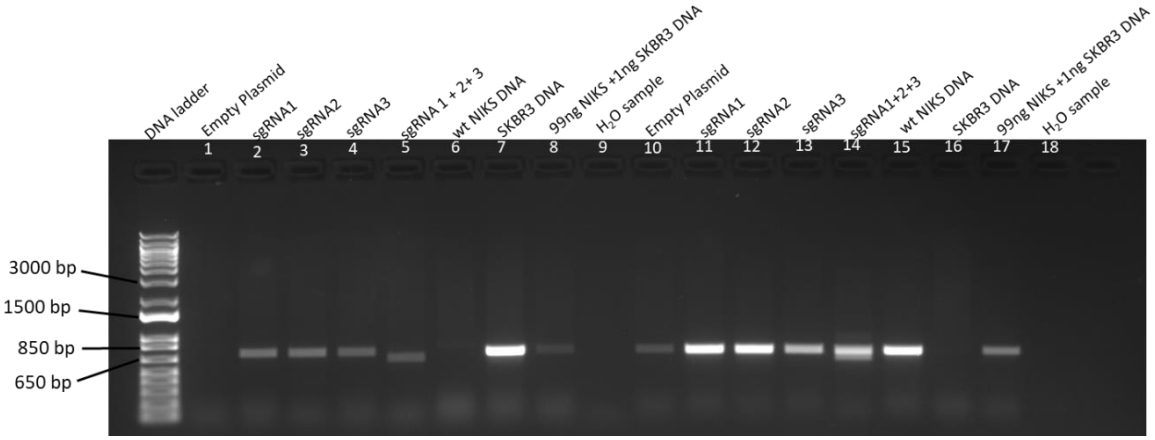


Figure 13. **Agarose gel showing PCR genotyping assay of transfected NIKS, post cell sorting**

PCR genotyping was carried out after transfected cells were sorted. Deletion primers were used in lines 1-9; insertion primers were used in lines 10-18. For more details on the primers see Material and Methods, section 2.6.

3.6 Single-cell cloning to isolate cells with the A3A_B deletion

FACS enriched the pooled population for transfected cells. However, the gating setting used was not stringent (see section 3.4), resulting in a heterogeneous population, formed by cells that could be heterozygous or homozygous for the deletion, or that have not had the deletion at all. In order to isolate cells with A3B deletion and characterize the type of deletion (heterozygous, homozygous), single cell cloning was carried out. After sorting cells were left to recover for ~2 weeks, and plated in a 96-well plate at a density of 0.5 cell per well, (Material and Methods section 2.9). Due to time restriction, only 3 sgRNA-transfected clones from a total of 130 were harvested for DNA extraction and PCR genotyping, while all the other clones were frozen for future genotyping and use. The agarose gel obtained with the PCR products is shown in **Figure 14**. Not surprisingly, the deletion was not present in these samples, as deletion-specific PCR products cannot be seen in lines 2-4. Interestingly, a band caused by larger PCR products is seen in line 11, when the insertion primers were used.

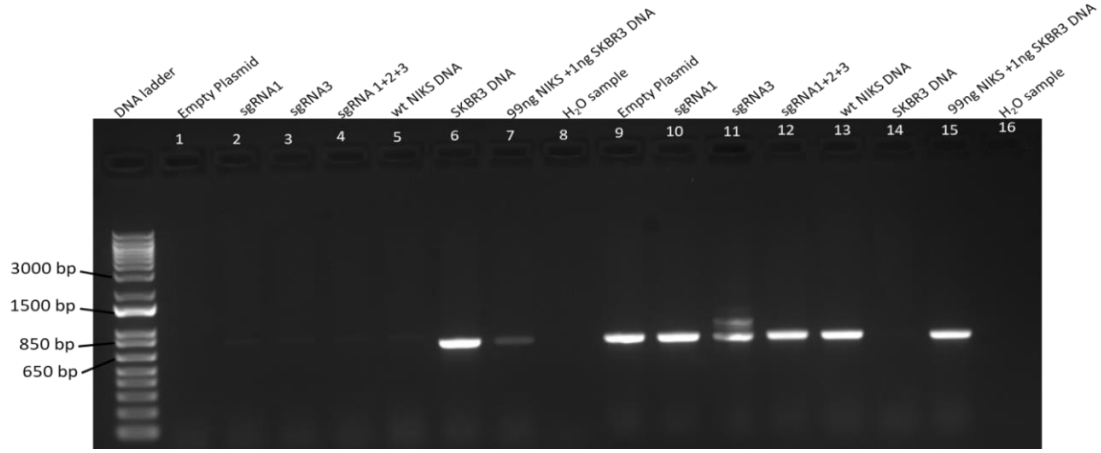


Figure 14. **Agarose gel showing PCR genotyping of single cell clones**

Post transfection, NIKS were sorted via FACS and single cells were isolated from the pooled samples in order to detect either homozygous or heterozygous deletion in the clones. Deletion primers were used in lines 1-9; insertion primers were used in lines 10-18. For more details on primers see Material and Methods, section 2.6.

3.7 GCV selection in TK-NIKS

In order to further enrich the population of cell, so that it is composed only of transfected cell with A3B deletion, TK-NIKS were treated with GCV. The previously constructed dual puro-ΔTK selection cassette, allows for both positive and negative cell selection (Chen and Bradley, 2000). Cells with this cassette are sensitive to thymidine (TK endogenous substrate) analogues such as FIAU and GCV; hence, GCV treatment leads to cell death. NIKS were modified and this cassette was added in intron 1 of A3B (Material and Methods section 2.1.2).

Post- transfection and flow cytometry, TK-NIKS were left to recover for ~2 weeks and treated with 63.5 μM of GCV for 11 days. This concentration is the IC_{50} of GCV, when a 1:2 dilution was used in a previous MTS (data not shown). However, the treatment was extended to 11 days, as no visible changes appeared after 5 days.

Figure 15 shows TK-NIKS before GCV treatment **(A-F)** and after treatment **(G-N)**. Cells appear slightly different, however after 11 days, they reached maximum confluency, as it is clear in the picture below **Figure 15 (G-N)**, and GCV treatment was stopped. All non-transfected TK-NIKS were expected to die after GCV treatment, but this was not observed in the current experiments. When analysed under the microscope, however, non-transfected TK-NIKS appeared more affected by the drug compared to the transfected cells **[Figure 15 (B)** and **(H)]**; a higher amount of cells appeared to be dead and floating in the flask. wt NIKS were expected to not be affected by the treatment, and this seems to be true, as cells do not appear much different before and after treatment **Figure 15 (A)** and **(G)**.

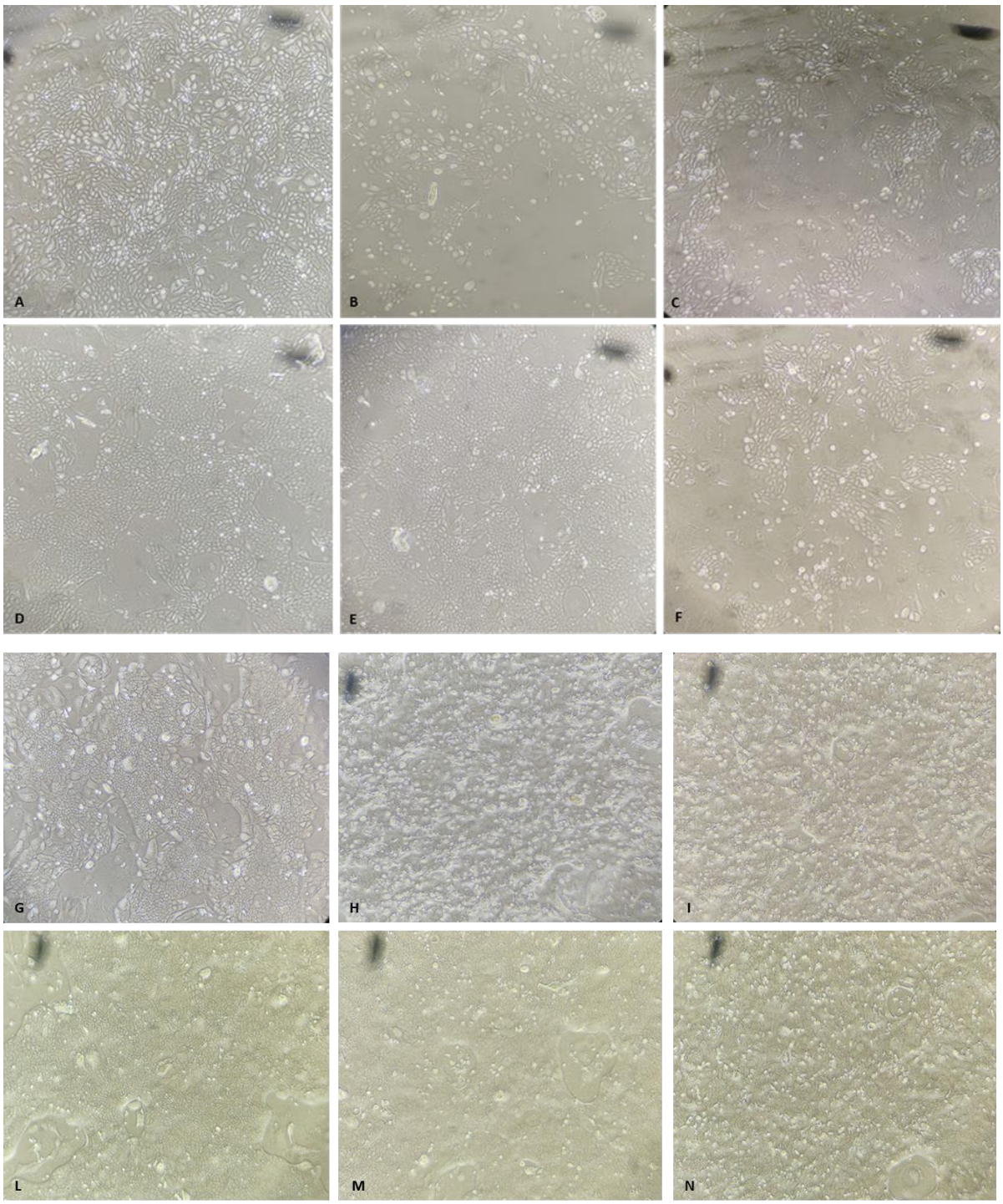


Figure 15. TK-NIKS before and after GCV treatment

Cells were treated with 63.5 μM GCV for 11 days. **(A-F)** Cells before GCV treatment. **(G-N)** Cells after GCV treatment. **(A.) (G.)** wt NIKS control. **(B.) (H.)** non-transfected TK- NIKS. **(C.) (I.)** TK-NIKS transfected with empty plasmid **(D.) (L.)** TK-NIKS transfected with sgRNA1 **(E.) (M.)** TK-NIKS transfected with sgRNA2 **(F.) (N.)** TK-NIKS transfected with sgRNA3.

