Investigating the diverse cellular functions of APOBEC3A and APOBEC3B in Urothelial Carcinoma

Fiona Alice Want

The Institute of Cancer Research

University of London

This thesis is submitted for the degree of PhD

Declaration

I confirm that the work presented in this thesis is my own. The use of materials and support from other sources has been fully acknowledged throughout the text.

Fiona Alice Want

Abstract

Urothelial carcinoma (UCC) is characterised by a high mutational burden, with many of the identified mutations resulting from the mutagenic activity of the APOBEC3 (A3) family of cytidine deaminases. *A3B* expression is frequently elevated in tumours and correlates with mutational load, suggesting that A3B is the primary driver of these A3 mutations. However, despite *A3A* expression being low and not elevated relative to normal tissue, several studies have shown that it is the more dominant mutagen. Experimental evidence suggests that mutations are not accumulated at a constant rate but are instead generated in bursts, and this may explain the disconnect between expression and the A3-signature. Anticancer agents with diverse mechanisms of action induce *A3A/B* expression in UCC cells, and this is likely driving episodic mutagenesis in patients. Investigation into the mechanism of induction revealed that *A3A* and *A3B* are differentially regulated in response to gemcitabine treatment in UCC cells; *A3A* induction is mediated by NF-κB while *A3B* induction occurs via ATR activation in response to replication stress. Expression of both *A3A* and *A3B* is attenuated with receptor tyrosine kinase inhibitors, highlighting that induction can be curbed with clinically available drugs. RNA sequencing revealed that acute expression of A3A and A3B has subtle effects on gene expression, but the identification of putative RNA editing events provides initial evidence that A3B, like A3A, is an RNA editing enzyme, and suggests that A3A/B may be driving adaptability independently of genomic alteration. While A3B can cause DNA damage, a proliferation defect and increase sensitivity to DNA damage response inhibitors in several cell lines, acute A3B expression had no effect on growth or drug sensitivity in UCC cells suggesting they tolerate transient elevated expression. Finally, A3A's interacting partners were identified using the proximity-labelling technique, BioID, and this revealed interactions with R-loop-binding proteins, suggesting that A3A has a novel role in R-loop homeostasis.

Acknowledgements

Firstly, I would like to thank Cancer Research UK and The Institute of Cancer Research for funding and supporting my PhD studies.

A special thanks goes to my primary supervisor, Dr Olivia Rossanese. I am eternally grateful to Olivia for the chance to work on such an interesting and challenging project, for all her training and mentoring over the years and for being an inspiring advocate for all women in STEM. I am especially thankful to her for giving up her time on evenings and weekends to read my thesis, and always offering support and feedback, no matter how hectic things are for her.

A huge thanks goes to my associate supervisor Dr Mike Walton, a source of never-ending quotes and wisdom. He has been in the lab with me through the whole project, the good days and the bad, and I cannot thank him enough for all his support and encouragement over the years. I am especially thankful for his ongoing support and help with my thesis even after he had retired.

Thank you to everyone who has helped me at the ICR with my work. To those who showed me how to use equipment, gave feedback on presentations and posters, have taken time out their day to teach me a new technique and share their knowledge; I will always be grateful for your help in developing my skills as a researcher. Thank you to members of the APOBEC project team for showing me how to work as part of a multidisciplinary drug discovery project, especially to the biology sub-team, Mike, Caitlin, Sam, Paul C and Chi for all the interesting discussions about APOBEC that helped me generate new ideas for my own work.

Thank you to Dr Pradeep Ramagir and Dr Konstantinos Mitsopoulos for all their work on my RNA sequencing dataset and for taking the time to explain everything to me in a way a non-bioinformatician can understand. A special thanks goes to Pradeep for all the time he spent analysing my data, especially for the time he

spent establishing the RNA editing pipeline. A big thank you also goes to Dr Lu Yu and Dr Theo Roumeliotis for conducting the mass spectrometry and analysing the BioID data.

Thank you to all members of the Target Evaluation and Molecular Therapeutics team, past and present. I've had so many good times with you all, from chats in the office or the lab, to when we've been out and about for team socials. My PhD would not have been the same without you. A special mention to Caitlin, LeAnne, Sam, and Brad. Caitlin and LeAnne, you've been there from the start and instantly made me feel part of the team, always being there for me for whatever I needed. The ICR lads, Sam and Brad, thank you for all the coffee that helped me through the long days, and the banter that never failed to make me laugh and cheer me up on the toughest of days. Sam, we've only known each other for 2 years but I know I've made a friend for life.

I must also thank everyone who helped me when I had my accident and broke my leg at the end of my second year. To everyone who checked in on me during the four months I was stuck at home, to LeAnne and Caitlin who bought me a care package, Olivia and everyone in registry who sorted my interruption of studies and made things as easy as possible for me while I was still in hospital, and my colleagues who helped me around the lab while I was still recovering. You are all the best!

A huge thank you to all my family and friends for the support over the years, their love, and faith in me, especially my parents and sister. Mum, thank you for always being there to help in any way you can to make things easier for me and dad, I'm forever grateful for you teaching me how to properly format large documents on Word, it definitely helped with the thesis. Thank you to my sister, who is the kindest person I will ever know and is there for me whenever I need her. I'm very lucky to have you all encouraging me to do my best and follow my interests.

And finally, thank you to my wonderful husband, Jack. I will forever be in debt to you for your support over the past 9 and a half years, for sticking by me through it all, and for always believing in me, even when I do not believe in myself. None of this would have been possible without you.

Table of contents

Table of contents **6**

List of figures **12**

List of tables **18**

Abbreviations **20**

Chapter 1 Introduction **33**

1.1 Hallmarks of cancer 33

1.1.1 Inflammation and cancer 33

1.2 Cancer, heterogeneity, and evolution 37

1.2.1 Mutational signatures 38

1.3 APOBEC enzymes 40

1.4 APOBEC: endogenous mutagens 43

1.4.1 Uracil excision 44

1.4.2 Repairing DNA breaks 46

1.4.3 The DNA damage response 47

1.4.4 APOBEC3A or B? 49

1.4.5 Subcellular localisation 53

1.4.6 Substrate availability 54

1.5 Roles of A3A and A3B in cancer 58

1.6 Urothelial Carcinoma 60

1.6.1 APOBECs in UCC 63

1.7 Thesis aims 66

Chapter 2 Materials and methods **69**

2.1 Cell lines 69

2.1.1 Culture conditions 69

2.1.2 Passaging 70

2.1.3 Cryopreservation 70
2.1.4 Doubling time 70
2.1.5 Optimal seeding density 70
2.2 Sulforhodamine B (SRB) assay 71
2.3 Incucyte growth assay 71
2.4 Compounds 71
2.5 Inhibitory Concentration (IC) determination 72
2.6 Treatments 73
2.7 Small interfering RNA (siRNA) transfection 73
2.8 Cell lysis 74
2.8.1 Gel-based APOBEC deamination assay 74
2.8.2 Western Blot 74
2.9 Determination of protein concentration 74
2.10 Gel-based APOBEC deamination assay 75
2.11 Western blotting 75
2.11.1 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) 75
2.11.2 Transfer 76
2.11.3 Detection 76
2.12 RNA isolation 77
2.13 Reverse Transcription and quantitative PCR (RT-qPCR) 77
2.14 Statistical analysis 78
2.14.1 qPCR statistical analysis 79
2.15 Polymerase Chain Reaction (PCR) 79
2.16 NEBuilder® HiFi DNA Assembly 80
2.17 Restriction Enzyme Digestion 81
2.18 Ligation 81
2.19 Isopropanol precipitation 82

2.20 Transformation 82
2.20.1 Colony PCR 82
2.20.2 Starter cultures and stocks 83
2.20.3 Bacterial Stocks 83
2.21 Plasmid extraction 83
2.21.1 Miniprep 83
2.21.2 Maxiprep 83
2.22 Plasmid Site-directed mutagenesis 84
2.23 Sequencing 84
2.24 Software 84
2.25 Mammalian cell genomic DNA extraction 84
2.26 Transient transfection 84
2.27 Generating stable cell lines 84
2.27.1 Linear DNA donors 85
2.27.2 Antibiotic selection 85
2.28 BioID identification of interacting partners 86
2.28.1 Cell lysis 86
2.28.2 Affinity purification of biotinylated proteins 86
2.28.3 Western blot 87
2.28.4 Sample preparation 87
2.28.5 Tandem Mass Tag Liquid Chromatography Mass Spectrometry (TMT-LC-MS/MS) 88
2.28.6 Statistical analysis 88
2.29 RNA sequencing 88
2.29.1 QC, read alignment and count generation 89
2.29.2 Differential expression and pathway enrichment analysis 89
2.29.3 RNA editing analysis 89

Chapter 3 Chemotherapy induces expression of *A3A* and *A3B* in UCC.... 92

3.1 Introduction.... 92

3.1.1 APOBEC3s are induced in response to cellular stress.... 92

3.1.2 Chapter aims.... 94

3.2 Results.... 96

3.2.1 Initial characterisation of a human UCC cell line panel.... 96

3.2.2 Genotoxic stress induces expression of *A3A* and *A3B* independent of p53 status.... 110

3.2.3 Drug-induced increases in deamination activity correlate with *A3B* expression.... 116

3.2.4 *A3A* and *A3B* induction is not a normal response in urothelium 119

3.2.5 *A3B* induction is concentration-dependent.... 120

3.3 Discussion.... 125

Chapter 4 *A3A* and *A3B* are differentially regulated in UCC.... 130

4.1 Introduction.... 130

4.1.1 Chapter aims.... 134

4.2 Results.... 135

4.2.1 Chemotherapy drugs induce expression of the canonical NF-κB target gene, *TNFα*.... 135

4.2.2 Inhibition of PKC attenuates gemcitabine-mediated induction of *A3A, A3B* and *TNFα*.... 136

4.2.3 Inhibition of NF-kB signalling attenuates gemcitabine-induced expression of *A3A* but not *A3B*.... 138

4.2.4 Induction of *A3A* is dependent on both canonical and alternative NF-kB signalling.... 140

4.2.5 STING inhibition does not attenuate gemcitabine-mediated induction.... 144

4.2.6 ATR inhibition attenuates induction of *A3B* but not *A3A*.... 146

4.2.7 RTKi treatment attenuates induction of *A3A, B* and *TNFα* 148

4.3 Discussion 153

Chapter 5 Establishing APOBEC3 cell line models 161

5.1 Introduction 161

5.2 Results 168

5.2.1 A3B, but not A3A, biotin-ligase fusions are toxic to *E. coli* 168

5.2.2 Disruption of the *A3B* ORF prevents expression of toxic A3B fusions enabling cloning in *E. coli* 171

5.2.3 Sub-cloning into the AAVS1-TetON system 173

5.2.4 A3A and A3B expression cell line models 175

5.2.5 Preliminary *AAVS1* integration 176

5.2.6 Validation of A3A miniTurbo constructs 177

5.2.7 *AAVS1* targeting using plasmid donors 179

5.2.8 Optimisation of *AAVS1* integration 182

5.2.9 Stable cell line troubleshooting 189

5.2.10 Randomly integrated stables for further studies 192

5.3 Discussion 195

Chapter 6 The cellular response to acute A3A and A3B exposure in UCC 201

6.1 Introduction 201

6.1.1 A3B is required for ER gene transcription in breast cancer 201

6.1.2 A3A is an RNA editing enzyme 203

6.1.3 A3A and A3B cause DNA damage, sensitising cells to DDR inhibition 205

6.1.4 Chapter aims 206

6.2 Results 207

6.2.1 Elevated A3A and A3B expression has subtle effects on the transcriptome 207

6.2.2 A3A and A3B are putative RNA editing enzymes in UCC.....219
6.2.3 Acute A3B exposure does not affect cell growth or drug sensitivity 221
6.2.4 Identification of interacting partners using BioID 224
6.3 Discussion 236
Chapter 7 General discussion 246
7.1 Introduction.....246
7.2 Regulation of *A3A* and *A3B* in UCC 248
7.3 Cellular response to transient exposure.....251
7.4 Are A3s drug targets? 254
7.5 Future studies 256
7.6 Concluding comments 258
References.....260
Appendix.....285
Primers 285
Plasmids.....290

List of figures

Figure 1.1 NF-kB signalling. 36

Figure 1.2 Intratumour heterogeneity and evolution. 38

Figure 1.3 APOBEC3 family members. 41

Figure 1.4 A3-mediated mutational mechanisms. 44

Figure 1.5 The DNA damage response (DDR). 49

Figure 1.6 Expression of *APOBEC3A* (A) and *APOBEC3B* (B) in tumour and normal samples 51

Figure 1.7 Proposed A3 activity on double-strand breaks (DSBs) 56

Figure 1.8 Proposed A3 activity on R-loops. 57

Figure 1.9 Proposed A3 activity on the replication fork. 58

Figure 1.10 Urothelial carcinoma (UCC) stages and key molecular subtypes. . 61

Figure 2.1 GATK best practices for variant analysis of RNA sequencing data sets. 90

Figure 3.1 Growth curves for the UCC cell line panel. 98

Figure 3.2 Seeding density determinations for the UCC cell line panel. 99

Figure 3.3 Functional p53 response to gemcitabine treatment in the UCC cell line panel 101

Figure 3.4 Endogenous *A3A* and *A3B* expression in a panel of 12 UCC cell lines. 102

Figure 3.5 Deamination activity detected in lysates from the UCC panel is primarily attributed to A3B. 104

Figure 3.6 Concentration-response curves for cisplatin IC_{50} determination of 10 UCC cell lines. 106

Figure 3.7 Concentration-response curves for gemcitabine IC_{50} determination of 10 UCC cell lines. 107

Figure 3.8 Concentration-response curves for paclitaxel IC_{50} determination of 10 UCC cell lines. 108

Figure 3.9 Heat map summary of IC_{50} values for cisplatin, gemcitabine, and paclitaxel for 10 cell lines in the UCC panel. 109

Figure 3.10 Induction of *A3B* in UCC cell lines in response to chemotherapeutics is not dependent on p53 status, disease type or molecular subtype. 112

Figure 3.11 Induction of *A3A* in UCC cell lines in response to chemotherapeutics is not dependent on p53 status, disease type or molecular subtype. 114

Figure 3.12 Ionising radiation only modestly induces *A3A/B* expression in BFTC-905 cells. 115

Figure 3.13 Commonly used anticancer agents cause an increase in A3 deamination activity. 117

Figure 3.14 *A3B* mRNA induction correlates with A3 activity in response to chemotherapeutics. 118

Figure 3.15 Chemotherapeutics do not induce *A3A* or *A3B* mRNA expression in a 'normal' immortalised urothelial cell line. 119

Figure 3.16 Concentration-response curves showing sensitivity of UCC cell lines, BFTC-905 and SW780, to gemcitabine and bleomycin. 121

Figure 3.17 Concentration-dependent increases in *A3B* in response to gemcitabine and bleomycin. 123

Figure 4.1 Receptor tyrosine kinase signalling. 132

Figure 4.2 Gemcitabine and bleomycin treatment induces expression of the NF-κB target gene, *TNFα.* 135

Figure 4.3 Protein Kinase C (PKC) inhibition attenuates gemcitabine-induction of *A3A, A3B* and *TNFα*. 137

Figure 4.4 Small molecule inhibition of NF-κB signalling attenuates gemcitabine-induction of *A3A* and *TNFα* but not *A3B.* 139

Figure 4.5 Validation of siRNA knockdown of NF-κB transcription factors, RelA (p65) and RelB. 141

Figure 4.6 siRNA knockdown of NF-κB transcription factors p65 and RelB attenuates gemcitabine induction of *A3A* but not *A3B.* 143

Figure 4.7 Inhibition of STING with the inhibitor, H-151, does not prevent gemcitabine-induced *A3A*, *A3B* or *TNFα* expression in BFTC-905 cells.....145

Figure 4.8 ATR inhibition attenuates gemcitabine-induction of *A3B* but not *A3A*. 147

Figure 4.9 Concentration-response curves showing sensitivity of UCC cell lines, BFTC-905 and SW780, to receptor tyrosine kinase inhibitors. 150

Figure 4.10 Receptor tyrosine kinase inhibitor (RTKi) treatment partially attenuates gemcitabine-induction of *A3A, A3B* and *TNFα* when used in combination for 48 hours. 151

Figure 4.11 *A3A* and *A3B* transcriptional regulation in response to genotoxic stress in UCC. 158

Figure 5.1 CRISPR/Cas9 system for integration of transgenes into the *AAVS1* locus. 166

Figure 5.2 miniTurbo/TurboID-tagged A3B is toxic to *E. coli* and recombinants were not recovered. 169

Figure 5.3 Successful ligation of miniTurbo tagged A3B and GFP into pTRIPZ. 170

Figure 5.4 miniTurbo-tagged A3A is not toxic to *E. coli*. 171

Figure 5.5 Disruption of the *A3B* open reading frame facilitates cloning of miniTurbo-tagged A3B. 172

Figure 5.6 Subcloning miniTurbo-tagged A3A and A3B into the AAVS1_TetON vector. 174

Figure 5.7 Successful stuffer removal and open reading frame restoration of A3B constructs. 175

Figure 5.8 *AAVS1* knock-in by homology-directed repair from a donor plasmid. 177

Figure 5.9 Doxycycline induces expression of miniTurbo-tagged A3A, and addition of biotin induces biotinylation of interacting partners. 178

Figure 5.10 Pre-selection integration screening of stable cell line pools shows very low frequency of integration into the *AAVS1* locus. 179

Figure 5.11 Expression of the A3B-miniTurbo-V5 (A), 3xHA-miniTurbo-A3A (B) and A3A-miniTurbo-V5 (C) is doxycycline concentration dependent. 180

Figure 5.12 Stable cell lines made using plasmid DNA as a donor have randomly integrated the transgene into the genome and are not *AAVS1* knock-ins. 181

Figure 5.13 Doxycycline treatment induces expression of 3xHA-miniTurbo-A3B but not 3xHA-miniTurbo-GFP in transiently transfected BFTC-905 cells. 182

Figure 5.14 Optimisation of long ssDNA production using the TakaraBio Guide-it Long ssDNA Production System v2. 183

Figure 5.15 dsDNA and ssDNA transgene templates for transfection. 184

Figure 5.16 PCR screening of BFTC-905 stable cell lines made with ssDNA HDR templates. 185

Figure 5.17 *AAVS1* integration of the V5 tagged A3A and A3B transgenes is seen in all BFTC-905 stable cell line pools made with dsDNA HDR templates. 186

Figure 5.18 *AAVS1* integration of the miniTurbo-tagged transgenes is seen in only one BFTC-905 stable cell line pool made with dsDNA HDR templates. .. 187

Figure 5.19 One AAVS1-BFTC-905 stable cell line pool made with dsDNA HDR donor templates expresses after induction with doxycycline for 48 hours. 188

Figure 5.20 The lack of expression seen in the *AAVS1*-BFTC-905 stable cell lines is not due to epigenetic silencing of the transgene promoter. 190

Figure 5.21 Exogenous supply of the Tet3G transactivator protein via transient transfection followed with doxycycline treatment induces protein expression in *AAVS1*-BFTC-905 stable cell lines. 192

Figure 5.22 Expression of V5-tagged A3B is doxycycline concentration-dependent in BFTC_A3B cells without transient Tet3G transfection. 193

Figure 5.23 Expression of V5-tagged A3A is doxycycline concentration-dependent in BFTC_A3A cells without transient Tet3G transfection. 194

Figure 6.1 A3B-mediated transcription of ER target genes. 203

Figure 6.2 Western blot to check expression of A3A*, A3A, A3B** and A3B** in response to doxycycline after Tet3G transient transfection prior to sending samples for RNAseq. 208

Figure 6.3 Principal component analysis (PCA) plot for 30 samples sent for RNAseq. 210

Figure 6.4 MA-plots showing transcriptional changes in all stable cell line pools transfected with the Tet3G transactivator plasmid and protein expression induced with doxycycline. 211

Figure 6.5 Gene set enrichment analysis (GSEA) of differentially expressed genes in Tet3G transactivator plasmid transfected cells (+Dox) vs mock transfected (+ Dox) for all stable cell lines. 213

Figure 6.6 Acute expression of A3A and A3B has modest effects on the transcriptome. 215

Figure 6.7 Top 100 differentially expressed genes in response to acute A3A expression. 216

Figure 6.8 The 62 differentially expressed genes in response to acute A3B expression. 217

Figure 6.9 Venn diagram showing common and unique differentially expressed genes between A3A-V5 and A3B-V5 expressing cells. 218

Figure 6.10 DE analysis between wild-type and mutant (catalytically inactive) A3A and A3B. 219

Figure 6.11 Acute A3B exposure does not affect growth in BFTC_A3B cells. 222

Figure 6.12 Acute A3B exposure does not affect sensitivity to two cytotoxic drugs or two DDR inhibitors in BFTC_A3B cells. 223

Figure 6.13 BioID workflow. 225

Figure 6.14 100 ng/mL and 10 ng/mL doxycycline provide low, matched expression levels of the A3A bait proteins. 226

Figure 6.15 Biotinylation time optimisation of three stable cell lines expressing A3B-miniTurbo-V5 (A), 3xHA-miniTurbo-A3A (B) and A3A-miniTurbo-V5 (C). 227

Figure 6.16 Affinity purification of biotinylated proteins using Strep-Tactin® Sepharose® beads shows self-biotinylation of the bait proteins in lysates. 229

Figure 6.17 A3A's interactome is enriched with RNA binding proteins. 233

Figure 6.18 Top statistically significant A3A-miniTurbo-V5 interactors.....234

List of tables

Table 2.1 All compounds used with source, stock, and storage information. 71

Table 2.2 siRNA information. 73

Table 2.3 Optimised siRNA transfection conditions. 73

Table 2.4 Antibodies used for immunoblotting. 77

Table 2.5 TaqMan assays (Applied Biosystems, USA) used in RT-qPCR. 78

Table 2.6 Typical PCR reaction mixture. 80

Table 2.7 Typical PCR cycle conditions. 80

Table 2.8 Typical restriction enzyme digestion reaction mixture. 81

Table 2.9 Typical ligation reaction mixture and modified mixture for stuffer removal of $A3B^{Split}$ constructs. 81

Table 2.10 Typical reaction mixture for colony PCR. 82

Table 2.11 Typical cycle conditions for colony PCR. 83

Table 2.12 Concentration of puromycin and blasticidin used for selection of UCC lines. 85

Table 3.1 UCC cell line panel information. 96

Table 3.2 Doubling times and optimal seeding densities of the UCC cell line panel. 97

Table 3.3 Summary of IC_{50} values for cisplatin, gemcitabine, and paclitaxel for 10 cell lines in the UCC panel. 109

Table 3.4 Key characteristics of UCC cell lines used for investigation of induction in response to chemotherapy drugs. 110

Table 3.5 Summary of concentrations used (nM). G, gemcitabine; B, bleomycin. 122

Table 3.6 *P* values obtained from Tukey's multiple comparisons following a repeated-measures one-way ANOVA. 124

Table 4.1 *P* values obtained from Sidak's multiple comparisons following a repeated-measures two-way ANOVA. 146

Table 5.1 Key constructs used for making stable cell lines in this chapter..... 176

Table 5.2 Summary of stable cell lines that will be used for further work in this thesis. 199

Table 6.1 Sample information for RNAseq analysis. 208

Table 6.2 Summary of DE genes between Tet3G and mock transfected pools. padj < 0.05. 212

Table 6.3 Summary of potential RNA editing events resulting from acute exposure to A3A and A3B. 220

Table 6.4 Gene ontology enrichment analysis (GO) of putative A3A interactors. 230

Abbreviations

2’3’-cGAMP	Cyclic guanosine monophosphate adenosine monophosphate
5-FU	5-fluorouracil
A	
A1	APOBEC1
A2	APOBEC2
A3	APOBEC3
A4	APOBEC4
A3A	APOBEC3A
A3B	APOBEC3B
A3B**	Catalytically inactive A3B (E68Q/E255Q)
A3D	APOBEC3D
A3F	APOBEC3F
A3G	APOBEC3G
A3H	APOBEC3H
ADAR	Adenosine deaminase acting on RNA
AID	Activation induced cytidine deaminase
AKAP9	A-kinase anchoring protein 9
AKT	Protein kinase B
alt-EJ	Alternative end-joining
Amp	Ampicillin
AP site	Apurinic/apyrimidinic site
APE	Apurinic/apyrimidinic endonuclease
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
AR	Androgen receptor
ATCC	American Type Culture Collection
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3-related protein
ATRIP	ATR-interacting protein
AAVS1	Adeno-associated virus integration site 1

B

BARD1	BRCA1 associated RING domain 1
Ba/Sq	Basal/squamous
BCA	Bicinchoninic acid
BCG	Bacillus Calmette–Guérin
BCL	B-cell lymphoma
BCR	B-cell receptor
BER	Base excision repair
BIR	Break-induced repair
Blast	Blasticidin
BLM	Bloom syndrome helicase
BRCA	Breast cancer gene
BSA	Bovine serum albumin
Bsd^R	Blasticidin S deaminase resistance

C

CBS	Clustered-base-substitution
CCLE	Cancer Cell Line Encyclopaedia
CDC	Cell division cycle
CDK	Cyclin dependent kinase
CDKN2A	Cyclin dependent kinase inhibitor 2A
cDNA	Complementary DNA
cGAS	2’3’-cGAMP synthase
CHK1 (or CHEK1)	Checkpoint kinase 1
CHK2	Checkpoint kinase 2
CIS	Carcinoma *in situ*
c-PARP	Cleaved-PARP
CRISPR	Clustered regularly interspaced short palindromic repeats
CRS	Cytoplasmic retention signal
CSR	Class-switch recombination
CTD	C-terminal domain
C-terminal	Carboxyl-terminus

CtBP1	C-terminal binding protein 1
Ctrl	Control
D	
DAG	Diacylglycerol
DBS	Doublet-base-substitution
DE	Differentially expressed
ddH_20	Double-distilled water
DDR	DNA damage response
DFS	Disease-free survival
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
Dox	Doxycycline
DREAM	Dimerization partner, RB-like, E2F and multi-vulval class B
DSB	Double-strand break
DSBR	Double-strand break repair
dsDNA	Double stranded DNA
DSMZ	German Collection of Microorganisms and Cell Cultures
DTT	Dithiothreitol
E	
E2F	E2 transcription factor
ECACC	European Collection of Authenticated Cell Cultures
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidemical growth factor receptor
ER	Oestrogen receptor
ERCC	Excision repair cross complementing protein 1
ERE	Oestrogen receptor response element
ERAA1	Ewing tumour-associated antigen 1
EXO1	Exonuclease 1

F

FBS	Foetal Bovine Serum
FGFR	Fibroblast growth factor receptor
FHIT	Fragile histidine triad diadenosine triphosphatase

G

GATA3	GATA binding protein 3
GATK	The Genome Analysis Toolkit
Gem	Gemcitabine
GFP	Green fluorescent protein
GO	Gene ontology
GRB2	Growth factor receptor bound protein 2
GSEA	Gene set enrichment analysis
GSK3	Glycogen synthase kinase 3

H

H2AX	H2A histone family, member X
HBV	Hepatitis B virus
HDR	Homology-directed repair
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HER2 (ERBB2)	Human epidermal growth factor receptor
HIF	Hypoxia-inducible factor
HIV	Human immunodeficiency virus
HPV	Human papillomavirus
HR	Homologous recombination
HRP	Horseradish peroxidase
HU	Hydroxyurea

I

IC_{50}	The half maximal inhibitory concentration
ICR	The Institute of Cancer Research
ID	Small insertion-and-deletion
IFN	Interferon
Ig	Immunoglobulin

IκB	Inhibitor of NF-κB
IKK	IκB kinase
IL	Interleukin
IP_3	Inositol triphosphate
IR	Ionising radiation
IRF3	Interferon Regulatory Factor 3
K	
Ku70 (XRCC6)	X-ray repair cross complementing 6
Ku80 (XRCC5)	X-ray repair cross complementing 5
L	
LB	Lysogeny broth
LIG	DNA ligase
LTβR	Lymphotoxin-β receptor
LumP	Luminal-papillary
LumNS	Luminal non-specified
LumU	Luminal-unstable
M	
MAP4	Microtubule associated protein 4
MAPK (ERK)	Mitogen-activated protein kinase
MAPKK	Mitogen-activated protein kinase kinase
MDM2	Mouse double minute 2 homolog
MEM	Minimum Essential Medium
MFS	Metastasis-free survival
MI-UCC	Muscle-invasive urothelial carcinoma
MMR	Mismatch repair
Mock	Lipid transfected control
MSH	MutS homolog
MLH	MutL homolog
MRE11	Meiotic recombination 11
mRNA	Messenger RNA
mTORC	Mechanistic target of rapamycin complex

MutL	Mutator L mismatch repair protein
MutS	Mutator S mismatch repair protein
MVAC	Methotrexate, vinblastine, doxorubicin (Adriamycin), and cisplatin
MW	Molecular weight
MYBL2 (b-Myb)	Myb-related protein B
N	
NaCl	Sodium Chloride
NaF	Sodium fluoride
NaOH	Sodium hydroxide
NBS1	Nibrin
NEAA	Non-essential amino acids
NE-like	Neuroendocrine-like
NER	Nucleotide excision repair
NEMO	NF-κB essential modulator
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NF1	Neurofibromin 1
NHEJ	Non-homologous end joining
NIK	NF-κB-inducing kinase
NLS	Nuclear localisation signal
NMI-UCC	Non-muscle-invasive urothelial carcinoma
NP-40	Nonidet P-40
NPM1	Nucleophosmin 1
NSCLC	Non-small lung cell carcinoma
NT	Non-targeting
NTD	N-terminal domain
N-terminal	Amino-terminus
O	
ORF	Open reading frame
OS	Overall survival
P	
P	Probability value

p21	Cyclin dependent kinase inhibitor 1A
p100	NF-κB precursor protein p100
p105	NF-κB precursor protein p105
p50	NF-κB protein p50
p52	NF-κB protein p52
p53 (TP53)	Tumour protein P53
PALB2	Partner and localizer of BRCA2
PARP	Poly (ADP-ribose) polymerase
PAXX	Paralogue of XRCC4 XLF
PBS	Phosphate-buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PGR	Progesterone receptor
PI3K	Phosphoinositide 3-kinase
PIK3CA	PI3K catalytic, alpha polypeptide
PIKK	PI3K-kinase-like protein kinase
PIP_2	phosphatidylinositol 4,5-bisphosphate
PIP_3	phosphatidylinositol 3,4,5-trisphosphate
PKC	Protein kinase C
PLC	Phospholipase C
PMA	Phorbol 12-myristate 13-acetate
PMS2	Post meiotic Segregation Increased 2
Pol	Polymerase
PPARγ (PPARG)	Proliferator-activated receptor gamma
PTEN	Phosphatase and tensin homolog
PTPRD	Tyrosine phosphatase receptor type D
PVDF	Polyvinylidene difluoride
Q	
QC	Quality control
R	
r^2	Coefficient of determination

RAD51	DNA repair protein RAD51 homolog
RAD52	DNA repair protein RAD52 homolog
RAF	Raf proto-oncogene, serine/threonine kinase
RAS	RAS GTPase
RB	Retinoblastoma protein
RelA p65)	REL-associated protein A
RelB	REL-associated protein B
REV1	REV1 DNA Directed Polymerase
RNA	Ribonucleic acid
RNase	Ribonuclease
RNAseq	RNA sequencing
RNP	Ribonucleoprotein
RPA	Replication protein A
RPMI	Roswell Park Memorial Institute 1640
RR	Ribonucleotide reductase
RS	Replication stress
RTK	Receptor tyrosine kinase
S	
SAFB	Scaffold attachment factor B
SBS	Single-base-substitution
SD	Standard deviation
SDS	Sodium-dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SDSA	Synthesis-dependent strand annealing
SHM	Somatic hypermutation
shRNA	Shirt hairpin ribonucleic acid
siRNA	Small interfering ribonucleic acid
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SOC	Stable Outgrowth Medium
SRB	Sulforhodamine B dye

SSB	Single strand break
SSBR	Single-strand break repair
ssDNA	Single stranded DNA
STAT	Signal transducer and activator of transcription
STING	Stimulator of interferon genes
T	
TAE	Tris-acetate-EDTA
TAK1	Transforming growth factor beta-Activated Kinase 1
TARDBP	TAR DNA binding protein
TBE	Tris-borate-EDTA
TBK1	TANK-Binding Kinase 1
TBS	Tris buffered saline
TBS-T	TBS plus tween 20
TCA	Trichloroacetic acid
TCGA	The Cancer Genome Atlas project
TCOF1	Treacle ribosome biogenesis factor 1
TCR	T-cell receptor
TGFBR1	Transforming growth factor beta receptor 1
Tet	Tetracycline
TLS	Translesion synthesis
TMT-LC-MS/MS	Tandem Mass Tag Liquid Chromatography Mass Spectrometry
TNF	Tumour necrosis factor
TNFR	TNF receptor
TOPBP1	DNA topoisomerase 2-binding protein 1
TREX1	Three prime repair exonuclease 1
TRIM29	Tripartite motif protein 29
TSC	Tuberous sclerosis complex
TURBT	Transurethral resection of the bladder tumour
U	
U	Uracil
UDG	Uracil DNA glycosylase

UN	Untreated
UNG	Uracil DNA glycosylase
USA	United States of America
UV	Ultraviolet
V	
Veh	Vehicle
Vif	Viral infectivity factor
v/v	Volume/volume
X	
XLF	XRCC4-like factor
XRCC	X-Ray Repair Cross Complementing
TCGA study	
LAML	Acute Myeloid Leukaemia
ACC	Adrenocortical carcinoma
BLCA	Bladder Urothelial Carcinoma
LGG	Brain Lower Grade Glioma
BRCA	Breast invasive carcinoma
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL	Cholangiocarcinoma
LCML	Chronic Myelogenous Leukaemia
COAD	Colon adenocarcinoma
CNTL	Controls
ESCA	Oesophageal carcinoma
FPPP	FFPE Pilot Phase II
GBM	Glioblastoma multiforme
HNSC	Head and Neck squamous cell carcinoma
KICH	Kidney Chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma

LUSC	Lung squamous cell carcinoma
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
MESO	Mesothelioma
MISC	Miscellaneous
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PCPG	Pheochromocytoma and Paraganglioma
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
SARC	Sarcoma
SKCM	Skin Cutaneous Melanoma
STAD	Stomach adenocarcinoma
TGCT	Testicular Germ Cell Tumours
THYM	Thymoma
THCA	Thyroid carcinoma
UCS	Uterine Carcinosarcoma
UCEC	Uterine Corpus Endometrial Carcinoma
UVM	Uveal Melanoma

Units

rpm	Revolutions per minute
x g	Relative centrifugal force. Times gravity.
kDa	Kilodalton
kb	Kilobases
bp	Base pair
µg	Microgram
mg	Milligram
g	Gram
µL	Microlitre
mL	Millilitre
L	Litre
nM	Nanomolar

µM	Micromolar
mM	Millimolar
M	Molar

Nucleic acid bases

5meC	5 methyl cytosine
A	Adenine
C	Cytosine
G	Guanine
N	Any base
T	Thymine
R	Pyrimidine base (C or T)
Y	Purine base (A or G)
W	A or T

Chapter 1

Introduction

Chapter 1 Introduction

1.1 Hallmarks of cancer

Cancer is a term for a large group of diseases characterised by the development of abnormal cells that proliferate uncontrollably and invade and destroy healthy tissue. Cancer is a leading cause of morbidity and mortality worldwide with an estimated 19.3 million new cancer cases and 10 million deaths in 2020 (Sung et al., 2021). Breast, lung, colorectal, liver and stomach cancer are the most commonly diagnosed and most common causes of cancer-related death (Sung et al., 2021). Hanahan and Weinberg first posed in 2000 that most, if not all, cancers can be characterised by a set of six core characteristics: limitless replicative potential, self-sufficiency in growth signals, insensitivity to anti-growth signals, evasion of programmed cell death, sustained angiogenesis, and tissue invasion and metastasis (Hanahan & Weinberg, 2000). Over the past 22 years, the core set of hallmarks has been expanded to include two additional hallmarks, reprogramming of energy metabolism and evading immune destruction (Hanahan & Weinberg, 2011), and two new emerging hallmarks, unlocking phenotypic plasticity, and senescent cells (Hanahan, 2022). Genomic instability and mutation, tumour-promoting inflammation (Hanahan & Weinberg, 2011), non-mutational epigenetic reprogramming and polymorphic microbiomes (Hanahan, 2022) have been described as enabling characteristics that facilitate the acquisition of the core hallmarks.

1.1.1 Inflammation and cancer

Inflammation plays important roles in all stages of tumourigenesis: initiation, promotion, progression and metastasis (Grivennikov et al., 2010; Pitot et al., 1981; Taniguchi & Karin, 2018), and is now recognised as one of the most important enabling characteristics that contributes to the emergence of cancer and its hallmarks (Hanahan & Weinberg, 2011). Chronic inflammation due to persistent infection, autoimmunity or exposure to chemicals can substantially increase cancer risk (Kuper et al., 2001; Plummer et al., 2016), and chronic inflammation may be responsible for up to 20% of cancer deaths (Grivennikov et al., 2010; Taniguchi & Karin, 2018). Some examples include *Helicobacter pylori* and gastric cancer risk, hepatitis B or C virus and liver cancer risk, inflammatory bowel disease (Crohn's disease and Ulcerative Colitis) and bowel cancer risk and

Schistosoma infections with bladder cancer risk (Grivennikov et al., 2010). Obesity is another cause of chronic inflammation and is a major contributor to cancer death (Calle et al., 2003). Necrotic cell death that occurs within solid tumours due to lack of blood supply can further contribute to chronic inflammation (Grivennikov et al., 2010). Arguably, the most important pathway implicated in inflammation driven carcinogenesis is the Nuclear Factor kappa-light-chain-enhancer of activated B cells (NF-κB) pathway where the NF-κB family of transcription factors act as central mediators of the inflammatory process (DiDonato et al., 2012). These proteins are key regulators of both innate and adaptive immune responses, and can have profound pro-tumorigenic effects via inhibition of apoptosis, increased proliferation, migration, invasion, metastasis and angiogenesis (Karin & Greten, 2005). The NF-κB pathway can be activated by a range of cellular stress including infection, DNA damage, oxidative stress and pro-inflammatory cytokines, and is often constitutively active in both the tumour itself and the surrounding microenvironment (Karin & Greten, 2005; Staudt, 2010).

1.1.1.1 Canonical vs alternative signalling

The NF-κB family of transcription factors consists of five proteins: RelA (p65), RelB, c-Rel, p50 and p52. The latter two are generated by processing of the inactive precursor proteins p105 and p100, respectively. These proteins dimerise in different combinations to form several distinct dimers, each with a unique transcriptional role (Oeckinghaus & Ghosh, 2009). The family of inhibitor of NF-κB (IκB) proteins is largely responsible for their regulation. Under normal, non-stimulated conditions, NF-κB transcription factors are sequestered in the cytoplasm but upon stimulation, IκB proteins are degraded, the transcription factors are released and they translocate to the nucleus (Liu et al., 2017). The precursor proteins p100 and p105 have IκB-like C-terminal portions and can also function as IκB proteins prior to processing. NF-κB activation can proceed via two different pathways: the canonical or the alternative (Figure 1.1). The canonical pathway is induced by ligand activation of a range of cell surface receptors including various cytokine receptors, tumour necrosis factor receptor (TNFR) superfamily members, T and B cell receptors (TCR/BCR) and pattern-recognition receptors (PRRs). Upon stimulation, phosphorylation of IκBα by the IκB kinase (IKK) complex composed of IKKα, IKKβ and NF-κB essential modulator (NEMO)

triggers IκBα ubiquitin-dependent degradation and rapid, but transient, nuclear accumulation of the key canonical dimers, p50/p65, p65/p65 and p50/c-Rel (Israël, 2010; Liu et al., 2017; Sun, 2017). Typical gene targets include those encoding the pro-inflammatory cytokines, TNFα, IL-6 and IL-1; various chemokines; pro-proliferative proteins, cyclin D1 and Myc*;* anti-apoptotic proteins, Bcl2 and Bcl-X_L; proinflammatory enzymes; angiogenic factors; adhesion molecules and negative regulators of the pathway.

In contrast to the rapid activation of the canonical pathway, alternative pathway activation is slower and more persistent, as p100, unlike p105, is not constitutively processed. The production of active p52 is dependent on accumulation of *de novo* synthesised NF-κB-inducing kinase (NIK) and subsequent IKKα-mediated phosphorylation of p100's C-terminal domain to trigger polyubiquitination and processing (Xiao et al., 2001). The alternative pathway is typically activated by ligands of a subset of the TNFR super family including the lymphotoxin β receptor (LTβR), BAFF-receptor (BAFFR), cluster of differentiation 40 (CD40) and RANK, and results in nuclear translocation of the alternative dimer p52/RelB (Liu et al., 2017; Sun, 2017). The alternative pathway is largely required for lymphoid organ development and B-cell maturation but also plays a role in the innate and adaptive immune responses (Sun, 2017). Many activators of the alternative pathway also activate the canonical pathway, demonstrating that the two pathways are intricately linked (Taniguchi & Karin, 2018).

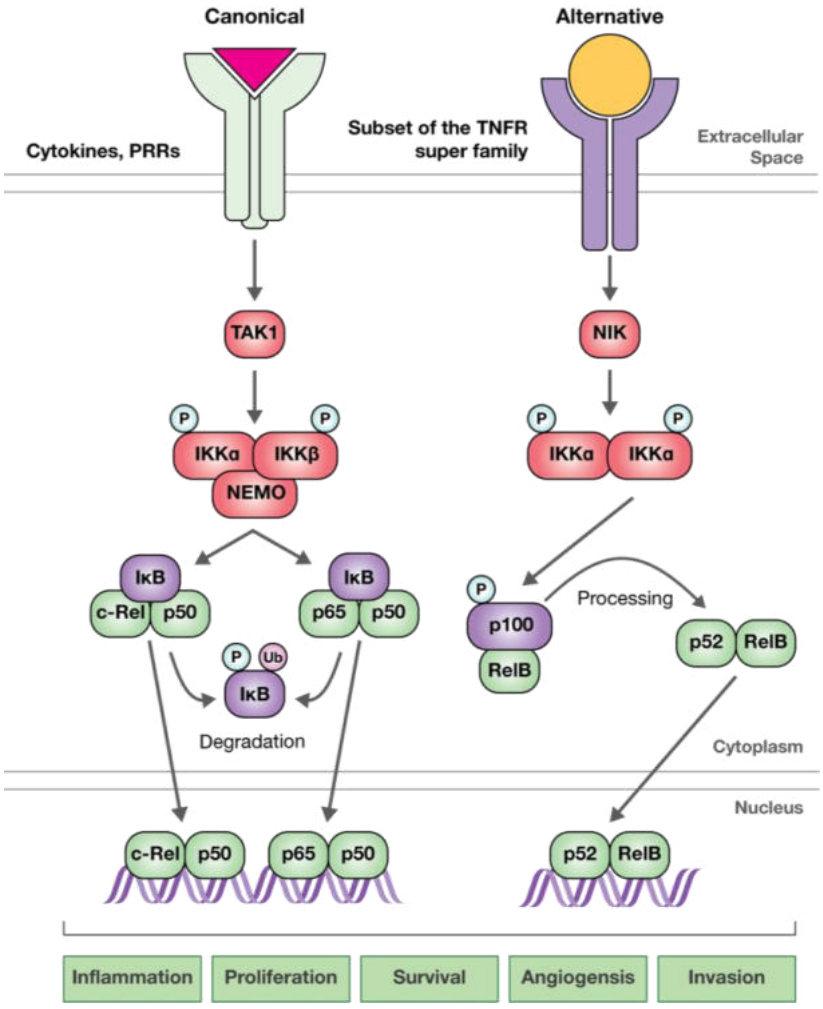


Figure 1.1 NF-kB signalling.

Canonical signalling: ligand binding activates the receptor and transforming growth factor beta-activated kinase 1 (TAK1) is activated. TAK1 phosphorylates and activates the canonical inhibitor of kappa-B (IκB) kinase complex (IKKα/β/NEMO) which phosphorylates IκB proteins, targeting them for degradation through ubiquitination. Canonical dimers (c-Rel/p50; p65/p50) are released and translocate to the nucleus. Alternative signalling: ligand binding activates the receptor and downstream NF-kappa-B-inducing kinase (NIK). NIK phosphorylates and activates the alternative IKK complex (IKKα dimer) which phosphorylates inactive p100, stimulating processing into active p52. The alternative dimer (p52/RelB) is released and translocates to the nucleus. Kinases are shown in red, transcription factors in green, inhibitor proteins in purple. P = phosphorylation; Ub = ubiquitination.

1.1.1.2 Cytosolic DNA sensing

Cytosolic DNA is a danger signal as it usually originates from infection with intracellular pathogens. DNA in the cytosol is bound by cytoplasmic sensor proteins to activate a signalling cascade that stimulates an innate immune

response to control the infection. Cyclic guanosine monophosphate adenosine monophosphate (2'3'-cGAMP) synthase (cGAS) is a key sensor that upon binding DNA, produces the second messenger cyclic 2'3'-cGAMP (Ablasser et al., 2013; Diner et al., 2013; Gao et al., 2013; Sun et al., 2013; Wu et al., 2013; Zhang et al., 2013). 2'3'-cGAMP binding to stimulator of interferon genes (STING) induces a conformational change and after a range of post-translational modifications including phosphorylation, SUMOylation, palmitoylation and polyubiquitination (Chiang & Gack, 2017; Liu et al., 2015; Mukai et al., 2016), STING translocates from the endoplasmic reticulum to the Golgi (Dobbs et al., 2015; Ishikawa & Barber, 2008; Saitoh et al., 2009; Zhang et al., 2013). Activated STING promotes autophosphorylation and activation of TANK-binding kinase 1 (TBK1) that phosphorylates the transcription factor interferon regulatory factor 3 (IRF3) to induce expression of type 1 interferons, and the IKK complex to activate NF-kB (Abe & Barber, 2014; Liu et al., 2015). While the cGAS-STING pathway primarily exists for sensing DNA in the cytosol due to intracellular infection, it can also initiate an innate immune response after recognition of self-DNA. Genomic instability in tumour cells can cause fragmented DNA or micronuclei to leak into the cytoplasm and consequently promote an inflammatory environment (Dou et al., 2017; Glück et al., 2017; Mackenzie et al., 2017).

1.2 Cancer, heterogeneity, and evolution

Cancer development is fundamentally an evolutionary process driven by the sequential acquisition of DNA alterations (mutations) that facilitate the shift from a benign to malignant state. Cancer genomes often contain huge numbers of mutations, ranging from single base pair substitutions, insertions and deletions to chromosome copy number alterations and chromosomal rearrangements. This increased tendency of cancer cells to acquire genomic alterations is known as genomic instability. The outcome of genomic instability is that the tumour is comprised of groups of cells (clones) with a diverse set of genetic alterations, referred to as intratumour heterogeneity (Figure 1.2). The concept of cancer evolution is fundamentally rooted in Darwinian selection with the underlying principle being that cells bearing mutations with a selective advantage survive and outcompete their neighbours that do not (Nowell, 1976). Those that provide a selective advantage are termed 'drivers', while mutations that do not have any

effect on cell fitness are called ‘passengers’, as they are carried along in the genome simply because they coexist with driver mutations (Stratton et al., 2009). Therefore, heterogeneity provides the fuel for cancer evolution and is a major obstacle to successful cancer treatment. Previously inert passenger mutations may become selectively advantageous in the face of therapy stress and drive drug resistance and relapse.

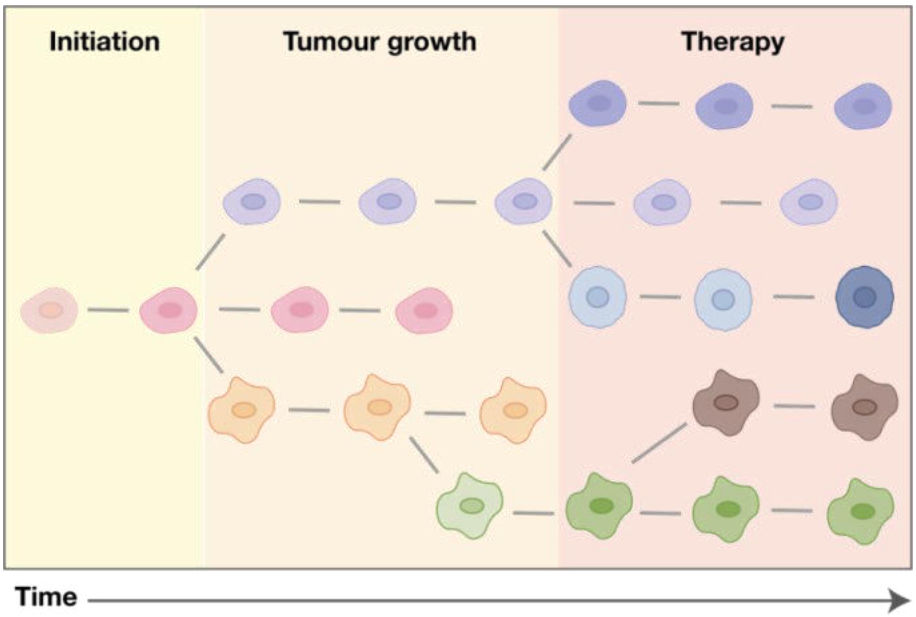


Figure 1.2 Intratumour heterogeneity and evolution.

The tumour is initiated from a common ancestral clone. Genomic instability generates multiple clones and a heterogenous tumour. Upon therapy treatment, sensitive clones die while clones with intrinsic, pre-existing resistance survive and expand due to a selective advantage, and new resistant clones emerge through adaptation and acquired resistance. Therefore, the therapy-resistant tumour and metastases are often largely different to the initial tumour.

1.2.1 Mutational signatures

Study of the mutations in cancer genomes allows identification of the mutational processes that have occurred over the evolutionary history of a particular tumour. Analysis of mutations and their surrounding sequence context in known cancer genes first demonstrated that mutations can be specifically attributed to mutagen exposure, the mechanism of DNA damage and the subsequent repair processes. For example, G-to-T transversion mutations commonly identified in lung cancer are a result of bulky guanine adducts caused by tobacco carcinogens (Hainaut &

Pfeifer, 2001; Pfeifer et al., 2002), while C-to-T transitions in melanoma are a result of UV-mediated pyrimidine dimer formation and subsequent repair (Pfeifer et al., 2005). Recently, next generation sequencing (NGS) technology has allowed unbiased, large-scale investigation into the mutational processes occurring within tumours. In 2012, 21 primary breast cancers and normal matched tissue were subject to whole genome sequencing (WGS) and a new algorithm was used to catalogue the mutational spectrum of human breast cancers and analyse their underlying mutational processes (Nik-Zainal et al., 2012). This algorithm was then used to investigate the mutational signatures of 30 different types of cancer using whole genome sequences from 507, and whole exome sequences from 6535, primary cancer samples and matched normal controls (Alexandrov et al., 2013). This landmark study identified 21 mutational signatures and proposed their underlying mechanisms, although many have still unknown aetiologies. Further work in 2020 using 4645 whole genome and 19184 whole exome sequences enabled the discovery of new signatures as well as deconvolution of previously identified overlapping signatures into distinct profiles (Alexandrov et al., 2020). There are now 49 single-base-substitution (SBS), 11 doublet-base-substitution (DBS), four clustered-base-substitution (CBS) and 17 small insertion-and-deletion (ID) signatures reported across over 30 tumour types. SBS1 is the most common signature, found in almost all tumours and is primarily composed of C-to-T transitions at NpCpG trinucleotides (where N is any of the four bases) and is due to spontaneous deamination of 5-methylcytosine (5mC) in CpG islands. SBS1 and SBS5 are ‘clock-like’ signatures, and positively correlate with patient age at diagnosis. The number of mutations per Mb attributed to them is different between tumour types but correlates with the estimated rate of stem cell division in the tissue of origin. Therefore, these signatures are likely a natural consequence of aging and represent a normal mutational process occurring in somatic cells, with the resulting mutations being acquired throughout the patient’s life at a relatively constant rate. Most other identified signatures do not correlate with age. The number of mutations attributed to them, and the number of patients and cancer types in which they are identified, varies significantly suggesting that these signatures are reflective of differential mutagen exposure, tissue-context specificities, individual genetic background, and repair capacity. Cancers associated with environmental

carcinogen exposure, such as lung and melanoma, had the highest numbers of somatic mutations and displayed the associated signatures (SBS4 and DBS2 attributed to tobacco smoking, and SBS7 and DSB1 attributed to UV light). SBS2 was identified as the second most common signature in human cancers and like SBS1, is primarily composed of C-to-T transition mutations (Alexandrov et al., 2013). However, analysis of the surrounding sequence context shows the mutated cytosine occurs at TpCpN trinucleotides, the preferred sequence context for the apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3 (APOBEC3) family of cytosine deaminases (Nik-Zainal et al., 2012). A second signature, SBS13, composed primarily of C-to-G mutations, again at TpCpN trinucleotides, is also attributed to this family. DBS11, characterised by CC-to-TT mutations, positively correlates with SBS2 suggesting it is also a direct or indirect consequence of APOBEC activity. SBS2 was initially identified in half of all tumour types analysed while SBS13 was only identified in breast and bladder tumours (Alexandrov et al., 2013). However, as of 2020, SBS2 has been identified in almost two-thirds of all cancer types analysed with SBS13 detected in almost all tumours that have SBS2 (Alexandrov et al., 2020).

1.3 APOBEC enzymes

The APOBEC family of deaminases catalyse the deamination of cytosine to uracil within ssDNA and/or RNA. The family is comprised of 11 genes: activation-induced cytidine deaminase *(AICDA* encoding AID*)* and *APOBEC1* on chromosome 12, *APOBEC2* and *4* on chromosome 6 and 1, respectively, and the seven tandemly arranged *APOBEC3 (A3)* genes *(A-D; F-H)* on chromosome 22 (Figure 1.3). Deamination activity is conferred by a Z-domain ($Hx_1Ex_{23-28}Cx_{2-4}C$ motif) where the conserved histidine and cysteines coordinate a single zinc atom, and glutamate promotes hydroxide ion formation that is key for the deamination reaction (Conticello et al., 2005; LaRue et al., 2009; LaRue et al., 2008). Z-domains can be grouped into three clusters, Z1, Z2 and Z3, with family members being either single- or double-domain enzymes. The minimal *A3* gene set, comprised of Z1, Z2 and Z3, present in the common ancestor of placental mammals underwent five duplication and three deletion events to generate the seven highly homologous A3 family members found in primates (LaRue et al., 2008). A3A, C and H are single-domain enzymes while A3B, D, F and G are

double-domain, with the gene originally designed *A3E* now recognised to encode the second domain of A3D (Conticello et al., 2005; LaRue et al., 2009; LaRue et al., 2008). The C-terminal Z-domain of double-domain enzymes confers catalytic activity, and while the N-terminal domain can coordinate zinc, it is considered to be catalytically inactive (Haché et al., 2005). Instead, the N-terminal domain appears to perform a regulatory function, being key for localisation and nucleic acid binding (Adolph et al., 2017; Fu et al., 2015; Pak et al., 2011; Salamango et al., 2018; Stenglein et al., 2008; Xiao et al., 2017).

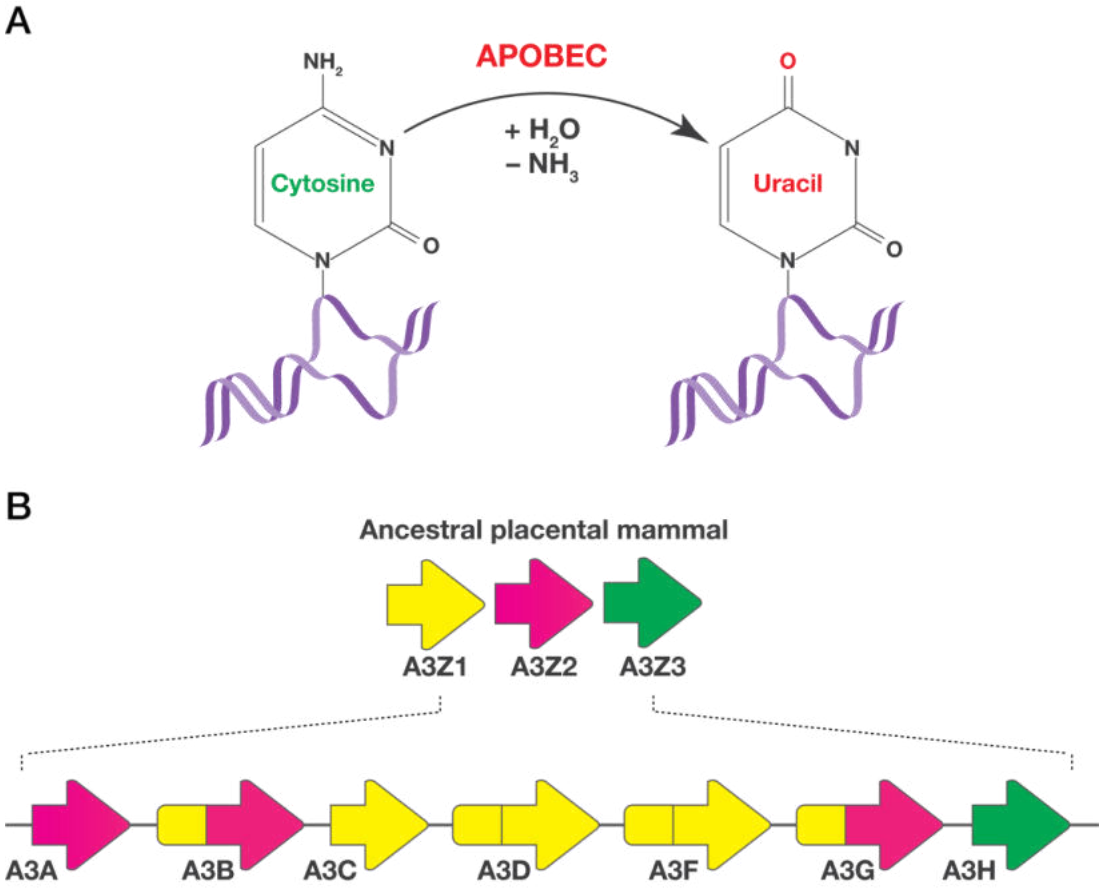


Figure 1.3 APOBEC3 family members.

(A) APOBEC family members catalyse the deamination of cytosine to uracil. (B) Seven *A3* family members are arranged in tandem on Chromosome 22. The current A3s arose from gene duplication events from the minimal gene set, Z1, Z2 and Z3, present in the common ancestor of placental mammals. Single-domain enzymes: A3A, A3C and A3H. Double-domain enzymes: A3B, A3D, A3F and A3G. Based on figures from Lackey et al., (2012) and Swanton et al., (2015).

The founding member of the APOBEC family, A1, was first identified in 1987 as an RNA editing enzyme due to its ability to generate a shorter ApoB protein through the deamination of cytosine^{666} to uracil that creates a premature stop codon (Powell et al., 1987; Teng et al., 1993). Currently, ApoB is the only well-established edited target of A1, though it may have other functions in modulating RNA stability independent of its editing activity (Anant & Davidson, 2000; Anant et al., 2004). A1 also deaminates ssDNA and there is evidence that this function predates its RNA editing ability (Severi et al., 2011). While homologous to A1, AID does not edit mRNA but instead plays a key role in the adaptive immune response facilitating antibody diversification in B-cells (Muramatsu et al., 1999). Deamination by AID occurs in regions of ssDNA, stimulating the formation of double strand breaks (DSBs) that are required for somatic hypermutation (SHM) and class-switch recombination (CSR) (Bransteitter et al., 2003; Dickerson et al., 2003; Martin et al., 2002). A2 is expressed in skeletal and cardiac muscle and is proposed to have a role in myoblast differentiation, as deficiency causes altered muscle fibre formation and myopathy (Liao et al., 1999; Sato et al., 2010); interestingly, A2 cannot deaminate DNA nor RNA and is thought to be acting as a transcriptional repressor (Lorenzo et al., 2021). A4 is expressed in the testis but very little is known about its function (Rogozin et al., 2005).

Extensive study of the A3 family member, A3G, has revealed a key role for A3s in the innate immune response to viral infection. A3G was first identified as a human immunodeficiency virus type 1 (HIV-1) restriction factor. A3G expression *in vitro* blocks replication of viral infectivity factor (Vif)-deficient HIV-1, as Vif binds to and targets A3G for ubiquitination and proteasomal degradation (Sheehy et al., 2002). Shortly after, A3G was shown to deaminate cytosines in DNA (Harris et al., 2002) and a deaminase-dependent viral restriction mechanism was confirmed (Harris et al., 2003; Mangeat et al., 2003). A3G is packaged into viral particles where it deaminates cytosines in CC motifs on the minus DNA strand after reverse transcription. Subsequent second strand synthesis generates G-to-A mutations in the coding strand, many of which convert tryptophan codons to a stop. Hypermutation results in cDNA instability and degradation, and ultimately provirus inactivation, while non-lethal A3-mutation of viral genomes can be the substrate for viral evolution and adaptation (Harris & Dudley, 2015). Other A3 family members are now implicated in the restriction of a variety of viruses and

can act on viral RNA substrates (Harris & Dudley, 2015; Milewska et al., 2018). Viral restriction can also occur independent of deaminase activity by steric hindrance of reverse transcriptase by A3 assembly on viral cDNA (Bishop et al., 2006; Chaurasiya et al., 2014; Holmes et al., 2007; Newman et al., 2005; Stenglein & Harris, 2006).

1.4 APOBEC: endogenous mutagens

The A3-mutational signatures were first identified in 2012 with the discovery of strand-coordinated, non-random regions of hypermutation in human cancers (Nik-Zainal et al., 2012; Roberts et al., 2012). These clustered mutations, termed *kataegis* (Greek for thunderstorm; >5 mutations per cluster), are often located near chromosome rearrangement breakpoints and their formation is dependent on available ssDNA (Nik-Zainal et al., 2012; Roberts et al., 2012). Both A3-signatures occur at 5'TC dinucleotide motifs, the preferred motif for all A3 deaminases except A3G, which targets CC dinucleotides. SBS2 is characterised by C-to-T mutations while SBS13 is primarily composed of C-to-G mutations, and the two signatures reflect differential processing of the deaminated cytosine (Figure 1.4). Cells have diverse and efficient pathways to repair DNA lesions faithfully and it is likely that most deamination events are resolved without mutation; cytosine deamination alone is not sufficient to generate the A3-mutational signature identified in tumours.

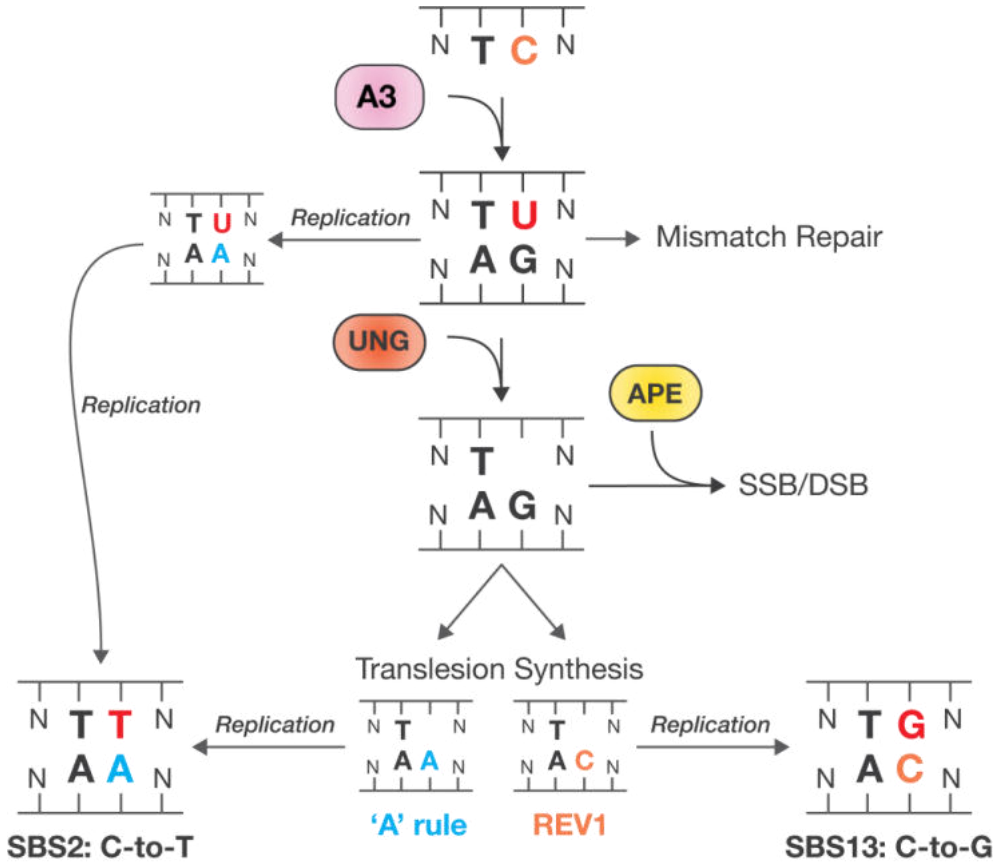


Figure 1.4 A3-mediated mutational mechanisms.

Cytosine is deaminated to uracil within TC motifs in ssDNA by A3 deaminases. If the U is not repaired prior to replication, an A is inserted opposite. During the next replication cycle, T is inserted opposite the A resulting in fixation of C-to-T mutations (single-base-substitution, SBS2). The base excision repair (BER) enzyme uracil DNA glycosylase (UNG) removes U to form an abasic site. Alternatively, the U:G mismatch is less frequently repaired using the mismatch repair (MMR) pathway. Processing of the abasic site by AP endonuclease (APE) generates a single strand break (SSB) that can be further converted to a double strand break (DSB). The abasic site can be bypassed during replication by translesion synthesis resulting in either insertion of A ('A' rule) or C (by REV1), and subsequent replication fixes C-to-T and C-to-G mutations generating SBS2 and SBS13, respectively.

1.4.1 Uracil excision

Uracil in DNA is not itself a serious lesion, as both DNA and RNA polymerases will recognise it as a thymine and replication and transcription will not stall. However, during replication, adenine is inserted opposite the template uracil and subsequent replication fixes C-to-T mutations. Therefore, cells have efficient

ways of coping with genomic uracil with the primary method of repair being the base excision repair (BER) pathway (Beard et al., 2019; Dianov & Hübscher, 2013). BER is initiated by damage-specific DNA glycosylases that both sense the altered base and perform the first step of the repair. Uracil is specifically removed by uracil DNA glycosylases (UNG), the most important of which is nuclear UNG2 in humans. The uracil is flipped out and the *N*-glycosidic bond between the sugar phosphate backbone and uracil base is cleaved, leaving an abasic (apyrimidinic/apurinic (AP)) site. AP endonuclease 1 (APE1) then cleaves the phosphodiester bond 5' of the AP site to generate a single strand break (SSB) that has a 3' hydroxyl and a 5' deoxyribose phosphate. DNA polymerase β (Polβ) inserts the correct base and removes the ligase-blocking 5' deoxyribose phosphate. Finally, the X-ray repair cross-complementing protein 1 and DNA ligase III (XRCC1-LIG3) complex ligates the DNA ends to seal the break.

AP sites that are not repaired prior to DNA replication are also tolerated through the action of translesion polymerases in a process called translesion synthesis (TLS). The action of TLS is linked to the A3-mutational signature; Polα and Polδ with proliferating cell nuclear antigen (PCNA) preferentially insert an A ('A' rule) opposite the AP site, while DNA directed polymerase REV1, preferentially inserts a C, leading to C-to-T and C-to-G mutations, respectively (Choi et al., 2010). However, there are translesion polymerases that preferentially insert T or G opposite the abasic site suggesting that A3 deamination generates a wider repertoire of mutations depending on polymerase use. It is worth noting that the A3-signature does not represent every deamination event that has occurred, but the ones that were not faithfully repaired. Experimentally-induced *kataegis* and SBS13 is substantially reduced by UNG- or REV1-deficiency demonstrating that downstream processing is key for generating part of the A3-mutational signature (Taylor et al., 2013).

Uracil is primarily removed by BER but U:G mismatches can also be repaired by the mismatch repair (MMR) pathway (Larson et al., 2008). Single base mismatches are detected by the MSH2-MSH6 heterodimer (MutSα) followed by recruitment of MLH1-PMS2 (MutLα), PCNA and exonuclease 1 (EXO1) (Huang & Zhou, 2021; Pećina-Šlaus et al., 2020). The endonuclease activity of PMS2 generates a nick facilitating degradation of the mismatch-containing strand by EXO1, Polδ fills in the gap and the nick is sealed by DNA ligase.

1.4.2 Repairing DNA breaks

BER of uracil requires the formation of a SSB and failure to repair these prior to replication can convert them into DSBs. Deamination events occurring in close proximity on opposing DNA strands that are processed by APE1 also facilitate DSB formation. Non-homologous end joining (NHEJ) allows the rapid repair of DSBs by directly ligating the DNA ends in the absence of a homologous repair template and therefore, functions throughout the cell cycle (Huang & Zhou, 2021; Scully et al., 2019). When the overhangs of the DSB are compatible, NHEJ can repair the break faithfully using regions of microhomology. However, if the ends are incompatible and require further processing, NHEJ is considered an error-prone process, often generating small insertions and deletions. Classical NHEJ (cNHEJ) requires binding of the Ku complex (Ku70 and Ku80) to the DNA ends and subsequent recruitment of DNA-dependent protein kinase catalytic subunit (DNA-PK). Ku and DNA-PK work cooperatively to bring the DNA ends together and protect them from nuclease degradation. The nuclease Artemis resects incompatible ends, with Polλ and Polμ performing gap synthesis prior to ligation by the LIG4, XRCC4, XRCC4-like factor (XLF), paralogue of XRCC4 XLF (PAXX) complex (Pannunzio et al., 2018). The alternative end joining pathway (alt-EJ) is Ku-independent and requires machinery traditionally involved in homologous recombination (HR) repair (Sallmyr & Tomkinson, 2018). This includes the MRN (MRE11-RAD50-NSB1) complex, C-terminal binding protein 1 (CtBP1) and EXO1 for resection, poly (ADP-ribose) polymerase (PARP) and RAD52 for end binding and bridging, Polθ for gap synthesis, and LIG3-XRCC1 or LIG1 for ligation.

Homologous recombination, in comparison to NHEJ and alt-EJ, is far more complex and only occurs during the S- and G2-phases of the cell cycle when a sister chromatid is available for error-free recombinational repair (Her & Bunting, 2018; Scully et al., 2019). The MRN complex, activated by CtBP1, initiates end resection and displaces Ku to prevent NHEJ. Endonuclease DNA2, EXO1 and Bloom syndrome helicase (BLM) carry out ‘long-range’ resection, degrading the 5’ strand to form long 3-terminal ssDNA overhangs, which are bound and protected by replication protein A (RPA). The HR mediators, breast cancer gene (BRCA) 1 and 2, RAD52 and partner and localizer of BRCA2 (PALB2) facilitate RPA displacement and RAD51 filament formation on the 3’ ssDNA tail. The

RAD51 nucleofilament invades homologous dsDNA forming a three-strand helix. Base pairing causes displacement of the non-complementary strand and formation of a displacement loop (D-loop) that is followed with nascent strand synthesis by Polδ. Repair is primarily completed by the non-crossover, error-free synthesis-dependent strand annealing (SDSA) pathway in somatic cells but error-prone long-tract gene conversion (LTGC) and break-induced replication (BIR) pathways can also be used.

1.4.3 The DNA damage response

DNA damage must be sensed, cell-cycle checkpoints activated, and the correct repair process initiated to prevent genome instability and initiate programmed cell death in cells with extensive damage to suppress tumorigenesis; this is known as the DNA damage response (DDR) (Figure 1.5). The central mediators of the DDR are the members of the phosphoinositide 3-kinase (PI3K)-related kinases (PIKK) family, ataxia-telangiectasia mutated (ATM), ATM-RAD3-related protein (ATR) and DNA-PK (Blackford & Jackson, 2017). ATM and DNA-PK are both activated in response to DSBs while ATR activation is involved in the response to DNA replication stress and SSBs. DNA-PK is activated by Ku bound to DNA ends and its primary role is to promote NHEJ of the DSB. ATM, on the other hand, is considered the key kinase for mediating the overall cellular response to DSBs. Under typical conditions, ATM exists as an inactive dimer but upon MRN binding a DSB, ATM is recruited, its autophosphorylation is triggered, it dissociates into monomers and is activated (Smith et al., 2020). ATM then phosphorylates several downstream signal transducers primarily resulting in G1/S checkpoint activation. Checkpoint kinase 2 (CHK2) is phosphorylated, dimerises, and phosphorylates the cell division cycle phosphatase 25A (CDC25A) causing cyclin dependent kinase 2, 4 or 6 (CDK2, CDK4/6) to remain in its inactive state preventing progression into S phase. CHK2 also phosphorylates tumour protein p53 (*TP53*), preventing interaction with and ubiquitination by its negative regulator, mouse double minute 2 (MDM2), that usually targets it for proteasomal degradation. p53 accumulates and initiates transcription of *CDKN1A* (encoding p21) and several pro-apoptotic genes leading to cell cycle arrest and apoptosis in extensively damaged cells. The ATM-CHK2 signalling pathway also regulates the G2/M checkpoint via CDC25C and CDK1.

Replication stress (RS) is defined as the slowing and/or stalling of replication that affects replication fork stability and is a main contributor to genomic instability (Gaillard et al., 2015). The ATR signalling pathway is key for controlling repair, replication timing, and checkpoint activation in response to RS (Gaillard et al., 2015; Smith et al., 2020). RS generates ssDNA that is bound and protected by RPA, that then recruits ATR via ATR interacting protein (ATIRP). Topoisomerase II binding protein 1 (TOPBP1) and Ewing tumour-associated antigen 1 (ETAA1) activate ATR which then phosphorylates and activates CHK1. CHK1 phosphorylates CDC25C and CDC25A, causing CDK1 and CDK2 inactivation, respectively, and activation of the intra-S and G2/M checkpoints. Like CHK2, CHK1 can also phosphorylate and activate p53. ATR also controls origin firing and fork restart to ensure complete faithful replication of the genome, while simultaneously preventing RPA exhaustion (Saldivar et al., 2017). Therefore, ATR-CHK1 signalling is key for preventing catastrophic damage and premature entry into mitosis.

The PARP family of enzymes are also key components of the DDR (Ray Chaudhuri & Nussenzweig, 2017). ssDNA, SSBs and DSBs are all rapidly bound by PARP1, and its activity enhances the recruitment of repair enzymes and signal transducers, MRN, XRCC1-LIG3, BRCA1, ATM, ATR, and DNA-PK, amongst numerous others, to the damaged site. PARP1 also protects SSBs from nucleases until repair commences and has roles in stabilising and restarting replication forks (Ray Chaudhuri & Nussenzweig, 2017).

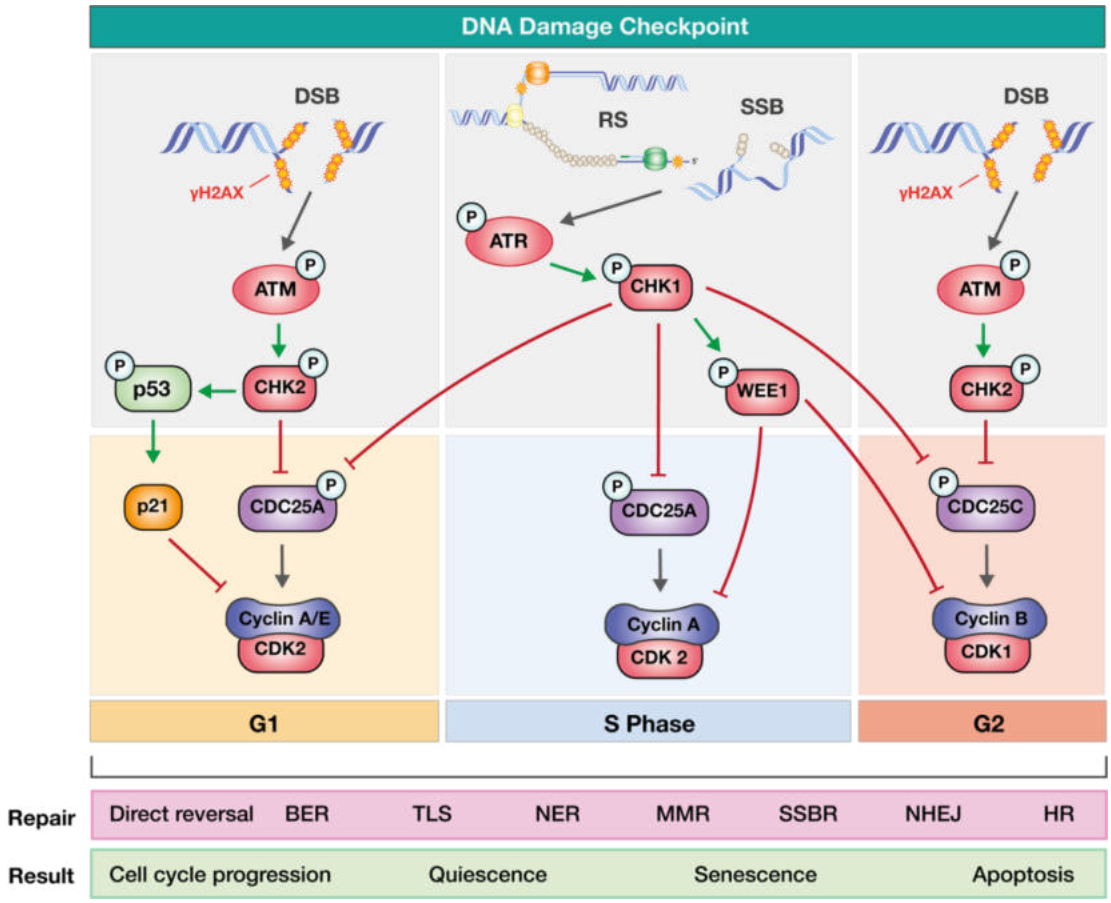


Figure 1.5 The DNA damage response (DDR).

Replication stress (RS) and single strand breaks (SSBs) activate ATR/CHK1 signalling while double strand breaks (DSBs) activate ATM/CHK2. CHK1 and CHK2 phosphorylate CDC25A and CDC25C to prevent activation of CDKs, activating the cell cycle checkpoints. CHK1 also phosphorylates and activates WEE1, and both CHK1 and 2 phosphorylate p53 relieving it of repression by MDM2 resulting in transcription and accumulation of the key cell cycle regulator p21. Cell cycle arrest allows time to repair DNA damage via several pathways: base excision repair (BER), translesion synthesis (TLS), nucleotide excision repair (NER), mismatch repair (MMR), SSB repair (SSBR) and DSB repair by non-homologous end joining (NHEJ) or homologous recombination (HR). Depending on the level of damage and if it was repaired successfully, the cell cycle can either resume, cells can quiesce, senesce or undergo programmed cell death.

1.4.4 APOBEC3A or B?

SBS2 and SBS13 have been attributed to the activity of the APOBEC3 family of enzymes based on the sequence context of the mutated cytosine (Alexandrov et al., 2013). These signatures occur within TpCpN trinucleotides, the motif preference for six of the seven A3 family members, referred to as the ‘TC’ specific

deaminases. This raises the question, the activity of which family member or combination of them is responsible for the signatures? Initial work proposed that A3B was the most likely candidate (Harris, 2015). A3B was first suggested to be the predominant mutagen in breast cancer, as it was the only family member upregulated in breast tumour versus normal tissue samples, and was upregulated in two-thirds of all breast cancer cell lines compared to the MCF10A cell line used as a normal control (Burns, Lackey, et al., 2013). *A3B* expression in breast cancer samples correlates with the C-to-T mutational load, with high *A3B* expressing tumours having twice as many mutations. Early findings showed that the mutations were specifically enriched within 5'TCW (W is A or T) motifs (Roberts et al., 2012), and analysis of A3B substrate preference *in vitro* using recombinant A3B C-terminal domain (CTD) protein revealed a strong preference for 5'TCA (Burns, Lackey, et al., 2013). Subsequent studies demonstrated upregulation of *A3B*, its correlation with mutational load, and enrichment in 5'TCW motif mutations in a number of other tumour types including those with the strongest A3-mutational signatures, bladder, head and neck, cervix and lung cancer (Burns, Temiz, et al., 2013; Roberts et al., 2013).

A3A activity has also been proposed to be a source of mutagenesis in human cancers but the literature data are conflicting. Deletion of *A3B* and fusion of *A3A* to the 3' untranslated region (UTR) of *A3B* (*ΔA3B*) is a common polymorphism identified in 22% of all humans, although its prevalence varies significantly geographically (Kidd et al., 2007). This chimeric mRNA encodes a protein identical to A3A that is more stable and results in higher A3A protein levels (Caval et al., 2014). Sequencing of breast cancer tumours showed that carriers of the deletion allele have a higher number of A3 mutations than non-carriers, suggesting that A3A can contribute to generating SBS2 and 13 (Nik-Zainal et al., 2014). While *A3B* is commonly upregulated in numerous tumour types relative to normal tissue, *A3A* is not (Figure 1.6), and correlations between *A3A* expression and the mutational signatures can be identified but they are weak at best (Cortez et al., 2019; Roberts et al., 2013). While the lack of correlation between *A3A* expression and the mutational signature initially suggested that A3A is not a prominent cause of SBS2 and 13, there is now emerging evidence that A3A-mediated mutagenesis is far more common than A3B-mediated. Expression of human A3A and A3B in UNG-deficient yeast, followed by WGS, revealed that the

signatures generated by the two family members can be differentiated (Chan et al., 2015). A3A strongly favours a 5'YTCA motif while A3B prefers 5'RTCA, where Y is a pyrimidine base and R is a purine. The genomes of 243 cancers with strong 5'TCA mutational signatures were categorised into A3A- and A3B-like groups by comparing Y/RTCA enrichment: 32.5% of the tumours could not be categorised but 41.6% were determined to be A3A-like with the remaining 25.9% A3B-like. The A3A-like tumours also had over ten-fold more A3 signature mutations and therefore, the authors concluded that A3A-mediated mutagenesis is more common. Two forms of SBS2 and SBS13 have since been identified in a pan cancer study of mutational signatures by Alexandrov et al., (2020) with the 5'YTC motif being more common, supporting the conclusions of Chan et al., (2015).