Carrying out genotyping PCR of GCV treated TK-NIKS, might suggest that GCV does not have much of an effect on the cells. PCR products were analysed on agarose gel and results are shown in **Figure 16**. There are bands showing the expected PCR product ~700-bp in lines 2-4, with the deletion primers, however these are very weak. The band seen in line 15 with the insertion primers in non-transfected TK-NIKS, might indicate that a large portion of these cells is still alive and are not subject to GCV effect.

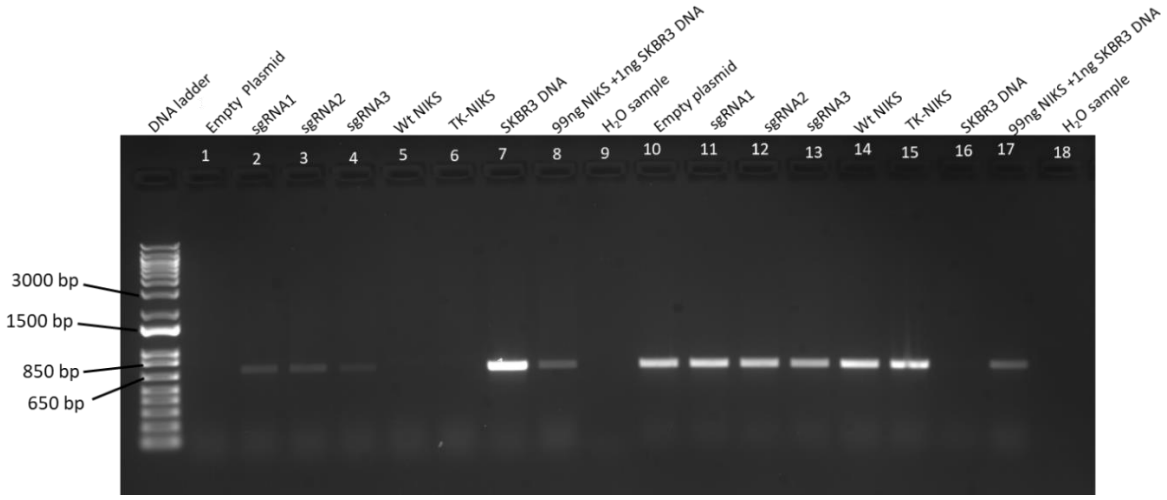


Figure 16. Agarose gel showing PCR genotyping of transfected TK-NIKS treated with 63.5 µM GCV

Genotyping of transfected TK-NIKS. Cells were treated with 63.5 µM ganciclovir for 11 days, DNA extracted and PCR genotyping was carried out. Deletion primers were used in lines 1-9; insertion primers were used in lines 10-18. For more details on the primers see Material and Methods, section 2.6.

Because the hypothesized result was not seen with 63.5 µM of GCV, a second cell viability assay was done and the IC_{90} was calculated, as shown in **Figure 8.** Therefore, another section of transfected TK-NIKS was treated with 186.5 µM of GCV for 7 days. Pictures of cells before and after GCV are not shown, however, results of the genotyping PCR are shown in picture **Figure 17**. The agarose gel shows bands of approximately 700-bp in lines 4-6 with the deletion primers; however, the prominent band shown in line 11 with the insertion primers, might indicate that not all the non-transfected TK-NIKS died as a result of GCV treatment.

A comparison of the results obtained with the two GCV concentrations, shows that the bands in lines 4-6 in **Figure 17** are more prominent than those in lines 2-4 in **Figure16.** This indicates that a higher GCV concentration enrich for cell with A3B deletion and therefore some selection occurs, however, this is still inefficient.

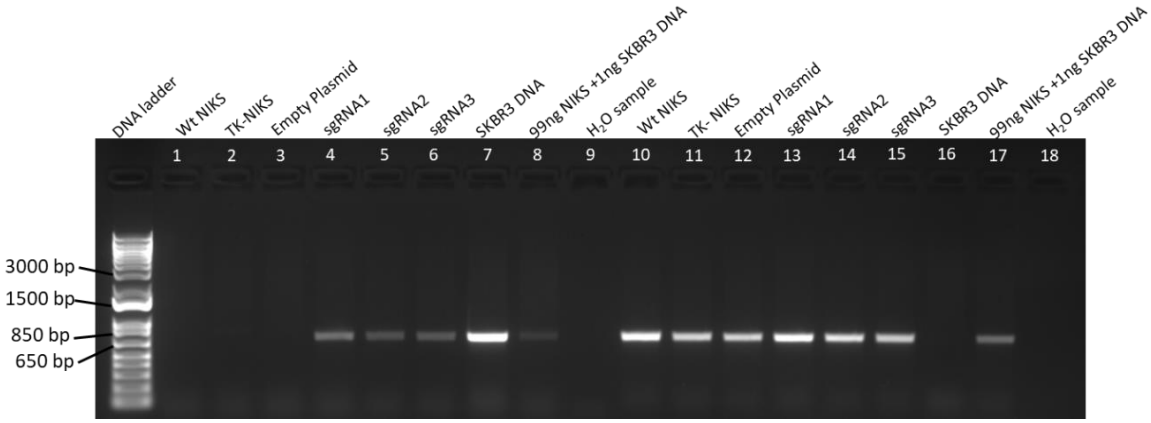


Figure 17. Agarose gel showing PCR genotyping of transfected TK- NIKS treated with 186.5 µM GCV

Transfected TK-NIKS were incubated with 186.5 µM ganciclovir for 7 days, DNA extraction and PCR genotyping were then carried out. Deletion primers were used in lines 1-9; insertion primers were used in lines 10-18. For more details on the primers, see Material and Methods, section 2.6.

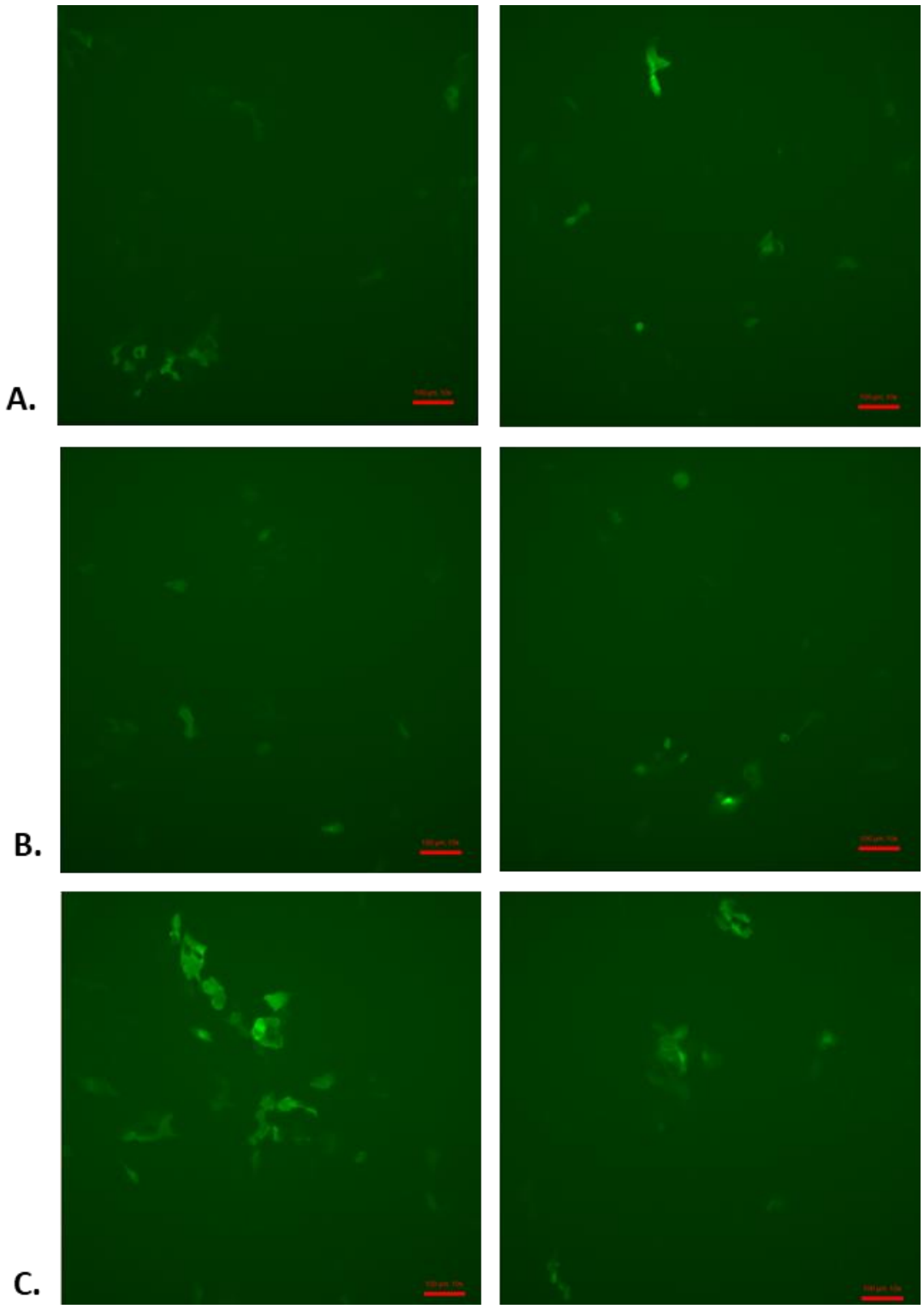
3.8 MCF10A transfection using different DOX concentrations

Until recently, A3B action was believed to cause genomic mutations that shape the evolution of a range of tumors, negatively contributing to prognosis and clinical outcomes (Nik-Zainal *et al.*, 2012; Burns *et al.*, 2013b; Burns *et al.*, 2013c; Roberts *et al.*, 2013; Sieuwerts *et al.*, 2014). However, APOBEC3 signature mutations are still detected in breast tumors with the ~29.5kb deletion from intron 4 of A3A to intron 7 of A3B, which removes A3B coding region (Chan *et al.*, 2015) (Nik-Zainal *et al.*, 2014) (Long *et al.*, 2013) (Xuan *et al.*, 2013). The deletion frequency is different among different populations (Klonowska *et al.*, 2017) (Kidd *et al.*, 2007) (Xuan *et al.*, 2013). For this reason, the epithelial breast cell line MCF10A was used as isogenic model, to reproduce the deletion, which would allow for testing the effect of the deletion on mutations propensity and tumorigenesis.

Previously modified, MFC10A cells had doxycycline (DOX)-inducible Cas9 GFP inserted into their genome, under the same promoter. Therefore, in order to express Cas9, cells were treated with different DOX concentrations (see Material and Methods, section 2.1.5). GFP allows to have a visual representation of whether Cas9 was successfully expressed.

Figure 18 shows images of MCF10A cells 2 days after dox treatment, at different concentrations. This was done before cell transfection.

Cas9 expression was successful with all the concentrations, however, from the images it is difficult to detect whether DOX concentrations had an impact on the mount of Cas9 synthesized, and which one was the optimal concentration for the cell line and the purpose of this work.


A.
B.
C.

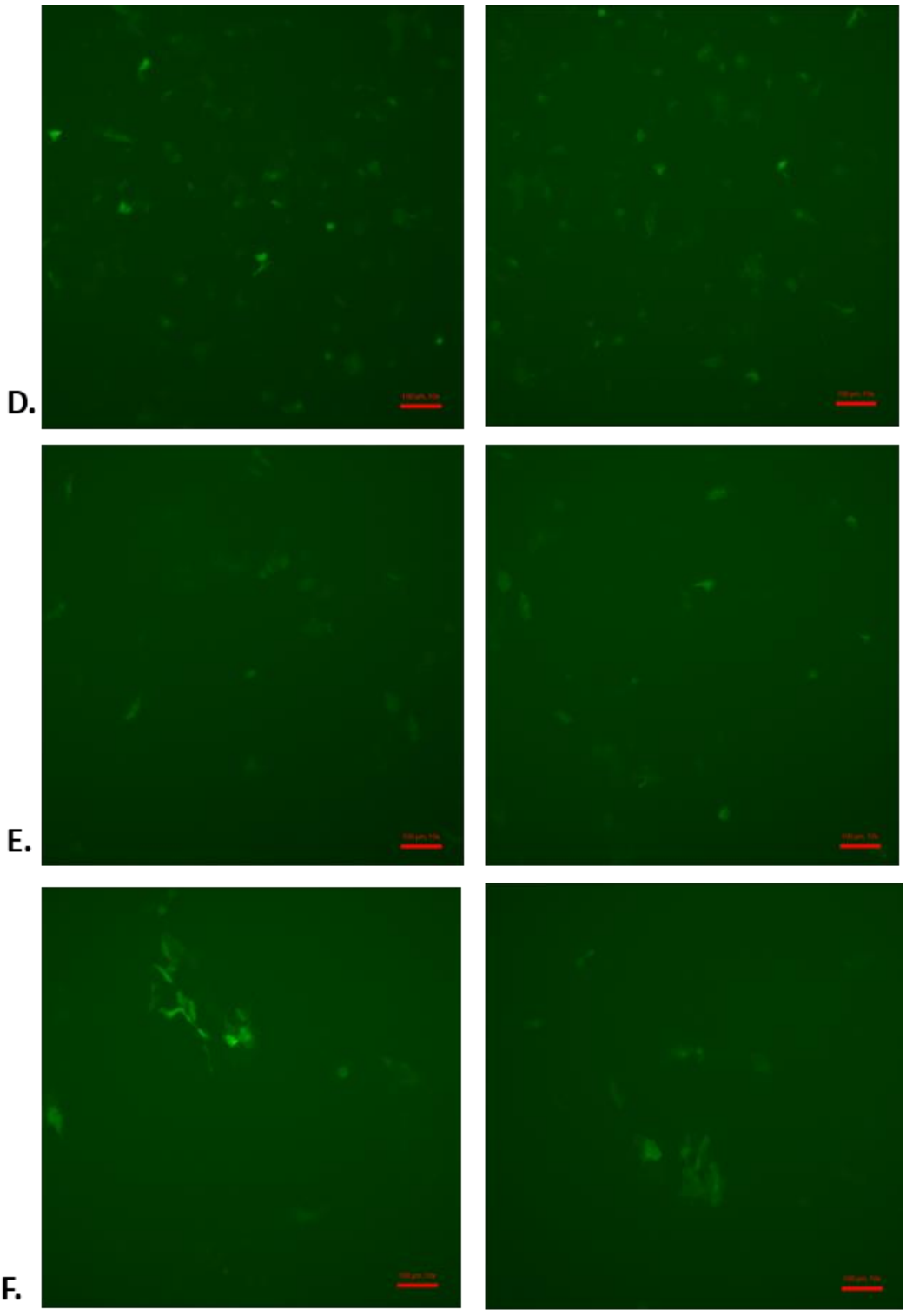


Figure 18. **Transfected MCF10A.**

Different concentrations of doxycycline were used to induce Cas9 expression. GFP insertion under the same dox-inducible Cas9 promoter, allows GFP expression upon DOX treatment, allowing to detect Cas9 expression. Two pictures from a large selection, were chosen for each DOX concentration, to show Cas9 expression efficiency. **(A)** 10ng/ml of DOX**; (B)** 50ng/ml of doxycycline; **(C)** 100ng/ml of doxycycline; **(D)** 250ng/ml of doxycycline; **(E)** 500ng/ml of doxycycline; **(F)** 1000ng/ml of doxycycline.

MCF10A cell treated with DOX for 2 days, were transfected with gRNA (tracrRNA:gRNA complex) (see Material and Methods, section 2.4). Because Cas9 was inserted in the cells' genome, it was assumed that all cells in the sample would express it and the deletion would occur, if transfection was successful.

48h post transfection, a section of the cells was harvested for genomic DNA extraction and the rest was kept growing. A PCR, using insertion and deletion primes as previously described (Material and Methods section 2.6), was carried out. PCR products were analysed on agarose gel and the results are shown in **Figure 19**. Despite expression of Cas9 seen in **Figure 18**, no successful deletion has occurred as no bands are seen with the deletion primers, lines 2-7.

Lines 1 and 13 are used as a negative control; the cells were transfected but no DOX treatment was done. A band of approximately 700-bp in lines 10 (positive control) and 11, indicates that primers are functional and other factors might be accountable for the results obtained.

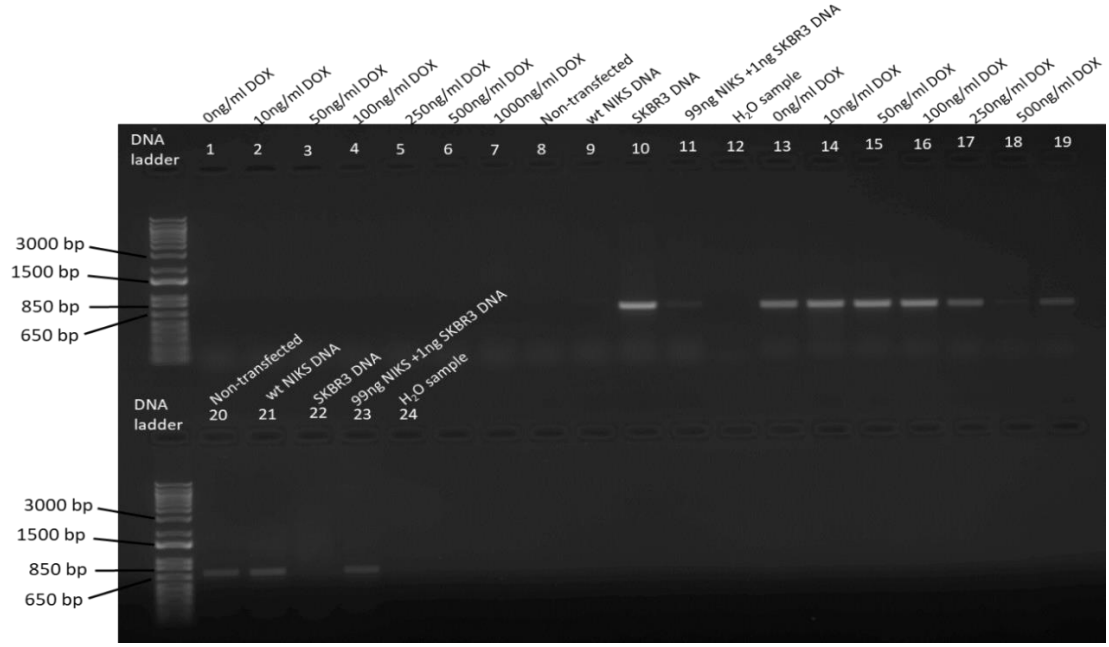


***Figure 19. Agarose gel showing* PCR genotyping assay for transfected MCF10A.**

MCF10A cells were treated with different concentrations of DOX, to induce Cas9 expression. Deletion primers were used in lines 1-12; insertion primers were used in lines 13-24. For more details on the primers, see Material and Methods section 2.6.