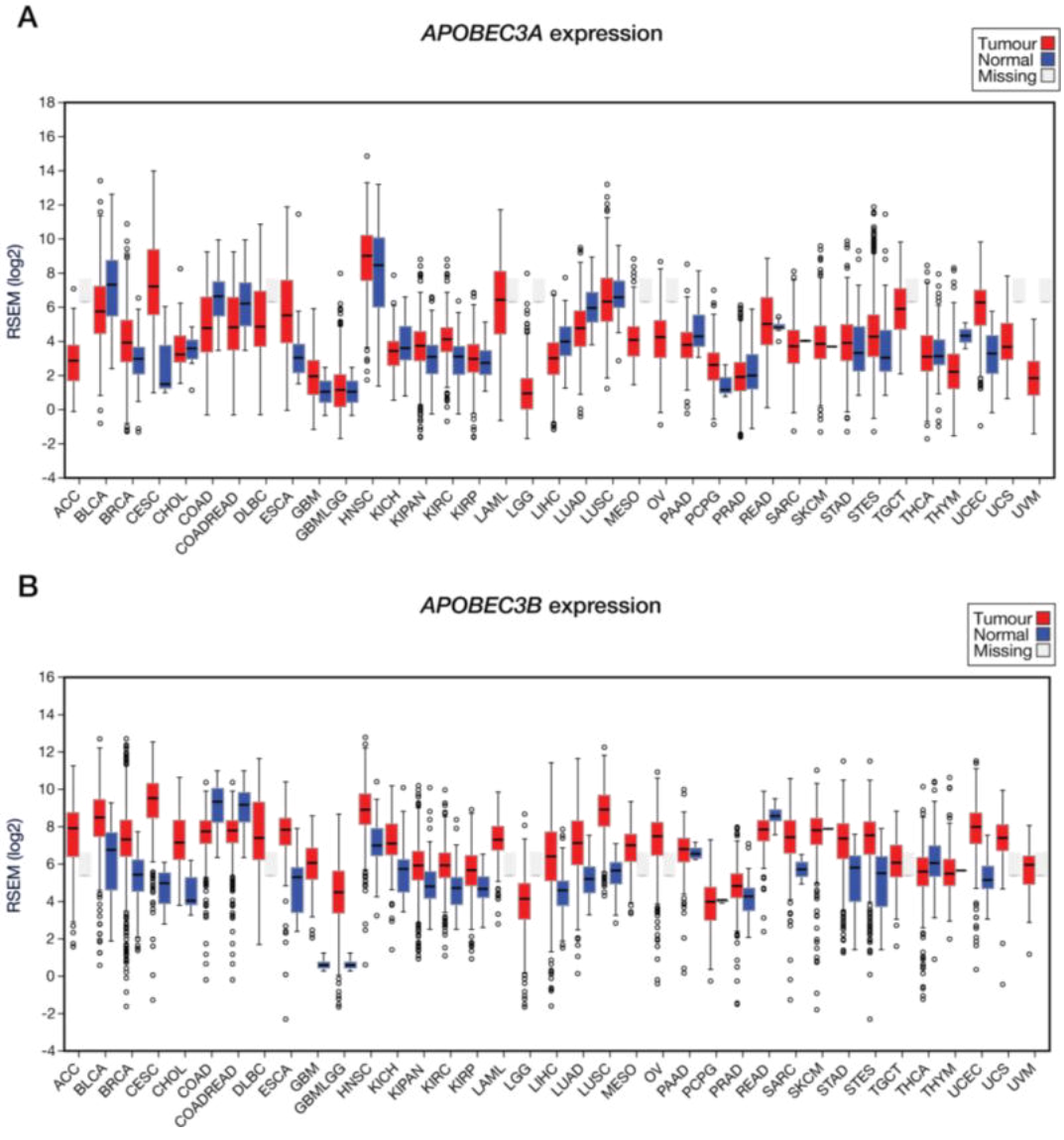


Figure 1.6 Expression of *APOBEC3A* (A) and *APOBEC3B* (B) in tumour and normal samples.

Expression data from The Cancer Genome Atlas project (TCGA), The Broad Institute of MIT and Harvard, 2015. RSEM, RNASeq by Expectation-Maximization. Figure made using plots generated by FireBrowse Gene Expression viewer. Cancer types can be found in Abbreviations.

Compelling evidence for the role of A3A and/or A3B in generating the A3 mutational signatures, genomic instability and tumorigenesis comes from *in vitro* and *in vivo* models. Early cancer cell line work demonstrated that nuclear deamination activity, genomic uracil load and mutation in cell lines could be reduced by small hairpin RNA (shRNA) knockdown of *A3B*, leading the authors to suggest A3B as the predominant deaminase active in breast cancer (Burns, Lackey, et al., 2013). However, the shRNA used is not specific to *A3B*, reducing *A3A* mRNA levels by ~14-fold in a more recent study, and the authors instead proposed that A3A is the prominent deaminase in breast cancer (Cortez et al., 2019). The breast cancer cell lines SKBR3 and AU565 have A3 mutational signatures despite being homozygous for the *ΔA3B* polymorphism, and deamination activity was abrogated by shRNA knockdown of *A3A,* but not *A3B,* in both *A3B-null* and high *A3B*-expressing breast cancer cell lines. Despite this, A3B has been demonstrated to generate the A3 mutational signature, cause DNA damage and genomic instability experimentally. HEK293 cells expressing A3B tagged with enhanced green fluorescent protein (eGFP) under the control of a doxycycline-inducible promoter had increased C-to-T and C-to-G mutations, some within A3 motif contexts, as well as increased copy number alternations, compared to eGFP-expressing control cells after 10 rounds of sublethal exposure (Akre et al., 2016). A3B-eGFP expression in HEK293 cells depleted of p53 by shRNA results in increased genomic uracil and a kataegic-like mutational signature characterised by C-to-T and C-to-G mutations within the preferred 5'TCW motif (Nikkilä et al., 2017). Like A3B, expression of A3A in model systems also generates SBS2-like and SBS13-like mutational signatures. A recent study used doxycycline-inducible A3A expression in an avian cell line (DT40) that lacks A3 orthologs and is genetically stable in long-term culture, to experimentally characterise the mutational signature generated by A3A (DeWeerd et al., 2022). The authors first exposed an ancestral DT40-A3A clone to A3A for 30 days prior to isolation of 16 daughter clones. WGS of the daughter clones revealed significantly higher and more variable mutational burden than uninduced,

parental, or catalytically inactive A3A clones. The identified mutations were primarily C-to-T and C-to-G mutations within 5'YTCW motifs with further preference for a G at the +2 position. While the mutations clustered, they more accurately represented *omikli* (Greek for fog; defined as 2 – 4 mutations per cluster by (Mas-Ponte & Supek, 2020)) than *kataegis*. The authors also identified an insertion and/or deletion (indel)-dominated signature after A3A exposure predominantly composed of single C deletions within the minimal 5'TC motif and suggested this could be a result of polymerase slippage at abasic sites. Interestingly, the ID9 signature previously identified by Alexandrov et al., (2020) with unknown aetiology, is primarily composed of single C deletions and is frequently found in A3-associated tumours suggesting that an additional cancer mutational signature may be attributable to A3A activity. A3A-like SBS2 and SBS13 signatures can also be identified in liver tumours from mice expressing human A3A, demonstrating that A3A can generate A3-mutational signatures *in vivo* as well as *in vitro* (Law et al., 2020).

1.4.5 Subcellular localisation

The ability of an A3 enzyme to generate the observed mutations in cancer genomes relies on its ability to access the genomic DNA substrate. The double-domain A3 enzymes, A3D, A3G and A3F, are too large to passively enter the nucleus, and as they lack a nuclear localisation signal (NLS) are kept in the cytoplasm limiting their genotoxicity (Lackey et al., 2013). A3G also has a cytoplasmic retention signal, further limiting its ability to interact with genomic DNA (Bennett et al., 2008). A3B, on the other hand, has an N-terminal NLS and is the only A3 enzyme that is constitutively localised to the nucleus (Lackey et al., 2012; Lackey et al., 2013). This was a major factor contributing to the initial conclusions that A3B was the predominant A3 in cancer mutagenesis. The single-domain enzymes, A3A, A3C and A3H are small enough to passively diffuse through the nuclear pore and have been shown to be distributed throughout the cell, potentially having access to genomic DNA (Lackey et al., 2013). However, endogenous A3A in $CD14^{+}$ cells or the THP-1 cell line is localised to the cytoplasm and is different to the cell-wide localisation observed in previous studies after transient or stable transfection (Land et al., 2013). Therefore, there are outstanding questions regarding A3A's localisation, and

ability to access genomic DNA and generate the observed mutational signatures. One possibility for the differences identified is that cell types that usually express A3A (myeloid lineages) have robust mechanisms to prevent A3A mutagenesis during immune activation that other cell types, because they are not usually exposed to A3A, lack. It may be that loss of mechanisms that repress induction of A3A in tumour cells coupled with the inability to regulate its genotoxicity via cytoplasmic retention facilitates A3A-genotoxicity.

1.4.6 Substrate availability

A3 enzymes deaminate ssDNA and therefore their ability to generate the mutational signatures in human cancers is dependent on its formation and its accessibility. ssDNA is usually bound by ssDNA binding proteins, primarily RPA, to protect it from degradation and damage, suggesting that under most conditions, ssDNA, even when present, is inaccessible to A3 enzymes. Therefore, A3s must either compete with RPA for binding and displace bound RPA or genomic DNA deamination can only occur when ssDNA is left exposed. Recombinant A3A and full-length A3B can both deaminate RPA-saturated oligonucleotide substrates *in vitro,* suggesting they can indeed displace RPA (Adolph et al., 2017). By comparing the deamination activity on the RPA-saturated substrate to a naked substrate, the authors could determine whether RPA hinders deamination ability. A3B is a processive enzyme, meaning it can slide across DNA or make intersegmental jumps to deaminate multiple cytosines in a single encounter. A3A, on the other hand, is considered non-processive and rapidly cycles on and off DNA more efficiently than A3B. Processivity and deamination activity of A3B was hindered by RPA 2-fold, suggesting that RPA acts as a roadblock and forces the enzyme to cycle off and back onto the DNA. In contrast to A3B, A3A activity was not hindered. The authors concluded that the ability to displace RPA and deaminate bound substrates requires enzyme cycling. When RPA is in excess, it can be readily displaced by other enzymes to facilitate repair and therefore, the authors proposed that during conditions of RS when RPA is in excess and ssDNA is saturated, RPA dissociation by A3A and A3B is possible. Another study also demonstrated that A3A can deaminate ssDNA substrates in the presence of whole cell lysates or purified RPA but that deamination activity specifically on linear but not hairpin substrates is reduced,

demonstrating that RPA binds less efficiently to hairpin structures (Brown et al., 2021). The authors suggested that non-hairpin forming ssDNA sequences will be bound by RPA and protected from mutation while hairpin-forming sequences will be preferentially mutated, supporting the observations that A3A mutational signatures are often enriched in sequences predicted to form secondary structures (Buisson et al., 2019; Langenbucher et al., 2021). Together, these findings highlight that RPA is unlikely to sufficiently protect DNA from A3 mutagenesis.

As discussed previously, cells have complex pathways for sensing DNA damage and ensuring its faithful repair. However, many of these repair processes generate regions of ssDNA that are vulnerable to A3 targeting. Enrichment of the A3 signature and *kataegis* at rearrangement sites points to A3 activity at long regions of ssDNA generated during end resection in DSB repair (Nik-Zainal et al., 2012; Roberts et al., 2012; Taylor et al., 2013) (Figure 1.7). Three prime repair exonuclease 1 (TREX1) degradation of DNA bridges occurring during telomere crisis also generates ssDNA and is required for A3 *kataegis* at chromothripsis break points (Maciejowski et al., 2020). Another type of hypermutation, *omikli*, is associated with MMR (Mas-Ponte & Supek, 2020). MMR also generates ssDNA intermediates, although these regions are much shorter than those for DSB repair and therefore, shorter cluster tracts are formed. A3 mutations are often enriched in DNA flanking a repaired U:G pair, and form via a mechanism that involves hijacking of BER intermediates by MMR proteins to generate ssDNA substrates (Chen et al., 2014). As A3 deamination is a major source of genomic uracil in cancer, this suggests that repair of A3-deamination events can generate more substrate and enhance A3 mutation.

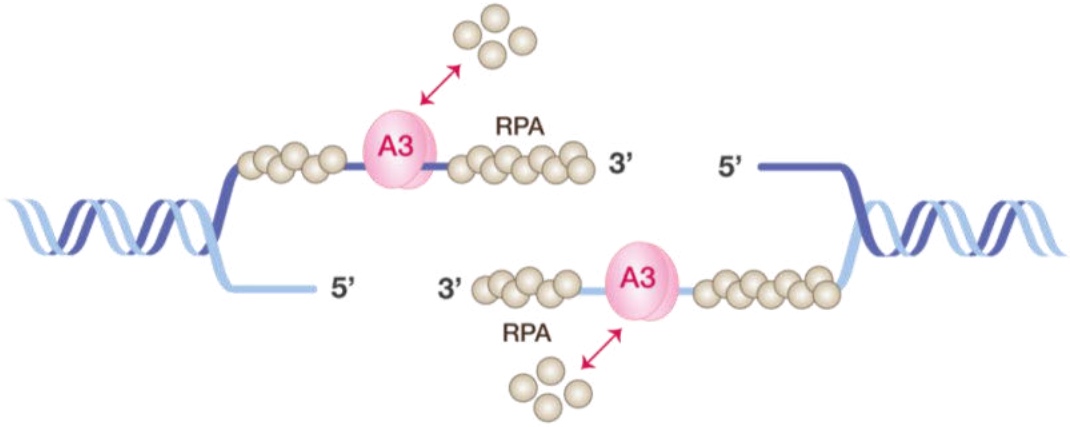


Figure 1.7 Proposed A3 activity on double-strand breaks (DSBs).

Resection of 3' ends during DSB repair generates long ssDNA regions that A3 enzymes can act on.

Transcription also generates long stable regions of ssDNA. The nascent mRNA hybridises with the template strand to form an RNA-DNA duplex structure called an R-loop, in which the non-template strand is displaced and remains single-stranded (Belotserkovskii et al., 2018). Therefore, A3s could preferentially deaminate the non-transcribed ssDNA strand of transcription-associated R-loops as has been observed for AID during SHM (Chaudhuri et al., 2003; Ramiro et al., 2003; Sohail et al., 2003) (Figure 1.8). A3A deaminates the non-transcribed strand of *in vitro* transcription reactions (Adolph et al., 2017; Love et al., 2012; Pham et al., 2013) but cannot deaminate when RNA polymerase is stalled and the target C is part of the transcription bubble in close proximity to RNA polymerase (Brown et al., 2021). This suggests that RNA polymerase sterically hinders A3 activity, and deamination of the non-transcribed strand requires R-loop structures with longer regions of ssDNA. A3A and B both deaminate R-loop structures *in vitro* (Adolph et al., 2017) and a recent preprint has demonstrated a role for A3B in transcription-associated R-loop homeostasis (McCann et al., 2021). In this study, shRNA depletion of *A3B* increases global R-loop levels while A3B overexpression reduces them, and the authors proposed a model where A3B deamination of the displaced, non-transcribed strand facilitates R-loop resolution. In support of this model, the A3B 5'RTCA signature was enriched on the non-transcribed strand in breast cancer samples and overlapped with R-loop

regions. Enrichment of the A3 mutational signatures in the non-transcribed strand has also been identified in bladder cancer (Nordentoft et al., 2014) but enrichment was not identified in a pan cancer analysis (Alexandrov et al., 2020).

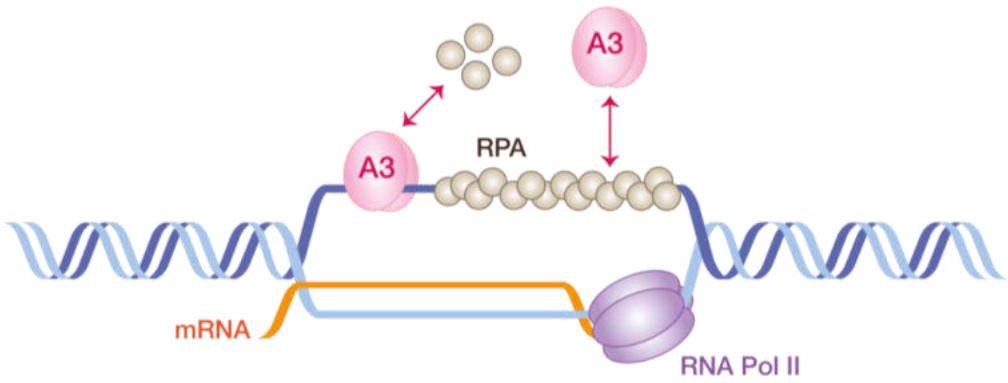


Figure 1.8 Proposed A3 activity on R-loops.

Hybridisation of nascent mRNA with the template strand forms an R-loop structure with the displaced ssDNA being a substrate for A3 enzymes.

During replication, small ssDNA gaps exist due to the discontinuous synthesis of the lagging strand. However, during RS, leading and lagging strand synthesis can become uncoupled, the replication fork can stall, and long regions of ssDNA can form that A3 enzymes can act on (Figure 1.9). The A3 mutational signatures identified in tumours are more commonly enriched in the lagging than leading strand (DeWeerd et al., 2022; Haradhvala et al., 2016; Hoopes et al., 2016; Seplyarskiy et al., 2016) and are enriched in early replicating regions of the genome (Kazanov et al., 2015). Interestingly, both early replicating regions and lagging strand synthesis are associated with increased MMR activity suggesting that MMR of lesions during replication also generates additional substrate for A3 deamination (Mas-Ponte & Supek, 2020). As RS further increases ssDNA substrate availability, it is strongly associated with A3 mutational signatures. Loss of fragile histidine triad protein (FHIT) results in nucleotide imbalance and RS, and is associated with the A3 mutational signature in lung tumours (Waters et al., 2015). In addition, RS can exhaust RPA leading to ssDNA that is unbound and exposed (Toledo et al., 2013). Therefore, in conditions of extreme RS where RPA is exhausted, A3 enzymes will be able to deaminate ssDNA without having to

displace RPA. Replication fork collapse generates one-ended DSBs that have to be repaired using a specific repair pathway, BIR, as there is not a second DNA end available for direct end joining or SDSA (Scully et al., 2019). BIR involves migrating bubble DNA synthesis that results in long stretches of ssDNA, and yeast studies have suggested that A3 deamination at BIR sites is a more common cause of *kataegis* than at traditional two-ended DSBs that have ssDNA as result of resection (Elango et al., 2019; Sakofsky et al., 2014; Sui et al., 2020).

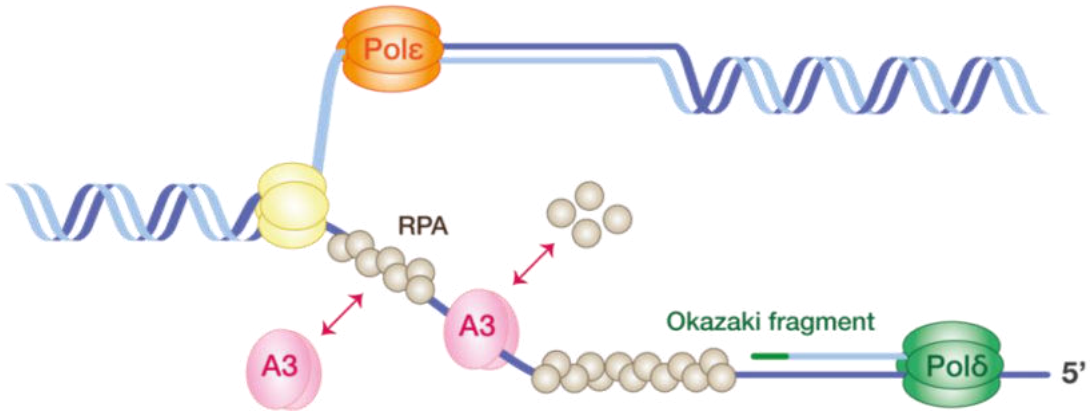


Figure 1.9 Proposed A3 activity on the replication fork.

During replication long regions of ssDNA substrate for A3 enzymes can form on the lagging strand when the replicative polymerases uncouple, and the replication fork stalls.

1.5 Roles of A3A and A3B in cancer

Deamination of viral RNA and DNA by A3 enzymes is known to contribute to viral evolution, immune escape and drug resistance (Pecori et al., 2022; Venkatesan et al., 2018). A3 deamination of viral genomes restricts viral replication when it causes lethal levels of mutation, inactivation, and degradation of the viral genome. However, mutations in viral genomes, if sublethal, rather than restricting virus replication can instead provide the mutagenic fuel required for adaptation and evolution. A3G-mediated G-to-A mutations in the coding strand of the HIV viral genome have been observed in drug-resistant variants and are associated with mutation of antigens resulting in reduced immune responses. There is now a mounting body of evidence that, analogous to A3s role in viral evolution, A3 activity in cancer can drive mutation, heterogeneity and ultimately tumour

progression, adaptability, and therapy resistance. Mutations in a number of known cancer driver genes occur within the A3 sequence context, including protein tyrosine phosphatase receptor type D (*PTPRD*), PI3K catalytic alpha polypeptide (*PIK3CA*), E1A binding protein P300 (*EP300*) (de Bruin et al., 2014), phosphatase and tensin homolog (*PTEN*), epidermal growth factor receptor (*EGFR*) and *TP53* (McGranahan et al., 2015). In the first 100 patients of the tracking non-small cell lung cancer evolution through therapy (TRACERx) cohort, 487 driver mutations were identified (Venkatesan et al., 2021). Of these, 71 were within an A3 sequence context and 37 of the tumours harboured at least one A3 context driver mutation. *Omikli* is also enriched in gene-dense, early replicating regions of the genome and these clustered events can be identified in tumour suppressor and chromatin remodelling genes (Mas-Ponte & Supek, 2020). In addition to generating single nucleotide variants (SNVs), A3 activity can generate larger scale genomic alternations that also fuel genomic instability. Overexpression of A3A and A3B in numerous cell line models increases serine 139 phosphorylated H2A histone family member X (γ-H2AX) foci, DNA fragmentation and RPA phosphorylation, markers of DSBs and RS, suggesting deamination and subsequent lesion processing generates other forms of DNA damage aside from SNVs (Akre et al., 2016; Buisson et al., 2017; Burns, Lackey, et al., 2013; Green et al., 2017; Lackey et al., 2013; Landry et al., 2011; Nikkilä et al., 2017). The findings that the damage is more profound in replicating cells and does not occur when catalytically inactive mutants are expressed suggests that deamination at replication forks is the primary mode of DSB formation. A3B overexpression causes accumulation of under-replicated DNA that results in chromosomal missegregation and micronuclei formation, suggesting that A3B activity can contribute to chromosomal instability (Venkatesan et al., 2021). Knockdown of endogenous *A3B* in high-*A3B* expressing myeloma cell lines decreases basal levels of γ-H2AX and prevents loss of an integrated transgene via A3B-mediated mutation and DSB formation (Yamazaki et al., 2019). Altogether these data suggest that A3 expressing tumours have increased heterogeneity and genomic instability that provides the substrate for evolution.

A3 expression and mutational signatures are linked to prognosis but associations vary depending on the tumour type and therapy used, suggesting that there are context-specific factors at play. Elevated *A3B* expression is associated with poor

outcomes such as lymph node involvement, tumour size and grade in breast cancer, and worse disease-free (DFS), metastasis-free (MFS) and overall survival (OS) in breast cancer patients that have not had neoadjuvant therapy suggesting A3B plays a direct role in breast cancer progression (Sieuwerts et al., 2014). These associations are particularly striking within the ER+ subtype (Periyasamy et al., 2015; Sieuwerts et al., 2014). A3B expression is also negatively correlated with tamoxifen benefit in breast cancer patients with metastatic disease, and depletion of A3B in both *in vitro* and *in vivo* models results in more durable responses to tamoxifen (Law et al., 2016). Increased *A3B* expression is also associated with poor outcomes in non-small cell lung cancer (NSCLC), again being associated with lymph node involvement, grade, and shorter DFS and OS (Wang et al., 2018; Yan et al., 2016), poor prognosis translocations in myeloma (Walker et al., 2015) and poor prognosis and treatment response in low-grade glioma (Luo et al., 2021).

1.6 Urothelial Carcinoma

Bladder cancer is the 10th most commonly diagnosed cancer worldwide, with 573,000 new cases and 213,000 deaths in 2020 (Sung et al., 2021). Bladder cancer incidence and mortality rates are ~4 times higher in men, making it the 6th most common cancer diagnosed in men and the 9th leading cause of male cancer death. The majority, 90%, of bladder cancers are urothelial carcinomas (UCC) with the remaining 10% being made up of less common types; squamous cell carcinoma (SCC) and adenocarcinomas. UCC arises from the innermost epithelial layer (urothelium) of the bladder and can present as non-muscle-invasive UCC (NMI-UCC) or muscle-invasive UCC (MI-UCC) (Figure 1.10), depending on the level of growth into the layers of the bladder (Tran et al., 2021). NMI-UCC accounts for 80% of UCC diagnoses and is comprised of carcinoma *in situ* (CIS), stage Ta and T1 tumours. CIS is a flat, aggressive tumour confined to the urothelium, Ta tumours are still confined to the urothelium but have started to grow into the lumen of the bladder (papillary carcinoma) while T1 tumours have begun to invade the next layer of the bladder, the lamina propria. MI-UCC diagnoses make up the remaining 20% and are comprised of T2, T3 and T4 tumours. Stage 2 tumours have moved through the lamina propria into the muscle layers, stage 3 tumours have grown through the muscle into the surrounding fat

tissue, and stage 4 tumours have broken through the fat layer and have spread outside the bladder. Approximately 90% of NMI-UCCs are low grade and are unlikely to develop into MI-UCC. As a result 5-year survival rates are generally very good, more than 88% for Ta and T1 stage tumours (Berdik, 2017). However, over 60% of NMI-UCCs recur and 10% progress to MI disease, with high-grade NMI-UCC being more aggressive and more likely to progress (Le Goux et al., 2020). Prognosis for patients with MI-UCC, on the other hand, is poor, being less than 50% for ≥ T2 tumours, with the highest stage, T4, having a 5-year survival rate of only 10%.

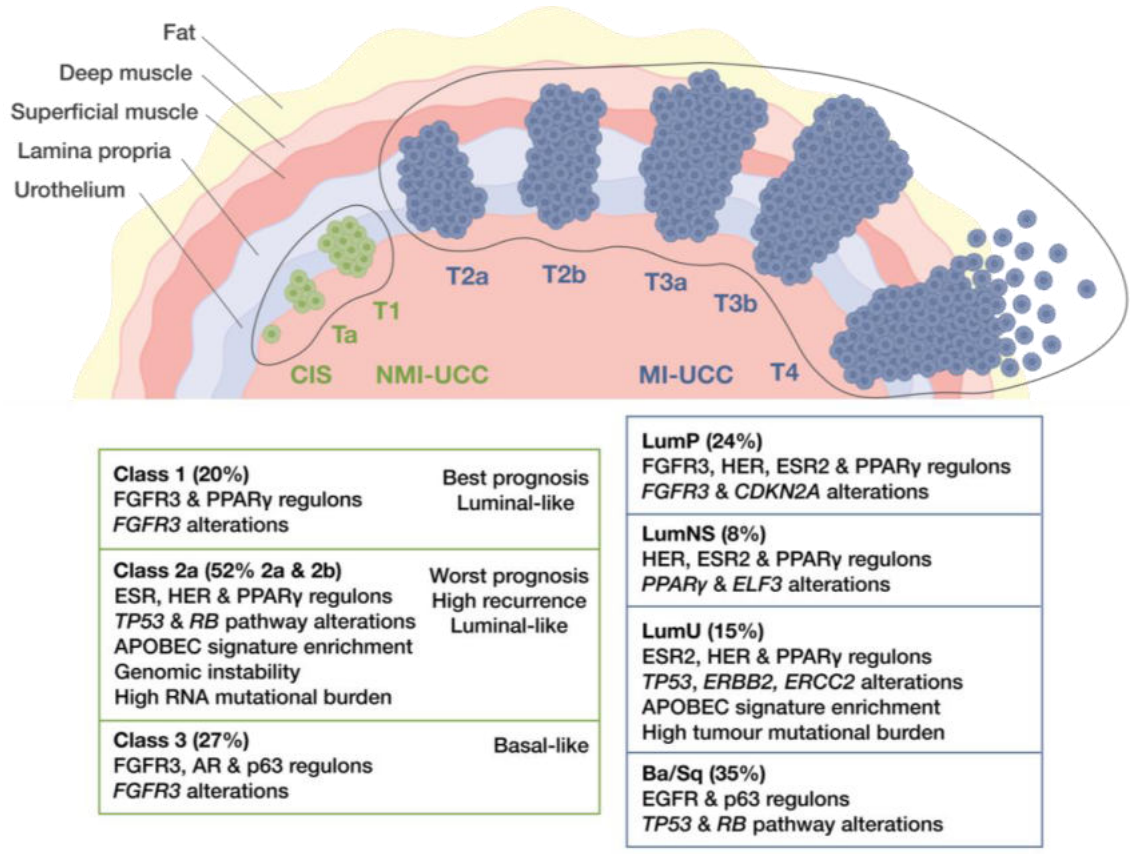


Figure 1.10 Urothelial carcinoma (UCC) stages and key molecular subtypes.

Non-muscle-invasive UCC (NMI-UCC) subtypes are from the study by Lindskrog et al., (2021) and muscle-invasive UCC subtypes are from the study by Kamoun et al., (2020). LumP, luminal-papillary; LumNS, luminal non-specific; LumU, luminal-unstable; Ba/Sq, basal-squamous. Class 2b, neuroendocrine-like and stroma-rich subtypes are not shown. Based on figure from Tran et al., (2021).

As NMI-UCC has not invaded the muscle wall, non-systemic localised treatment is an option for reducing toxicity (Berdik, 2017; Hurst et al., 2018; Tran et al., 2021). Low-risk NMI-UCC is most commonly treated with transurethral resection of the bladder tumour (TURBT) followed by a single intravesical chemotherapy treatment with classical cytotoxic drugs such as mitomycin C, epirubicin, doxorubicin or gemcitabine and frequent cystoscopic surveillance. Intermediate-risk NMI-UCC requires a follow up TURBT procedure and multiple doses of intravesical chemotherapy, while high-risk NMI-UCC requires Bacillus Calmette–Guérin (BCG) treatment and potential bladder removal (cystectomy). Standard MI-UCC treatment involves neoadjuvant cisplatin (or other platinum-based) chemotherapy and radical cystectomy. Bladder-preserving trimodal therapy that combines TURBT, radiotherapy and chemotherapy can be used but 30% of patients will still require radical cystectomy. Metastatic disease is still primarily treated with cisplatin combinations, most commonly methotrexate, vinblastine, doxorubicin (Adriamycin), and cisplatin together (MVAC), or gemcitabine and cisplatin (GC) (Stecca et al., 2021). Unfortunately, chemotherapy resistance and disease progression are common, and median survival is only 12 – 15 months. Repeated recurrences of NMI-UCC and progression to MI disease requires frequent patient monitoring, and the common emergence of chemotherapy resistance means that from diagnosis to patient death, bladder cancer is one of the most expensive cancers to treat.

Bladder cancer survival rates have remained relatively unchanged for over three decades due to minimal changes in treatment options. However, recent progress in understanding the molecular basis of UCC is facilitating the discovery of targetable cancer drivers and advancing treatment options for the first time in decades (Tran et al., 2021). Molecular subtypes of MI-UCC have now been extensively studied, defining two major subtypes, luminal- and basal-like, classified based on their similarity to breast cancer subtypes (Choi et al., 2014; Damrauer et al., 2014; Network, 2014; Sjödahl et al., 2012). Additional studies then refined these major subtypes into five or six more specific groups (Robertson et al., 2017; Sjödahl et al., 2017). Unfortunately, the ability of these molecular subtypes to be routinely used for clinical guidance is limited by the lack of a unified system. Therefore, a recent study analysed transcriptomic datasets from 1750 MI-UCC across 16 published studies alongside two additional cohorts to

generate six consensus subtypes: luminal-papillary (LumP; 24%), luminal non-specified (LumNS; 8%), luminal-unstable (LumU; 15%), stroma-rich (15%), basal/squamous (Ba/Sq 35%) and neuroendocrine-like (NE-like; 3%) (Kamoun et al., 2020) The luminal subtypes are enriched for signatures of urothelial differentiation and gene sets ('regulons') regulated by nuclear receptors, primarily the proliferator-activated receptor γ (*PPARG* encoding PPARγ) and the oestrogen-receptor (*ESR*), in addition to those regulated by human epidermal growth factor receptor 2 (HER2 encoded by *ERBB2*). LumP tumours are specifically associated with a fibroblast growth factor receptor 3 (*FGFR3*) regulon, and frequently harbour *FGFR3* alterations (amplification, mutations, and translocations) and cyclin dependent kinase inhibitor 2A (*CDKN2A* encoding $p16^{INK4A}$ and $p14^{ARF}$) deletion. LumU tumours have unstable genomes and are characterised by *PPARG* and *TP53* alterations, and *ERBB2* amplification. Ba/Sq tumours frequently harbour mutations in *TP53* and retinoblastoma protein (*RB*) and are enriched in *EGFR* regulons. NMI-UCC tumours have also been classified based on their genetic alterations and transcriptional gene signatures. NMI-UCCs were initially grouped into three classes (Hedegaard et al., 2016) that have been further refined into four distinct classes, 1, 2a, 2b and 3 (Lindskrog et al., 2021). Majority of the 834 tumours when classified according to the consensus MI-UCC groupings (Kamoun et al., 2020) were characterised as LumP, showing that early stage tumours are mainly luminal-like. Class 3 tumours have enrichment of basal cell markers, but the co-expression of luminal markers led the authors to conclude that these NMI-UCCs should not be compared to the Ba/Sq subtype of MI-UCC. Class 1 and 3 tumours are both associated with *FGFR3* alterations and regulons, with class 1 tumours also enriched in *PPARG* gene sets. Class 2a tumours have unstable genomes, are associated with p53 and RB pathway mutations, and *ESR* and *ERBB2* regulons. Altogether, the molecular characterisation of UCC demonstrates that they have gene signatures and genetic alterations that are targetable by clinically available drugs including erlotinib, infigratinib and lapatinib for EGFR, FGFR3 and HER2 associated alterations/regulons, respectively.

1.6.1 APOBECs in UCC

UCC has one of the highest somatic mutation rates of all tumour types and is only surpassed by those with known causes of high mutational burden including lung

cancer, melanoma, and colorectal cancer (Alexandrov et al., 2020; Alexandrov et al., 2013; Network, 2014). UCC has the strongest A3 mutational signature of all tumour types analysed; almost all UCC tumours have both SBS2 and 13 (proportion of tumours with the signature ~1.0), with SBS2 accounting for ~1 and SBS13 accounting for ~2.5 median mutations per Mb (Alexandrov et al., 2020). Overall, A3-mediated mutagenesis in UCC is thought to account for 67% of all SNVs making it the most prominent mutagenic process occurring in UCC tumours (Robertson et al., 2017). The A3 mutational signatures are identified in early stage NMI-UCC and become enriched in later stage ≥T2 MI-UCC tumours (Nordentoft et al., 2014). Over half of the identified A3 signatures are clonal (present in >90% of tumour cells), suggesting that A3 mutagenesis occurs early but is ongoing and stable (Robertson et al., 2017). Interestingly, SBS13 in bladder cancer is more significantly associated with early mutations than late ones suggesting that it is generated early in tumour development, while SBS2 is associated with both early and late mutations, and accumulates over time (McGranahan et al., 2015). This suggests there is a temporal difference in the two signatures in bladder cancer resulting from differential downstream processing of the uracil lesion in early vs late tumours. A3A-like mutational signatures are also increased in chemotherapy-resistant advanced UCC and these mutations occur in pathways known to be important for drug resistance, for example the ABC family of transporters, suggesting that A3A mutagenesis is a driver of evolution and drug resistance in this tumour type (Faltas et al., 2016).

In UCC, *A3B* expression correlates with stage and is increased in lymph node metastases relative to the primary tumour (Nordentoft et al., 2014). Class 2a NMI-UCCs enriched in A3 mutational signatures are more heterogenous, have a higher rate of recurrence, more often progress to MI-UCC and therefore, have the worst prognosis of the four classes (Hedegaard et al., 2016; Lindskrog et al., 2021). This, taken with the findings that 40% of clonal and 45% of subclonal mutations identified in bladder cancer occur in an A3 sequence motif suggests that A3-mediated mutagenesis is a strong driver of progression in NMI-UCC (McGranahan et al., 2015). Conversely, high A3 expression and enrichment of A3-mutational signatures is associated with a strikingly good overall prognosis in MI-UCC and advanced metastatic disease in several studies (Glaser et al., 2018; Middlebrooks et al., 2016; Natesan et al., 2022; Robertson et al., 2017). These

findings demonstrate that the selective advantage of A3-mediated mutagenesis is dependent on the selective pressure acting upon the tumour. In NMI-UCC, increased A3 mutagenesis provides the genomic instability required to drive disease progression. However, A3 mutagenesis likely generates neoantigens and it is therefore possible that past a threshold of mutagenesis, A3 activity increases tumour visibility and clearance by the immune system. In support of this, A3 mutational signatures are associated with higher neoantigen loads in MI-UCC (Robertson et al., 2017). There is also emerging evidence that a high burden of clonal neoantigens is associated with both improved patient survival and more durable responses to immunotherapy (McGranahan et al., 2016). T-cells that are reactive to clonal neoantigens will target all the tumour cells, improving their ability to control the tumour. However, T-cells that are reactive to subclonal neoantigens will only target a subset of the tumour cells and tumour control is hindered. While UCCs have high mutational burdens, over half of the A3 mutational signature is generated early in UCC development and therefore, the neoantigen repertoire is both large and clonal, enabling better tumour control by the immune system. This predicted high clonal neoantigen load in UCC may be behind the success of traditional BCG therapy and more recently, immune checkpoint inhibitors. These findings suggest that there is an optimal time for immunotherapy treatment; when UCCs have a high burden of clonal A3 mutations and neoantigens, immunotherapy is likely to be beneficial. However, when the A3 mutational burden increases neoantigen heterogeneity and the number of subclonal neoantigens, immunotherapy benefit will be reduced. Chemotherapy treatment is also associated with an increase in subclonal neoantigen expression and this highlights that using immunotherapy as a last-line treatment after chemotherapy failure may be limiting its efficacy (McGranahan et al., 2016). As chemotherapy drugs can induce expression of A3 family members (Kanu et al., 2016; Middlebrooks et al., 2016; Periyasamy et al., 2021; Yamazaki et al., 2020), it is likely that the increase in subclonal neoantigen burden is, at least partially, mediated by A3 activity.

1.7 Thesis aims

Almost two-thirds of all cancer types are enriched for two mutational signatures, SBS2 and SBS13, that have been attributed to the deamination activity of the A3 family of cytosine deaminases (Alexandrov et al., 2020). When the signatures were first identified and characterised in 2012 (Nik-Zainal et al., 2012), A3B was quickly proposed to be the leading candidate. *A3B* expression is elevated in many tumours relative to normal tissue and expression correlates well with the mutational signatures identified in human cancers (Burns, Lackey, et al., 2013; Burns, Temiz, et al., 2013; Roberts et al., 2013; Swanton et al., 2015). Further studies into the motif preference of two family members, A3A and A3B, revealed that their signatures can be differentiated (Chan et al., 2015). Interestingly, reanalysis of tumour samples found that A3A-like signatures are more common, and A3A-like tumours have more mutations than A3B-like tumours, suggesting instead that A3A is the prominent mutagen (Alexandrov et al., 2020; Chan et al., 2015). However, this raises the question: how can A3A be responsible for generating the mutational signature when its expression is often undetectable, not elevated relative to normal tissues, and does not correlate well with the mutational load? A hypothesis, now supported by some experimental evidence, is that the A3 mutational signature is generated in bursts during transiently elevated expression of A3A/B (Petljak et al., 2019). Therefore, elevated *A3A* expression is not identified at the time of sample processing, but the signature is a molecular footprint of previous elevated expression and activity. In further support of this, *A3A* and/or *A3B* are transiently upregulated in response to a range of cellular stresses, including hypoxia, cellular crowding and RS (Alqassim et al., 2021; Bader et al., 2021; Sharma et al., 2017; Sharma et al., 2016; Sharma et al., 2015; Sharma et al., 2019), and this could be driving bursts of expression in tumours. Of particular clinical relevance is the finding that A3 expression and activity is upregulated in response to chemotherapy drug treatment (Kanu et al., 2016; Middlebrooks et al., 2016; Periyasamy et al., 2021; Yamazaki et al., 2020) and this demonstrates that drug treatment itself could drive heterogeneity, evolution and adaptation via A3-mediated mutation. A second outstanding question is: why would cells want to upregulate mutagens in the face of stress, which could cause further DNA damage, genotoxic stress, and genomic instability? It may be that they are upregulated to speed up adaptation, analogous

to bacterial stress-induced mutagenesis (SIM), or that they function in DNA repair, but the evidence for this is lacking.

Until recently, most published studies have focused on the role of A3B in breast cancer and characterisation of A3B's role in other tumour types, and A3A's role more generally, is lacking. UCC has the strongest A3 mutational signature of all types analysed, and A3 mutagenesis is enriched in the earliest stage tumours to those that are chemotherapy-resistant, suggesting that A3s may be major drivers of progression, adaptation, and drug resistance in UCC. Therefore, the overall goal of this thesis was to shed light on the regulation and function of both A3A and A3B in UCC. To do this, *A3A* and *A3B* induction in response to chemotherapy was characterised, and both small molecule inhibition and knockdown studies were used to interrogate the signalling pathways responsible for the observed induction. To date, there are no published studies describing the interactomes of A3A and A3B, so work aimed to characterise them in UCC using a proximity-labelling technique that circumvents issues with the lack of good commercially available A3 antibodies that limits use of other techniques. While exogenously expressed A3A is distributed throughout the cell and its expression induces DNA damage, endogenously upregulated A3A appears to be retained in the cytoplasm and non-genotoxic, suggesting that the observed effects in response to exogenous A3A may be a result of extreme overexpression or GFP tagging, and not physiologically relevant (Land et al., 2013). Therefore, identification of A3A's interacting partners using lower, more physiologically relevant levels of expression will give insight into where A3A is localised and if it can act on genomic DNA. RNA sequencing was used to determine the cellular response to transiently elevated expression of A3A and A3B, and alongside identification of interacting partners, determine whether there is evidence of a transcriptional function similar to that previously identified in breast cancer (Periyasamy et al., 2015). Finally, A3A and the CTD of A3B are highly homologous and functional redundancy may exist. Therefore, similar regulatory mechanisms, cellular responses to elevated expression, and interacting partners would be strong evidence of functional redundancy between these family members.

Chapter 2

Materials and methods

Chapter 2 Materials and methods

2.1 Cell lines

Human UCC cell lines HT-1376 (ATCC® CRL-1472™), HT-1197 (ATCC® CRL-1473™), 5637 (ATCC® HTB-9™), RT4 (ATCC® HTB-2™), SW780 (ATCC® CRL-2169™), T-24 (ATCC® HTB-4™), TCCSUP (ATCC® HTB-5™) and UM-UC-3 (ATCC® CRL-1749™) were obtained from the American Type Culture Collection (ATCC). UCC cell lines BFTC-905 (ACC 361), 647-V (ACC 414) and RT-112 (ACC 418) were obtained from DSMZ (German Collection of Microorganisms and Cell Cultures, Germany), and UM-UC-6 (08090503) was obtained from ECACC (European Collection of Authenticated Cell Cultures). The human immortalised normal urothelial cell line, NHU-TERT B, was derived from normal ureter tissue from a male patient and was kindly provided by Professor M. A Knowles (University of Leeds, UK).

2.1.1 Culture conditions

HT-1376, HT-1197, UM-UC-3, UM-UC-6 and TCCSUP were maintained in Minimum Essential Medium (MEM) with 2 mM L-glutamine (Gibco, USA) supplemented with 1 mM sodium pyruvate (Sigma-Aldrich, USA), non-essential amino acids (Sigma-Aldrich, USA) and 10% (v/v) foetal bovine serum (FBS; Gibco, USA). BFTC-905 and 647-V were maintained in Dulbecco's Modified Eagle Medium (DMEM; Sigma-Aldrich, USA) supplemented with 2 mM L-glutamine, 1 mM sodium pyruvate, non-essential amino acids and 10% (v/v) FBS. RT4, 5637, RT-112, SW780 and T-24 were maintained in Roswell Park Memorial Institute (RPMI) 1640 (Sigma-Aldrich, USA) supplemented with 10% FBS. The human immortalised normal urothelial cell line, NHU-TERT B, was cultured and passaged as previously described (Chapman et al., 2006). All cell lines were incubated at 37˚C in a humidified atmosphere of 5% CO_2. All cell lines, except NHU-TERT B, were cultured in Nunc™ tissue-culture flasks and plates (Thermo Fisher Scientific, USA). Cells were screened for *Mycoplasma* contamination using the VenorGeM® *Mycoplasma* PCR detection kit (Minerva Labs, UK) according to manufacturer instructions and were Short Tandem Repeat (STR) profiled to authenticate their identity (ICR, UK).

2.1.2 Passaging

Cells were washed in phosphate-buffered saline (PBS), detached from the flask with trypsin-ethylenediaminetetraacetic acid (EDTA) 0.25% solution, phenol red (Sigma-Aldrich, USA or Gibco, USA), neutralised in medium and split at the appropriate ratio for each cell line. Cells were counted using a Z2 Coulter Particle Count and Size Analyser (Beckman Coulter, USA).

2.1.3 Cryopreservation

For long-term storage cells were cryopreserved in liquid nitrogen. Cells grown in T-75 tissue culture flasks were detached using trypsin-EDTA when cultures had reached 70% confluence and neutralised with addition of medium. Cells were centrifuged for 4 minutes at 400 x g, the pellet was washed with PBS, cells were resuspended in their normal growth medium supplemented with 10% (v/v) DMSO media and aliquoted into an appropriate number of Nalgene® Cryo-Tubes (Thermo Fisher Scientific, USA). Cells were precooled to -80°C in a Mr. Frosty™ Freezing Container (Thermo Fisher Scientific, USA) overnight prior to liquid nitrogen storage. Cells were recovered by incubating cryotubes in a 37°C water bath with gentle agitation until completely thawed. Cells were diluted in medium and centrifuged for 4 minutes at 400 x g after which the cell pellet was resuspended and transferred to a T-25 or T-75 tissue-culture flask containing the appropriate volume of prewarmed medium.

2.1.4 Doubling time

Cells were seeded at 2 – 4 x 10^5 cells/well depending on cell line in T-25 tissue-culture flasks in 5 mL medium. Cells were detached and counted approximately every 24 hours for one week using a Z2 Coulter Particle Count and Size Analyser (Beckman Coulter, USA). Doubling times were calculated on GraphPad Prism 9 using the “Exponential growth equation” nonlinear fit function for counts generated during the log growth phase.

2.1.5 Optimal seeding density

Each cell line was seeded at a range of densities in two 96-well plates in 100 µL standard growth media and left to attach for a minimum of one cell doubling after which cells were treated with 50 µL of media and one plate was harvested (T_0).

The other plate was harvested 96 hours post-treatment (T_{96}). Optimal seeding density was determined from the linear part of the curve before plateau.

2.2 Sulforhodamine B (SRB) assay

Cells grown and treated in 96-well plates were fixed with 10% (w/v) trichloroacetic acid (TCA), stained with 0.4% (w/v) SRB in 1% (v/v) acetic acid for 1 hour and then washed four times with 1% (v/v) acetic acid. Bound SRB, a proxy for cell mass, was solubilised with 10 mM tris-base and absorbance was read at 490 nm in a Wallac Victor X4 Multilabel Plate Reader (PerkinElmer Life Sciences, USA).

2.3 Incucyte growth assay

Cells were grown in Corning® black wall, clear bottom, optical 96-well plates. Five images (10X objective) per well were acquired using the Incucyte® S3 Live-Cell Analysis System (Essen BioScience, USA). Percentage cell confluence was calculated from the acquired phase contrast images using the Incucyte® S3 analysis software.

2.4 Compounds

All compounds used in this thesis with stock and storage information is shown in Table 2.1.

Table 2.1 All compounds used with source, stock, and storage information.

Compound	Source	Solvent	Stock	Storage
Paclitaxel	Sigma-Aldrich, USA	DMSO	1 mM	-80°C
Cisplatin	Sigma-Aldrich, USA	Saline (0.9% NaCl)	2 mM	-80°C
5-fluorouracil	Sigma-Aldrich, USA	Sterile H_2O	10 mM	-80°C
Gemcitabine	Eli Lilly, USA	Saline (0.9% NaCl)	10 mM	-80°C
Bleomycin (sulphate)	Cayman Chemicals, UK	DMSO	10 mM	-80°C
Sotrastaurin (AEB071)	Cayman Chemicals, UK	DMSO	2 mM	-80°C
BAY 11-7082	Abcam, UK	DMSO	10 mM	-80°C
TPCA-1	Abcam, UK	DMSO	10 mM	-80°C
H-151	Cayman Chemicals, UK	DMSO	1 mM	-80°C
KU-60019	MedChem Express, USA	DMSO	10 mM	-80°C

Ceralasertib (AZD6738)	MedChem Express, USA	DMSO	10 mM	-80°C
Erlotinib (Hydrochloride)	MedChem Express, USA	DMSO	10 mM	-80°C
Lapatinib	Cayman Chemicals, UK	DMSO	10 mM	-80°C
Infigratinib (Phosphate)	MedChem Express, USA	DMSO	5 mM	-80°C
Doxycycline	Takara Bio, Japan	Sterile H_2O	50 mg/mL	-20°C
Blasticidin S hydrochloride	Sigma-Aldrich, USA	Sterile H_2O	1 mg/mL	-20°C
Puromycin dihydrochloride	Sigma-Aldrich, USA	Sterile H_2O	1 mg/mL	-20°C
Biotin	Sigma-Aldrich, USA	DMSO	100 mM	-20°C
Sodium butyrate	Sigma-Aldrich, USA	Sterile H_2O	500 mM	4°C
Panobinostat	MedChem Express, USA	DMSO	10 mM	-80°C

2.5 Inhibitory Concentration (IC) determination

Cells were seeded at their optimum density in 100 µL standard growth media and left to attach for at least one cell doubling (most cell lines ~24 hours). A range of drug concentrations were added to the cells in 50 µL growth media and after 96 hours (at least three population doublings for most cell lines) an SRB assay was performed (2.2). All drug exposures were performed in triplicate. Growth was determined as:

$$\boldsymbol{Growth}\ (\%\ \boldsymbol{Vehicle}) = \left[\frac{\boldsymbol{T_{96} - T_0}}{\boldsymbol{V_{96} - T_0}}\right] \times \mathbf{100} \ \text{ or } \ \boldsymbol{Growth}\ (\%\ \boldsymbol{T_0}) = \left[\frac{\boldsymbol{T_{96}}}{\boldsymbol{T_0}}\right] \times \mathbf{100}$$

where T_0 = absorbance value at drug treatment, T_{96} = absorbance value at 96 hours treatment with test drug, and V_{96} = absorbance value at 96 hours vehicle control treatment. Both methods of analysis produce similar IC_{50} values, but growth (%T_0) considers the number of population doublings since T_0. IC_{50} values were determined on GraphPad Prism version 9 using the "log(inhibitor) vs. response -- Variable slope (four parameters)" function while IC_{20} and IC_{80} values were determined using "log(agonist) vs. response -- Find ECanything" function with the F constant set to 80 and 20, respectively.

2.6 Treatments

Cells were seeded in plates and left to attach for a minimum of one cell doubling prior to drug treatment. Cells were washed with ice-cold PBS and plates were stored at -80°C until further processing.

2.7 Small interfering RNA (siRNA) transfection

ON-TARGETplus SMARTpool siRNAs (Horizon Discovery, UK; Table 2.2) were resuspended in nuclease-free water to make 50 µM stocks and single use aliquots were stored at -20°C. Cells were reverse transfected with the siRNAs in 6-well plates using Lipofectamine™ RNAiMAX (Invitrogen, USA) according to manufacturer instructions. Media was changed 24 hours post-transfection. Seeding density, Lipofectamine™ and siRNA concentration was optimised using negative (non-targeting) and positive (GAPDH) control siRNAs. Optimal conditions were those where cell viability exceeded 80% of the untransfected control with > 80% knockdown of GAPDH (Table 2.3).

Table 2.2 siRNA information.

Target	Catalogue number	Individual siRNA components
Non-targeting	D-001810-10	N/A
GAPDH	D-001830-10	N/A
RelA (p65)	L-003533-00	J-003533-06; J-003533-07; J-003533-08; J-003533-09
RelB	L-004767-00	J-004767-06; J-004767-07; J-004767-08; J-004767-09

Table 2.3 Optimised siRNA transfection conditions.

Cell Line	Seeding Density	Lipid Concentration	siRNA concentration
BFTC-905	1.5 x 10^5 cells/well	0.2%	20 nM
SW780	2 x 10^5 cells/well	0.1%	10 nM

2.8 Cell lysis

Lysis buffer was added directly to plates/wells and cells were scraped with a cell scraper to ensure efficient detachment. All lysates were frozen on dry ice and stored at -80°C until use.

2.8.1 Gel-based APOBEC deamination assay

Lysates were prepared using 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (25 mM HEPES pH 7.4, 125 mM sodium chloride (NaCl), 1 mM dithiothreitol (DTT), 5 mM EDTA, 0.3% Nonidet™ P 40 (NP-40) substitute and cOmplete™ protease inhibitor cocktail (1 tablet/10 mL; Roche, Switzerland)). Lysates were pipetted up and down 10 times prior to two rounds of sonication for 5 seconds using an MSE Soniprep 150, and were incubated on ice for 10 minutes prior to clarification.

2.8.2 Western Blot

For western blots in Chapter 3, total cell lysates were prepared using NP-40 lysis buffer (50 mM HEPES pH 7.4, 250 mM NaCl, 0.3% NP-40 substitute, 1 mM DTT, 1 mM EDTA, 1 mM sodium fluoride (NaF), 10 mM β-glycerophosphate, 0.1 mM activated sodium orthovanadate and cOmplete™ protease inhibitor cocktail (1 tablet/10mL)). Lysates were incubated on ice for 10 minutes before clarification.

For western blots in Chapters 5 and 6, total cell lysates were prepared using RIPA lysis buffer (50 mM tris pH 8.0, 150 mM NaCl, 0.1% sodium dodecyl sulphate (SDS), 0.5% sodium deoxycholate (SDC), 1% Triton X-100 and 1X Halt™ protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific, USA)). Lysates were sonicated for 10 seconds using an MSE Soniprep 150 followed by 10 minutes incubation on ice.

2.9 Determination of protein concentration

Protein concentration was determined using the bicinchoninic acid (BCA) assay (Thermo Fisher Scientific, USA). Cell lysates were diluted 1:20 in double-distilled H_2O (ddH_2O) and 10 µL was added to 200 µL of BCA and copper (II) sulphate (reagents mixed at 50:1 ratio) in a 96-well plate. Bovine serum albumin (BSA) diluted in ddH_2O (0.05 – 2 mg/mL) was used as a protein standard. Test samples and protein standards were plated in duplicate. Plates were mixed and incubated

at 37°C for 30 minutes prior to reading absorbance at 570 nm with a Wallac Victor X4 Multilabel Plate Reader (PerkinElmer Life Sciences, USA). Test sample concentrations were determined using the standard curve generated from the protein standards.

2.10 Gel-based APOBEC deamination assay

Protein concentration of all lysates was equalised by diluting in lysis buffer (2.8.1) and deamination reactions were performed at 37°C for 3 hours. Each reaction contained: 0.6 µM of the fluorescently labelled APOBEC3B substrate 5'-IR700-ATTTATATTATTTATTCATATTTATATTTA (Integrated DNA Technologies (IDT), USA), 1 µL RNase A (diluted to 2 U/µL in ddH_2O; Sigma-Aldrich, USA), 1 µL Uracil-DNA Glycosylase (UDG; diluted to 1.5 U/µL in ddH2O; New England Biolabs, USA) in 1X UDG reaction buffer (New England Biolabs, USA) and was made up to 20 µL with HEPES buffer. After the initial incubation, 2 µL of 1 M sodium hydroxide (NaOH) was added and samples were incubated at 95°C for 10 minutes to stop the reaction and cleave the substrate probe. Samples were run on an 18% tris-borate-EDTA (TBE) polyacrylamide gel for 15 minutes at 75 V followed by 35 minutes at 160 V. An X2 formamide buffer was used for loading. Imaging of the gel was performed using the Odyssey-Fc imaging system (LI-COR Biosciences, USA) at 700 nm and signal quantification was performed using the ImageStudio™ software (LI-COR Biosciences, USA). Purified APOBEC3B C-terminal domain enzyme produced in house (ICR, UK) was used as a positive control.

2.11 Western blotting

2.11.1 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Lysates were mixed with NuPAGE™ lithium dodecyl sulphate (LDS) sample buffer (Invitrogen, USA) and NuPAGE™ Sample Reducing Agent (Invitrogen, USA) and heated to 70°C for 10 minutes. Lysates were centrifuged prior to loading and SDS-PAGE was performed using precast polyacrylamide NuPAGE™ bis-tris protein gels (Invitrogen, USA) with NuPAGE™ MOPS or MES SDS Running Buffer (Invitrogen, USA). Either the SeeBlue™ Plus2 Pre-stained Protein Standard (Invitrogen, USA) or Chameleon Duo Pre-Stained Protein ladder (LI-COR Biosciences, USA) was used for estimation of protein size. Gels

were run in a XCell SureLock Mini-Cell Electrophoresis System (Thermo Fisher Scientific, USA) at 175 V until the dye front had reached the bottom of the gel; the running buffer in the upper chamber contained NuPAGE™ Antioxidant (Invitrogen, USA).

2.11.2 Transfer

Proteins were transferred to a methanol-activated Immobilon-FL polyvinylidene fluoride (PVDF) membrane (Merck Millipore, USA) using a Mini Trans-Blot Cell (BioRad, USA) and NuPAGE™ Transfer Buffer (Invitrogen, USA) with NuPAGE™ Antioxidant (Invitrogen, USA) and 10% methanol at 100 V for 90 minutes.

2.11.3 Detection

After transfer, membranes were dried at room temperature for 30 minutes. Prior to blocking, membranes were reactivated in methanol, rinsed with ddH_2O, and washed in tris-buffered saline with 0.1% Tween-20 (TBS-T) for 5 minutes. Membranes were blocked in 5% (w/v) BSA in TBS-T for 1 hour at room temperature with agitation prior to incubation with primary antibodies diluted in blocking buffer overnight at 4°C. Membranes were washed with TBS-T for 5 minutes, six times before incubation with horseradish peroxidase (HRP) conjugated secondary antibodies for 1 hour at room temperature. Membranes were washed in TBS-T for 5 minutes, six times and then washed in TBS for 5 minutes, twice prior to incubation with either Pierce™ ECL Western Blotting Substrate, SuperSignal™ West Pico PLUS Chemiluminescent Substrate or SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific, USA) depending on the target. Bound proteins were visualised after 5 minutes using either the Odyssey-Fc imaging system (LI-COR Biosciences, USA) or the ChemiDoc MP system (BioRad, USA). Antibodies used are shown in Table 2.4.

Table 2.4 Antibodies used for immunoblotting.

Antibody	Dilution	Catalogue #	Source	Species
Primary antibodies				
Vinculin	1:10K	V9264	Sigma	Mouse
p53 (DO-1)	1:1K	sc-126	Santa Cruz	Mouse
p21 Waf1/Cip1 (12D1)	1:1K	2947	CST	Rabbit
Cleaved-PARP (c-PARP) (Asp214)	1:1K	5625	CST	Rabbit
V5-Tag	1:2K	R960-25	Invitrogen	Mouse
HA-tag (C29F4)	1:2K	#3724	CST	Rabbit
Pierce™ High Sensitivity Streptavidin-HRP	1:20K	21130	Thermo	-
Secondary antibodies				
Anti-mouse HRP	1:10K	170-6516	Bio-Rad	Goat
Anti-rabbit HRP	1:10K	170-6515	Bio-Rad	Goat

2.12 RNA isolation

Total RNA was isolated from cells using the RNeasy Plus Mini Kit (Qiagen, Germany) or the ReliaPrep™ RNA Miniprep System (Promega, USA). Cell lysates were homogenised using QIAshredders (Qiagen, Germany) when using the RNeasy Kit. RNA quantity and quality was evaluated with a NanoDrop 8000 (Thermo Fisher Scientific, USA) or NanoPhotometer® NP80 UV/Vis Spectrophotometer (Geneflow, UK).

2.13 Reverse Transcription and quantitative PCR (RT-qPCR)

Matched amounts of RNA were reverse transcribed to cDNA in 20 µL reactions using the High-Capacity cDNA Reverse Transcription kit (Invitrogen, USA) according to manufacturer instructions. No RNA template and no reverse transcriptase controls were used as negative controls to ensure samples were free from genomic DNA and air contamination. cDNA was diluted to 5 ng/µL in nuclease-free water and qPCR was performed using 10 ng of cDNA in 10 µL reactions with TaqMan™ Fast Advanced Master Mix (Applied Biosystems, USA). Reactions were performed in MicroAmp™ Optical 384-Well Reaction Plates sealed with MicroAmp™ Optical Adhesive Film (Applied Biosystems, USA) using

the ViiA 7 Real-Time PCR System (Applied Biosystems, USA). All reactions were performed in three technical replicates. Gene of interest (GOI) and endogenous control (EC) TaqMan expression assays (Applied Biosystems, USA) used are shown in (Table 2.5).

Table 2.5 TaqMan assays (Applied Biosystems, USA) used in RT-qPCR.

Target	TaqMan Assay ID	Dye
18s rRNA	Hs99999901_s1 (4319413E)	VIC-MGB
APOBEC3A	Hs02572821_s1	FAM-MGB
APOBEC3B	Hs00358981_m1	FAM-MGB
TNFα	Hs00174128_m1	FAM-MGB
GAPDH	Hs02786624_g1	FAM-MGB
RELA	Hs01042014_m1	FAM-MGB
RELB	Hs00232399_m1	FAM-MGB

Expression was measured in Ct values that are distributed on a log2 scale. Expression of the target gene was normalised to the endogenous control; Delta-Ct (ΔCt) = Ct (EC) − Ct (GOI). The difference in expression between the experimental sample and a reference sample was calculated as Delta-Delta-Ct (ΔΔCt) = ΔCt (experimental) – ΔCt (reference) and relative change in expression to the reference was calculated as $2^{\Delta\Delta Ct}$. The total number of cycles used for amplification was 40. For initial determination of endogenous expression in the UCC panel, a strict cut-off threshold of Ct = 37 (ΔCt = -29.5 with an average EC Ct = 7.5) was used to eliminate false positives due to air contamination during plate loading; ΔCt ≤ -29.5 was considered as undetectable expression.

2.14 Statistical analysis

The control and treated conditions of each biological repeat of a cell culture experiment were seeded at the same time from the same passage/flask of cells and consequently, lack experimental independence and paired or repeated-measured statistical tests are most appropriate. Parametric paired t-test, repeated-measures one-way or two-way ANOVA were used to compare treated and control groups. After ANOVA, the most appropriate multiple comparisons test

was performed. For repeated-measures one-way ANOVA: Tukey's when comparing all groups, Dunnett's when comparing treated to control and Sidak's when comparing preselected pairs. For repeated-measures two-way ANOVA: Sidak's multiple comparisons test was used when comparing all groups.

Repeated-measures ANOVA cannot handle missing values so where values were missing in a dataset due to random effects, the data was instead analysed by fitting a mixed-effects model. In the presence of missing values, occurring entirely at random, the results can be interpreted like repeated-measures ANOVA (GraphPad Prism, USA). It is impossible to violate the assumption of sphericity when each row of data represents matched data (a biological repeat) and as such, the Geisser and Greenhouse correction was not applied when performing ANOVA or fitting mixed-effects models.

2.14.1 qPCR statistical analysis

Parametric statistical tests were performed on ΔCt values as they and more generally, gene expression, inherently follow a log-normal distribution and the assumption of normality/sphericity can be satisfied. Relative quantification or fold change values are not normally distributed and cannot be used for statistical analysis.

2.15 Polymerase Chain Reaction (PCR)

25 µL reactions were used for screening reactions while 100 µL reactions were used for amplification of plasmids or fragments for cloning. Typical reaction conditions are shown in Table 2.6 and Table 2.7. PCR products were run on an appropriate percentage tris-acetate-EDTA (TAE) or TBE agarose gel containing GelRed® Nucleic Acid Gel Stain (100X) (Biotium, USA) at 80-100 V for 1 hour to check successful amplification. DNA Gel Loading Dye (6X) (Thermo Fisher Scientific, USA) was added to the PCR sample to a final 1X concentration prior to gel loading. The Quick-Load® 1 kb Plus DNA Ladder (New England Biolabs, USA) was ran alongside samples for DNA size estimation. Full details of plasmids and primers used in this thesis can be found in the Appendix.

Table 2.6 Typical PCR reaction mixture.

Component	Stock	Final Concentration
Q5® Hot Start High-Fidelity 2X Master Mix (M0494)	2X	1X
Forward primer	10 µM	0.5 µM
Reverse primer	10 µM	0.5 µM
Template DNA	-	Genomic DNA: 8 ng/µL Plasmid DNA: 5 pg/µL
Nuclease-free water	-	To final reaction volume

Table 2.7 Typical PCR cycle conditions.

Step	Temperature	Time	Cycles
Denaturation	98°C	10 seconds	1
Denaturation	98°C	10 seconds	25
Annealing	*Primer specific*	20 seconds	
Extension	72°C	30 seconds/kb	
Final Extension	72°C	2 minutes	1
Hold	4°C	∞	

2.16 NEBuilder® HiFi DNA Assembly

Cloning was performed using the NEBuilder® HiFi DNA Assembly master mix (New England Biolabs, USA) according to manufacturer instructions with minor modifications (fragment excess 7:1) using PCR generated vector backbone and fragments. PCR primers were generated using the NEBuilder Assembly Tool v.2.2.7 (New England Biolabs, USA). Prior to assembly, PCR products were digested with DpnI to remove residual template DNA and purified using the Monarch® PCR & DNA Cleanup Kit (New England Biolabs, USA) or the NucleoSpin™ Gel and PCR Clean-up Kit (Macherey-Nagel GmbH, Germany) according to manufacturer instructions. When assembling two or more fragments, fragments were incubated at 50°C for 15 minutes prior to addition of the vector backbone and further incubation at 50°C for 45 minutes to prevent the formation of linear products that cannot circularise.

2.17 Restriction Enzyme Digestion

Standard reactions were set up as in Table 2.8 and incubated at 37°C for 1 hour. All restriction enzymes were purchased from New England Biolabs, USA. Restriction enzymes were heat-inactivated according to manufacturer recommendations. 1 µL QuickCIP (New England Biolabs, USA) was added to vector backbone digestions prior to ligation of inserts to prevent backbone recircularization.

Table 2.8 Typical restriction enzyme digestion reaction mixture.

Component	Stock	Final Concentration
Restriction Enzyme	20,000 units/mL	10 units (1 µL)
CutSmart® Buffer	10X	1X
DNA Template	-	1 µg
Nuclease-free water	-	To 50 µL final volume

2.18 Ligation

Plasmid ligations were performed using T4 DNA Ligase (New England Biolabs, USA); standard reaction conditions and modified conditions for removal of the stuffer fragment and restoration of the *A3B* open reading frame (ORF) are shown in Table 2.9. T4 DNA ligase was heat-inactivated at 65°C for 10 minutes prior to downstream use.

Table 2.9 Typical ligation reaction mixture and modified mixture for stuffer removal of A3B^{Split} constructs.

Component	Stock	Final Concentration	A3B _Split
T4 Ligase Buffer	10X	1X	1X
Vector DNA	-	0.02 pmol	5 µg
Insert DNA	-	0.10 pmol	-
Nuclease-free water	-	To 20 µL	To 400 µL
T4 DNA ligase	400,000 units/mL	1 µL (400 units)	4 µL (1600 units)

2.19 Isopropanol precipitation

DNA was precipitated with ⅒ volume 3 M sodium acetate (NaOAc) and an equal volume of 100% isopropanol. The mixture was incubated on dry ice for 15 minutes prior to centrifugation at 14000 x g for 10 minutes. The supernatant was removed, and the pellet was washed with 80% ice-cold ethanol. The pellet was left to air-dry for 5 minutes before being resuspended in endotoxin-free tris-EDTA (TE) buffer (10 mM tris, 0.1 mM EDTA).

2.20 Transformation

Transformation of competent cells was performed according to manufacturer instructions. NEB Stable Competent *E. coli* (New England Biolabs, USA) were used routinely for constructs containing repetitive regions or regions of homology (pTRIPZ and AAVS1_Puro_Tet3G_3xFLAG_Twin_Strep). NEB® 5-alpha Competent *E. coli* (New England Biolabs, USA) was used as an alternative when transformation and selection of NEB Stable transformants was unsuccessful. Transformants were plated onto lysogeny broth (LB) agar plates containing ampicillin (100 µg/mL) and were incubated overnight at 37°C.

2.20.1 Colony PCR

Single bacterial colonies were picked and resuspended in 50 µL sterile water in a 96-well plate and were used in the PCR reaction shown in Table 2.10 with conditions shown in Table 2.11.

Table 2.10 Typical reaction mixture for colony PCR.

Component	Stock	Final Concentration
OneTaq® Quick-Load® 2X Master Mix with Standard Buffer (New England Biolabs, USA)	2X	1X
Forward primer	10 µM	0.2 µM
Reverse primer	10 µM	0.2 µM
Template DNA	-	1 µL resuspended colony
Nuclease-free water	-	To final reaction volume

Table 2.11 Typical cycle conditions for colony PCR.

Step	Temperature	Time	Cycles
Denaturation	94°C	2 minutes	1
Denaturation	94°C	30 seconds	30
Annealing	*Primer specific*	20 seconds	
Extension	72°C	1 minute/kb	
Final Extension	72°C	5 minutes	1
Hold	4°C	∞	

2.20.2 Starter cultures and stocks

The remaining bacterial suspension from 2.20.1 or single colonies from agar plates were used to inoculate 5 mL LB containing ampicillin (100 µg/mL). Cultures were incubated overnight at 37°C with shaking (220 revolutions per minute (rpm)).

2.20.3 Bacterial Stocks

200 µL of an overnight culture was added to 200 µL 40% sterile glycerol, thoroughly mixed and stored at -80°C. Stocks were streaked out onto LB agar plates to obtain single colonies prior to initiating starter cultures.

2.21 Plasmid extraction

2.21.1 Miniprep

Plasmid DNA was extracted from 1 – 5 mL overnight cultures using the Monarch® Plasmid Miniprep Kit (New England Biolabs, USA) according to manufacturer instructions.

2.21.2 Maxiprep

500 µL to 1 mL of a starter culture grown for 8 hours at 37°C with shaking (220 rpm) was used to inoculate 100 mL LB + ampicillin (100 µg/mL). The maxi-culture was grown overnight at 37°C with shaking (220 rpm). Plasmid DNA was extracted and prepared for cell transfections using the EndoFree Plasmid Maxi Kit (Qiagen, USA) according to manufacturer instructions.

2.22 Plasmid Site-directed mutagenesis

Site-directed mutagenesis was performed using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs, USA) according to manufacturer instructions.

2.23 Sequencing

All sequencing was performed by Source BioScience, UK. Plasmids were diluted to 100 ng/µL and PCR products were diluted to 10 ng/µL. Primers were supplied at 3.2 µM.

2.24 Software

Constructs were designed and sequence alignment was performed using SnapGene software (Insightful Science; available at snapgene.com). GraphPad Prism version 9.0 was used for making graphs and statistical analysis.

2.25 Mammalian cell genomic DNA extraction

Genomic DNA was extracted from cells using the Monarch® Genomic DNA Purification Kit (New England Biolabs, USA) according to manufacturer instructions.

2.26 Transient transfection

All cell lines were plated 24 hours prior to forward transfection such that they were 70-90% confluent at time of transfection. BFTC-905 cells were transfected with Lipofectamine™ 3000 Transfection Reagent (Invitrogen, USA) according to manufacturer instructions. SW780, RT-112 and 5637 cells were transfected with FuGENE® 6 Transfection Reagent (Promega, USA) according to manufacturer instructions.

2.27 Generating stable cell lines

To generate *AAVS1* knock-in stable cell lines, cells were reverse transfected with ribonucleoprotein (RNP) complexes using Lipofectamine™ RNAiMAX (Invitrogen, USA) according to manufacturer instructions (Integrated DNA Technologies, USA; online protocol Alt-R CRISPR-Cas9 system—RNP transfections). All experiments used Alt-R® CRISPR-Cas9 crRNA (T2-guide

sequence ggggccactagggacaggatTGG (Mali et al., 2013)) complexed with Alt-R® CRISPR-Cas9 tracrRNA fluorescently tagged with ATTO™ 550, assembled into an ribonucleoprotein complex (RNP) with Alt-R® S.p. HiFi Cas9 Nuclease V3 protein (Integrated DNA Technologies, USA). Final RNP concentration was 10 nM. When using plasmid DNA homology-directed repair (HDR) donors, cells were imaged 18 hours post-transfection to monitor RNP transfection, media was replaced with fresh media supplemented with 10% tetracycline-free HyClone™ FBS (GE Healthcare, USA; referred to as Tet-free) and cells were forward transfected with plasmid as described in section 2.26. When using linear dsDNA or ssDNA templates, cells were reverse co-transfected with T2-RNP and 1µg DNA.

2.27.1 Linear DNA donors

dsDNA donors for generating stable cell lines were generated by PCR (2.15). ssDNA donors were generated using the Guide-it Long ssDNA Production System v2 (Takara Bio, Japan) according to manufacturer instructions. Linear donor DNA was isopropanol precipitated prior to transfection (2.19).

2.27.2 Antibiotic selection

After CRISPR/Cas9 RNP/HDR or plasmid transient transfection, antibiotic selection was used to obtain a pool of cells that had integrated the transgene into their genome. 48 hours post-plasmid transfection media was changed to Tet-free media containing the concentration of selection drug determined for each cell line (Table 2.12). Cells were grown in media containing selection until all untransfected control cells were dead. Once stable lines were established and stocks frozen, they were routinely grown in media supplemented with 10% tetracycline-free FBS (PAN-Biotech GmbH, Germany; referred to as Tet-low) containing maintenance selection at half the concentration used during selection. Maintenance selection was removed prior to using the stable cell lines for experiments and doxycycline was used to induce gene expression.

Table 2.12 Concentration of puromycin and blasticidin used for selection of UCC lines.

Cell Line	Puromycin	Blasticidin
BFTC-905	1 µg/mL	8 µg/mL

RT-112	1 µg/mL	6 µg/mL
SW780	1 µg/mL	4 µg/mL
NHU-TERT B	2 µg/mL	4 µg/mL

2.28 BioID identification of interacting partners

2.3 x 10^6 cells were grown in 150 mm tissue culture dishes (Thermo Fisher Scientific, USA) for 24 hours before the addition of the indicated concentration of doxycycline to induce protein expression. 48 hours after the addition of doxycycline, the media was changed to contain 500 µM biotin for the indicated times. Biotinylation was stopped by washing the dishes with 10 mL ice-cold PBS on ice, four times. Dishes were stored at -80°C until sample preparation.

2.28.1 Cell lysis

1.5 mL lysis buffer (50 mM tris pH 7.5, 150 mM NaCl, 0.5% SDS, 1% SDC, 1% Triton X-100) was added and the dish was scraped with a cell scraper to ensure efficient detachment of cells. The lysate was moved to a 5 mL microcentrifuge tube and pipetted gently 10 times. The lysate was sonicated on ice for 2 cycles of 30 seconds using an MSE Soniprep 150 prior to the addition of 1 mL of 50 mM tris pH 7.5 and a further 30 second sonication cycle on ice. Protein concentration was determined using the BCA assay. An aliquot of the lysate was saved for input.

2.28.2 Affinity purification of biotinylated proteins

150 µL Strep-Tactin® Sepharose® beads (IBA Lifesciences, Germany) were centrifuged at 600 x g for 2 minutes and washed twice with Buffer W (100 mM tris pH 8.0, 150 mM NaCl, 1 mM EDTA) (IBA Lifesciences, Germany). The beads were then incubated with 5 mg protein lysate over night with rotation at 4°C. The beads and supernatant were separated by centrifugation and the supernatant was saved as the unbound fraction. The beads were washed once in wash buffer 1 (50 mM tris pH 7.5, 150 mM NaCl, 2% SDS, 1% SDC, 1% Triton X-100), once in lysis buffer and twice in wash buffer 2 (50 mM tris pH 7.5, 250 mM lithium chloride (LiCl), 0.5% SDC, 0.5% NP-40 substitute). The beads were washed with 50 mM triethylammonium bicarbonate (TEAB; Sigma-Aldrich, USA) three times

and stored 'dry' at -20°C prior to shipment on dry ice to the Core Proteomic and Metabolomics Facility (ICR). Sample preparation, mass spectrometry and data processing were performed by Dr Lu Yu. Statistical analysis was performed by Dr Lu Yu and Dr Theo Roumeliotis.

2.28.3 Western blot

The beads were resuspended in 20 µL 2X NuPAGE™ LDS sample buffer and 1X and NuPAGE™ Sample Reducing Agent and heated to 70°C for 10 minutes. The beads were centrifuged at 600 x g for 2 minutes and the supernatant was moved to a new tube. The supernatant was used for western blot as previously described (2.11).

2.28.4 Sample preparation

Sample preparation was performed, and methodology details provided, by Dr Lu Yu (ICR). Dry beads were resuspended in 180 µl 5 mM tris(2-carboxyethyl) phosphine hydrochloride solution in 100 mM TEAB and incubated for 10 minutes at 25°C with shaking at 1,000 rpm. Iodoacetamide (Sigma-Aldrich, USA) was added to a final concentration of 10 mM and the beads were incubated for 30 minutes at 25°C with shaking at 1,000 rpm in the dark. 1 µg trypsin (MS grade, Thermo Fisher Scientific, USA) was added, and protein digestion proceeded for 18 hours at 37°C with shaking at 1000 rpm. Peptides were completely dried using a SpeedVac Vacuum Concentrator, resuspended in ddH_2O, and dried again. Peptides were resuspended in ddH_2O, and protein concentration was estimated by measuring absorbance at 280 nm using a Nanodrop. One third of the sample was labelled with 0.25 mg TMTproTM reagents (Thermo Fisher Scientific, USA) according to manufacturer instructions. After 1 hour incubation at room temperature (RT), the reaction was quenched for 15 minutes with 2 µl 5% hydroxylamine (Thermo Fisher Scientific, USA), the labelled peptides for each biological replicate were combined and dried in a SpeedVac. The dried sample was resuspended in in 0.1% ammonium hydroxide (NH_4OH) and fractionated on a XBridge BEH C18 column (2.1 mm i.d. x 150 mm; Waters Corporation, USA); fractions were concentrated to 8 fractions and dried in a SpeedVac.

2.28.5 Tandem Mass Tag Liquid Chromatography Mass Spectrometry (TMT-LC-MS/MS)

LC-MS/MS was performed, and methodology details were provided, by Dr Lu Yu (ICR). Fractions were resuspended in 112 µl 0.5% formic acid (FA) and 10 µl was injected for LC-MS/MS analysis on the Orbitrap Fusion Lumos mass spectrometer coupled with U3000 RSLCnano UHPLC system. Raw files were processed in Proteome Discoverer 2.4 (Thermo Fisher Scientific, USA) using the Sequest HT search engine to search against the reviewed Homo Sapiens Uniprot database (Version January 2022). Search results were validated by Percolator and uniquely identified peptides were considered for quantification.

2.28.6 Statistical analysis

Statistical analysis was performed, and methodology details were provided, by Dr Theo Roumeliotis (ICR). For each biological replicate, raw signal intensity for each protein was scaled by dividing by the mean signal intensity across the five treatment conditions for each bait and $\log_2$ transformed. To determine putative interactors across the whole experiment using both the N- and C-terminally tagged baits, a linear model with the bait as the independent variable was generated for each protein identified and its correlation to A3A was caluclated. Putative interactors were defined as those having a strong correlation with A3A (correlation ≥ 0.85).

To generate a heatmap of identified proteins, raw protein abundances (signal-to-noise values exported from Proteome Discoverer) were log2 scaled to the mean of each bait, protein-wise. Scaled values for the two replicates were averaged and subjected to hierarchical clustering with Euclidean distance using the Phantasus tool (https://artyomovlab.wustl.edu/phantasus/). For statistical testing, log2 ratios + Dox vs - Dox for 4 and 6 hour biotin treatments were calculated and used as replicates for a one-sample t test using the Perseus platform (Tyanova et al., 2016).

2.29 RNA sequencing

Total RNA was prepared from cells using the ReliaPrep™ RNA Miniprep System (Promega, USA) and sent to the Beijing Genomics Institute (BGI) Inc., Hong Kong on dry ice. Quality control was performed by BGI using the BioAnalyzer 2100

(Agilent). DNBSEQ strand-specific mRNA libraries were prepared and sequenced on the BGISEQ-500 platform with a paired-end read length of 100 (PE100) and 30 million reads per sample. All RNA sequencing (RNAseq) data analysis was performed by Dr Pradeep Ramagiri with input and guidance from Dr Konstantinos Mitsopoulos.

2.29.1 QC, read alignment and count generation

Raw FastQ files were processed using the nf-core/rnaseq pipeline (version 21.04.3). Briefly, reads were QC checked using FastQC v0.11.9 (Andrews, 2010) and MultiQC v1.12 (Ewels et al., 2016) and UMI extraction was performed using UMI-tools (Smith et al., 2017). Adapter and quality trimming was done with Trim Galore! (Krueger) and genome contaminants were removed (BBSplit), prior to alignment to the reference genome (GRCh38) using STAR v.2.7.6a (Dobin et al., 2013). Once the reads were aligned, HTSeq-count (Putri et al., 2022) was used to count the number of reads mapping unambiguously to genomic features in each sample.

2.29.2 Differential expression and pathway enrichment analysis

Differential expression analysis of the count data was performed in R using the Bioconductor package DESeq2 v1.20.0 (Love et al., 2014). Pathway and Gene Set Enrichment Analysis was carried out in R using package fgsea v1.0 (Korotkevich et al., 2021).

2.29.3 RNA editing analysis

The Genome Analysis Toolkit (GATK) best practices guidelines were followed in the creation and implementation of the pipeline for calling mutations in RNASeq data (Figure 2.1) (Van der Auwera & O'Connor, 2020). Variants were called by Mutect2 (Benjamin et al., 2019), filtered by ‘Variant Filtering’ function using GATK v4.1.9.0. and mutations were annotated using snpEff version 5.1 (Cingolani et al., 2012).