3.9 Electroporation and Fluorescence microscopy

Electroporation is yet another technique used to deliver CRISPR/Cas9 system into the cell. Highly efficient and quick, it is a relatively harsh method for the cells due to high voltage and partial membrane repair. It relies on electric waves that are pulsed through the sample, to make the cell membrane more permeable by creating temporary holes, so that small molecules (in this case CRISPR/Cas9 system) can be delivered into the cell.

To understand the efficiency of this technique, we decided to conduct a control experiment using a deactivated form of Cas9 (dCas9), tagged with red halo-tag. dCas9 was transfected into the wt NIKS and MCF10A cells via electroporation. No sgRNA, were inserted at this point (see Material and Methods section 2.11 for more details on electroporation).

Cells were left to recover, fixed and stained and observed under the microscope (see Material and Methods sections 2.12 and 2.13).

Figure 20 and **Figure 21**, show images of MCF10A and NIKS, respectively, after electroporation. Two different concentrations of dCas9 were used, 0.06 µg/µl of dCas9 [**Figure 20** (A) and **Figure 21** (B)] and 0.12 µg/µl of dCas9 [**Figure 20**(A) and **Figure 21** (B)]. Surprisingly though, not significant difference was observed between the two concentrations and the cell lines. Additionally, the transfection efficiency observed was very low and not beneficial to the work. As a result, this technique was not carried out any further.

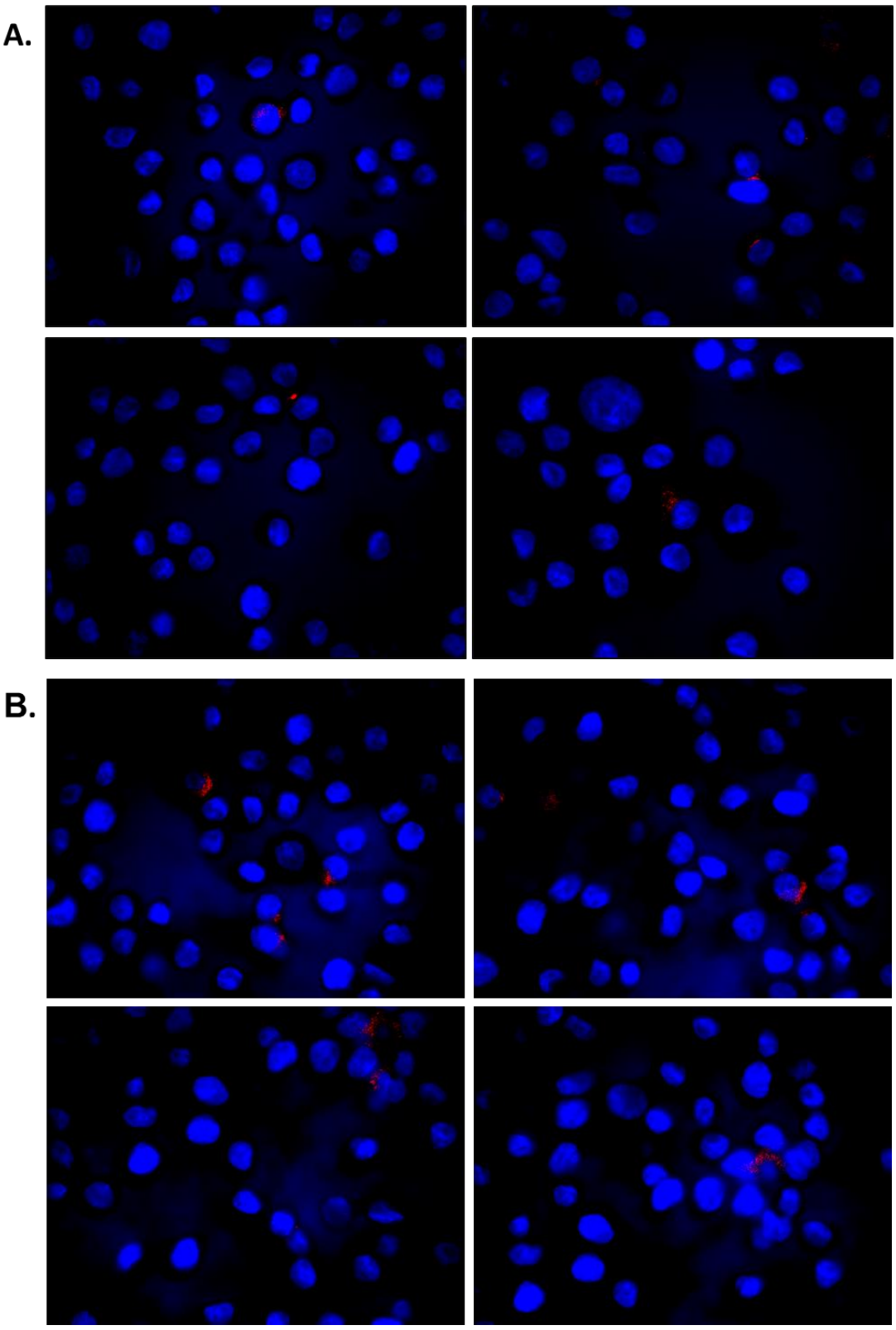
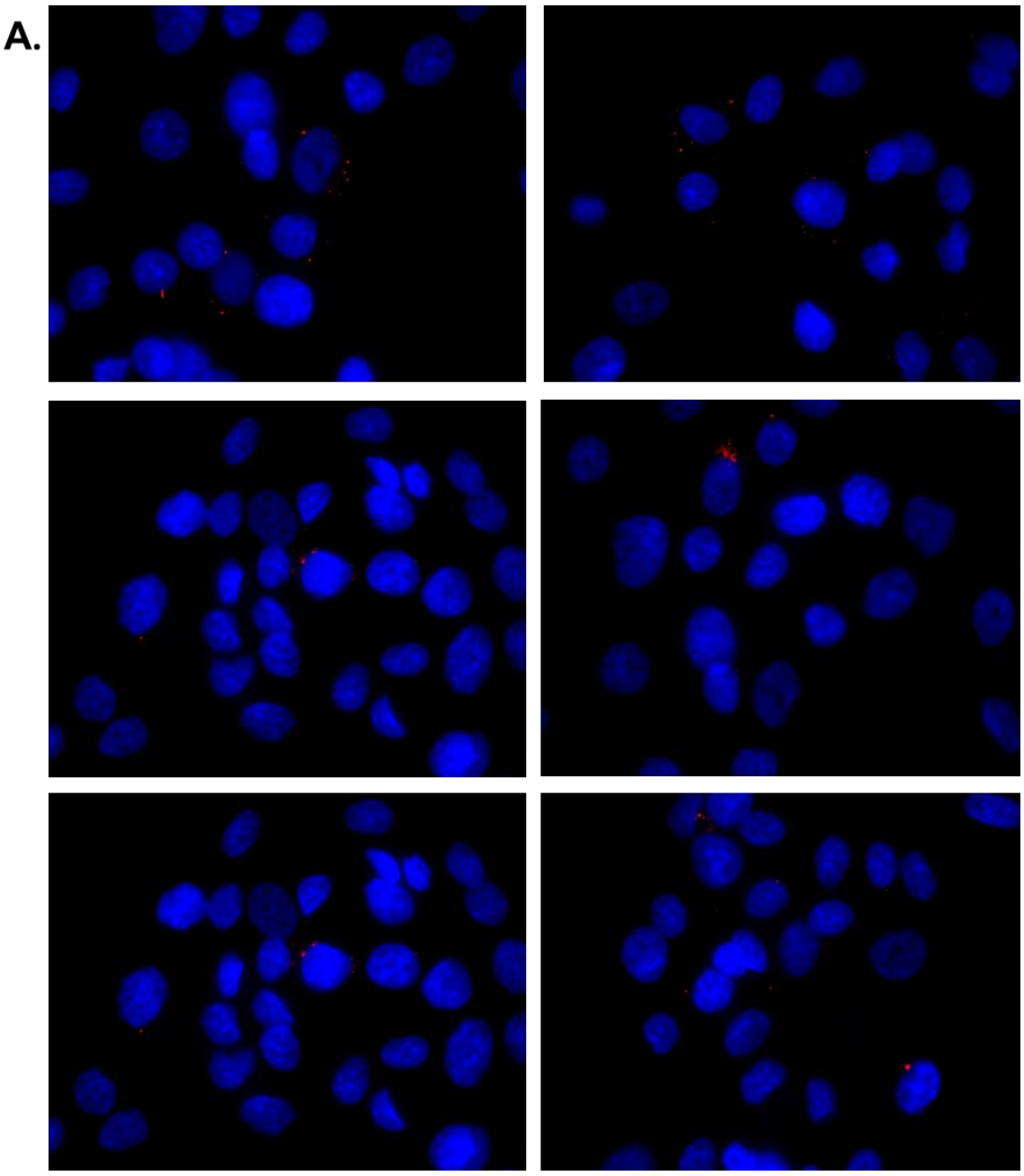


Figure 20. **dCas9 localization in MCF10A**

MCF10A cells were transfected via electroporation to deliver dCas9. Nuclei were stained with Hoechst stain and dCas9 was tagged with red halo-tag (shown by the red clusters in the images). **(A.)** Cells were transfected using 0.06 µg/µl of dCas9. **(B.)** Cells were transfected using 0.12 µg/µl of dCas9. Images were taken after cells were left to recover for a day and were fixed and stained. Imaging was done on an Olympus BX-61 microscope with DAPI and Cy5 filters.


A.