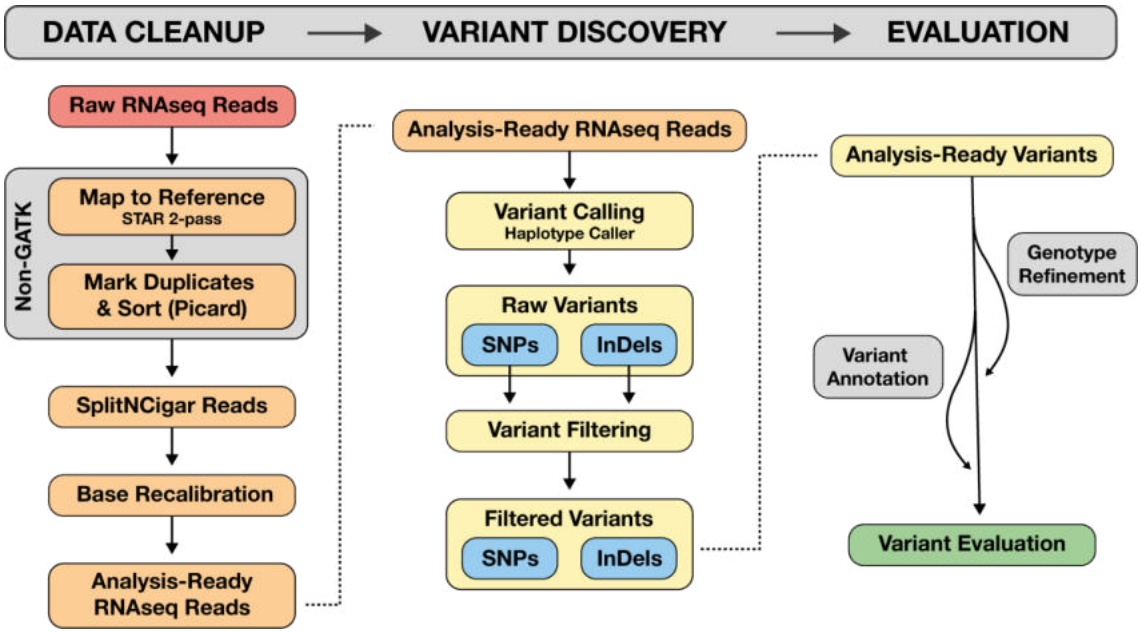


Figure 2.1 GATK best practices for variant analysis of RNA sequencing data sets.

Figure adapted from https://gatk.broadinstitute.org/hc/en-us/articles/360035531192?id=3891.

Chapter 3

Chemotherapy induces expression of *A3A* and *A3B* in UCC

Chapter 3 Chemotherapy induces expression of *A3A* and *A3B* in UCC

3.1 Introduction

Many human cancers have an A3 mutational signature that acts as a molecular footprint of A3 deamination activity (Alexandrov et al., 2020; Alexandrov et al., 2013). As discussed in detail in Chapter 1, these signatures are primarily assigned to A3A and A3B based on their sequence context, and while *A3B* expression correlates with mutational load, *A3A* does not, despite emerging evidence that the majority of the A3 mutational signature is attributed to A3A (Chan et al., 2015; Cortez et al., 2019; Law et al., 2020). It has been proposed that the lack of correlation with *A3A* is due to episodic bursts of expression that generate mutations, but that the expression is transient, quickly returning to baseline, and is largely undetectable at the time of sample processing (Oh et al., 2021; Petljak et al., 2019). In support of this, Petljak et al., (2019) isolated 'parental' clones from the starting cell line and cultured them for 161 days, after which this parental line was subcloned to isolate 'daughter' clones. Whole exome sequencing (WES) or WGS was performed on both the parental and daughter clones, and to ensure any detected mutations were not present in the starting clone, the parental clones were used as the reference sequence. A striking result was that the prevalence of the A3 mutational signatures was highly variable even from daughters derived from the same parental clone. The authors also found that within a daughter clone, the signature was not accumulated at a constant rate. Instead, serial sampling revealed that large numbers of mutations are generated in a short period of time, followed by long stable periods of minimal signature accumulation.

3.1.1 APOBEC3s are induced in response to cellular stress

There are numerous studies on how *A3A* and/or *A3B* expression becomes dysregulated in tumour cells and what may be driving the episodic bursts in expression. Several A3 family members are upregulated in response to a diverse range of cellular stresses commonly experienced by tumour cells during their development and progression, offering insight into how tumours accumulate large numbers of A3-mediated mutations. The most well characterised role for A3 proteins is as a key part of the innate immune response via viral restriction and

their upregulation in response to viral insult is widely accepted to be primarily mediated by the interferon response (Moris et al., 2014). *A3s* are also upregulated and/or their deamination activity potentiated in response to hypoxia and cellular crowding. Widespread RNA editing occurring in monocytes and macrophages in response to hypoxic conditions or cellular crowding is largely mediated by A3A (Alqassim et al., 2021; Sharma et al., 2015), while mitochondrial hypoxic stress and cellular crowding induces RNA editing by A3G in natural killer cells (Sharma et al., 2019). In addition, cyclic hypoxia increases both the expression and activity of A3B (Bader et al., 2021). Of relevance is the recent discovery that commonly used chemotherapy drugs induce the expression of *A3A* and/or *A3B* (Kanu et al., 2016; Middlebrooks et al., 2016; Oh et al., 2021; Periyasamy et al., 2021; Yamazaki et al., 2020). This has conceivable clinical implications regarding drug-induced mutational activity, heterogeneity and potentially tumour evolution and promotion of drug resistance. This induction is likely responsible for the increased mutational signature seen in chemotherapy resistant UCC (Faltas et al., 2016).

At the start of this project, two studies had demonstrated that commonly used anticancer agents induce expression and activity of *A3A* and/or *A3B* in both breast and bladder cancer cell lines, but characterisation of the mechanism behind this induction was limited. Middlebrooks et al., (2016) showed that treatment of cells with the DSB-inducing drug, bleomycin, induces expression of both *A3A* and *A3B* in multiple breast and bladder cancer cell lines. While the authors did not investigate the mechanism behind the induction, they did show that interferon-stimulated genes upregulated with *A3A* and *A3B* after Sendai Virus infection, were not induced in response to bleomycin treatment. This suggests that DNA damage-mediated induction occurs via a different process than viral infection and is not mediated by the interferon response. Kanu et al., (2016) showed that a range of anticancer agents induce expression of *A3B* and *A3G*, with drugs inducing RS (hydroxyurea, aphidicolin and gemcitabine) being the most robust, suggesting that RS is a major source of *A3B* expression in breast cancer. During the course of the current work, three additional studies were published, describing A3 upregulation in response to genotoxic stress by anticancer agents with investigations into the mechanisms (Oh et al., 2021; Periyasamy et al., 2021; Yamazaki et al., 2020).

p53 is commonly referred to as the “guardian of the genome” and plays an important role in a diverse set of cellular processes in response to stress (Hafner et al., 2019) including *A3* regulation. *A3B* overexpression correlates with *TP53* inactivation/mutation in many tumour types (Burns, Lackey, et al., 2013; Cescon et al., 2015; Kanu et al., 2016; Periyasamy et al., 2017; Roper et al., 2019; Silwal-Pandit et al., 2014). The mechanistic link between p53 and *A3B* regulation was first established when two studies demonstrated that knockdown of E6, a viral oncoprotein known to inactivate p53, decreases expression of *A3B* in human papillomavirus (HPV) positive cell lines (Ohba et al., 2014; Vieira et al., 2014). p53 has since been shown to be a key transcriptional repressor of *A3B*. Menendez et al., (2017) showed that *A3B* induction occurs in response to the p53 activating drugs, nutlin and doxorubicin, in a range of p53-deficient but not in p53-proficient cell lines, that *A3B* repression in wild-type p53 lines is alleviated through p53 knockdown, and expression of wild-type p53 in p53-null lines inhibits induction of *A3B.* Periyasamy et al., (2017) specifically investigated the functional role of p53 in *A3B* repression in breast cancer cell lines and showed that nutlin treatment reduces *A3B* expression in wild-type p53 lines, but has no effect in those with mutant p53. The repression is not through direct p53 binding to the *A3B* promoter but rather by p21-dependent recruitment of the repressive E2F4/p107/p130-containing DREAM complex (Periyasamy et al., 2017). A *kataegic*-like A3 mutational signature is seen in A3B-overexpressing HEK293 cells depleted of p53, suggesting that p53 pathway loss facilitates accumulation of the mutational signature (Nikkilä et al., 2017). In contrast to *A3B*, *A3A* upregulation in response to DNA damage is p53-independent (Menendez et al., 2017) suggesting *A3A* is not regulated by the p53 pathway in the same way as *A3B*. Induction of *A3A* and *A3B* occurs in response to the DSB-inducing drug, bleomycin in both a mutant and wild-type UCC cell line suggesting that induction in this tumour type is not strictly dependent on p53 status (Middlebrooks et al., 2016).

3.1.2 Chapter aims

The work in this chapter aimed to fully characterise a panel of UCC cell lines and use appropriate models to investigate induction of *A3A* and *A3B* in response to commonly used chemotherapy drugs. Multiple drugs with differing mechanisms

of action were used to determine whether induction is related to the type of DNA damage generated or if it is a more general response to drug-induced stress. Determination of p53 response will also allow further investigation into whether p53 status impacts the ability to induce *A3* expression in response to chemotherapy treatment in this tumour type.

3.2 Results

3.2.1 Initial characterisation of a human UCC cell line panel

A panel of 12 UCC cell lines with genetic alterations commonly identified in UCC patients, including *TP53* and *RB* pathway mutations and *FGFR3* fusions, were chosen for characterisation to determine suitability for the work in this thesis (summarised in Table 3.1). Nine cell lines are derived from MI-UCC tumours, and three are derived from early stage (Ta or T1) tumours and can be used as models for NMI-UCC. The cell lines are classified as either luminal, basal, or non-type (mixed expression of both luminal and basal markers) as determined by Warrick et al., (2016). Whole exome sequencing of 25 UCC cell lines identified the presence of SBS2 and SBS13 suggesting that commercially available cell lines have mutational signatures concordant with those identified in patients and are suitable for studying A3s (Nickerson et al., 2017). The UCC panel used in this thesis consists of cell lines characterised by Nickerson et al., (2017) (except BFTC-905 and 647-V).

Table 3.1 UCC cell line panel information.

Information collated from the CCLE, Williams et al., (2013), Earl et al., (2015), Warrick et al., (2016), and Nickerson et al., (2017).

Cell line	Sex	Subtype	Key genetic alterations	Model	Driver status (Earl et al., 2015)
UM-UC-3	M	Non-type	*CDKN2A, KRAS, PTEN, TP53, TERT*	MI	Tumour suppressor driven
UM-UC-6	M	-	*FGFR3, PIK3CA*	Disputed	FGFR3-driven
HT-1197	M	Basal	*FGFR3, NRAS, PIK3CA, TERT, TP53*	MI	-
HT-1376	F	Basal	*RB1, TP53, TERT*	MI	Tumour suppressor driven
RT-112	F	Luminal	*FGFR3-TACC3*	NMI	FGFR3-driven
T-24	F	Non-type	*HRAS, TP53, TERT*	MI	Tumour suppressor driven
647-V	M	Basal	*TP53, RB1, TERT*	MI	Tumour suppressor driven
BFTC-905	F	Basal	*NRAS, TP53*	MI	-
SW780	F	Luminal	*CDKN2A, FGFR3-BAIAP2L1, TERT*	NMI	FGFR3-driven
TCCSUP	F	Non-type	*PIK3CA, RB1, TP53, TERT*	MI	Tumour suppressor driven

5637	M	Basal	*RB1, TP53, TERT, ATM, ERRB2*	MI	Tumour suppressor driven
RT4	M	Luminal	*CDKN2A, TSC1, TERT, FGFR3-TACC3*	NMI	FGFR3-driven

Doubling times and optimal seeding density for future IC_{50} experiments using the SRB assay were determined. Doubling time was calculated during the log phase of growth (Figure 3.1). Cells were seeded at several densities and grown for 6 days, the approximate length of future IC_{50} experiments, the SRB assay was performed, and optimal seeding density was determined from the linear part of the graph (Figure 3.2). A summary of doubling times and optimal seeding densities for IC_{50} determinations is shown in Table 3.2.

Table 3.2 Doubling times and optimal seeding densities of the UCC cell line panel.

Cell line	Doubling time (hours)	96-well plate (cells/well)
647-V	26	2000 – 4000
UM-UC-6	21	500 – 1000
HT-1197	43	4000
RT-112	24	1000 – 2000
BFTC-905	26	1000 – 2000
SW780	21	1000 – 2000
T-24	20	500 – 1000
RT4	38	4000 – 8000
TCCSUP	30	4000 – 8000
5637	24	2000
HT-1376	39	4000 – 8000
UM-UC-3	25	1000 – 2000

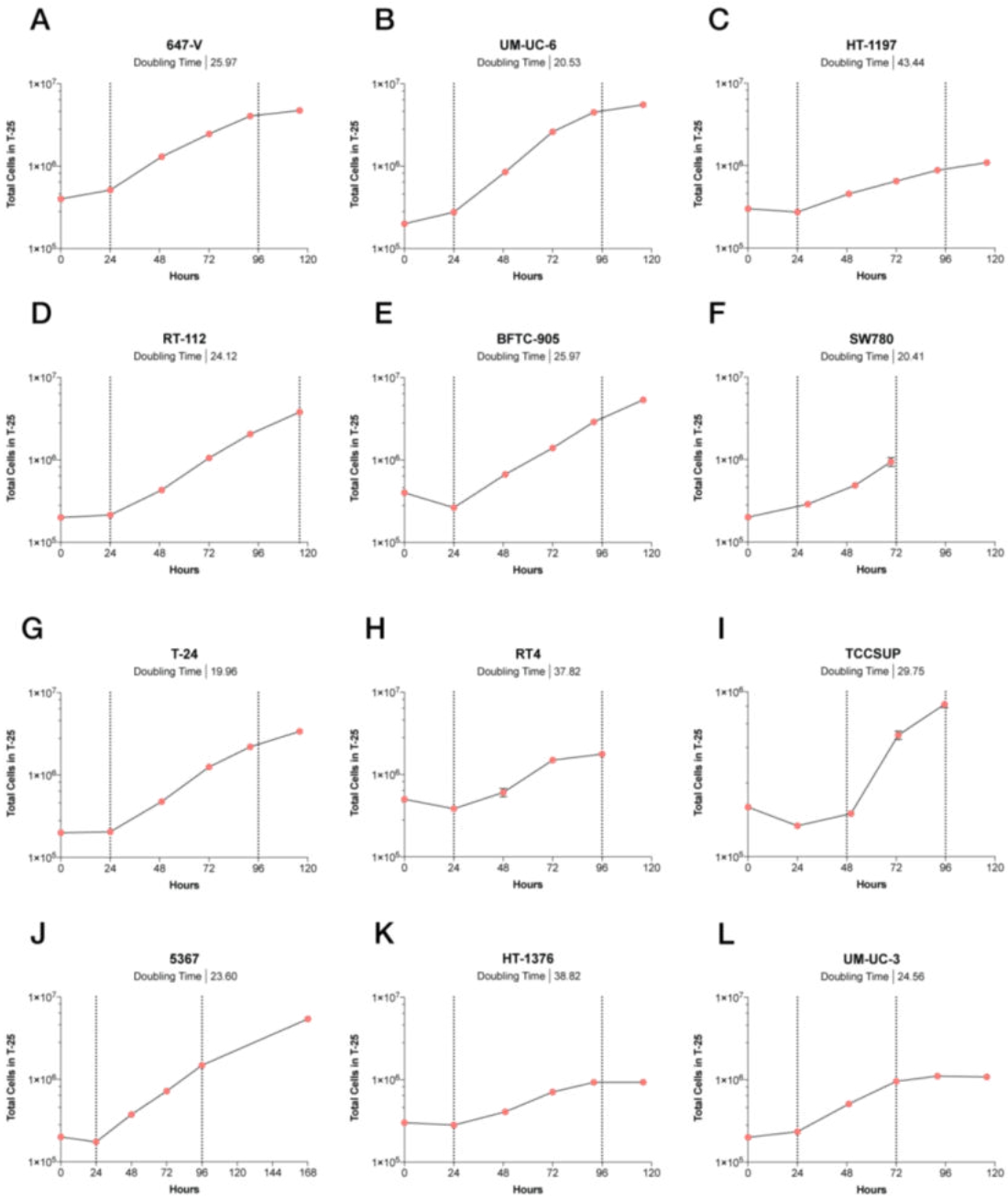


Figure 3.1 Growth curves for the UCC cell line panel.

647-V (A), UM-UC-6 (B), HT-1197 (C), RT-112 (D), BFTC-905 (E), SW780 (F), T-24 (G), RT4 (H), TCCSUP (I), 5637(J), HT-1376 (K) and UM-UC-3 (L) cells were seeded in two T-25 flasks and both flasks were counted approximately every 24 hours. Graphs and doubling times were generated using GraphPad Prism 9. Dotted lines show period of growth used to calculate doubling time.

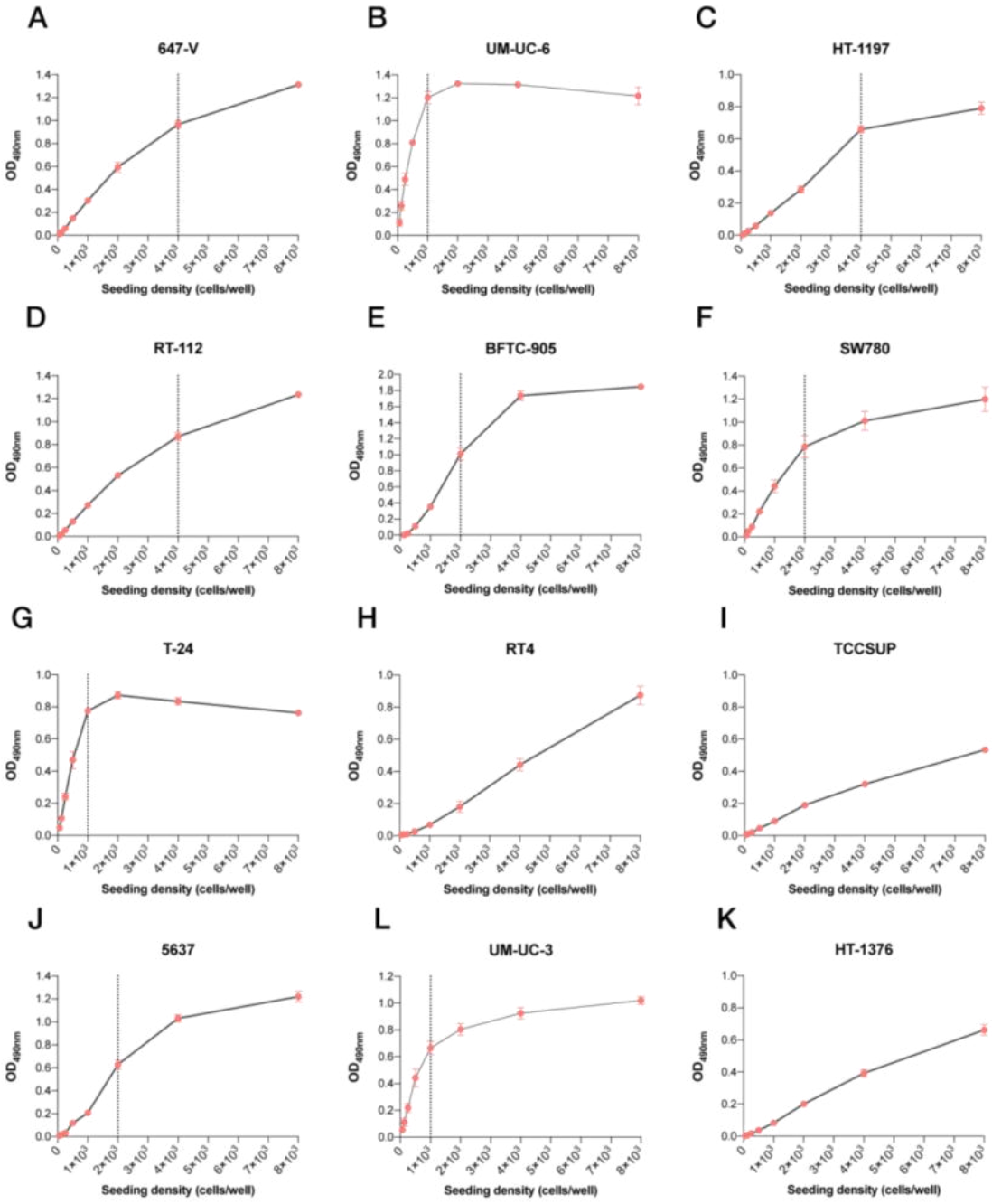


Figure 3.2 Seeding density determinations for the UCC cell line panel.

647-V (A), UM-UC-6 (B), HT-1197 (C), RT-112 (D), BFTC-905 (E), SW780 (F), T-24 (G), RT4 (H), TCCSUP (I), 5637(J), HT-1376 (K) and UM-UC-3 (L) cells were seeded at a range of densities in 96-well plates and left for 6 days. Optimal seeding density was determined using the SRB assay, from the linear part of the graph with maximal seeding density as the seeding density at which OD_{490nm} did not exceed 1.2 (dotted line). Data is from a single experiment with 6 technical replicates.

The tumour suppressor p53 is implicated in the transcriptional regulation of *A3B* (Menendez et al., 2017; Periyasamy et al., 2017) and the loss or mutation of p53 is associated with the A3 mutational signatures in different tumour types (Burns, Lackey, et al., 2013; Cescon et al., 2015; Kanu et al., 2016; Periyasamy et al., 2017; Roper et al., 2019; Silwal-Pandit et al., 2014). To determine whether p53 status influences baseline expression of *A3B* and induction in response to chemotherapy treatment in UCC as seen in other tumour types, the functional status of p53 in each of the UCC cell lines was determined (Figure 3.3). In the presence of DNA damage, p53 is phosphorylated, the negative regulator MDM2 is displaced and p53 rapidly accumulates, inducing expression of the downstream target, *CDKNIA* (encoding p21). In response to DNA damage induced by treatment with the chemotherapy drug, gemcitabine, accumulation of p53 and p21 is seen in UM-UC-6, RT4, SW780 and HT-1197 and therefore, these cell lines have a wild-type response. There is no p53 or p21 accumulation in UM-UC-3, 5637, HT-1376, TCCSUP, T-24, RT-112 and 647-V cells and these cell lines are designated as having a mutant response. Despite a small induction of p21, p53 is not detected in BFTC-905 and is designated as having a mutant response. This is consistent with the literature where this cell line has been shown to overexpress *MDM2* (Cheng et al., 1995). For all cell lines, except HT-1197 that has been previously reported as being mutant p53, the results presented here are largely consistent with the mutational status reported in the literature and cell line databases (Table 3.1).

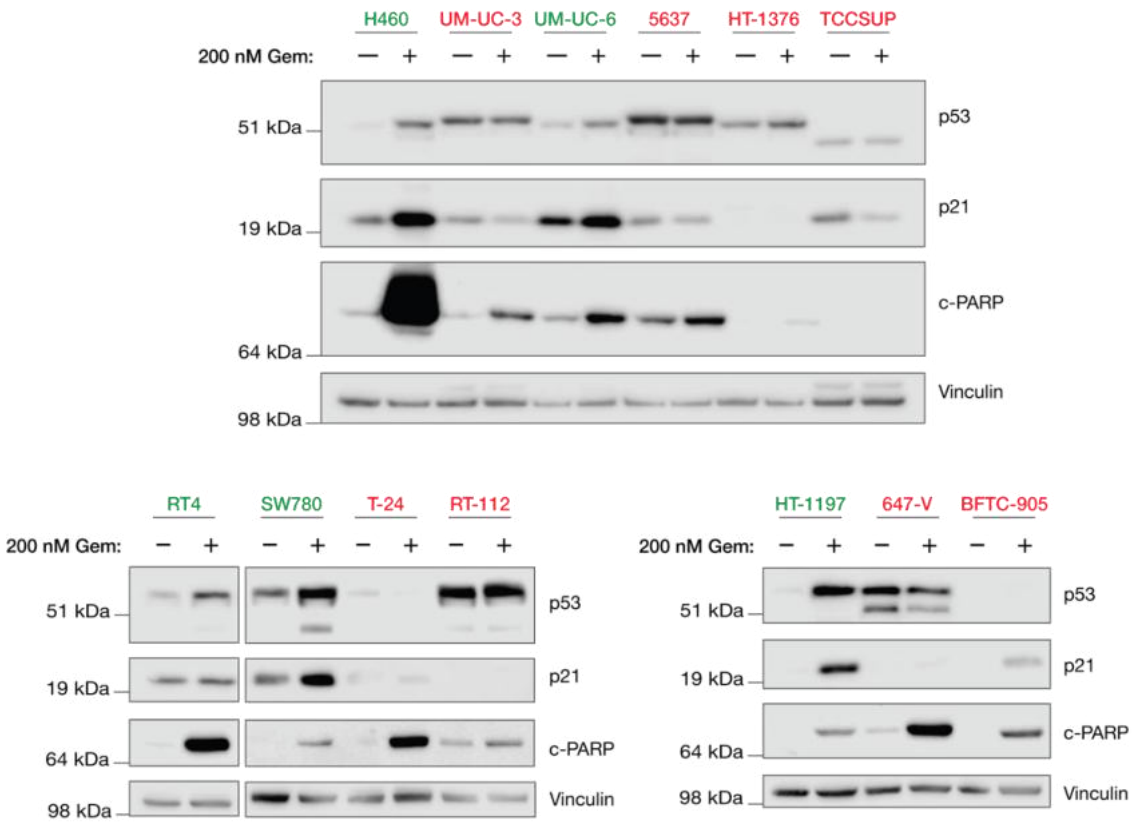


Figure 3.3 Functional p53 response to gemcitabine treatment in the UCC cell line panel.

Cells were seeded in 6-well plates and left to attach for one cell doubling prior to treatment with 200 nM gemcitabine (Gem). Cells were harvested after 24 hours, and 25 µg protein was used for western blotting. The H460 cell line confirmed to be wild-type p53 by Dr Caitlin McCarthy (ICR) was used as a positive control. Cleaved PARP (c-PARP) was used to monitor apoptosis and successful drug treatment; vinculin was used as a loading control. Cell lines with a wild-type p53 response are shown in green; cell lines with a mutant response are shown in red. Blots representative of three independent experiments.

3.2.1.1 The UCC panel has a range of *A3A* and *A3B* mRNA expression

UCC has the strongest A3 signature of all tumour types analysed and has elevated expression of *A3B* relative to normal tissue (Alexandrov et al., 2020; Burns, Temiz, et al., 2013; Swanton et al., 2015). To determine if expression is elevated in the cell line panel, endogenous *A3A* and *A3B* expression was measured by RT-qPCR and compared to a telomerase reverse transcriptase (TERT) immortalised cell line, NHU-TERT B (Chapman et al., 2006). This line is derived from normal urothelium and lacks genetic alterations associated with UCC development. Their karyotype, morphology and phenotype largely mimic

that of normal urothelium and as such they are a useful control for normal urothelial tissue. Four cell lines in the panel (BFTC-905, SW780, 5637 and HT-1376) express *A3A* (Figure 3.4 A). *A3B* is expressed in 11 cell lines in the panel and expression is elevated relative to the normal control in nine; *A3B* expression was undetectable in 647-V cells (Figure 3.4 B and C). Comparison of the ΔCt values of the two family members shows that *A3A* is expressed at much lower levels than *A3B.* Elevated *A3B* expression is seen in two NMI-UCC cell lines with wild-type p53 (SW780 and RT4) demonstrating that loss of p53 is not strictly required for elevated *A3B* expression in this tumour type and is seen in models of early-stage disease. *A3B* is expressed in all molecular subtypes confirming subtype is not a strict determinant of expression consistent with reported findings using patient tumour samples (Glaser et al., 2018). *A3A* is expressed in one luminal cell line but is more common in basal cell lines, again consistent with the literature.

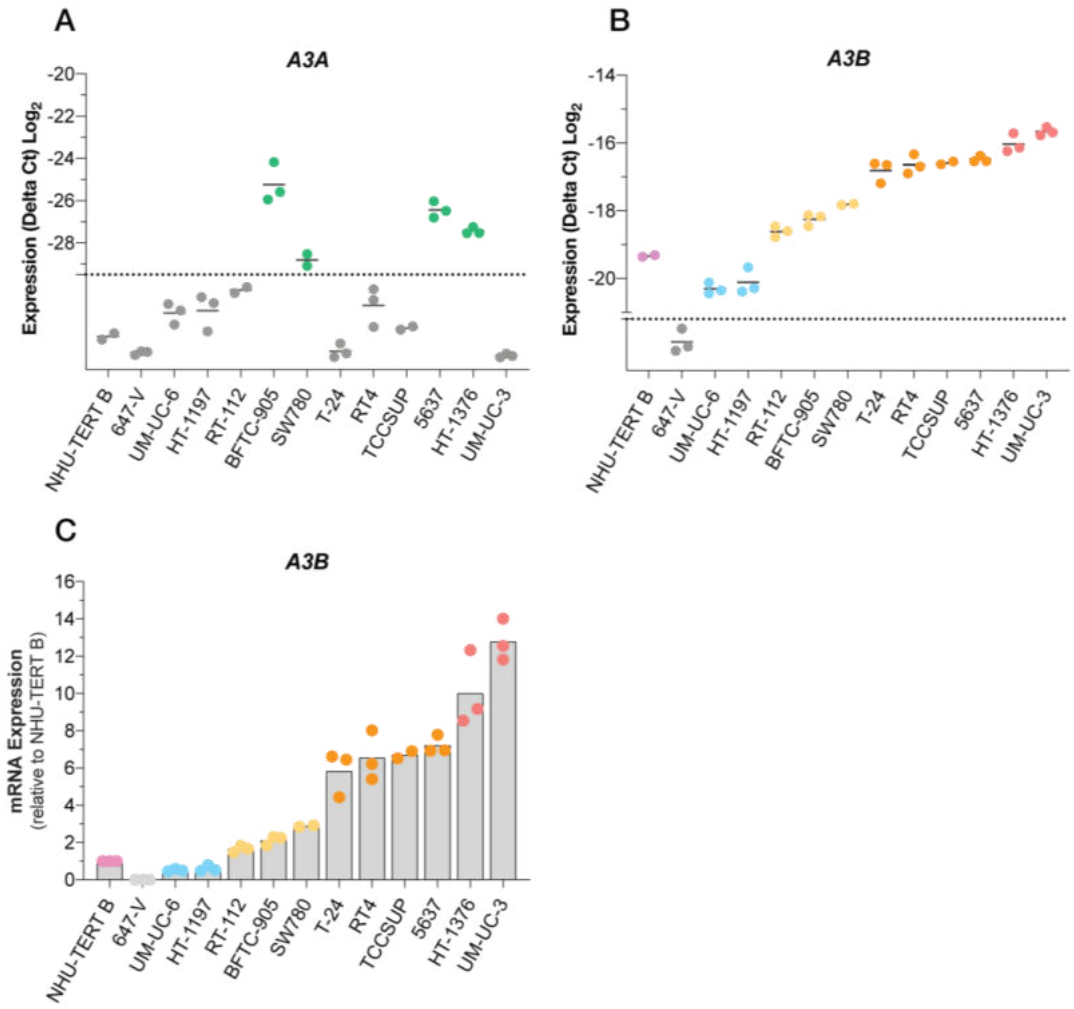


Figure 3.4 Endogenous *A3A* and *A3B* expression in a panel of 12 UCC cell lines.

Cells were harvested when 60-70% confluent prior to RNA extraction. mRNA expression was measured with RT-qPCR from the same cDNA samples for both genes and analysed using the ΔΔCt method. *A3A* (A) and *A3B* (B) expression are shown as ΔCt values to allow comparison of expression. (C) *A3B* expression shown relative to the 'normal' immortalised urothelial cell line control, NHU-TERT B. Dotted line denotes undetectable expression; dots represent independent experiments, line/bar represents mean. Cell lines with detectable *A3A* expression are shown in green. For B and C, colours indicate low, moderate, high, and very high expression of *A3B.*

3.2.1.2 *In vitro* A3 deamination activity correlates with *A3B* expression

Deamination activity in cellular lysates was determined using an *in vitro* deamination assay to establish whether the elevated mRNA expression is correlated with enzyme activity (Figure 3.5). As the substrate for the deamination assay contains a single cytosine within a 5'TCA motif that six of the seven (all but A3G) family members can act on (McDaniel et al., 2020), it is not possible to definitively determine which family member is responsible for the activity seen. However, as expression of *A3A* is either undetectable or low in comparison to *A3B,* and there is a significant positive correlation between *A3B* expression and activity (R^2 = 0.79; *P* = 0.00004), we can assume that the activity measured in lysates using this deamination assay is primarily due to A3B activity (Figure 3.5 D). siRNA knockdown was not used to validate the assumption that A3B is primarily responsible for the activity measured in lysates as at the time this work was conducted, there were concerns over the specificity of commercial siRNAs due to the large homology between the A3 family members. Protein expression detection by western blot could also not be conducted due to the lack of commercially available antibodies sensitive enough to detect endogenous A3A and A3B reliably and specifically.

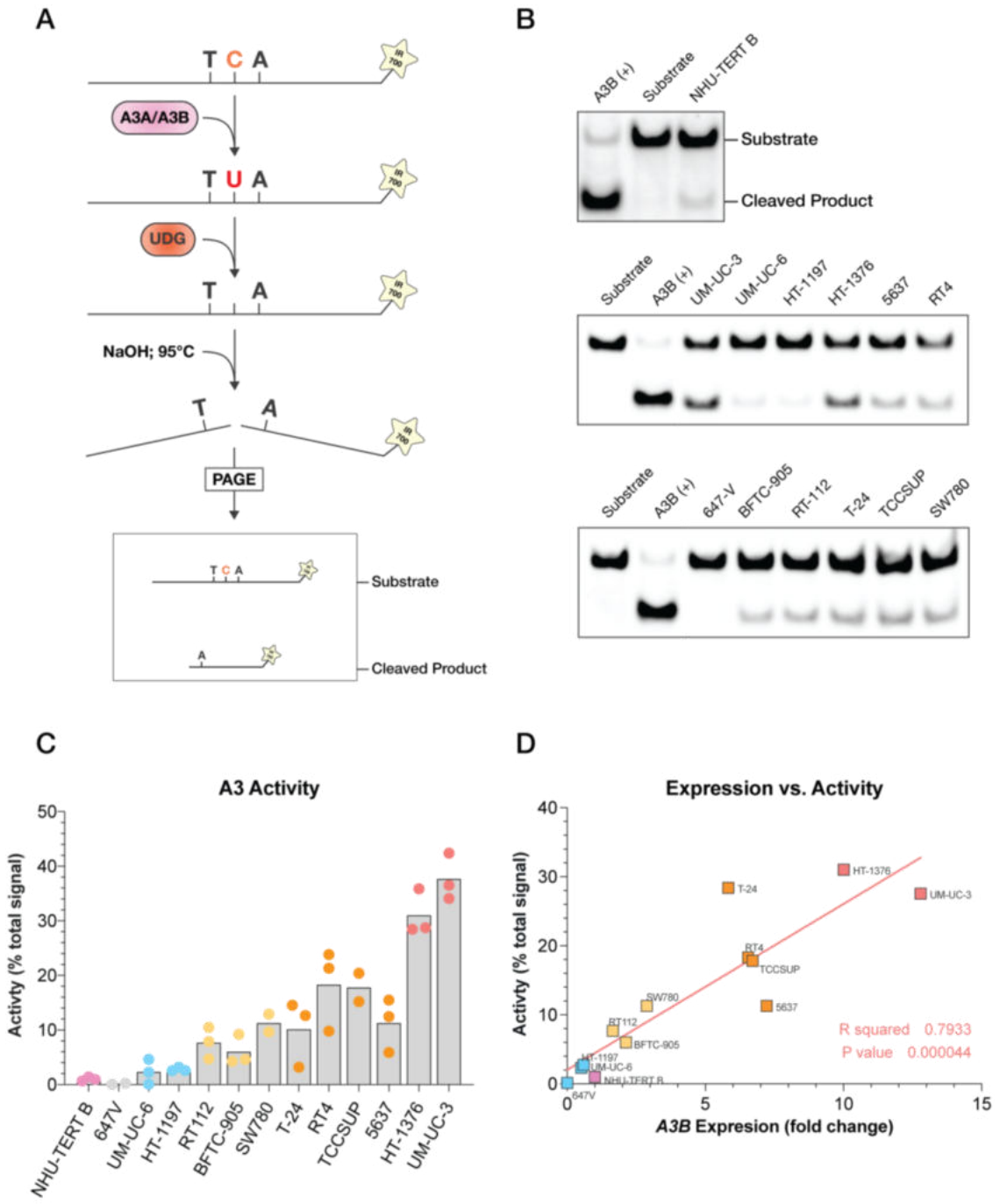


Figure 3.5 Deamination activity detected in lysates from the UCC panel is primarily attributed to A3B.

(A) Schematic demonstrating the *in vitro* deamination assay. A fluorophore-tagged ssDNA substrate containing a single cytosine within an A3 motif (TCA) is incubated with cellular lysates; if A3 activity is present the cytosine is deaminated. The resulting uracil is removed by uracil DNA glycosylase (UDG) leaving an abasic (AP) site. Addition of sodium hydroxide (NaOH) and heating to 95°C cleaves the substrate at the AP site. The substrate and cleaved products are size resolved using PAGE and visualised. (B) Representative polyacrylamide gel images showing deamination activity in lysates. Cells were harvested when 60-70% confluent prior to use of 25 µg protein lysate in the deamination assay. (C) A3 activity was quantified by representing the cleaved signal

as a percentage of the total signal in each lane. Dots represent independent experiments, bars represent mean. All cell lines, except 647-V, show detectable deamination activity in lysates. (D) A3 activity was plotted against *A3B* mRNA expression. Simple linear regression analysis was performed using GraphPad Prism 9 to determine the correlation co-efficient (R^2). Cell lines are coloured as in Figure 3.4.

3.2.1.3 Characterising sensitivity of the panel to three commonly used chemotherapy drugs

Chemotherapeutics induce expression of *A3A* and/or *A3B* in breast and bladder cancer cell lines (Kanu et al., 2016; Middlebrooks et al., 2016). However, induction in UCC has only been demonstrated in response to treatment with bleomycin, and it is not known whether drugs with other mechanisms of action also induce expression in this tumour type. To investigate this, sensitivity of the UCC cell lines to three commonly used agents, with differing mechanisms of action, was determined and the results used to guide the concentrations used in later experiments. Cell lines were treated with a range of concentrations of the DNA cross-linker, cisplatin (Figure 3.6), the antimetabolite, gemcitabine (Figure 3.7) and the microtubule-stabilising drug, paclitaxel (Figure 3.8), and IC_{50} values were determined using the SRB assay (summarised in Table 3.3).

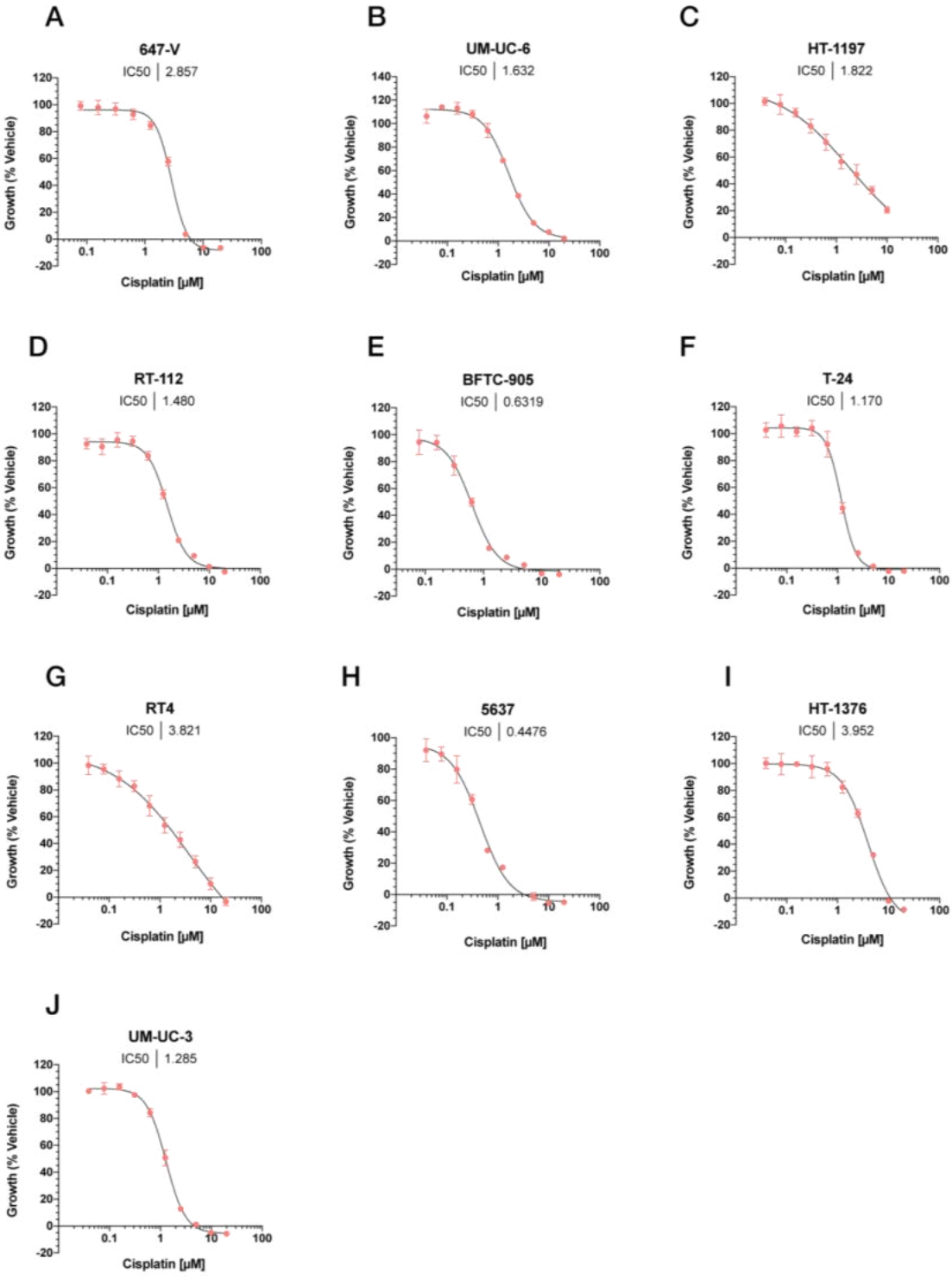


Figure 3.6 Concentration-response curves for cisplatin IC_{50} determination of 10 UCC cell lines.

647-V (A), UM-UC-6 (B), HT-1197 (C), RT-112 (D), BFTC-905 (E), T-24 (F), RT4 (G), 5637 (H), HT-1376 (I) and UM-UC-3 (J) cells were seeded at their optimal seeding densities and left to attach for 43 hrs prior to treatment with a range of cisplatin concentrations. The SRB assay was performed after 96 hrs drug treatment. Data shown as Growth (% Vehicle). IC_{50} was calculated by non-linear regression using GraphPad Prism 9. Data from a single experiment and values are mean of the four technical replicates ± SD.

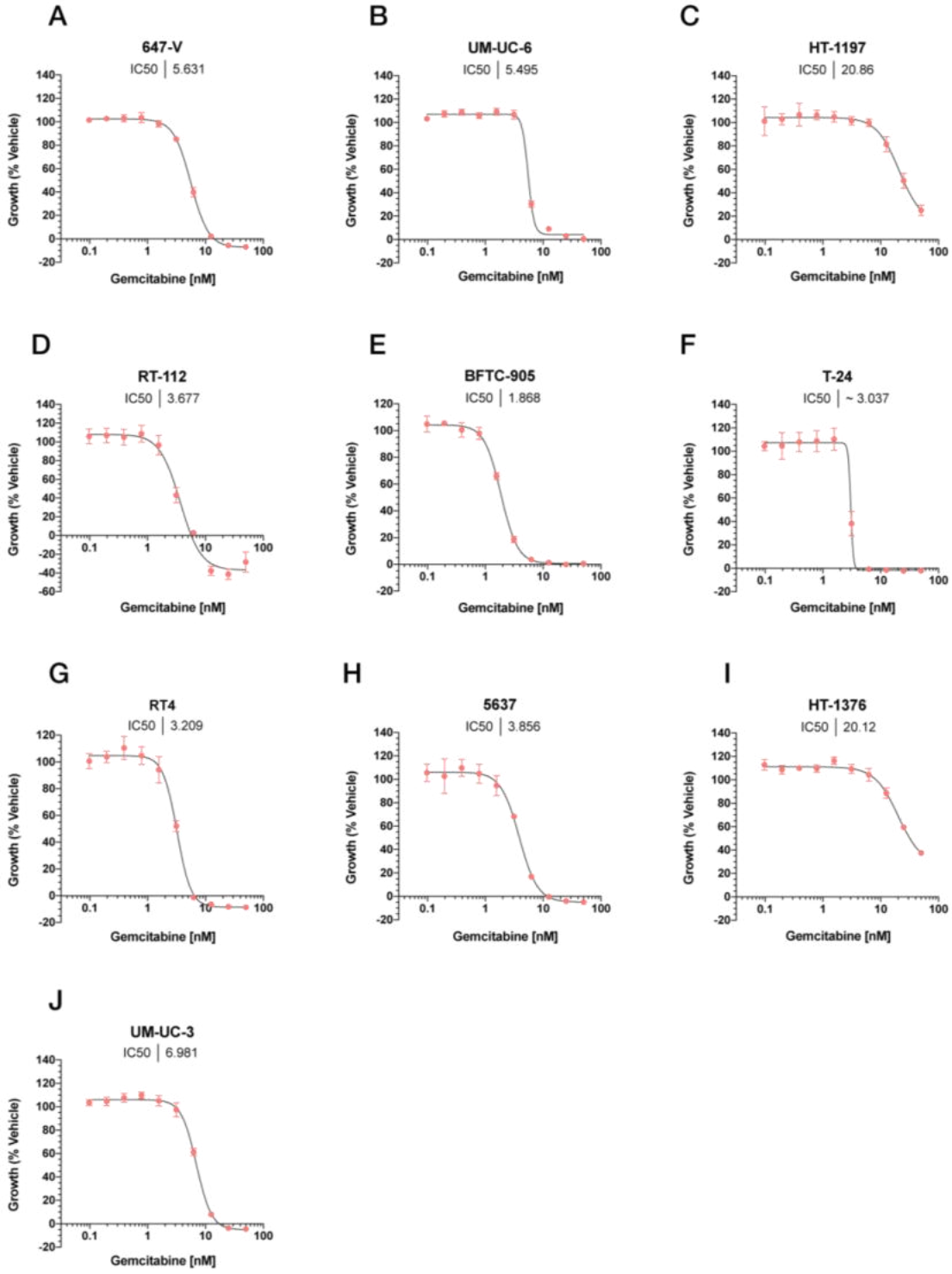


Figure 3.7 Concentration-response curves for gemcitabine IC_{50} determination of 10 UCC cell lines.

647-V (A), UM-UC-6 (B), HT-1197 (C), RT-112 (D), BFTC-905 (E), T-24 (F), RT4 (G), 5637 (H), HT-1376 (I) and UM-UC-3 (J) cells were seeded at their optimal seeding densities and left to attach for 43 hrs prior to treatment with a range of gemcitabine concentrations. The SRB assay was performed after 96 hrs drug treatment. Data shown as Growth (% Vehicle). IC_{50} was calculated by non-linear regression using GraphPad Prism 9. Data from a single experiment and values are mean of the four technical replicates ± SD.

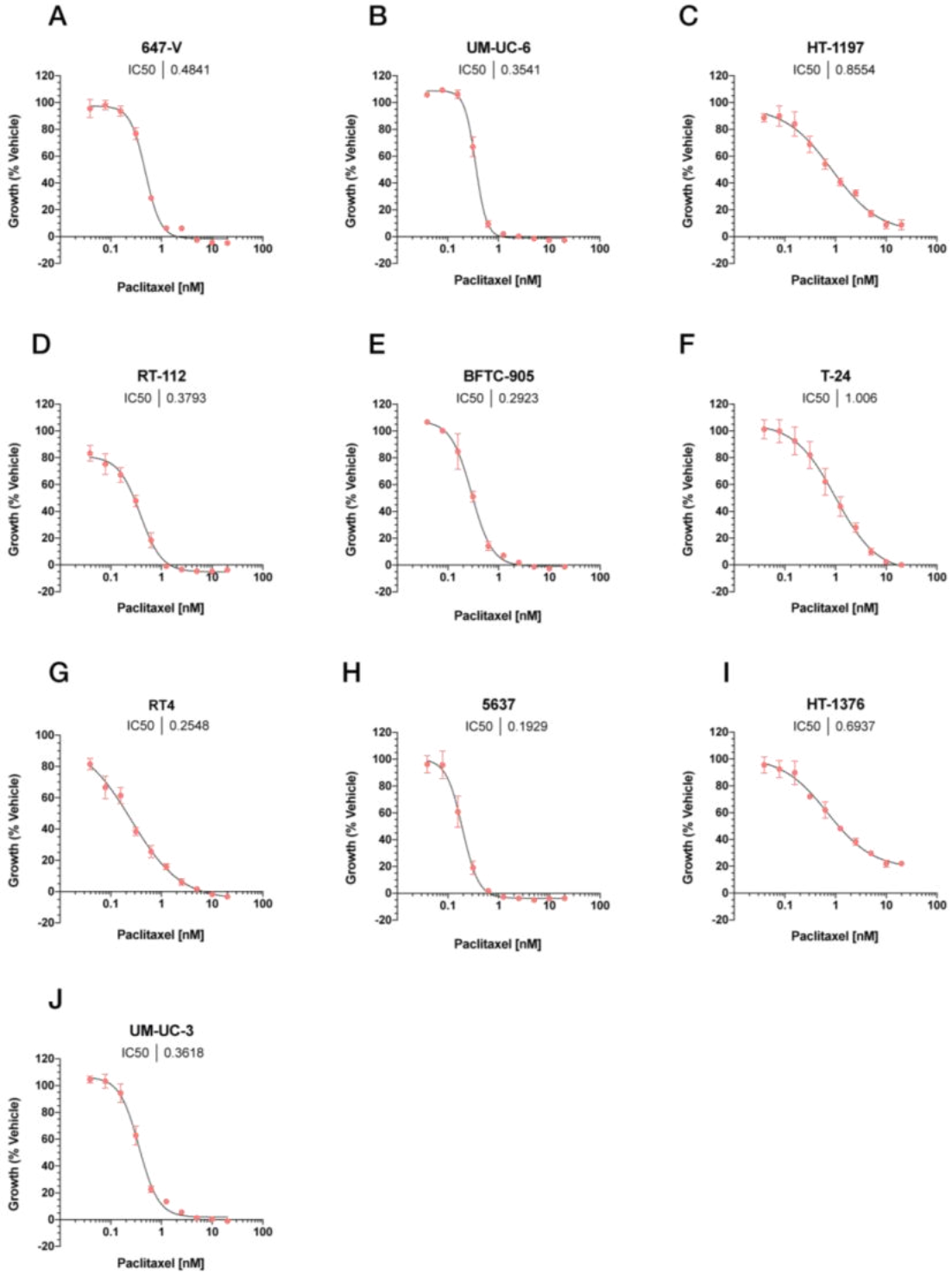


Figure 3.8 Concentration-response curves for paclitaxel IC_{50} determination of 10 UCC cell lines.

647-V (A), UM-UC-6 (B), HT-1197 (C), RT-112 (D), BFTC-905 (E), T-24 (F), RT4 (G), 5637 (H), HT-1376 (I) and UM-UC-3 (J) cells were seeded at their optimal seeding densities and left to attach for 43 hrs prior to treatment with a range of paclitaxel concentrations. The SRB assay was performed after 96 hrs drug treatment. Data shown as Growth (% Vehicle). IC_{50} was calculated by non-linear regression using GraphPad Prism 9. Data from a single experiment and values are mean of the four technical replicates ± SD.

Table 3.3 Summary of IC_{50} values for cisplatin, gemcitabine, and paclitaxel for 10 cell lines in the UCC panel.

Cell line	Cisplatin (μM)	Gemcitabine (nM)	Paclitaxel (nM)
647-V	2.9	5.7	0.5
UM-UC-6	1.6	5.5	0.4
HT-1197	1.8	20.9	0.9
RT-112	1.5	3.7	0.4
BFTC-905	0.6	1.9	0.3
T-24	1.2	~3.0	1.0
RT4	3.8	3.2	0.3
5637	0.4	3.9	0.2
HT-1376	4.0	20.1	0.7
UM-UC-3	1.3	7.0	0.4
Panel Mean ± SD	**1.9 ± 1.3**	**7.5 ± 7**	**0.5 ± 0.3**

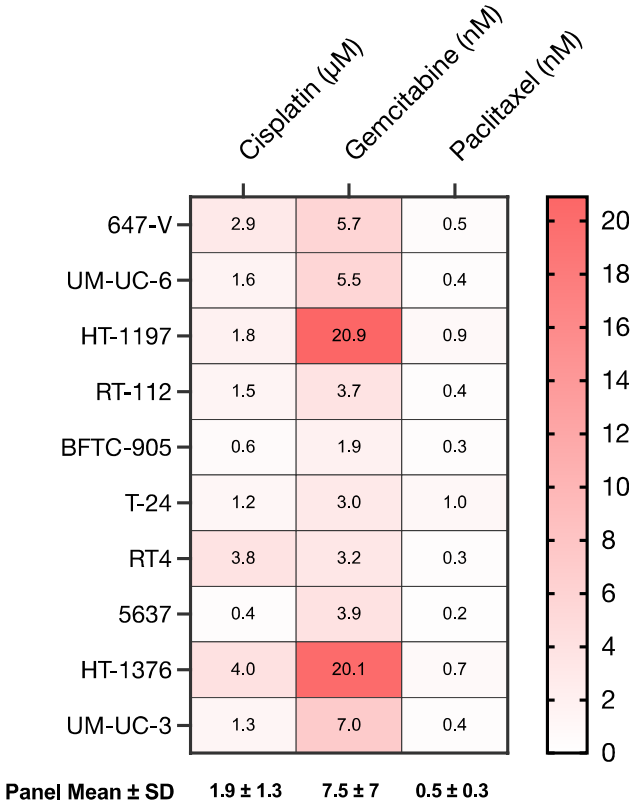


Figure 3.9 Heat map summary of IC_{50} values for cisplatin, gemcitabine, and paclitaxel for 10 cell lines in the UCC panel.

3.2.2 Genotoxic stress induces expression of *A3A* and *A3B* independent of p53 status

To investigate induction of A3A/B in response to therapeutics, five UCC cell lines were treated with commonly used chemotherapy drugs for 48 hours and mRNA expression using RT-qPCR analysis, or activity using the gel-based deamination assay, was examined. The cell lines were chosen such that the mini panel contains cell lines with different levels of baseline endogenous *A3A* and *A3B* expression, p53 status, molecular subtype, and disease stage (Table 3.4). This will allow investigation into whether any of these factors impact induction in response to chemotherapy treatment. Five different chemotherapy drugs and ionising radiation (IR), in the form of X-rays, were investigated to determine whether induction of A3A/B is related to the drug's mechanism of action. Gemcitabine and paclitaxel were used at concentrations approximately 10 x the highest IC_{50} of panel (Table 3.3). Cisplatin was used at 10 µM (approximately 5 x mean IC_{50} of panel) to ensure vehicle concentration remained at 0.5%. Bleomycin and the antimetabolite, 5-fluorouracil were used at concentrations used previously in the literature (Kanu et al., 2016; Middlebrooks et al., 2016). RT4 cells were treated with half the concentration of the other cell lines to ensure recovery of sufficient amounts of quality RNA for RT-qPCR. The dose of IR used was guided by previous work conducted in our lab in NSCLC lines (Dr Michael Walton, ICR).

Table 3.4 Key characteristics of UCC cell lines used for investigation of induction in response to chemotherapy drugs.

A3A, *A3B* expression and p53 status determined by work in this chapter.

Cell line	Baseline *A3B*	Baseline *A3A*	p53 status	Disease	Subtype
HT-1376	Very high	Yes	Mutant	MI	Basal
BFTC-905	Moderate	Yes	Mutant	MI	Basal
SW780	Moderate	Yes	Wild-type	NMI	Luminal
RT4	High	No	Wild-type	NMI	Luminal
HT-1197	Low	No	Wild-type	MI	Basal

Chemotherapeutics with differing mechanisms of action induce expression of *A3B* in four of the five cell lines tested (Figure 3.10). Significant induction of *A3B* is seen in response to gemcitabine, cisplatin, bleomycin, and 5-fluorouracil in both mutant p53 cell lines tested (HT-1376 and BFTC-905; Figure 3.10 A and B). Paclitaxel also significantly induces expression in BFTC-905 cells. Gemcitabine significantly induces expression of *A3B* in both NMI-UCC models harbouring wild-type p53 (SW780 and RT4; Figure 3.10 C and D). Cisplatin and bleomycin also induce expression in SW780 cells. In contrast to the mutant p53 cell lines, 5-fluorouracil significantly downregulates expression in SW780 and RT4 cells. These results demonstrate that induction is not related to the drugs mechanism of action, with induction seen in response to all the drugs tested in BFTC-905 cells and multiple drugs in the other cell lines. While the other four cell lines have elevated *A3B* expression, HT-1197 has low baseline levels and does not induce expression in response to any of the drugs tested (Figure 3.10 D). This may suggest that cells with elevated expression have lost regulatory networks that would usually supress expression facilitating induction in the presence of drug. Although induction is less consistent in the p53 wild-type, NMI models, robust induction occurs in response to gemcitabine treatment, and this demonstrates that p53 status alone does not determine the ability to induce in response to chemotherapy treatment. Induction is also not dependent on molecular subtype with induction seen in both luminal and basal cell lines.

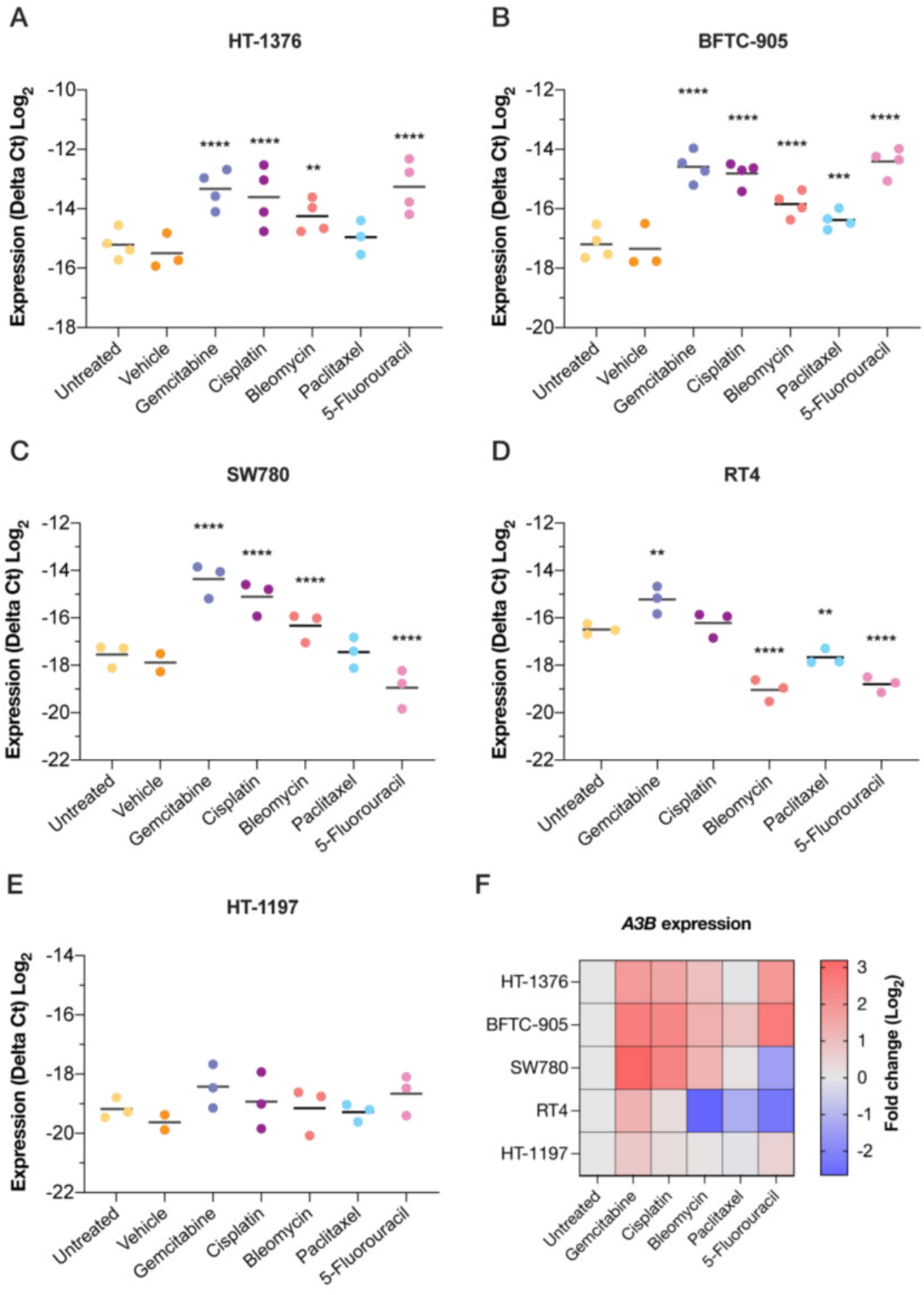


Figure 3.10 Induction of *A3B* in UCC cell lines in response to chemotherapeutics is not dependent on p53 status, disease type or molecular subtype.

Cells were seeded in 6-well plates and treated 24 hours later with the indicated drugs for 48 hours prior to RNA extraction; mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method. (A – E) *A3B* expression shown as ΔCt values for the five cell lines tested; dots represent independent experiments, line represents mean. A repeated-measures one-way

mixed-effects model (drug vs. untreated) was performed using GraphPad Prism 9 on ΔCt values. ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$. (F) Heatmap showing $\log_2$ fold changes in *A3B* expression relative to the untreated condition for each cell line. Vehicle, 0.5% DMSO; gemcitabine, 200 nM; cisplatin, 10 μM; bleomycin, 16.5 μM; paclitaxel, 10 nM; 5-fluorouracil, 10 μM (RT4 cells were treated with half the concentration).

Chemotherapy drugs also induce expression of *A3A* irrespective of p53 status, disease stage, molecular subtype, and baseline *A3A* expression level (Figure 3.11). Interestingly, the ability of a drug to induce expression of *A3B* does not predict its ability to induce expression of *A3A* suggesting they may be differentially regulated. HT-1376 cells only induce *A3A* in response to 5-fluorouracil but induction of *A3B* was seen with all other drugs except paclitaxel (Figure 3.11 A). As seen with *A3B, A3A* is induced in response to all drugs tested in BFTC-905 cells (Figure 3.11 B). SW780 cells induce expression of *A3A* in response to gemcitabine, cisplatin and bleomycin as seen with *A3B,* but paclitaxel that did not induce *A3B* expression does induce *A3A* (Figure 3.11 C). RT4 cells that do not have detectable baseline levels of *A3A* induce in response to gemcitabine, and both bleomycin and paclitaxel treatment, two drugs that downregulated expression of *A3B* (Figure 3.11 D). HT-1197 cells did not induce in response to any of the drugs tested (Figure 3.11 E). It is worth noting that while the fold changes in expression of *A3A* are large, the ΔCt values demonstrate that drug-induced levels of *A3A* are still lower than endogenous baseline levels of *A3B. A3A* expression was not reliably detected in untreated or DMSO treated SW780 cells despite being expressed in the initial characterisation and this shows how experimental set up can affect baseline expression levels. Of all the drugs tested, gemcitabine is the most robust inducer, significantly inducing expression of *A3A* in three, and *A3B* in four of the five cell lines.

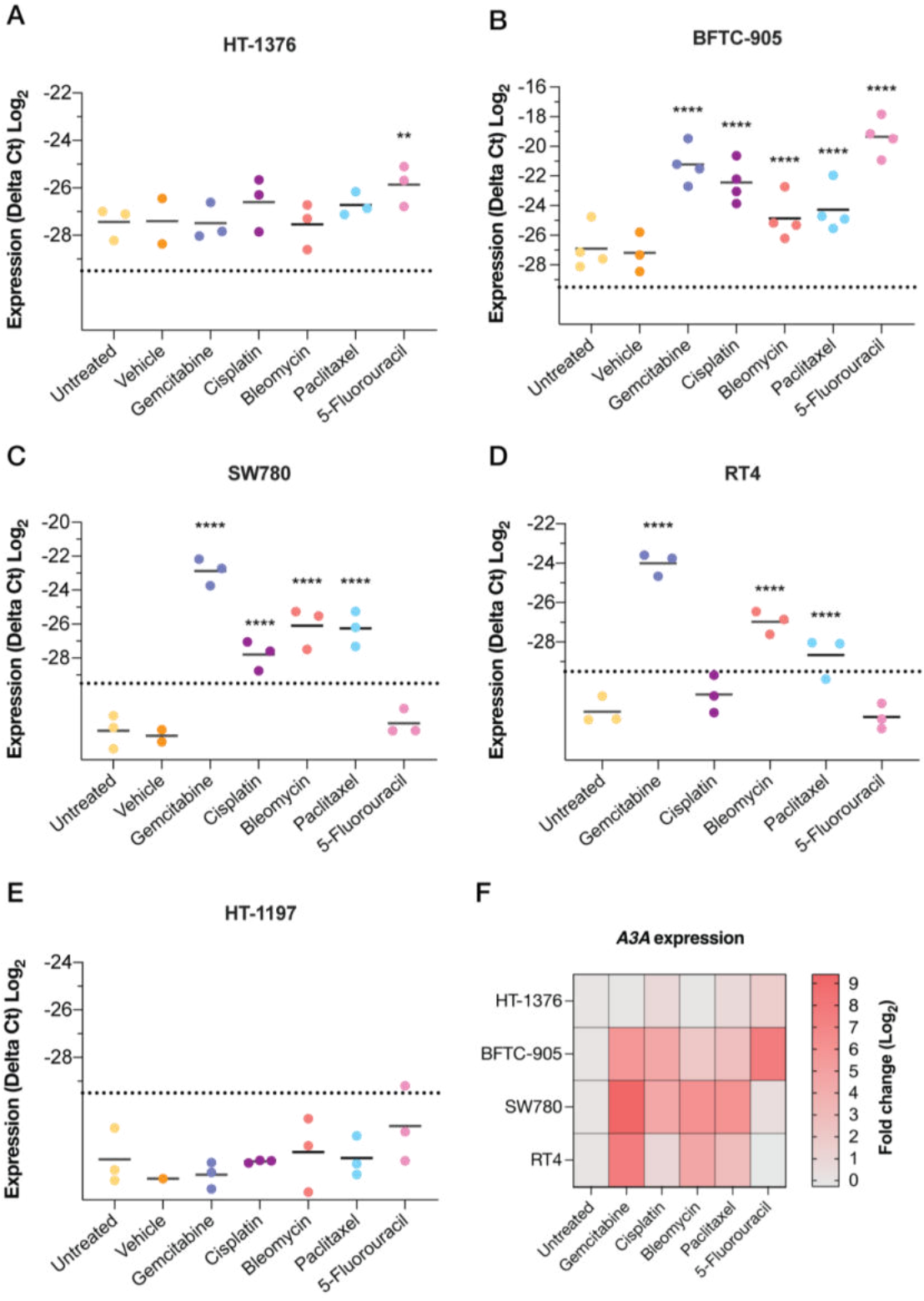


Figure 3.11 Induction of *A3A* in UCC cell lines in response to chemotherapeutics is not dependent on p53 status, disease type or molecular subtype.

Cells were seeded in 6-well plates and treated 24 hours later with the indicated drugs for 48 hours prior to RNA extraction; mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method. (A – E) *A3A* expression shown as ΔCt values for the five cell lines tested; dots represent independent experiments, line represents mean. A repeated-measures one-way

mixed-effects model (drug vs. untreated) was performed using GraphPad Prism 9 on ΔCt values. ** $P \leq 0.01$; **** $P \leq 0.0001$. (F) Heatmap showing $\log_2$ fold changes in *A3A* expression relative to the untreated condition for each cell line. Vehicle, 0.5% DMSO; gemcitabine, 200 nM; cisplatin, 10 μM; bleomycin, 16.5 μM; paclitaxel, 10 nM; 5-fluorouracil, 10 μM (RT4 cells were treated with half the concentration).

3.2.2.1 Ionising radiation is not a robust inducer of *A3A* or *A3B*

To determine whether induction of *A3A/B* occurs in response to non-drug forms of DNA damage and cellular stress, cells were treated with IR in the form of X-rays. While the induction of *A3A/B* in response to drug-induced stress occurred in several cell lines, IR only significantly induced expression of *A3A* and *A3B* in BFTC-905 cells and the level of induction was modest in comparison to that seen in response to chemotherapy drugs (Figure 3.12). IR did not induce expression of either *A3A* or *A3B* in HT-1376, HT-1197 or SW780 cells (data not shown) and therefore, IR is not a robust inducer of *A3* expression.

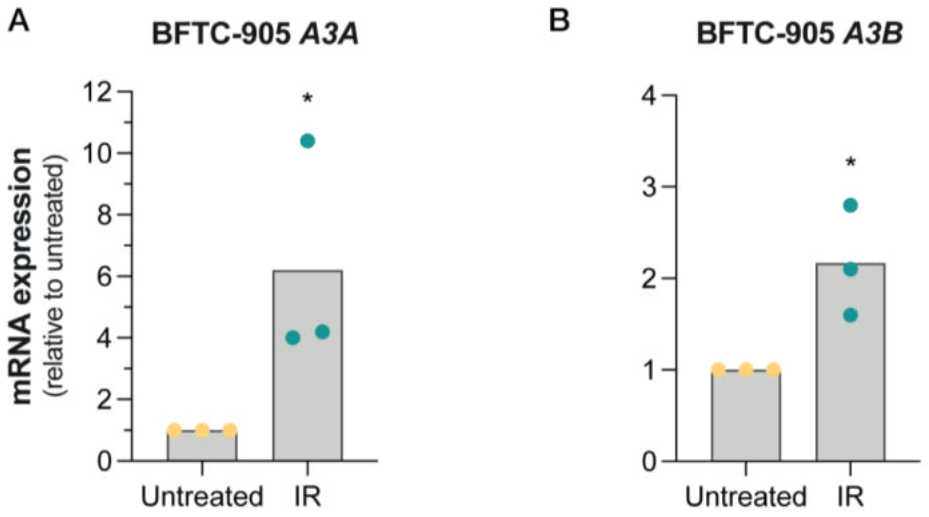


Figure 3.12 Ionising radiation only modestly induces *A3A/B* expression in BFTC-905 cells.

Cells were seeded in 6-well plates and treated with 10 Gy ionising radiation (IR; X-rays) 24 hours later. Cells were harvested, and RNA was extracted, after a further 48 hours. *A3A* (A) and *A3B* (B) mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method; dots represent independent experiments, bars represent mean. A paired t-test (IR vs. untreated) was performed using GraphPad Prism 9 on ΔCt values. * $P \leq 0.05$. Results for cell lines HT-1376, HT-1197 and SW780 are not shown as there were no measurable changes.

3.2.3 Drug-induced increases in deamination activity correlate with *A3B* expression

To determine whether the drug-induced increases in mRNA expression are accompanied with an increase in deamination activity, the *in vitro* assay was used to measure activity in lysates after drug treatment. As seen with mRNA expression, drugs with differing mechanisms of action increase deamination activity regardless of p53 status, disease stage or molecular subtype (Figure 3.13). In line with the observations made in section 3.2.1.2, where basal *A3B* expression correlates with A3 activity, A3 activity in response to drug treatment correlates well with *A3B* but not *A3A* expression (Figure 3.14). This further demonstrates that the A3 activity assay primarily measures A3B activity and not A3A. Interestingly, RT4 cells show an increase in activity but not mRNA expression upon cisplatin treatment (Figure 3.13 C); this discrepancy may be explained by the higher concentration of cisplatin used (10 µM vs 5 µM). HT-1197 cells do not induce expression of *A3A/B* in response to any of the drugs tested but gemcitabine and cisplatin treatment increases deamination activity suggesting that there may also be post-translational regulation of A3B occurring independent of mRNA changes (Figure 3.13 D).

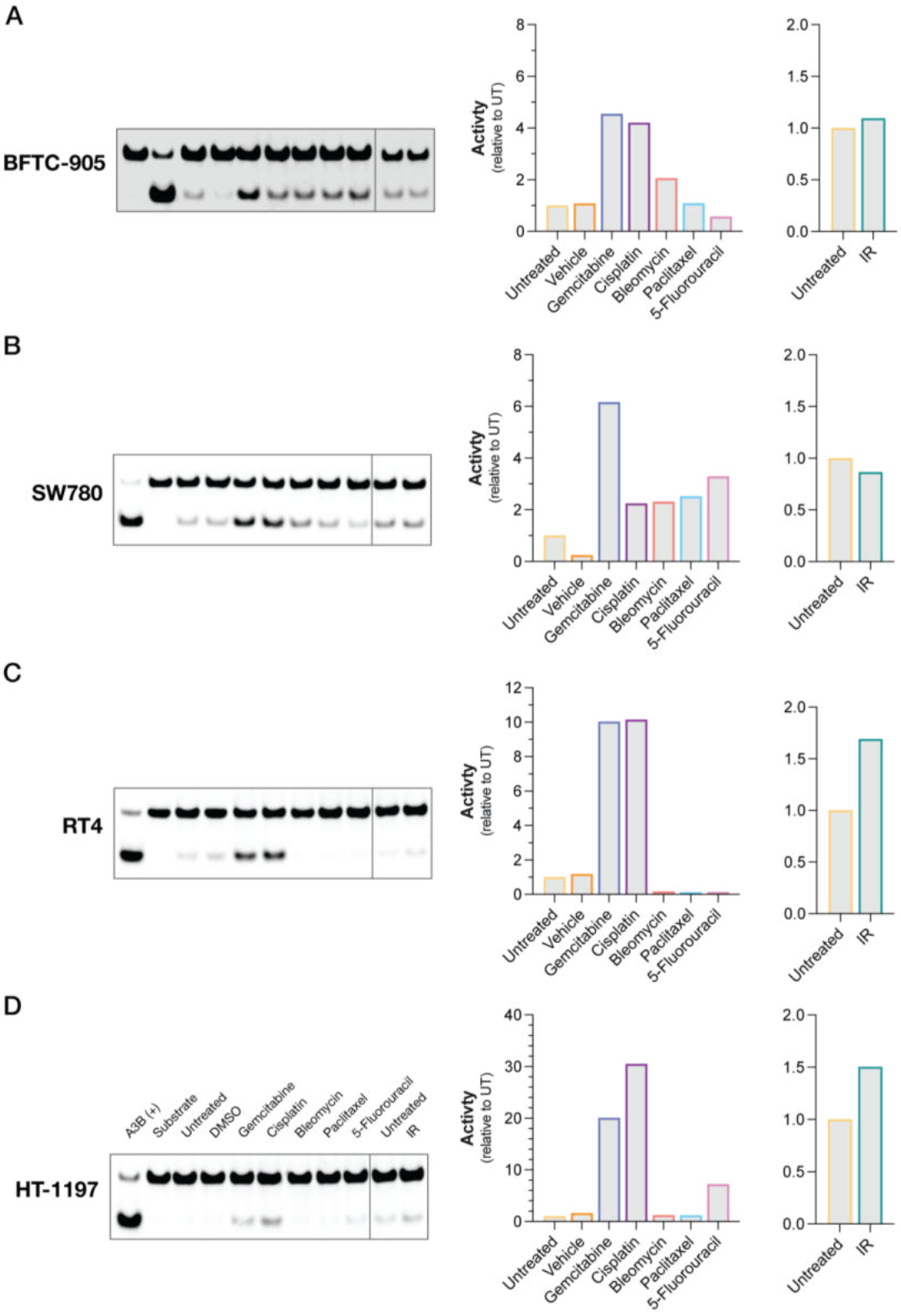


Figure 3.13 Commonly used anticancer agents cause an increase in A3 deamination activity.

BFTC-905 (A), SW780 (B), RT4 (C) and HT-1197 (D) cells were seeded in 6-well plates and treated 24 hours later with the indicated drugs for 48 hours prior to cell lysis and the deamination activity assay (25 µg protein). Graphs show quantification of activity (percentage of total signal in

each lane) relative to the untreated control. Gels and quantification from one experiment performed in parallel to one biological repeat from experiments described in 3.2.2. Vehicle, 0.5% DMSO; gemcitabine, 200 nM; cisplatin, 10 µM; bleomycin, 16.5 µM; paclitaxel, 10 nM; 5-fluorouracil, 10 µM.

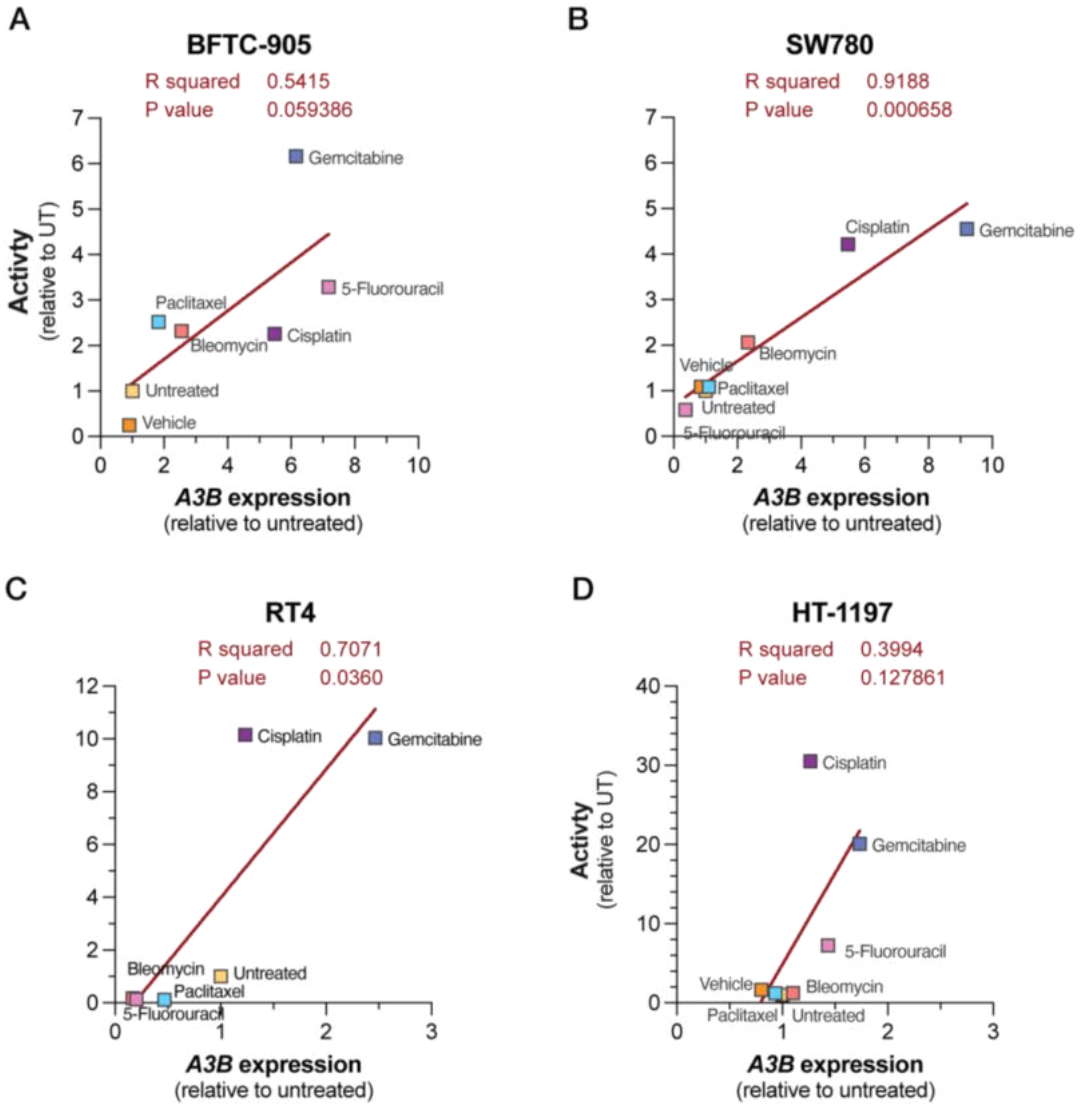


Figure 3.14 *A3B* mRNA induction correlates with A3 activity in response to chemotherapeutics.

A3 activity from Figure 3.13 for the cell lines BFTC-905 (A), SW780 (B), RT4 (C) and HT-1197 (D) was plotted against average *A3B* mRNA expression relative to the untreated control from Figure 3.10. Simple linear regression analysis was performed using GraphPad Prism 9 to determine the correlation co-efficient (R^2).

3.2.4 *A3A* and *A3B* induction is not a normal response in urothelium

To determine whether induction of *A3A* and *A3B* in response to chemotherapeutics is a normal response, NHU-TERT B cells were treated with the five drugs used previously (gemcitabine, cisplatin, paclitaxel, bleomycin, and 5-fluorouracil) at half the concentration used for the tumour cell lines. Higher concentrations could not be used due to insufficient recovery of quality RNA required for RT-qPCR; the RNA recovered after bleomycin treatment was of poor quality and could not be used. NHU-TERT B cells do not induce *A3B* or *A3A* expression in response to chemotherapeutics suggesting that induction is likely tumour-associated and normal cells have mechanisms to inhibit induction in conditions of stress (Figure 3.15).

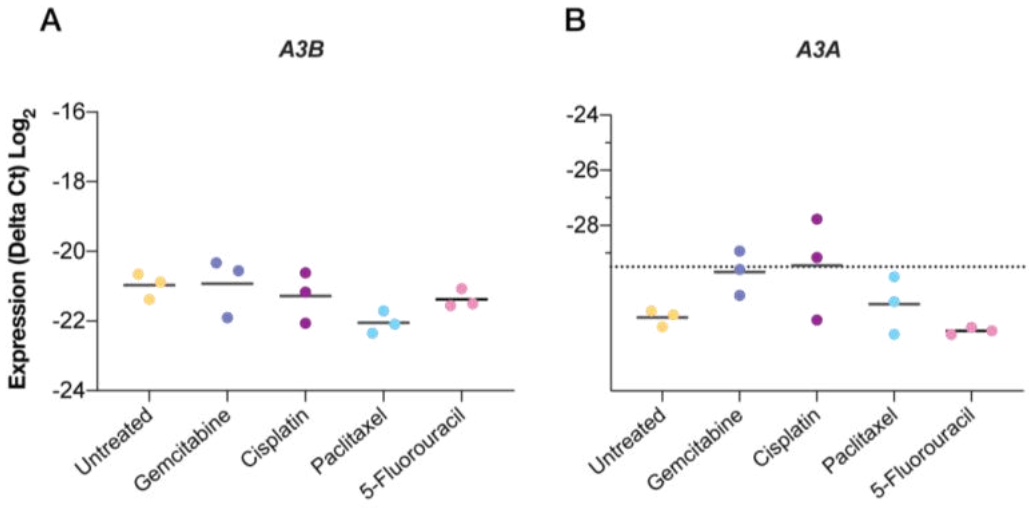


Figure 3.15 Chemotherapeutics do not induce *A3A* or *A3B* mRNA expression in a 'normal' immortalised urothelial cell line.

NHU-TERT B cells were seeded in 6-well plates and treated 24 hours later with the indicated drugs for 48 hours prior to RNA extraction. *A3B* (A) and *A3A* (B) mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method. Data shown as ΔCt values; dots represent independent experiments, line represents mean; dotted line denotes undetectable expression. A repeated-measures one-way ANOVA (drug vs. untreated) was performed using GraphPad Prism 9 on ΔCt values. Gemcitabine, 100nM; cisplatin, 5μM; paclitaxel, 5nM; 5-fluorouracil, 5μM.

3.2.5 *A3B* induction is concentration-dependent

Two cell lines, the wild-type p53, NMI SW780 and the mutant p53, MI BFTC-905 were taken forward for further investigation to determine whether the observed induction is concentration dependent, suggesting it is related to the level of DNA damage and stress generated. Gemcitabine and bleomycin were chosen as they both robustly induce expression in these lines and have distinct mechanisms of action. Gemcitabine primarily induces RS and triggers the ATR signalling cascade while bleomycin induces DSBs and ATM. Comparison of induction in response to the two drugs at equitoxic doses will determine whether the type of damage and subsequent DDR signalling impacts the observed level of induction. Prior to induction studies, the IC_{50} values for these drugs were determined by SRB (Figure 3.16). Values were initially obtained from curves where growth = % vehicle (Figure 3.16 A and C). An alternative method of calculating the IC_{50} values is to normalise the data to the T0 plate where growth = % T_0 (T_0 = 100%) (Figure 3.16 B and D). While the IC_{50} values obtained from the two methods are similar, the benefit of the % T_0 method is that it allows the number of population doublings occurring during the experiment to be determined and allows better characterisation of whether a drug effect is cytostatic or cytotoxic. Work presented in this in this section used values obtained from growth = % vehicle calculations, but all future work will use IC_{50} values obtained from curves using growth = % T_0.

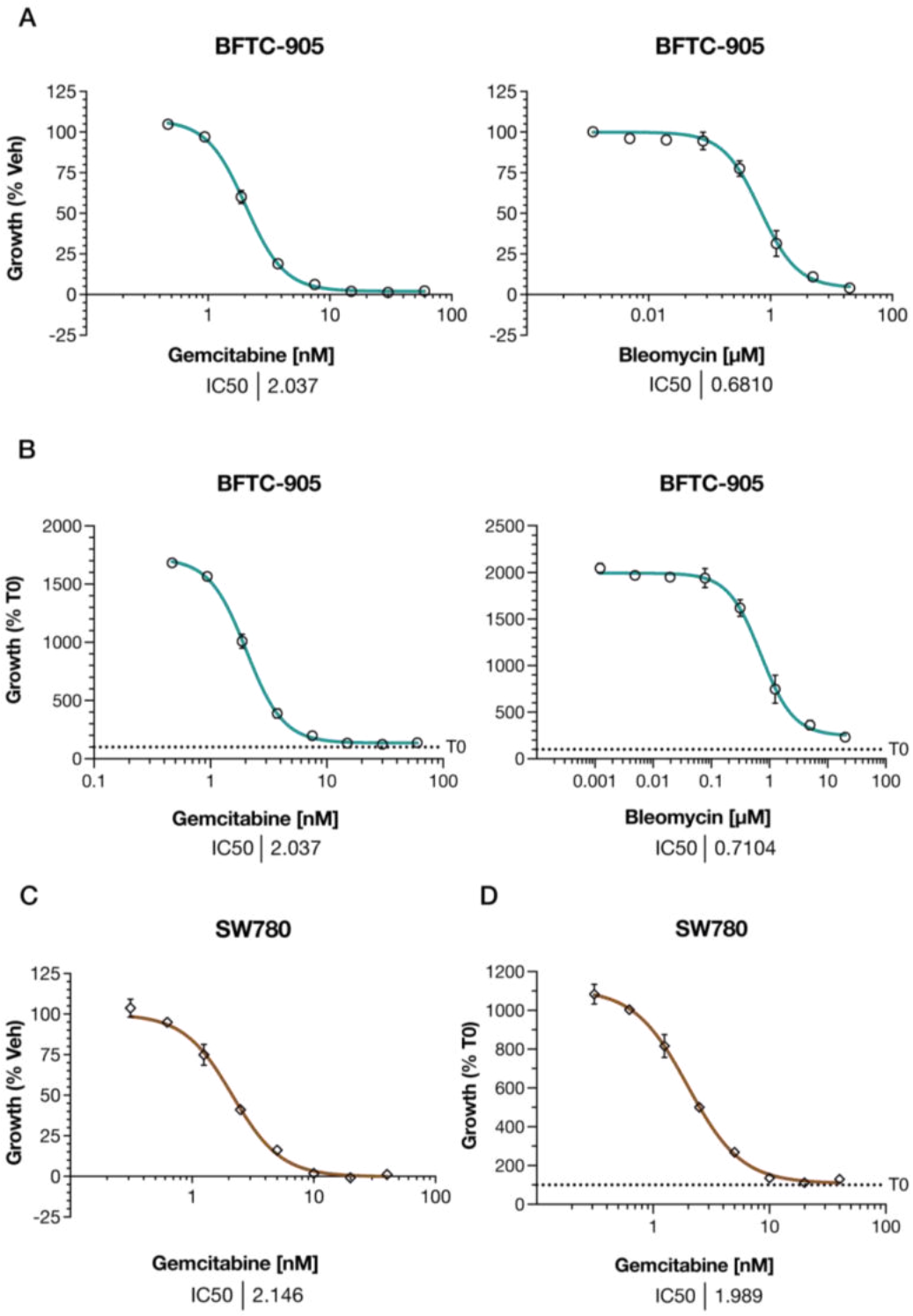


Figure 3.16 Concentration-response curves showing sensitivity of UCC cell lines, BFTC-905 and SW780, to gemcitabine and bleomycin.

Curves for gemcitabine and bleomycin in BFTC-905 cells calculated using the % vehicle method (A) and % T_0 method (B). Curves for gemcitabine in SW780 cells calculated using the % vehicle method (C) and % T_0 method (D). Response was analysed after 96 hours drug treatment using the SRB assay. Data is shown as mean ± SD for triplicate values in a single experiment. IC_{50} values were determined using GraphPad Prism 9.

BFTC-905 and SW780 cells were treated with a range of concentrations of gemcitabine and bleomycin corresponding to ⅒ x IC_{50}, IC_{20}, IC_{50}, IC_{80}, 10 x IC_{50}, 100 x IC_{50} and 200 x IC_{50} to investigate how *A3B* mRNA induction correlates with concentration (Table 3.5). Both cell lines show concentration-dependent increases in *A3B* expression in response to gemcitabine treatment (Figure 3.17 A and B). Repeated measures one-way ANOVA followed by Tukey's multiple comparisons tests revealed there were no significant differences in expression between 10 x, 100 x and 200 x IC_{50} (Table 3.6). These results show that maximal induction occurs at ~10 x IC_{50} with no/minimal increases in expression with concentration beyond this and that *A3B* induction is saturable. BFTC-905 cells also show concentration-dependent increases in response to bleomycin treatment (Figure 3.17 C) with no significant differences in expression between 10 x IC_{50} and 100 x IC_{50} (P = 0.8814). Accordingly, a concentration of 10 x IC_{50} was used for future work. Work presented earlier in this chapter suggested that gemcitabine is a more robust inducer of *A3B* than bleomycin. To test this hypothesis, a one-tailed t-test was conducted on *A3B* expression values obtained in response to treatment of BFTC-905 cells with equitoxic concentrations of both drugs (10 x IC_{50}). Gemcitabine (ΔCt -14.9 ± 0.32) is a more robust inducer of *A3B* than bleomycin (ΔCt -16.04 ± 0.25); t = 22.31(1), P = 0.0143). These findings corroborate those in the literature that suggest RS is a major driver of *A3B* expression (Kanu et al., 2016).

Table 3.5 Summary of concentrations used (nM). G, gemcitabine; B, bleomycin.

Cell line	**⅒ x IC_{50}**	**IC_{20}**	**IC_{50}**	**IC_{80}**	**10 x IC_{50}**	**100 x IC_{50}**	**200 x IC_{50}**
SW780 (G)	0.2	1.1	2.1	4.1	21	210	420
BFTC-905 (G)	0.2	1.2	2.0	3.4	20	200	400
BFTC-905 (B)	68	256	681	1800	6800	68000	-

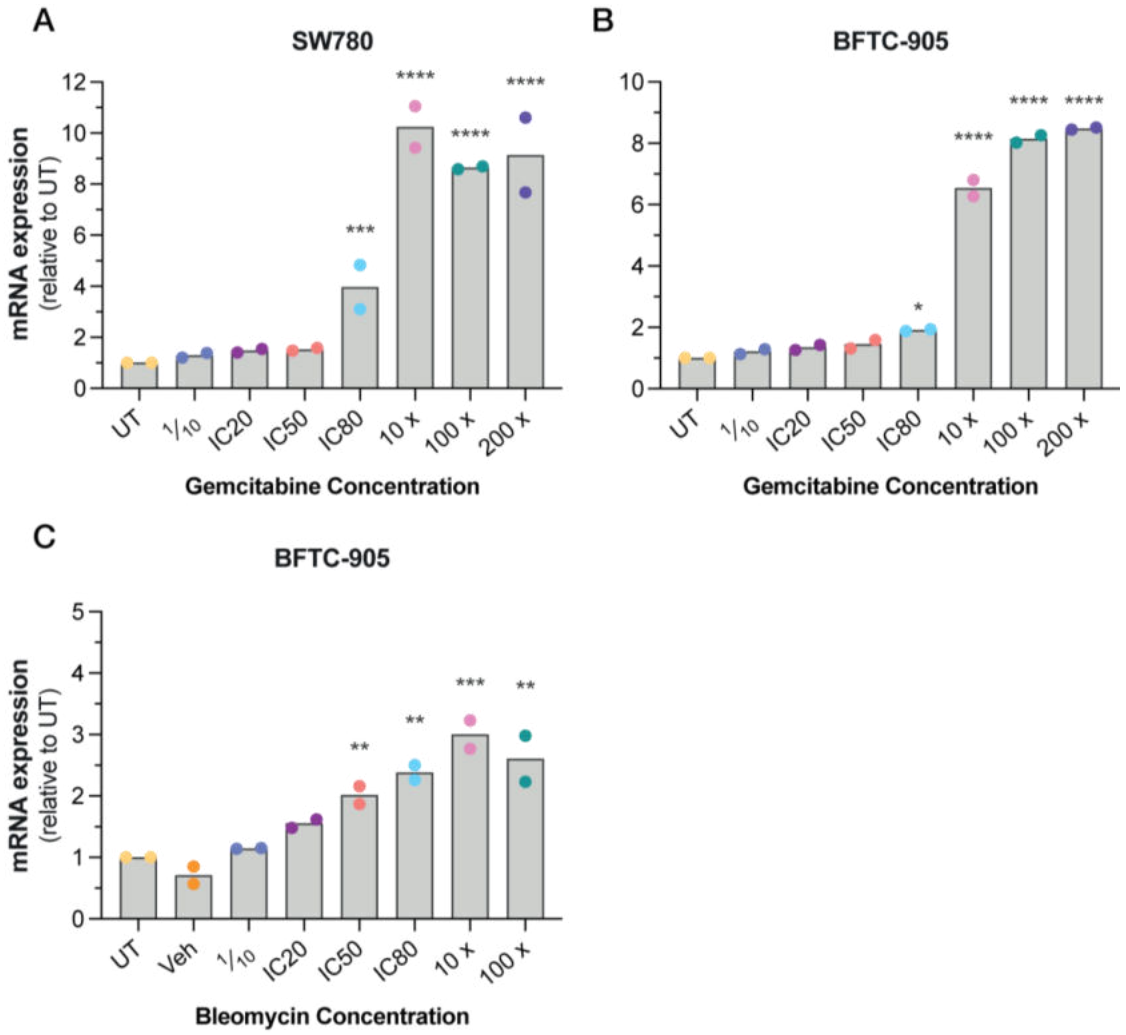


Figure 3.17 Concentration-dependent increases in *A3B* in response to gemcitabine and bleomycin.

SW780 (A) and BFTC-905 (B and C) cells were seeded in 6-well plates and treated 24 hours later with the indicated drugs for 48 hours prior to RNA extraction (see Table 3.5). mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method; dots represent independent experiments, bars represent mean. A repeated-measures one-way ANOVA (comparing all groups) followed by Tukey's multiple comparisons test was performed using GraphPad Prism 9 on ΔCt values; significance values shown are vs. UT. UT, untreated; Veh, 0.68% DMSO. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.

Table 3.6 *P* values obtained from Tukey's multiple comparisons following a repeated-measures one-way ANOVA.

SW780, F (8, 8) = 109.9 *P* <0.0001. BFTC-905, F (8, 8) = 128.1 *P* <0.0001.

Comparison	SW780 (*P* value)	BFTC-905 (*P* value)
10 x vs. 100 x	0.8857	0.6670
10 x vs. 200 x	0.9829	0.4963
100 x vs. 200 x	>0.9999	>0.9999

3.3 Discussion

The first aim of this chapter was to fully characterise expression and activity of two cytidine deaminases, A3A and A3B, implicated in cancer mutagenesis in a panel of UCC cell lines representing the different subtypes found in patients. This showed that *A3B* expression is elevated relative to the NHU-TERT B cell line model of normal urothelium in nine of the 12 cell lines suggesting, as is reported in the The Cancer Genome Atlas project (TCGA) dataset and additional studies, that *A3B* upregulation and APOBEC-mediated mutagenic signatures are common in this tumour type (Alexandrov et al., 2020; Alexandrov et al., 2013; Burns, Temiz, et al., 2013; Network, 2014). Elevated expression is seen in both basal and luminal subtype models of UCC demonstrating *A3B* upregulation is not subtype specific, again supporting previously reported findings in the literature (Glaser et al., 2018). *A3B* is also elevated in the two cell line models for early, NMI disease (RT4 and SW780) showing *A3B* upregulation is not limited to late-stage disease. This is in line with patient data in the literature; unlike other tumour types where APOBEC mutagenesis is more strongly associated with late, subclonal diversification (de Bruin et al., 2014), it occurs early in UCC development and may play a role in tumour initiation and development (McGranahan et al., 2015; Middlebrooks et al., 2016; Nordentoft et al., 2014; Robertson et al., 2017; Shi et al., 2020). *A3B* elevation and A3-mutagenesis have been associated with cell lines/tumours with defective p53 signalling in many cancer types, and functional p53 has been directly implicated in repressing *A3B* expression (Menendez et al., 2017; Nikkilä et al., 2017; Periyasamy et al., 2017). Two wild-type p53 cell lines, RT4 and SW780, have elevated *A3B* expression demonstrating that mechanisms exist in the context of UCC that facilitate *A3B* upregulation in the presence of a wild-type p53 response. In contrast to *A3B, A3A* expression is only detected in four cell lines and at levels much lower than that of *A3B,* consistent with data from patient samples that shows *A3A* is less commonly upregulated. While this may suggest that A3B is the predominant mutagen, the relationship between *A3A* expression and the A3-mutational signature identified in tumours is complex as previously discussed. Data presented in this chapter also shows that there is increased A3 activity in most of the UCC cell line panel as determined by an *in vitro* deamination assay containing a TCA motif, which both A3A and A3B are proposed to act on. *A3B* expression

correlates well with the activity seen in the deamination assay suggesting that *A3B* is dominant in these cell lines. Indeed, conditions that induce *A3A*, but not *A3B*, do not increase deamination activity measured with the assay used in this work. Recently, however, several studies show that measuring A3A and A3B activity is more complex than simply using a linear substrate containing the preferred motif. A3A is over 100-fold more active than A3B in the presence of cellular RNA (Cortez et al., 2019), and A3B is more active on linear substrates while A3A is more active on hairpin substrates (Langenbucher et al., 2021). In addition, knockdown of *A3B* abolished most of the deamination activity detected when using a polyA-TC linear substrate (Jalili et al., 2020). As the deamination assay used for the work presented in this chapter contains a linear substrate in the presence of RNase A, it is likely A3B activity is primarily being measured. However, very little is known about the functional regulation of these enzymes; it is possible that the sonication and RNase treatment required to detect activity *in vitro* disrupts endogenous negative regulation occurring in cells, possibly interactions with other proteins and nucleic acids, and artificially activates A3B. Accordingly, results from the *in vitro* deamination assay should be interpreted with caution as they may not represent the physiological state of A3B in cells. In conclusion, from the data presented in this chapter, neither A3A nor A3B can be definitively determined as the predominant mutagen in UCC cell lines.

The next aim of this chapter was to investigate induction of *A3A* and *A3B* in response to cytotoxic drugs commonly used in cancer treatment as this has been previously observed in bladder (Middlebrooks et al., 2016), breast (Kanu et al., 2016) and lung (Dr Michael Walton, ICR) cancer cell lines. Five UCC cell lines with varying levels of baseline endogenous *A3A/B* expression representing basal and luminal subtypes of UCC, MI and NMI disease, with both wild-type and mutant p53 responses were treated with five commonly used chemotherapeutics (gemcitabine, cisplatin, bleomycin, paclitaxel, and 5-fluorouracil) as well as IR. While induction of *A3A/A3B* was dependent on the drug and cell line tested, overall, drugs with diverse mechanisms stimulate expression in four of the five cell lines tested. This finding could have profound clinical implications as anticancer agents themselves may be stimulating tumour heterogeneity and adaption via drug-induced A3 activity and mutagenesis. In contrast to previously published work (Menendez et al., 2017) and work within our group in NSCLC

lines (Dr Michael Walton, ICR) where *A3B* induction was dependent on p53 status, significant increases in both *A3A* and *A3B* expression is seen in two wild-type p53 cell lines (SW780 and RT4) demonstrating that induction is independent of p53 status. This is consistent with recent findings in breast cancer, where induction in response to chemotherapeutics is seen in wild-type p53 lines, and demonstrates that functional p53 is not sufficient to prevent induction in some tumour contexts (Periyasamy et al., 2021). Interestingly, induction of both *A3A* and *A3B* did not always occur in response to the same drug suggesting that the two family members are differentially regulated in response to drug-induced stress, and this warrants further investigation into the underlying mechanisms driving their expression.

Drug-mediated stimulation of *A3A* and *A3B* expression is not seen in the NHU-TERT B normal urothelium cell line model suggesting induction in response to drug treatment is tumour-specific. This contrasts with the findings that MCF10A cells, commonly used as a model for normal breast epithelium (Kanu et al., 2016; Periyasamy et al., 2021), and human mammary epithelial cells (HMEC) do induce *A3B* expression in response to a number of chemotherapeutic agents including those used in this study (Periyasamy et al., 2021). While the concentrations used by Periyasamy et al., (2021) were higher (cisplatin 20 µM and 5-fluorouracil 50 µM vs 5 µM) and mRNA expression was measured at 24 hours in contrast to 48 hours, the concentrations used in this work are high enough to induce expression in the UCC tumour lines and therefore, the differences reported are likely due to fundamental tissue differences between urothelium and breast epithelium. This, taken with the findings that *A3* induction is p53-independent suggests that there are additional, currently unknown factors acting in UCC tumour cells that facilitate upregulation of *A3A* and *B* in response to drug-induced stress.

Two cell lines were taken forward for further characterisation of drug-induced expression: SW780 representing early stage, NMI disease with wild-type p53 and BFTC-905 representing later stage, MI disease with a mutant p53 response. *A3B* induction was investigated in the context of two drugs, the RS-inducing drug, gemcitabine, and the DSB-inducing drug, bleomycin, to elucidate the potential clinical implications of drug-induced expression. While maximal *A3B* induction occurs at approximately 10 x IC_{50} in response to gemcitabine (both lines) and bleomycin (BFTC-905), small inductions (~2-fold) are also seen with lower, IC_{80}

and IC_{50}, concentrations. The increases seen at a wide range of concentrations demonstrates that some level of *A3B* induction is likely occurring in patient tumours treated with clinical doses. As *A3A/B* is suggested to be a driver of heterogeneity and tumour adaptation (Alexandrov et al., 2020), have emerging roles in transcription (Periyasamy et al., 2015) and RNA editing (Alqassim et al., 2021; Asaoka et al., 2019; Sharma et al., 2017; Sharma et al., 2015), it is likely that clinical drug treatment itself could be stimulating the development of drug resistance and disease progression via a range of different mechanisms. A direct comparison of the induction seen in response to the two drugs showed that gemcitabine is the more robust inducer and that while cytotoxic drugs with a range of mechanisms of action can induce expression, the level of induction varies. These findings support those in the literature that suggest that RS is a major driver of *A3B* expression in cancer cells (Kanu et al., 2016; Venkatesan et al., 2021).

In summary, findings presented in this chapter demonstrate that elevated *A3A* and *A3B* expression and induction in response to chemotherapy drugs is independent of tumour stage, molecular subtype and p53 status. This is consistent with data from patient tumour samples showing that A3 dysregulation is common in this tumour type. A specific drug does not always induce expression of both family members, and this suggests they may be differentially regulated. Gemcitabine, a drug that induces RS, was found to be the most robust inducer of expression suggesting RS is a major driver of *A3A* and/or *A3B* expression in tumours consistent with the findings of others. Therefore, gemcitabine was chosen for further work to interrogate the mechanisms of induction and whether the two family members are differentially regulated in UCC.

Chapter 4

A3A and *A3B* are differentially regulated in UCC

Chapter 4 *A3A* and *A3B* are differentially regulated in UCC

4.1 Introduction

The A3 genomic region is not frequently associated with chromosomal rearrangements, copy number variation (CNV) or promoter mutations, and therefore the elevated levels of *A3B* observed in tumours are likely a result of upstream signal transduction events that drive expression (Burns, Lackey, et al., 2013). Inflammation is a key enabling characteristic that facilitates the hallmarks of cancer (Hanahan, 2022), with NF-κB acting as the central mediator of the inflammatory process (DiDonato et al., 2012). *A3B* is transcriptionally regulated by a protein kinase C (PKC)/NF-κB signalling axis and it has been proposed that chronic infection and inflammation could be driving elevated *A3B* expression in tumours (Leonard et al., 2015; Maruyama et al., 2016). The diacylglycerol (DAG) mimic phorbol-myristic acid (PMA) induces expression of *A3B* in multiple cell lines originating from different cancer types, and this induction can be attenuated with PKC inhibition (Leonard et al., 2015). The authors investigated known PKC-regulated transcription factors and through *in silico* analysis identified several NF-κB binding sites located upstream of the *A3B* transcriptional start site. PMA-induced expression of *A3B* is also attenuated through small molecule inhibition of NF-κB, and chromatin immunoprecipitation (ChIP) experiments demonstrated that the alternative heterodimer RelB/p52 is actively recruited to the *A3B* promoter. Maruyama et al., (2016) also showed that PMA induces *A3B* expression and induction is attenuated by PKC and NF-κB inhibition. However, in contrast to Leonard et al., (2015), experiments using reporter constructs and electrophoretic mobility shift assays (EMSA) showed that it was the canonical heterodimers, p65/p50 and p65/c-Rel, that were recruited to the *A3B* promoter, and that knockdown of *p65* prevents induction in response to PMA. To determine whether PKC and downstream NF-κB signalling is a major driver of the elevated *A3B* expression observed in tumours, Leonard et al., (2015) treated multiple cancer cell lines with the PKC inhibitor sotrastaurin. PKC inhibition only reduced endogenous expression of *A3B* by > 50% in half of the cell lines tested and this demonstrates that other signalling pathways act alongside PKC/NF-κB to drive elevated expression of *A3B.* Therefore, while PKC and NF-κB signalling is clearly

implicated in *A3B* regulation, there is conflicting evidence as to whether the canonical or alternative branch of NF-κB is required, and it is not the sole pathway responsible for the upregulation seen in tumour cells.

Constitutive activation of signal transduction cascades is a well characterised contributor to oncogenesis. Many of the genes commonly mutated or amplified in cancer encode components of the PI3K or mitogen-activated protein kinase (MAPK) pathways, and this results in constitutive, ligand-independent activation (Figure 4.1). PKC acts downstream of receptor tyrosine kinases (RTKs) and several RTK-related signalling pathways are associated with elevated *A3B* expression suggesting oncogenic signalling is a major driver of A3 mutagenesis in tumours. *ERBB2* amplification and *PTEN* mutations are associated with high numbers of A3-mediated mutations (Kanu et al., 2016; Roberts et al., 2013), while HER2+ and *EGFR*-amplified cell lines commonly have elevated *A3B* expression (Chou et al., 2017; Kanu et al., 2016). siRNA knockdown of *ERBB2* reduces basal A3B mRNA, protein, and deamination activity in HER2+ breast cancer cell lines and treatment with two RTK inhibitors (RTKi) afatinib and lapatinib, demonstrated that this is due to signalling inhibition (Kanu et al., 2016). Basal *A3B* expression is also reduced in ER+ breast cancer cells by afatinib and two other RTKis, osimertinib and PD153035 (Chou et al., 2017). The transcription factor, Myb-related protein B (encoded by *MYBL2;* known as B-Myb) binds to the *A3B* promoter and knockdown reduces baseline endogenous expression of *A3B.* RTKi treatment reduces expression of *MYBL2* suggesting that RTK signalling mediates *A3B* expression via the transcription factor B-Myb. PKC inhibition with sotrastaurin did not alter expression of *MYBL2* and this confirmed that multiple pathways are responsible for the elevated *A3B* expression observed in tumours.

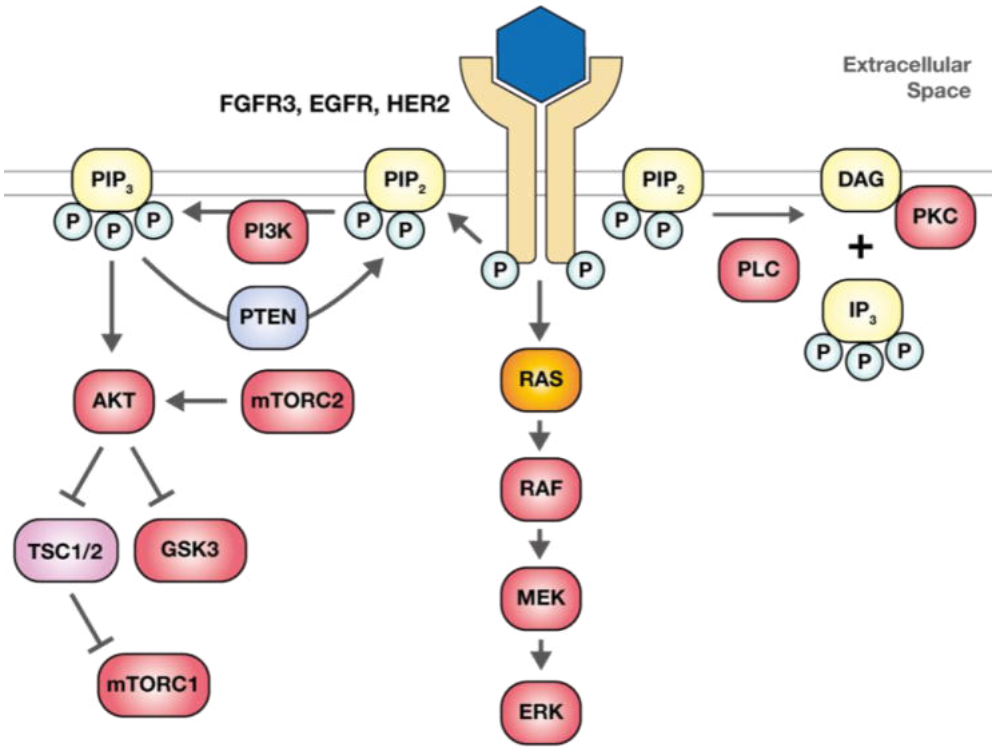


Figure 4.1 Receptor tyrosine kinase signalling.

Both the downstream PI3K and MAPK signalling cascades are involved in A3 regulation. Small molecule inhibition of PI3K, MEK, and mTOR reduces basal A3 activity, although it is worth noting that the substrate used in this study can be deaminated by A3G as well as the TC-specific deaminases (Kanu et al., 2016). However, siRNA knockdown of *PTEN*, a negative regulator of the PI3K signalling cascade, increases A3B protein expression and deamination activity, demonstrating that PI3K signalling can specifically regulate A3B. Interestingly, the common activating helical domain mutations in *PIK3CA* (E542K and E545K) occur within the consensus TC-specific A3 motif, suggesting that A3 activity can generate mutations that cause dysregulated PI3K signalling (Henderson et al., 2014), which could, in turn, further increase A3 activity in tumour cells. In addition to PKC, the other downstream signalling cascades, MAPK (Bassères et al., 2010; Finco et al., 1997; Schulze-Osthoff et al., 1997) and PI3K/AKT/mTOR (Dan et al., 2008; Hutti et al., 2012; Kane et al., 1999; Madrid et al., 2001), also feed into and activate NF-κB. This suggests that RTK signalling may drive *A3B* expression via multiple downstream pathways, with expression activated by various transcription factors including NF-κB and B-Myb.

A3 enzymes are proposed to act at sites of DNA damage, and findings presented in Chapter 3 and by others, demonstrated that chemotherapy drugs induce expression of *A3A* and *A3B*. While the level of induction varied between drugs with different mechanisms of action, all the drugs tested induce some form of DNA damage and this suggests that DNA damage is the initiating event required for *A3A* and *A3B* expression. Of the five drugs tested, gemcitabine, a drug that primarily causes RS, appears to be the most robust inducer and this is consistent with findings in breast cancer that proposed that RS is a major driver of A3B mutagenesis (Kanu et al., 2016). Exogenous supply of nucleosides, which is known to relieve RS, attenuated hydroxyurea-mediated induction of A3. In addition, inhibition of ATR and CHK1 signalling with both small molecule inhibitors and siRNA knockdown attenuates hydroxyurea-mediated induction of A3 activity and/or mRNA expression in breast cancer cell lines and the non-tumorigenic line, MCF10A. Inhibition of ATM also partially attenuates induction in breast cell lines, although the effect is modest, and the authors concluded that the ATR/CHK1 pathway is the primary mechanism of upregulation in response to hydroxyurea-mediated RS. ATR and DNA-PK inhibition also attenuates hydroxyurea-mediated induction of *A3B* in myeloma cell lines with a GFP reporter inserted at the end of the endogenous *A3B* coding region (Yamazaki et al., 2020). In addition to reducing baseline expression, inhibition of RTK signalling with lapatinib attenuates hydroxyurea-induction of *A3B* (Kanu et al., 2016). Hydroxyurea stimulated expression was also attenuated with inhibition of downstream MEK and mTOR signalling. Dysregulated RTK signalling is a known driver of RS and therefore, it is possible that these drugs reduce hydroxyurea-mediated induction either via a reduction in RS or by inhibiting activation of downstream transcription factors implicated in A3 regulation such as B-Myb and NF-κB.

Activation of NF-κB occurs in response to DNA damage and the DDR proteins ATM and ATR play key roles in this process. Cells from ataxia-telangiectasia (AT) patients with inactivating mutations in *ATM*, do not activate NF-κB in response to IR and camptothecin (CPT), and NF-κB activation can be restored through exogenous expression of wild-type *ATM* (Lee et al., 1998; Li et al., 2001; Piret et al., 1999). Upon detection of a DSB, ATM, alongside PARP-1 (Stilmann et al., 2009) signals via a series of NEMO post-translational modifications to activate

the canonical IKK complex (Huang et al., 2003; Wu et al., 2006) via interactions with TNF receptor-associated factor 6 (TRAF6) (Hinz et al., 2010). RS that is classically sensed by ATR can also activate NF-kB via an ATM-dependent pathway (Wu & Miyamoto, 2008), with ATR itself acting as a negative regulator of NF-kB activation (Crawley et al., 2015; Wu & Miyamoto, 2008). The adaptor protein, STING, that is classically involved in activation of the innate immune response in response to cytosolic DNA, also has a non-canonical, cGAS-independent function (Dunphy et al., 2018). STING has recently been shown to act alongside ATM, PARP-1, p53 and TRAF6 to activate NF-kB in response to etoposide-induced DNA damage. This highlights the link between DNA damage, the DDR, RTK signalling and the innate immune response.

4.1.1 Chapter aims

Findings presented in Chapter 3 showed that both *A3A* and *A3B* are upregulated in response to chemotherapy drugs. While several studies had begun investigating the mechanism of drug-mediated induction of *A3B*, little was known about induction of *A3A*. In addition, most of the work was conducted in breast cancer cell lines, and it is not known whether these signalling pathways also regulate *A3B* in the context of UCC. Therefore, this chapter aimed to characterise the mechanism of induction of both *A3A* and *A3B* in response to chemotherapy drugs in UCC cell lines, assess whether clinically available targeted therapies can curb expression, and if pathways previously shown to be involved in *A3B* regulation are also required for induction of *A3A*. At the start of this project, a PKC/NF-κB signalling axis had recently been identified as a key pathway contributing to the elevated *A3B* expression identified in many human cancers (Leonard et al., 2015; Maruyama et al., 2016). As DNA damage activates NF-κB signalling, and pathways downstream of RTKs feed into NF-κB, it was hypothesised that drug-mediated NF-κB activation was contributing to the observed induction. As such, work in this chapter aimed to specifically determine whether NF-κB is required for drug-mediated induction, and if so, which arm of the NF-κB pathway is responsible.

4.2 Results

4.2.1 Chemotherapy drugs induce expression of the canonical NF-κB target gene, *TNFα*

Anticancer agents activate the *A3B* transcriptional regulator, NF-κB (Leonard et al., 2015; Maruyama et al., 2016; Nakanishi & Toi, 2005) and therefore, it was hypothesised that induction of *A3A/B* occurs because of activated NF-κB signalling. Initially, to investigate if NF-κB pathway activation occurs in response to gemcitabine and bleomycin treatment in UCC, mRNA expression of the canonical target gene, tumour necrosis factor alpha (*TNFα),* was measured. Gemcitabine treatment induces *TNFα* expression in BFTC-905 and SW780 cells in a concentration-dependent manner (Figure 4.2 A and B). There were no significant differences in expression when cells were treated with 10 x IC_{50} and 100 x IC_{50} demonstrating maximal induction occurs at ~10 x IC_{50} (SW780 P = 0.0686; BFTC-905 P = 0.9501), as observed for *A3B* induction (Chapter 3). Bleomycin also induces *TNFα* expression in BFTC-905 cells, again with no significant difference in expression between 10 x IC_{50} and 100 x IC_{50} (P = 0.5640).

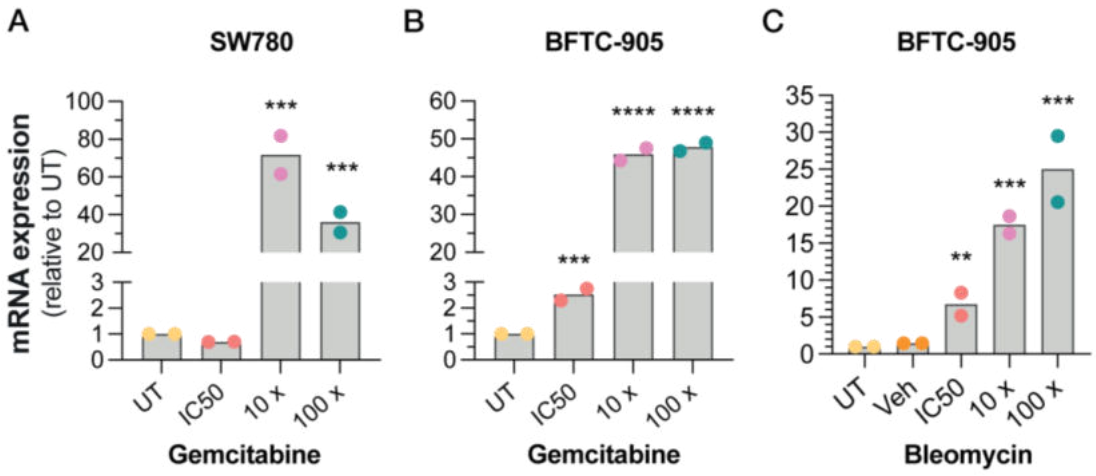


Figure 4.2 Gemcitabine and bleomycin treatment induces expression of the NF-κB target gene, *TNFα.*

SW780 (A) and BFTC-905 (B and C) cells were seeded in 6-well plates and treated 24 hours later with the indicated drugs for 48 hours prior to RNA extraction (see Table 3.5 for concentrations). mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method; dots represent independent experiments, bars represent mean. A repeated-measures one-way ANOVA (comparing all groups) followed by Tukey's multiple comparisons test was performed

using GraphPad Prism 9 on ΔCt values; significance values shown are vs. UT. UT, untreated; Veh, 0.68% DMSO. ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.

To examine whether gemcitabine is also a more robust inducer of *TNFα* expression as demonstrated for *A3B*, a one-tailed t-test was conducted on *TNFα* expression values obtained in response to treatment of BFTC-905 cells with equitoxic doses of both drugs (10 x IC_{50}). Gemcitabine (ΔCt -19.61 ± 0.64) is a more robust inducer of *TNFα* than bleomycin (ΔCt -21.01 ± 0.43); $t = 9.40(1)$, $P = 0.0337$). This provides further evidence that NF-κB pathway activation may be responsible for the observed induction, as the more robust inducer of an NF-κB target gene is also the more robust inducer of *A3B.*

4.2.2 Inhibition of PKC attenuates gemcitabine-mediated induction of *A3A, A3B* and *TNFα*

A3B upregulation in a number of tumour types is mediated by PKC and either downstream canonical (Maruyama et al., 2016) or alternative (Leonard et al., 2015) NF-κB signalling. Inhibition of the PKC/NF-κB signalling axis with the PKC inhibitor, sotrastaurin, modestly reduces basal expression of *A3B* in four UCC cell lines (Leonard et al., 2015) but the role of PKC/NF-κB in *A3A* regulation and drug-induced expression is unknown. To investigate this, BFTC-905 and SW780 cells were pre-treated with sotrastaurin for two hours followed by gemcitabine treatment for 48 hours, in the continued presence of sotrastaurin. Sotrastaurin was used at 10 μM, the same concentration used by Leonard et al., (2015). PKC inhibition with sotrastaurin does not change basal *A3A* expression in either cell line (Figure 4.3 A and B) but modestly reduces basal *A3B* expression (Figure 4.3 C and D; 45% and 37% in BFTC-905 and SW780 cells, respectively) consistent with the findings of Leonard et al., (2015). PKC inhibition attenuates gemcitabine-mediated induction of *A3A* and *A3B* in both cell lines (Figure 4.3). The attenuation of the NF-κB target gene, *TNFα,* suggests that induction occurs via canonical NF-κB signalling downstream of PKC. However, a gene target of the alternative arm of NF-κB was not assessed and therefore, it cannot be ruled out that alternative NF-κB signalling is also required.

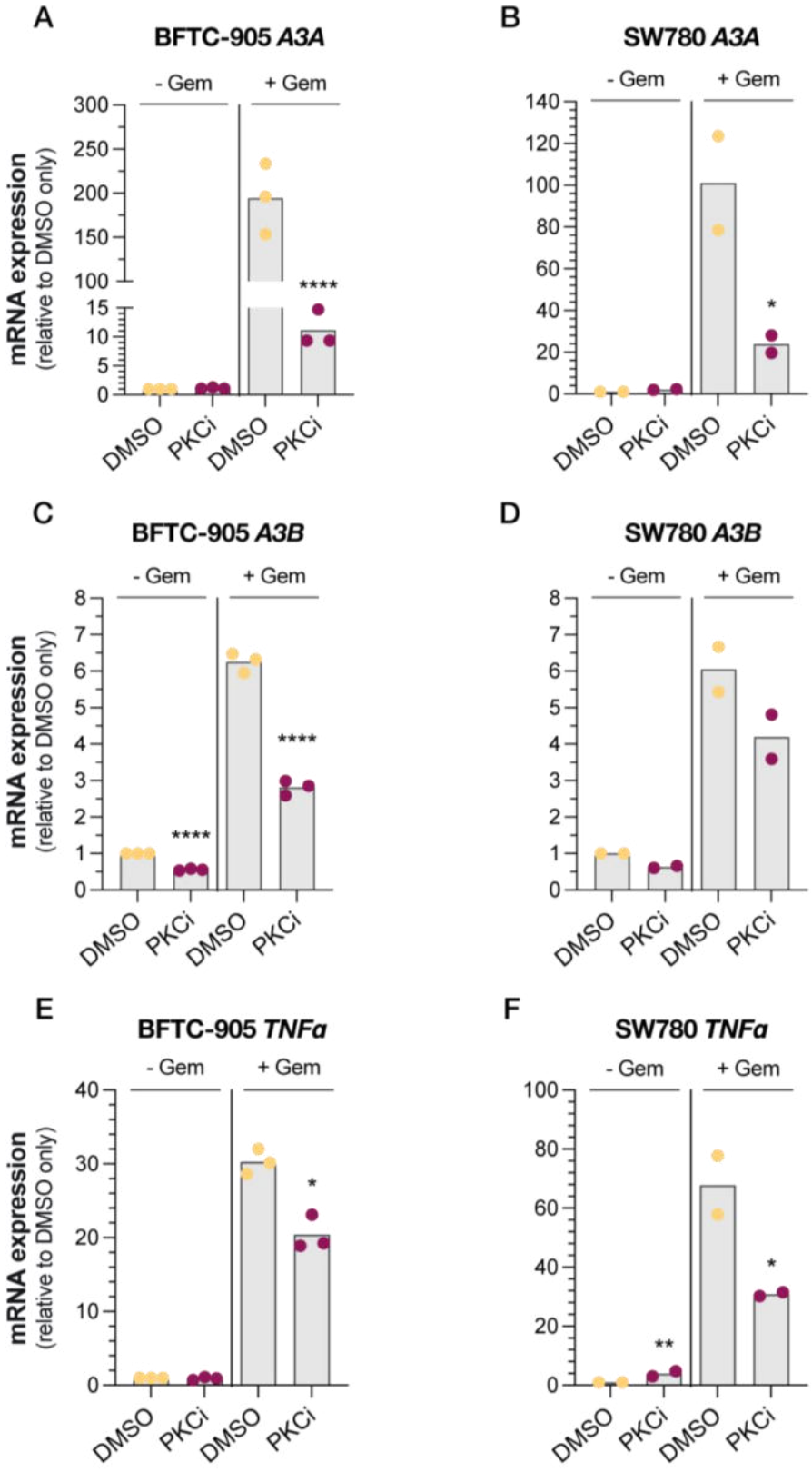


Figure 4.3 Protein Kinase C (PKC) inhibition attenuates gemcitabine-induction of *A3A, A3B* and *TNFα*.

BFTC-905 and SW780 cells were seeded at 1.5 x 10^5 cells per well in 6-well plates and after 24 hours were pre-treated with either DMSO (0.5%) or 10 µM PKCi, sotrastaurin, for 2 hours followed by 20 nM gemcitabine (Gem) treatment for 48 hours; sotrastaurin was present throughout the

experiment. *A3A* (A and B), *A3B* (C and D), and *TNFα* (E and F) mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method; dots represent independent experiments, bars represent mean. A repeated-measures one-way ANOVA followed by Sidak's multiple comparisons test was performed using GraphPad Prism 9 on ΔCt values. Within each panel (– Gem or + Gem), significance values are vs. the DMSO condition. * $P \leq 0.05$; ** $P \leq 0.01$; **** $P \leq 0.0001$.

4.2.3 Inhibition of NF-kB signalling attenuates gemcitabine-induced expression of *A3A* but not *A3B*

Taken together, these results suggest that PKC-mediated NF-κB signalling drives induction of *A3A* and *A3B* in response to drug stress. To confirm this BFTC-905 and SW780 cells were pre-treated with two NF-κB pathway inhibitors that inhibit IKK and prevent IκBα phosphorylation, BAY 11-7082 (Pierce et al., 1997) and TPCA-1 (Podolin et al., 2005), for one hour followed by gemcitabine treatment for 48 hours, in the continued presence of inhibitor. TPCA-1 attenuates induction of *A3A* in both cell lines while BAY 11-7082 only attenuates induction in BFTC-905 cells (Figure 4.4 A and B). In contrast, no attenuation of *A3B* induction is observed with either agent, in either cell line (Figure 4.4 C and D). TPCA-1 attenuates expression of *TNFα* in both cell lines, confirming inhibition of NF-κB (Figure 4.4 E and F). BAY 11-7082 treatment attenuates induction of *TNFα* in BFTC-905 cells as expected, but potentiation is seen in SW780 cells (Figure 4.4 E and F). BAY 11-7082 has been proposed to be a broad spectrum inhibitor with numerous targets (Lee et al., 2012; Rauert-Wunderlich et al., 2013) and that may contribute to the lack of *TNFα* attenuation in SW780 cells. TPCA-1 is 22-fold more selective for IKKβ (required for canonical NF-κB signalling) than IKKα (required for both arms) (Podolin et al., 2005). Further, *A3A* and *TNFα* show similar induction and attenuation patterns, suggesting that *A3A* induction is primarily mediated by canonical NF-κB signalling. Surprisingly, *A3B* induction is unchanged with NF-κB inhibition, seeming to rule out a role for this pathway. As an alternative pathway NF-κB gene target was not assessed, it cannot be ruled out that a PKC/alternative NF-κB signalling cascade is required.

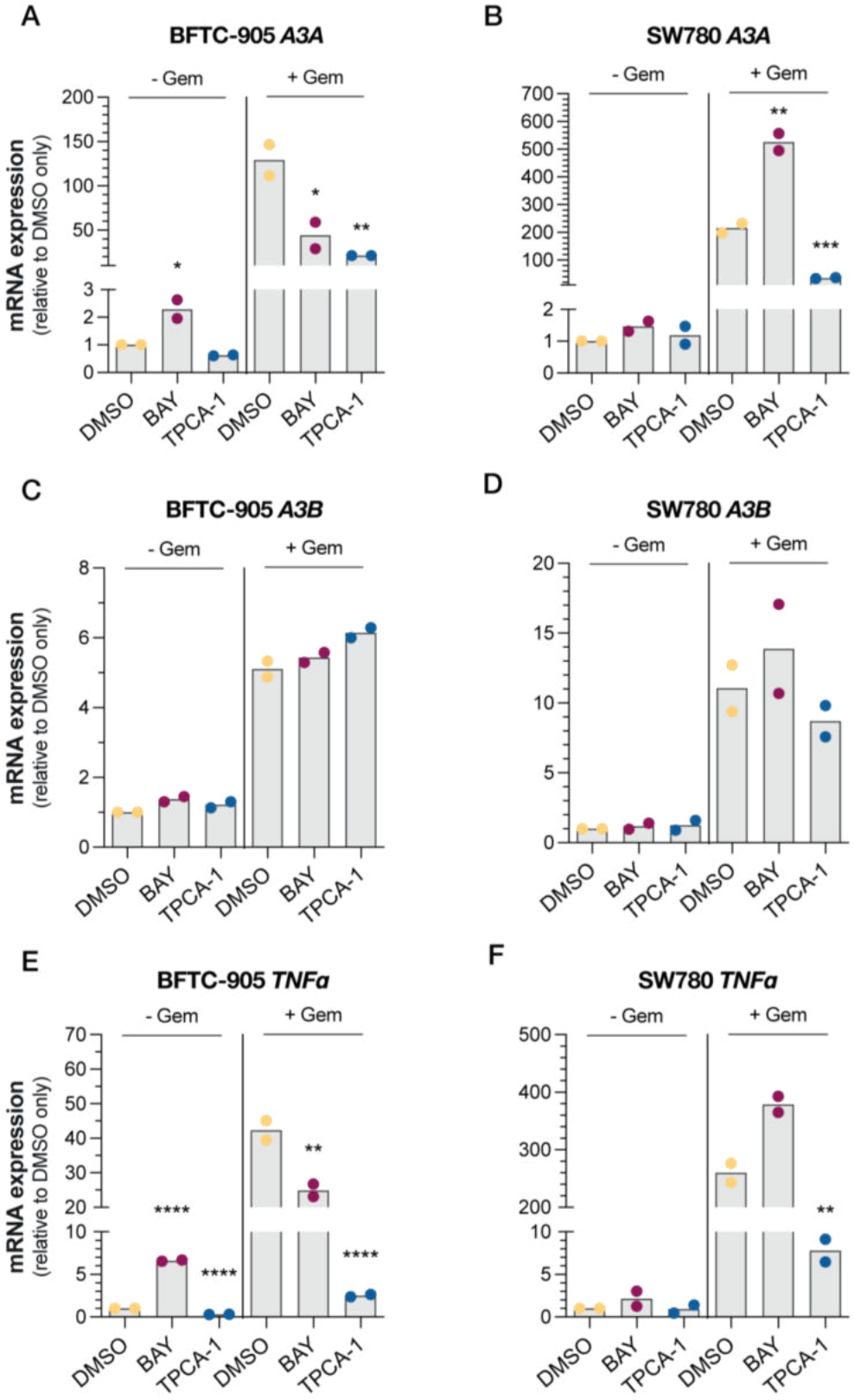


Figure 4.4 Small molecule inhibition of NF-κB signalling attenuates gemcitabine-induction of *A3A* and *TNFa* but not *A3B*.

BFTC-905 and SW780 cells were seeded at 1.5 x 10^5 cells per well in 6-well plates and after 24 hours were pre-treated with either DMSO (0.05%) or the NF-κB inhibitors, 5 μM BAY 11-7082

(BAY) and 2 μM TPCA-1 for 1 hour followed by 20 nM gemcitabine (Gem) treatment for 48 hours; NF-κB inhibitors were present throughout the experiment. *A3A* (A and B), *A3B* (C and D), and *TNFa* (E and F) mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method; dots represent independent experiments, bars represent mean. A repeated-measures one-way ANOVA followed by Sidak's multiple comparisons test was performed using GraphPad Prism 9 on ΔCt values. Within each panel (– Gem or + Gem), significance values are vs. the DMSO condition. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.

4.2.4 Induction of *A3A* is dependent on both canonical and alternative NF-kB signalling

As an alternative NF-κB gene target was not assessed and TPCA-1 is known to be more selective at inhibiting canonical NF-κB signalling, inhibition of alternative signalling was not confirmed, and it cannot be ruled out that it contributes to induction. Therefore, to fully characterise the role of NF-κB in gemcitabine-mediated induction of *A3A* and *A3B,* the canonical and alternative NF-κB transcription factors, *RELA (p65)* and *RELB*, respectively, were knocked down using siRNA prior to gemcitabine treatment for 48 hours. Assessment of knockdown efficiency with RT-qPCR showed that in the untreated conditions, *RELA* expression was reduced by 91% and 84% (Figure 4.5 A and B) while *RELB* expression was reduced by 93% and 85% (Figure 4.5 C and D) in BFTC-905 and SW780 cells, respectively. Gemcitabine treatment induces expression of *RELB* (BFTC-905, 6.8-fold; SW780, 5.8-fold) and as expected this is attenuated by knockdown of *RELB* (Figure 4.5 C and D). Knockdown of *RELA* reduced baseline levels of *RELB* by ~ 57% in BFTC-905 cells (Figure 4.5 C) and gemcitabine-mediated induction in both cell lines (Figure 4.5 C and D). While the activation of the canonical NF-κB pathway is rapid and transient, activation of the alternative arm is slow and persistent. Canonical NF-κB heterodimers are sequestered in the cytoplasm and in response to stress, are activated and rapidly accumulate in the nucleus within minutes. In contrast, the activation of the alternative arm of NF-κB requires *de novo* synthesis, and induction of the canonical NF-κB pathway results in increased expression of both RelB mRNA and protein (Bren et al., 2001). This may explain the reduction in *RELB* expression with *RELA* knockdown.

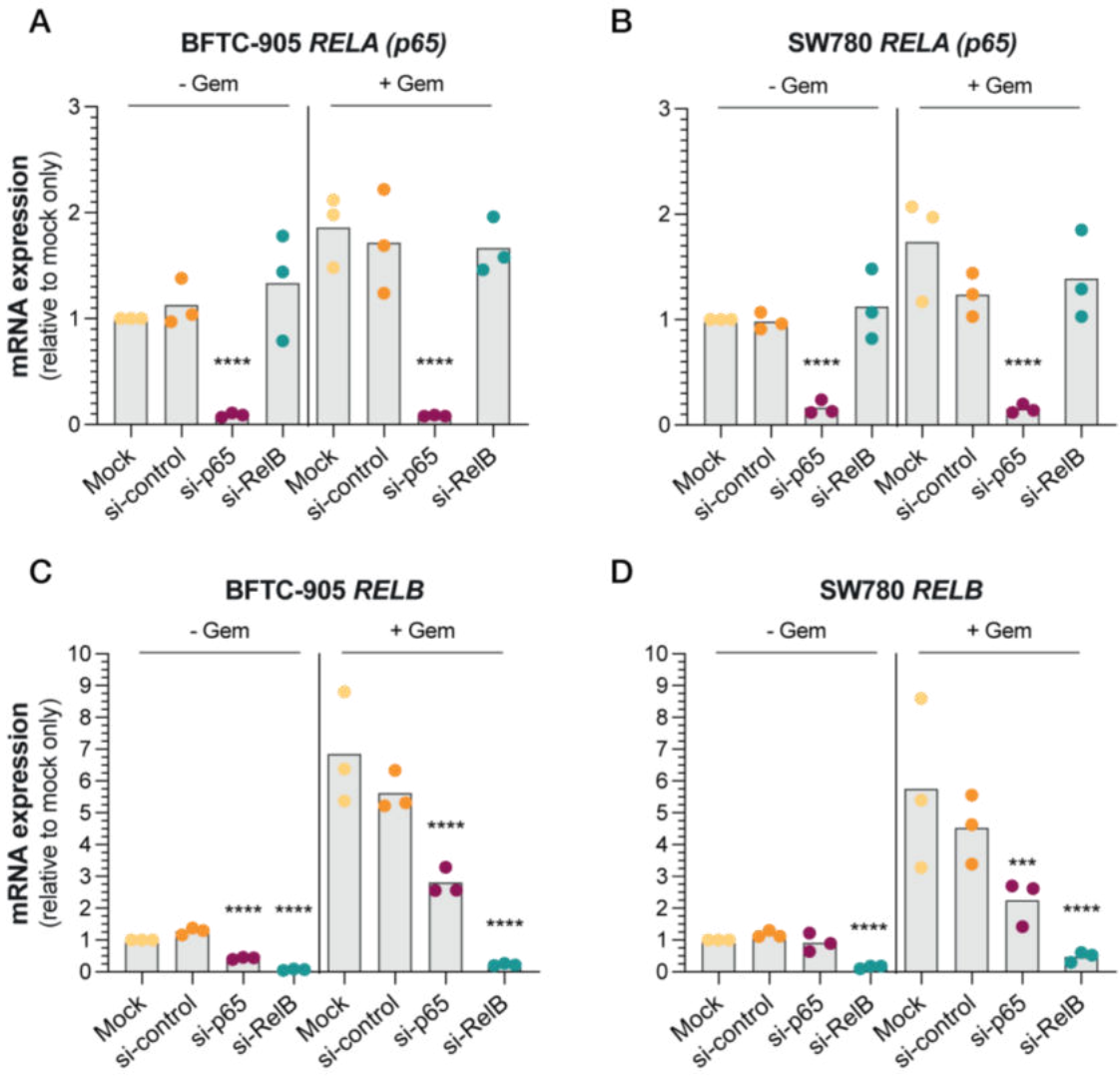


Figure 4.5 Validation of siRNA knockdown of NF-κB transcription factors, RelA (p65) and RelB.

BFTC-905 and SW780 cells were reverse transfected with SMARTpool siRNA in 6-well plates. After 24 hours, the media was changed, and cells were treated with 20nM gemcitabine for 48 hours. *RELA* (p65) (A and B) and *RELB* (C and D) mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method; dots represent independent experiments, bars represent mean. A repeated-measures one-way ANOVA followed by Tukey's multiple comparisons test was performed using GraphPad Prism 9 on ΔCt values. Within each panel (– Gem or + Gem), significance values are vs. the DMSO condition. *** $P \leq 0.001$; **** $P \leq 0.0001$. si-control = non-targeting, scrambled siRNA.

siRNA knockdown of both *RELA* and *RELB* attenuates induction of *A3A* demonstrating that both the canonical and alternative arms of NF-κB signalling are required for *A3A* induction (Figure 4.6 A and B). Taken with the finding that

gemcitabine induced expression of *RELB,* it may be that activation of canonical signalling proceeds alternative pathway activation.

In contrast to *A3A,* gemcitabine-mediated induction of *A3B* was not attenuated with knockdown *RELA* and this supports the hypothesis that canonical NF-κB signalling is not required for induction of *A3B* in UCC cells (Figure 4.6 C and D). As expression of an alternative NF-κB target gene was not assessed in the previous experiments, it was not known whether the alternative pathway is required for *A3B* induction. Induction of *A3B* was also not attenuated by *RELB* knockdown in BFTC-905 cells (Figure 4.6 C). While knockdown of *RELB* significantly attenuated *A3B* induction in SW780 cells relative to the mock transfected control, when compared to the control siRNA, this change was not significant (Figure 4.6 D). These findings demonstrate that neither canonical nor alternative NF-κB signalling is required for gemcitabine-mediated induction of *A3B* in UCC cells. This suggests that the attenuation of *A3B* seen with sotrastaurin (PKCi) is due to attenuation of an NF-κB-independent signalling pathway downstream of PKC. In conclusion, these findings show that *A3A* and *A3B* are differentially regulated in response to drug stress in UCC.

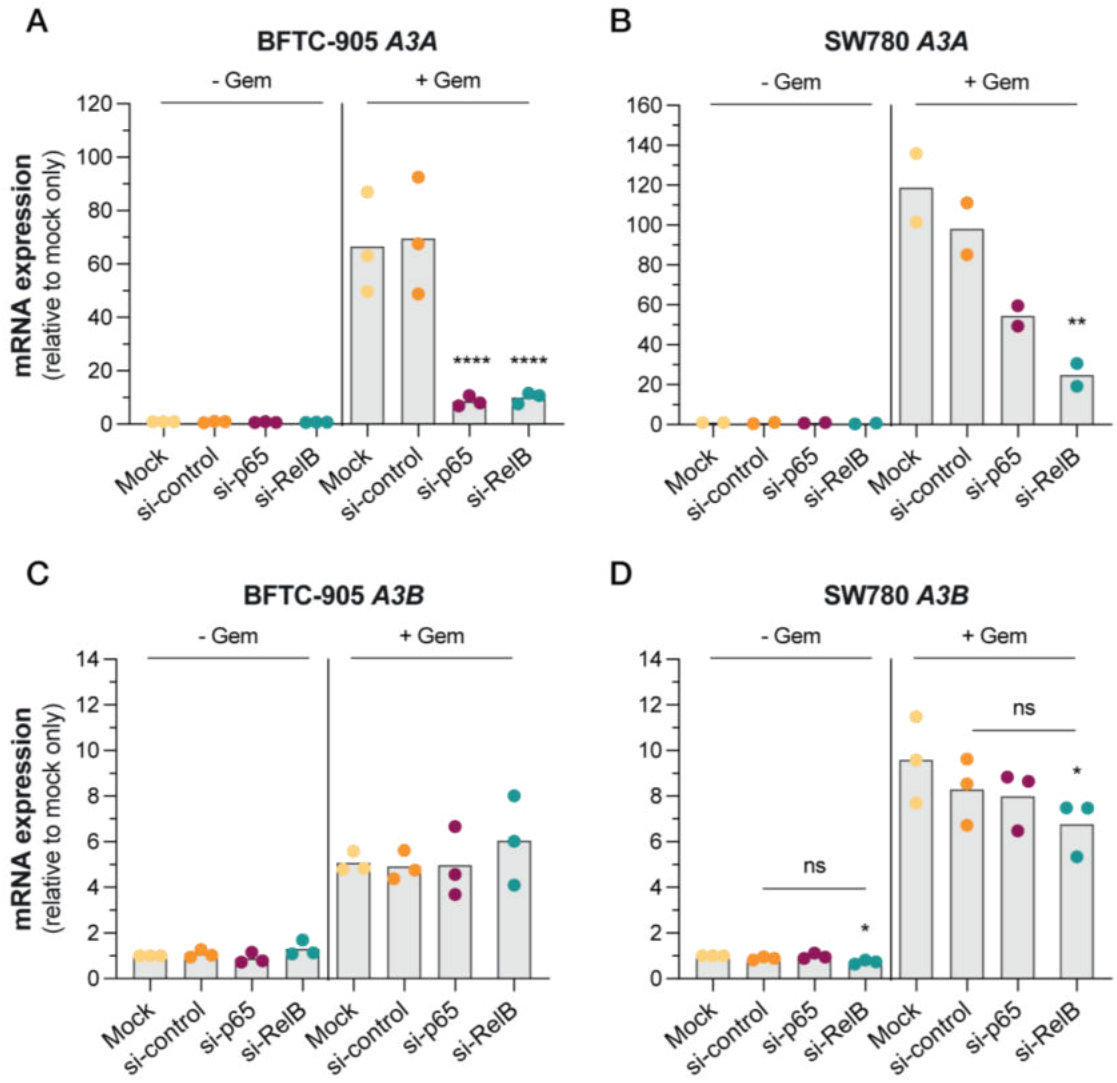


Figure 4.6 siRNA knockdown of NF-κB transcription factors p65 and RelB attenuates gemcitabine induction of *A3A* but not *A3B*.

BFTC-905 and SW780 cells were reverse transfected with SMARTpool siRNA in 6-well plates. After 24 hours, the media was changed, and cells were treated with 20nM gemcitabine for 48 hours. *A3A* (A and B) and *A3B* (C and D) mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method; dots represent independent experiments, bars represent mean. A repeated-measures one-way ANOVA followed by Tukey's multiple comparisons test was performed using GraphPad Prism 9 on ΔCt values. Within each panel (– Gem or + Gem), significance values are vs. the DMSO condition. * $P \leq 0.05$; ** $P \leq 0.01$; **** $P \leq 0.0001$. si-control = non-targeting, scrambled siRNA.

4.2.5 STING inhibition does not attenuate gemcitabine-mediated induction

Previous work in our group with NSCLC cell lines (Dr Michael Walton, ICR) has shown that short drug treatments (one to two hours) are sufficient to generate peak *A3B* induction 48 hours later, with no induction seen at early timepoints, suggesting that early DNA damage signalling is the induction-initiating event. NF-κB can be activated by cGAS-STING sensing of cytosolic DNA fragments generated during DNA damage and repair. Therefore, it is possible that extensive DNA damage caused by gemcitabine generates cytosolic DNA and this stimulates expression of *A3A* via cGAS-STING. To investigate this, BFTC-905 cells were pre-treated with the STING inhibitor, H-151, for two hours followed by gemcitabine treatment for either two or 48 hours, in the continued presence of H-151. *A3A, A3B* and *TNFα* expression was then measured 48 hours post treatment. H-151 inhibits palmitoylation of STING at the Golgi (Haag et al., 2018), a key post-translational modification required for canonical cGAS-STING signalling in response to cytosolic DNA (Mukai et al., 2016). Consistent with prior observations, induction of all three genes was seen at 48 hours, regardless of initial treatment time (Figure 4.7), and there were no significant differences in expression between two and 48 hour treatments (Table 4.1). Small molecule inhibition of STING did not attenuate short- or long-term gemcitabine stimulated expression of *A3A, A3B* or *TNFα* (Figure 4.7 A, B, C, respectively; Table 4.1). Therefore, these results suggest that gemcitabine is not driving NF-κB activation and subsequent *A3A* expression via cGAS-STING sensing of cytosolic DNA in UCC cells.

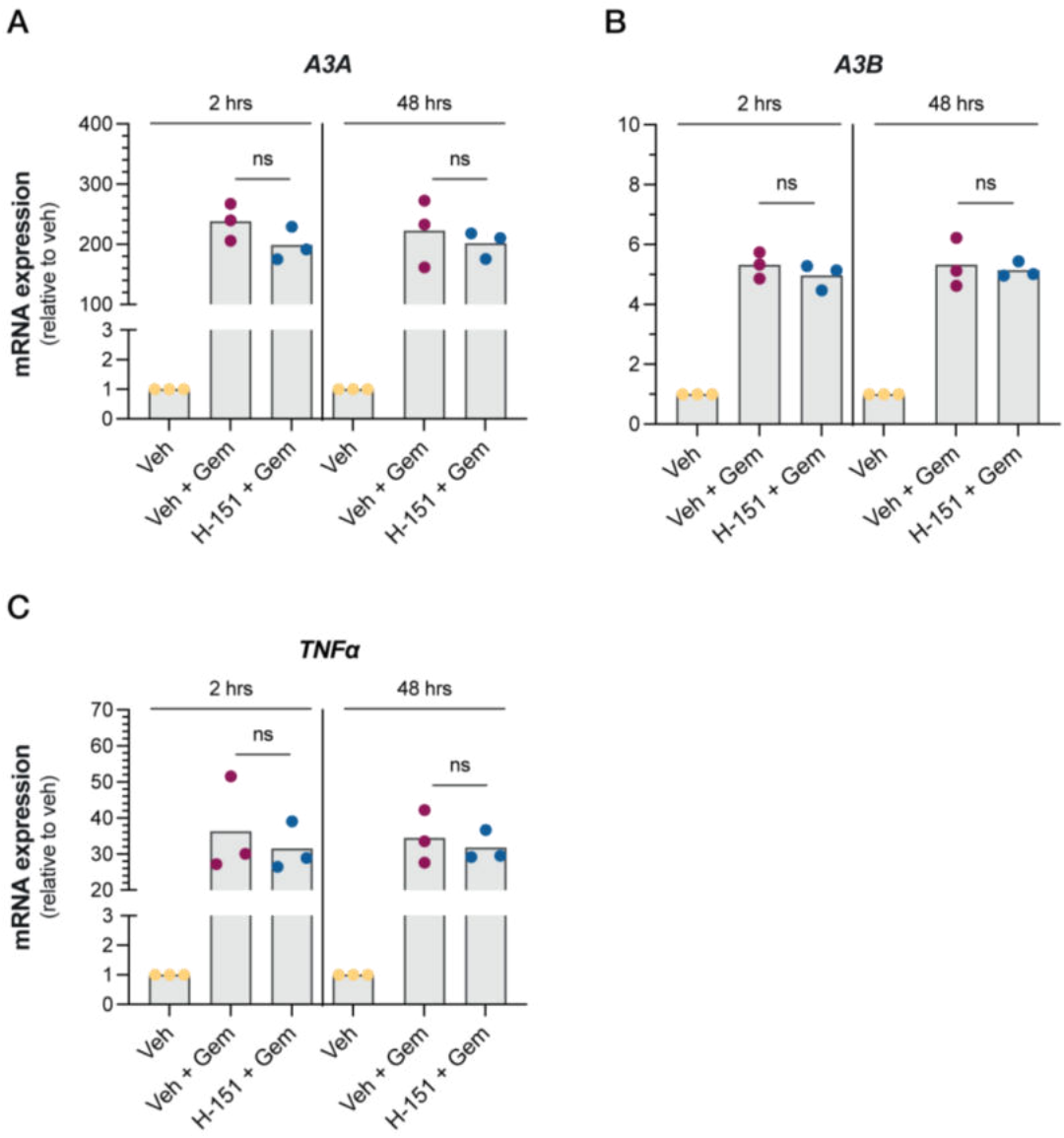


Figure 4.7 Inhibition of STING with the inhibitor, H-151, does not prevent gemcitabine-induced *A3A*, *A3B* or *TNFα* expression in BFTC-905 cells.

Cells were seeded in 6-well plates and after 24 hours were pre-treated with either vehicle (veh, DMSO, 0.1%) or 1 μM STING inhibitor, H-151, for 2 hours before gemcitabine treatment. Cells were treated for 2 hours (200 nM gemcitabine) after which media was changed and RNA was harvested after a further 46 hours, or cells were treated continuously for 48 hours (20 nM gemcitabine); H-151 was present throughout. *A3A* (A), *A3B* (B) and *TNFα* (C) mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method; dots represent independent experiments, bars represent mean. A repeated-measures one-way ANOVA followed by Sidak's multiple comparisons test was performed using GraphPad Prism 9 on ΔCt values; ns $P > 0.05$.

Table 4.1 *P* values obtained from Sidak's multiple comparisons following a repeated-measures two-way ANOVA.

Comparison	*A3B* (*P* value)	*A3A* (*P* value)	*TNFα* (*P* value)
2-hour treatment			
Vehicle vs. Vehicle + Gem	0.0002	<0.0001	<0.0001
Vehicle vs. H-151 + Gem	0.0003	<0.0001	<0.0001
Vehicle + Gem vs. H-151 + Gem	0.9987	0.8508	0.6404
48-hour treatment			
Vehicle vs. Vehicle + Gem	0.0002	<0.0001	<0.0001
Vehicle vs. H-151 + Gem	0.0002	<0.0001	<0.0001
Vehicle + Gem vs. H-151 + Gem	>0.9999	0.9998	0.9627
2-hour vs. 48-hour			
Vehicle + Gem vs. Vehicle + Gem	>0.9999	0.9997	>0.9999
H-151 + Gem vs. H-151 + Gem	>0.9999	>0.9999	>0.9999

4.2.6 ATR inhibition attenuates induction of *A3B* but not *A3A*

Drug-mediated induction of *A3B* can be attenuated through inhibition of DDR signalling in both breast and myeloma cell lines (Kanu et al., 2016; Yamazaki et al., 2020), and findings within the group in NSCLC (Dr Michael Walton, ICR) and UCC cell lines demonstrated that short drug exposures are sufficient to induce peak expression of *A3A*, *A3B* and *TNFα* at 48 hours post treatment. Together this demonstrates that DNA damage is a major driver of *A3B* expression with early DNA damage signalling as the initiating event. While DDR signalling has been implicated in the regulation of *A3B*, it is not known whether it also drives expression of *A3A* in response to drug treatment. To investigate the role of the DDR in gemcitabine-mediated induction, BFTC-905 and SW780 cells were pre-treated with either the ATM inhibitor, KU-60019, or the ATR inhibitor, AZD6738, for one hour followed by gemcitabine treatment for 48 hours, in continued presence of the inhibitor. ATM and ATR inhibition does not attenuate gemcitabine-mediated induction of *A3A* in either cell line (Figure 4.8 A and B). Interestingly, ATM inhibition and ATR inhibition potentiate *A3A* induction in BFTC-905 and SW780 cells, respectively. In contrast to *A3A*, consistent with

previous findings (Kanu et al., 2016; Yamazaki et al., 2020), ATR pathway inhibition significantly attenuates induction of *A3B* in both cell lines while ATM has no effect (Figure 4.8 C and D). Induction of *A3A* was also observed with single agent KU-60019 and AZD6738 treatment (Figure 4.8 A and B), and this is consistent with induction of *A3A* seen with ATR inhibition in NSCLC lines (Dr Caitlin McCarthy, ICR). Together, these findings suggest that in contrast to *A3B,* where DNA damage and DDR signalling activates expression, DDR signalling represses *A3A*, and further highlights that the two family members are differentially regulated.

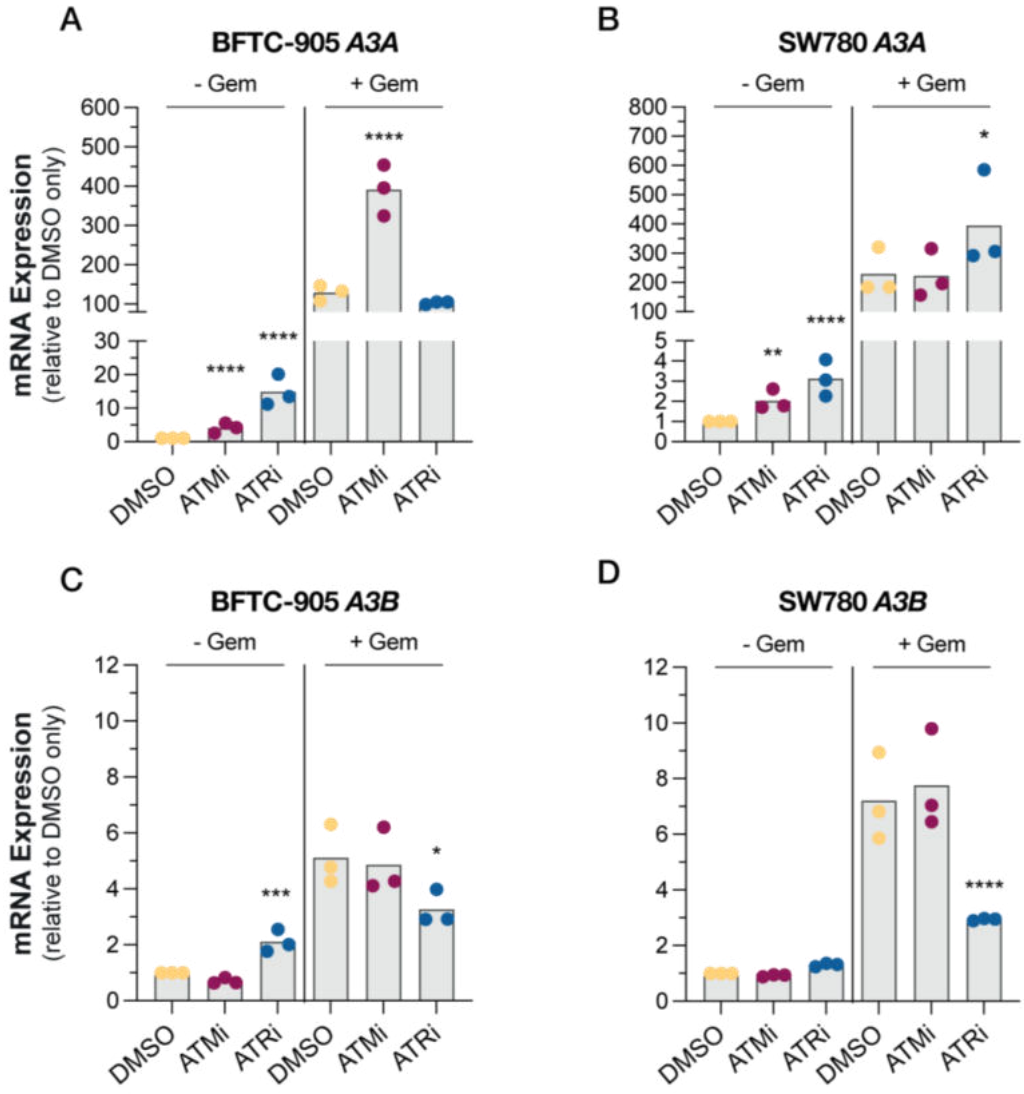


Figure 4.8 ATR inhibition attenuates gemcitabine-induction of *A3B* but not *A3A*.

BFTC-905 and SW780 cells were seeded at 1.5 x 10^5 cells per well in 6-well plates and after 24 hours were pre-treated with either DMSO (0.05%), 2 µM KU-60019 (ATMi) or 2 µM AZD6738

(ATRi) for 1 hour followed by gemcitabine (Gem) treatment for 48 hours; inhibitors were present throughout the experiment. *A3A* (A and B) and *A3B* (C and D) mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method; dots represent independent experiments, bars represent mean. A repeated-measures one-way ANOVA followed by Sidak's multiple comparisons test was performed using GraphPad Prism 9 on ΔCt values. Within each panel (– Gem or + Gem), significance values are vs. the DMSO condition. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.

4.2.7 RTKi treatment attenuates induction of *A3A, B* and *TNFα*

Dysregulated RTK signalling resulting from amplifications, mutations and oncogenic fusion is a common feature of bladder cancers (Kamoun et al., 2020; Lindskrog et al., 2021). Luminal subtype tumours and cell lines are commonly associated with FGFR3 and HER2 regulon activity and frequently harbour *FGFR* or *ERBB2* alterations (amplification, activating mutations and oncogenic fusions), while basal subtypes are commonly associated with EGFR regulon activity. The two cell lines used in this chapter were chosen as they robustly induce both *A3A* and *A3*B in response to gemcitabine, are models for a basal subtype MI-UCC (BFTC-905) and luminal subtype NMI-UCC (SW780) and therefore, are useful for investigating whether RTKs are involved in *A3* regulation. The SW780 cell line harbours the *FGFR3-BAIAP2L1* fusion (Williams et al., 2013) and has moderate cell surface expression of both EGFR and HER2 (Siddiqui et al., 2019). Inhibition of HER2 signalling with lapatinib reduces both basal and hydroxyurea-mediated expression of *A3B* and A3 activity in breast cancer cell lines (Kanu et al., 2016). However, at the time this work was conducted, no studies had investigated the role of RTK signalling in *A3A* regulation. Therefore, to investigate the role of RTK signalling in both baseline and gemcitabine-induced expression of *A3A* and *A3B,* expression after treatment with three RTKis was measured.

Growth IC_{50} values for erlotinib (EGFRi), lapatinib (HER2i) and infigratinib (FGFR3i) were determined for BFTC-905 and SW780 (Figure 4.9). Cells were then pre-treated with erlotinib, lapatinib or infigratinib at 2.5 x IC_{50} for one hour followed by treatment with gemcitabine for 48 hours, in the continued presence of the inhibitors. As RTK signalling is linked to NF-κB activation, expression of the canonical target gene*, TNFα*, was also measured. Single-agent infigratinib treatment increased expression of *A3A* in both BFTC-905 and SW780 cells while

lapatinib also modestly induced expression in BFTC-905 cells (Figure 4.10 A and B). Inhibition of RTK signalling pathways reduces baseline expression of *A3B* in breast cell lines but single-agent RTKi treatment had no effect on baseline expression of *A3B* in UCC lines (Figure 4.10 A and B). This shows that, in contrast to findings in breast (Kanu et al., 2016), RTK signalling is likely not responsible for the elevated levels of *A3B* in UCC tumours relative to normal tissue.

While single-agent RTK treatment modestly induces *A3A* expression and has no effect on expression of *A3B*, RTKi use in combination with gemcitabine attenuates gemcitabine-mediated induction of both *A3A* and *A3B.* Lapatinib and infigratinib significantly attenuate induction of both *A3A* (Figure 4.10 A and B) and *A3B* (Figure 4.10 C and D) in both cell lines. Erlotinib also attenuates induction of *A3A* and *A3B* although the attenuation did not reach statistical significance in BFTC-905 cells.

Infigratinib treatment attenuates induction in response to gemcitabine in both cells lacking any FGFR3 alterations (BFTC-905), and the FGFR3 fusion-containing cell line (SW780). This suggests that whether a particular RTKi induces or attenuates expression of *A3A* and/or *A3B* is not solely dependent on the genetic background of the cell line. The lack of full abrogation of gemcitabine-mediated induction observed for each of the RTKis suggests that gemcitabine is more generally activating RTK signalling, and downstream signalling cascades activate the expression of both *A3A* and *A3B*. RTKi treatment also attenuates gemcitabine-mediated induction of *TNFα* (Figure 4.10 E and F). Of note, *A3A* and *TNFα* expression patterns once again mirror each other in both the single-agent and combination treatment conditions suggesting that the effect of RTKi treatment on *A3A* expression is mediated via canonical NF-κB signalling. Interestingly, single-agent RTKi treatment induces NF-κB activity and downstream *A3A* and *TNFα* expression but attenuates gemcitabine-mediated induction when used in combination. Taken with the findings that NF-κB is not required for gemcitabine-mediated induction of *A3B,* this suggests that RTKis can attenuate induction of both family members, but the downstream pathways are different.

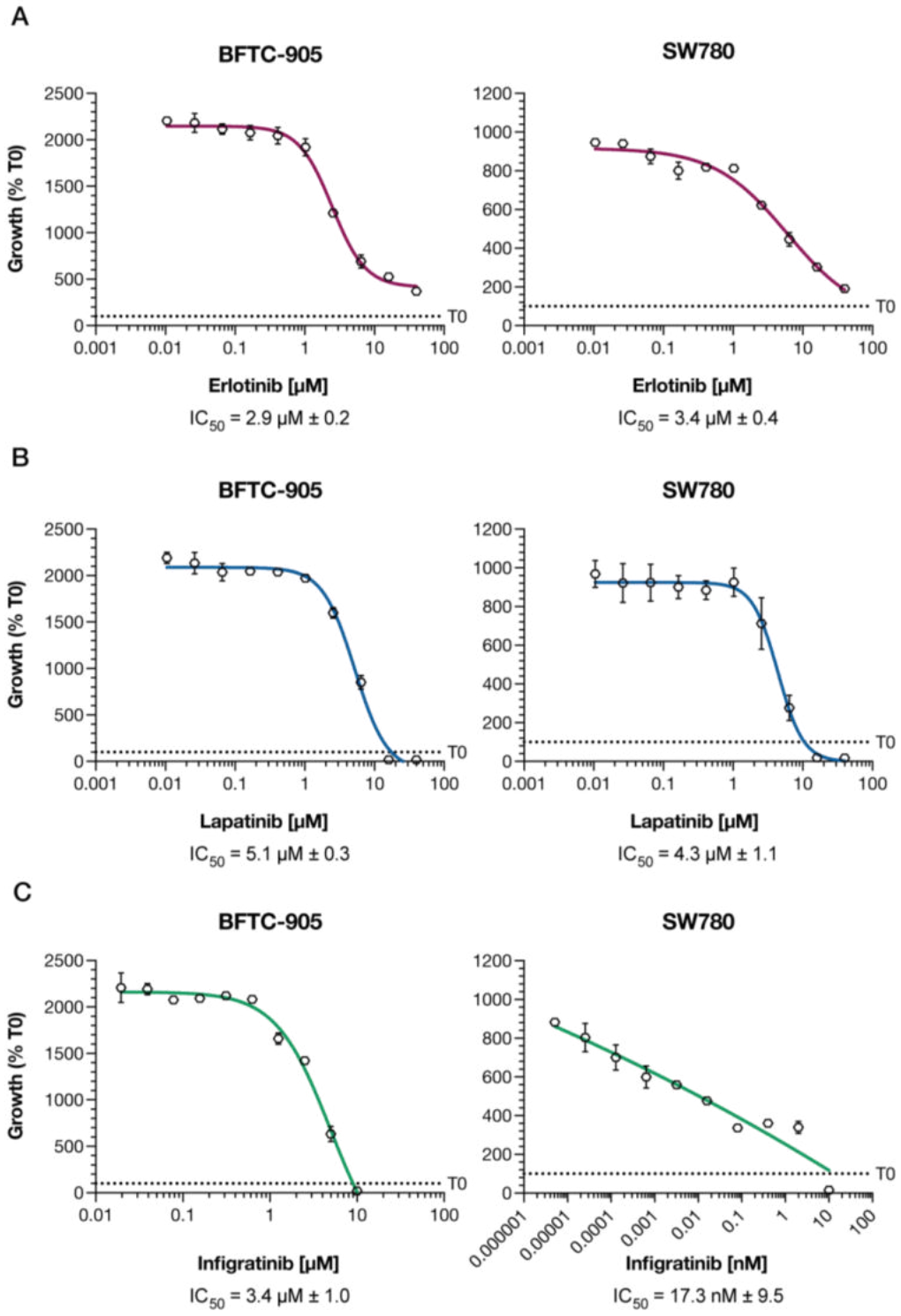


Figure 4.9 Concentration-response curves showing sensitivity of UCC cell lines, BFTC-905 and SW780, to receptor tyrosine kinase inhibitors.

Curves for BFTC-905 and SW780 cells for erlotinib, EGFR inhibitor (A), lapatinib, HER2 inhibitor (B) and infigratinib, FGFR3 inhibitor (C). Response was analysed after 96 hours drug treatment using the SRB assay. Data is shown as mean ± SD for two independent experiments. IC_{50} values shown represent the average and standard deviation of the values obtained from each independent experiment using non-linear regression on GraphPad Prism 9.