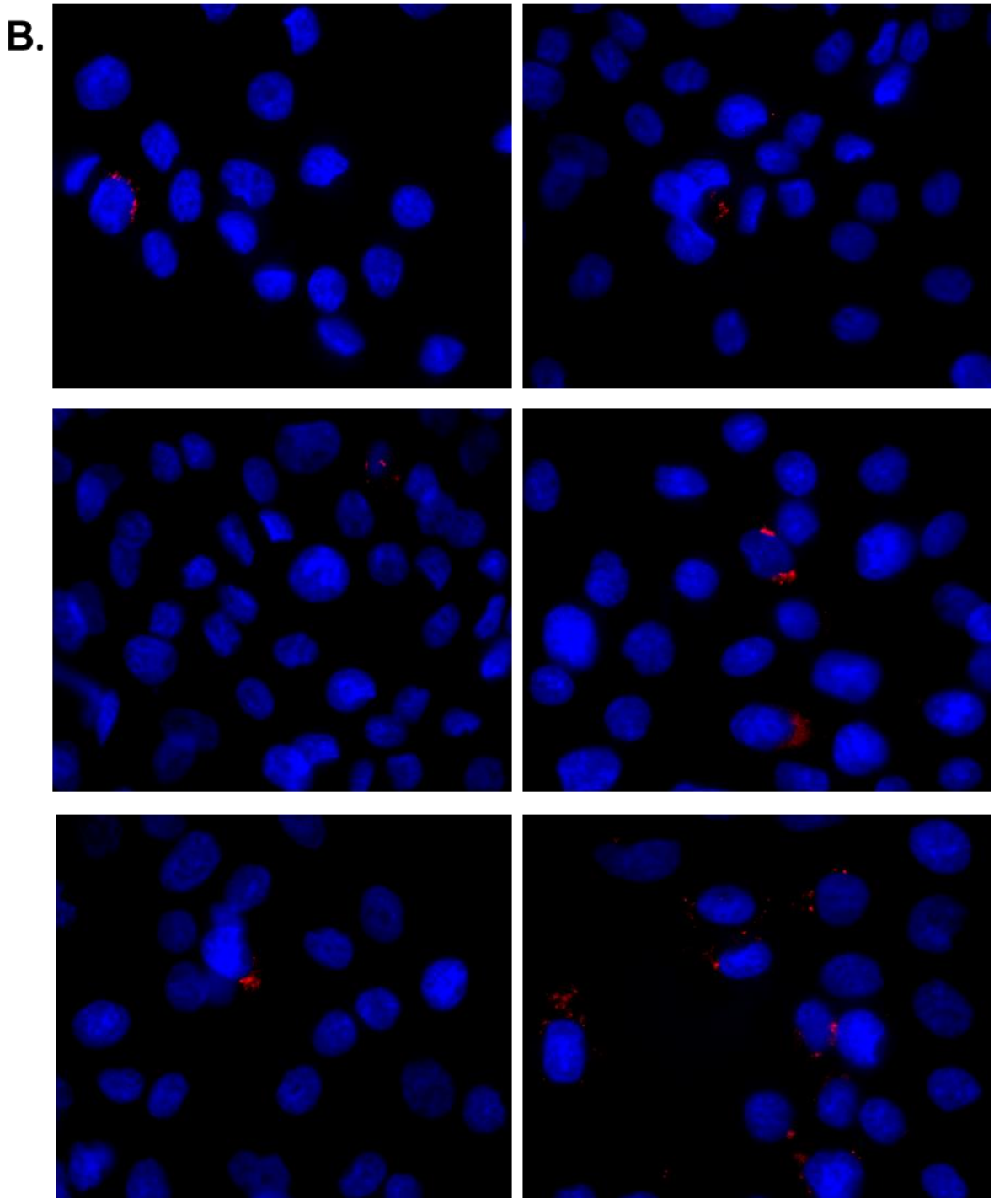


Figure 21. **dCas9 localization in NIKS**

wt NIKS cells were transfected via electroporation to deliver dCas9. Nuclei were stained with Hoechst stain and dCas9 was tagged with red halo-tag (shown by the red clusters in the images). **(A.)** Cells were transfected using 0.06 µg/µl of dCas9. **(B.)** Cells were transfected using 0.12 µg/µl of dCas9. Images were taken after cells were left to recover for a day and were fixed and stained. Imaging was done on an Olympus BX-61 microscope with DAPI and Cy5 filters.

4 Discussion

APOBEC3 enzymes play an important role in human health. These enzymes have shown to be vital against both RNA viruses such as HIV (Mangeat *et al.*, 2003), and DNA viruses such as HPV and HBV (Vartanian *et al.*, 2008; Vartanian *et al.*, 2010b). A3s are expressed in many healthy tissues (Salter *et al.*, 2016), with no harmful side effects when present at low levels, however, high expression of these enzymes can cause mutations in the host DNA.

Viral infection causes the activation of other defence mechanisms by the host, such as type I interferons; these have been shown to upregulate A3, particularly A3A and A3B (Bonvin *et al.*, 2006; Wang *et al.*, 2014; Cescon *et al.*, 2015; Suspène *et al.*, 2016) which leads to high mutations in host genome and ultimately tumorigenesis. However, upregulation and off target activity has not been fully outlined in non-viral cancers. Attention has been focused on A3A and A3B, which are the main A3 enzymes that play a pivotal role in cancer formation. Activation of these enzymes causes high signature mutations in oncogenes and/or tumour suppressor genes, characteristic of several cancers, particularly in breast cancer (Nik-Zainal *et al.*, 2012; Burns *et al.*, 2013b; Alexandrov *et al.*, 2013) Due to high mRNA expression levels, A3B has been thought to be the major cause of signature mutations in tumours (Burns *et al.*, 2013c; Roberts *et al.*, 2013; Sieuwerts *et al.*, 2017b) Surprisingly, though, a germline deletion spanning from intron 4 of A3A to intron 7 of A3B, increases cancer formation (Long *et al.*, 2013; Xuan *et al.*, 2013; Middlebrooks *et al.*, 2016) It is therefore clear that more studies need to be carried out to gain a better understanding of A3A and A3B role in the cell and in cancer.

4.1 Creating an isogenic model for the A3A_B deletion

During this project, the epithelial cell line NIKS was used as an isogenic model to recreate the naturally occurring A3B germline deletion. (**Figure 13**). The population of cells obtained after transfection and sorting was highly heterogenic, due to low transfection efficiency and non-stringent gating used during cell sorting; however, it was assumed that the deletion occurred in more than 1% of cells, due to more prominent bands seen in lines 2-5 as opposed to line 8 in **Figure 13**. Line 8 was obtained by combining 1 ng of SKBR3 DNA with 99ng of wt NIKS DNA, as explained in **Figure 7**, where SKBR3 DNA was assumed to represent 1% of the total DNA present in the PCR reaction. This shows that even when a small amount of DNA contains the A3B deletion, it can be detected in a PCR reaction. On the other hand, comparing lines 2-5 to line 7 (**Figure 13**), which consists of SKB3 DNA (homozygous for A3B deletion), indicates that only a minority of the cells in the sample have had the deletion. This can also be assumed by comparing lines 2-5 to lines 10-14 (**Figure 13**), which show DNA from the samples analysed with the insertion primers.

Interestingly, NIKS transfected with a mix of the three sgRNAs present a smaller sized PCR product, generating a lower band, as shown in **Figure 13**, line 5. It has been outlined that when making a large genomic deletion, combination of multiple sgRNA (termed multiplex) increases targeting efficiency, with a higher probability of generating the desired result (Song *et al.*, 2017). Despite, size of the genomic deletion and targeting efficiency are inversely proportional (Song *et al.*, 2017), multiplex sgRNAs were successful in generating genomic deletions ranging from 23kb to 100 kb, in different organisms such as rabbit, *C. elegans* and humans (Chen *et al.*, 2014; Han *et al.*, 2014; Song *et al.*, 2016). Therefore, the presence of the

lower band in **Figure 13** might be due to the combined effect of the sgRNAs used during transfection. The sgRNAs cut the DNA at different points, creating different sized bands, as shown in **Figure 22**, via non-homologous end joining repair.

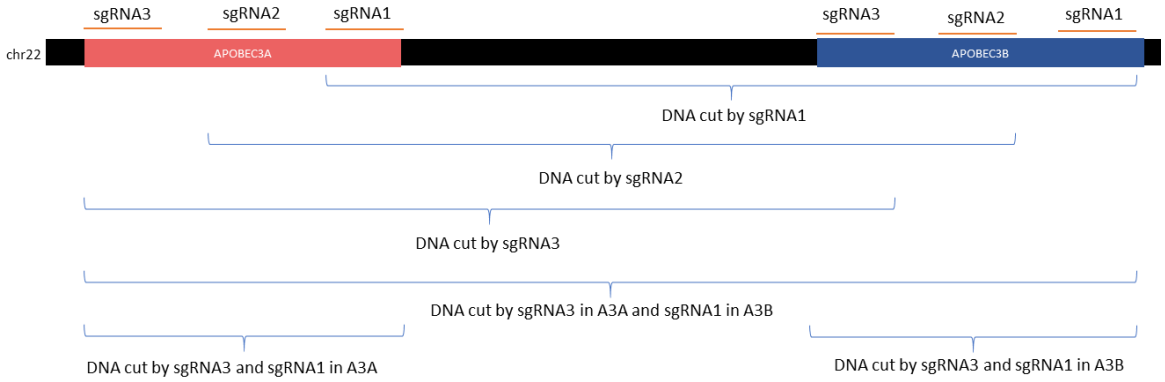


Figure 22. Representation of the different cut sites when using all three sgRNA in combination.

The diagram represents some of the possible cut sites (shown under the *A3A* and *A3B* genes) when the sgRNAs are used in combination. The cut done when the sgRNAs are used individually, are seen in lines 2-4 in **Figure 13**. The combination of sgRNAs can cause either larger or smaller DNA cuts, than the sgRNA individually. The diagram is not drawn to scale, as the sgRNA are found much closer together both in A3A intron and in A3B intron (see **Figure 6**); however, this diagram enables understanding of bands seen in lines 5 and 14 in **Figure 13**.

The band in line 5 is lower than the bands in lines 2-4 in **Figure 13**. This is the result of a larger cut created in the DNA, which spans from the sgRNA3 on A3A to the sgRNA1 on A3B, as shown in **Figure 22** (DNA cut by sgRNA3 in A3A and sgRNA1 in A3B). This generates a PCR product that is slightly smaller than the ones originated when the sgRNAs were used individually (DNA cut by sgRNA1, sgRNA2, sgRNA3 in **Figure 22**). Additionally, using the three sgRNAs together, causes the formation of the satellite band seen in line 14 of **Figure 13**. This is because sgRNAs can cause deletions spanning from sgRNA3 to sgRNA1 within each individual intron of A3A and A3B, forming small PCR products (**Figure 22** DNA cut by sgRNA3 and sgRNA1 in A3B). This rearrangement is still detected by the insertion primers, as they are positioned on either side of the sgRNAs in A3B (see **Figure 6**). DNA sequencing might aid the identification of the exact cut points within the genome,

which can provide a better understanding of the cells' genomic profile after deletion has occurred.

To isolate cells that show the deletion, single-cell cloning was carried out, as outlined in section 3.6. This will enable us to identify which clones present the deletion and whether this is homozygous or heterozygous.

It is not surprising that the results we observed in **Figure 14**, from the three clones picked did not show the deletion; however we would expect several of the remaining 130 clones to show it. This occurs because CRISPR/Cas9 system might not get delivered to all cells but these are still selected by the flow cytometer due to non-stringent gating applied (see section 3.4.1); moreover, even if the system was delivered to the cells, there is still a possibility that the CRISPR-induced double strand break was quickly repaired before the mutation occurred. In fact, rearrangement occurred at the CRISPR/Cas9 cut point is seen in line 11 in **Figure 14**, which shows a band generated by a larger PCR product than the one expected by insertion primers (~705bp). Performing RT-qPCR with the remaining clones can confirming that *A3B* expression loss can provide insight on the expression level of the other *A3* genes, such as upregulation of A3A, and what impact the deletion has had on their regulation. It would be particularly interesting to assess *A3H* expression in comparison to *A3A*, as the study from *Starrett et al (2016),* debunks A3A role in breast and lung cancers, indicating A3H-I the main cause of the APOBEC-signature mutations, when A3B is deleted.

4.2 Using *puro-ΔTK* did not enrich for cells with the deletion

In order to enrich for cells that have the desired ~29.5kb deletion, TK-NIKS cells were used. This cell line was obtained by inserting the *puro-ΔTK* selection cassette

in intron 1 of A3B NIKS (**Figure 2**). A successful deletion between intron 4 of A3A and intron 7 of A3B would impair expression of the *puro-ΔTK* selection cassette, and cells would be resistant to GCV, as shown in **Figure 23**.

Additionally, using this construct, allows for a quicker section of cells that present the deletion; GCV is applied to cell culture after cell sorting without the need to carry out single-cell cloning.

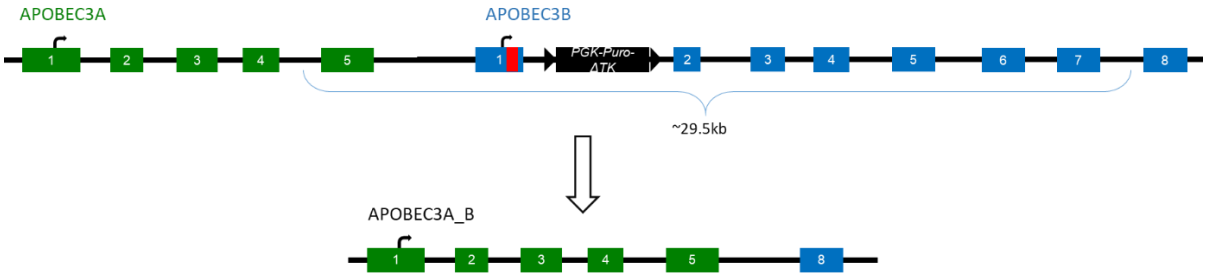


Figure 23. The ~29.5 deletion in TK-NIKS stops puro-ΔTK expression

The puro-ΔTK selection cassette is not expressed in the hybrid *APOBEC3A_B* gene, after deletion has happened, as shown in the diagram above. This makes cells resistant to GCV.

Treating sorted cells with GCV, should eliminate cells that did not have a successful deletion and leave a population of cells presenting the desired ~29.5kb deletion between A3A and A3B. This is possible because viral thymidine kinase and human thymidine kinase might have had a different evolution, resulting in HSV-TK to recognises a wider range of different substrates (Harrison *et al.*, 1991). Therefore, nucleoside analogues such as GCV, are substrates for HSV-TK (Borrelli *et al.*, 1988). TK converts the prodrug GCV to the toxic form known as GCV-triphosphate, which interferes with DNA elongation, eventually leading to cell death (Greco *et al.*, 2015; Hossain *et al.*, 2019). The *puro-ΔTK* selection cassette, exploits the differences in substrate specificity between human and viral TK to eliminate undesired cells (Hossain *et al.*, 2019). HSV-TK has also been widely used as gene

therapy in cancer, known as suicide gene therapy, with promising results (Greco *et al.*, 2015).

Our results show that GCV selection did not yield the desired outcome. Genotyping of GCV treatment of non-transfected TK-NIKS is not expected to show a PCR product with the insertion-specific primers on the agarose gel. However, the ~705bp product seen in lines 15 and 11 in **Figure 16** and **Figure 17** respectively, suggests the opposite. Thus, it can be assumed that in the transfected TK-NIKS samples, negative selection by GCV did not occur and the population is still heterogeneous.

GCV used in our experiments was dissolved in DMSO, therefore treatment with high drug concentrations results in a high percentage of DMSO which is not tolerated by cells, as it is show in **Figure 8**. At GCV concentrations higher than 100 μM it is difficult to separate GCV and DMSO action on the cells, so we cannot say with certainty what causes cell death above this concentration. Interestingly however, a comparison of the results obtained with the two GCV concentrations, shows that the bands in lines 4-6 in **Figure 17** are more prominent than those in lines 2-4 in **Figure16.** This indicates that a higher GCV concentration enrich for cell with A3B deletion and therefore some selection occurs, however, this is still inefficient. Thus, it would be worthwhile to test GCV solubility in different solvents, which might be less toxic to cells.

As an explanation for the failed selection, we hypothesized that the cells might expel the drug, however this has not been shown before and no mechanism is known for this behaviour. In a parallel MSc project in the laboratory, this method was successful for selecting cells in which the FRT-flanked selection cassette was removed by Flippase transfection, following introduction of a transcription terminator to block expression of the A3A gene in a bladder cancer cell line (BFTC). These

cells originally expressed the *puro-ΔTK* selection cassette in A3A intron 1, and were susceptible to low doses of GCV, allowing for quick negative selection of cells that still expressed the cassette. This would suggest that different cell lines expressing HSV-TK have different sensitivities to GCV. Perhaps, genomic instability and high proliferation rates of cancer cells, allows for an easier incorporation of phosphorylated GCV into DNA during replication, making cancer cells more vulnerable to GCV action when expressing HSV-TK. It would be interesting to understand the effect of GCV on other non-cancerous cell lines expressing HSV-TK, and how these results compare to NIKS. Another possibility for the failed selection using *puro-ΔTK* in our experiments, might be the presence of random mutations within the *TK* fusion gene. DNA sequencing of TK-NIKS, would allow us to gain a better understanding of the cells' genetic profile and what influences our results. Further investigation to explore how to conduct negative selection in NIKS expressing HSV-TK, is therefore required.

4.3 A3B deletion polymorphism in breast tissues and alternative ways to deliver CRISPR/Cas9 into cells.

A3B deletion polymorphism has been shown to increase deamination mutations compared to when the deletion was not present. This was mostly observed in breast tumours, with varying incidence among different populations (Long *et al.*, 2013; Chan *et al.*, 2015; Revathidevi *et al.*, 2016; Klonowska *et al.*, 2017; Gansmo *et al.*, 2018).

With this in mind, we thought that it would be interesting to reproduce the ~29.5kb deletion between A3A and A3B in non-cancerous epithelial breast cells; for this reason, we used the MCF10A cell line. The cells used in this project, present a stable tetracycline-inducible Cas9 expression, which is synthesised upon cell

treatment with tetracycline or its analogues, such as doxycycline. Despite not being able to establish the optimal DOX concentration, Cas9 is expressed at all concentrations tested (**Figure 18**). However, when the PCR products of the transfected cells where analysed on agarose gel, no bands indicative of the deletion were observed (lines 2-7 in **Figure 19**). Due to time restrictions, it was not possible to optimize and/or repeat this experiment. The results obtained could be due to several reasons: the designed gRNA might have not been specific enough; transfection efficiency might have been poor; Cas9 and gRNA were not in the right proportions to generate the desired deletion. While this approach did not yield the wanted results in this project, it has been proven to be a successful way to generate genomic mutations via CRISPR/Cas9 (Cao *et al.*, 2016). Off target activity it is also limited by using doxycycline-induced Cas9 expressing cells, due to the limited time the genome is exposed to Cas9 activity, while still enabling target activity (Cao *et al.*, 2016). As a result of the relative simplicity and advantages of this system, it would be valuable to further investigate how it can be exploited to create the desired ~29.5kb deletion in MCF10A.

Another technique to deliver CRISPR/Cas9 into cells is electroporation. This method creates small pores in the cell membranes, allowing CRISPR system to enter the cell. While this is a quick and easy technique, which does not discriminate between cell types, it is harsh on mammalian cells, which might find it difficult to recover (Lino *et al.*, 2018). However, due to the high throughput and speed of this technique, it will continue to be improved and utilized for research purposes (Lino *et al.*, 2018). Perhaps due to the cells being sensitive to the voltage and current used, we did not obtain the hoped results by using electroporation (section 3.9). An inactive form of

Cas9 was delivered to MFC10A and NIKS, at different concentrations; the transfection efficiency observed via fluorescence microscopy was very low with little, if any, difference between the different concentrations used (**Figure 20** and **Figure 21**). For this reason and due to time restriction, we decided not to take this work forward.