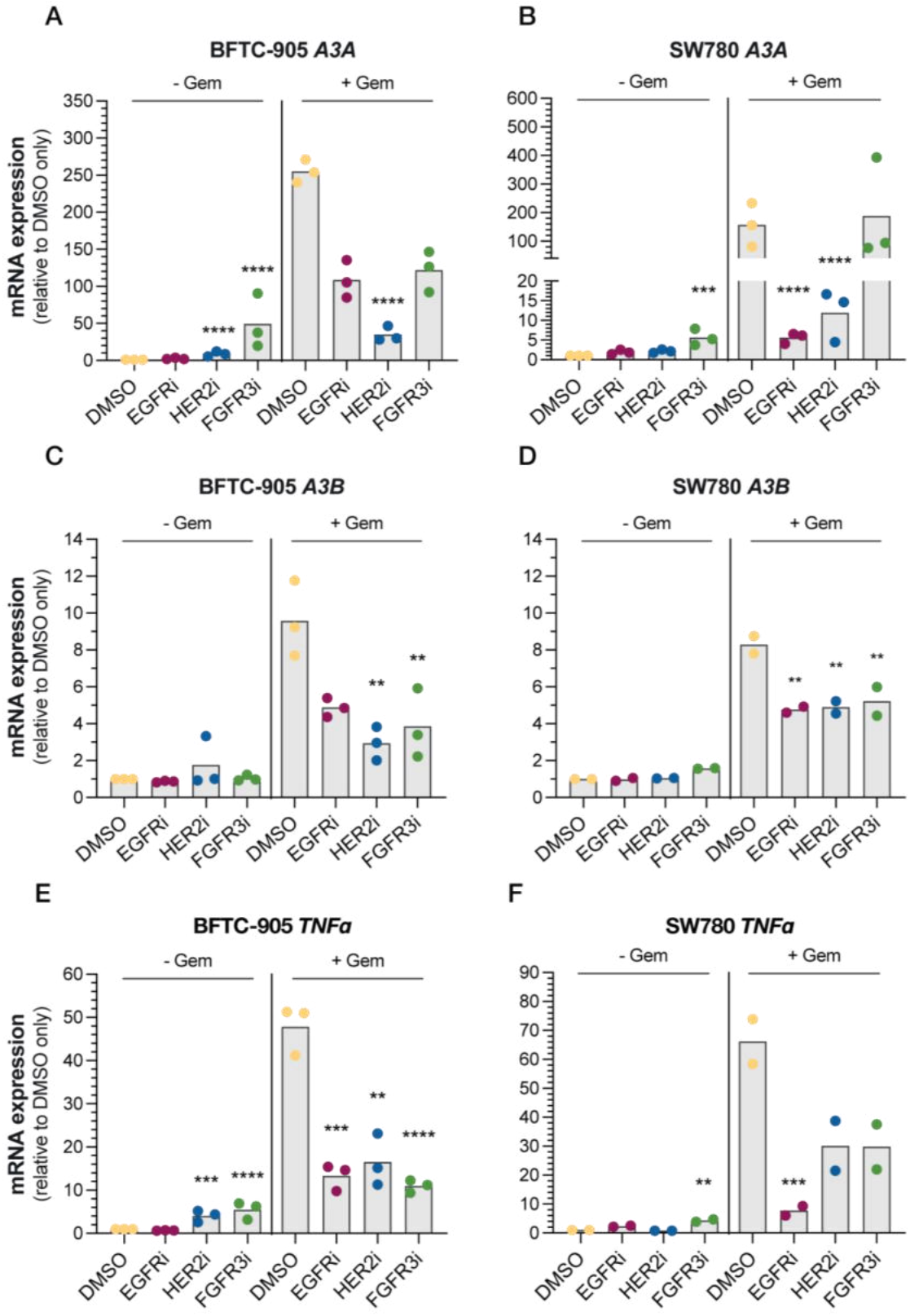


Figure 4.10 Receptor tyrosine kinase inhibitor (RTKi) treatment partially attenuates gemcitabine-induction of *A3A, A3B* and *TNFα* when used in combination for 48 hours.

BFTC-905 (A, C and E) and SW780 (B, D and F) cells were seeded at 1.5 x 10^5 cells per well in 6-well plates and after 24 hours were pre-treated with either DMSO (0.2%), erlotinib (EGFRi), lapatinib (HER2i) or infigratinib (FGFR3i) for 1 hour followed by gemcitabine (Gem) treatment for

48 hours; inhibitors were present throughout the experiment. *A3A,* (A and B), *A3B* (C and D) and *TNFα* (E and F) mRNA expression was measured with RT-qPCR and analysed using the ΔΔCt method. Data shown as fold changes relative to the DMSO control; dots represent independent experiments, bars represent mean. A repeated-measures one-way ANOVA followed by Sidak's multiple comparisons test was performed using GraphPad Prism 9 on ΔCt values. Within each panel (– Gem or + Gem), significance values are vs. the DMSO condition. ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.

4.3 Discussion

When this work began, a PKC-mediated NF-κB signalling cascade had been identified as a driver of elevated *A3B* in tumours but there was uncertainty as to whether this cascade is dependent on the canonical (Maruyama et al., 2016) or alternative (Leonard et al., 2015) arms of NF-κB. As numerous anticancer drugs have been shown to activate NF-κB (Nakanishi & Toi, 2005), it was hypothesised here that the observed drug-mediated induction of *A3B*, and potentially *A3A*, was occurring due to activation of NF-κB signalling. In support of this, work presented in this chapter showed that induction of the canonical NF-κB target gene, *TNFα*, occurs in response to gemcitabine and bleomycin treatment in a concentration-dependent manner. Of the two drugs tested, gemcitabine was identified as the most robust inducer of both *A3B* (Chapter 3) and *TNFα,* and this suggested that the level of *A3B* induction is proportional to the level of NF-κB activation. Further work showed that PKC inhibition with sotrastaurin, attenuates induction of both *A3A* and *A3B*, and together these findings suggested that the PKC/NF-κB signalling axis is key in regulating, not only *A3B* as previously reported (Leonard et al., 2015; Maruyama et al., 2016), but also *A3A.* To confirm that NF-κB signalling is required for the observed induction, cells were co-treated with gemcitabine and the NF-κB inhibitors, BAY 11-7082 and TPCA-1, and unexpectedly, this revealed differences in how *A3A* and *A3B* are regulated in response to chemotherapy. While small molecule inhibition of NF-κB attenuates induction of *A3A*, *A3B* expression is unaffected. These findings contrast with those in a recent study where small molecule inhibition of NF-κB signalling with the two inhibitors used in this work, does attenuate induction of *A3B* in response to chemotherapy in breast tissue derived cell lines (Periyasamy et al., 2021). The patterns of induction and attenuation for *A3A* and *TNFα* expression were strikingly similar, and this suggested that canonical NF-κB signalling is driving *A3A* induction. A gene target of the alternative arm of NF-κB was not assessed and the involvement of this branch of the pathway in *A3A* or *A3B* induction could not be determined but these results did demonstrate that *A3A* and *A3B* are differentially regulated. Further interrogation into the two arms of NF-κB using siRNA to knockdown the canonical, *RELA*, and alternative, *RELB*, transcription factors confirmed that NF-κB was required for induction of *A3A*. This is consistent with recent findings that demonstrated that RelA is enriched at *A3A*'s

transcriptional start site after hydroxyurea treatment, and *RELA* depletion attenuates induction (Oh et al., 2021). Attenuation of *A3A* expression was achieved by knockdown of both *RELA* and *RELB* demonstrating that both arms of the pathway contribute to the observed induction. Gemcitabine also induced expression of *RELB*, which could be attenuated by *RELA* knockdown. This, taken with the knowledge that activation of the canonical arm is rapid, while activation of the alternative arm is slow, requires *de novo* synthesis and RelA stimulates expression of *RELB* (Bren et al., 2001), suggests that canonical activation may first occur followed by sustained activation of the alternative pathway. To investigate this further, western blotting should be used to determine which factors are accumulating in the nucleus in response to gemcitabine treatment over time. The attenuation of *A3B* induction observed with PKC inhibition was initially thought to occur via inhibition of downstream NF-κB signalling, but this is not the case. In contrast to *A3A*, induction of *A3B* was not attenuated with *RELA* or *RELB* knockdown and this confirmed that *A3B* induction occurs independently of NF-κB signalling in UCC cells. This shows that another factor downstream of PKC drives *A3B* expression and other downstream PKC-regulated transcription factors should be investigated. Again, these results contrast to findings in breast cancer cells where knockdown of *RELA* and *NFKB1* (encoding p105/p50) with siRNA prevents cisplatin- and etoposide-mediated expression of *A3B* (Periyasamy et al., 2021). Together, these findings demonstrate that the pathways involved in regulation of *A3B* (and *A3A*), likely vary among cancer types.

Hydroxyurea-stimulation of *A3B* expression and A3 activity in breast cancer and myeloma cell lines can be attenuated with ATR, CHK1 and DNA-PK inhibition (Kanu et al., 2016; Yamazaki et al., 2020), and findings presented in this chapter and by others (Dr Michael Walton, ICR) showed that short drug treatments induce peak expression at 48 hours. Together, these findings demonstrate that early DNA damage is the initiating event. Anticancer agents can activate NF-κB through the DDR or by cGAS-STING activation in response to cytosolic DNA fragments generated during DNA damage and subsequent repair. Therefore, the role of DNA damage cGAS-STING signalling in *A3* induction was assessed. Inhibition of STING with the small molecule inhibitor, H-151 (Haag et al., 2018), did not attenuate gemcitabine-mediated induction of either *A3A* or *TNFα*

suggesting that NF-κB activation occurs via DDR signalling of DNA damage, not cytosolic DNA sensing. It is worth noting that stimulation of cGAS-STING with cGAMP was not used in this study and therefore, it cannot be confirmed that the inhibitor was indeed inhibiting STING. STING also has a non-canonical, cGAS-independent function, acting alongside ATM to activate NF-κB in response to DNA damage by etoposide and this occurs independently of canonical STING trafficking and palmitoylation (Dunphy et al., 2018). As H-151 prevents palmitoylation of STING at the Golgi (Haag et al., 2018), and knockout studies were not conducted, this non-canonical role of STING in *A3A* or *A3B* induction could not be evaluated. However, a recent study showed that hydroxyurea-mediated induction of *A3A* occurs in STING knockout cells (Oh et al., 2021). Thus, it seems that neither canonical nor non-canonical functions of STING are required for induction of *A3A* in response to gemcitabine.

Investigation of DDR signalling shows that ATR inhibition attenuates gemcitabine-mediated *A3B* expression in UCC cell lines. This is consistent with the findings of others and confirms that RS is a major driver of *A3B* expression in tumour cells (Kanu et al., 2016; Yamazaki et al., 2020). ATM inhibition did not attenuate induction of *A3B* in UCC cells in response to gemcitabine, and this is consistent with the findings of Yamazaki et al., (2020) who did not observe attenuation of hydroxyurea stimulated expression of *A3B* in myeloma cells with ATMi. However, ATM inhibition did attenuate cisplatin and etoposide stimulation of *A3B* in breast cancer cells (Periyasamy et al., 2021). The authors showed that inhibition and knockdown of ATM significantly reduces induction of A3B mRNA and protein, while ATR inhibition or knockdown has minimal or no effect. They also showed that DNA-PK, a DDR kinase that was not investigated in this thesis, is required for *A3B* induction, and that induction is mediated by DNA-PK/ATM activation of the canonical NF-κB signalling pathway. These results differ to those presented in this thesis where ATM and NF-κB are not required for *A3B* induction in response to gemcitabine in UCC. This is likely a result of tumour specific differences in *A3B* regulation in addition to the type of DNA damage generated by the drugs investigated. Gemcitabine is a nucleoside analogue that inhibits DNA synthesis when it is incorporated into DNA and, like hydroxyurea, inhibits ribonucleotide reductase leading to depletion of the nucleotide pool. Therefore, these drugs primarily prevent DNA replication and induce RS. On the other hand,

etoposide inhibits topoisomerase II, an enzyme that simultaneously cuts both strands of DNA to resolve supercoiling, while cisplatin generates intra- and interstrand crosslinks, the repair of which can generate DSBs. As such, the dependence on DNA-PK, ATM and ATR is likely related to the type of genotoxic stress and lesion generated by the drug under investigation. In support of this, DNA-PK but not ATM was required for hydroxyurea-induction of *A3B* (Yamazaki et al., 2020) and this suggests that RS inducing drugs require DNA-PK and ATR, while cisplatin- and etoposide-mediated induction is dependent on DNA-PK and ATM. Further investigation into whether DNA-PK is required for gemcitabine-mediated induction, and the dependence on DDR kinases with drugs that generate different types of DNA damage, will address whether this is indeed the case. Finally, while ATR is clearly implicated in regulation of *A3B*, the downstream transcription factors required to activate expression are currently unidentified.

In contrast to *A3B*, induction of *A3A* is not attenuated by ATR inhibition, further demonstrating that these family members are differentially regulated in response to stress. Instead, ATR inhibition further increased *A3A* expression in SW780 cells. This is consistent with the findings of Oh et al., (2021), who saw that ATR inhibition does not attenuate hydroxyurea or gemcitabine induction of *A3A* and instead potentiates it. Interestingly, the authors showed that cells treated with both ATR and hydroxyurea treatment had much higher levels of γ-H2AX than those treated with hydroxyurea alone, suggesting that increased expression of *A3A* requires fork collapse, DSB formation and replication catastrophe. Therefore, the lack of potentiation seen in BFTC-905 cells may be because they are more tolerant of RS and had not been pushed into replication catastrophe. Combined hydroxyurea and ATRi treatment reduces levels of the NF-κB inhibitor protein, IκBα, confirming that ATR negatively regulates NF-κB and as such, the increased *A3A* expression observed is due to relief of NF-κB repression (Oh et al., 2021). Therefore, under conditions of RS, ATR acts to minimise NF-κB driven *A3A* expression and limit further damage but is paradoxically required for the upregulation of another cellular threat, *A3B*. ATM inhibition does not attenuate induction of *A3A* despite it having a known key role in activation of NF-κB in response to DNA damage. However, activation of NF-κB and *A3B* expression in other cancer types is dependent on DNA-PK (Periyasamy et al., 2021; Yamazaki

et al., 2020) and it is possible that this is also the case for *A3A* in UCC. Further work should evaluate the role of DNA-PK in *A3A* induction.

RS and NF-κB were confirmed as being key drivers of *A3B* and *A3A* in UCC cells, respectively, and RTK signalling is implicated in both these processes. Therefore, work was conducted to assess whether RTK inhibitors could attenuate gemcitabine-mediated induction. Indeed, three RTKis targeting EGFR, HER2 or FGFR3 attenuate induction of both family members. Other work in this chapter demonstrated that PKC and NF-κB are required for induction of *A3A,* and the attenuation of *TNFα* observed with RTKi treatment suggests that attenuation of *A3A* is via a reduction in NF-κB activation. On the other hand, NF-κB is not required for induction of *A3B*, but PKC seems to play an important role. Other pathways downstream of RTKs were not investigated and as such, their contribution to *A3B* induction should be investigated. The ability of a particular RTKi to attenuate induction appeared to occur independently of the genetic background of the cell lines tested, as BFTC-905 cells do not have FGFR3 alterations and are not sensitive to the FGFR3i, infigratinib, but induction could be attenuated with FGFR3 inhibition. Interestingly, a number of anticancer agents can activate RTKs (Benhar et al., 2002; Chun et al., 2006; Furugaki et al., 2010; Miyabayashi et al., 2013; Van Schaeybroeck et al., 2005) and upregulate their expression in a process potentially mediated by NF-κB (Kan, Koido, Okamoto, Hayashi, Ito, Kamata, Komita, Ishidao, et al., 2015). Together, this suggests that gemcitabine activates RTKs more generally, rather than a specific one, and RTKi treatment counteracts this activation, with inhibition of divergent downstream signalling resulting in attenuation of both *A3A* and *A3B* expression. This hypothesis is supported by the findings that RTK inhibition only ever attenuated, not completely abrogated, induction and it is likely that multiple RTKs are being activated simultaneously. RTK signalling can also drive RS through increased origin firing, replication, transcription, and nucleotide depletion. Therefore, it is also possible that RTKi treatment alleviates RS and induction is curbed in this manner.

An interesting observation made in this work is that RTK inhibitors, when used as single agents, modestly induce expression of *A3A* despite attenuating gemcitabine-mediated induction. Where this was seen, *TNFα* induction was also observed, and this suggests that the general stress caused by the inhibitors is

activating NF-κB, likely independently of the drugs mechanism of action. Induction of *A3A* in NSCLC cell lines in response to the EGFRi, osimertinib, has recently been reported in a preprint (Isozaki et al., 2021). The authors demonstrated that induction was dependent on NF-κB signalling as knockdown of *IKKa* and *NFkB1* attenuated *A3A* expression. These findings demonstrate the complex mechanisms at play where targeted therapies can both induce and attenuate expression depending on if they are used as single agents or in combination with a DNA damaging drug.

In conclusion, the data presented in this chapter shows that *A3A* and *A3B* are differentially regulated in UCC cells and that multiple pathways are working in combination to induce these family members in response to gemcitabine RS. Figure 4.11 shows the proposed model of regulation of *A3A* and *A3B* in response to genotoxic stress in UCC cells from the findings presented in this chapter.

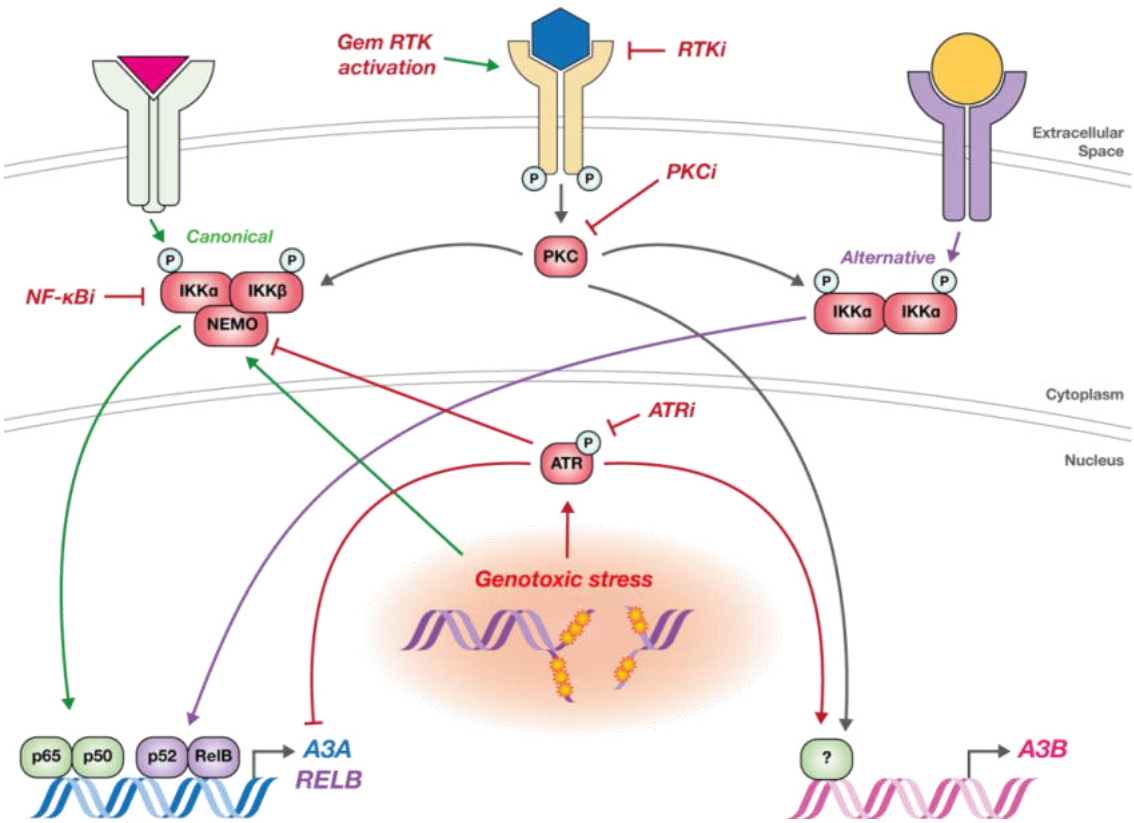


Figure 4.11 *A3A* and *A3B* transcriptional regulation in response to genotoxic stress in UCC.

Gemcitabine induces both replication stress (RS) and activates receptor tyrosine kinases (RTKs). Downstream PKC signalling activates NF-κB and subsequent expression of *A3A*. PKC can also induce expression *A3B* although the factors involved are currently unknown. RS activates ATR signalling that activates expression of *A3B*, although again the factors involved downstream of ATR are uncharacterised, while simultaneously negatively regulating NF-κB signalling and minimising expression of *A3A*.

Chapter 5

Establishing APOBEC3 cell line models

Chapter 5 Establishing APOBEC3 cell line models

5.1 Introduction

Identification of a protein's interactome gives significant insight into where the protein is localised and what it is doing in cells. To date, very little is known about the interacting partners of A3A and A3B, limiting the understanding of substrates they act on, what their function in cells is and why tumour cells upregulate them. In addition, the cellular localisation of A3A is currently disputed; expression of exogenous A3A revealed a cell-wide distribution (Lackey et al., 2013) but endogenously induced A3A is retained in the cytoplasm and is non-genotoxic (Land et al., 2013). Therefore, it is possible that the observed cell-wide distribution of exogenous A3A is a result of mislocalisation caused by GFP tagging or massively increased levels. Further, A3A is highly homologous to the CTD of A3B suggesting they could have redundant functions. Two approaches were chosen to examine their localisation, function in cells, and if there is evidence of redundancy: 1) the proximity-labelling technique BioID, where proteins are tagged with a biotin ligase that biotinylates proximal proteins, followed by mass spectrometry to identify interacting partners; and 2) expression of exogenous A3A or A3B followed by transcriptomics (Chapter 6). However, our own experience and that of others suggested that A3B is toxic to *E. coli* and generating the required constructs for these studies can be challenging. In contrast to small epitope tagged or tag-free A3B, GFP tagged A3B is cloned and expressed with relative ease, suggesting that GFP is attenuating its function, leading to mitigation of the toxicity. It is now widely accepted that tagging with fluorescent proteins has limitations; they can affect protein localisation, folding and function, and many have a tendency to oligomerise which can lead to protein aggregation and toxicity (Chudakov et al., 2010; Snapp, 2005). N-terminal GFP-tagging of the A3 cousins, AID and A1, affects their intracellular localisation resulting in cytoplasmic retention (Patenaude et al., 2009; Severi & Conticello, 2015). In the case of A1, N-terminal tagging with GFP reduces efficiency of RNA editing (Severi & Conticello, 2015), while C-terminal tagging affects its ability to interact with cofactors and dimerise (Chieca et al., 2021). Therefore, A3A and B tagged with a small epitope that is unlikely to interfere with function is desirable as the proteins can be detected by western blotting in the absence of good

commercially available antibodies, without the potential issues arising from GFP tagging. To overcome the toxicity issues associated with A3B, many groups have disrupted the cDNA open reading frame with an intron to prevent production wild-type, full-length active protein in *E. coli* (Akre et al., 2016; Hultquist et al., 2011; St. Martin et al., 2018).

At the start of this project the constructs available to conduct the required experiments were limited. Lentiviral vectors containing GFP tagged wild-type and catalytically inactive (A3B**; E68Q/E255Q) A3B, and A3B N- and C-terminally tagged with the biotin ligase, BirA, were available but there were concerns with their suitability, and there were no A3A constructs available to use in the group. Therefore, this chapter describes work generating a set of constructs with both A3A and A3B tagged with a biotin-ligase and the small epitope V5 to generate cell line models for transcript- and proteomic studies.

Transient transfection is a useful but limited technique; protein expression only occurs for up to 96 hours due to plasmid loss, expression levels are hard to control, and variability in transfection efficiency can cause reproducibility issues between experimental repeats. While generating cell lines that have stably integrated DNA encoding the protein of interest into the genome is far more labour intensive, it allows for constitutive expression, facilitating long term studies of protein function with reduced experimental variability. Isolation of stable clones and use of inducible systems allows further control of expression levels. In the simplest way of making stable cell lines, a small number of cells transiently transfected with plasmid DNA will randomly integrate the plasmid into the genome, but this method is inefficient and time-consuming. The use of drug resistance markers is key for selective screening of cells that have integrated the exogenous DNA as integration frequencies are low, often in the range of 0.01% of transfected cells. Therefore, it can take months to expand cells with the proper integration sufficiently to use in experiments and not all cell lines are amenable to expansion from single cells. Use of linearised plasmids can increase integration frequency and reduce the likelihood of plasmid breakage and gene disruption during integration, but linearisation also increases DNA-mediated toxicity and can reduce transfection efficiency.

Lentiviral transduction involves simultaneous transfection of multiple plasmids into producer cell lines; a transfer plasmid containing the insert of interest flanked by long terminal repeat (LTR) sequences that facilitate integration into the genome of target cells, alongside packaging and envelope plasmids required for the producer cells to generate viral particles (Cooray et al., 2012). The resulting viral particles are released into the cell culture medium of the producer cell line and are used to infect and integrate the insert into the genome of the target cell (Bulcha et al., 2021; Cooray et al., 2012; Sakuma et al., 2012). While lentiviral transduction is more efficient than transfection-based methods of generating stable cell lines (Tandon et al., 2018), there are concerns around safety to the user, primarily around the random generation of replication-competent lentivirus, oncogenic potential of inserts and insertional mutagenesis (Cooray et al., 2012; Elegheert et al., 2018). With both methods, the transgenes integrate into the genome randomly and can cause unpredictable effects through insertional mutagenesis. Intergenic insertion can cause dysregulation of adjacent genes through transgene promoter action while intragenic insertions can cause gene inactivation or the production of aberrant gene products (Papapetrou & Schambach, 2016; Sadelain et al., 2012). The ideal integration locus would be one that was transcriptionally active and facilitates stable expression of an integrated transgene while not disrupting the genome or causing discernible phenotypic effects; such sites have been designated genomic safe harbour (GSH) loci. While it can be argued that no true GSH locus exists, targeted integration into a defined genomic locus has benefits over random integration. As integration is at a defined locus, the resulting stable cell line pool can, to some extent, be considered isogenic; therefore, clonal isolation to account for differential insertional mutagenesis is not absolutely required. The main concern is that random off-target integration may have occurred elsewhere in the genome, although the risk of this is small and can be identified by Southern Blotting. At GSH loci, transgene copy number is limited (to the number of alleles per cell; 1 or 2 in diploid lines) and this enables homogenous and physiological expression, in comparison to the heterogenous expression seen in virally transduced pools resulting from copy number and integration site variation. Three loci have been extensively used as defined integration sites and are referred to as GSHs: adeno-associated virus site 1 *(AAVS1)*, chemokine C-C motif receptor 5 *(CCR5)* and the

human orthologous sequence of the mouse loci *Rosa26* (Papapetrou & Schambach, 2016; Sadelain et al., 2012). The *AAVS1* locus is located in intron one of the *PPP1R12C* gene on chromosome 19 and was initially identified as a natural integration site for AAV (Kotin et al., 1992; Tan et al., 2001). This locus has been used extensively and is a good candidate for a genomic safe harbour; *PPP1R12C* has open chromatin, is constitutively expressed in most cell types and as such, allows stable and homogenous expression of integrated transgenes (Lombardo et al., 2011; Shin et al., 2020). In addition, insertion into *AAVS1* does not increase expression of neighbouring genes, likely due to the presence of a transcriptional insulator (Li et al., 2009; Lombardo et al., 2011; Ogata et al., 2003).

A3 cytidine deaminases are a key part of the innate immune response and are upregulated in response to viral infection and cytoplasmic DNA/RNA in a process largely mediated by interferons (Stavrou & Ross, 2015). Viral infection induces expression of A3 enzymes and deamination of retroviral ssDNA intermediates generates G-to-A mutations in the coding strand that can trigger ssDNA degradation, in addition to nonsense and missense mutations in the integrated provirus. A3 enzymes are also upregulated *in vitro* upon RNA and DNA transfection (Oh et al., 2021; Suspène et al., 2017), and deaminate and destabilise plasmid DNA *in vitro* (Stenglein et al., 2010) and *in vivo* (Kostrzak et al., 2015), reducing transient gene expression and stable gene transfer efficiency (Stenglein et al., 2010). Therefore, due to their normal biological function, their action complicates the process of making stable cell lines, with lentiviral transduction methods potentially more problematic than simple plasmid transfection. A3 activity in packaging cell lines, through endogenous induction or constitutive expression of encoded A3s from the transfer plasmid, and subsequent plasmid deamination and destabilisation results in vector hypermutation and reduced viral titres (Delviks-Frankenberry et al., 2019; Kremer et al., 2006; Miller & Metzger, 2011). In addition, viral transduction of the target cell mimics natural viral infection and induces A3 expression. The subsequent deamination of ssDNA intermediates can generate mutations in integrated transgenes or stimulate degradation resulting in reduced integration frequencies. shRNA depletion of A3B in myeloma cells prevents loss of fluorescence in lentiviral-transduced cells with stably integrated mCherry (Yamazaki et al., 2019). By 21-days post-transduction, PCR of the transgene revealed that it was present

in A3B-depleted cells but was almost completely lost in control cells, suggesting that endogenous A3 actively drives loss of integrated lentiviral transgenes. Therefore, this is something to consider when working with cell lines with high basal A3 expression and activity.

The use of inducible promoters can minimise the risk of introducing mutations and the effects of A3 activity on gene-transfer efficiency when working with plasmids encoding A3 family members. Keeping expression off also prevents accumulation of the A3 signature during stable cell line selection and mitigates against isolation of a population that has adapted to elevated A3 expression prior to conducting experiments. The Tet-On inducible system is one of the most widely used; upon binding doxycycline, the transactivator protein undergoes a conformational change and binds to a tetracycline responsive promoter to induce target gene expression, which can be precisely tailored as induction is dependent on doxycycline concentration (Gossen & Bujard, 1992; Gossen et al., 1995). This is particularly useful when lower, more physiologically relevant levels of expression are required. For example, when conducting BioID experiments, lower expression levels are required to reduce mislocalisation and false interactions that can occur upon high level overexpression (Roux et al., 2018). Therefore, it was decided that the most appropriate models would be those with A3A/B under the control of a doxycycline-inducible promoter with the transgene stably integrated into the *AAVS1* locus. Copy number limitation coupled with doxycycline-inducibility will facilitate expression of elevated, but more physiologically relevant levels.

Targeted integration into the *AAVS1* locus can be achieved through the use of TAL effector nucleases (TALENs), zinc-finger nucleases (ZFNs) or the clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 system that uses small RNAs to guide the Cas9 nuclease to cleave specific genomic sequences (Jinek et al., 2012). In this thesis, CRISPR/Cas9 was used with a donor repair template based on that built by Dalvai et al., (2015) using a validated guide RNA (gRNA) sequence (Mali et al., 2013). CRISPR/Cas9 induces a DSB in *AAVS1* and the transgene is subsequently integrated into the genome via repair from the Tet-On autoregulated donor template; antibiotic resistance gene-trapping, allows selective recovery of targeted cells (Figure 5.1) (Dalvai et al., 2015).

Like other methods, CRISPR/Cas9 is not without its limitations: it is well-known that off-targeting editing events can occur (Fu et al., 2013), which can be reduced through the use of high-fidelity Cas9 mutants (Vakulskas et al., 2018). Prolonged Cas9 activity, as seen with plasmid-based methods, is associated with increased off-target effects; thus, the most effective way to reduce off-target editing is to limit Cas9 activity (Chen et al., 2020). This can be done effectively through direct delivery of the precomplexed gRNA and Cas9 protein to cells as ribonucleoprotein (RNP) complexes; the levels of RNP are initially high, facilitating rapid genome-editing while subsequent rapid degradation limits off-target editing (Kim et al., 2014).

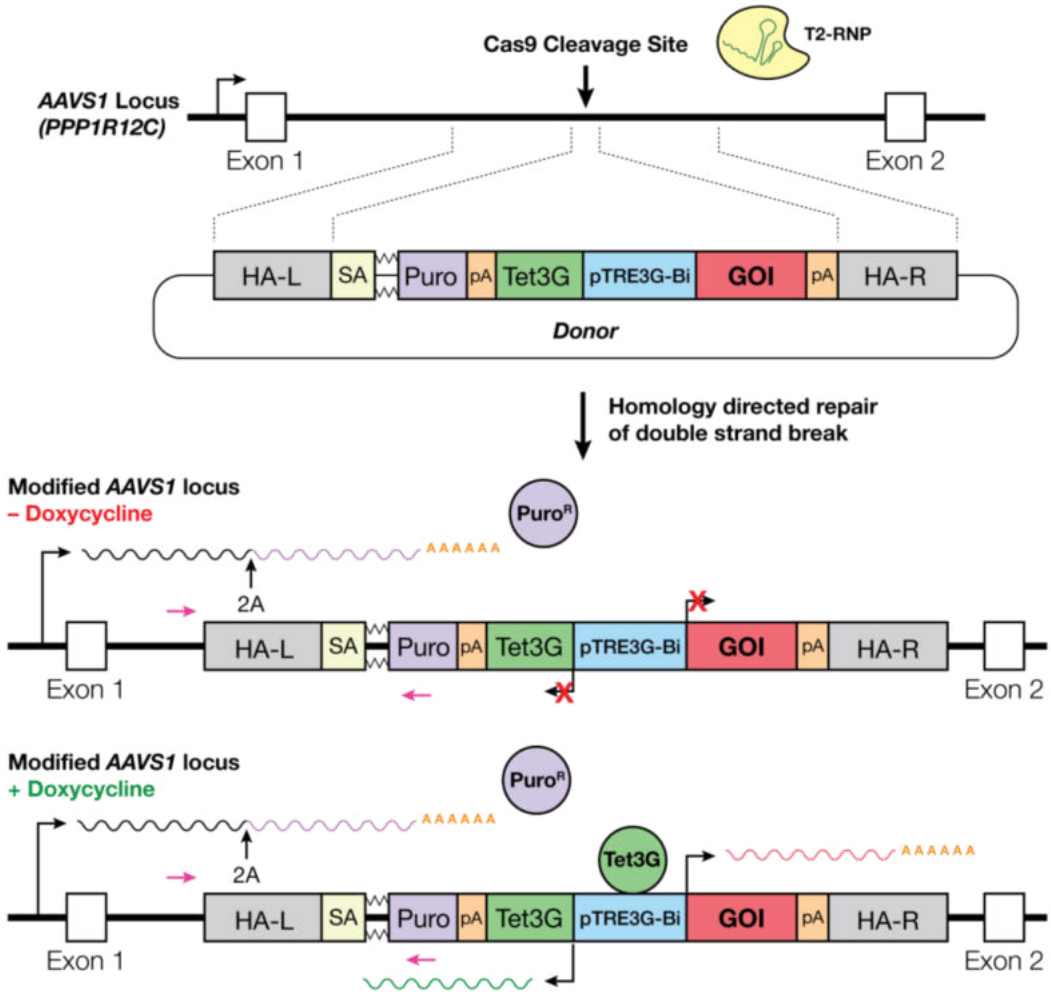


Figure 5.1 CRISPR/Cas9 system for integration of transgenes into the *AAVS1* locus.

CRISPR/Cas9 induced DSBs are generated, and integration of the transgene occurs through the repair of the DSB via homologous recombination from the donor template containing left and right homology arms (HA-L/HA-R). A splice acceptor site (SA) allows constitutive expression of the

resistance cassette ($Puro^R$ shown in figure) from the endogenous *PPP1R12C* gene promoter, and a 2A self-cleaving peptide sequence (2A) allows production of functional $Puro^R$ protein and selection of edited cells. In the presence of doxycycline, the Tet3G transactivator binds to the bidirectional Tet-responsive promoter (pTRE3G-Bi), inducing expression of itself and the gene-of-interest (GOI). PCR primers used for integration screening are show in pink. pA; polyadenylation signal. System and figure based on that by Dalvai et al., (2015).

5.2 Results

To identify interacting partners of A3A and A3B, a proximity labelling technique called BioID will be used (Roux et al., 2012). This involves expressing the protein of interest tagged with a biotin-ligase that biotinylates proximal proteins that can then be purified and identified using mass spectrometry. For this study, the newer biotin-ligase mutants with improved biotinylation properties, TurboID and miniTurbo, will be used (Branon et al., 2018).

The initial strategy for generating the required cell lines was to first clone TurboID and miniTurbo tagged A3A, A3B and GFP into the pcDNA3.1 vector for validation of expression and biotinylation activity via transient transfection experiments. The results would determine which biotin ligase was most appropriate for use in the BioID experiments and these constructs would then be used to make doxycycline-inducible *AAVS1*-integrated stable cell lines. The BioID constructs would then be used to generate the A3A and A3B expression constructs through removal of TurboID/miniTurbo to generate wild-type and catalytic dead A3A and B tagged with the V5 epitope. However, the cloning process was extremely difficult, several obstacles had to be overcome and the process changed drastically during the project.

5.2.1 A3B, but not A3A, biotin-ligase fusions are toxic to *E. coli*

A3B, GFP, TurboID and miniTurbo fragments with overlapping ends were generated by PCR and were assembled into pcDNA3.1 using the NEBuilder HiFi DNA Assembly Method. PCR of the resulting NEB assembly reaction mixtures using vector-specific primers confirmed successful assembly (Figure 5.2 A). The reaction mixture was transformed into NEB® 5-alpha Competent *E. coli* and ampicillin-resistant colonies were screened by colony PCR. All GFP transformants analysed generated products consistent with successful cloning (Figure 5.2 B) and Sanger Sequencing confirmed correct assembly. A3B recombinants could not be recovered despite multiple transformation attempts with varying conditions, including growth at lower temperatures, suggesting *E. coli* toxicity (Figure 5.2 C).

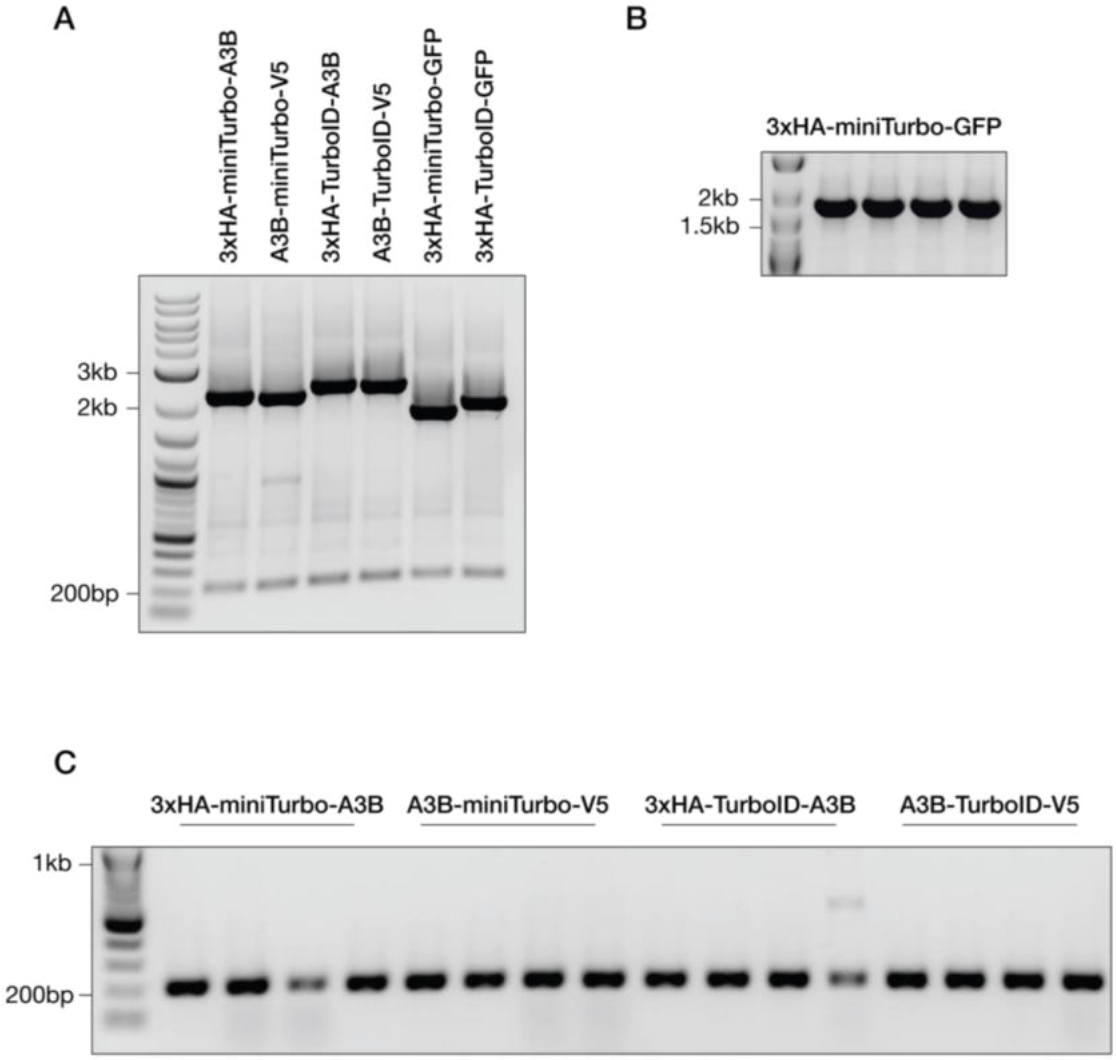


Figure 5.2 miniTurbo/TurboID-tagged A3B is toxic to *E. coli* and recombinants were not recovered.

(A) PCR with 0.5 µL of the NEB assembly reaction mixture shows successful assembly into pcDNA3.1. (B) Colony PCR was performed on ampicillin-resistant colonies using vector-specific primers. 3xHA-miniTurbo-GFP transformants with a PCR product consistent with successful cloning were recovered (1766 bp). (C) miniTurbo/TurboID-tagged A3B transformants produced a PCR product consistent with vector-only nonrecombinants (212 bp).

While our lab and others have been unable to easily clone untagged full length A3B, A3B tagged with TurboGFP and BirA has successfully been cloned in the doxycycline-inducible pTRIPZ lentiviral vector. As TurboID (35 kDa) and miniTurbo (28 kDa) are a similar size to BirA (35 kDa) and TurboGFP (26 kDa), the inability to recover A3B recombinants was unexpected. To investigate whether the issues encountered were due to the pcDNA3.1 backbone, the

transgenes were subcloned into pTRIPZ (Figure 5.3). Hundreds of ampicillin-resistant colonies were recovered after transformation with GFP ligation mixtures, but ampicillin-resistant colonies could not be recovered from A3B transformants. This suggests that the promoters in both backbones are “leaky” and that the problems encountered are due to toxicity of tagged A3B rather than an issue with the pcDNA3.1 vector. To make troubleshooting the toxicity issues more manageable, only miniTurbo tagged constructs were progressed further. miniTurbo was chosen over TurboID because its smaller size means it is less likely to interfere with the function and localisation of A3A/B, and it has less background biotinylation activity in the absence of biotin compared to TurboID (Branon et al., 2018).

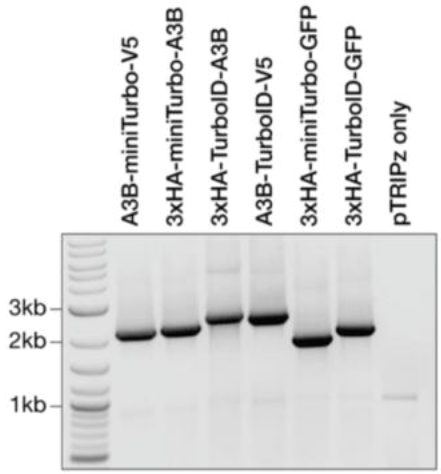


Figure 5.3 Successful ligation of miniTurbo tagged A3B and GFP into pTRIPZ.

Transgenes were PCR amplified from the NEB assembly reaction mixtures with a forward primer containing an AgeI site and a reverse primer containing an ClaI site and subcloned into pTRIPZ. PCR using vector-specific primers with 0.5 µL of the ligation reaction mixture shows successful insertion.

To determine whether tagged A3A is also toxic, N- and C-terminally miniTurbo-tagged A3A were assembled into pTRIPZ, and ampicillin-resistant colonies were screened by colony PCR. Transformants with a product consistent with successful cloning were identified, and Sanger Sequencing confirmed correct assembly in select clones demonstrating that these A3A fusions are not toxic to *E. coli* (Figure 5.4).

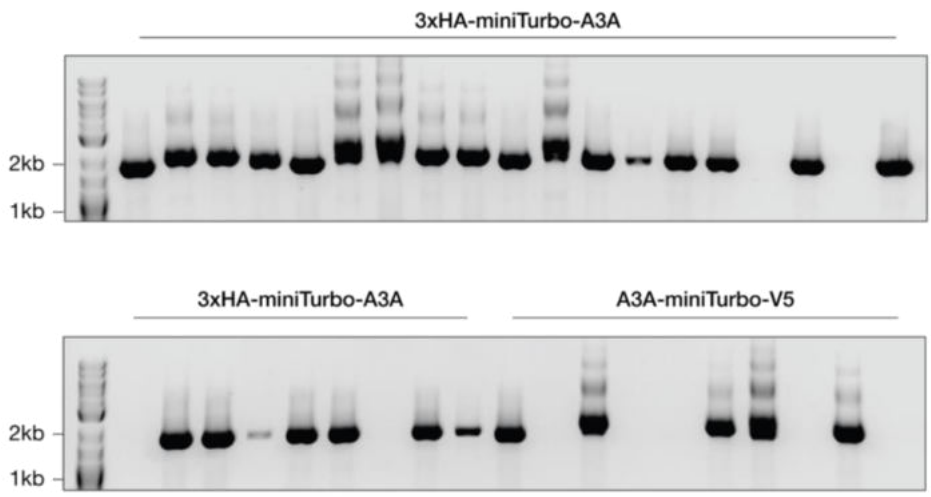


Figure 5.4 miniTurbo-tagged A3A is not toxic to *E. coli*.

Colony PCR was performed on ampicillin-resistant colonies using vector-specific primers. Transformants with PCR products consistent with successful cloning were identified (3xHA-miniTurbo-A3A = 1692 bp; A3A-miniTurbo-V5 = 1644 bp).

5.2.2 Disruption of the *A3B* ORF prevents expression of toxic A3B fusions enabling cloning in *E. coli*

To circumvent A3B toxicity in *E. coli,* a novel cloning technique was employed (Krela et al., 2019). This method involves inserting a stuffer fragment into a unique native restriction enzyme site within the cDNA to disrupt the reading frame and prevent translation of full length A3B in case of promoter leakage ($A3B^{Split}$). The stuffer is removed with an *in vitro* digestion and ligation reaction to restore the open reading frame (ORF) prior to transfection (Figure 5.5 A). A unique BsrGI site was identified in the ORF of *A3B*, and a stuffer fragment was inserted into pTRIPZ_A3B-GFP to form the pTRIPZ_$A3B^{Split}$-GFP construct (Figure 5.5 B). N- and C-terminally miniTurbo tagged $A3B^{Split}$ were assembled into linearised pcDNA3.1, and transformants with a product consistent with successful cloning were identified and verified by Sanger Sequencing (Figure 5.5 C). Therefore, disrupting the ORF of *A3B* circumvents A3B toxicity.

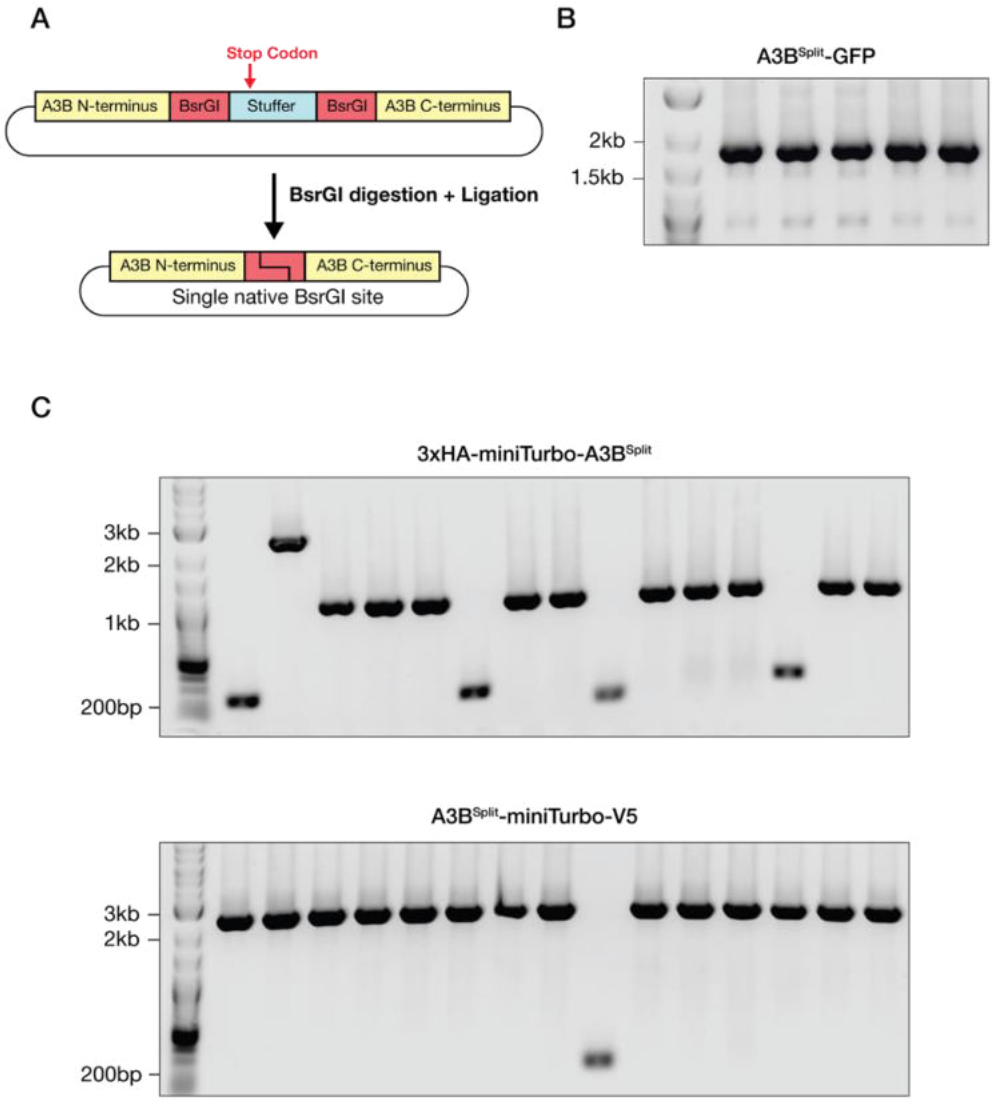


Figure 5.5 Disruption of the *A3B* open reading frame facilitates cloning of miniTurbo-tagged A3B.

(A) Schematic showing disruption of the open reading frame of *A3B* through insertion of a stuffer into a native BsrGI restriction site. (B) A 230 bp stuffer was PCR amplified and inserted into BsrGI linearised pTRIPz_A3B-GFP to generate pTIRPz_A3B^{Split}-GFP. Transformants with PCR products consistent with successful cloning were identified (1815 bp). (C) 3xHA-miniTurbo-A3B^{Split} and A3B^{Split}-miniTurbo-V5 transformants had PCR products consistent with successful cloning into pcDNA3.1 (3xHA-miniTurbo-A3B^{Split} = 2440 bp; A3B^{Split}-miniTurbo-V5 = 2395 bp; nonrecombinants = 212 bp). The 1119 bp product corresponds to 3xHA-miniTurbo-NLS_pCDNA3 parental template carryover.

5.2.3 Sub-cloning into the AAVS1-TetON system

To generate doxycycline-inducible stable cell lines, miniTurbo tagged A3A, A3B and GFP were PCR amplified from the previously made, sequence verified plasmids and assembled into the AAVS1_Puro_Tet3G_3xFLAG_Twin_Strep backbone (Dalvai et al., 2015). This vector targets the transgene to the *AAVS1* safe-harbour locus and utilises a Tet3G® promoter that lacks binding sites for endogenous mammalian transcription factors to ensure there is no basal transcription in the absence of doxycycline. During PCR linearisation of the backbone, the 3xFLAG_Twin_Strep tag was removed as this would interfere with affinity purification of biotinylated proteins. PCR of the NEB assembly reaction mixtures using vector-specific primers confirmed successful assembly (Figure 5.6 A), ampicillin-resistant colonies were screened by colony PCR and successful cloning events were identified for all constructs (Figure 5.6 B). Two clones for each were sent for sequencing and all, except 3xHA-miniTurbo-A3A that contained an A-deletion in the A3A start codon, were correct. The A deletion was restored using the Q5® Site-Directed Mutagenesis Kit. All plasmids were maxi-prepped using an endotoxin-free kit ready for mammalian cell transfections and were re-sequenced. PCR using vector-specific primers and sequencing confirmed successful removal of the stuffer and restoration of the *A3B* ORF after digestion with BsrGI and ligation with T4 ligase (Figure 5.7). The plasmids were purified by isopropanol precipitation with endotoxin-free reagents and were confirmed to have the correct sequence by Sanger Sequencing prior to mammalian cell transfections.

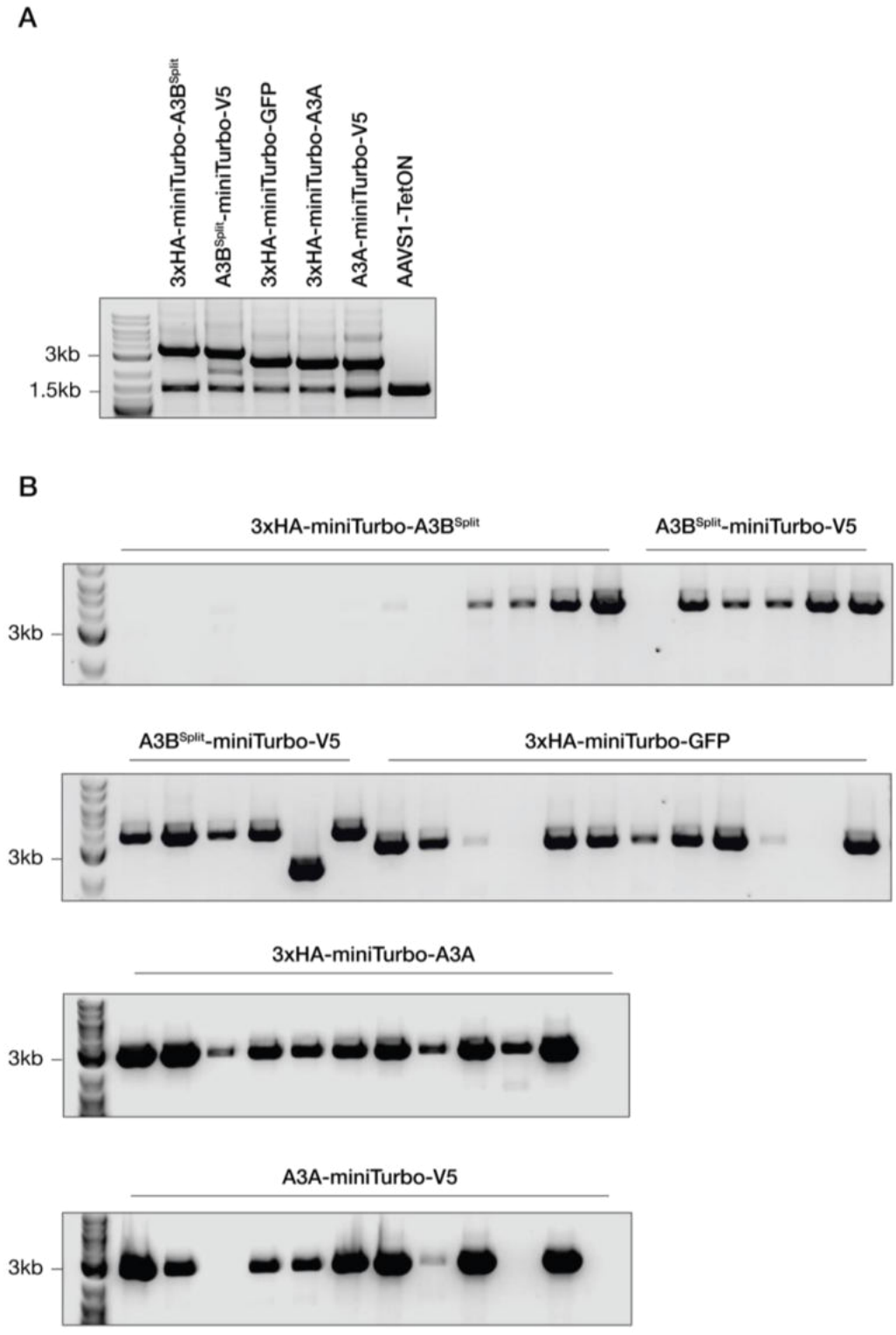


Figure 5.6 Subcloning miniTurbo-tagged A3A and A3B into the AAVS1_TetON vector.

(A) PCR with 0.5 µL of the NEB assembly reaction mixture shows successful assembly into AAVS1_TetON. (B) Colony PCR was performed on ampicillin-resistant colonies using vector-specific primers. Transformants had PCR products consistent with successful cloning (3xHA-miniTurbo- A3B^{Split} = 3391 bp; A3B^{Split}-miniTurbo-V5 = 3343 bp; 3xHA-miniTurbo-GFP = 2717 bp; 3xHA-miniTurbo-A3A = 2612 bp; A3A-miniTurbo-V5 = 2564 bp).

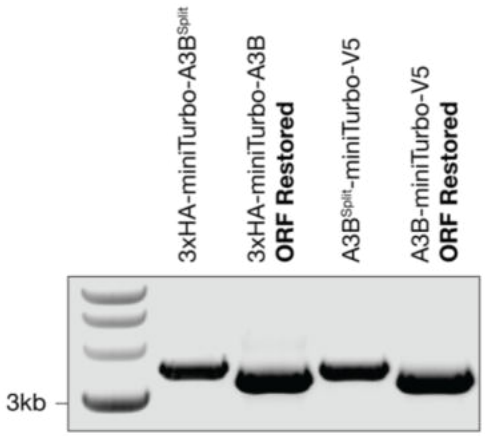


Figure 5.7 Successful stuffer removal and open reading frame restoration of A3B constructs.

3xHA-miniTurbo-A3B^{Split} and A3B^{Split}-miniTurbo-V5 plasmids were digested with BsrGI and ligated to remove the stuffer fragment and restore the *A3B* ORF. PCR on the undigested and digested/ligated plasmids show a size shift consistent with successful removal of the 230 bp stuffer fragment (3xHA-miniTurbo-A3B^{Split}, 3449 bp; 3xHA-miniTurbo-A3B, 3219 bp; A3B^{Split}-miniTurbo-V5, 3401 bp; A3B-miniTurbo-V5, 3171 bp).

5.2.4 A3A and A3B expression cell line models

To study the cellular response to elevated A3A and A3B expression in UCC cell lines, V5-tagged A3A and A3B constructs were generated from the miniTurbo-tagged A3A/B constructs. To enable the constructs to be used in the already puromycin resistant NHU-TERT B cell line, *Bsd^R* that confers resistance to blasticidin was assembled into A3B^{Split}-miniTurbo-V5 and A3A-miniTurbo-V5 plasmids to replace *Puro^R* using the NEBuilder HiFi DNA Assembly Method. The resulting plasmids were PCR linearised to remove miniTurbo and blunt end ligated to generate V5-tagged A3A and A3B^{Split} plasmids. To generate V5-tagged catalytically inactive A3A (A3A*-V5; E72A), the glutamic acid codon (GAG) at position 72 was mutated to encode an alanine (GCC) using the Q5® Site-Directed Mutagenesis Kit. V5-tagged catalytically inactive A3B (A3B**; E68Q/E255Q) was PCR amplified from the pTRIPZ_A3B**-GFP plasmid and assembled into the *Bsd^R* AAVS1-TetON backbone. All plasmids were maxi-prepped and sequenced prior to use; yields of the A3A-V5 plasmid were low suggesting some *E. coli* toxicity. The stuffer fragment was removed from the A3B^{Split} -V5 plasmid to restore the A3B open reading frame in an *in vitro* reaction

previously described (5.2.2). The key constructs made are summarised in Table 5.1.

Table 5.1 Key constructs used for making stable cell lines in this chapter.

Construct	Description
BioID	
AAVS1-TetON_3xHA-miniTurbo-A3B^{Split}	Stuffer is removed and plasmid is used to make cell lines expressing A3B N-terminally tagged with miniTurbo (mT-A3B) Confers puromycin resistance
AAVS1-TetON_A3B^{Split}-miniTurbo-V5	Stuffer is removed and plasmid is used to make cell lines expressing A3B C-terminally tagged with miniTurbo (A3B-mT) Confers puromycin resistance
AAVS1-TetON_3xHA-miniTurbo-A3A	Cell lines expressing A3A N-terminally tagged with miniTurbo (mT-A3A) Confers puromycin resistance
AAVS1-TetON_A3A-miniTurbo-V5	Cell lines expressing A3A C-terminally tagged with miniTurbo (A3A-mT) Confers puromycin resistance
A3A and A3B expression	
AAVS1_TetON_A3A-V5	Cell lines expressing wild-type A3A tagged with V5 (A3A-V5) Confers blasticidin resistance
AAVS1_TetON_A3A*-V5	Cell lines expressing catalytically inactive A3A (E72A) tagged with V5 (A3A*-V5) Confers blasticidin resistance
AAVS1_TetON_A3B^{Split}-V5	Stuffer is removed and plasmid is used to make cell lines expressing wild-type A3B terminally tagged with V5 (A3B-mT) Confers blasticidin resistance
AAVS1_TetON_A3B**-V5	Cell lines expressing catalytically inactive A3B (E68Q/E255Q) tagged with V5 (A3B**-V5) Confers blasticidin resistance

5.2.5 Preliminary *AAVS1* integration

Prior to the preliminary knock-in experiment, the genomic region surrounding the T2-guide site was PCR amplified and the product was sent for sequencing. This confirmed there are no single nucleotide polymorphisms (SNPs) in the guide site or surrounding region that could affect CRISPR/Cas9 cleavage. To determine the conditions required for integration of the transgene into the *AAVS1* locus, a

preliminary CRISPR/Cas9 integration experiment was performed in BFTC-905 and 5637 cells. These cell lines were chosen as they endogenously express both *A3A* and *A3B*. Cells were reverse transfected with T2-RNPs targeting the *AAVS1* locus to induce DSB formation and then forward transfected with the HDR-donor plasmid 18 hours later. Repair of the DSB by homologous recombination from the repair template integrates the transgene into the locus. Integration at the locus was assessed by PCR with a forward primer specific to a region outside the integration site and a reverse primer specific to the Puro^{R} cassette; amplification only occurs when there has been integration. Successful integration occurred in both cell lines transfected with T2-RNP and donor plasmid with no/minimal integration seen in cells transfected with control RNP (NT-RNP) and donor plasmid (Figure 5.8).

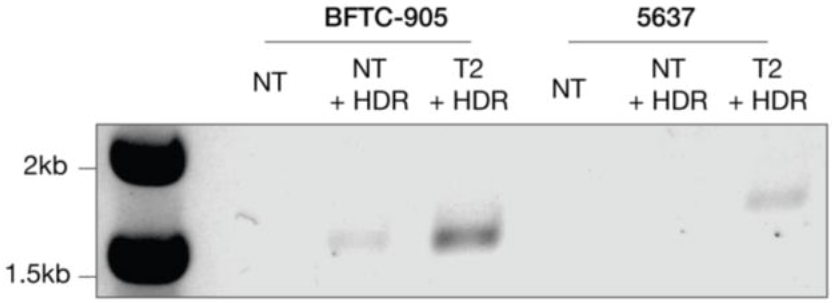


Figure 5.8 *AAVS1* knock-in by homology-directed repair from a donor plasmid.

PCR was performed using integration screening PCR primers on genomic DNA extracted from two cell lines, BFTC-905 and 5637, transfected with non-targeting, NT-RNP (tracrRNA only), NT-RNP with the AAVS1-TetON_3xHA-miniTurbo-GFP donor plasmid or the targeting, T2-RNP with the 3xHA-miniTurbo-GFP donor plasmid. When the T2-RNP is transfected with the donor plasmid, both cell line pools show integration of the transgene at the *AAVS1* locus (1521 bp).

5.2.6 Validation of A3A miniTurbo constructs

Prior to making stable cell lines, the two miniTurbo-tagged A3A constructs were validated in BFTC-905 cells by transient transfection. Cells were forward transfected and treated with doxycycline 24 hours later; this results in expression of both 3xHA-miniTurbo-A3A (mT-A3A) and A3A-miniTurbo-V5 (A3A-mT) with no expression seen in the absence of doxycycline. Addition of exogenous biotin for

2 hours stimulated biotinylation activity confirming the constructs are functional (Figure 5.9).

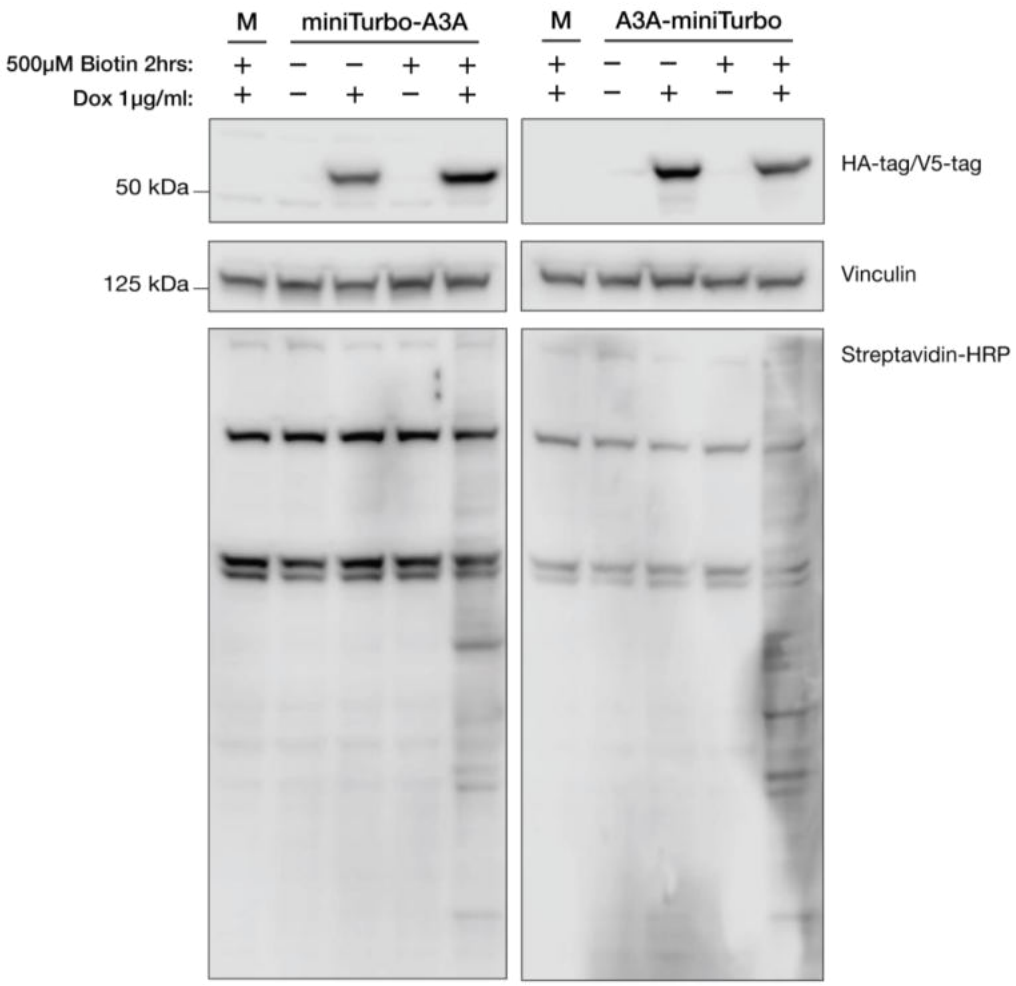


Figure 5.9 Doxycycline induces expression of miniTurbo-tagged A3A, and addition of biotin induces biotinylation of interacting partners.

Western blots showing induction of protein expression upon addition of doxycycline, and biotinylation activity upon addition of biotin. BFTC-905 cells were seeded in 6-well plates at 3 x 10^5 cells/well and transfected with 2.5 µg plasmid 24 hours later at 70% confluence. 1 µg/mL doxycycline was added 24 hours post-transfection and cells were harvested for western blot 48 hours later, after 2-hour treatment with 500 µM biotin. 30 µg protein was used for western blot; vinculin was used as a loading control, streptavidin-HRP blotting detects biotinylated proteins, and HA-tag or V5-tag blotting detects tagged protein expression. miniTurbo-A3A, cells transfected with AAVS1-TetON_3xHA-miniTurbo-A3A; A3A-miniTurbo, cells transfected with AAVS1-TetON_A3A-miniTurbo-V5; M, mock transfected cells, Lipofectamine3000 only. Blot representative of two biological repeats.

5.2.7 *AAVS1* targeting using plasmid donors

BFTC-905 cells were transfected with Cas9-RNPs targeting the *AAVS1* locus and one of the five plasmids containing miniTurbo-tagged A3A, A3B or GFP. 72 hours post-transfection, an aliquot of cells was taken for genomic DNA extraction and integration screening prior to selection with puromycin; this revealed that integration had occurred but at a very low frequency (Figure 5.10).

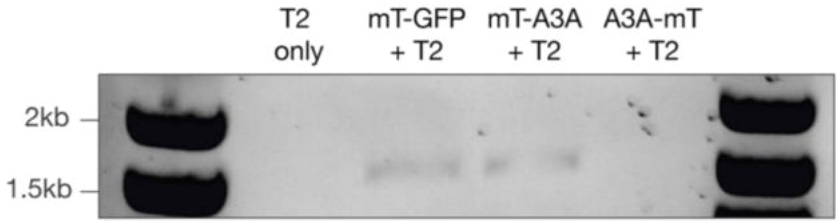


Figure 5.10 Pre-selection integration screening of stable cell line pools shows very low frequency of integration into the *AAVS1* locus.

PCR using integration screening primers was performed using genomic DNA extracted from BFTC-905 cells transfected with T2-guide RNPs and plasmid donor templates targeting the *AAVS1* locus. T2 only = T2-guide RNP only; mT-GFP + T2 = T2-guide RNP and AAVS1-TetON_3xHA-miniTurbo-GFP; mT-A3A + T2 = T2-guide RNP + AAVS1-TetON_3xHA-miniTurbo-A3A; A3A-mT + T2 = T2-guide RNP + AAVS1-TetON_A3A-miniTurbo-V5.

The pools were selected with puromycin to recover low frequency integration events and the resulting puromycin resistant pools were screened for transgene expression with a range of doxycycline concentrations. Expression of A3B-miniTurbo-V5 (A3B-mT) (Figure 5.11 A), mT-A3A (Figure 5.11 B), and A3A-mT (Figure 5.11 C), is seen and is doxycycline concentration dependent. However, no expression of 3xHA-miniTurbo-A3B (mT-A3B) or 3xHA-miniTurbo-GFP (mT-GFP) was observed (data not shown).

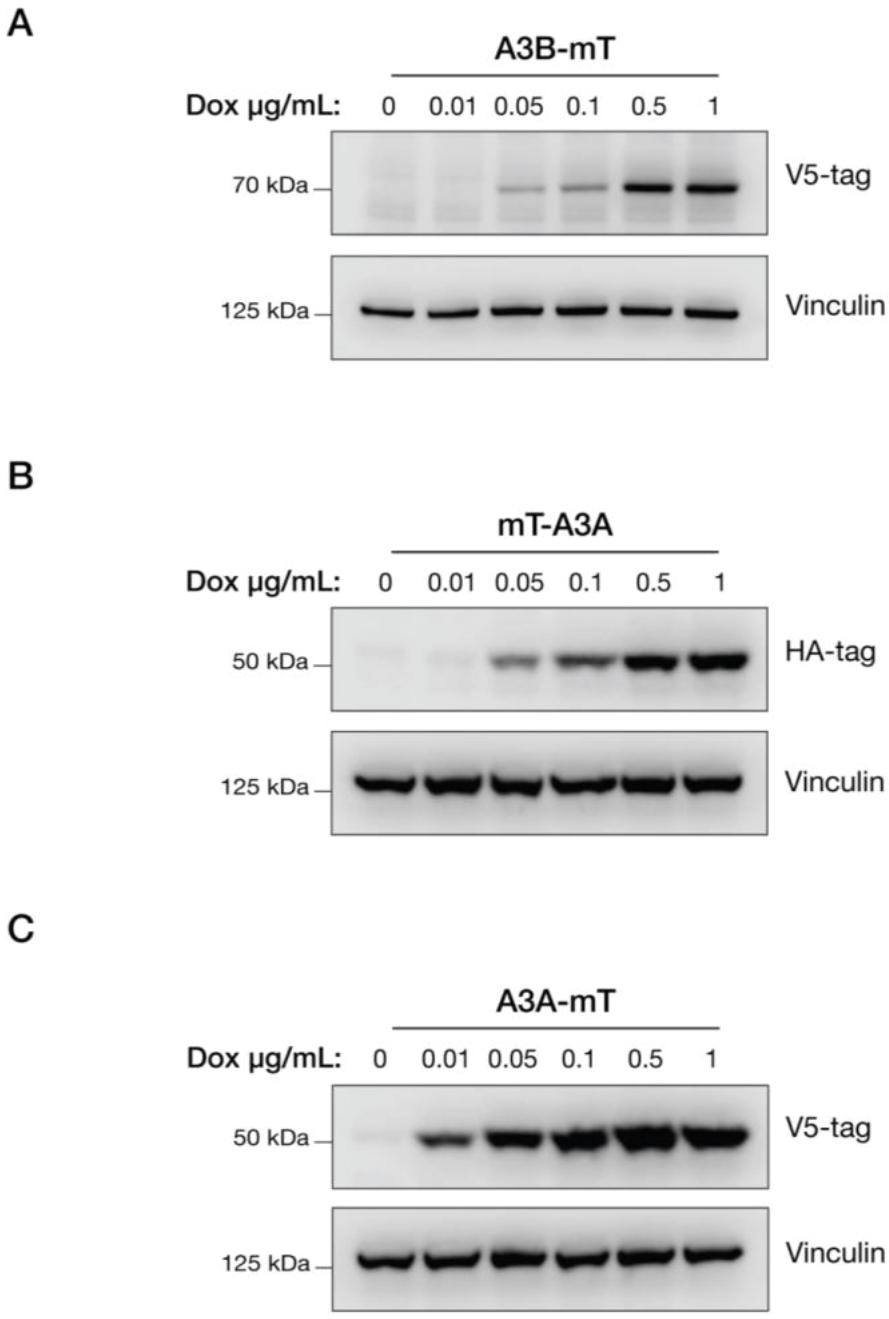


Figure 5.11 Expression of the A3B-miniTurbo-V5 (A), 3xHA-miniTurbo-A3A (B) and A3A-miniTurbo-V5 (C) is doxycycline concentration dependent.

The stable cell line pools were seeded in 6-well plates at 1.5 x 10^5 cells/well, the concentration of doxycycline (dox) indicated was added 24 hours later and cells were harvested after a further 48 hours for western blot. 30 µg protein was used for western blot; vinculin was used as a loading control, V5-tag and HA-tag blotting detects tagged protein expression. mT = miniTurbo. Blots representative of two biological repeats.

Post-selection PCR screening of the pools that induced expression in the presence of doxycycline showed that all cell lines had integrated the transgene. However, integration had occurred randomly (Figure 5.12 A) and not specifically into the *AAVS1* locus (Figure 5.12 B).

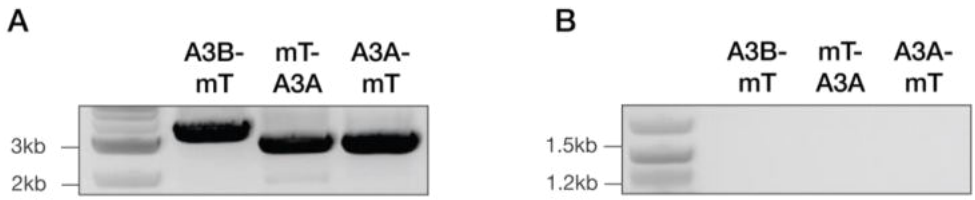


Figure 5.12 Stable cell lines made using plasmid DNA as a donor have randomly integrated the transgene into the genome and are not *AAVS1* knock-ins.

Genomic DNA from BFTC-905 stable cell line pools was extracted and used for transgene specific (A) and *AAVS1* knock-in detection (B) PCR using primers shown in Figure 5.1. The presence of a PCR product in A but not B shows the stable cell lines have randomly integrated the transgene and it has not been specifically integrated at the *AAVS1* locus. mT = miniTurbo.

To investigate the lack of expression of mT-A3B and mT-GFP in the stable cell line pools, parental BFTC-905 cells were transiently transfected with the HDR donor plasmids. Induction of mT-A3B occurs in response to doxycycline treatment suggesting that the lack of expression is due to promoter/gene disruption during integration (Figure 5.13). However, multiple attempts to make a stable cell line expressing mT-A3B were unsuccessful so the project was progressed without this cell line. Induction of mT-GFP did not occur in response to doxycycline treatment (Figure 5.13). Sanger Sequencing confirmed no mutational reason for the lack of doxycycline-inducibility. Multiple attempts to generate a functioning mT-GFP construct were unsuccessful, so the project was progressed without this control cell line.

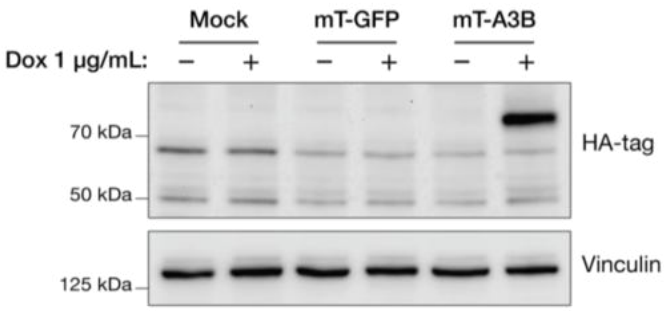


Figure 5.13 Doxycycline treatment induces expression of 3xHA-miniTurbo-A3B but not 3xHA-miniTurbo-GFP in transiently transfected BFTC-905 cells.

BFTC-905 cells were seeded in 6-well plates at 3 x 10^5 cells/well and transfected with 2.5 µg plasmid 24 hours later at 70% confluence. 1 µg/mL doxycycline was added 24 hours post-transfection and cells were harvested for western blot 48 hours later. 30 µg protein was used for western blot; vinculin was used as a loading control and anti-HA blotting detects fusion protein expression. mT-GFP, cells transfected with AAVS1-TetON_3xHA-miniTurbo-GFP; mT-A3A, cells transfected with AAVS1-TetON_3xHA-miniTurbo-A3B; mock, cells transfected with Lipofectamine3000 only. Blot representative of two biological repeats.

5.2.8 Optimisation of *AAVS1* integration

Using plasmid DNA as the homologous recombination template did not result in successful integration of the transgenes into the *AAVS1* locus. This is not unexpected as knock-in from plasmid templates occurs at very low frequency. Linear DNA templates, including dsDNA PCR products and ssDNA, containing relatively short homology arms have been shown to be more efficient at stimulating targeted knock-in. ssDNA in particular has been shown to be less toxic and reduce the risk of off-target integration events (Codner et al., 2018; Miura et al., 2015; Zhang et al., 2017) but producing sufficient amounts of long ssDNA for transfection can be problematic. To obtain *AAVS1* knock-in stable cell lines, both dsDNA in the form of PCR products and long ssDNA was used as a HDR template.

5.2.8.1 Making linear donors

dsDNA templates were generated by PCR from the plasmids described earlier in this chapter (see Table 5.1). The dsDNA templates were amplified using primers that resulted in a product containing homology arms of ~300 bp, both primers

also had two phosphorothioate bonds at the 5' end to provide protection against exonuclease degradation and enhance stability. Previous studies have demonstrated this homology arm length is sufficient (Zhang et al., 2017) and it ensures that the full PCR products are no larger than ~5 kb such that they can be used to generate long ssDNA templates using the TakaraBio Guide-it Long ssDNA Production System v2. This system makes long ssDNA templates via specific degradation of the DNA strand amplified with a phosphorylated primer. Amplification with a phosphorylated forward primer generates antisense ssDNA while a phosphorylated reverse primer generates sense ssDNA. Generating both sense and antisense ssDNA is recommended to determine which primer generates the cleanest product and highest yield. Optimisation was performed using the longest of the templates (3xHA-A3B-miniTurbo; 5 kb). Reactions were run on a 2% TBE agarose gel to determine if strand degradation was successful; ssDNA runs approximately half the size of its dsDNA template. While the PCR with both primer sets was successful, ssDNA was only efficiently made when using a phosphorylated forward primer to make antisense ssDNA. When using the phosphorylated reverse primer, degradation was incomplete (Figure 5.14).

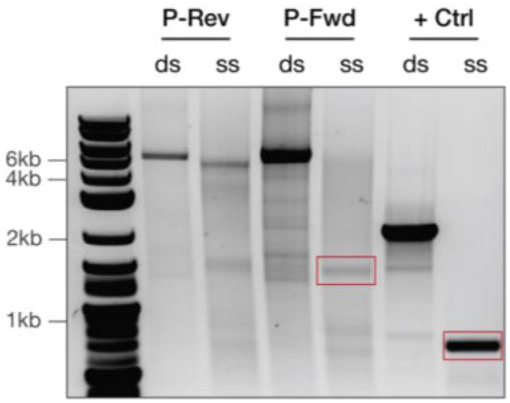


Figure 5.14 Optimisation of long ssDNA production using the TakaraBio Guide-it Long ssDNA Production System v2.

Long ssDNA is efficiently produced when using a phosphorylated forward primer to generate antisense ssDNA. Degradation of the phosphorylated strand is incomplete when using a phosphorylated reverse primer. ssDNA is highlighted with the red boxes. P-Rev = phosphorylated reverse primer; P-Fwd = phosphorylated forward primer.

PCR products for all constructs were made using a phosphorylated forward primer and sequenced prior to ssDNA production. Long ssDNA was successfully made and purified for all transgenes (Figure 5.15).

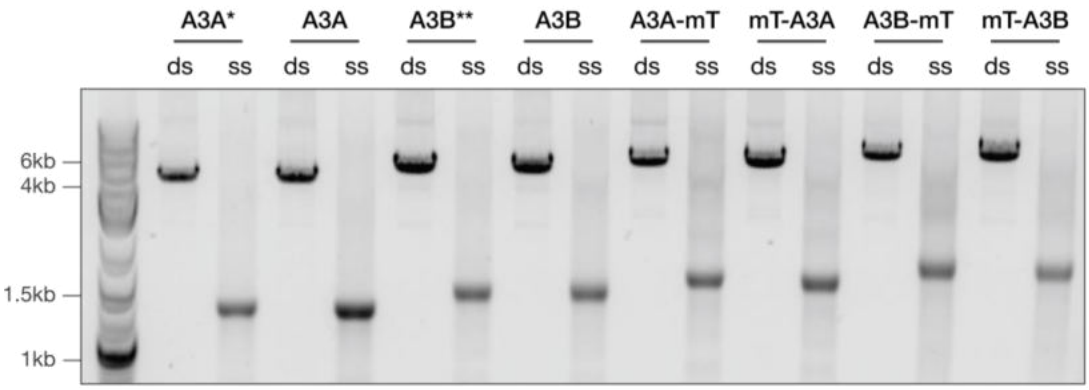


Figure 5.15 dsDNA and ssDNA transgene templates for transfection.

dsDNA (ds) PCR and ssDNA (ss) templates for transfection were made and purified for making stable cell lines. The agarose gel shows complete conversion to ssDNA for all constructs. A3A*, A3A*-V5; A3A, A3A-V5; A3B**, A3B**-V5; A3B, A3B-V5; A3A-mT, A3A-miniTurbo-V5; mT-A3A, 3xHA-miniTurbo-A3A; A3B-V5, A3B-miniTurbo-V5; mT-A3B, 3xHA-miniTurbo-A3B.

5.2.8.2 *AAVS1* integration with linear donors

BFTC-905 cells were co-transfected with T2-RNPs targeting the *AAVS1* locus and a HDR template as either dsDNA or long ssDNA as described in Materials and methods 2.27. 48 hours post-transfection cells were selected with the appropriate concentration of puromycin or blasticidin. The resistant pools were then screened for transgene integration, both random and *AAVS1*-targeted, by PCR using matched amount of genomic DNA such that a qualitative comparison of integration frequency could be seen; a stronger band is indicative of more copies of the target in the starting material. SW780, RT4 (models for NMI-UCC) and the 'normal' immortalised cell line, NHU-TERT B, were also transfected with the constructs using these methods, but either resistant pools were not recovered, or transgene integration was not observed. Despite optimisation of conditions, toxicity and expanding cells from low numbers was a major issue and no further work was conducted.

5.2.8.2.1 Characterisation of stable cell line pools made with ssDNA HDR templates

AAVS1-targeted integration only occurred in one expression (A3A-V5) and two BioID (mT-A3A and mT-A3B) pools transfected with long ssDNA HDR templates (Figure 5.16). Random transgene integration was seen in all the ssDNA stable pools, but the product was very faint in the four A3 expression and two of the BioID stables (A3B-mT and mT-A3B) suggesting only a small percentage of the cells had integrated the full-length transgene. It is likely that the ssDNA was partially degraded prior to integration resulting in the appearance of *AAVS1* knock-in but lack of full transgene detection. Due to the lack of targeted *AAVS1* knock-in and suggestion of low transgene integration in the pools derived from ssDNA HDR templates, these pools were not taken forward for any further screening or work.

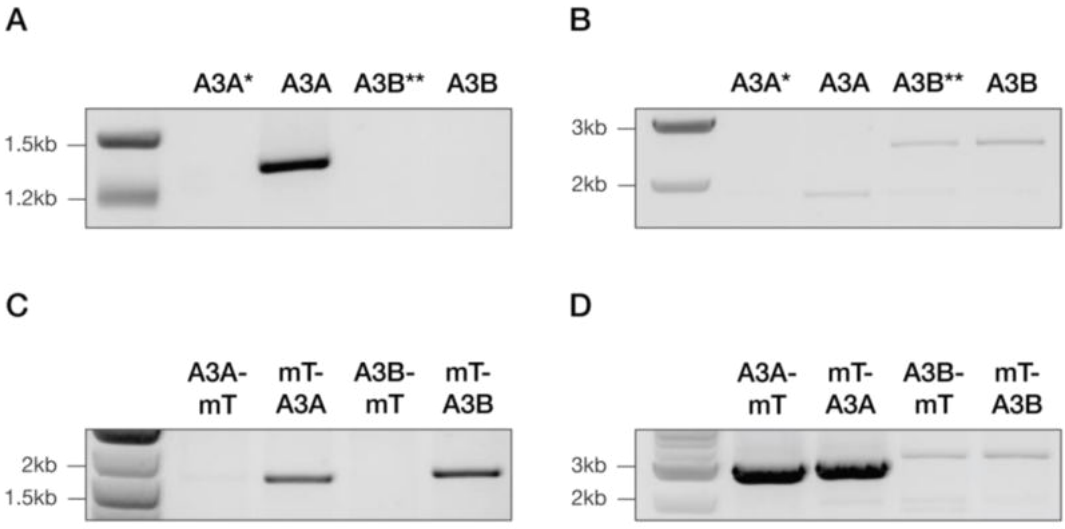


Figure 5.16 PCR screening of BFTC-905 stable cell lines made with ssDNA HDR templates.

(A and C) PCR detecting *AAVS1*-integration using primers described in Figure 5.1. *AAVS1* integration is seen in one expression, and two BioID stable cell lines. (B and D) Transgene specific PCR detecting full-length integration into the genome. The faint bands suggest low levels of random integration of the full-length transgene. A3A*, catalytically inactive A3A*-V5; A3A, A3A-V5; A3B**, catalytically inactive A3B**-V5; A3B, A3B-V5; A3A-mT, A3A-miniTurbo-V5; mT-A3A, 3xHA-miniTurbo-A3A; A3B-V5, A3B-miniTurbo-V5; mT-A3B, 3xHA-miniTurbo-A3B.

5.2.8.2.2 Characterisation of stable cell line pools made with dsDNA HDR templates

Surprisingly, PCR screening showed that *AAVS1* integration when using dsDNA as a HDR template was more successful than when using ssDNA (Figure 5.17). *AAVS1* knock-in detected by PCR occurred in all expression pools derived from dsDNA templates and the full-length transgene was also detected. While the specific *AAVS1* screening PCR revealed knock-in at the locus, there is a small chance that cells containing *AAVS1* integrations may also contain the transgene integrated at a random locus. The only way to fully characterise all integration sites is to perform Southern Blotting; due to time constraints this was not performed for the work described in this thesis and the pool was used for all further experiments.

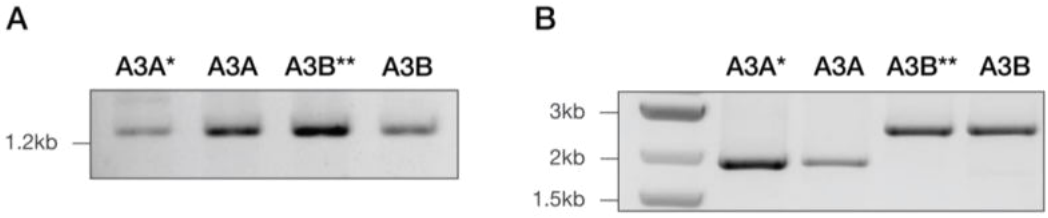


Figure 5.17 *AAVS1* integration of the V5 tagged A3A and A3B transgenes is seen in all BFTC-905 stable cell line pools made with dsDNA HDR templates.

(A) PCR detecting *AAVS1*-integration screening using primers described in Figure 5.1**.** Integration at the *AAVS1* locus is seen in all stable cell line pools. (B) Transgene specific PCR detecting full length integration into the genome. PCR was performed on matched amounts of genomic DNA. A3A* = catalytically inactive A3A*-V5; A3A, A3A-V5; A3B**, A3B**-V5; A3B, A3B-V5.

AAVS1 integration was less successful with the miniTurbo-tagged constructs than the V5 tagged A3A/A3B expression constructs, most likely due to their larger size (Figure 5.18). Targeted *AAVS1* integration only occurred in one pool (mT-A3A), but the full-length transgene was detected in all pools to differing degrees.

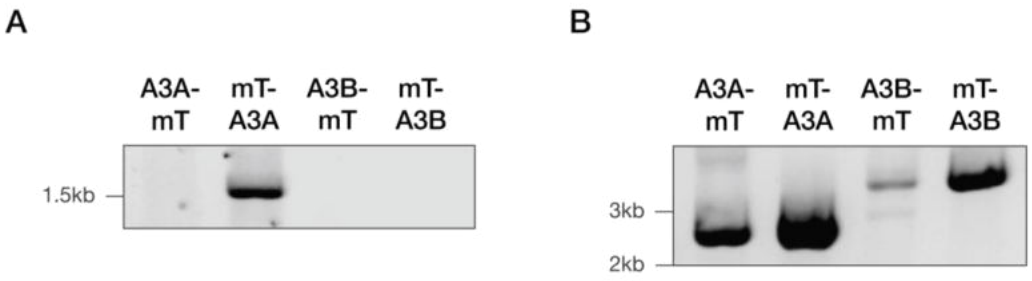


Figure 5.18 *AAVS1* integration of the miniTurbo-tagged transgenes is seen in only one BFTC-905 stable cell line pool made with dsDNA HDR templates.

(A) PCR detecting *AAVS1*-integration screening using primers described in Figure 5.1**.** Integration at the *AAVS1* locus is seen in one stable cell line pool (3xHA-miniTurbo-A3A). (B) Transgene specific PCR detecting full length integration into the genome. PCR was performed on matched amounts of genomic DNA. A3A-mT, A3A-miniTurbo-V5; mT-A3A, 3xHA-miniTurbo-A3A; A3B-V5, A3B-miniTurbo-V5; mT-A3B, 3xHA-miniTurbo-A3B.

Despite the eight stable cell lines made with dsDNA HDR templates showing integration of the transgene (either *AAVS1* knock-in or random), after 48 hours treatment with doxycycline, induction of the tagged protein was only seen in one BioID (mT-A3A) cell line; no expression was seen in the other seven lines (Figure 5.19). Due to the lack of specific *AAVS1* integration and protein expression in these BioID stable cell line pools, no further work was conducted; the three doxycycline-inducible randomly integrated cell lines characterised earlier in this chapter will be used for all BioID experiments (Figure 5.11).

The band from the transgene detection PCR for the one cell line that expressed mT-A3A with doxycycline treatment was much stronger than that of the other cell lines. This suggests that the lack of expression seen in response to doxycycline could be a result of low copy number or a low percentage of cells in the pool having integrated the full transgene.

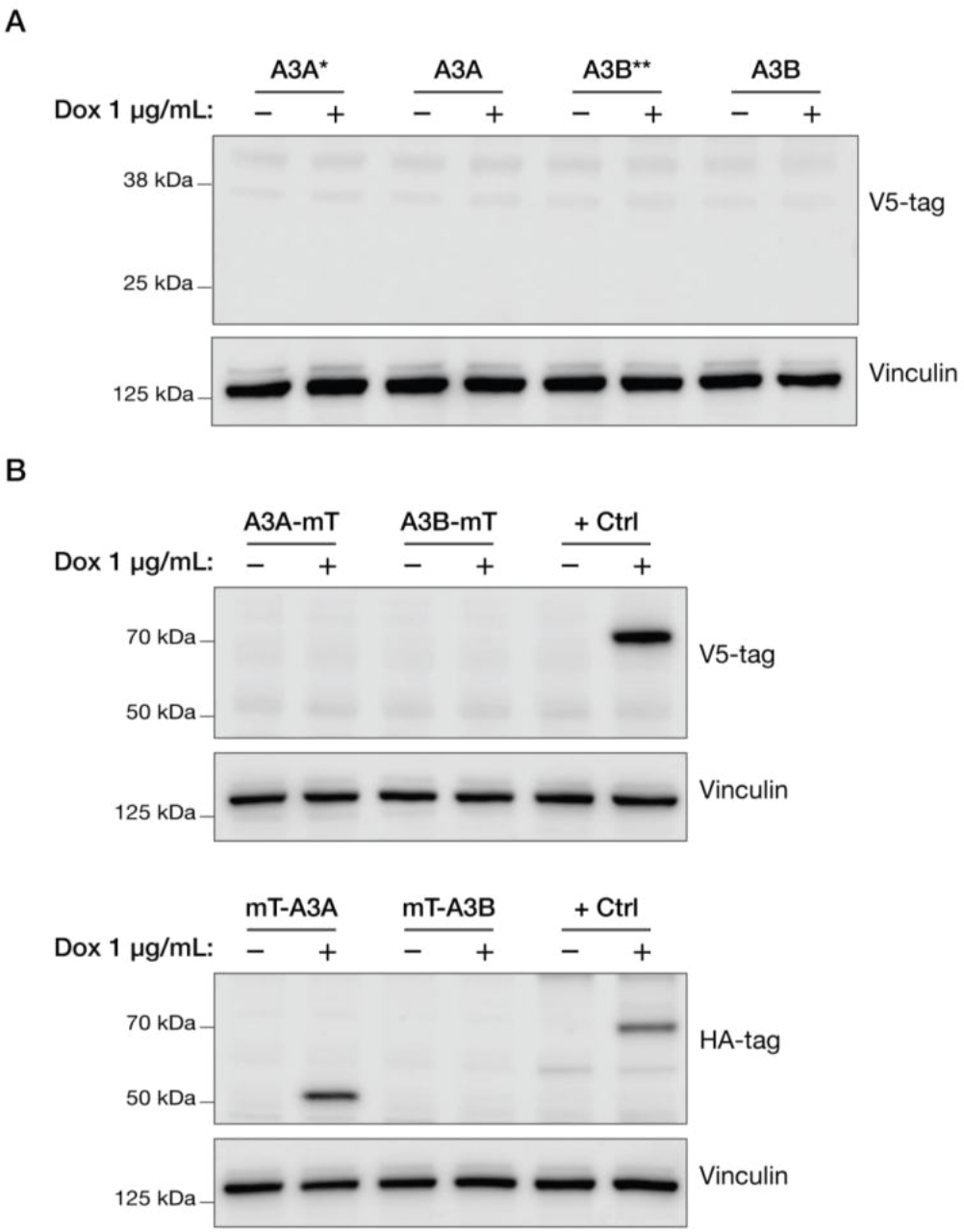


Figure 5.19 One AAVS1-BFTC-905 stable cell line pool made with dsDNA HDR donor templates expresses after induction with doxycycline for 48 hours.

The stable cell line pools were seeded in 6-well plates at 1.5 x 10^5 cells/well, doxycycline (dox) was added 24 hours later, and cells were harvested after a further 48 hours for western blotting. 30 µg protein was used for western blot; vinculin was used as a loading control, V5-tag and HA-tag blotting detects tagged protein expression. (A) Protein expression is not seen in any of the A3A and A3B expressing stable cell lines after doxycycline induction. (B) One BioID stable cell line (mT-A3A) expresses the tagged protein after doxycycline induction. A3A*, catalytically inactive A3A*-V5; A3A, A3A-V5; A3B**, catalytically inactive A3B**-V5; A3B, A3B-V5; A3A-mT, A3A-miniTurbo-V5; mT-A3A, 3xHA-miniTurbo-A3A; A3B-V5, A3B-miniTurbo-V5; mT-A3B, 3xHA-

miniTurbo-A3B. + Ctrl = BFTC-905 cells transiently transfected with either A3B-mT or mT-A3B. Blots representative of two biological repeats.

5.2.9 Stable cell line troubleshooting

It was unexpected that the V5 tagged A3A and A3B stable cell lines were not expressing the tagged proteins in the presence of doxycycline despite having integrated the full transgenes. The transgenes were PCR amplified from genomic DNA and sent for Sanger sequencing, which confirmed the transgene sequence for all stables was correct and the lack of expression was not due to mutation.

5.2.9.1 The lack of expression seen in the stable cell lines is not due to epigenetic silencing of the promoter

A major issue with generating stable cell lines is lack of expression of the protein of interest because of epigenetic promoter silencing. The silencing process is usually initiated by histone deacetylation and can be reversed through treatment with histone deacetylase inhibitors (HDACi) (Chen et al., 1997; Oyer et al., 2009; Yu et al., 2016). To investigate whether the lack of expression was occurring due to epigenetic silencing of the transgene promoter, the A3A and A3B expressing cell lines were treated with the HDAC inhibitors, sodium butyrate and panobinostat, in the presence of doxycycline for 72 hours. There was no expression of the V5 tagged proteins in any of the stable cell line pools when treated with pan-HDACi in the presence of doxycycline, demonstrating that the lack of expression is not due to epigenetic silencing of the promoter region (Figure 5.20).

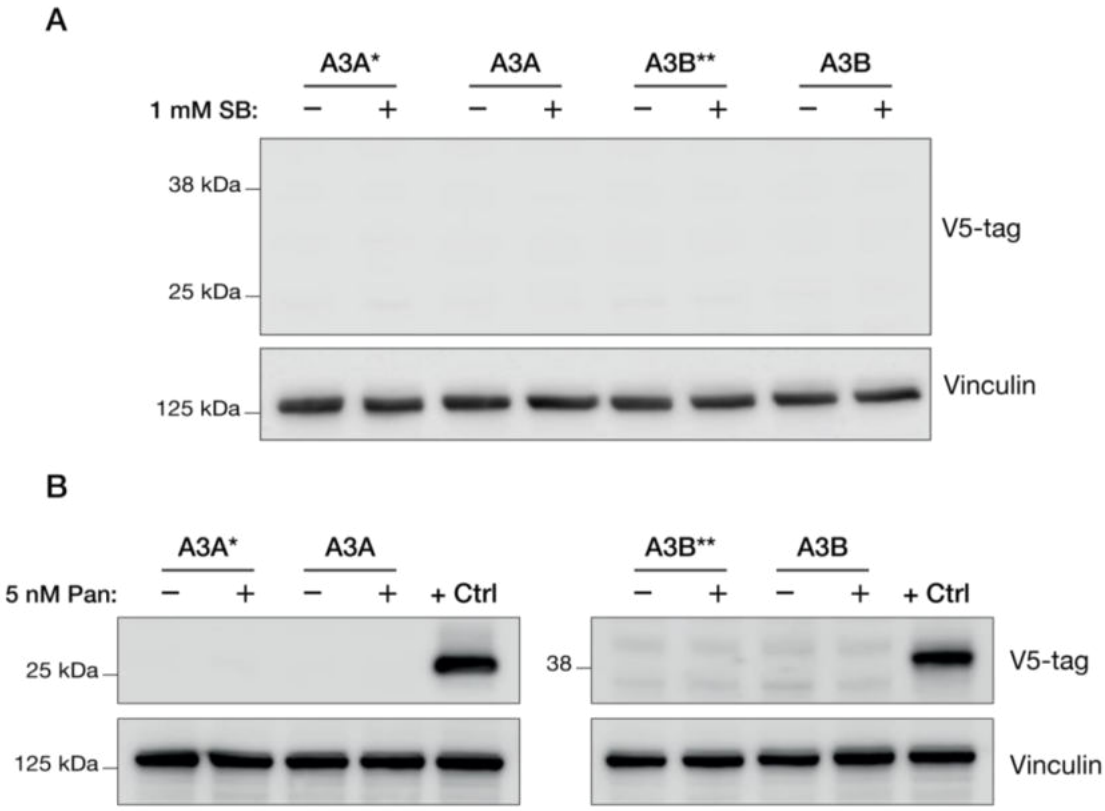


Figure 5.20 The lack of expression seen in the *AAVS1*-BFTC-905 stable cell lines is not due to epigenetic silencing of the transgene promoter.

The stable cell lines were seeded in 6-well plates at 1.5 x 10^5 cells/well and cells were treated with either 1 mM sodium butyrate (SB) (A) or 5 nM panobinostat (B) for 72 hours after which cells were harvested for western blot. Doxycycline was present in all conditions throughout the experiment at 1 µg/mL. The concentration of SB used was the same used in Yu et al., (2016); the panobinostat concentration used was recommended by a colleague (LeAnne Carmichael). + Ctrl = lysates from BFTC-905 cells transiently transfected with either A3A-V5 or A3B-V5 plasmids. Blots representative of two biological repeats.

5.2.9.2 Exogenous supply of the Tet3G transactivator protein induces expression of the fusion proteins in the presence of doxycycline

With autoregulated, bidirectional Tet-on expression systems, expression of both the transactivator and the gene of interest (GOI) occur from a bidirectional tetracycline responsive promoter. Baseline expression of the transactivator is required to initiate the positive feedback loop and induce doxycycline concentration-dependent expression of the transgene (Baron et al., 1995; Markusic et al., 2005; Paulus et al., 1996). Therefore, the ability to initiate the

positive feedback loop is dependent on both the cellular abundance of the transactivator protein and the availability of its ligand, doxycycline. Heinz et al., (2013) showed that lack of transgene expression in a subset of cells that had integrated an autoregulated Tet-on expression system could be overcome through increasing the cellular abundance of the transactivator. The authors suggested that baseline activity of the promoter is dependent on the genomic integration site and therefore, only integration loci with sufficient basal promoter activity can initiate the positive feedback loop and induce transgene expression in the presence of doxycycline. To examine this, cells were transiently transfected with a plasmid containing only the Tet3G transactivator to kick start the positive feedback loop. Transient transfection of the transactivator plasmid followed by doxycycline treatment induces concentration-dependent expression of all tagged proteins (Figure 5.21). This demonstrates that the lack of expression is due to insufficient basal levels of transactivator to initiate the positive feedback loop. This could be due to the genomic landscape of the integration site, or in the context of *AAVS1*-targeted integration where the maximum theoretical number of integration events and copy number per cell is two, a result of low transgene copy number.

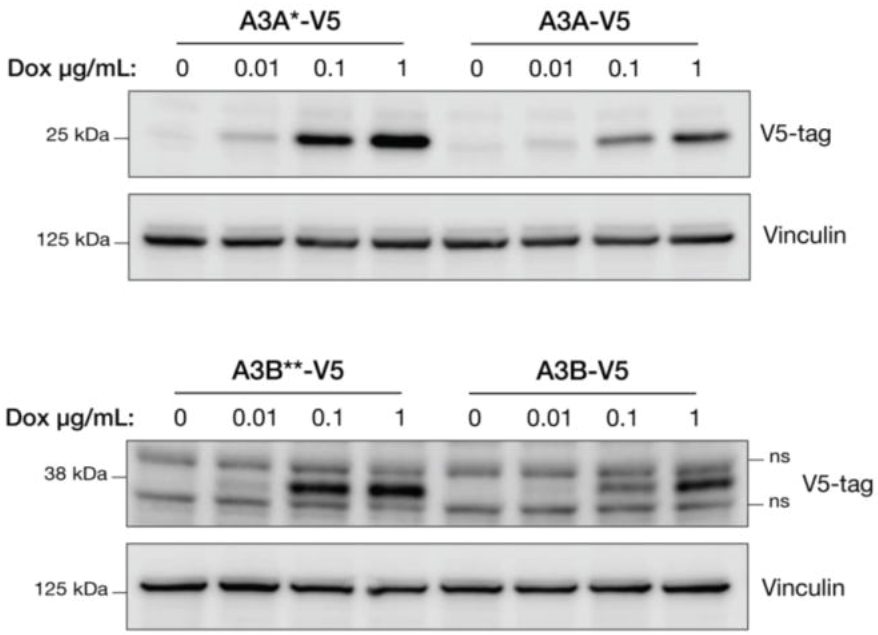


Figure 5.21 Exogenous supply of the Tet3G transactivator protein via transient transfection followed with doxycycline treatment induces protein expression in *AAVS1*-BFTC-905 stable cell lines.

BFTC-905 cells were seeded in T-12.5cm flasks at 5 x 10^5 cells/flask and transfected with 3.25 µg Tet3G-only plasmid 24 hours later at 60% confluence. Transfection complexes were removed 6 hours post-transfection and 24 hours later cells were re-seeded into media containing the indicated amounts of doxycycline (dox) at 1 x 10^5 cells/well in 6-well plates. Cells were harvested 72 hours later, and 30 µg protein was used for western blot; vinculin was used as a loading control, V5-tag blotting detects tagged protein expression. Blots representative of two biological repeats.

5.2.10 Randomly integrated stables for further studies

Due to time constraints, the *AAVS1* stable cell lines that require transient transfection with the Tet3G transactivator were used for the RNAseq experiments. While waiting for sequencing and data to be returned and analysed, stable cell lines that do not require transient transfection were made. To generate both basal and luminal subtype cell lines expressing A3A and A3B, BFTC-905 and RT-112 cells were transfected with the expression plasmids (A3A, A3A*, A3B and A3B**), and selected with blasticidin until resistant pools were obtained. The BFTC-905 A3B-V5 (BFTC_A3B) pool expressed the protein upon the addition of doxycycline without requiring transient transfection (Figure 5.22 A) but expression of A3B**-V5 was barely detectable (Figure 5.22 B). No expression

was seen upon the addition of doxycycline in any of the other blasticidin resistant BFTC-905 or RT-112 pools (data not shown).

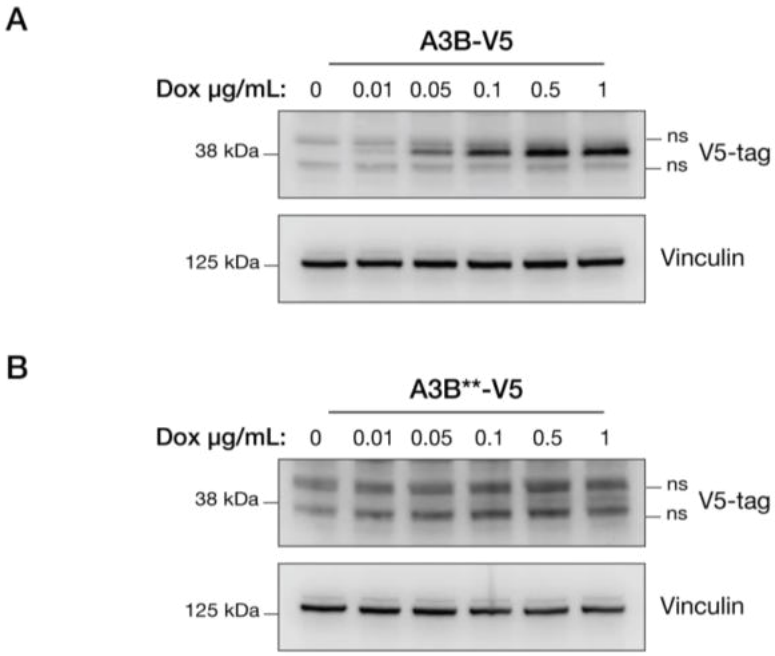


Figure 5.22 Expression of V5-tagged A3B is doxycycline concentration-dependent in BFTC_A3B cells without transient Tet3G transfection.

BFTC_A3B and BFTC_A3B** cells were seeded in 6-well plates at 1.5 x 10^5 cells/well, the concentration of doxycycline (dox) indicated was added 24 hours later and cells were harvested after a further 48 hours for western blot. 30 µg protein was used for western blot; vinculin was used as a loading control, V5-tag blotting detects tagged protein expression. (B) Expression of A3B**-V5 is barely detectable in randomly integrated blasticidin resistant BFTC_A3B** cells.

The NHU-TERT B cell line is already puromycin and neomycin resistant, so to make stables using this line, *Bsd*^{R} conferring resistance to the selectable marker, blasticidin, was originally used in the A3 expression constructs. A recent study has shown that the choice of selectable marker can affect protein expression levels in stable cell line pools (Guo et al., 2021). The authors found that pools selected with blasticidin showed the lowest level of protein expression and displayed high cell-to-cell variability, while selection with puromycin (*Puro*^{R}*)* resulted in much higher expression and less variability within the pool. This may explain why despite recovering eight blasticidin resistant pools, substantial expression of the protein of interest was only seen in one, while the others had

barely detectable or no expression. Due to concerns about using *Bsd*^{R} as a selection marker and the need for an A3A expressing BFTC-905 stable cell line that does not require transient transfection, the selectable marker was reverted to *Puro*^{R}. BFTC-905 cells were transfected and selected in puromycin until a resistant pool was obtained and the BFTC_A3A line was confirmed to express A3A-V5 in the presence of doxycycline (Figure 5.23).

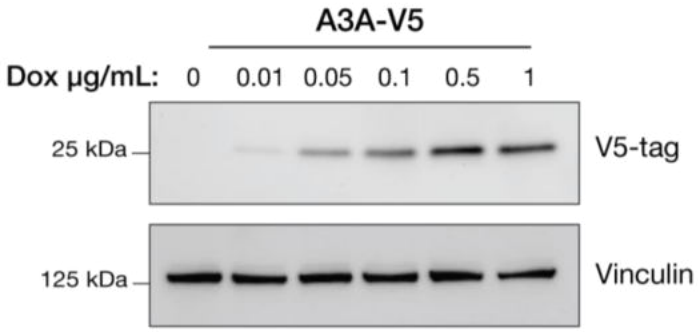


Figure 5.23 Expression of V5-tagged A3A is doxycycline concentration-dependent in BFTC_A3A cells without transient Tet3G transfection.

Puromycin resistant BFTC_A3A cells were seeded in 6-well plates at 1.5 x 10^5 cells/well, the concentration of doxycycline (dox) indicated was added 24 hours later and cells were harvested after a further 48 hours for western blot. 30 µg protein was used for western blot; vinculin was used as a loading control, V5-tag blotting detects tagged protein expression.

5.3 Discussion

This chapter describes work generating a set of stable cell line models that express A3A and A3B tagged with the small epitope, V5, for RNA sequencing studies and to evaluate the cellular response to elevated expression. Stable cell lines that express A3A and A3B tagged with the biotin ligase mutant, miniTurbo, were also made to enable identification of interacting partners using the BioID method. The cloning process was challenging and time consuming as A3B toxicity in *E. coli* was a major issue that had to be overcome. This has been reported in the literature previously and most have overcome the toxicity issues through insertion of an intronic sequence into the ORF that prevents production of A3B in *E. coli*, while facilitating expression in mammalian cells (Akre et al., 2016; Hultquist et al., 2011; St. Martin et al., 2018). However, this technique is not optimal for use with the *AAVS1*-targeting system used in this work. The puromycin resistance cassette is expressed from the endogenous *PPP1R12C* promoter, and therefore, a splice acceptor site is already present in the construct and any transcriptional read through from the puromycin cassette could generate aberrant transcripts. Instead, the toxicity issues were overcome by disrupting the ORF with a stuffer fragment inserted into a native restriction site that can be removed prior to transfection in an *in vitro* digestion and ligation reaction (Krela et al., 2019). This technique was very effective at overcoming A3B toxicity, even with the technical limitation requiring a unique native restriction site within the cDNA. This method is also more labour intensive as it requires additional plasmid processing and purification prior to use and scaling up to obtain large amounts of plasmid to work with is difficult. While A3A was not so toxic to *E. coli* as to prevent cloning, yields were low, and multiple plasmids showed evidence of A3A deamination events. Therefore, plasmid sequencing at every step was essential and this again lengthened the cloning process.

Due to concerns over using viral transduction methods, integration of the transgenes into the *AAVS1* genomic safe harbour locus was attempted. Initial attempts to generate *AAVS1* knock-ins using plasmid DNA were unsuccessful. Three of the five (mT-A3A, A3A-mT and A3B-mT) puromycin resistant pools expressed the proteins upon induction with doxycycline but PCR screening revealed that the transgenes had been integrated randomly rather than into the *AAVS1* locus. The GFP-mT and mT-A3B pools did not express the tagged

proteins. Transient transfections with the constructs revealed that the mT-A3B protein was expressed in the presence of doxycycline suggesting the lack of expression seen in the stable cell line pool was due to issues during integration, likely linearisation of the plasmid at a key region. The mT-GFP protein was not expressed in response to doxycycline suggesting an issue with the plasmid itself. Multiple attempts to make a functional mT-GFP plasmid were unsuccessful, and the project was progressed without this construct. While the lack of a GFP control for the BioID experiments has its limitations as the ability to identify false positives due to non-specific interactions with miniTurbo is reduced, it does mean that any real interacting proteins that may have been excluded due to false interactions with GFP will not be excluded from analysis.

Integration into the *AAVS1* locus using plasmid DNA as the HDR template was initially unsuccessful and there are several reports in the literature showing that that HDR integration is more efficient using linear DNA substrates, with ssDNA being more efficient than dsDNA. Accordingly, the transgenes were provided as dsDNA and ssDNA. Interestingly, integration was more efficient when delivering the HDR template as linear dsDNA than ssDNA. *AAVS1* integration was seen in five of the eight stable cell line pools (A3A*-V5, A3A-V5, A3B**-V5, A3B-V5 and mT-A3A) generated with dsDNA templates in comparison to three of the eight (A3A-V5, mT-A3A and mT-A3B) made with ssDNA templates. This may be due to the difficulty in generating long ssDNAs, reduced stability and degradation of ssDNA or, as it is the substrate for A3 deaminases, A3-mediated deamination.

The stable cell line pools generated using dsDNA templates were screened for expression of the proteins in the presence of doxycycline and unexpectedly, the proteins were not expressed. Due to the lack of integration and expression, no further work was conducted with the *AAVS1* BioID cell lines, and the experiment was progressed using the three cell lines expressing mT-A3A, A3A-mT and A3B-mT, that had randomly integrated the transgenes into the genome. Multiple attempts were made to generate the N-terminally tagged A3B cell line, but these were unsuccessful and disappointingly, the work had to progress without it. This limited the study as interacting partners may not be identified if C-terminal tagging prevents interactions.

To investigate why the four V5 tagged A3A and A3B *AASV1*-knockin cell lines were not expressing the proteins upon induction with doxycycline, the transgene was amplified and sent for Sanger sequencing which showed that the sequence was correct, ruling out mutation as a cause of the lack of expression. Treatment of these cell lines with two HDACi revealed that the lack of expression was also not due to epigenetic silencing of the transgenes.

The doxycycline inducible system relies on the presence of a transactivator protein to bind doxycycline, undergo a conformational change and bind to the promoter to initiate gene expression (Gossen & Bujard, 1992; Gossen et al., 1995). The system used in this study is composed of an autoregulated bidirectional promoter that drives expression of both the Tet3G transactivator and the GOI (Dalvai et al., 2015). This ensures tight regulation and essentially no expression in the absence of doxycycline, which is ideal when working with A3s. However, a small amount of promoter leakiness is required to produce constitutive low levels of the transactivator for the positive feedback loop to be initiated upon the addition of doxycycline. Lack of transgene expression driven from an autoregulated doxycycline-inducible promoter in a subset of cells can be overcome through increasing the cellular abundance of the transactivator protein (Heinz et al., 2013). The authors of this study suggested that this was because the basal activity of the promoter is dependent on the integration locus, and only loci permitting sufficient constitutive expression of the transactivator to initiate the positive feedback loop will induce transgene expression in the presence of doxycycline. Indeed, transient transfection with a Tet3G transactivator expressing plasmid and subsequent doxycycline treatment induced protein expression in the four *AAVS1* knock-in A3A/B expressing cell lines. This suggests that the *AAVS1* integration locus in BFTC-905 cells does not facilitate induction of the system due to insufficient baseline transactivator expression. Transiently transfecting the stable cell lines prior to doxycycline induction has both pros and cons. The system is very tightly regulated with no leakiness in the absence of doxycycline and as the transgene is stably integrated, the induction seen in response to doxycycline is reproducible and eliminates variation arising because of transfection efficiencies. However, experiments requiring transient transfection prior to doxycycline-induction are more costly, labour intensive and complicate experiments due to the requirement for additional controls. Due to time

constraints, the RNA sequencing experiment was progressed with these cell lines.

While the RNA sequencing experiment and data analysis was ongoing, stable cell lines that do not require transient transfection and are easier to work with were made for further studies. BFTC-905 and RT-112 blasticidin resistant pools were recovered, but disappointingly, only one of the lines (BFTC_A3B) expressed the protein, with the other seven showing no/minimal expression. A recent publication has demonstrated that the selectable marker used can affect protein expression levels in stable cell lines, with *Bsd*^{R} conferring resistance to blasticidin, showing the lowest level of protein expression and high cell-to-cell variability (Guo et al., 2021), offering a potential explanation as to why only one line had detectable expression. In a final attempt to obtain an A3A-V5 expressing cell line that does not require transient transfection with transactivator, *Bsd*^{R} was replaced with *Puro*^{R}, and the A3A-V5 expressing BFTC_A3A line was made.

In conclusion, while the process of generating suitable stable cell line models was extremely time consuming and several unexpected obstacles had to be overcome, usable cell line models were made (Table 5.2). The A3A and A3B expression constructs were successfully integrated into the *AAVS1* locus, cell lines that express A3A and A3B without Tet3G transactivator supplement were made, and miniTurbo tagged A3A and A3B expressing lines can be used for BioID identification of interacting partners.

Table 5.2 Summary of stable cell lines that will be used for further work in this thesis.

BFTC, BFTC-905 UCC cell line; Dox, doxycycline.

Cell line	Description
BioID	
BFTC_A3B-mT	Dox-inducible expression of A3B C-terminally tagged with miniTurbo (A3B-mT)
BFTC_mT-A3A	Dox-inducible expression of A3A N-terminally tagged with miniTurbo (mT-A3A)
BFTC_A3A-mT	Dox-inducible expression of A3A C-terminally tagged with miniTurbo (A3A-mT)
A3A and A3B expression	
BFTC-AAVS1_A3A	Requires Tet3G transactivator transfection for dox-inducible expression of wild-type A3A tagged with V5 (A3A-V5) from the *AAVS1* locus
BFTC-AAVS1_A3A*	Requires Tet3G transactivator transfection for dox-inducible expression of catalytically inactive A3A (E72A) tagged with V5 (A3A*-V5) from the *AAVS1* locus
BFTC-AAVS1_A3B	Requires Tet3G transactivator transfection for dox-inducible expression of wild-type A3B tagged with V5 (A3B-V5) from the *AAVS1* locus
BFTC-AAVS1_A3B**	Requires Tet3G transactivator transfection for dox-inducible expression of catalytically inactive A3B (E68Q/E255Q) tagged with V5 (A3B**-V5) from the *AAVS1* locus
BFTC_A3B	Dox-inducible expression of A3B-V5
BFTC_A3A	Dox-inducible expression of A3A-V5

Chapter 6

The cellular response to acute A3A and A3B exposure in UCC

Chapter 6 The cellular response to acute A3A and A3B exposure in UCC

6.1 Introduction

Findings presented in Chapter 3 and Chapter 4, and by others (Isozaki et al., 2021; Kanu et al., 2016; Mayekar et al., 2020; Middlebrooks et al., 2016; Oh et al., 2021; Periyasamy et al., 2021; Yamazaki et al., 2020) revealed that commonly used anticancer agents transiently induce expression of both *A3A* and *A3B*. Currently, it is not entirely clear why cells would induce potential mutagens in the face of stress. However, there is emerging evidence that A3A and A3B upregulation confers a selective advantage in several tumour types and can contribute to therapy resistance. The cisplatin-resistant PEA2 ovarian cancer cell line has elevated A3B expression and activity relative to the sensitive PEA1 cell line derived from the same patient prior to cisplatin treatment, and knockdown of *A3B* restores PEA2 sensitivity (Periyasamy et al., 2021). RTKi treatment induces A3B expression and activity in NSCLC cell lines, and exogenous expression of *A3B* in a mouse model of EGFR mutant lung cancer results in higher tumour burden and tumour volume after cyclical RTKi treatment suggesting a role for A3B in RTKi resistance (Mayekar et al., 2020). RTKi treatment also increases expression and activity of A3A in NSCLC cells, and CRISPR knockout or shRNA knockdown of *A3A* prevents the emergence of RTKi resistant clones suggesting A3A, as well as A3B, has a role in RTKi resistance (Isozaki et al., 2021). Infection of human melanoma cells with the oncolytic virus, vesicular stomatitis virus (VSV), also induces expression of *A3B* and shRNA knockdown prevents the emergence of VSV-resistance both *in vitro* and *in vivo,* while A3B overexpression reduces VSV viral titre and efficacy (Huff et al., 2018).

6.1.1 A3B is required for ER gene transcription in breast cancer

Evidence for a selective benefit of A3B upregulation is particularly compelling in breast cancer. High expression of *A3B* in ER+ breast cancer patient samples is associated with aggressive phenotypes (Cescon et al., 2015), worse prognosis, and poor response to therapy (Law et al., 2016; Periyasamy et al., 2015; Periyasamy et al., 2021; Sieuwerts et al., 2014). *A3B* expression is strongly associated with PFS in ER+ breast cancer patients receiving tamoxifen therapy;

patients with low levels of *A3B* have more durable responses than those with high *A3B* suggesting A3B plays a key role in the development of tamoxifen resistance (Law et al., 2016). Indeed, *in vivo* mouse experiments demonstrated that A3B is required for the development of tamoxifen-resistant tumours. Knockdown of *A3B* in ER+ xenograft tumours resulted in a durable response to tamoxifen with limited tumour growth, while tumours overexpressing wild-type, but not catalytically inactive A3B, rapidly developed tamoxifen resistance (Law et al., 2016). The association between *A3B* expression, prognosis, and tamoxifen resistance is due to its role in regulating growth of ER+ breast cancer cells. siRNA knockdown of *A3B* reduces oestrogen-induced expression of ER target genes and growth of ER+ cell lines *in vitro* and *in vivo* (Periyasamy et al., 2015). A3B is recruited to oestrogen-response elements (EREs) through interaction with ER and promotes transcription through its deamination activity. C-to-U deamination followed by uracil excision by the BER enzyme, UNG, generates an abasic site that is converted to a DSB after further processing; subsequent repair of the DSB initiates chromatin remodelling that facilitates ER gene transcription (Figure 6.1).

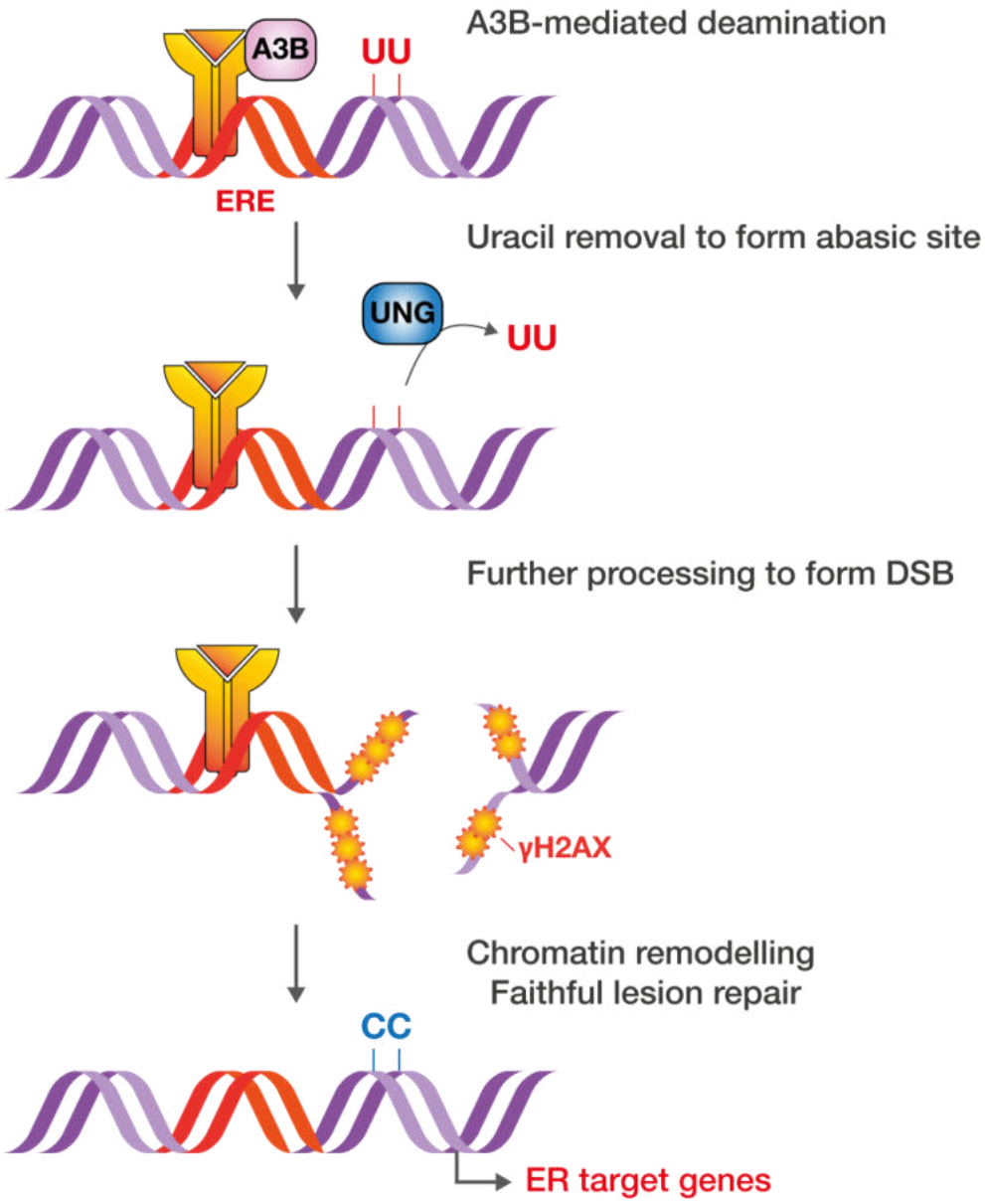


Figure 6.1 A3B-mediated transcription of ER target genes.

A3B is recruited to oestrogen-response elements (ERE) by the oestrogen receptor (ER). Localised deamination events and subsequent downstream processing generates a double strand break (DSB). DSB repair stimulates chromatin remodelling that facilitates gene transcription. Adapted from Periyasamy et al., (2015).

6.1.2 A3A is an RNA editing enzyme

While there is a clear role for A3B in ER+ breast cancer, it is less clear whether a similar function of A3A exists as no functional studies to date have shown a mechanistic role for A3A in ER gene transcription. However, like its cousin, A1, A3A is an RNA editing enzyme. RNA editing exists to transiently alter the coding

sequence of a transcript without changing the sequence of the heritable genome. Editing is primarily performed by the adenosine deaminase acting on RNA (ADAR) family of enzymes that deaminate adenosine (A) to inosines (I) and the AID/APOBEC family that catalyse C-to-U editing (Kung et al., 2018). RNA editing is an additional layer of post-transcriptional gene regulation as mutations in RNA can affect mRNA stability and transiently alter the cellular proteome through amino acid changes. Widespread C-to-U mRNA editing occurring in monocytes and macrophages in response to hypoxia and interferon stimulation has been attributed to A3A as RNA editing frequencies were reduced with A3A siRNA knockdown (Sharma et al., 2015). Transient exogenous overexpression of A3A in HEK293T cells further demonstrated an RNA editing function (Sharma et al., 2017). Editing was observed in thousands of genes and many of those identified previously as A3A-edited mRNAs in monocytes and macrophages were detected in addition to novel sites (Sharma et al., 2017; Sharma et al., 2015). The double-domain enzyme, A3G, also induces widespread C-to-U RNA editing when transiently expressed in HEK293T cells demonstrating multiple A3 family members have the capability to edit RNA (Sharma et al., 2016). As observed with A3A in monocytes and macrophages, hypoxia and cellular crowding induces A3G-mediated RNA editing in natural killer cells and this results in metabolic remodelling (Alqassim et al., 2021; Sharma et al., 2019). In contrast to A3A and A3G, there is currently no evidence supporting an RNA editing function of A3B. Jalili et al., (2020) found that mRNA mutations identified in patient tumours strongly correlate with *A3A* expression but not *A3B.* The authors also demonstrated that C-to-U editing of the two most frequently edited sites identified in tumours, $DDOST^{558}$ and $CYPIF1^{3222}$, does not occur in cells exogenously expressing A3B. However, despite the extensive homology of the A3 family, when comparing the genes edited by A3A and A3G after overexpression in HEK293T cells, there was only a small overlap in the genes edited and this shows that the RNA editing profiles of A3A and A3G are distinct (Sharma et al., 2016). The A3 family members have different preferences for nucleic acid secondary structure, and it is likely that the secondary structure of the mRNA surrounding the target C influences whether it will be edited (Hou et al., 2021). Therefore, it is possible that if A3B has an RNA editing function, its targets will be different to those of A3A.

6.1.3 A3A and A3B cause DNA damage, sensitising cells to DDR inhibition

While the immortalised embryonic kidney cell line HEK293 is not a true model for normal cells, it is easy to manipulate and has low endogenous A3 expression and activity levels, making it a useful model for studying how normal cells may respond to elevated levels of A3s (Akre et al., 2016; Nikkilä et al., 2017). Expression of wild-type, but not catalytically inactive (E68Q/E225Q), A3B for 6 days in HEK293 clones reduces cell viability by ~40% (Burns, Lackey, et al., 2013), while expression in clones derived from a clonal T-REx-293 parental line reduces cell viability by over 80% (Akre et al., 2016; Nikkilä et al., 2017). Exogenous A3B expression in HEK cells causes increases in γ-H2AX foci, p53 accumulation and PARP cleavage, G2/M cell cycle arrest and the formation of multi- and anucleate cells suggesting activation of the DDR (Burns, Lackey, et al., 2013; Lackey et al., 2013; Nikkilä et al., 2017). *A3B* expression is elevated in tumours relative to normal tissues; this suggests that tumour cells have to adapt to cope with elevated levels of DNA damage and overcome the A3-induced cell cycle arrest observed in HEK293 cells (Burns, Lackey, et al., 2013; Burns, Temiz, et al., 2013; Swanton et al., 2015). In support of this, A3B expression in p53-depleted HEK293 cells still induces DNA damage and markers of the DDR but cell cycle arrest does not occur suggesting that p53 loss facilitates proliferation in the presence of A3B-mediated damage (Nikkilä et al., 2017).

Exogenous A3A overexpression in HEK293 cells also causes DNA damage, activation of the DDR and cell cycle arrest (Lackey et al., 2013; Land et al., 2013; Landry et al., 2011). Interestingly, cell cycle arrest in response to A3A expression is more rapid than that seen with A3B and occurs in S-phase suggesting the cellular response to A3A- and A3B-induced damage is different (Burns, Lackey, et al., 2013; Lackey et al., 2013; Landry et al., 2011). In addition, after 6 days of A3A expression, viability is reduced by ~80%, more than that seen with A3B in the same study, suggesting that A3A is more toxic than A3B (Burns, Lackey, et al., 2013). However, it has been proposed that the A3A-mediated toxicity seen in HEK293 cells is due to mislocalisation of A3A in stably or transiently transfected cells (Land et al., 2013). The authors showed that exogenous GFP-tagged A3A has a cell-wide distribution while interferon-induced expression of endogenous A3A in immune-lineage cells is primarily cytoplasmic. Thus, they suggested cells

that usually express A3A use cytoplasmic retention mechanisms to limit its genotoxicity.

The DNA damage induced by A3A/A3B activity can increase sensitivity of cells to anticancer agents. Exogenous expression of A3A induces cell cycle arrest in U2OS and THP-1 cells and sensitises them to ATR and CHK1 inhibition (Buisson et al., 2017; Green et al., 2017) while exogenous A3B expression in p53-depleted HEK293 cells sensitises them to inhibition of WEE1, CHK2 and PARP in addition to ATR and CHK1 (Nikkilä et al., 2017). Inhibition of DDR signalling abrogates the cell cycle checkpoints enabling cells to replicate with extensive DNA damage and replication stress leading to RPA exhaustion, replication catastrophe, and cell death (Buisson et al., 2017; Green et al., 2017; Nikkilä et al., 2017; Toledo et al., 2013).

6.1.4 Chapter aims

The aim of this chapter was to use the cell line models generated in Chapter 5 to determine whether A3B has a transcriptional function in UCC similar to that identified in breast cancer (Periyasamy et al., 2015). In addition, A3A is highly homologous to the CTD of A3B, and therefore, it is possible that these two family members have redundant functions and A3A also has a role in transcriptional activation. To investigate this, RNA sequencing after acute exposure to elevated levels of A3A and A3B will be used alongside identification of their interacting partners using the proximity-labelling technique, BioID. Characterisation of A3A's interactome will also provide clues as to whether A3A is actively recruited into the nucleus or is primarily retained in the cytoplasm. Finally, cell growth and IC_{50} assays will be used to determine if elevated expression of A3A and A3B induces a proliferation defect or alters drug sensitivity in UCC, as has been previously reported in other models.

6.2 Results

6.2.1 Elevated A3A and A3B expression has subtle effects on the transcriptome

Analysis of UCC tumour samples has revealed that there are two main subtypes that are molecularly similar to the basal and luminal subtypes seen in breast cancer (Choi et al., 2014; Damrauer et al., 2014). Luminal subtypes of UCC have evidence of ER gene transcriptional profiles (Choi et al., 2014; Damrauer et al., 2014) opening the possibility that A3B may have an ER transcriptional role in bladder cancer as well as in breast (Periyasamy et al., 2015). This chapter aimed to use RNAseq to compare the transcriptional profiles of one basal cell line, BFTC-905, and one luminal cell line, SW780, after acute A3B expression to determine if there was evidence of an ER or novel gene transcriptional profile. While both cell lines are derived from female patients, attempts to make luminal subtype expressing cell lines were unsuccessful and this limits the ability of the study to identify ER transcriptional activity. RNAseq was also performed after acute exposure to A3A to determine whether A3A has a novel transcriptional role. Prior to the RNAseq experiments, the four BFTC-905-*AAVS1* stable cell line pools (A3A*, A3A, A3B** and A3B) were confirmed to be negative for *Mycoplasma* infection. The stable cell lines were transiently transfected with Tet3G transactivator plasmid and then seeded in 6-well plates with doxycycline to induce expression of the V5-tagged proteins. Induction in response to doxycycline is seen in western blot samples set up in parallel to those used for RNAseq (Figure 6.2). RNA was prepared and sent to BGI for RNA sequencing; all 30 RNA samples passed quality control with RIN ≥ 9.7. Table 6.1 shows details of the samples including codes which will be used in all RNAseq data analysis described in this chapter. All RNAseq bioinformatic analysis described was kindly performed by Dr Pradeep Ramagiri.

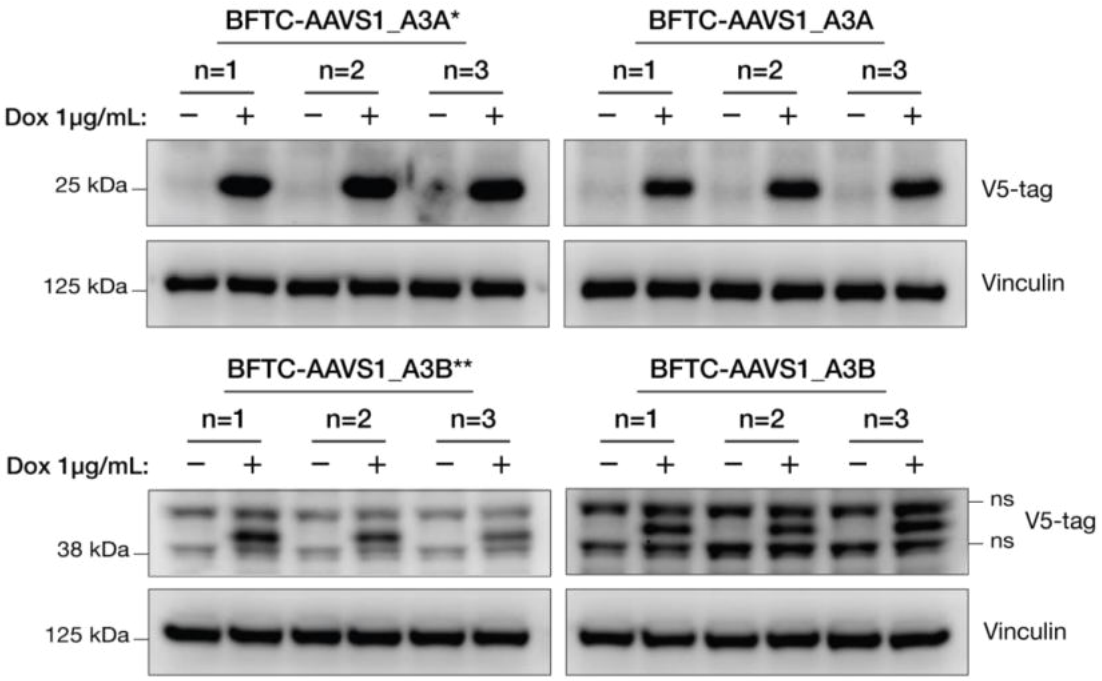


Figure 6.2 Western blot to check expression of A3A*, A3A, A3B and A3B** in response to doxycycline after Tet3G transient transfection prior to sending samples for RNAseq.**

BFTC_AAVS1 cells were seeded in T12.5cm flasks at 5 x 10^5 cells/flask and transfected with 3.25 µg Tet3G-only plasmid 24 hours later at 60% confluence. Transfection complexes were removed 6 hours post-transfection and 18 hours later cells were re-seeded into media containing the indicated amounts of doxycycline in duplicate 6-well plates at 1 x 10^5 cells/well for RNA preparation to send for RNAseq, and protein to check for expression. Cells were harvested after 72 hours doxycycline treatment and 30 µg protein was used for western blot; vinculin was used as a loading control, V5-tag blotting detects tagged protein expression.

Table 6.1 Sample information for RNAseq analysis.

Cell Line	Code	Sample Group	Sample Number
BFTC-AAVS1_A3A*	A3A* (+Dox)	A3A*-V5 (E72A) Tet3G plasmid transfected; 1 µg/mL doxycycline	Sample_1 – 3
	A3A* (mock)	A3A*-V5 (E72A) Mock transfected; 1 µg/mL doxycycline	Sample_19 – 21
BFTC-AAVS1_A3A	A3A (−)	A3A-V5 (wild-type) Tet3G transfected; no doxycycline	Sample_4 – 6
	A3A (+Dox)	A3A-V5 (wild-type) Tet3G plasmid transfected; 1 µg/mL doxycycline	Sample_7 – 9

	A3A (mock)	A3A-V5 (wild-type) Mock transfected; 1 µg/mL doxycycline	Sample_22 – 24
BFTC-AAVS1_A3B**	A3B** (+Dox)	A3B**-V5 (E68Q/E255Q mutant) Tet3G plasmid transfected; 1 µg/mL doxycycline	Sample_10 – 12
	A3B** (mock)	A3B**-V5 (mutant) Mock transfected; 1 µg/mL doxycycline	Sample_25 – 27
BFTC-AAVS1_A3B	A3B (−)	A3B-V5 (wild-type) Tet3G plasmid transfected; no doxycycline	Sample_13 – 15
	A3B (+Dox)	A3B-V5 (wild-type) Tet3G transfected; 1 µg/mL doxycycline	Sample_15 – 18
	A3B (mock)	A3B-V5 (wild-type) Mock transfected; 1 µg/mL doxycycline	Sample_28 – 30

All sequencing runs passed QC with > 97.7% bases having an inferred base call accuracy of 99% (Q20). After data filtering and alignment, initial exploratory data analysis was performed using principal component analysis (PCA). PCA allows easy visualisation of similarities or differences between treatment groups in large datasets. PCA analysis revealed that the biological replicates for each treatment clustered together; this shows that there is reproducibility between the repeats (Figure 6.3). The lack of clustering between the mock transfected control groups (A3A*, A3A, A3B** and A3B (mock)) represents differences in baseline gene expression between the stable cell line pools. The lack of clustering between mock and Tet3G transactivator transfected pools is likely due to transcriptome changes arising in response to exogenous DNA transfection in addition to changes resulting from tagged protein of interest expression; this will be discussed in more detail later in this chapter. The close clustering of the samples with and without doxycycline treatment indicates that expression of wild-type A3A or A3B has subtle effects on the transcriptome; this is not entirely unexpected as the models used are designed to have lower, more physiologically relevant levels of elevated expression. While A3A and A3A* (+Dox) cluster closely together, A3B and A3B** (+Dox) are less closely clustered and this could indicate that changes occurring in response to A3A expression are less dependent on its catalytic function compared to A3B.

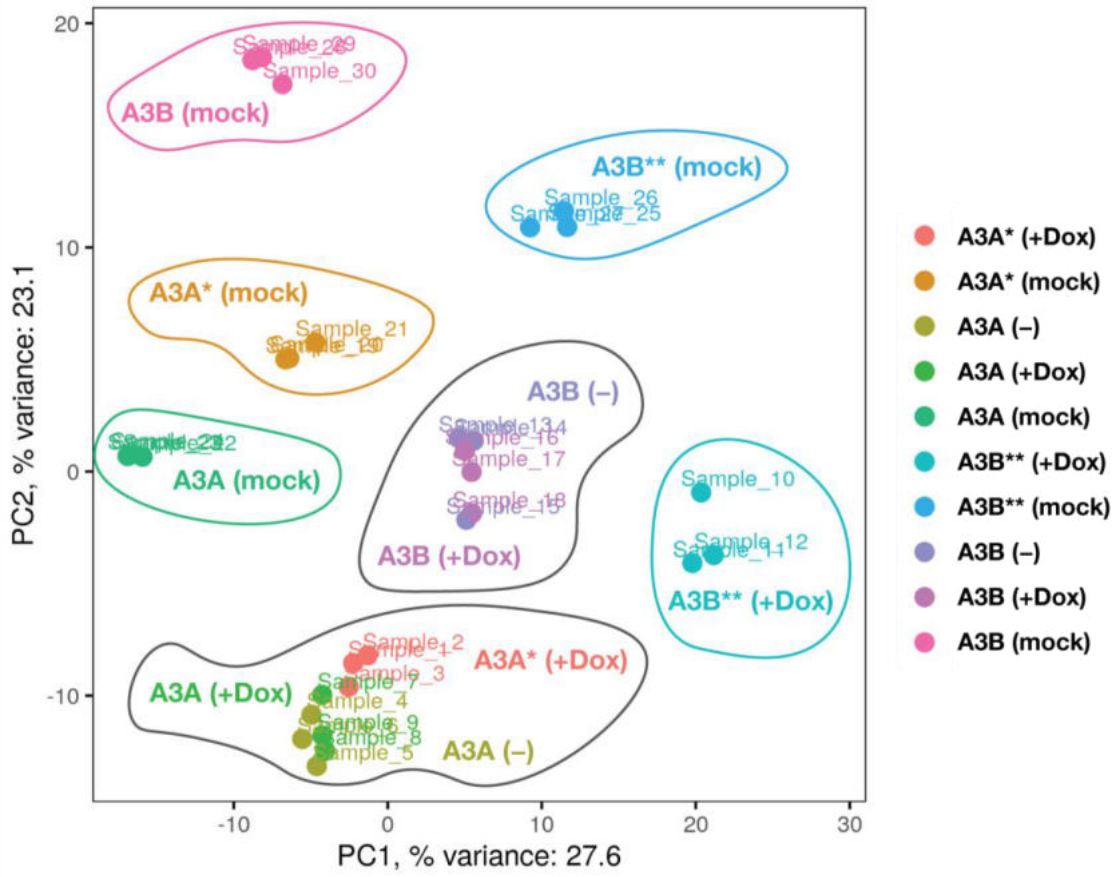


Figure 6.3 Principal component analysis (PCA) plot for 30 samples sent for RNAseq.

A3A* = A3A*-V5 (E72A) catalytically inactive mutant; A3A = A3A-V5 wild-type; A3B** = A3B**-V5 (E68Q/E255Q) catalytically inactive mutant; A3B = A3B-V5 wild-type. (+ Dox) = cells transfected with Tet3G plasmid and treated with 1 µg/mL doxycycline for 72 hours to induce expression; (–) = cells transfected with Tet3G plasmid only; (mock) = cells mock transfected with Lipofectamine only and treated with 1 µg/mL doxycycline for 72 hours. PCA analysis was performed by Dr Pradeep Ramagiri.

The raw count data for each sample was normalised to correct for library size and RNA composition bias before being divided into two groups, with one used as the baseline, to determine differentially expressed (DE) genes between the groups. DE analysis was performed in R using the Bioconductor package, DEseq2. To investigate the changes occurring in response to A3A and A3B expression several comparisons were made. The stable cell line pools used for this work require transient transfection with a Tet3G transactivator expressing plasmid prior to doxycycline treatment to induce expression of the V5-tagged proteins of interest (Chapter 5). PCA analysis showed that there were baseline differences between the four cell lines and therefore, to investigate the response to

transfection and protein induction, the induced conditions (transfected + Dox) were compared to their mock transfected (+ Dox) controls. This revealed profound changes to the transcriptome (Figure 6.4). Table 6.2 summarises the numbers of statistically significant DE genes (adjusted *P* value (padj) > 0.05). These changes encompass DE genes from both transfection and expression of the protein of interest.

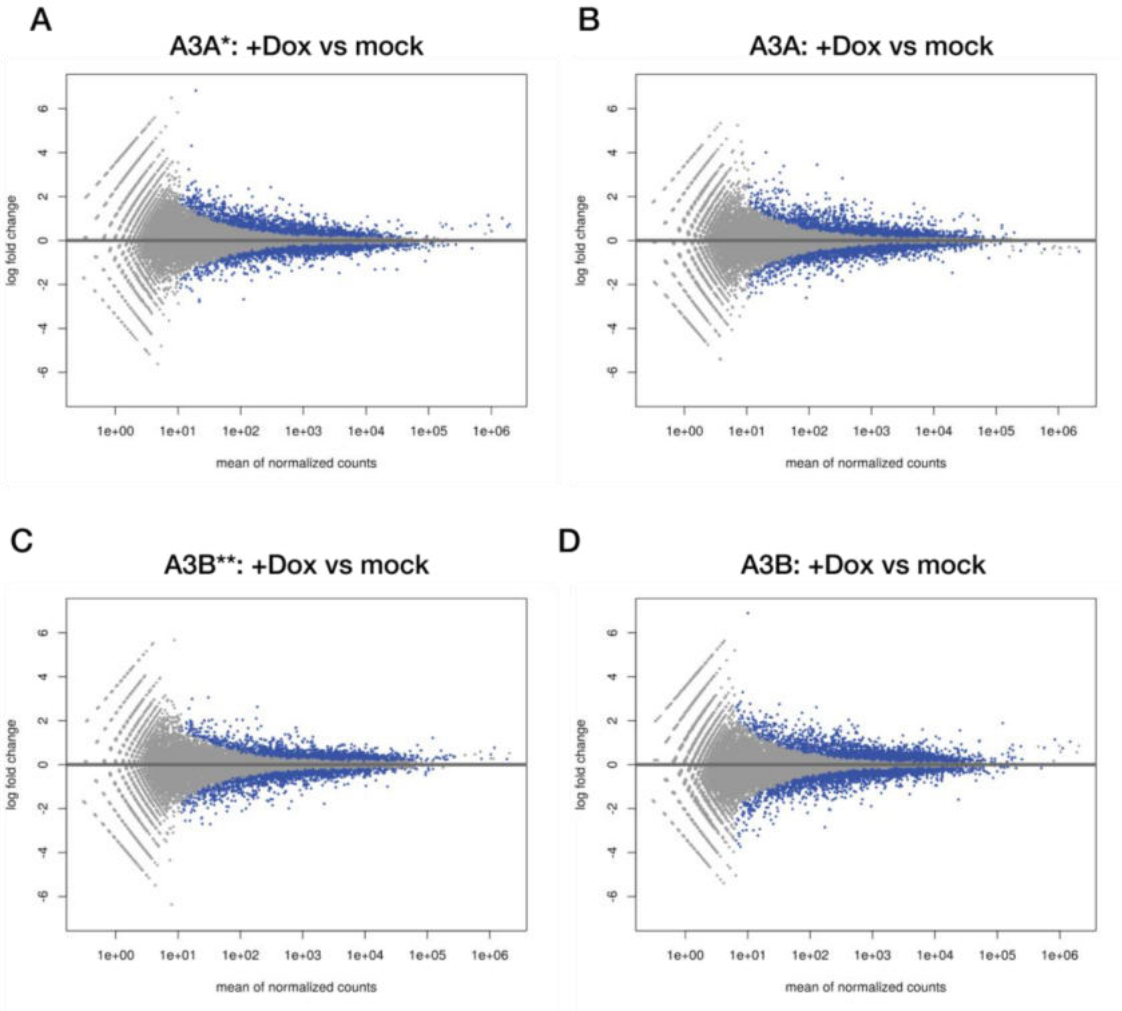


Figure 6.4 MA-plots showing transcriptional changes in all stable cell line pools transfected with the Tet3G transactivator plasmid and protein expression induced with doxycycline.

The plots show log2 fold changes over the mean of normalized counts where the x-axis is the average expression over both samples and the y-axis is the log2 fold change between the condition groups. Each gene is a dot and differentially expressed genes with a significant padj value are in blue. DE genes in the BFTC-905 pools (A) A3A* (+Dox) vs A3A* (mock). (B) A3A (+Dox) vs A3A (mock). (C) A3B** (+Dox) vs A3B** (mock). (D) A3B (+Dox) vs A3B (mock). Analysis by Dr Pradeep Ramagiri.

Table 6.2 Summary of DE genes between Tet3G and mock transfected pools. padj < 0.05.

Comparison	Significantly DE genes	Up	Down
A3A*: +Dox vs mock	3355	1610	1745
A3A: +Dox vs mock	3774	1977	1797
A3B**: +Dox vs mock	3079	1705	1374
A3B: +Dox vs mock	5085	2722	2363

Gene set enrichment analysis (GSEA) uses the Molecular Signatures Database (MSigDB) that contains annotated gene sets of well-characterised biological processes to provide biological insight into the changes occurring between two datasets. There are two main methods of performing GSEA. The first involves using DE genes predetermined as being of interest, for example, based on their significance value or $\log_2$ fold change. The other method involves simply ranking all significantly DE genes regardless of their *P* value or $\log_2$ fold change. Both methods have pros and cons; as the input genes are predetermined, the first method is limited in that genes that are biologically relevant may have been excluded because they did not meet a predetermined threshold for inclusion. Alternatively, with the second method, while ensuring potential interesting changes are not excluded, the data is ranked and will identify enriched pathways even when $\log_2$ fold changes are small, but the biological significance of this enrichment is questionable. For this study, method two was employed and the list of DE genes was ranked according to their DE significance (*P* value) and their fold change (including the sign value to determine whether they are up- or downregulated), and GSEA was performed in R using the Bioconductor package fgsea using Hallmark gene sets from MsigDB (Figure 6.5). This revealed enrichment of similar biological processes in response to Tet3G plasmid transfection and doxycycline protein induction for the four cell lines. Many of the enriched pathways are involved in the innate immune response and cellular stress (TNFα signalling, complement, interferon, TGFβ).

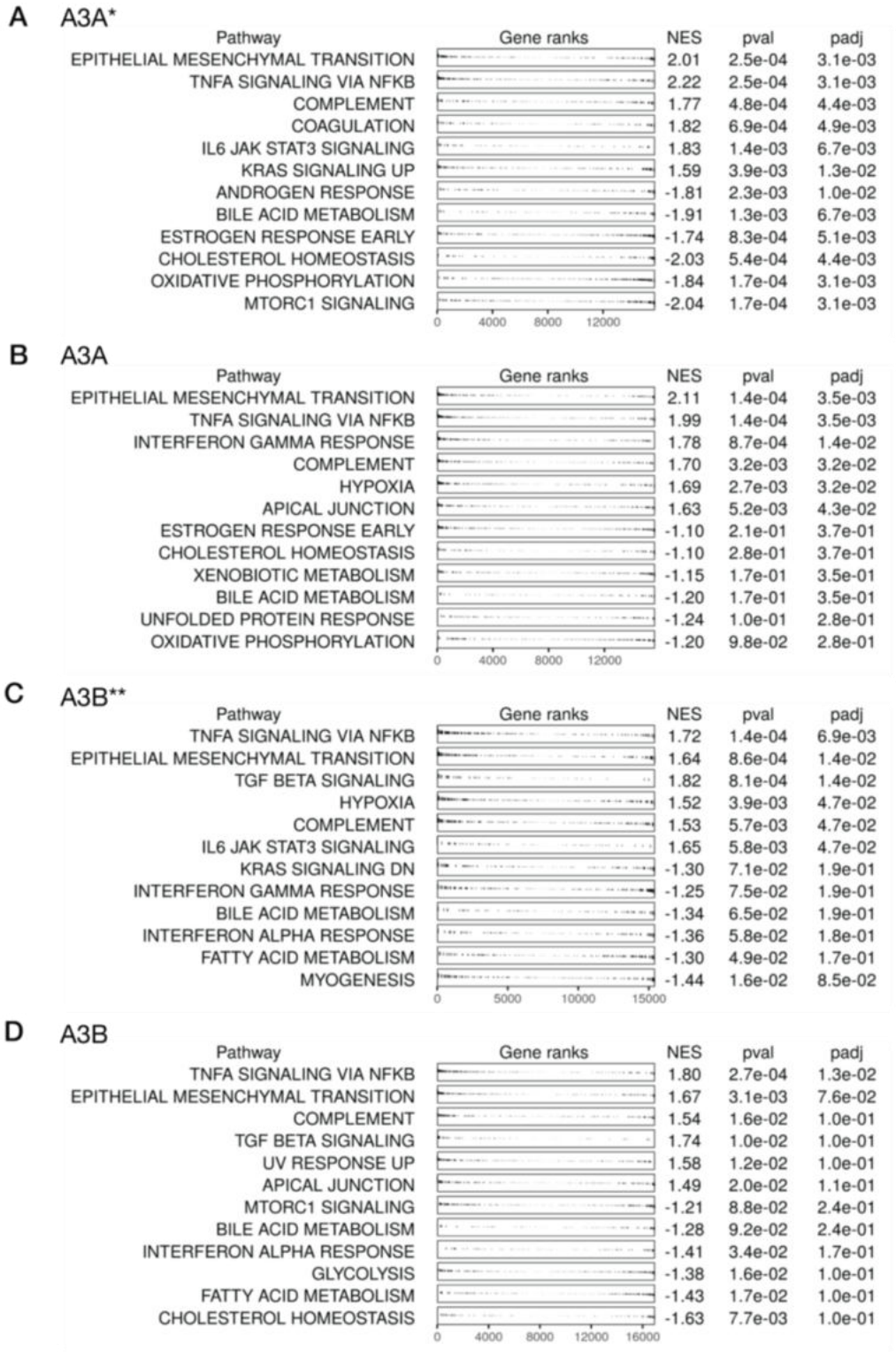


Figure 6.5 Gene set enrichment analysis (GSEA) of differentially expressed genes in Tet3G transactivator plasmid transfected cells (+Dox) vs mock transfected (+ Dox) for all stable cell lines.

Lists of the most enriched pathways identified using GSEA and the MSigDB hallmark gene sets based on their normalised enrichment score (NES). A3A* (A), A3A (B), A3B** (C) and A3B (D). Analysis by Dr Pradeep Ramagiri.

To investigate the transcriptome changes occurring because of exogenous acute expression of A3A or A3B, DE analysis was performed comparing A3A or A3B (+Dox) to (− Dox) samples, where (−) was used as the baseline. Cells were transfected in a flask prior to being split into minus and plus doxycycline treatments. Therefore, all DE changes can largely be attributed to the expression of A3A or A3B and are summarised in Figure 6.6. Expression of A3A resulted in 218 statistically significant DE genes (padj ≤ 0.05); of these, 25 were upregulated and 193 were downregulated. Figure 6.7 shows the top 100 DE genes with the lowest adjusted *P* values in a heatmap. Expression of A3B resulted in 62 statistically significant DE genes (padj ≤ 0.05); of the differentially expressed genes, 16 were upregulated and 46 were downregulated. Figure 6.8 shows the 62 DE genes in a heatmap. The small number of genes identified in these comparisons demonstrates that majority of the changes identified when comparing to the mock transfected controls (Figure 6.4) are a result of plasmid transfection and not A3A*, A3A, A3B** or A3B protein expression. To characterise changes occurring only because of plasmid transfection, comparisons between A3A (– dox) and A3A (mock), and A3B (– dox) and A3B (mock) could be conducted.

In conclusion, the changes occurring as a result of acute elevated expression of A3A and A3B are subtle. The number of DE genes and their $\log_2$ fold changes are small and as discussed previously, GSEA is not appropriate.

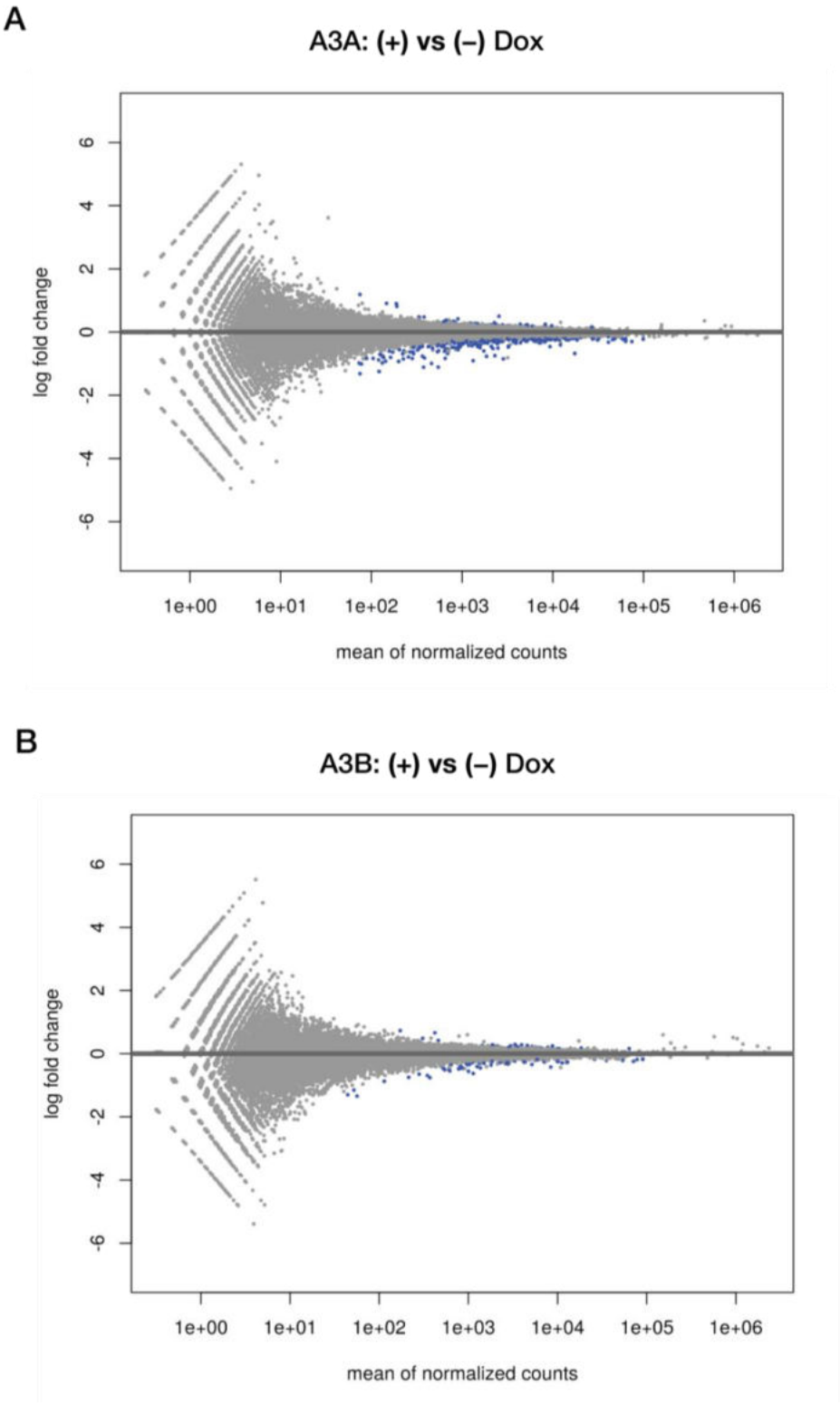


Figure 6.6 Acute expression of A3A and A3B has modest effects on the transcriptome.

MA-plot showing log2 fold changes over the mean of normalized counts where the x-axis is the average expression over all samples and the y-axis is the log2 fold change between the condition groups. Each gene is a dot and differentially expressed genes with a significant padj value are in blue. (A) A3A (wild-type): + vs – doxycycline. (B) A3A (wild-type): + vs – doxycycline. Analysis performed by Dr Pradeep Ramagiri.

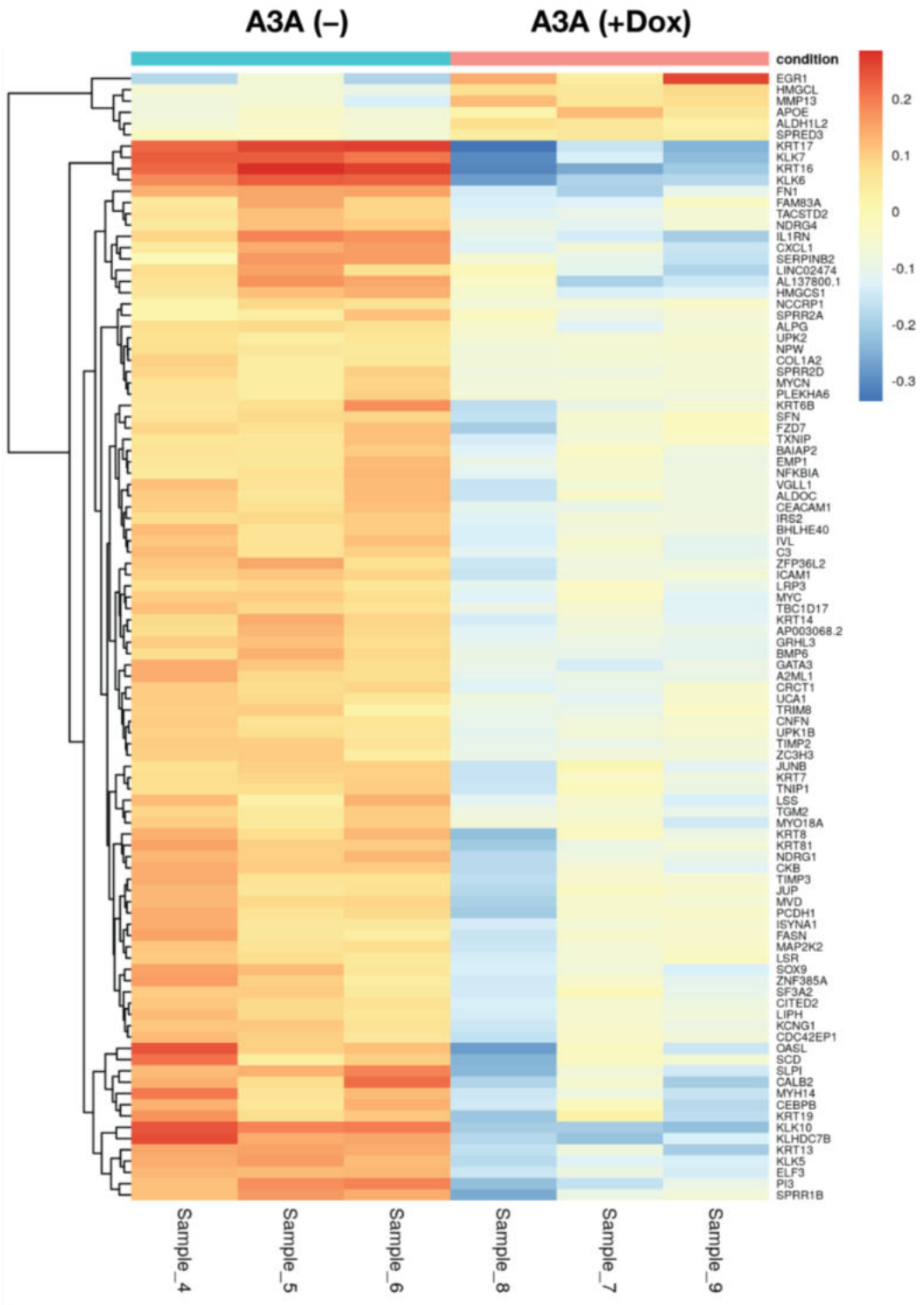


Figure 6.7 Top 100 differentially expressed genes in response to acute A3A expression.