4.4 Current knowledge and future directions

Analysis on cell line and tumour tissue have shown upregulation of *A3B* compared to normal corresponding tissues, which correlates to increased mutations in cancer driver genes, such as *TP53* and *PIK3CA* (Cescon *et al.*, 2015). This seems to be associated with several types of breast cancers, as well as other types of carcinomas, as shown in previous studies (Nik-Zainal *et al.*, 2012; Burns *et al.*, 2013b; Burns *et al.*, 2013c; Roberts *et al.*, 2013; Alexandrov *et al.*, 2013; Sieuwerts *et al.*, 2017b). However, the processes and mechanisms that lead to the deregulation of A3B, subsequent increased expression and generation of APOBEC-signature mutations have not been identified yet. Depicted as the major cause of mutations seen in cancers, resulting in C>T transition, developing A3 inhibitors seems a tempting idea. Chemotherapy adjuvant or administered independently, the potential inhibitors, might decrease A3B-driven tumoral heterogeneity which contributes to resistance of existing therapies and decreased survival. The germline APOBEC3B deletion polymorphism surely supports this hypothesis, suggesting that A3B might not be essential, however, different considerations need to be made and further research is required.

Other members of the A3 family, such as A3A, A3H and A3G, cannot be fully excluded; furthermore, single nucleotide polymorphism (SNP) can also be

associated with increased APOBEC-signature mutation and cancer risk (Middlebrooks *et al.*, 2016). SNPs increase of bladder cancer risk might occur through A3B upregulation, which prevailed over A3A_B associated risk. The opposite was true in breast cancer, which was strongly modulated by A3A_B presence. Additionally, A3A expression was significantly higher in breast than in bladder cancer cells, when both were virally infected, indicating that cell and tissue specificity affect cancer formation (Middlebrooks *et al.*, 2016). High A3A was detected in A3A_B tumours which expressed high inflammation-related genes, causing continuous immune system activation and high A3A mutations that might not be detectable in later stages of tumour (Cescon *et al.*, 2015). Conversely, higher APOBEC mutagenesis, might contribute to neoantigens formation and more immune response. A3A_B tumours show increased immune infiltration in breast tumour and the phenotype associated with this mutation can be exploited for development of immunotherapy (Cescon *et al.*, 2015). This would benefit a large portion of the population, as over 40% presents the A3A_B polymorphism, although this might be an over-estimation.

A few recent studies have proposed A3A as the favoured A3 enzyme in creating cancer mutagenesis, when the 29.5kb deletion is present. These studies support the idea that the chimeric A3A_B gene formed accounts for the observed A3A increased action, as A3A_B is composed by the entire coding region of A3A and A3B 3' untranslated region (UTR) (Caval *et al.*, 2014; Chan *et al.*, 2015; Middlebrooks *et al.*, 2016). Perhaps A3B 3' UTR (or maybe a downstream enhancer of A3B), allows for a different regulation of the hybrid gene, stabilizing the transcript. *A3B* mRNA has been shown to be more abundant than *A3A,* therefore, the increased A3A seen in A3A_B cancers might be an effect of the A3B 3'UTR

regulation. However, these assumptions have been made using transfected reporters in model organisms with different genetic make-up than human cells and hence, different responses to genomic alterations.

Despite studies indicating A3A-induced mutations in a A3A_B background, *Starrett et al (2016)*, showed that A3H haplotype I (A 3H-I) might account for the APOBEC-signature mutations seen in breast and lung cancer when the deletion polymorphism in present. In fact, A3H-I is merely present when A3B is expressed, and A3H-I levels rise when A3B is absent (Starrett *et al.*, 2016). Moreover, the study showed no evidence of the *A3A_B* gene, when depleting A3B using CRISPR/Cas9 in MCF7 cells (Starrett *et al.*, 2016). This was in line with a previous study that was not able to detect A3A_B (Cescon *et al.*, 2015). Perhaps, both hybrid A3A and A3H-I might influence mutagenesis when A3B is deleted, though further studies that analyse this effect need to be carried out, as well as understanding the role of the *A3A_B* gene. However, none of the A3 enzymes can be fully excluded and no conclusion or direct links can be made, as it is extremely difficult to determine mutations in cancer and the exact time point at which they happened. The replica systems created in laboratories do not portray the true environment present within the body, thus unravelling the many processes involved in carcinogenesis is rather challenging.

Further research on the expression and activity of APOBEC3 enzymes in normal and tumour cells is essential to understand what drives their deregulation, which A3 generates mutations in tumours that develop in individuals homozygous for the A3A_B polymorphism, how and if the polymorphism can be exploited for generation of future therapies and what individuals might benefit from them.

4.5 Conclusion

APOBEC mutations might occur at different points throughout tumour development; A3B activity might be the leading cause of tumorigenesis and its action shapes the tumour throughout its lifetime; A3A might also cause mutations in cancers, but these are not seen in later stages and A3H might also be involved. On the A3A_B background, A3B is not expressed, which does not reduce mutation formation and cancer risk. A3A might transiently cause mutations at the beginning of tumour formation and/or A3H-I might drive tumorigenesis too. Therefore, APOBEC-signature mutations heavily shape and modulate tumour formation and progression and influence response to treatment. Furthermore, population background also influences A3 expression and role in cancer.

This study generates an isogenic model of the A3A_B deletion in a non-cancerous cell line (NIKS). This is a valuable tool which allows determination of the other APOBEC3 enzymes expression, and what role these might play in tumorigenesis. This prompts further studies to understand what are the processes and mechanisms that regulate A3 expression and if these are linked. Additionally, it provides a platform for novel therapies development, such as A3B inhibitors as well as understanding which individuals will benefit from them; perhaps it would be worth analysing the possibility of developing inhibitors that target enzymes such as A3A and A3H alone or in combination with A3B. More importantly this model, might aid the determination of what causes deregulation of A3 enzymes and what drives tumorigenic. Lastly, but surely not least important, it would be possible to study the implication of viral infections such as HPV, when the A3B deletion is present and what is the association with HPV-driven tumours.

5 References

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A. V., *et al.* (2013) Signatures of mutational processes in human cancer. *Nature* **500(7463)**: 415–421.

Baker, S.J., and Reddy, E.P. (2012) CDK4: A key player in the cell cycle, development, and cancer. *Genes and Cancer* .

Barretina, J., Taylor, B.S., Banerji, S., Ramos, A.H., Lagos-Quintana, M., DeCarolis, P.L., *et al.* (2010) Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nat Genet* **42**: 715–721.

Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J. V., and Cullen, B.R. (2006) Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci U S A* **103**: 8780–8785.

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

Borrelli, E., Heyman, R., Hsi, M., and Evans, R.M. (1988) Targeting of an inducible toxic phenotype in animal cells. *Proc Natl Acad Sci* .

Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., *et al.* (2013a) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* .

Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., *et al.* (2013b) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**: 366–370.

Burns, M.B., Leonard, B., and Harris, R.S. (2015) APOBEC3B: Pathological consequences of an innate immune DNA mutator. *Biomed J* **38**: 102–110.

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

Cao, J., Wu, L., Zhang, S.M., Lu, M., Cheung, W.K.C., Cai, W., *et al.* (2016) An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Res* **44**: e149.

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

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

Chan, K., Roberts, S.A., Klimczak, L.J., Sterling, J.F., Saini, N., Malc, E.P., *et al.* (2015) An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet* .

Chen, T.W., Lee, C.C., Liu, H., Wu, C.S., Pickering, C.R., Huang, P.J., *et al.* (2017) APOBEC3A is an oral cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion polymorphism. *Nat Commun* .

Chen, X., Xu, F., Zhu, C., Ji, J., Zhou, X., Feng, X., and Guang, S. (2014) Dual sgRNA-directed gene knockout using CRISPR/Cas9 technology in Caenorhabditis elegans. *Sci Rep* **4**: 1–7.

Chen, Y., and Bradley, A. (2000) TECHNOLOGY REPORT A New Positive / Negative Selectable Marker , pu tk , for Use in Embryonic Stem Cells. *Curr Biol* **35**: 31–35.

Cong, L., and Zhang, F. (2015) Genome engineering using CRISPR-Cas9 system. *Methods Mol Biol* **1239**: 197–217.

Davies, C., Pan, H., Godwin, J., Gray, R., Arriagada, R., Raina, V., *et al.* (2013) Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet*

Day, P.J., Cleasby, A., Tickle, I.J., O'Reilly, M., Coyle, J.E., Holding, F.P., *et al.* (2009) Crystal structure of human CDK4 in complex with a D-type cyclin. *Proc Natl Acad Sci* .

Faltas, B.M., Prandi, D., Tagawa, S.T., Molina, A.M., Nanus, D.M., Sternberg, C., *et al.* (2016a) Clonal evolution of chemotherapy-resistant urothelial carcinoma. *Nat Genet* .

Faltas, B.M., Prandi, D., Tagawa, S.T., Molina, A.M., Nanus, D.M., Sternberg, C., *et al.* (2016b) Clonal evolution of chemotherapy-resistant urothelial carcinoma. *Nat Genet* **48**: 1490–1499.

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**: 118–124.

Göhler, S., Silva Filho, M.I. Da, Johansson, R., Enquist-Olsson, K., Henriksson, R., Hemminki, K., *et al.* (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. *J Cancer Res Clin Oncol* .

Greco, R., Oliveira, G., Lupo Stanghellini, M.T., Vago, L., Bondanza, A., Peccatori, J., *et al.* (2015) Improving the safety of cell therapy with the TK-suicide gene. *Front Pharmacol* .

Han, J., Zhang, J., Chen, L., Shen, B., Zhou, J., Hu, B., *et al.* (2014) Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9. *RNA Biol* **11**: 829–835.

Harari, A., Ooms, M., Mulder, L.C.F., and Simon, V. (2009) Polymorphisms and Splice Variants Influence the Antiretroviral Activity of Human APOBEC3H. *J Virol* .

Harrison, P.T., Thompson, R., and Davidson, A.J. (1991) Evolution of herpesvirus thymidine kinases from cellular deoxycytidine kinase. *J Gen Virol* .

Hatch, E.M. (2018) Nuclear envelope rupture: little holes, big openings. *Curr Opin Cell Biol* **52**: 66–72.

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

Henderson, S., and Fenton, T. (2015) APOBEC3 genes: Retroviral restriction factors to cancer drivers. *Trends Mol Med* .

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

Hossain, J.A., Riecken, K., Miletic, H., and Fehse, B. (2019) Cancer suicide gene therapy with TK.007. In *Methods in Molecular Biology*. .