Heatmap showing the expression data of the top 100 differentially regulated genes with the lowest padj value (data has been rlog transformed). The A3A (−) group was used as a baseline to calculate the differentially expressed genes between the two groups. Analysis performed by Dr Pradeep Ramagiri.

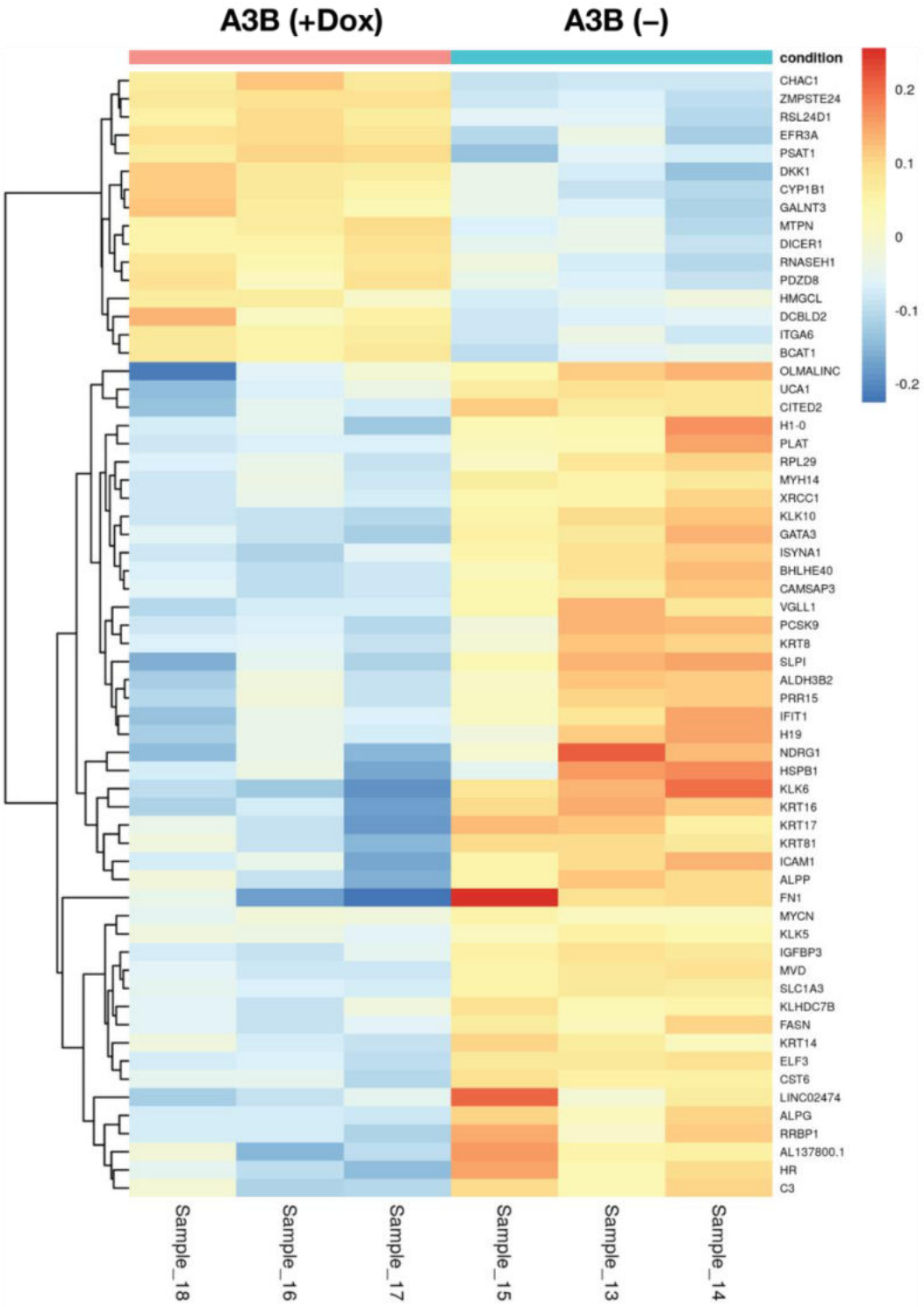


Figure 6.8 The 62 differentially expressed genes in response to acute A3B expression.

Heatmap showing the expression data of the top 100 differentially regulated genes with the lowest padj value (data has been rlog transformed). The A3B (−) group was used as a baseline to calculate the differentially expressed genes between the two groups. Analysis performed by Dr Pradeep Ramagiri.

Comparison of the DE genes identified in response to A3A and A3B expression revealed 40 commonly DE genes (Figure 6.9). Interestingly, several basal cell markers, *KRT16, KRT15*, *KRT14*, and *MYCN* were downregulated in response to both A3A and A3B expression suggesting that their expression may be stimulating the emergence of a more luminal phenotype.

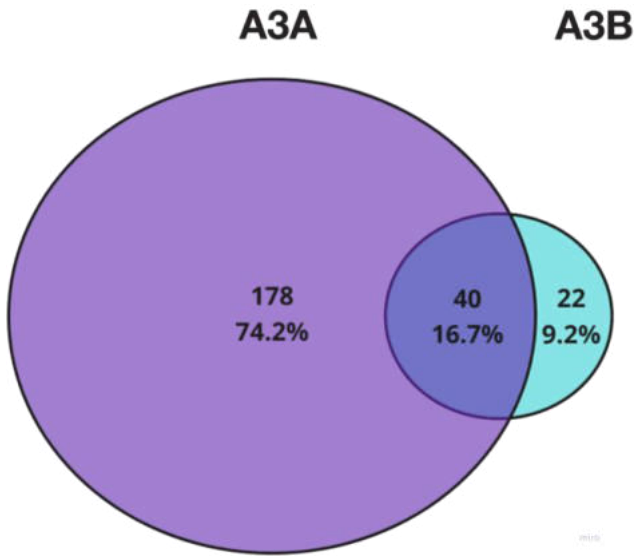


Figure 6.9 Venn diagram showing common and unique differentially expressed genes between A3A-V5 and A3B-V5 expressing cells.

DE genes identified in response to A3A-V5 expression shown in purple and A3B-V5 expression are shown in blue. Overlapping gene sets identified with Venny (Oliveros, 2007).

DE analysis was also performed comparing wild-type A3A and A3B (+Dox) to catalytically inactive A3A* and A3B** (+Dox), where the mutant (A3A* or A3B**) was used as the baseline, to determine if transcriptome changes specifically related to deamination activity could be identified. The transcriptome changes are summarised in Figure 6.10. Comparisons between A3A and A3A* revealed 1167 significantly DE genes (514 upregulated and 653 downregulated) while comparisons between A3B and A3B** revealed 3997 significantly DE genes (1757 upregulated and 2240 downregulated). As expected, the ability of these comparisons to identify genes differentially expressed because of catalytic activity is limited as it encompasses DE genes in response to both catalytic activity and pre-existing baseline differences between the cell lines, as discussed previously.

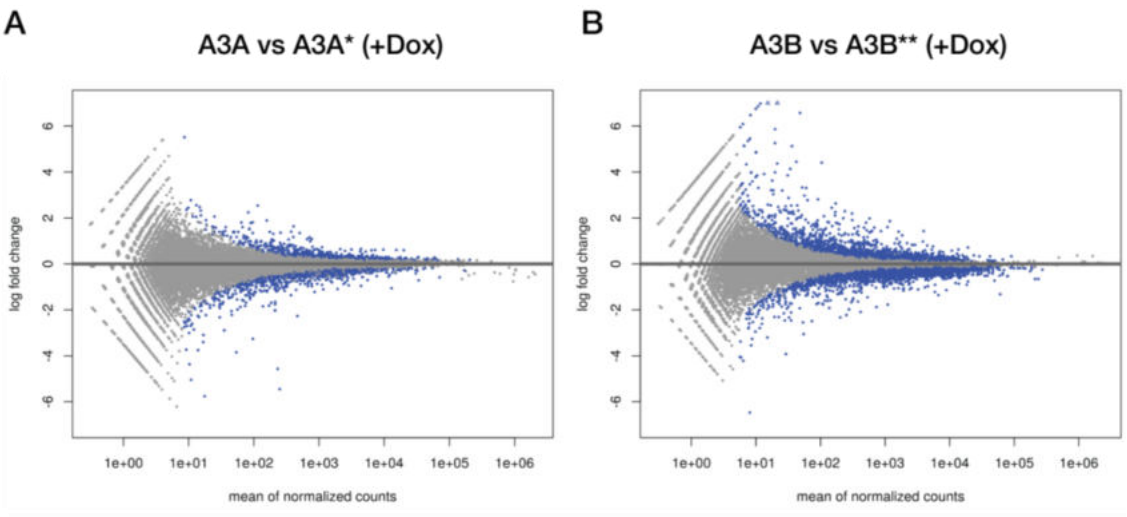


Figure 6.10 DE analysis between wild-type and mutant (catalytically inactive) A3A and A3B.

MA-plot showing log2 fold changes over the mean of normalized counts where the x-axis is the average expression over all samples and the y-axis is the log2 fold change between the condition groups. Each gene is a dot and differentially expressed genes with a significant padj value are in blue. (A) A3A wild-type vs A3A* mutant (+Dox). (B) A3B wild-type vs A3AB* mutant (+Dox). Analysis performed by Dr Pradeep Ramagiri.

6.2.2 A3A and A3B are putative RNA editing enzymes in UCC

A3A has been shown to edit RNA in several model systems (Jalili et al., 2020; Sharma et al., 2017; Sharma et al., 2015). Therefore, the RNAseq dataset was used for a preliminary investigation into the potential RNA editing functions of A3A and A3B. RNA sequencing raw reads were mapped to the reference genome and subject to the Mutect2 variant calling pipeline to detect putative mRNA mutations; RNAseq library preparation requires mRNA to cDNA synthesis and therefore, C-to-U mRNA editing can be detected as a C-to-T change relative to the reference sequence. A3A/A3B doxycycline-treated (+Dox) samples were compared to untreated (−) to identify RNA editing events specifically related to the induction of A3A/B. As the comparisons are made within a cell line pool that was split into the two treatment groups, any changes detected are likely to be a result of RNA editing rather than SNPs already present in genome. Comparisons between cell lines were not conducted as any changes identified could either be SNPs present in the genome or mRNA editing events. RNA variant calling is a relatively new method of analysis and is open to interpretation. To be more confident that any identified mutations were real events and not an artifact of the

analysis, a stringent threshold was used. A mutation was called where depth read exceeded 50 (DP > 50), the mutation was detected in at least 10% of the reads (allelic frequency (AF) > 10%) and was present in the + Dox but not the – Dox condition. This analysis revealed evidence of RNA editing after acute exposure to both A3A and A3B (Table 6.3). A total of 911 mutations were identified after acute expression of A3A; 17.1% were C-to-T suggestive of A3 activity, while 12.5% were A-to-G, evidence of A-to-I editing by the ADAR family of enzymes. Similar numbers of mutations were identified after A3B expression; a total of 788 mutations were identified, of which 19.5% were C-to-T and 12.6% were A-to-G.

Table 6.3 Summary of potential RNA editing events resulting from acute exposure to A3A and A3B.

RNA mutations detected at a threshold of read depth > 50 and allelic frequency > 10% shown for the group comparison and each biological repeat. Analysis performed by Dr Pradeep Ramagiri.

Sample comparison	Total mutations	C-to-T (% total)	A-to-G (% total)
A3A (+Dox) vs (−)			
All repeats (samples 7 – 9 vs 4 – 6)	911	156 **(17.1%)**	114 **(12.5%)**
N=1: sample 7 vs 4	372	59 **(15.9%)**	55 **(14.8%)**
N=2: sample 8 vs 5	399	53 **(13.3%)**	43 **(10.8%)**
N=3: sample 9 vs 6	367	59 **(16.1%)**	53 **(14.4%)**
A3B (+Dox) vs (−)			
All repeats (samples 16 – 18 vs 13 – 15)	788	154 **(19.5%)**	99 **(12.6%)**
N=1: sample 16 vs 13	345	64 **(18.6%)**	47 **(13.6%)**
N=2: sample 17 vs 14	312	52 **(16.7%)**	43 **(13.8%)**
N=3: sample 18 vs 15	366	65 **(17.8%)**	52 **(14.2%)**

In conclusion, there is a suggestion that RNA editing is occurring in response to A3A/B exposure but to fully characterise whether the identified C-to-T mutations detected in the RNAseq dataset can be attributed to A3A and/or A3B activity, the sequence context surrounding the mutation needs to be characterised. However, this is beyond the scope of this project due to limitations with the dataset that will be discussed fully later.

6.2.3 Acute A3B exposure does not affect cell growth or drug sensitivity

Acute A3A and A3B exposure induces DNA damage and causes a proliferation defect in HEK293 cells (Akre et al., 2016; Burns, Lackey, et al., 2013; Nikkilä et al., 2017). This has also been reported in the tumour cell lines U2OS and HeLa (Burns, Lackey, et al., 2013; Lackey et al., 2013; Landry et al., 2011), and in response to A3A in several leukaemia lines (Green et al., 2017). A3B expression in HEK293 cells depleted of p53 sensitises them to inhibition of the DDR proteins, ATR, CHK1, CHK2, WEE1 and PARP (Nikkilä et al., 2017), while A3A expression sensitises leukaemia cells to ATR inhibition (Buisson et al., 2017; Green et al., 2017). To further investigate the cellular effects of acute elevated expression of A3B, growth and drug sensitivity to two chemotherapy drugs and two DDR inhibitors was assessed. A3B-V5 expressing BFTC-905 cells that do not require Tet3G transfection (BFTC_A3B; Chapter 5) were used for these experiments. Growth of both induced and uninduced cells was measured over 7 days using the Incucyte® S3 to measure cell confluence. Expression of A3B in BFTC-905 cells did not alter cell confluence (Figure 6.11 A) or growth, with no significant difference between non-induced and induced cells at day 5 (paired t-test, P = 0.343) (Figure 6.11 B). Expression of A3B was verified by western blot (Figure 6.11 C) and treatment of the parental BFTC-905 cell line with 1 µg/ml doxycycline confirmed that doxycycline itself has no effect on cell growth (Figure 6.11 D).

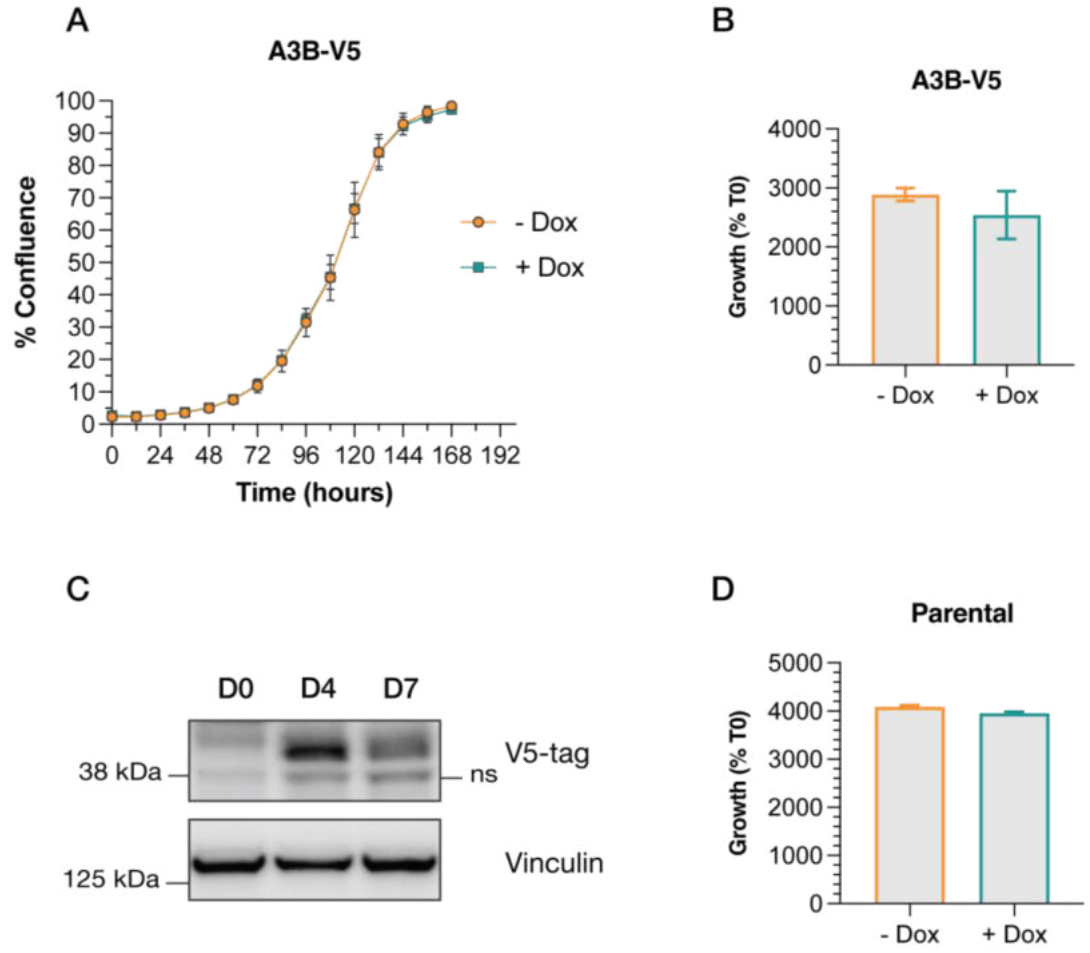


Figure 6.11 Acute A3B exposure does not affect growth in BFTC_A3B cells.

(A) BFTC_A3B cells were seeded at 500 cells/well in 96-well plates in Tet-low media and treated with 1 µg/ml doxycycline (dox) 24 hours later. Cells were imaged every 12 hours with an Incucyte® S3 to measure cell confluence. Data represents mean ± SD of two independent experiments each with 5 technical replicates. (B) Data shown as growth (% T0) for confluence readings taken at day 5. (C) BFTC-905_A3B cells were seeded in 6-well plates and treated with 1 µg/ml dox in parallel to one biological repeat from A and harvested at days 0, 4 and 7 for western blot to show expression of A3B-V5 during the Incucyte assay. V5-tag detects tagged A3B; vinculin was used as a loading control. (D) BFTC-905 parental cells were seeded at 500 cells/well in 96-well plates in Tet-low media and treated with 1 µg/ml doxycycline (dox) 24 hours later. Growth (% T0) was determined using the SRB assay after 5 days.

Next, IC_{50} values were determined for the chemotherapy drugs, gemcitabine (Figure 6.12 A) and cisplatin (Figure 6.12 B), and two DDR inhibitors, the ATM inhibitor KU-60019 (Figure 6.12 C) and the ATR inhibitor AZD6738 (Figure 6.12 D) in induced and uninduced cells. Comparison of fit of the concentration-response curves with and without A3B induction show that acute A3B exposure

does not alter sensitivity to any of the four drugs tested. These results suggest that A3B is not playing a role in DNA damage repair or generating DNA damage that renders cells more susceptible to DDR inhibition as previously suggested in other cell line models.

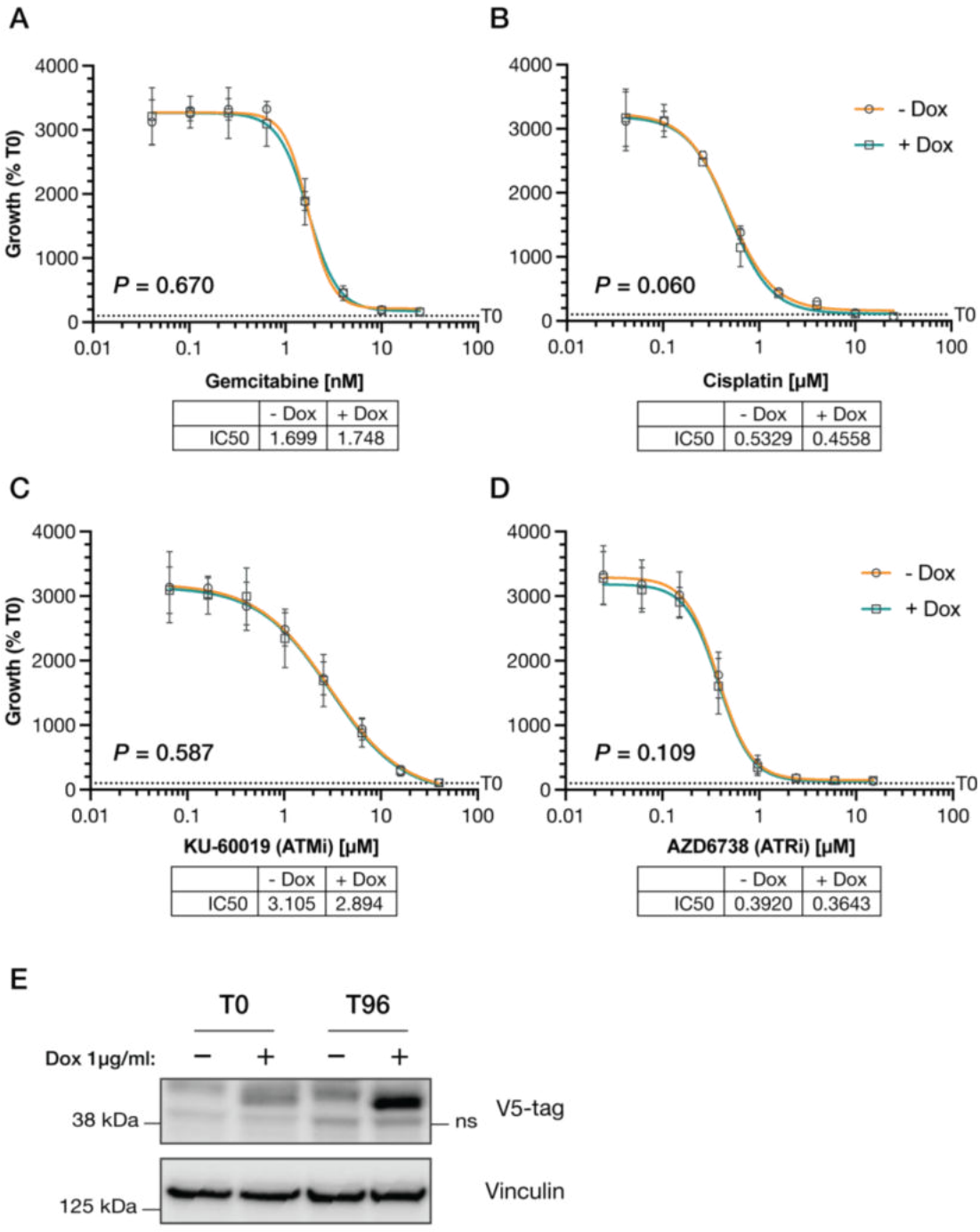


Figure 6.12 Acute A3B exposure does not affect sensitivity to two cytotoxic drugs or two DDR inhibitors in BFTC_A3B cells.

Concentration-response curves showing sensitivity of BFTC-905 A3B-V5 expressing cells to gemcitabine (A), cisplatin (B), the ATM inhibitor, KU-60019 (C) and the ATR inhibitor, AZD6738 (D). 1000 cells/well were seeded in 96-well plates in Tet-low media containing 1 µg/ml doxycycline (dox). 24 hours later, cells were treated with the indicated drugs and response was analysed after 96 hours drug treatment using the SRB assay. Data is shown as mean ± SD for two independent experiments each with 3 technical replicates. IC_{50} values are determined from the average curves using non-linear regression on GraphPad Prism 9. *P* values are shown for comparison of curve fits between – and + dox. (E) BFTC-905 A3B-V5 cells were seeded in 6-well plates and in Tet-low media containing 1 µg/ml dox in parallel to one biological repeat from (A – D) and harvested at T0 and T96 (endpoint) for western blot to show expression of A3B-V5 during the assay. V5-tag detects tagged A3B; vinculin was used as a loading control.

6.2.4 Identification of interacting partners using BioID

The lack of specific commercially available antibodies to A3A and A3B has limited the identification of interacting partners as many methods of identifying protein-protein interactions are antibody based. However, the use of proximity-labelling methods allows the antibody-independent identification of interactors and has several benefits over traditional immunoprecipitation-based methods. One such proximity-labelling method is BioID where the protein of interest is fused to a promiscuous biotin ligase (the bait) that, upon the addition of exogenous biotin, will biotinylate proximal proteins (the prey) (Roux et al., 2012). The biotinylated proteins can then be affinity purified and subject to mass spectrometry analysis for identification of interacting partners (Figure 6.13). BioID allows identification of both weak and transient interactions that may not be identified using traditional methods, and labelling can be temporally regulated. The major disadvantage of the original BioID method is the slow labelling kinetics of the biotin-ligase, BirA; efficient labelling requires treatment with exogenous biotin for at least 18 hours to produce sufficient material for analysis. Since the initial method was published in 2012, several studies have generated newer biotin ligase mutants with improved labelling characteristics, including TurboID and miniTurboID (Branon et al., 2018). These can label sufficient material for analysis in as little as 10 minutes, allowing identification of interacting partners in the context of dynamic processes that occur over short periods of time (Branon et al., 2018). Stable cell lines expressing A3A (N- and C-terminally tagged), and A3B (C-terminally tagged only) tagged with the biotin-ligase mutant miniTurbo were used to identify

interacting partners in the UCC cell line, BFTC-905 (cell lines fully characterised in Chapter 5).

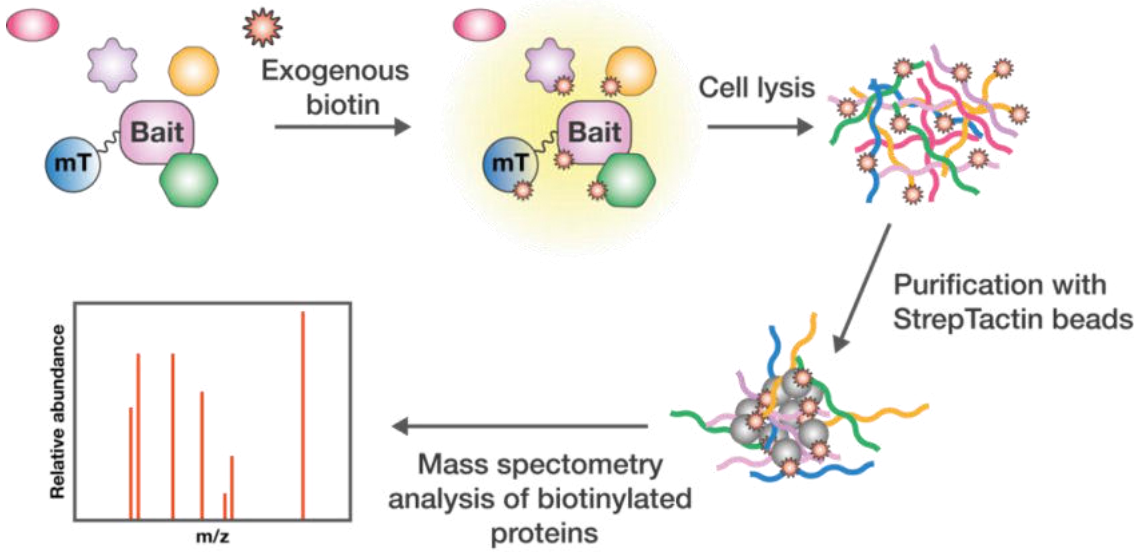


Figure 6.13 BioID workflow.

mT, miniTurbo biotin-ligase. Based on figure by Cheerathodi & Meckes (2020).

6.2.4.1 Biotinylation optimisation

Bait protein expression for proximity labelling experiments should be low to avoid mislocalisation and false discoveries (Roux et al., 2018). Expression of A3B-miniTurbo was substantially lower than the two A3A bait fusions so 500 ng/mL doxycycline was used to induce maximal expression. The expression of the 3xHA-miniTurbo-A3A (mT-A3A) bait seen in response to 100 ng/mL doxycycline is similar to the induction of the A3A-miniTurbo-V5 (A3A-mT) bait seen at the lowest doxycycline concentration tested (10 ng/mL) (Figure 6.14); these concentrations are used for all BioID experiments.

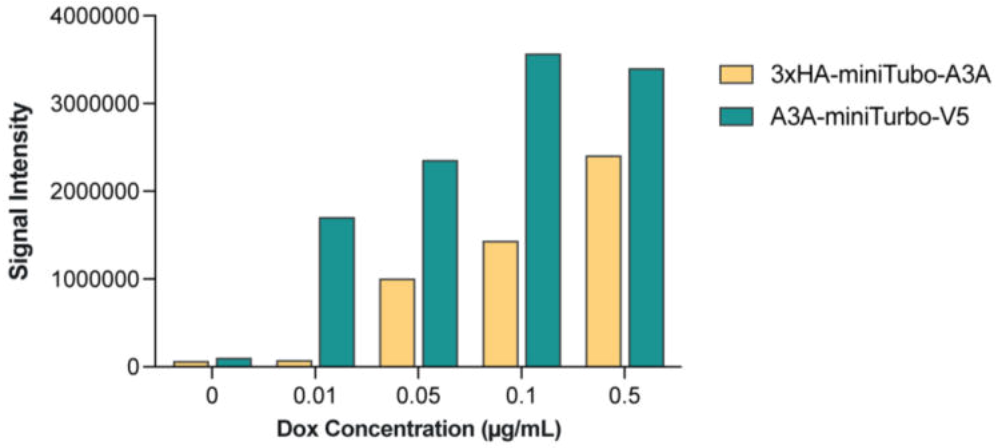


Figure 6.14 100 ng/mL and 10 ng/mL doxycycline provide low, matched expression levels of the A3A bait proteins.

Quantification of western blot signals from Chapter 5, Figure 5.11; signal intensities of bait protein bands were normalised to the loading control vinculin.

The newer BirA biotin ligase mutants have been shown to label proximal proteins in as little as 10 minutes (Branon et al., 2018). However, fusing miniTurbo to a bait protein can alter labelling kinetics and labelling time must be optimised for each bait protein. Too short labelling times can result in insufficient labelled material for affinity purification and mass spectrometry, and this hinders the ability to identify hits. On the other hand, excessive labelling times can increase non-specific biotinylation events and increase false positives. The stable cell lines expressing the three bait proteins were treated with biotin and cells were harvested at multiple different time points (Figure 6.15). While biotinylation was saturating around 8 hours treatment in all three cell lines, more biotinylation activity was seen at earlier time points in BFTC_A3A-mT (Figure 6.15 C) than BFTC_mT-A3A (Figure 6.15 B) cells, suggesting differences in ligase activity as a result of N- or C-terminal bait tagging. Two timepoints, 4 and 6 hours, were chosen for further optimisation.

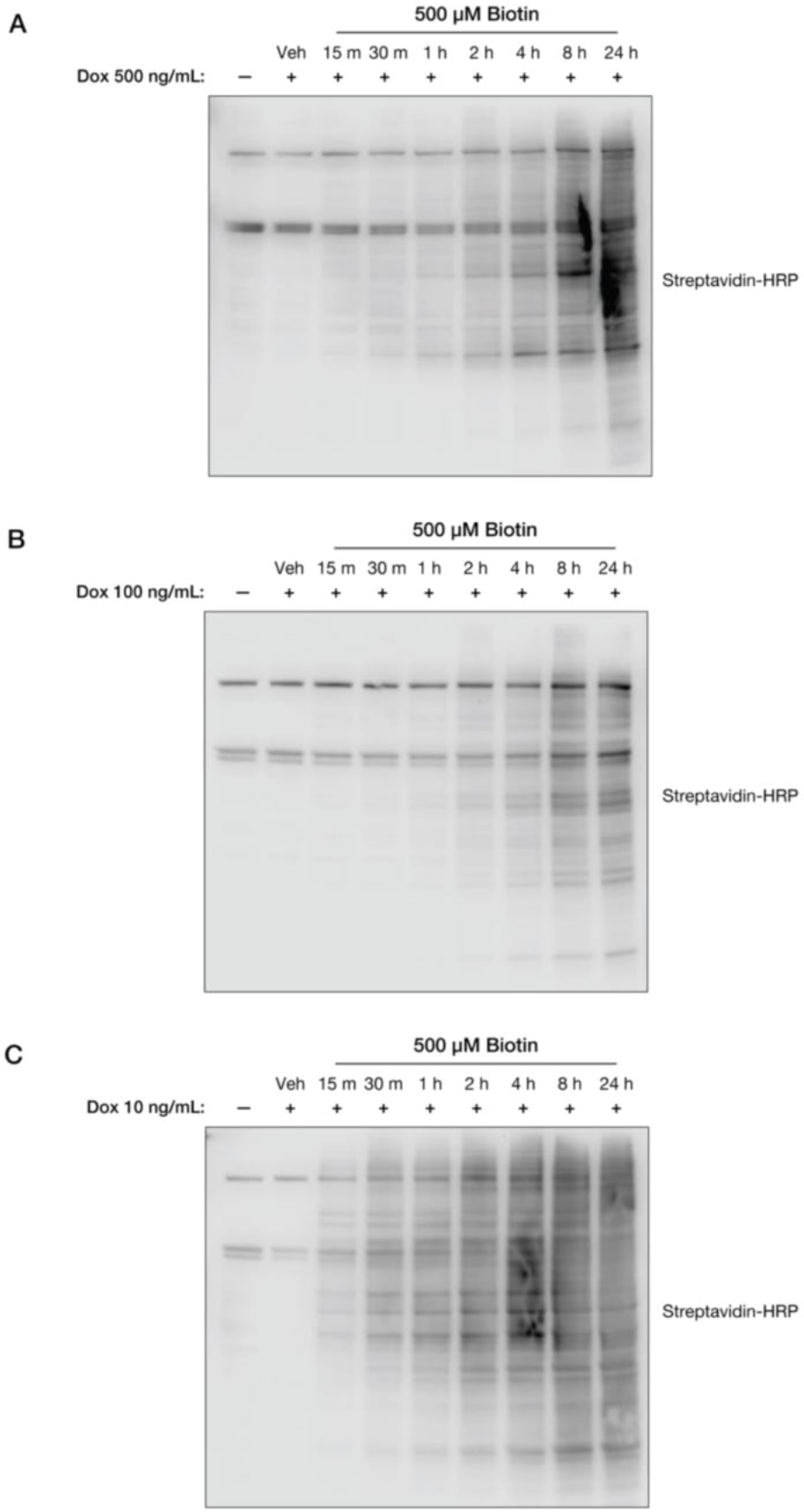


Figure 6.15 Biotinylation time optimisation of three stable cell lines expressing A3B-miniTurbo-V5 (A), 3xHA-miniTurbo-A3A (B) and A3A-miniTurbo-V5 (C).

The stable cell line pools were seeded in 6-well plates at 1.5 x 10^5 cells/well, the concentration of doxycycline (dox) indicated was added 24 hours later. After 48 hours dox treatment, 500 uM biotin was added and incubated for the indicated time before harvest for western blot. 30 µg protein was used for western blot; Streptavidin-HRP detects biotinylated proteins. Veh = cells treated with DMSO for 24 hours; m = minutes; h = hours. One biological repeat.

6.2.4.2 Affinity purification of biotinylated proteins confirms self-biotinylation of bait

The ideal way to check that the miniTurbo-tagged bait protein retains its endogenous functions, and that the biotin labelling time is sufficient, is to affinity purify biotinylated proteins and confirm biotinylation of a known interacting partner by western blot. While interacting partners of A3B have been identified in other cell line contexts, well-characterised interacting partners of A3A and A3B that can be used for this purpose are not known in UCC. However, it is well established that bait proteins are self-biotinylated directly by miniTurbo and proteins that oligomerise, like A3A (Bohn et al., 2015) and A3B (Li et al., 2014), will also be biotinylated when in complex with themselves. This self-biotinylation can be used to measure whether proximal biotinylation is occurring. Affinity purification of biotinylated proteins from cells treated with biotin for either 4 or 6 hours followed by western blot probing for either the V5 or HA tag revealed self-biotinylation and enrichment of all three bait proteins, and depletion of the bait in the unbound compared to the input fraction (Figure 6.16). The presence of a faint band when affinity purifying biotinylated proteins from BFTC_mT-A3A cells treated with biotin but not doxycycline, shows that there is some expression of this construct in the absence of doxycycline, and this may cause problems with downstream data analysis (Figure 6.16 B).

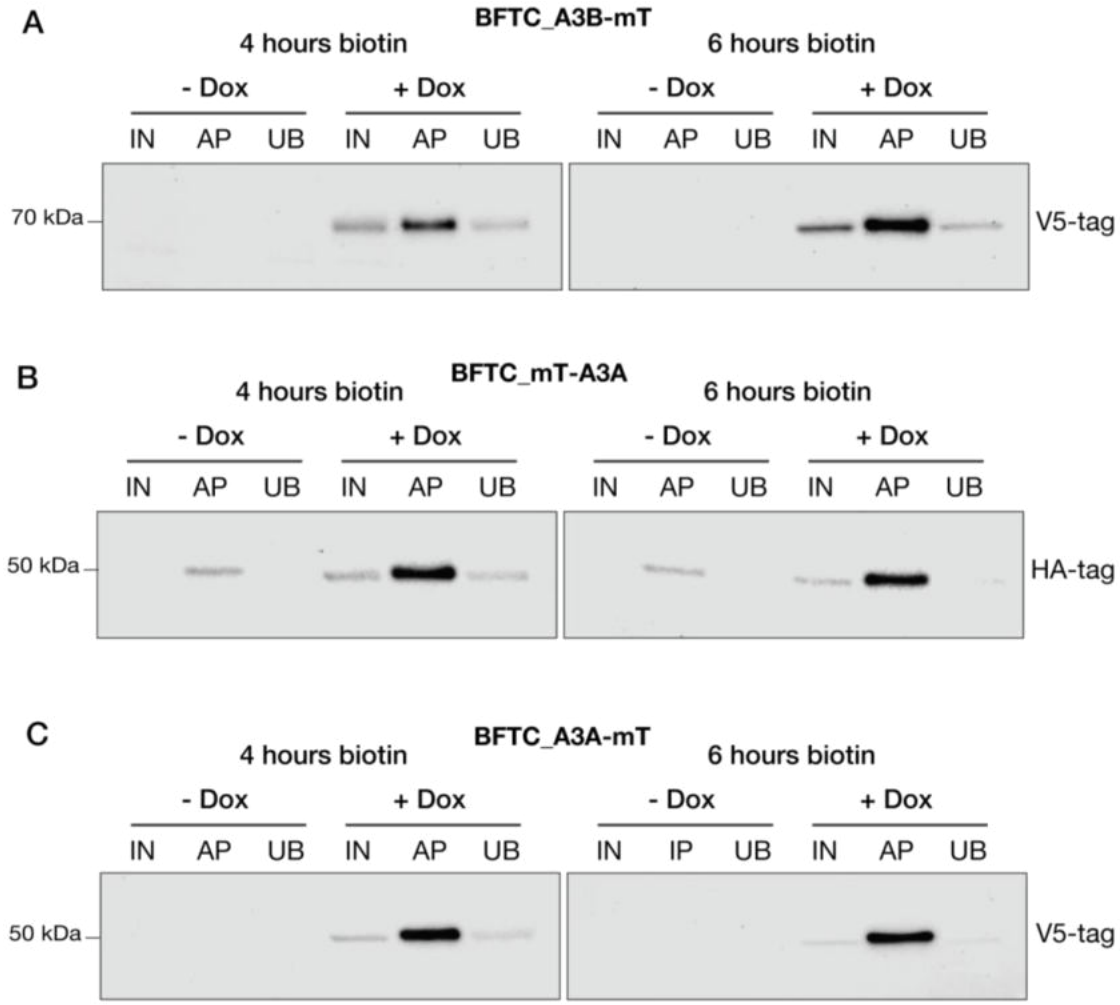


Figure 6.16 Affinity purification of biotinylated proteins using Strep-Tactin® Sepharose® beads shows self-biotinylation of the bait proteins in lysates.

The stable cell line pools BFTC_A3B-mT (A), BFTC_mT-A3A (B) and BFTC_A3A-mT (C) were seeded in 150 mm dishes (2.3 x 10^6 cells) and doxycycline (500 ng/mL (A), 100 ng/mL (B), 10 ng/mL (C)) was added 24 hours later. After 48 hours dox treatment, 500uM biotin was added and incubated for the indicated time before lysis for affinity purification and western blot. Biotinylated proteins were affinity purified by incubating 1 mg total protein with Strep-Tactin® Sepharose® beads. Beads were stringently washed before proteins were eluted in 2X LDS sample buffer and used for western blotting. V5-tag and HA-tag blotting detects tagged protein expression. IN = 2.5% input; AP = affinity purification; UB = unbound; same volume as input.

6.2.4.3 A3A's interactome is enriched in RNA binding proteins

A3A-mT, mT-A3A and A3B-mT were induced with doxycycline for 48 hours and biotinylation was accomplished by treatment with exogenous biotin for 4 and 6 hours. Biotinylated proteins were affinity purified using Strep-Tactin®

Sepharose® beads and the beads were sent to the ICR core proteomics facility for TMT mass spectrometry. Four peptides in the data set mapped to both A3A and A3B; thus, they could originate from either. One unique peptide for A3A was identified but there were no peptides unique to A3B. As A3B pulldown could not be confirmed, analysis could not be conducted. In addition, the number of identified proteins and reproducibility between the biological repeats for A3B-mT was low. This could be due to lower expression of the A3B-mT protein, insufficient biotinylated material, or pulldown efficiency. Initially, to identify putative interactors across the whole experiment, using data from both time points and both the N- and C-terminally tagged A3A baits, a linear model for each protein was generated using the values for A3A as the independent variable and each protein's correlation to A3A was calculated. The correlation of each of the 533 identified proteins to A3A is shown in Figure 6.17 A. Putative interactors were defined as those having a strong correlation (≥ 0.85) to A3A, and the 81 proteins identified are shown in (Figure 6.17 B). STRING network and gene ontology (GO) pathway analysis revealed a strong enrichment in terms associated with RNA metabolism, RNA binding and the ribonucleoprotein complex formation (Table 6.4). Many of the proteins (54 of the 81) were also enriched with the GO term "nucleus" (FDR = 7.62×10^{-6}).

Table 6.4 Gene ontology enrichment analysis (GO) of putative A3A interactors.

Top three significantly enriched pathways identified after STRING network analysis are shown for three GO classes (Szklarczyk et al., 2019).

GO term	Pathway Description	False Discovery Rate (FDR)
Biological Processes		
GO:0016071	mRNA metabolic process	3.16×10^{-21}
GO:0016070	RNA metabolic process	1.49×10^{-18}
GO:0090304	Nucleic acid metabolic process	1.18×10^{-15}
Molecular Function		
GO:0003723	RNA binding	5.85×10^{-32}
GO:0045296	Cadherin binding	7.80×10^{-24}
GO:0003676	Nucleic acid binding	1.33×10^{-15}

Cellular Component		
GOCC:1990904	Ribonucleoprotein complex	2.56×10^{-17}
GOCC:0005829	Cytosol	6.74×10^{-16}
GOCC:0043228	Extracellular exosome	1.66×10^{-13}

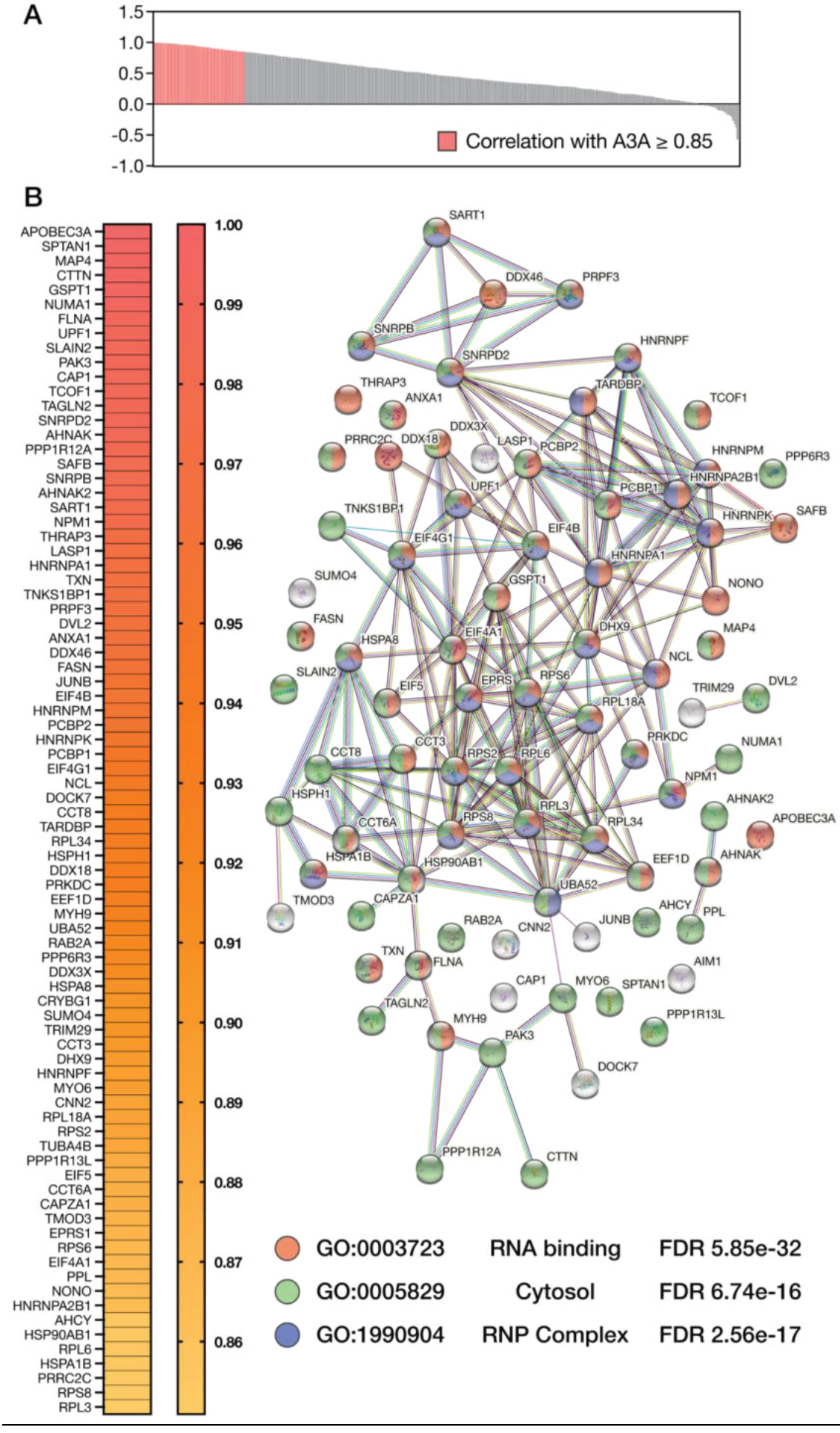
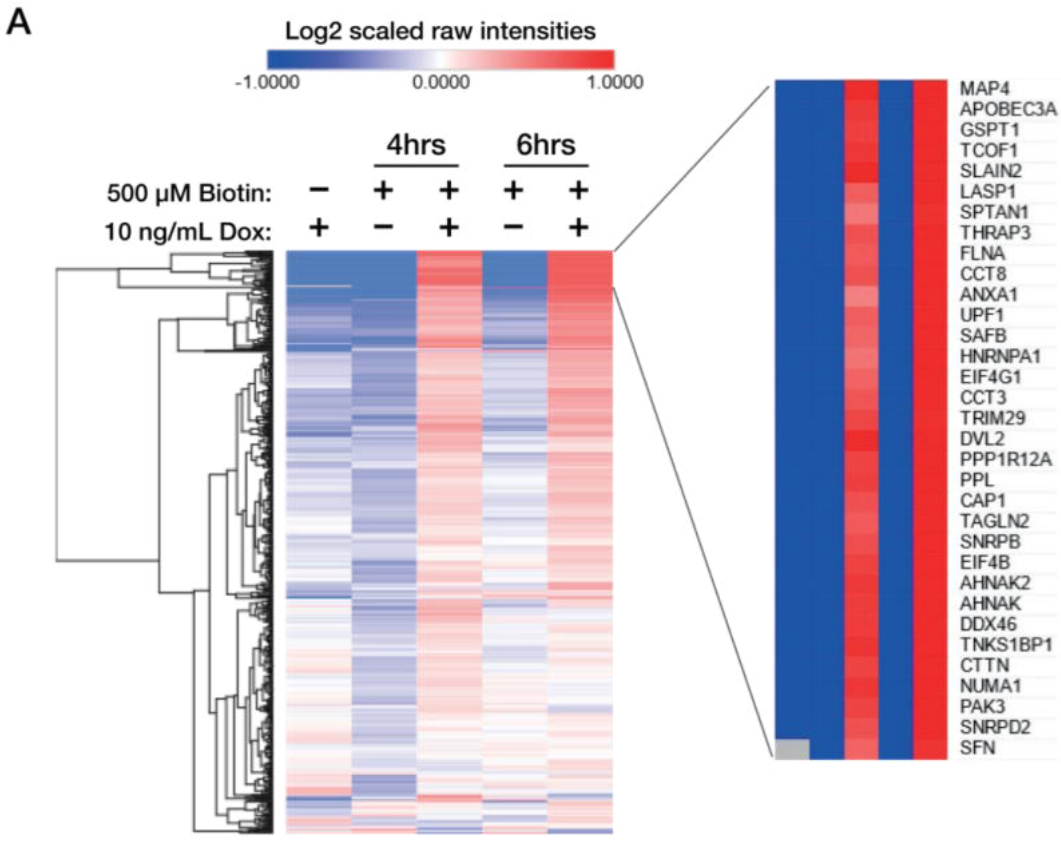

A
1.5
1.0
0.5
0.0
-0.5
-1.0
Correlation with A3A ≥ 0.85
B
APOBEC3A
SPTAN1
MAP4
CTTN
GSPT1
NUMA1
FLNA
UPF1
SLAIN2
PAK3
CAP1
TCOF1
TAGLN2
SNRPD2
AHNAK
PPP1R12A
SAFB
SNRPB
AHNAK2
SART1
NPM1
THRAP3
LASP1
HNRNPA1
TXN
TNKS1BP1
PRPF3
DVL2
ANXA1
DDX46
FASN
JUNB
EIF4B
HNRNPM
PCBP2
HNRNPK
PCBP1
EIF4G1
NCL
DOCK7
CCT8
TARDBP
RPL34
HSPH1
DDX18
PRKDC
EEF1D
MYH9
UBA52
RAB2A
PPP6R3
DDX3X
HSPA8
CRYBG1
SUMO4
TRIM29
CCT3
DHX9
HNRNPF
MYO6
CNN2
RPL18A
RPS2
TUBA4B
PPP1R13L
EIF5
CCT6A
CAPZA1
TMOD3
EPRS1
RPS6
EIF4A1
PPL
NONO
HNRNPA2B1
AHCY
HSP90AB1
RPL6
HSPA1B
PRRC2C
RPS8
RPL3
1.00
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.90
0.89
0.88
0.87
0.86
SART1
DDX46
PRPF3
SNRPB
HNRNPF
SNRPD2
THRAP3
TARDBP
ANXA1
TCOF1
DDX3X
PRRC2C
DDX18
LASP1
PCBP2
HNRNPM
PPP6R3
HNRNPA2B1
UPF1
PCBP1
TNKS1BP1
HNRNPK
SAFB
EIF4G1
EIF4B
HNRNPA1
SUMO4
GSPT1
NONO
FASN
EIF4A1
DHX9
MAP4
HSPA8
NCL
SLAIN2
EIF5
EPRS
RPS6
TRIM29
DVL2
RPL18A
PRKDC
NUMA1
CCT8
CCT3
RPS2
RPL6
NPM1
HSPH1
AHNAK2
RPL3
CCT6A
RPS8
RPL34
APOBEC3A
HSPA1B
HSP90AB1
EEF1D
AHNAK
UBA52
TMOD3
CAPZA1
AHCY
PPL
RAB2A
CNN2
JUNB
TXN
FLNA
AIM1
CAP1
MYO6
SPTAN1
TAGLN2
PPP1R13L
MYH9
PAK3
DOCK7
PPP1R12A
CTTN
GO:0003723 RNA binding FDR 5.85e-32
GO:0005829 Cytosol FDR 6.74e-16
GO:1990904 RNP Complex FDR 2.56e-17

Figure 6.17 A3A's interactome is enriched with RNA binding proteins.

(A) Correlation of the 533 biotinylated proteins identified across the whole experiment with two independent repeats by mass spectrometry with A3A. Proteins with a correlation ≥ 0.85 are likely strong interactors and are shown in red. (B) STRING network analysis of the 81 proteins identified as putative strong interactors with A3A (Szklarczyk et al., 2019). Statistical analysis performed by Dr Theo Roumeliotis; figures made myself. RNP, ribonucleoprotein.

The detection of mT-A3A in the pulldown in the uninduced condition (Figure 6.16 B) made statistical analysis of the changes between the induced and uninduced (-/+ Dox) conditions difficult. While proteins with a strong correlation to A3A could be identified, the $\log_2$ ratios +/- Dox were low. Therefore, further analysis was not conducted with this dataset. The $\log_2$ ratios +/- Dox for the A3A-mT dataset were used for a one sample t test and significant hits were defined as those with a $\log_2$ fold change ≥ 1 and $P < 0.05$. 20 proteins were identified as top hits for being A3A interactors (Figure 6.18 B). STRING network and GO pathway analysis, again, identified enrichment in terms associated with RNA binding and regulation.



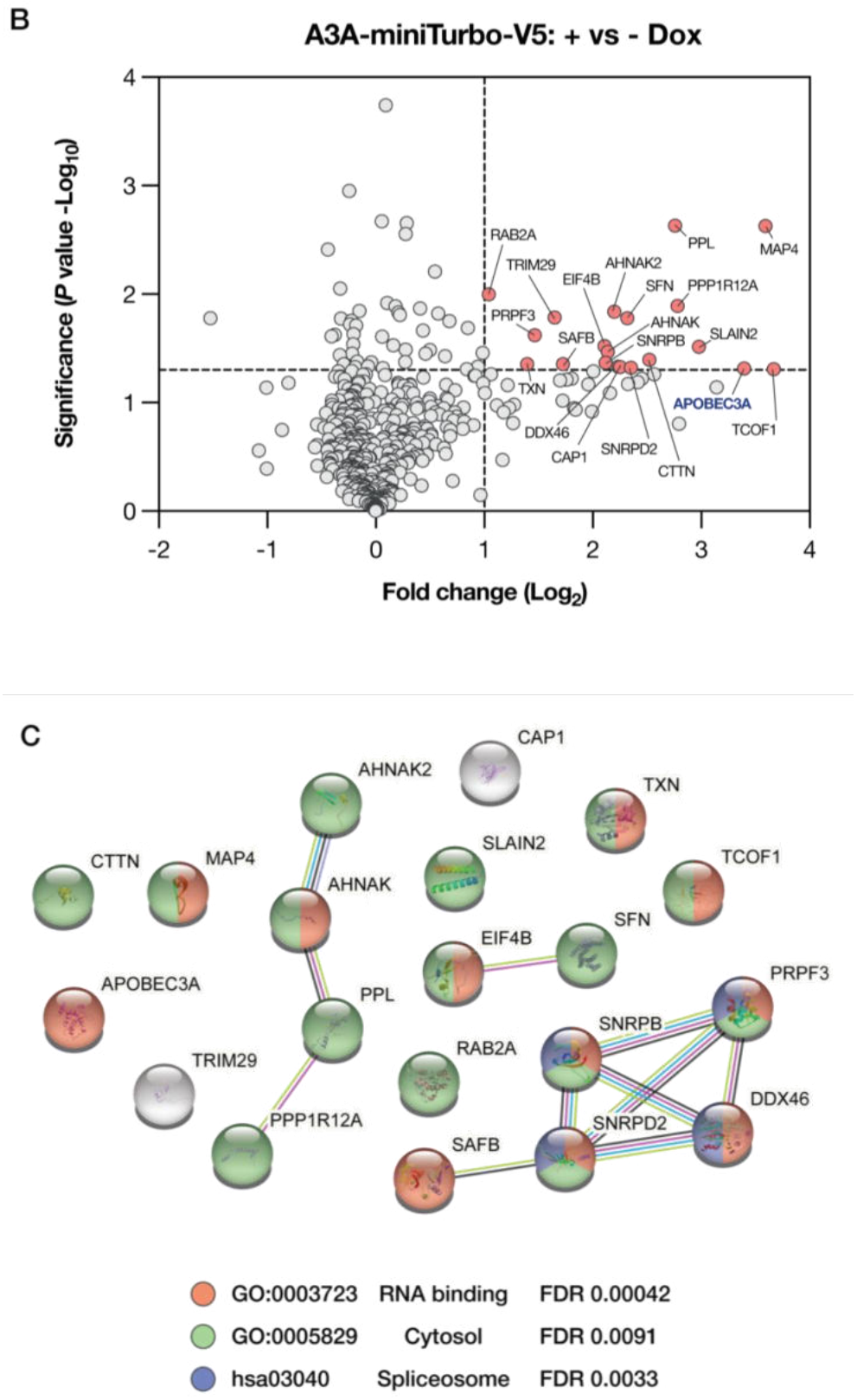


Figure 6.18 Top statistically significant A3A-miniTurbo-V5 interactors.

(A) Heat map showing raw signal intensities of the 533 identified proteins identified for the five treatment conditions. Time points show duration of biotin treatment after 48 hours induction with doxycycline (Dox). Scaled values for the two replicates were averaged and subjected to

hierarchical clustering with Euclidean distance. (B) Log_2 ratios +Dox vs -Dox were calculated and the 4- and 6-hour samples were used as replicates for a one sample t test. Significant hits are shown in red and were defined as proteins with a Log_2 fold change ≥ 1 with $p < 0.05$. (C) STRING network analysis of the 20 statistically significant hits showing top pathway enrichment (Szklarczyk et al., 2019). Statistical analysis performed by, and figure A made by Dr Theo Roumeliotis.

6.3 Discussion

To date, A3B deamination activity has only been shown to be required for the transcription of oestrogen-responsive genes in ER+ breast cancer (Periyasamy et al., 2015). It is not known whether this function is unique to breast cancer or whether A3B is required for nuclear receptor transcription in other cancer types, in particular, whether it is involved in ER or PPARγ gene set transcription in luminal UCC. In addition, there are also questions as to whether A3B is involved in transcription more generally and if A3A has similar functions. Unfortunately, attempts to generate A3A/A3B expressing stable cell lines in luminal models of UCC were unsuccessful and this study was conducted in a basal model of UCC, potentially limiting the ability to identify ER or PPARγ transcriptional activity. RNAseq of cells acutely exposed to elevated expression of A3A and A3B revealed only subtle changes in the transcriptome with the overall number of DE genes and their log_2 fold changes being small. Therefore, GSEA was not appropriate, and the identification of enriched gene sets that would facilitate assessment of transcription factor cooperation was not possible. Several basal subtype markers were downregulated in response to both A3A and A3B expression and this may suggest that expression in basal subtype cells can facilitate the emergence of a more luminal subtype. Considering that the A3 mutational signature is more strongly enriched in luminal subtypes of UCC, and the luminal molecular subtype of UCC is largely driven by the nuclear receptors ER and PPARγ, it is possible that A3A/B act with nuclear receptors to drive the luminal phenotype (Choi et al., 2014; Damrauer et al., 2014; Warrick et al., 2016). Overall, there was not strong evidence that A3A and A3B have a role in gene transcription in BFTC-905 cells. However, cells were only acutely exposed to elevated expression, and it would be interesting to investigate the general response to chronic upregulation and whether prolonged expression increases a luminal gene profile or phenotype. Additional work should also include a luminal UCC model enriched for ER and PPARγ regulon activity as this is the most appropriate model for investigating A3A/B's role in nuclear receptor gene transcription.

It is worth noting that there are several limitations to the RNAseq study that may have impacted the ability to identify novel transcription factor functions and the general response to elevated expression. PCA analysis of the samples revealed

that despite the stable cell line pools being derived from the same parental cell line stock, there are differences in the baseline transcriptional profile of the four stable cell line pools. This is likely a result of the pools being derived from a very small number of transfected parental cells that had integrated the transgene (approximately 0.01%), and represents existing clonal variation already present in the parental cell line. In support of this, a recent preprint has demonstrated that CRISPR clones are not isogenic when derived from the same parental stock (Panda et al., 2022). For example, in this study a chromosome 1 duplication event was present in 10% of the parental cells but only one of the resulting three clones had this CNV. This highlights the risks inherent in generating genetically modified cell lines from a heterogenous parental population as it results in variation due to existing heterogeneity already present in the parental stock. Although the work presented in this chapter did not use clones, integration events were rare, and the pools were generated from a very small number of parental cells that had integrated the transgene and represent different pre-existing parental clones. This pre-existing baseline variation adds a level of complexity when comparing the pools; the identified DE genes between wild-type and mutant expressing lines encompasses changes because of catalytic activity of the protein as well as the existing differences between the cell lines. Therefore, for future studies, it may be more appropriate to first clone the parental cell line, and then generate stable cells from this clonal parental line to minimise variation between them and facilitate comparisons.

Further, the stable cell line pools had to be transiently transfected with the Tet3G transactivator prior to doxycycline treatment to express the proteins of interest, which added a further level of complexity to the analysis. Comparisons between mock transfected (+ Dox) and plasmid transfected (+ Dox) groups revealed profound transcriptome changes. The number of changes when inducing expression of A3A or A3B were small and therefore, the large number of changes occurring are mostly a result of transient plasmid transfection. This demonstrates that plasmid transfection itself can drastically alter the transcriptome and care should be taken when analysing RNAseq datasets derived from experiments involving transient transfection. In all four cell line comparisons (+ Dox vs mock), GSEA revealed enrichment in pathways involved in the innate immune response and cellular stress (e.g., TNFα signalling, complement, interferon, TGFβ).

However, APOBEC proteins have known roles in the innate immune response and there is the possibility that transcriptome changes that occur because of A3 exposure were not identified as they were altered by transient transfection and present in both the induced and uninduced conditions. Therefore, to determine whether the study design limited the detection of APOBEC-induced transcriptional changes, stable cell lines that do not require plasmid transfection should be used for further studies.

An interesting finding of this work was the suggestion that both A3A and A3B edit mRNA. A variant base calling pipeline was used to look for evidence of mutations in the transcriptome in response to acute A3A and A3B expression and revealed that most mutations identified were C-to-T transitions (A3A, 17.1%; A3B, 19.5%). Considering the number of potential mutations that can arise, the identification of such a high proportion of one type is good evidence for A3 mRNA editing. However, to determine with certainty whether these identified C-to-T mutations are a result of A3A/B deamination activity, the sequence context surrounding the mutated C needs to be fully characterised. In addition, mRNA to cDNA conversion followed by Sanger Sequencing can be used to validate sites of particular interest. Therefore, while this dataset could be used for a preliminary investigation into whether there is evidence of RNA editing, the experimental design is not ideal for full investigation into the RNA editing activities of these enzymes. A lack of C-to-T events when expressing the mutant, catalytically inactive enzymes would increase confidence that the observed mutations upon acute exposure of wild-type A3A/B are true deamination events. However, the uninduced (−Dox) condition for the catalytically inactive mutants was not included in this study. The wild type and mutant conditions (+ Dox) could not be compared due to the observed changes as a result of subclonal variation. Any mutations identified could be due to SNPs already present in the genome rather than mRNA editing events. In addition, while the depth of sequencing used (30M reads per sample) is sufficient for DE analysis, the ability to identify low frequency editing events with this level of coverage is limited. As library generation for RNAseq requires cDNA synthesis and PCR amplification, there is the possibility that identified mutations are an artifact of library preparation. However, the high stringent threshold chosen for this analysis (DP > 50; AF > 10, i.e., 10% of the reads are edited with at least five reads containing the variant base call) helped

mitigate this risk. The number of editing events identified was much lower than those identified after exogenous expression of A3A in HEK293 cells (Sharma et al., 2017). However, the stringency used in this study was low (5% editing with at least one read containing a variant base call). Regardless, a higher depth of sequencing (>100M) would further increase confidence in the identified mutations. The work presented here was conducted using a stable cell line pool and there is likely heterogenous expression of, and response to, A3A/B. In addition, A3 mutagenesis appears to be a stochastic process with different mutations occurring in individual cells (Akre et al., 2016). During stress, in particular drug treatment, some cells survive while others die, and therefore single-cell sequencing may be better placed to interrogate whether the level of RNA editing or cellular response to elevated A3A/B varies between cells and if this affects which cells have a survival benefit under conditions of stress. Interestingly, A-to-G transitions (A3A, 12.6%; A3B, 12.5%) suggestive of ADAR deamination activity were also identified after acute A3A/B exposure. This was unexpected and raises the question as to whether A3 activity stimulates ADAR activity. Interestingly, expression of A3C and ADARp150 correlate in stem cells from patients with myeloproliferative neoplasms and the two deaminases have been shown to co-localise and co-immunoprecipitate (co-IP) suggesting A3s and ADAR may work together to edit RNA (Jiang et al., 2021).

To further investigate the acute response to A3B exposure, expression of A3B-V5 was induced in BFTC-905 cells and the effect on growth and drug sensitivity was assessed. A3B overexpression has been shown to induce a proliferation defect resulting from DNA damage and cell cycle arrest in the non-tumourigenic line, HEK293 (Akre et al., 2016; Burns, Lackey, et al., 2013; Nikkilä et al., 2017). A3B expression is also detrimental to tumour initiation in mutant EGFR mouse models of lung cancer suggesting that A3B DNA damage leads to increased tumour death early in development (Mayekar et al., 2020). Studies investigating how tumour cells respond to increased A3B expression are lacking, but the high levels observed in many human cancers suggests that cells can tolerate A3B-induced DNA damage and/or that they regulate the enzymes mutagenic potential. The proliferation defect observed in HEK293 cells exposed to A3B overexpression was rescued by p53 inactivation (Nikkilä et al., 2017) but while BFTC-905 cells have a mutant p53 response, this is unlikely the sole reason that

cells tolerate elevated expression. Studies with NSCLC cell lines in our group showed neither p53 mutant nor wild-type cells had a marked proliferation defect in response to A3B overexpression (Dr Caitlin McCarthy, ICR), and elevated baseline *A3B* expression and induction in response to chemotherapy drugs occurs in both wild-type and mutant p53 UCC cell lines (Chapter 3). Overexpression of A3B in p53-defective HEK293 cells renders them sensitive to DDR inhibition (Nikkilä et al., 2017) but acute expression of A3B in BFTC-905 UCC cells has no effect on sensitivity to the DDR inhibitors, KU-60019 (ATMi) or AZD6738 (ATRi). While markers of DNA damage were not assessed in this work, the lack of a proliferation defect or increased sensitivity to DDR inhibition suggests that the cells may be negatively regulating A3B, inhibiting its ability to generate DNA damage, or that damage is rapidly repaired and tolerated. In support of this, in contrast to HEK293 cells, overexpression of A3B in NSCLC cell lines does not induce markers of DNA damage or RS, or sensitise them to DDR inhibition (Dr Caitlin McCarthy, ICR). These findings, taken with the subtle transcriptome changes observed, support a hypothesis where tumour cells can tolerate elevated levels, most likely through regulating the mutagenic activity of the enzyme. Phosphorylation of A3B by protein kinase A (PKA) inhibits its deamination activity, demonstrating that regulatory post-translational modifications are used by cells to modulate activity (Matsumoto et al., 2019). Further investigation into markers of DNA damage and RS in UCC cells in response to elevated expression and the post-translational regulation of A3B are required. The effect on growth and drug sensitivity in response to A3A expression in BFTC-905 cells was not conducted due to time constraints and further work should address this. However, the subtle transcriptome changes observed in response to elevated A3A expression suggest that BFTC-905 cells also tolerate A3A. While many other studies have used A3A and A3B tagged with GFP, this study used A3A and A3B tagged with a small epitope to try and minimise disruption of normal protein function, expressed at elevated, but not massively overexpressed levels. Therefore, it is possible that some of the differences reported are because of differences in expression levels, and how the proteins are regulated and localised. Therefore, it is likely that results in this study more accurately reflect the cellular response to elevated levels of A3A and A3B, and

the minimal effects observed are due to regulation of the enzymes, limitation of their genotoxic capacity and/or tolerance to any damage generated.

To further investigate the function of A3A and A3B in UCC, work aimed to identify their interacting partners using the proximity-labelling technique, BioID. A3B C-terminally tagged, and A3A N- and C-terminally tagged, with the biotin-ligase miniTurbo were expressed in BFTC-905 cells and biotinylated proteins were affinity purified and identified using mass spec. Unfortunately, there were several technical issues with A3B and interacting partners could not be identified. While biotinylation activity was observed at the chosen time points and self-biotinylation of the bait was confirmed, no unique A3B peptides were identified in the dataset. As the presence of A3B could not be confirmed and the signal intensities of the detected proteins were low, no further analysis on this dataset was conducted. The lack of an N-terminally tagged A3B bait was also a limiting factor. Therefore, further work should attempt to make an N-terminally tagged A3B bait and optimise conditions required for successful identification of A3B's interacting partners. On the other hand, A3A was uniquely identified in the data set and putative A3A interactors were identified. Analysis of biotinylated proteins identified with both A3A baits at two timepoints revealed numerous proteins correlated with A3A; 81 proteins have a strong correlation with A3A and are likely interactors. STRING network and GO pathway enrichment analysis revealed significant enrichment in pathways involved in RNA binding, mRNA metabolism and ribonucleoprotein complexes. These findings are particularly interesting considering the suggestion of RNA editing activity and supports further investigation into this potential function. While there was evidence of increased RNA editing by ADAR after expression of A3A, ADAR was not identified as a putative interactor of A3A using BioID. This suggests that A3A does not directly interact with and regulate ADAR but does not exclude the possibility that A3A is in some way involved in ADAR regulation, and further studies are warranted.

The putative interactors were also enriched in the GO terms "cytosol" and "nucleus" suggesting that A3A is not exclusively located in the cytoplasm and has access to genomic DNA. *PRKDC* encoding the catalytic subunit of DNA-PK, tripartite motif protein 29 (TRIM29) and nucleophosmin 1 (NPM1), were identified as putative interactors, proteins with known roles in the DDR and damage repair. DNA-PK is recruited to the ends of DSBs and is key in NHEJ repair (Blackford &

Jackson, 2017), and TRIM29 interacts with DNA-PK and is suggested to act as a scaffold protein to recruit DNA repair proteins to sites of damage (Masuda et al., 2015). NPM1 is a chaperone protein and the NPM family have important roles in modulation of genome stability, replication and transcription (Box et al., 2016). NPM1 is recruited to DSBs, colocalising with key repair proteins such as BRCA1, BARD1, γ-H2AX (Koike et al., 2010; Sato et al., 2004) and BRCA2 (Wang et al., 2011). NPM1 also interacts with APE1, which is required for further processing of abasic sites in BER, and this interaction is thought to functionally regulate APE1 activity (Vascotto et al., 2009; Vascotto et al., 2014). Together these results suggest that A3A does indeed have access to the nucleus and that A3A acts at sites of DNA damage, in particular, DSBs. This is consistent with the hypothesis that DSBs provide ssDNA substrates for A3 deamination and the association of the mutational signature with breakpoints (Nik-Zainal et al., 2012).

One particularly interesting finding is the identification of microtubule associated protein 4 (MAP4) and scaffold attachment factor B (SAFB) as putative A3A interactors. MAP4 and SAFB have been identified as A3B interactors in a recent preprint (McCann et al., 2021) and this suggests that A3A and A3B, have to some extent, an overlapping interactome and is further evidence that they may have overlapping functions. In this preprint, the authors found 15 proteins shared between A3B and the S9.6 monoclonal antibody (mAb) that binds strongly to RNA/DNA hybrid (R-loop) structures. They demonstrated that A3B interacts with R-loops and R-loop-associated proteins and plays a key role in R-loop homeostasis by deaminating the displaced ssDNA strand, promoting R-loop resolution. MAP4 and SAFB were identified in both the A3B and S9.6 mass spec datasets and this suggests that, like A3B, A3A may also play a role in R-loop resolution. SAFB, treacle ribosome biogenesis factor 1 (TCOF1) and TAR DNA binding protein (TARDBP) act at transcription complexes, and this is further evidence of A3A activity at transcription-associated R-loops. DExH-box helicase 9 (DHX9) was also identified as a putative A3A interactor. DHX9 is an RNA helicase involved in several cellular processes, including transcription and replication, that acts to maintain genome stability (Gulliver et al., 2020). It preferentially acts at R-loops and G-quadraplexes to resolve secondary structures that may hinder replication or transcription (Chakraborty & Grosse, 2011) and was identified as the top RNA/DNA hybrid interacting candidate in a

pulldown mass spec experiment using the S9.6 mAb (Cristini et al., 2018). Further overexpression and knockdown studies will be able to determine if A3A activity effects R-loop abundance. In addition, the interaction of MAP4, SAFB and DHX9 with A3A should be confirmed using an orthogonal approach and their role in R-loop homeostasis should be evaluated.

A key aim of this study was to use BioID in combination with RNAseq to determine whether A3A and A3B have a role in nuclear receptor transcriptional regulation. The evidence for this in the RNAseq data set was minimal and the nuclear receptors ER or PPARγ, were not identified as putative interacting partners in BFTC-905 cells. While several proteins involved in transcriptional regulation were identified, the involvement of many of these in R-loop biology suggests that A3A's role in transcription is to ensure transcription can proceed unhindered through resolving R-loops rather than by activating gene transcription as seen for A3B in breast cancer (Periyasamy et al., 2015). Overall, this data, taken with the enrichment of RNA binding proteins and those associated with R-loops, supports a hypothesis that A3A is not directly involved in transcriptional initiation and rather, mediates adaptability via RNA modulation and contributes to genomic stability by R-loop interactions.

It is worth noting that there were some limitations to the BioID experiment that may have impacted the results. BioID experiments traditionally include controls that were not used in this study and include either miniTurbo alone or miniTurbo fused to an unrelated protein, often GFP or a protein localised to the same cellular compartment as the protein of interest. The use of a non-related bait localised to the same cellular compartment requires prior knowledge of the localisation of the proteins of interest, which is not always the case. The localisation of A3A is still disputed and was a question under investigation in this work; as such, this control was not appropriate. Attempts to generate stable cell lines expressing miniTurbo tagged GFP were unsuccessful and due to time constraints, the project had to progress without this control. While not including a GFP or miniTurbo only control limits the ability to identify false-interactors due to interaction with either miniTurbo or the solid support used for affinity purification, there are some benefits. GFP is a foreign protein, is known to non-specifically interact with numerous proteins and can be toxic. Including GFP fused to miniTurbo, while decreasing likelihood of identifying false positives, increases the likelihood that

true interacting proteins will be excluded from the analysis due to non-specific interactions and increases false negatives. To minimise the identification of non-specific binding of proteins to the solid support, in this case agarose beads, stringent washing with detergents was conducted and uninduced controls were included. The identification of the A3B interacting partners, MAP4 and SAFB, and the known R-loop associated protein, DHX9, in this study improves confidence that the interacting partners identified are indeed real. Nonetheless, BioID is a broad screen for putative interacting partners and hits should always be confirmed by other methods.

In conclusion, minimal transcriptome changes were observed with elevated A3A and A3B expression, and A3B does not induce a proliferation defect or alter sensitivity to two chemotherapy or two DDR inhibitor drugs in BFTC-905 UCC cells. Altogether, this suggests that tumour cells tolerate acute elevated expression and A3B does not play a role in DNA damage repair. Increased numbers of C-to-T changes were detected in mRNA of cells transiently exposed to elevated A3A and A3B expression, suggesting they may edit RNA. This was supported by the extensive number of identified A3A interactors that have known roles in RNA binding and metabolism. There was no compelling evidence for a transcriptional function of A3A and A3B from the RNAseq experiments and transcriptionally active nuclear receptors, such as ER, were not identified as A3A interactors supporting the conclusion that A3A and A3B likely do not have a transcriptional function in basal subtype UCC cells. While A3B's interacting partners were not identified, several novel putative interactors of A3A were identified that suggest that A3A is localised to sites of DNA damage and, like A3B, interacts with and acts on R-loop structures.

Chapter 7

General Discussion

Chapter 7 General discussion

7.1 Introduction

The A3 family of cytosine deaminases play a key role in the innate immune response primarily by restricting viral replication through hypermutation of viral genomes (Harris & Dudley, 2015). However, there is now evidence that their off-target activity is a major source of mutation and genomic instability in many human cancers (Swanton et al., 2015). Two mutational signatures characterised by C-to-T and C-to-G mutations within 5'TC motifs, have been identified in two-thirds of all human cancers and are attributed to A3-mediated deamination and subsequent processing of the resulting uracil lesions (Alexandrov et al., 2020). Initial studies suggested that out of the six A3 family members that target cytosines within 5'TC motifs, A3B was the most likely candidate for generating the signature in human cancers. A3B is often elevated in tumours relative to normal tissue, its expression correlates well with A3 mutational load in numerous tumour types and is the only family member known to be constitutively localised to the nucleus due to an N-terminal NLS (Burns, Lackey, et al., 2013; Burns, Temiz, et al., 2013; Lackey et al., 2012; Lackey et al., 2013; Roberts et al., 2013; Swanton et al., 2015). Despite lacking an NLS, A3A is small enough to passively enter the nucleus and has a cell-wide distribution (Lackey et al., 2013). Overexpression of the two family members in yeast revealed that the signatures generated by them are distinct (Chan et al., 2015). The A3A-like 5'YTCA signature is more common in human cancers and A3A-like tumours have over ten times more mutations than those with A3B 5'RTCA signatures (Alexandrov et al., 2020; Chan et al., 2015). However, A3A is rarely upregulated in tumours, and this raises the question of how A3A can be the predominant mutagen when there is a disconnect between its expression and the mutational signature? One theory, that is now gaining experimental evidence, is that the mutational signature is generated from transient bursts in A3A expression (Petljak et al., 2019). If such a burst in expression is not occurring at the time of tumour sampling, the mutational signature, evidence of prior expression and activity, will be detected but not expression itself, providing an explanation for the disconnect. The work by Petljak et al., (2019) suggests that the constitutive high expression of *A3B* seen in many tumour types is not responsible for a constant rate of mutation and

cells have mechanisms to limit the genotoxic effects of constitutively high *A3B.* Interestingly, a recent study showed that differential splicing of A3B generates a protein lacking exon 5 that is catalytically inactive and non-mutagenic (Rouf Banday et al., 2021), demonstrating that post-transcriptional and post-translational modifications can modulate mutagenic activity and facilitate tolerance to high levels of expression. Tumour cells in patients are exposed to a wide range of cellular stresses and it is possible these stresses are driving bursts in expression. Indeed, A3 expression and activity has been shown to be induced in response to hypoxia, cellular crowding, and chemotherapy treatment (Alqassim et al., 2021; Bader et al., 2021; Kanu et al., 2016; Oh et al., 2021; Periyasamy et al., 2021; Sharma et al., 2017; Sharma et al., 2016; Sharma et al., 2015; Yamazaki et al., 2020). These findings have strong clinical implications as drug treatment itself could be fuelling heterogeneity and emergence of drug resistance.

The overall aim of this thesis was to specifically investigate the regulation of, and cellular response to, the most likely candidates for generating the mutational signatures, A3A and A3B, in UCC. UCC has the strongest A3 mutational signature of all tumour types analysed, and it is identified in both early stage, late stage, and chemotherapy resistant tumours suggesting key roles in initiation, progression, and drug resistance. Therefore, UCC is an interesting and unique disease context for the study of A3 enzymes. Little work has been done to investigate the induction of A3s in response to chemotherapy and the mechanism of induction in this tumour type, so a major aim of this thesis was to thoroughly characterise induction, identify putative regulatory mechanisms and determine whether induction can be modulated with clinically available drugs. Induction of A3s in response to chemotherapy is transient, and this transient drug-induced expression is likely a driver of episodic bursts in expression and signature accumulation in patients. However, a major question remains: why do tumour cells upregulate an endogenous mutagen under conditions of stress? This feeds into the final aim of the thesis which was to characterise the cellular response to acute A3A/B exposure to give insight into the role of these enzymes during periods of acute upregulation.

7.2 Regulation of *A3A* and *A3B* in UCC

Numerous studies have now demonstrated that chemotherapy drugs (Kanu et al., 2016; Oh et al., 2021; Periyasamy et al., 2021; Yamazaki et al., 2020), and more recently, some targeted therapies (Isozaki et al., 2021; Mayekar et al., 2020), can induce expression of *A3A* and/or *A3B* in several tumour contexts. In this thesis, *A3A* and *A3B* were found to be upregulated in response to numerous chemotherapy drugs with a range of mechanisms of action, in UCC cell lines representing NMI and MI disease, luminal and basal molecular subtypes, and with functional p53, demonstrating that these factors are not differentiators for the ability of cells to induce expression (Chapter 3). This is consistent with observations in UCC patient samples, where A3 expression and mutagenesis is widespread. Induction, where observed, was minimal in cells treated with targeted therapies such as erlotinib, lapatinib and infigratinib, suggesting that induction is primarily a response to DNA damage rather than anticancer agents more generally. Induction in response to chemotherapy treatment was not observed in a cell line model of normal urothelium, demonstrating that there are factors at play that prevent expression of these endogenous mutagens under conditions of stress in healthy urothelium. This contrasts with the breast, where induction is observed in several models of normal breast tissue and suggests that regulation of *A3B* is, to some extent, tissue specific (Kanu et al., 2016; Periyasamy et al., 2021). Subsequent investigation into the mechanism of induction in response to gemcitabine treatment further highlighted both similarities and differences in regulation between bladder and breast (Chapter 4). When this work began, NF-κB signalling had recently been implicated in the regulation of *A3B* but there was conflicting evidence as to whether this involved the canonical (Maruyama et al., 2016) or alternative (Leonard et al., 2015) branches. This, together with the knowledge that anticancer agents often activate NF-κB signalling, suggested that the observed *A3A/B* induction was mediated by NF-κB. Another group also had the same hypothesis and has recently published their work (Periyasamy et al., 2021). While the authors found that *A3B* induction in response to chemotherapy drug treatment was dependent on a DNA-damage induced NF-κB signalling pathway in breast cancer cell lines, this was unexpectedly not the case in UCC. Neither small molecule inhibition of NF-κB signalling or knockdown of two key transcription factors, *RELA* and *RELB*,

attenuated induction of *A3B* in response to gemcitabine. These findings are particularly interesting as bladder cancers subtypes are named based on their similarity to those identified in breast yet, *A3B* regulation is different.

An additional novel finding is that despite their recent common evolutionary origin and homology, *A3A* and *A3B* are differentially regulated (Chapter 4). While neither canonical nor the alternative NF-κB signalling pathways were required for induction of *A3B,* they are both involved in the induction of *A3A*. Small molecule inhibition of NF-κB and knockdown of both the canonical, *RELA*, and alternative, *RELB,* transcription factors attenuated induction in response to gemcitabine. A further interesting observation is that gemcitabine also induces *RELB,* and this is attenuated by *RELA* knockdown, suggesting that *A3A* induction initially requires rapid activation of canonical signalling followed by transcription of *RELB* and subsequent activation of the delayed, alternative pathway.