Jha, P., Sinha, S., Kanchan, K., Qidwai, T., Narang, A., Singh, P.K., *et al.* (2012) Deletion of the APOBEC3B gene strongly impacts susceptibility to falciparum malaria. *Infect Genet Evol*

Kanu, N., Cerone, M.A., Goh, G., Zalmas, L.P., Bartkova, J., Dietzen, M., *et al.* (2016) DNA

replication stress mediates APOBEC3 family mutagenesis in breast cancer. *Genome Biol* .

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

Klonowska, K., Kluzniak, W., Rusak, B., Jakubowska, A., Ratajska, M., Krawczynska, N., *et al.* (2017) The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population. *Oncotarget* **8**: 76357–76374.

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

Lackey, L., Demorest, Z.L., Land, A.M., Hultquist, J.F., Brown, W.L., and Harris, R.S. (2012) APOBEC3B and AID have similar nuclear import mechanisms. *J Mol Biol* .

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

Law, E.K., Sieuwerts, A.M., Lapara, K., Leonard, B., Starrett, G.J., Molan, A.M., *et al.* (2016a) The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci Adv* .

Law, E.K., Sieuwerts, A.M., Lapara, K., Leonard, B., Starrett, G.J., Molan, A.M., *et al.* (2016b) The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci Adv* **2**.

Leonard, B., Hart, S.N., Burns, M.B., Carpenter, M.A., Temiz, N.A., Rathore, A., *et al.* (2013) APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer Res* .

Li, M.M.H., and Emerman, M. (2011) Polymorphism in Human APOBEC3H Affects a Phenotype Dominant for Subcellular Localization and Antiviral Activity. *J Virol* .

Lino, C.A., Harper, J.C., Carney, J.P., and Timlin, J.A. (2018) Delivering crispr: A review of the challenges and approaches. *Drug Deliv* .

Long, J., Delahanty, R.J., Li, G., Gao, Y.T., Lu, W., Cai, Q., *et al.* (2013) A common deletion in the APOBEC3 genes and breast cancer risk. *J Natl Cancer Inst* .

Lovin, N., and Peterlin, B.M. (2009) APOBEC3 proteins inhibit LINE-1 retrotransposition in the absence of ORF1p binding. In *Annals of the New York Academy of Sciences*. pp. 268–275.

Maciejowski, J., Li, Y., Bosco, N., Campbell, P.J., and Lange, T. De (2015) Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell* **163**: 1641–1654.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* .

Marouf, C., Göhler, S., Filho, M.I.D.S., 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* .

McCann, J.M., Klein, M.M., Leland, E.M., Law, E.K., Brown, W.L., Salamango, D.J., and Harris, R.S. (2019) The DNA deaminase APOBEC3B interacts with the cell cycle protein CDK4 and disrupts CDK4-mediated nuclear import of Cyclin D1. *J Biol Chem* .

McGranahan, N., Favero, F., Bruin, E.C. De, Birkbak, N.J., Szallasi, Z., and Swanton, C. (2015a) Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med* .

McGranahan, N., Favero, F., Bruin, E.C. De, Birkbak, N.J., Szallasi, Z., and Swanton, C. (2015b) Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med* **7**.

Mehta, A., Kinter, M.T., Sherman, N.E., and Driscoll, D.M. (2000) Molecular Cloning of Apobec-1 Complementation Factor, a Novel RNA-Binding Protein Involved in the Editing of Apolipoprotein B mRNA. *Mol Cell Biol* **20**: 1846–1854.

Middlebrooks, C.D., Banday, A.R., Matsuda, K., Udquim, K.I., Onabajo, O.O., Paquin, A., *et al.* (2016) Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. *Nat Genet* .

Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Löwer, J., Cichutek, K., *et al.* (2006) APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem* .

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

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Loo, P. Van, Greenman, C.D., Raine, K., *et al.* (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell* .

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

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

OhAinle, M., Kerns, J.A., Li, M.M.H., Malik, H.S., and Emerman, M. (2008) Antiretroelement Activity of APOBEC3H Was Lost Twice in Recent Human Evolution. *Cell Host Microbe* **4**: 249–

259.

Pak, V., Heidecker, G., Pathak, V.K., and Derse, D. (2011) The Role of Amino-Terminal Sequences in Cellular Localization and Antiviral Activity of APOBEC3B. *J Virol* **85**: 8538–8547.

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

Revathidevi, S., Manikandan, M., Rao, A.K.D.M., Vinothkumar, V., Arunkumar, G., Rajkumar, K.S., *et al.* (2016) Analysis of APOBEC3A/3B germline deletion polymorphism in breast, cervical and oral cancers from South India and its impact on miRNA regulation. *Tumor Biol* **37**: 11983–11990.

Rezaei, M., Hashemi, M., Hashemi, S.M., Mashhadi, M.A., and Taheri, M. (2015) APOBEC3 Deletion is Associated with Breast Cancer Risk in a Sample of Southeast Iranian Population. *Int J Mol Cell Med* **4**: 103–8.

Richardson, S.R., Narvaiza, I., Planegger, R.A., Weitzman, M.D., and Moran, J. V. (2014) APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *Elife* **3**.

Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., *et al.* (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human

cancers. *Nat Genet* .

Roberts, S.A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., *et al.* (2012) Clustered Mutations in Yeast and in Human Cancers Can Arise from Damaged Long Single-Strand DNA Regions. *Mol Cell* **46**: 424–435.

Salter, J.D., Bennett, R.P., and Smith, H.C. (2016) The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem Sci* .

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

Shinohara, M., Io, K., Shindo, K., Matsui, M., Sakamoto, T., Tada, K., *et al.* (2012) APOBEC3B can impair genomic stability by inducing base substitutions in genomic DNA in human cells. *Sci Rep* .

Sieuwerts, A.M., Schrijver, W.A.M.E., Dalm, S.U., Weerd, V. De, Moelans, C.B., Hoeve, N. Ter, *et al.* (2017a) Progressive APOBEC3B mRNA expression in distant breast cancer metastases. *PLoS One* .

Sieuwerts, A.M., Schrijver, W.A.M.E., Dalm, S.U., Weerd, V. De, Moelans, C.B., Hoeve, N. Ter, *et al.* (2017b) Progressive APOBEC3B mRNA expression in distant breast cancer metastases. *PLoS One* **12**.

Sieuwerts, A.M., Willis, S., Burns, M.B., Look, M.P., Gelder, M.E.M. Van, Schlicker, A., *et al.* (2014) Elevated APOBEC3B Correlates with Poor Outcomes for Estrogen-Receptor-Positive Breast Cancers. *Horm Cancer* **5**: 405–413.

Smith, N.J., and Fenton, T.R. (2019) The APOBEC3 genes and their role in cancer: Insights from human papillomavirus. *J Mol Endocrinol* **62**: R269–R287.

Song, Y., Lai, L., and Li, Z. (2017) Large-scale genomic deletions mediated by CRISPR/Cas9 system. *Oncotarget* **8**: 5647–5647.

Song, Y., Yuan, L., Wang, Y., Chen, M., Deng, J., Lv, Q., *et al.* (2016) Efficient dual sgRNA-directed large gene deletion in rabbit with CRISPR/Cas9 system. *Cell Mol Life Sci* **73**: 2959–2968.

Starrett, G.J., Luengas, E.M., McCann, J.L., Ebrahimi, D., Temiz, N.A., Love, R.P., *et al.* (2016) The DNA cytosine deaminase APOBEC3H haplotype i likely contributes to breast and lung cancer mutagenesis. *Nat Commun* .

Stenglein, M.D., and Harris, R.S. (2006) APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J Biol Chem* **281**: 16837–16841.

Stopak, K., Noronha, C. De, Yonemoto, W., and Greene, W.C. (2003) HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability.

Mol Cell .

Suspene, R., Aynaud, M.-M., Guetard, D., Henry, M., Eckhoff, G., Marchio, A., *et al.* (2011) Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc Natl Acad Sci* .

Suspène, R., Mussil, B., Laude, H., Caval, V., Berry, N., Bouzidi, M.S., *et al.* (2016) Self-cytoplasmic DNA upregulates the mutator enzyme APOBEC3A leading to chromosomal DNA damage. *Nucleic Acids Res* .

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

Vartanian, J.P., Guétard, D., Henry, M., and Wain-Hobson, S. (2008) Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science (80-)*

Vartanian, J.P., Henry, M., Marchio, A., Suspène, R., Aynaud, M.M., Guétard, D., *et al.* (2010a) Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog* **6**: 1–9.

Vartanian, J.P., Henry, M., Marchio, A., Suspène, R., Aynaud, M.M., Guétard, D., *et al.* (2010b) Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog* .

Vasudevan, A.A.J., Kreimer, U., Schulz, W.A., Krikoni, A., Schumann, G.G., Häussinger, D., *et al.* (2018) APOBEC3B activity is prevalent in urothelial carcinoma cells and only slightly

affected by LINE-1 expression. *Front Microbiol* .

Vieira, V.C., and Soares, M.A. (2013) The Role of Cytidine Deaminases on Innate Immune Responses against Human Viral Infections. *Biomed Res Int* **2013**: 1–18.

Walker, B.A., Wardell, C.P., Murison, A., Boyle, E.M., Begum, D.B., Dahir, N.M., *et al.* (2015) APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nat Commun* .

Wang, Z., Wakae, K., Kitamura, K., Aoyama, S., Liu, G., Koura, M., *et al.* (2014) APOBEC3 Deaminases Induce Hypermutation in Human Papillomavirus 16 DNA upon Beta Interferon Stimulation. *J Virol* .

Wen, W.X., Soo, J.S.S., Kwan, P.Y., Hong, E., Khang, T.F., Mariapun, S., *et al.* (2016) Germline APOBEC3B deletion is associated with breast cancer risk in an Asian multi-ethnic cohort and with immune cell presentation. *Breast Cancer Res* .

Xuan, D., Li, G., Cai, Q., Deming-Halverson, S., Shrubsole, M.J., Shu, X.O., *et al.* (2013) APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis* .

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. *Int J Clin Exp Pathol* .