The RS-inducing drug, gemcitabine, was found to be the most robust inducer of expression corroborating findings in the literature that RS is a major driver of A3B expression, activity and subsequent mutagenesis (Kanu et al., 2016; Periyasamy et al., 2021; Venkatesan et al., 2021; Yamazaki et al., 2020). Induction of both *A3A* and *A3B* was seen at 48 hours, even after only two hours gemcitabine treatment demonstrating that early DNA damage is the key initiating event and this, with the link to NF-κB activation, stimulated investigation into the DDR. Gemcitabine primarily induces RS and activates the ATR/CHK1 signalling cascade suggesting ATR is a primary mediator of induction as seen in breast cancer cell lines (Kanu et al., 2016). Indeed, *A3B* induction is attenuated by small molecule inhibition of ATR demonstrating that ATR activation is required for *A3B* expression during RS. In contrast to other work, ATM is not required and this may be a result of the type of DNA lesion generated by the different drugs investigated (Periyasamy et al., 2021). While induction of *A3B* is attenuated with ATRi, some potentiation of *A3A* was seen. ATR is known to negatively regulate NF-κB activation in response to DNA damage and is likely that inhibition of ATR prevents this negative regulation, resulting in increased expression. These findings are consistent with those recently published (Oh et al., 2021).

Despite their differential regulation (*A3A* by NF-κB and *A3B* by RS/ATR), attenuation of both family members can be achieved with either PKCi or RTKi

treatment, and this has interesting clinical implications that are discussed in detail later. RTK signalling feeds into the NF-κB pathway and as RTKi treatment attenuates induction of *TNFα* as well as *A3A,* attenuation is likely occurring via inhibition of downstream NF-κB signalling. PKC is one of many key signal transducers downstream of RTKs that feeds into NF-κB, and it is currently unknown whether PKC is the key player or whether others such as AKT are also required. In any case, the likelihood is that the attenuation is ultimately a result of inhibition of NF-κB activation. On the other hand, in the case of *A3B*, the downstream pathway largely remains uncharacterised, although PKC appears to be an important player, the downstream transcription factors that drive *A3B* expression are currently unknown. Several anticancer agents, including gemcitabine, are known to either activate or upregulate RTKs (Benhar et al., 2002; Chun et al., 2006; Furugaki et al., 2010; Kan, Koido, Okamoto, Hayashi, Ito, Kamata, Komita, Nagasaki, et al., 2015; Miyabayashi et al., 2013; Van Schaeybroeck et al., 2005) and oncogenic signalling is a well-established cause of RS due to dysregulated origin firing, nucleotide depletion and transcription-replication collisions (Primo & Teixeira, 2019). Therefore, it is also possible that RTK inhibition is counteracting gemcitabine activation and *A3B* is attenuated because of reduced downstream RS.

A major question exists as to why cells would upregulate these endogenous mutagens under conditions of stress. One possibility is that they stimulate DNA damage repair through generating abasic sites and recruitment of repair enzymes. Lymphoid cells expressing A3G are less sensitive to IR due to more efficient DSB repair, and it is proposed that A3G plays a key role in this process (Botvinnik et al., 2021; Nowarski et al., 2012). However, mechanistic evidence for how A3G participates in repair and overcomes its strong cytoplasmic retention signal to accumulate in the nucleus is limited, and conflicts with other studies showing that A3G causes DNA damage (Talluri et al., 2021). While the induction of *A3B* in response to ATR activation suggests it may play a role in repair, *A3A* expression is repressed by ATR, and taken with findings of others showing that A3A and A3B are genotoxic (Akre et al., 2016; Burns, Lackey, et al., 2013; Lackey et al., 2013; Nikkilä et al., 2017), a direct role in DNA damage repair for these family members seems unlikely. This is further supported by other findings presented in this thesis showing that elevated A3B expression in BFTC-905 cells

that robustly induce expression in response to chemotherapy, does not alter BFTC-905 sensitivity to two chemotherapy drugs (Chapter 6). The extensive A3 mutational signatures identified in human cancers suggest that the upregulation of *A3A* and/or *A3B* has a selective benefit to tumour cells and experimental evidence for this is now emerging. A3B activity is linked to tamoxifen resistance in ER+ breast cancer (Law et al., 2016), resistance to oncolytic viral therapy (Huff et al., 2018), cisplatin-resistance in ovarian cancer (Periyasamy et al., 2021), and both A3A and A3B are proposed to play a role in RTKi resistance in mouse models of lung cancer (Isozaki et al., 2021; Mayekar et al., 2020). The common theme is that all these stresses induce expression. Therefore, it is possible that, analogous to stress-induced mutagenesis (SIM) in bacterial cells, that tumour cells under conditions of stress increase their ability to adapt by increasing mutation rates through induction of *A3A/B*. Interestingly, an increase in mutability in response to targeted therapy is seen in drug-tolerant persister cells (DTPs) where error-prone polymerases are upregulated, and MMR and HR pathways are downregulated (Russo et al., 2019). Further investigation into whether A3A/B are part of a stress-response and involved in the drug-tolerant persister cell phenotype is warranted.

7.3 Cellular response to transient exposure

Results presented in Chapter 3 demonstrated that *A3A* and *A3B* are transiently upregulated during drug-induced stress, but little is known about their role in cancer and specifically, during these transient exposures. A3B deamination is required for transcription of oestrogen-responsive genes in ER+ breast cancer, and this study was the first demonstration of a novel, innate immune response-independent function (Periyasamy et al., 2015). To date, this transcriptional function of A3B has only been demonstrated in breast cancer, but it is possible it may also occur in other tumour types with ER-dependent transcriptional profiles or, that A3B may cooperate with other nuclear receptors to drive gene transcription. A3A and the C-terminus of A3B and are highly homologous, raising the possibility that A3A may also have yet undiscovered transcriptional functions. Bladder cancer subtypes are related to those identified in breast cancer, with ER gene transcriptional profiles enriched in the luminal subtype, suggesting A3B may also drive ER gene transcription in UCC. Therefore, a main aim of this study was

to use overexpression, RNAseq and proteomic studies to determine whether there is evidence of a transcriptional function of A3A and/or A3B in UCC cells. These studies required generating appropriate cell line models expressing V5-tagged A3A and A3B for the RNAseq studies and biotin-ligase, miniTurbo-tagged A3A and B for BioID identification of interacting partners. Due to concerns with viral transduction methods of generating cell lines (discussed in detail in Chapter 5), knock-in of the transgenes into the GSH locus, *AAVS1*, was attempted. This work was challenging, laborious, and required extensive troubleshooting to overcome toxicity and expression issues. Nonetheless, A3A, B, and their corresponding catalytically inactive mutants tagged with V5 were successfully integrated into the *AAVS1* locus and the resulting BFTC-905 (basal subtype) stable cell lines were used for several studies. An expressing luminal cell line was not successfully made, and this limited the ability of the study to identify a putative ER-dependent transcriptional function. RNAseq of BFTC-905 cells transiently exposed to A3A or A3B for three days revealed only subtle transcriptomic changes and there was no striking evidence of a transcriptional co-factor function for either family member in basal subtype UCC. A3A and A3G are RNA editing enzymes (Alqassim et al., 2021; Sharma et al., 2017; Sharma et al., 2016; Sharma et al., 2015) but an RNA editing function of A3B has not been described and it is not known if A3A edits RNA in UCC. A novel finding of this thesis was that there was an enrichment in C-to-T changes in mRNA upon doxycycline induction of both A3A and A3B. There was also an increase in ADAR-mediated A-to-G mutations suggesting that A3s may co-regulate ADAR activity. These findings are particularly interesting considering that Class 2a NMI-UCC tumours that are enriched for the A3-mutational signatures are also enriched in RNA editing signatures (Lindskrog et al., 2021), and suggests that A3A/B activity may be contributing to UCC heterogeneity via both genetic alteration and non-heritable RNA editing. While it is beyond the scope of this project to fully characterise the sequence context of the mutations to confirm the observed C-to-T changes were occurring within the A3 motif, and the experimental system and sequencing depth limited this study, the findings do give preliminary evidence for A3A and A3B RNA editing during transient exposure in UCC cells and provides rational for further, more in depth, studies.

Acute elevated expression of A3B in BFTC-905 also had no effect on growth or sensitivity to two chemotherapy drugs (gemcitabine and cisplatin) and two DDR inhibitors (ATRi and ATMi), consistent with results in NSCLC lines in our group (Dr Caitlin McCarthy, ICR). These results suggest that acute exposure to A3A and A3B has minimal cellular effects and that elevated levels are tolerated well by tumour cells. The lack of effect on acute drug sensitivity suggests that these enzymes are not upregulated to help with direct repair of chemotherapy-induced DNA lesions. Instead, it is more likely that they facilitate adaptability via RNA editing, and that drug resistance is mediated by accumulation of mutations, changing drug sensitivity over time, rather than by transcriptional regulation. In addition, elevated A3B expression did not alter sensitivity to ATRi suggesting that an acute increase either does not exacerbate replication stress or that cells can tolerate it. A3A is thought to be the major driver of the mutational signature and DNA damage induced by A3A renders cells uniquely sensitive to ATRi (Buisson et al., 2017; Green et al., 2017) so additional studies investigating whether A3A modulates acute sensitivity to drugs in UCC should also be conducted. The minimal effects of elevated expression observed in BFTC-905 UCC cells during routine growth conditions suggests that these enzymes function differently under conditions of stress and their transcriptomes and interactomes should be investigated in this context.

The proximity-labelling technique, BioID, was used to identify interacting partners. Unfortunately, A3B's interacting partners could not be identified due to technical limitations but numerous putative A3A interacting partners were successfully identified (Chapter 6). STRING network and GO analysis of the identified proteins revealed strong enrichment in RNA binding proteins and pathways involved in mRNA metabolism, supporting the findings from the RNAseq study that suggested A3A can act on RNA substrates. The cellular localisation of A3A has been contradictory and an aim of this thesis was to determine whether A3A interacted with nuclear proteins, confirming that it could access genomic DNA. Several nuclear proteins were identified including those involved in DSB repair (DNA-PK, TRIM29 and NPM1), and this supports the hypothesis that A3 enzymes act on regions of ssDNA generated during repair of DSBs and confirms that A3A can localise to the nucleus in UCC cells. Another key aim of this thesis was to investigate whether A3A and A3B have overlapping

functions and work presented in this thesis provides evidence that this is the case. MAP4 and SAFB were identified as A3B interactors in a recent preprint (McCann et al., 2021) and were also identified as putative A3A interactors in this thesis. MAP4 and SAFB are known to associate with R-loops, having been identified previously through affinity purification and mass spectrometry studies using the RNA/DNA hybrid-specific mAb, S9.6 (Cristini et al., 2018), and McCann et al., (2021) recently showed that A3B plays a key role in R-loop resolution and homeostasis as R-loops accumulate in A3B-depleted cells. DHX9 was also identified as a putative A3A interactor and was also found to be the top hit in the proteomic study by Cristini et al., (2018) using S9.6. Interestingly, DHX9 has been shown to prevent R-loop dependent DNA damage in response to the chemotherapy drug, camptothecin (CPT), by resolving CPT-induced R-loops (Cristini et al., 2018). Many chemotherapies cause DNA damage that facilitate R-loop formation. For example, gemcitabine used extensively in this thesis, causes RS, a state characterised by RNA polymerase stalling and replication-transcription machinery conflicts, events that facilitate R-loop formation. Together, this strongly suggests that A3A and A3B are upregulated in response to drug stress to resolve R-loops.

7.4 Are A3s drug targets?

The identification of widespread A3 mutational signatures in human cancers led many to the hypothesis that inhibiting the responsible enzyme will reduce tumour heterogeneity, slow evolution and ultimately, disease progression and prevent the emergence of drug resistance. It seemed feasible, for the first time, to actively target tumour evolution with a small molecule inhibitor. A3 expression and the mutational signature is associated with poor prognosis in several tumour types and A3A and/or A3B have been shown experimentally to play a role in resistance to tamoxifen (Law et al., 2016), oncolytic viruses (Huff et al., 2018), cisplatin (Periyasamy et al., 2021) and RTKis (Isozaki et al., 2021; Mayekar et al., 2020), providing therapeutic rational for an A3 inhibitor. However, several potential issues must be addressed. Early interest was in A3B, as its expression is elevated in many tumour types and correlates with mutational load, and several groups, including ours, initiated the search for an A3B inhibitor (Grillo et al., 2022). However, the A3 family members are highly homologous, especially A3A and the

CTD of A3B, meaning that designing a specific inhibitor is extremely challenging. In addition, there are questions as to whether efforts should instead focus on A3A. While A3B specifically has a role for driving ER-transcription, cell growth and tamoxifen resistance in ER+ breast cancer, recent studies have demonstrated that A3A-like mutational signatures are far more prevalent, meaning A3A or dual inhibitors may be required. Findings presented in this thesis show many pathways commonly dysregulated in cancer are involved in *A3B* regulation and this raises the question as to whether elevated A3B expression is a passenger event. It is also worth noting that the A3 family of enzymes have important roles in the innate immune response and there are unanswered questions as to whether inhibition will increase susceptibility to viral infection and cause immunosuppression. As discussed previously, the association of A3 expression/signature with prognosis is not clear cut. While the A3 signature is enriched in high-risk NMI-UCCs and is associated with disease progression, it is conversely associated with a strikingly good survival rate in MI-UCC. Over half of the A3 mutational signature is clonal in UCC and the increased survival rate is thought to be due to A3-mediated generation of clonal neoantigens resulting in good tumour control by the immune system (McGranahan et al., 2016; Robertson et al., 2017). Therefore, while an A3 inhibitor may be useful in preventing disease progression in early stage UCC, in patients that do progress, it may limit the generation of clonal neoantigens that increase immune clearance and could then hinder response to immunotherapy. On the other hand, A3 signatures are often enriched in late, subclonal populations in other tumour types (de Bruin et al., 2014; McGranahan et al., 2015) and preventing subclonal heterogeneity could not only reduce the generation of drug resistance mutations but also minimise subclonal neoantigen heterogeneity and increase immunotherapy success. Finally, A3A expression is not commonly elevated in tumours, and it appears to generate the signature during waves of transient expression. Taking all this into consideration, stratifying patients, scheduling, and measuring clinical success is going to be extremely challenging. With all this in mind, other approaches indirectly targeting A3 mutagenesis may be more fruitful. Findings presented in this thesis demonstrated that clinically available drugs can attenuate the induction of both A3A and A3B in response to chemotherapy treatment. Using already available drugs in combination with chemotherapy to modulate transient induction

rather than blanket A3 inhibition may be a better option that circumvents some of the issues presented above. An alternative option is to identify whether synthetic lethal interactions exist and focus drug discovery efforts there, or use *A3B* expression and/or the mutational signature as a predictive marker for immuno- or other therapy success (Wang et al., 2018).

7.5 Future studies

Findings presented in this thesis show that clinically available targeted therapies can attenuate the chemotherapy induced expression of *A3A* and *A3B*. As there are no A3 inhibitors currently available, combinations of chemotherapy drugs with inhibitors targeting the pathways regulating *A3A/B* could be a useful strategy to help combat drug-induced expression. RTKis are the most promising of those studied as they can limit induction of both *A3A* and *A3B* and therefore, may have a role in limiting tumour evolution, heterogeneity and drug resistance when used in combination with chemotherapies. Long-term studies both *in vitro* and *in vivo* should be conducted to determine whether the observed induction contributes to the emergence of chemotherapy resistant UCC. Combination treatments can then be investigated to determine if they curb episodic bursts of mutation and subsequent emergence of drug resistance.

Acute elevated expression of A3A and A3B demonstrated that cells can tolerate transient exposure under routine growth conditions and in basal UCC, it is unlikely that they have a transcriptional function. Therefore, it is more plausible that recurrent bursts of expression drive adaptation through mutational mechanisms. WGS was not performed in this study and therefore, it will be interesting to subject cells to pulsed drug exposure and see if accumulation of the signature correlates with drug-induced bursts in expression. Using pulsed drug exposure with *A3A* or *A3B* CRISPR knockout models will also help to establish which family member is driving mutagenesis in this tumour type and whether loss impacts emergence of drug resistance. It is also possible that if these family members are being upregulated as part of a stress response their function is different during conditions of drug-stress versus routine growth, and this should be investigated. BioID, or an alternative method of identifying interacting partners, could be used to identify their interactomes in response to gemcitabine treatment to see if it is altered and provide further clues as to why they are upregulated. In addition, little

is known about how these enzymes are post-transcriptionally or -translationally regulated but it seems unlikely that genotoxic enzymes are constitutively active. The minimal effects observed in response to elevated expression in BFTC-905 cells suggests that their genotoxicity is modulated during periods of relative stability as opposed to periods of extreme stress. High *A3B* expression is seen in cell lines both enriched for the A3-mutational signature and those that are not suggesting that constitutively high expression is not responsible for accumulation of the mutational signature (Cortez et al., 2019). In addition, non-mutagenic isoforms of A3A and A3B exist in cells and production of the mutagenic isoform is increased in UCC cells in response to bleomycin (Rouf Banday et al., 2021). While post-transcriptional regulation is not occurring in the experiments presented in this thesis due to expression of full-length wild-type cDNA, these findings by Banday et al., (2021) demonstrate that A3 enzymes are differentially regulated during conditions of stress. It is entirely possible that post-translational modifications or sequestering, for example by RNA for A3B (Cortez et al., 2019) or cytoplasmic retention for A3A (Land et al., 2013), regulates their activity ensuring that they are only active when and where they are required. Therefore, further work should characterise the post-translational regulation of A3A and A3B to determine whether these enzymes are held in an inactive state during stable conditions and specifically activated in response to stress. In addition, robust in cell reporter assays should be developed to enable physiologically relevant measurement of deamination. Recent findings suggest that preparation of lysates for the *in vitro* deamination assay can disrupt negative regulation of these enzymes (Cortez et al., 2019; Jalili et al., 2020) and therefore, while these *in vitro* assays can determine which isoform is being produced, the activity measured may not be truly reflective of what is happening in a cell.

A novel finding presented in this thesis is the putative RNA editing activity of A3A and A3B in UCC cells. The experimental design limited the ability to fully characterise RNA editing events in this study, but the preliminary findings presented here provide rational for follow up experiments. A3 mutagenesis is largely stochastic and therefore, low frequency mutational events are hard to identify in a heterogenous pool of cells. High-depth RNA sequencing of clones or ultimately, single-cell sequencing, will allow characterisation of how individual cells are adapting in the face of stress and give insight into why some cells survive

while others do not. Finally, further investigation into the increased ADAR RNA editing signature in response to A3A and A3B expression should be conducted to determine whether these enzyme families cooperate. As discussed above, additional post-translational modification or co-factors that regulate their RNA editing activity may also exist and should be investigated. RNA editing activity should also be assessed under conditions of stress to determine whether it is upregulated.

Arguably, the most interesting finding presented in this thesis is the identification of R-loop binding proteins as A3A interactors, suggesting that A3A plays a role in R-loop homeostasis. Initially, putative interacting partners should be confirmed by orthogonal techniques. Chemotherapy drugs induce expression of A3A and A3B and the DNA damage they cause facilitates R-loop formation. Therefore, further work depleting cells, and expressing increased levels, of A3A and A3B in the presence of chemotherapy drugs followed by measurement of R-loop abundance should determine whether A3A and A3B are being upregulated to resolve chemotherapy-induced R-loops.

Finally, luminal UCC cell line models expressing A3A and A3B should be generated and used for the proposed studies to compare the luminal and basal subtypes. Luminal subtypes are enriched for nuclear receptor (ER and PPARγ) transcriptional profiles and therefore, while a transcriptional function of A3A/B was not identified in the basal BFTC-905 line, this does not exclude the possibility that they have transcriptional functions in the luminal background.

7.6 Concluding comments

The results presented in this thesis show that chemotherapy drugs with differing mechanisms of action induce expression of *A3A* and *A3B* in UCC. *A3A* and *A3B* are differentially regulated in this tumour type but clinically available targeted therapies can attenuate induction of both in response to gemcitabine-induced stress. Acute exposure to elevated levels of A3A and A3B has minimal cellular and transcriptome effects suggesting that during periods of relative stability, UCC cells tolerate elevated expression and may regulate their mutagenic/genotoxic activity. A putative RNA editing function of both A3A and A3B was identified, and characterisation of A3A's interactome identified several proteins involved in RNA binding and regulation, supporting the conclusions that A3A binds and acts on

RNA. Finally, A3A interacts with several proteins associated with DSBs and R-loops confirming that A3A has access to the nucleus. This suggests that A3A acts at ssDNA generated during DSB repair and the displaced strand of R-loops; thus, A3A may have a role in R-loop homeostasis, like A3B.

References

Abe, T., & Barber, G. N. (2014). Cytosolic-DNA-mediated, STING-dependent proinflammatory gene induction necessitates canonical NF-κB activation through TBK1. *Journal of virology, 88*(10), 5328-5341.

Ablasser, A., Goldeck, M., Cavlar, T., Deimling, T., Witte, G., Röhl, I., . . . Hornung, V. (2013). cGAS produces a 2′-5′-linked cyclic dinucleotide second messenger that activates STING. *Nature, 498*(7454), 380-384.

Adolph, M. B., Love, R. P., Feng, Y., & Chelico, L. (2017). Enzyme cycling contributes to efficient induction of genome mutagenesis by the cytidine deaminase APOBEC3B. *Nucleic acids research, 45*(20), 11925-11940.

Akre, M. K., Starrett, G. J., Quist, J. S., Temiz, N. A., Carpenter, M. A., Tutt, A. N., . . . Harris, R. S. (2016). Mutation processes in 293-based clones overexpressing the DNA cytosine deaminase APOBEC3B. *PloS one, 11*(5), e0155391.

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Ng, A. W. T., Wu, Y., . . . Bergstrom, E. N. (2020). The repertoire of mutational signatures in human cancer. *Nature, 578*(7793), 94-101.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., . . . Børresen-Dale, A.-L. (2013). Signatures of mutational processes in human cancer. *Nature, 500*(7463), 415.

Alqassim, E. Y., Sharma, S., Khan, A., Emmons, T. R., Cortes Gomez, E., Alahmari, A., . . . Robert McGray, A. (2021). RNA editing enzyme APOBEC3A promotes pro-inflammatory M1 macrophage polarization. *Communications biology, 4*(1), 1-11.

Anant, S., & Davidson, N. O. (2000). An AU-rich sequence element (UUUN [A/U] U) downstream of the edited C in apolipoprotein B mRNA is a high-affinity binding site for Apobec-1: binding of Apobec-1 to this motif in the 3′ untranslated region of c-myc increases mRNA stability. *Molecular and cellular biology, 20*(6), 1982-1992.

Anant, S., Murmu, N., Houchen, C. W., Mukhopadhyay, D., Riehl, T. E., Young, S. G., . . . Davidson, N. O. (2004). Apobec-1 protects intestine from radiation injury through posttranscriptional regulation of cyclooxygenase-2 expression. *Gastroenterology, 127*(4), 1139-1149.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.* Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Asaoka, M., Ishikawa, T., Takabe, K., & Patnaik, S. K. (2019). APOBEC3-mediated RNA editing in breast cancer is associated with heightened immune activity and improved survival. *International journal of molecular sciences, 20*(22), 5621.

Bader, S. B., Ma, T. S., Simpson, C. J., Liang, J., Maezono, S. E. B., Olcina, M. M., . . . Hammond, E. M. (2021). Replication catastrophe induced by cyclic

hypoxia leads to increased APOBEC3B activity. *Nucleic acids research, 49*(13), 7492-7506.

Baron, U., Freundlieb, S., Gossen, M., & Bujard, H. (1995). Co-regulation of two gene activities by tetracycline via a bidirectional promoter. *Nucleic acids research, 23*(17), 3605.

Bassères, D. S., Ebbs, A., Levantini, E., & Baldwin, A. S. (2010). Requirement of the NF-κB subunit p65/RelA for K-Ras–induced lung tumorigenesis. *Cancer research, 70*(9), 3537-3546.

BBSplit. Retrieved from https://sourceforge.net/projects/bbmap/

Beard, W. A., Horton, J. K., Prasad, R., & Wilson, S. H. (2019). Eukaryotic base excision repair: new approaches shine light on mechanism. *Annual review of biochemistry, 88*, 137-162.

Belotserkovskii, B. P., Tornaletti, S., D'Souza, A. D., & Hanawalt, P. C. (2018). R-loop generation during transcription: Formation, processing and cellular outcomes. *DNA repair, 71*, 69-81.

Benhar, M., Engelberg, D., & Levitzki, A. (2002). Cisplatin-induced activation of the EGF receptor. *Oncogene, 21*(57), 8723-8731.

Benjamin, D., Sato, T., Cibulskis, K., Getz, G., Stewart, C., & Lichtenstein, L. (2019). Calling somatic SNVs and indels with Mutect2. *bioRxiv*, 861054.

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

Berdik, C. (2017). Unlocking bladder cancer. *Nature, 551*, S34. doi:10.1038/551S34a

Bishop, K. N., Holmes, R. K., & Malim, M. H. (2006). Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *Journal of virology, 80*(17), 8450-8458.

Blackford, A. N., & Jackson, S. P. (2017). ATM, ATR, and DNA-PK: the trinity at the heart of the DNA damage response. *Molecular cell, 66*(6), 801-817.

Bohn, M.-F., Shandilya, S. M., Silvas, T. V., Nalivaika, E. A., Kouno, T., Kelch, B. A., . . . Schiffer, C. A. (2015). The ssDNA mutator APOBEC3A is regulated by cooperative dimerization. *Structure, 23*(5), 903-911.

Botvinnik, A., Shivam, P., Smith, Y., Sharma, G., Olshevsky, U., Moshel, O., . . . Britan‐Rosich, E. (2021). APOBEC3G rescues cells from the deleterious effects of DNA damage. *The FEBS Journal, 288*(20), 6063-6077.

Box, J. K., Paquet, N., Adams, M. N., Boucher, D., Bolderson, E., O'Byrne, K. J., & Richard, D. J. (2016). Nucleophosmin: from structure and function to disease development. *BMC molecular biology, 17*(1), 1-12.

Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., . . . Ting, A. Y. (2018). Efficient proximity labeling in living cells and organisms with TurboID. *Nature biotechnology, 36*(9), 880-887.

Bransteitter, R., Pham, P., Scharff, M. D., & Goodman, M. F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded

DNA but requires the action of RNase. *Proceedings of the National Academy of Sciences, 100*(7), 4102-4107.

Bren, G. D., Solan, N. J., Miyoshi, H., Pennington, K. N., Pobst, L. J., & Paya, C. V. (2001). Transcription of the RelB gene is regulated by NF-κB. *Oncogene, 20*(53), 7722-7733.

Brown, A. L., Collins, C. D., Thompson, S., Coxon, M., Mertz, T. M., & Roberts, S. A. (2021). Single-stranded DNA binding proteins influence APOBEC3A substrate preference. *Scientific reports, 11*(1), 1-13.

Buisson, R., Langenbucher, A., Bowen, D., Kwan, E. E., Benes, C. H., Zou, L., & Lawrence, M. S. (2019). Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science, 364*(6447), eaaw2872.

Buisson, R., Lawrence, M. S., Benes, C. H., & Zou, L. (2017). APOBEC3A and APOBEC3B activities render cancer cells susceptible to ATR inhibition. *Cancer research*.

Bulcha, J. T., Wang, Y., Ma, H., Tai, P. W., & Gao, G. (2021). Viral vector platforms within the gene therapy landscape. *Signal transduction and targeted therapy, 6*(1), 1-24.

Burns, M. B., Lackey, L., Carpenter, M. A., Rathore, A., Land, A. M., Leonard, B., . . . Nikas, J. B. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature, 494*(7437), 366-370.

Burns, M. B., Temiz, N. A., & Harris, R. S. (2013). Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature genetics, 45*(9), 977.

Calle, E. E., Rodriguez, C., Walker-Thurmond, K., & Thun, M. J. (2003). Overweight, obesity, and mortality from cancer in a prospectively studied cohort of US adults. *New England Journal of Medicine, 348*(17), 1625-1638.

Caval, V., Suspène, R., Shapira, M., Vartanian, J.-P., & Wain-Hobson, S. (2014). A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3′ UTR enhances chromosomal DNA damage. *Nature communications, 5*(1), 1-7.

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

Chakraborty, P., & Grosse, F. (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA repair, 10*(6), 654-665.

Chan, K., Roberts, S. A., Klimczak, L. J., Sterling, J. F., Saini, N., Malc, E. P., . . . Mieczkowski, P. A. (2015). An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nature genetics, 47*(9), 1067.

Chapman, E., Hurst, C., Pitt, E., Chambers, P., Aveyard, J., & Knowles, M. (2006). Expression of hTERT immortalises normal human urothelial cells without inactivation of the p16/Rb pathway. *Oncogene, 25*(36), 5037.

Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., & Alt, F. W. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature, 422*(6933), 726-730.

Chaurasiya, K. R., McCauley, M. J., Wang, W., Qualley, D. F., Wu, T., Kitamura, S., . . . Iwatani, Y. (2014). Oligomerization transforms human APOBEC3G from an efficient enzyme to a slowly dissociating nucleic acid-binding protein. *Nature chemistry, 6*(1), 28-33.

Chen, F., Alphonse, M., & Liu, Q. (2020). Strategies for nonviral nanoparticle‐based delivery of CRISPR/Cas9 therapeutics. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology, 12*(3), e1609.

Chen, J., Miller, B. F., & Furano, A. V. (2014). Repair of naturally occurring mismatches can induce mutations in flanking DNA. *Elife, 3*, e02001.

Chen, W. Y., Bailey, E. C., McCune, S. L., Dong, J.-Y., & Townes, T. M. (1997). Reactivation of silenced, virally transduced genes by inhibitors of histone deacetylase. *Proceedings of the National Academy of Sciences, 94*(11), 5798-5803.

Cheng, Y. T., Li, Y. L., Wu, J. D., Long, S. B., Tzai, T. S., Tzeng, C. C., & Lai, M. D. (1995). Overexpression of MDM‐2 mRNA and mutation of the p53 tumor suppressor gene in bladder carcinoma cell lines. *Molecular carcinogenesis, 13*(3), 173-181.

Chiang, C., & Gack, M. U. (2017). Post-translational control of intracellular pathogen sensing pathways. *Trends in immunology, 38*(1), 39-52.

Chieca, M., Montini, M., Severi, F., Lembo, G., Donati, F., Pecori, R., & Conticello, S. G. (2021). Dimerisation of APOBEC1 is dispensable for its RNA/DNA editing activity and modulates its availability. *bioRxiv*, 410803.

Choi, J.-Y., Lim, S., Kim, E.-J., Jo, A., & Guengerich, F. P. (2010). Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases α, δ, η, ι, κ, and REV1. *Journal of molecular biology, 404*(1), 34-44.

Choi, W., Porten, S., Kim, S., Willis, D., Plimack, E. R., Hoffman-Censits, J., . . . Lee, I.-L. (2014). Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer cell, 25*(2), 152-165.

Chou, W.-C., Chen, W.-T., Hsiung, C.-N., Hu, L.-Y., Yu, J.-C., Hsu, H.-M., & Shen, C.-Y. (2017). B-Myb induces APOBEC3B expression leading to somatic mutation in multiple cancers. *Scientific reports, 7*, 44089.

Chudakov, D. M., Matz, M. V., Lukyanov, S., & Lukyanov, K. A. (2010). Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological reviews, 90*(3), 1103-1163.

Chun, P. Y., Feng, F. Y., Scheurer, A. M., Davis, M. A., Lawrence, T. S., & Nyati, M. K. (2006). Synergistic effects of gemcitabine and gefitinib in the treatment of head and neck carcinoma. *Cancer research, 66*(2), 981-988.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for annotating and predicting the effects of single

nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly, 6*(2), 80-92.

Codner, G. F., Mianné, J., Caulder, A., Loeffler, J., Fell, R., King, R., . . . McCabe, C. V. (2018). Application of long single-stranded DNA donors in genome editing: generation and validation of mouse mutants. *BMC biology, 16*(1), 1-16.

Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K., & Neuberger, M. S. (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. *Molecular biology and evolution, 22*(2), 367-377.

Cooray, S., Howe, S. J., & Thrasher, A. J. (2012). Retrovirus and lentivirus vector design and methods of cell conditioning. In *Methods in enzymology* (Vol. 507, pp. 29-57): Elsevier.

Cortez, L. M., Brown, A. L., Dennis, M. A., Collins, C. D., Brown, A. J., Mitchell, D., . . . Roberts, S. A. (2019). APOBEC3A is a prominent cytidine deaminase in breast cancer. *PLoS genetics, 15*(12), e1008545.

Crawley, C. D., Kang, S., Bernal, G. M., Wahlstrom, J. S., Voce, D. J., Cahill, K. E., . . . Yamini, B. (2015). S-phase-dependent p50/NF-κB1 phosphorylation in response to ATR and replication stress acts to maintain genomic stability. *Cell Cycle, 14*(4), 566-576.

Cristini, A., Groh, M., Kristiansen, M. S., & Gromak, N. (2018). RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. *Cell reports, 23*(6), 1891-1905.

Dalvai, M., Loehr, J., Jacquet, K., Huard, C. C., Roques, C., Herst, P., . . . Doyon, Y. (2015). A scalable genome-editing-based approach for mapping multiprotein complexes in human cells. *Cell reports, 13*(3), 621-633.

Damrauer, J. S., Hoadley, K. A., Chism, D. D., Fan, C., Tiganelli, C. J., Wobker, S. E., . . . Parker, J. S. (2014). Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proceedings of the National Academy of Sciences, 111*(8), 3110-3115.

Dan, H. C., Cooper, M. J., Cogswell, P. C., Duncan, J. A., Ting, J. P.-Y., & Baldwin, A. S. (2008). Akt-dependent regulation of NF-κB is controlled by mTOR and Raptor in association with IKK. *Genes & development, 22*(11), 1490-1500.

de Bruin, E. C., McGranahan, N., Mitter, R., Salm, M., Wedge, D. C., Yates, L., . . . Rowan, A. J. (2014). Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science, 346*(6206), 251-256.

Delviks-Frankenberry, K. A., Ackerman, D., Timberlake, N. D., Hamscher, M., Nikolaitchik, O. A., Hu, W.-S., . . . Pathak, V. K. (2019). Development of lentiviral vectors for HIV-1 gene therapy with Vif-resistant APOBEC3G. *Molecular Therapy-Nucleic Acids, 18*, 1023-1038.

DeWeerd, R. A., Németh, E., Póti, Á., Petryk, N., Chen, C.-L., Hyrien, O., . . . Green, A. M. (2022). Prospectively defined patterns of APOBEC3A mutagenesis are prevalent in human cancers. *Cell reports, 38*(12), 110555.

Dianov, G. L., & Hübscher, U. (2013). Mammalian base excision repair: the forgotten archangel. *Nucleic acids research, 41*(6), 3483-3490.

Dickerson, S. K., Market, E., Besmer, E., & Papavasiliou, F. N. (2003). AID mediates hypermutation by deaminating single stranded DNA. *The Journal of experimental medicine, 197*(10), 1291-1296.

DiDonato, J. A., Mercurio, F., & Karin, M. (2012). NF‐κB and the link between inflammation and cancer. *Immunological reviews, 246*(1), 379-400.

Diner, E. J., Burdette, D. L., Wilson, S. C., Monroe, K. M., Kellenberger, C. A., Hyodo, M., . . . Vance, R. E. (2013). The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell reports, 3*(5), 1355-1361.

Dobbs, N., Burnaevskiy, N., Chen, D., Gonugunta, V. K., Alto, N. M., & Yan, N. (2015). STING activation by translocation from the ER is associated with infection and autoinflammatory disease. *Cell host & microbe, 18*(2), 157-168.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21.

Dou, Z., Ghosh, K., Vizioli, M. G., Zhu, J., Sen, P., Wangensteen, K. J., . . . Zhou, Z. (2017). Cytoplasmic chromatin triggers inflammation in senescence and cancer. *Nature, 550*(7676), 402-406.

Dunphy, G., Flannery, S. M., Almine, J. F., Connolly, D. J., Paulus, C., Jønsson, K. L., . . . Unterholzner, L. (2018). Non-canonical activation of the DNA sensing adaptor STING by ATM and IFI16 mediates NF-κB signaling after nuclear DNA damage. *Molecular cell, 71*(5), 745-760. e745.

Earl, J., Rico, D., Carrillo-de-Santa-Pau, E., Rodríguez-Santiago, B., Méndez-Pertuz, M., Auer, H., . . . Schulz, W. A. (2015). The UBC-40 Urothelial Bladder Cancer cell line index: a genomic resource for functional studies. *BMC genomics, 16*(1), 403.

Elango, R., Osia, B., Harcy, V., Malc, E., Mieczkowski, P. A., Roberts, S. A., & Malkova, A. (2019). Repair of base damage within break-induced replication intermediates promotes kataegis associated with chromosome rearrangements. *Nucleic acids research, 47*(18), 9666-9684.

Elegheert, J., Behiels, E., Bishop, B., Scott, S., Woolley, R. E., Griffiths, S. C., . . . Jones, E. Y. (2018). Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. *Nature protocols, 13*(12), 2991-3017.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics, 32*(19), 3047-3048.

Faltas, B. M., Prandi, D., Tagawa, S. T., Molina, A. M., Nanus, D. M., Sternberg, C., . . . Elemento, O. (2016). Clonal evolution of chemotherapy-resistant urothelial carcinoma. *Nature genetics, 48*(12), 1490.

Finco, T. S., Westwick, J. K., Norris, J. L., Beg, A. A., Der, C. J., & Baldwin, A. S. (1997). Oncogenic Ha-Ras-induced signaling activates NF-κB

transcriptional activity, which is required for cellular transformation. *Journal of Biological Chemistry, 272*(39), 24113-24116.

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology, 31*(9), 822-826.

Fu, Y., Ito, F., Zhang, G., Fernandez, B., Yang, H., & Chen, X. S. (2015). DNA cytosine and methylcytosine deamination by APOBEC3B: enhancing methylcytosine deamination by engineering APOBEC3B. *Biochemical Journal, 471*(1), 25-35.

Furugaki, K., Iwai, T., Kondoh, K., Moriya, Y., & Mori, K. (2010). Antitumor activity of erlotinib in combination with gemcitabine in in vitro and in vivo models of KRAS-mutated pancreatic cancers. *Oncology letters, 1*(2), 231-235.

Gaillard, H., García-Muse, T., & Aguilera, A. (2015). Replication stress and cancer. *Nature Reviews Cancer, 15*(5), 276-289.

Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B. L., Zillinger, T., . . . Hartmann, G. (2013). Cyclic [G (2′, 5′) pA (3′, 5′) p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *cell, 153*(5), 1094-1107.

Glaser, A. P., Fantini, D., Wang, Y., Yu, Y., Rimar, K. J., Podojil, J. R., . . . Meeks, J. J. (2018). APOBEC-mediated mutagenesis in urothelial carcinoma is associated with improved survival, mutations in DNA damage response genes, and immune response. *Oncotarget, 9*(4), 4537.

Glück, S., Guey, B., Gulen, M. F., Wolter, K., Kang, T.-W., Schmacke, N. A., . . . Ablasser, A. (2017). Innate immune sensing of cytosolic chromatin fragments through cGAS promotes senescence. *Nature cell biology, 19*(9), 1061-1070.

Gossen, M., & Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences, 89*(12), 5547-5551.

Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., & Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. *Science, 268*(5218), 1766-1769.

Green, A. M., Budagyan, K., Hayer, K. E., Reed, M. A., Savani, M. R., Wertheim, G. B., & Weitzman, M. D. (2017). Cytosine deaminase APOBEC3A sensitizes leukemia cells to inhibition of the DNA replication checkpoint. *Cancer research*, canres. 3394.2016.

Grillo, M. J., Jones, K. F., Carpenter, M. A., Harris, R. S., & Harki, D. A. (2022). The current toolbox for APOBEC drug discovery. *Trends in Pharmacological Sciences*.

Grivennikov, S. I., Greten, F. R., & Karin, M. (2010). Immunity, inflammation, and cancer. *cell, 140*(6), 883-899.

Gulliver, C., Hoffmann, R., & Baillie, G. S. (2020). The enigmatic helicase DHX9 and its association with the hallmarks of cancer. *Future Science OA, 7*(2), FSO650.

Guo, C., Fordjour, F. K., Tsai, S. J., Morrell, J. C., & Gould, S. J. (2021). Choice of selectable marker affects recombinant protein expression in cells and exosomes. *Journal of Biological Chemistry*, 100838.

Haag, S. M., Gulen, M. F., Reymond, L., Gibelin, A., Abrami, L., Decout, A., . . . Behrendt, R. (2018). Targeting STING with covalent small-molecule inhibitors. *Nature, 559*(7713), 269-273.

Haché, G., Liddament, M. T., & Harris, R. S. (2005). The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. *Journal of Biological Chemistry, 280*(12), 10920-10924.

Hafner, A., Bulyk, M. L., Jambhekar, A., & Lahav, G. (2019). The multiple mechanisms that regulate p53 activity and cell fate. *Nature reviews Molecular cell biology, 20*(4), 199-210.

Hainaut, P., & Pfeifer, G. P. (2001). Patterns of p53 G→ T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. *Carcinogenesis, 22*(3), 367-374.

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer discovery, 12*(1), 31-46.

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *cell, 100*(1), 57-70.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *cell, 144*(5), 646-674.

Haradhvala, N. J., Polak, P., Stojanov, P., Covington, K. R., Shinbrot, E., Hess, J. M., . . . Braunstein, L. Z. (2016). Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. *cell, 164*(3), 538-549.

Harris, R. S. (2015). Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. *Breast Cancer Research, 17*(1), 1-10.

Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., . . . Malim, M. H. (2003). DNA deamination mediates innate immunity to retroviral infection. *cell, 113*(6), 803-809.

Harris, R. S., & Dudley, J. P. (2015). APOBECs and virus restriction. *Virology, 479*, 131-145.

Harris, R. S., Petersen-Mahrt, S. K., & Neuberger, M. S. (2002). RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Molecular cell, 10*(5), 1247-1253.

Hedegaard, J., Lamy, P., Nordentoft, I., Algaba, F., Høyer, S., Ulhøi, B. P., . . . Mogensen, K. (2016). Comprehensive transcriptional analysis of early-stage urothelial carcinoma. *Cancer cell, 30*(1), 27-42.

Heinz, N., Hennig, K., & Loew, R. (2013). Graded or threshold response of the tet-controlled gene expression: all depends on the concentration of the transactivator. *BMC biotechnology, 13*(1), 1-12.

Henderson, S., Chakravarthy, A., Su, X., Boshoff, C., & Fenton, T. R. (2014). APOBEC-mediated cytosine deamination links PIK3CA helical domain

mutations to human papillomavirus-driven tumor development. *Cell reports, 7*(6), 1833-1841.

Her, J., & Bunting, S. F. (2018). How cells ensure correct repair of DNA double-strand breaks. *Journal of Biological Chemistry, 293*(27), 10502-10511.

Hinz, M., Stilmann, M., Arslan, S. Ç., Khanna, K. K., Dittmar, G., & Scheidereit, C. (2010). A cytoplasmic ATM-TRAF6-cIAP1 module links nuclear DNA damage signaling to ubiquitin-mediated NF-κB activation. *Molecular cell, 40*(1), 63-74.

Holmes, R. K., Koning, F. A., Bishop, K. N., & Malim, M. H. (2007). APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation: comparisons with APOBEC3G. *Journal of Biological Chemistry, 282*(4), 2587-2595.

Hoopes, J. I., Cortez, L. M., Mertz, T. M., Malc, E. P., Mieczkowski, P. A., & Roberts, S. A. (2016). APOBEC3A and APOBEC3B preferentially deaminate the lagging strand template during DNA replication. *Cell reports, 14*(6), 1273-1282.

Hou, S., Lee, J. M., Myint, W., Matsuo, H., Yilmaz, N. K., & Schiffer, C. A. (2021). Structural basis of substrate specificity in human cytidine deaminase family APOBEC3s. *Journal of Biological Chemistry, 297*(2).

Huang, R., & Zhou, P.-K. (2021). DNA damage repair: Historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. *Signal transduction and targeted therapy, 6*(1), 1-35.

Huang, T. T., Wuerzberger-Davis, S. M., Wu, Z.-H., & Miyamoto, S. (2003). Sequential modification of NEMO/IKKγ by SUMO-1 and ubiquitin mediates NF-κB activation by genotoxic stress. *cell, 115*(5), 565-576.

Huff, A. L., Wongthida, P., Kottke, T., Thompson, J. M., Driscoll, C. B., Schuelke, M., . . . Pulido, J. S. (2018). APOBEC3 Mediates Resistance to Oncolytic Viral Therapy. *Molecular Therapy-Oncolytics, 11*, 1-13.

Hultquist, J. F., Lengyel, J. A., Refsland, E. W., LaRue, R. S., Lackey, L., Brown, W. L., & Harris, R. S. (2011). Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *Journal of virology, 85*(21), 11220-11234.

Hurst, C., Rosenberg, J., & Knowles, M. (2018). SnapShot: Bladder Cancer. *Cancer cell, 34*(2), 350-350. e351.

Hutti, J. E., Pfefferle, A. D., Russell, S. C., Sircar, M., Perou, C. M., & Baldwin, A. S. (2012). Oncogenic PI3K mutations lead to NF-κB–dependent cytokine expression following growth factor deprivation. *Cancer research, 72*(13), 3260-3269.

Ishikawa, H., & Barber, G. N. (2008). STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature, 455*(7213), 674-678.

Isozaki, H., Abbasi, A., Nikpour, N., Langenbucher, A., Su, W., Stanzione, M., . . . Jalili, P. (2021). APOBEC3A drives acquired resistance to targeted therapies in non-small cell lung cancer. *bioRxiv*.

Israël, A. (2010). The IKK complex, a central regulator of NF-κB activation. *Cold Spring Harbor perspectives in biology, 2*(3), a000158.

Jalili, P., Bowen, D., Langenbucher, A., Park, S., Aguirre, K., Corcoran, R. B., . . . Buisson, R. (2020). Quantification of ongoing APOBEC3A activity in tumor cells by monitoring RNA editing at hotspots. *Nature communications, 11*(1), 1-13.

Jiang, Q., Isquith, J., Ladel, L., Mark, A., Holm, F., Mason, C., . . . Pham, J. (2021). Inflammation-driven deaminase deregulation fuels human pre-leukemia stem cell evolution. *Cell reports, 34*(4), 108670.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816-821.

Kamoun, A., de Reyniès, A., Allory, Y., Sjödahl, G., Robertson, A. G., Seiler, R., . . . Choi, W. (2020). A consensus molecular classification of muscle-invasive bladder cancer. *European urology, 77*(4), 420-433.

Kan, S., Koido, S., Okamoto, M., Hayashi, K., Ito, M., Kamata, Y., . . . Homma, S. (2015). Gemcitabine treatment enhances HER2 expression in low HER2-expressing breast cancer cells and enhances the antitumor effects of trastuzumab emtansine. *Oncology reports, 34*(1), 504-510.

Kan, S., Koido, S., Okamoto, M., Hayashi, K., Ito, M., Kamata, Y., . . . Homma, S. (2015). Up-regulation of HER2 by gemcitabine enhances the antitumor effect of combined gemcitabine and trastuzumab emtansine treatment on pancreatic ductal adenocarcinoma cells. *BMC cancer, 15*(1), 1-9.

Kane, L. P., Shapiro, V. S., Stokoe, D., & Weiss, A. (1999). Induction of NF-κB by the Akt/PKB kinase. *Current Biology, 9*(11), 601-S601.

Kanu, N., Cerone, M. A., Goh, G., Zalmas, L.-P., Bartkova, J., Dietzen, M., . . . Gromova, I. (2016). DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. *Genome biology, 17*(1), 185.

Karin, M., & Greten, F. R. (2005). NF-κB: linking inflammation and immunity to cancer development and progression. *Nature Reviews Immunology, 5*(10), 749-759.

Kazanov, M. D., Roberts, S. A., Polak, P., Stamatoyannopoulos, J., Klimczak, L. J., Gordenin, D. A., & Sunyaev, S. R. (2015). APOBEC-induced cancer mutations are uniquely enriched in early-replicating, gene-dense, and active chromatin regions. *Cell reports, 13*(6), 1103-1109.

Kidd, J. M., Newman, T. L., Tuzun, E., Kaul, R., & Eichler, E. E. (2007). Population stratification of a common APOBEC gene deletion polymorphism. *PLoS genetics, 3*(4), e63.

Kim, S., Kim, D., Cho, S. W., Kim, J., & Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research, 24*(6), 1012-1019.

Koike, A., Nishikawa, H., Wu, W., Okada, Y., Venkitaraman, A. R., & Ohta, T. (2010). Recruitment of phosphorylated NPM1 to sites of DNA damage through RNF8-dependent ubiquitin conjugates. *Cancer research, 70*(17), 6746-6756.

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., & Sergushichev, A. (2021). Fast gene set enrichment analysis. *bioRxiv*, 060012.

Kostrzak, A., Henry, M., Demoyen, P., Wain-Hobson, S., & Vartanian, J. (2015). APOBEC3A catabolism of electroporated plasmid DNA in mouse muscle. *Gene therapy, 22*(1), 96-103.

Kotin, R. M., Linden, R. M., & Berns, K. I. (1992). Characterization of a preferred site on human chromosome 19q for integration of adeno‐associated virus DNA by non‐homologous recombination. *The EMBO journal, 11*(13), 5071-5078.

Krela, R., Poreba, E., Weglewska, M., Skrzypczak, T., & Lesniewicz, K. (2019). A novel method for cloning of coding sequences of highly toxic proteins. *Biochimica et Biophysica Acta (BBA)-General Subjects, 1863*(3), 521-527.

Kremer, M., Suezer, Y., Martinez-Fernandez, Y., Münk, C., Sutter, G., & Schnierle, B. S. (2006). Vaccinia virus replication is not affected by APOBEC3 family members. *Virology Journal, 3*(1), 1-8.

Krueger, F. (Producer). Trim Galore! *Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.* Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Kung, C.-P., Maggi Jr, L. B., & Weber, J. D. (2018). The role of RNA editing in cancer development and metabolic disorders. *Frontiers in Endocrinology*, 762.

Kuper, H., Adami, H. O., & Trichopoulos, D. (2001). Infections as a major preventable cause of human cancer. *Journal of internal medicine, 249*(S741), 61-74.

Lackey, L., Demorest, Z. L., Land, A. M., Hultquist, J. F., Brown, W. L., & Harris, R. S. (2012). APOBEC3B and AID have similar nuclear import mechanisms. *Journal of molecular biology, 419*(5), 301-314.

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

Land, A. M., Law, E. K., Carpenter, M. A., Lackey, L., Brown, W. L., & Harris, R. S. (2013). Endogenous APOBEC3A is cytoplasmic and non-genotoxic. *Journal of Biological Chemistry*, jbc. M113. 458661.

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

Langenbucher, A., Bowen, D., Sakhtemani, R., Bournique, E., Wise, J. F., Zou, L., . . . Lawrence, M. S. (2021). An extended APOBEC3A mutation signature in cancer. *Nature communications, 12*(1), 1-11.

Larson, E. D., Bednarski, D. W., & Maizels, N. (2008). High-fidelity correction of genomic uracil by human mismatch repair activities. *BMC molecular biology, 9*(1), 1-13.

LaRue, R. S., Andrésdóttir, V., Blanchard, Y., Conticello, S. G., Derse, D., Emerman, M., . . . Löchelt, M. (2009). Guidelines for naming nonprimate APOBEC3 genes and proteins. *Journal of virology, 83*(2), 494-497.

LaRue, R. S., Jónsson, S. R., Silverstein, K. A., Lajoie, M., Bertrand, D., El-Mabrouk, N., . . . Harris, R. S. (2008). The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. *BMC molecular biology, 9*(1), 1-20.

Law, E. K., Levin-Klein, R., Jarvis, M. C., Kim, H., Argyris, P. P., Carpenter, M. A., . . . Durfee, C. (2020). APOBEC3A catalyzes mutation and drives carcinogenesis in vivo. *Journal of Experimental Medicine, 217*(12).

Law, E. K., Sieuwerts, A. M., LaPara, K., Leonard, B., Starrett, G. J., Molan, A. M., . . . Sweep, F. C. (2016). The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Science advances, 2*(10), e1601737.

Le Goux, C., Vacher, S., Schnitzler, A., Barry Delongchamps, N., Zerbib, M., Peyromaure, M., . . . Damotte, D. (2020). Assessment of prognostic implication of a panel of oncogenes in bladder cancer and identification of a 3-gene signature associated with recurrence and progression risk in non-muscle-invasive bladder cancer. *Scientific reports, 10*(1), 1-11.

Lee, J., Rhee, M. H., Kim, E., & Cho, J. Y. (2012). BAY 11-7082 is a broad-spectrum inhibitor with anti-inflammatory activity against multiple targets. *Mediators of inflammation, 2012*.

Lee, S.-J., Dimtchev, A., Lavin, M. F., Dritschilo, A., & Jung, M. (1998). A novel ionizing radiation-induced signaling pathway that activates the transcription factor NF-κB. *Oncogene, 17*(14), 1821-1826.

Leonard, B., McCann, J. L., Starrett, G. J., Kosyakovsky, L., Luengas, E. M., Molan, A. M., . . . Brown, W. L. (2015). The PKC/NF-κB signaling pathway induces APOBEC3B expression in multiple human cancers. *Cancer research, 75*(21), 4538-4547.

Li, C., Hirsch, M., Carter, P., Asokan, A., Zhou, X., Wu, Z., & Samulski, R. (2009). A small regulatory element from chromosome 19 enhances liver-specific gene expression. *Gene therapy, 16*(1), 43-51.

Li, J., Chen, Y., Li, M., Carpenter, M. A., McDougle, R. M., Luengas, E. M., . . . Mueller, J. D. (2014). APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells. *Journal of molecular biology, 426*(6), 1296-1307.

Li, N., Banin, S., Ouyang, H., Li, G. C., Courtois, G., Shiloh, Y., . . . Rotman, G. (2001). ATM is required for IκB kinase (IKK) activation in response to DNA double strand breaks. *Journal of Biological Chemistry, 276*(12), 8898-8903.

Liao, W., Hong, S.-H., Chan, B. H.-J., Rudolph, F. B., Clark, S. C., & Chan, L. (1999). APOBEC-2, a cardiac-and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochemical and biophysical research communications, 260*(2), 398-404.

Lindskrog, S. V., Prip, F., Lamy, P., Taber, A., Groeneveld, C. S., Birkenkamp-Demtröder, K., . . . Christensen, E. (2021). An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer. *Nature communications, 12*(1), 1-18.

Liu, S., Cai, X., Wu, J., Cong, Q., Chen, X., Li, T., . . . Grishin, N. V. (2015). Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. *Science, 347*(6227), aaa2630.

Liu, T., Zhang, L., Joo, D., & Sun, S.-C. (2017). NF-κB signaling in inflammation. *Signal transduction and targeted therapy, 2*(1), 1-9.

Lombardo, A., Cesana, D., Genovese, P., Di Stefano, B., Provasi, E., Colombo, D. F., . . . Riso, P. L. (2011). Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nature methods, 8*(10), 861-869.

Lorenzo, J. P., Molla, L., Ibarra, I. L., Ruf, S., Ridani, J., Subramani, P. G., . . . Di Noia, J. M. (2021). APOBEC2 is a Transcriptional Repressor required for proper Myoblast Differentiation. *bioRxiv*, 2020.2007. 2029.223594.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology, 15*(12), 1-21.

Love, R. P., Xu, H., & Chelico, L. (2012). Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A. *Journal of Biological Chemistry, 287*(36), 30812-30822.

Luo, C., Wang, S., Liao, W., Zhang, S., Xu, N., Xie, W., & Zhang, Y. (2021). Upregulation of the APOBEC3 Family Is Associated with a Poor Prognosis and Influences Treatment Response to Raf Inhibitors in Low Grade Glioma. *International journal of molecular sciences, 22*(19), 10390.

Maciejowski, J., Chatzipli, A., Dananberg, A., Chu, K., Toufektchan, E., Klimczak, L. J., . . . de Lange, T. (2020). APOBEC3-dependent kataegis and TREX1-driven chromothripsis during telomere crisis. *Nature genetics, 52*(9), 884-890.

Mackenzie, K. J., Carroll, P., Martin, C.-A., Murina, O., Fluteau, A., Simpson, D. J., . . . Leitch, A. (2017). cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature, 548*(7668), 461-465.

Madrid, L. V., Mayo, M. W., Reuther, J. Y., & Baldwin, A. S. (2001). Akt stimulates the transactivation potential of the RelA/p65 subunit of NF-κB through utilization of the IκB kinase and activation of the mitogen-activated protein kinase p38. *Journal of Biological Chemistry, 276*(22), 18934-18940.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., . . . Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science, 339*(6121), 823-826.

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

Markusic, D., Oude-Elferink, R., Das, A. T., Berkhout, B., & Seppen, J. (2005). Comparison of single regulated lentiviral vectors with rtTA expression

driven by an autoregulatory loop or a constitutive promoter. *Nucleic acids research, 33*(6), e63-e63.

Martin, A., Bardwell, P. D., Woo, C. J., Fan, M., Shulman, M. J., & Scharff, M. D. (2002). Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. *Nature, 415*(6873), 802-806.

Maruyama, W., Shirakawa, K., Matsui, H., Matsumoto, T., Yamazaki, H., Sarca, A. D., . . . Takaori-Kondo, A. (2016). Classical NF-κB pathway is responsible for APOBEC3B expression in cancer cells. *Biochemical and biophysical research communications, 478*(3), 1466-1471.

Mas-Ponte, D., & Supek, F. (2020). DNA mismatch repair promotes APOBEC3-mediated diffuse hypermutation in human cancers. *Nature genetics, 52*(9), 958-968.

Masuda, Y., Takahashi, H., Sato, S., Tomomori-Sato, C., Saraf, A., Washburn, M. P., . . . Hatakeyama, S. (2015). TRIM29 regulates the assembly of DNA repair proteins into damaged chromatin. *Nature communications, 6*(1), 1-13.

Matsumoto, T., Shirakawa, K., Yokoyama, M., Fukuda, H., Sarca, A. D., Koyabu, S., . . . Maruyama, W. (2019). Protein kinase A inhibits tumor mutator APOBEC3B through phosphorylation. *Scientific reports, 9*(1), 1-12.

Mayekar, M. K., Caswell, D. R., Vokes, N. I., Law, E. K., Wu, W., Hill, W., . . . McCoach, C. E. (2020). Targeted cancer therapy induces APOBEC fuelling the evolution of drug resistance. *bioRxiv*.

McCann, J. L., Cristini, A., Law, E. K., Lee, S. Y., Tellier, M., Carpenter, M. A., . . . Stefanovska, B. (2021). R-loop homeostasis and cancer mutagenesis promoted by the DNA cytosine deaminase APOBEC3B. *bioRxiv*.

McDaniel, Y. Z., Wang, D., Love, R. P., Adolph, M. B., Mohammadzadeh, N., Chelico, L., & Mansky, L. M. (2020). Deamination hotspots among APOBEC3 family members are defined by both target site sequence context and ssDNA secondary structure. *Nucleic acids research, 48*(3), 1353-1371.

McGranahan, N., Favero, F., de Bruin, E. C., Birkbak, N. J., Szallasi, Z., & Swanton, C. (2015). Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Science translational medicine, 7*(283), 283ra254-283ra254.

McGranahan, N., Furness, A. J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S. K., . . . Hiley, C. T. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science, 351*(6280), 1463-1469.

Menendez, D., Nguyen, T.-A., Snipe, J., & Resnick, M. A. (2017). The cytidine deaminase APOBEC3 family is subject to transcriptional regulation by p53. *Molecular Cancer Research, 15*(6), 735-743.

Middlebrooks, C. D., Banday, A. R., Matsuda, K., Udquim, K.-I., Onabajo, O. O., Paquin, A., . . . Kubo, M. (2016). Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. *Nature genetics, 48*(11), 1330.

Milewska, A., Kindler, E., Vkovski, P., Zeglen, S., Ochman, M., Thiel, V., . . . Pyrc, K. (2018). APOBEC3-mediated restriction of RNA virus replication. *Scientific reports, 8*(1), 1-12.

Miller, A. D., & Metzger, M. J. (2011). APOBEC3-mediated hypermutation of retroviral vectors produced from some retrovirus packaging cell lines. *Gene therapy, 18*(5), 528-530.

Miura, H., Gurumurthy, C. B., Sato, T., Sato, M., & Ohtsuka, M. (2015). CRISPR/Cas9-based generation of knockdown mice by intronic insertion of artificial microRNA using longer single-stranded DNA. *Scientific reports, 5*(1), 1-11.

Miyabayashi, K., Ijichi, H., Mohri, D., Tada, M., Yamamoto, K., Asaoka, Y., . . . Isayama, H. (2013). Erlotinib prolongs survival in pancreatic cancer by blocking gemcitabine-induced MAPK signals. *Cancer research, 73*(7), 2221-2234.

Moris, A., Murray, S., & Cardinaud, S. (2014). AID and APOBECs span the gap between innate and adaptive immunity. *Frontiers in microbiology, 5*, 534.

Mukai, K., Konno, H., Akiba, T., Uemura, T., Waguri, S., Kobayashi, T., . . . Taguchi, T. (2016). Activation of STING requires palmitoylation at the Golgi. *Nature communications, 7*(1), 1-10.

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

Nakanishi, C., & Toi, M. (2005). Nuclear factor-κB inhibitors as sensitizers to anticancer drugs. *Nature Reviews Cancer, 5*(4), 297-309.

Natesan, D., Zhang, L., Martell, H. J., Jindal, T., Devine, P., Stohr, B., . . . Joseph, N. (2022). APOBEC Mutational Signature and Tumor Mutational Burden as Predictors of Clinical Outcomes and Treatment Response in Patients With Advanced Urothelial Cancer. *Frontiers in oncology, 12*.

Network, C. G. A. R. (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature, 507*(7492), 315.

Newman, E. N., Holmes, R. K., Craig, H. M., Klein, K. C., Lingappa, J. R., Malim, M. H., & Sheehy, A. M. (2005). Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Current Biology, 15*(2), 166-170.

Nickerson, M. L., Witte, N., Im, K. M., Turan, S., Owens, C., Misner, K., . . . Dean, M. (2017). Molecular analysis of urothelial cancer cell lines for modeling tumor biology and drug response. *Oncogene, 36*(1), 35-46.

Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., . . . Stebbings, L. A. (2012). Mutational processes molding the genomes of 21 breast cancers. *cell, 149*(5), 979-993.

Nik-Zainal, S., Wedge, D. C., Alexandrov, L. B., Petljak, M., Butler, A. P., Bolli, N., . . . Papaemmanuil, E. (2014). Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative

APOBEC-dependent mutations in breast cancer. *Nature genetics, 46*(5), 487-491.

Nikkilä, J., Kumar, R., Campbell, J., Brandsma, I., Pemberton, H. N., Wallberg, F., . . . Serebrenik, A. A. (2017). Elevated APOBEC3B expression drives a kataegic-like mutation signature and replication stress-related therapeutic vulnerabilities in p53-defective cells. *British journal of cancer, 117*(1), 113.

Nordentoft, I., Lamy, P., Birkenkamp-Demtröder, K., Shumansky, K., Vang, S., Hornshøj, H., . . . Roth, A. (2014). Mutational context and diverse clonal development in early and late bladder cancer. *Cell reports, 7*(5), 1649-1663.

Nowarski, R., Wilner, O. I., Cheshin, O., Shahar, O. D., Kenig, E., Baraz, L., . . . Goldberg, M. (2012). APOBEC3G enhances lymphoma cell radioresistance by promoting cytidine deaminase-dependent DNA repair. *Blood*, blood-2012-2001-402123.

Nowell, P. C. (1976). The Clonal Evolution of Tumor Cell Populations: Acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression. *Science, 194*(4260), 23-28.

Oeckinghaus, A., & Ghosh, S. (2009). The NF-κB family of transcription factors and its regulation. *Cold Spring Harbor perspectives in biology, 1*(4), a000034.

Ogata, T., Kozuka, T., & Kanda, T. (2003). Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. *Journal of virology, 77*(16), 9000-9007.

Oh, S., Bournique, E., Bowen, D., Jalili, P., Sanchez, A., Ward, I., . . . Semler, B. L. (2021). Genotoxic stress and viral infection induce transient expression of APOBEC3A and pro-inflammatory genes through two distinct pathways. *Nature communications, 12*(1), 1-17.

Ohba, K., Ichiyama, K., Yajima, M., Gemma, N., Nikaido, M., Wu, Q., . . . Wong, J. E. L. (2014). In vivo and in vitro studies suggest a possible involvement of HPV infection in the early stage of breast carcinogenesis via APOBEC3B induction. *PloS one, 9*(5), e97787.

Oliveros, J. C. (2007). VENNY. An interactive tool for comparing lists with Venn Diagrams. *http://bioinfogp.cnb.csic.es/tools/venny/index.html*.

Oyer, J. A., Chu, A., Brar, S., & Turker, M. S. (2009). Aberrant epigenetic silencing is triggered by a transient reduction in gene expression. *PloS one, 4*(3), e4832.

Pak, V., Heidecker, G., Pathak, V. K., & Derse, D. (2011). The role of amino-terminal sequences in cellular localization and antiviral activity of APOBEC3B. *Journal of virology, 85*(17), 8538-8547.

Panda, A., Suvakov, M., Mariani, J., Drucker, K. L., Park, Y. H., Jang, Y., . . . Kim, J. J. (2022). Clonally selected lines after CRISPR/Cas editing are not isogenic. *bioRxiv*.

Pannunzio, N. R., Watanabe, G., & Lieber, M. R. (2018). Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *Journal of Biological Chemistry, 293*(27), 10512-10523.

Papapetrou, E. P., & Schambach, A. (2016). Gene insertion into genomic safe harbors for human gene therapy. *Molecular Therapy, 24*(4), 678-684.

Patenaude, A.-M., Orthwein, A., Hu, Y., Campo, V. A., Kavli, B., Buschiazzo, A., & Di Noia, J. M. (2009). Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nature structural & molecular biology, 16*(5), 517-527.

Paulus, W., Baur, I., Boyce, F. M., Breakefield, X. O., & Reeves, S. A. (1996). Self-contained, tetracycline-regulated retroviral vector system for gene delivery to mammalian cells. *Journal of virology, 70*(1), 62-67.

Pećina-Šlaus, N., Kafka, A., Salamon, I., & Bukovac, A. (2020). Mismatch repair pathway, genome stability and cancer. *Frontiers in molecular biosciences*, 122.

Pecori, R., Di Giorgio, S., Paulo Lorenzo, J., & Nina Papavasiliou, F. (2022). Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. *Nature Reviews Genetics*, 1-14.

Periyasamy, M., Patel, H., Lai, C.-F., Nguyen, V. T., Nevedomskaya, E., Harrod, A., . . . Thomas, R. S. (2015). APOBEC3B-mediated cytidine deamination is required for estrogen receptor action in breast cancer. *Cell reports, 13*(1), 108-121.

Periyasamy, M., Singh, A. K., Gemma, C., Farzan, R., Allsopp, R. C., Shaw, J. A., . . . Coombes, R. C. (2021). Induction of APOBEC3B expression by chemotherapy drugs is mediated by DNA-PK-directed activation of NF-κB. *Oncogene, 40*(6), 1077-1090.

Periyasamy, M., Singh, A. K., Gemma, C., Kranjec, C., Farzan, R., Leach, D. A., . . . Fenton, T. R. (2017). p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells. *Nucleic acids research, 45*(19), 11056-11069.

Petljak, M., Alexandrov, L. B., Brammeld, J. S., Price, S., Wedge, D. C., Grossmann, S., . . . Tubio, J. M. (2019). Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. *cell, 176*(6), 1282-1294. e1220.

Pfeifer, G. P., Denissenko, M. F., Olivier, M., Tretyakova, N., Hecht, S. S., & Hainaut, P. (2002). Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene, 21*(48), 7435-7451.

Pfeifer, G. P., You, Y.-H., & Besaratinia, A. (2005). Mutations induced by ultraviolet light. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 571*(1-2), 19-31.

Pham, P., Landolph, A., Mendez, C., Li, N., & Goodman, M. F. (2013). A biochemical analysis linking APOBEC3A to disparate HIV-1 restriction and skin cancer. *Journal of Biological Chemistry, 288*(41), 29294-29304.

Pierce, J. W., Schoenleber, R., Jesmok, G., Best, J., Moore, S. A., Collins, T., & Gerritsen, M. E. (1997). Novel inhibitors of cytokine-induced IκBα phosphorylation and endothelial cell adhesion molecule expression show anti-inflammatory effects in vivo. *Journal of Biological Chemistry, 272*(34), 21096-21103.

Piret, B., Schoonbroodt, S., & Piette, J. (1999). The ATM protein is required for sustained activation of NF-κB following DNA damage. *Oncogene, 18*(13), 2261-2271.

Pitot, H. C., Goldsworthy, T., & Moran, S. (1981). The natural history of carcinogenesis: implications of experimental carcinogenesis in the genesis of human cancer. *Journal of supramolecular structure and cellular biochemistry, 17*(2), 133-146.

Plummer, M., de Martel, C., Vignat, J., Ferlay, J., Bray, F., & Franceschi, S. (2016). Global burden of cancers attributable to infections in 2012: a synthetic analysis. *The Lancet Global Health, 4*(9), e609-e616.

Podolin, P. L., Callahan, J. F., Bolognese, B. J., Li, Y. H., Carlson, K., Davis, T. G., . . . Roshak, A. K. (2005). Attenuation of murine collagen-induced arthritis by a novel, potent, selective small molecule inhibitor of IκB kinase 2, TPCA-1 (2-[(aminocarbonyl) amino]-5-(4-fluorophenyl)-3-thiophenecarboxamide), occurs via reduction of proinflammatory cytokines and antigen-induced T cell proliferation. *Journal of Pharmacology and Experimental Therapeutics, 312*(1), 373-381.

Powell, L. M., Wallis, S. C., Pease, R. J., Edwards, Y. H., Knott, T. J., & Scott, J. (1987). A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *cell, 50*(6), 831-840.

Primo, L. M., & Teixeira, L. K. (2019). DNA replication stress: oncogenes in the spotlight. *Genetics and molecular biology, 43*.

Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E., & Zanini, F. (2022). Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics, 38*(10), 2943-2945.

Ramiro, A. R., Stavropoulos, P., Jankovic, M., & Nussenzweig, M. C. (2003). Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nature immunology, 4*(5), 452-456.

Rauert-Wunderlich, H., Siegmund, D., Maier, E., Giner, T., Bargou, R. C., Wajant, H., & Stühmer, T. (2013). The IKK inhibitor Bay 11-7082 induces cell death independent from inhibition of activation of NFκB transcription factors. *PloS one, 8*(3), e59292.

Ray Chaudhuri, A., & Nussenzweig, A. (2017). The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nature reviews Molecular cell biology, 18*(10), 610-621.

Roberts, S. A., Lawrence, M. S., Klimczak, L. J., Grimm, S. A., Fargo, D., Stojanov, P., . . . Saksena, G. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature genetics, 45*(9), 970.

Roberts, S. A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., . . . Mieczkowski, P. A. (2012). Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Molecular cell, 46*(4), 424-435.

Robertson, A. G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A. D., . . . Akbani, R. (2017). Comprehensive molecular characterization of muscle-invasive bladder cancer. *cell, 171*(3), 540-556. e525.

Rogozin, I. B., Basu, M. K., Jordan, I. K., Pavlov, Y. I., & Koonin, E. V. (2005). APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases predicted by computational analysis. *Cell Cycle, 4*(9), 1281-1285.

Roper, N., Gao, S., Maity, T. K., Banday, A. R., Zhang, X., Venugopalan, A., . . . Brown, A.-L. (2019). APOBEC mutagenesis and copy-number alterations are drivers of proteogenomic tumor evolution and heterogeneity in metastatic thoracic tumors. *Cell reports, 26*(10), 2651-2666. e2656.

Rouf Banday, A., Onabajo, O. O., Lin, S. H.-Y., Obajemu, A., Vargas, J. M., Delviks-Frankenberry, K. A., . . . Florez-Vargas, O. (2021). Targeting natural splicing plasticity of APOBEC3B restricts its expression and mutagenic activity. *Communications biology, 4*(1), 1-16.

Roux, K. J., Kim, D. I., Burke, B., & May, D. G. (2018). BioID: a screen for protein‐protein interactions. *Current protocols in protein science, 91*(1), 19.23. 11-19.23. 15.

Roux, K. J., Kim, D. I., Raida, M., & Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of Cell Biology, 196*(6), 801-810.

Russo, M., Crisafulli, G., Sogari, A., Reilly, N. M., Arena, S., Lamba, S., . . . Novara, L. (2019). Adaptive mutability of colorectal cancers in response to targeted therapies. *Science, 366*(6472), 1473-1480.

Sadelain, M., Papapetrou, E. P., & Bushman, F. D. (2012). Safe harbours for the integration of new DNA in the human genome. *Nature Reviews Cancer, 12*(1), 51-58.

Saitoh, T., Fujita, N., Hayashi, T., Takahara, K., Satoh, T., Lee, H., . . . Noda, T. (2009). Atg9a controls dsDNA-driven dynamic translocation of STING and the innate immune response. *Proceedings of the National Academy of Sciences, 106*(49), 20842-20846.

Sakofsky, C. J., Roberts, S. A., Malc, E., Mieczkowski, P. A., Resnick, M. A., Gordenin, D. A., & Malkova, A. (2014). Break-induced replication is a source of mutation clusters underlying kataegis. *Cell reports, 7*(5), 1640-1648.

Sakuma, T., Barry, M. A., & Ikeda, Y. (2012). Lentiviral vectors: basic to translational. *Biochemical Journal, 443*(3), 603-618.

Salamango, D. J., McCann, J. L., Demir, Ö., Brown, W. L., Amaro, R. E., & Harris, R. S. (2018). APOBEC3B nuclear localization requires two distinct N-terminal domain surfaces. *Journal of molecular biology, 430*(17), 2695-2708.

Saldivar, J. C., Cortez, D., & Cimprich, K. A. (2017). The essential kinase ATR: ensuring faithful duplication of a challenging genome. *Nature reviews Molecular cell biology, 18*(10), 622-636.

Sallmyr, A., & Tomkinson, A. E. (2018). Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. *Journal of Biological Chemistry, 293*(27), 10536-10546.

Sato, K., Hayami, R., Wu, W., Nishikawa, T., Nishikawa, H., Okuda, Y., . . . Ohta, T. (2004). Nucleophosmin/B23 is a candidate substrate for the BRCA1-BARD1 ubiquitin ligase. *Journal of Biological Chemistry, 279*(30), 30919-30922.

Sato, Y., Probst, H. C., Tatsumi, R., Ikeuchi, Y., Neuberger, M. S., & Rada, C. (2010). Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *Journal of Biological Chemistry, 285*(10), 7111-7118.

Schulze-Osthoff, K., Ferrari, D., Riehemann, K., & Wesselborg, S. (1997). Regulation of NF-κB activation by MAP kinase cascades. *Immunobiology, 198*(1-3), 35-49.

Scully, R., Panday, A., Elango, R., & Willis, N. A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature reviews Molecular cell biology, 20*(11), 698-714.

Seplyarskiy, V. B., Soldatov, R. A., Popadin, K. Y., Antonarakis, S. E., Bazykin, G. A., & Nikolaev, S. I. (2016). APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. *Genome research, 26*(2), 174-182.

Severi, F., Chicca, A., & Conticello, S. G. (2011). Analysis of reptilian APOBEC1 suggests that RNA editing may not be its ancestral function. *Molecular biology and evolution, 28*(3), 1125-1129.

Severi, F., & Conticello, S. G. (2015). Flow-cytometric visualization of C> U mRNA editing reveals the dynamics of the process in live cells. *RNA biology, 12*(4), 389-397.

Sharma, S., Patnaik, S. K., Kemer, Z., & Baysal, B. E. (2017). Transient overexpression of exogenous APOBEC3A causes C-to-U RNA editing of thousands of genes. *RNA biology, 14*(5), 603-610.

Sharma, S., Patnaik, S. K., Taggart, R. T., & Baysal, B. E. (2016). The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. *Scientific reports, 6*(1), 1-12.

Sharma, S., Patnaik, S. K., Thomas Taggart, R., Kannisto, E. D., Enriquez, S. M., Gollnick, P., & Baysal, B. E. (2015). APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nature communications, 6*(1), 1-15.

Sharma, S., Wang, J., Alqassim, E., Portwood, S., Cortes Gomez, E., Maguire, O., . . . Baysal, B. E. (2019). Mitochondrial hypoxic stress induces widespread RNA editing by APOBEC3G in natural killer cells. *Genome biology, 20*(1), 1-17.

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

Shi, M.-J., Meng, X.-Y., Fontugne, J., Chen, C.-L., Radvanyi, F., & Bernard-Pierrot, I. (2020). Identification of new driver and passenger mutations

within APOBEC-induced hotspot mutations in bladder cancer. *Genome medicine, 12*(1), 1-20.

Shin, S., Kim, S. H., Shin, S. W., Grav, L. M., Pedersen, L. E., Lee, J. S., & Lee, G. M. (2020). Comprehensive analysis of genomic safe harbors as target sites for stable expression of the heterologous gene in HEK293 cells. *ACS Synthetic Biology, 9*(6), 1263-1269.

Siddiqui, M. R., Railkar, R., Sanford, T., Crooks, D. R., Eckhaus, M. A., Haines, D., . . . Agarwal, P. K. (2019). Targeting epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor 2 (HER2) expressing bladder cancer using combination photoimmunotherapy (PIT). *Scientific reports, 9*(1), 1-14.

Sieuwerts, A. M., Willis, S., Burns, M. B., Look, M. P., Gelder, M. E., Schlicker, A., . . . Leyland-Jones, B. (2014). Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Hormones and Cancer, 5*(6), 405-413.

Silwal-Pandit, L., Vollan, H. K. M., Chin, S.-F., Rueda, O. M., McKinney, S., Osako, T., . . . Børresen-Dale, A.-L. (2014). TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. *Clinical Cancer Research, 20*(13), 3569-3580.

Sjödahl, G., Eriksson, P., Liedberg, F., & Höglund, M. (2017). Molecular classification of urothelial carcinoma: global mRNA classification versus tumour‐cell phenotype classification. *The Journal of pathology, 242*(1), 113-125.

Sjödahl, G., Lauss, M., Lövgren, K., Chebil, G., Gudjonsson, S., Veerla, S., . . . Ringnér, M. (2012). A molecular taxonomy for urothelial carcinoma. *Clinical Cancer Research, 18*(12), 3377-3386.

Smith, H. L., Southgate, H., Tweddle, D. A., & Curtin, N. J. (2020). DNA damage checkpoint kinases in cancer. *Expert Reviews in Molecular Medicine, 22*.

Smith, T., Heger, A., & Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome research.* Retrieved from https://github.com/CGATOxford/UMI-tools

Snapp, E. (2005). Design and use of fluorescent fusion proteins in cell biology. *Current protocols in cell biology, 27*(1), 21.24. 21-21.24. 13.

Sohail, A., Klapacz, J., Samaranayake, M., Ullah, A., & Bhagwat, A. S. (2003). Human activation‐induced cytidine deaminase causes transcription‐dependent, strand‐biased C to U deaminations. *Nucleic acids research, 31*(12), 2990-2994.

St. Martin, A., Salamango, D., Serebrenik, A., Shaban, N., Brown, W. L., Donati, F., . . . Harris, R. S. (2018). A fluorescent reporter for quantification and enrichment of DNA editing by APOBEC–Cas9 or cleavage by Cas9 in living cells. *Nucleic acids research, 46*(14), e84-e84.

Staudt, L. M. (2010). Oncogenic activation of NF-κB. *Cold Spring Harbor perspectives in biology, 2*(6), a000109.

Stavrou, S., & Ross, S. R. (2015). APOBEC3 proteins in viral immunity. *The Journal of Immunology, 195*(10), 4565-4570.

Stecca, C., Abdeljalil, O., & Sridhar, S. S. (2021). Metastatic Urothelial Cancer: A rapidly changing treatment landscape. *Therapeutic Advances in Medical Oncology, 13*, 17588359211047352.

Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J., & Harris, R. S. (2010). APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nature structural & molecular biology, 17*(2), 222.

Stenglein, M. D., & Harris, R. S. (2006). APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *Journal of Biological Chemistry, 281*(25), 16837-16841.

Stenglein, M. D., Matsuo, H., & Harris, R. S. (2008). Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization. *Journal of virology, 82*(19), 9591-9599.

Stilmann, M., Hinz, M., Arslan, S. Ç., Zimmer, A., Schreiber, V., & Scheidereit, C. (2009). A nuclear poly (ADP-ribose)-dependent signalosome confers DNA damage-induced IκB kinase activation. *Molecular cell, 36*(3), 365-378.

Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature, 458*(7239), 719-724.

Sui, Y., Qi, L., Zhang, K., Saini, N., Klimczak, L. J., Sakofsky, C. J., . . . Zheng, D.-Q. (2020). Analysis of APOBEC-induced mutations in yeast strains with low levels of replicative DNA polymerases. *Proceedings of the National Academy of Sciences, 117*(17), 9440-9450.

Sun, L., Wu, J., Du, F., Chen, X., & Chen, Z. J. (2013). Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science, 339*(6121), 786-791.

Sun, S.-C. (2017). The non-canonical NF-κB pathway in immunity and inflammation. *Nature Reviews Immunology, 17*(9), 545.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians, 71*(3), 209-249.

Suspène, R., Mussil, B., Laude, H., Caval, V., Berry, N., Bouzidi, M. S., . . . Vartanian, J.-P. (2017). Self-cytoplasmic DNA upregulates the mutator enzyme APOBEC3A leading to chromosomal DNA damage. *Nucleic acids research, 45*(6), 3231-3241.

Swanton, C., McGranahan, N., Starrett, G. J., & Harris, R. S. (2015). APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer discovery*.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., . . . Bork, P. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research, 47*(D1), D607-D613.

Talluri, S., Samur, M. K., Buon, L., Kumar, S., Potluri, L. B., Shi, J., . . . Munshi, N. C. (2021). Dysregulated APOBEC3G causes DNA damage and

promotes genomic instability in multiple myeloma. *Blood cancer journal, 11*(10), 1-11.

Tan, I., Ng, C. H., Lim, L., & Leung, T. (2001). Phosphorylation of a novel myosin binding subunit of protein phosphatase 1 reveals a conserved mechanism in the regulation of actin cytoskeleton. *Journal of Biological Chemistry, 276*(24), 21209-21216.

Tandon, N., Thakkar, K. N., LaGory, E. L., Liu, Y., & Giaccia, A. J. (2018). Generation of stable expression mammalian cell lines using lentivirus. *Bio-protocol, 8*(21).

Taniguchi, K., & Karin, M. (2018). NF-κB, inflammation, immunity and cancer: coming of age. *Nature Reviews Immunology, 18*(5), 309-324.

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

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

Toledo, L. I., Altmeyer, M., Rask, M.-B., Lukas, C., Larsen, D. H., Povlsen, L. K., . . . Lukas, J. (2013). ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *cell, 155*(5), 1088-1103.

Tran, L., Xiao, J.-F., Agarwal, N., Duex, J. E., & Theodorescu, D. (2021). Advances in bladder cancer biology and therapy. *Nature Reviews Cancer, 21*(2), 104-121.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., . . . Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nature methods, 13*(9), 731-740.

Vakulskas, C. A., Dever, D. P., Rettig, G. R., Turk, R., Jacobi, A. M., Collingwood, M. A., . . . Camarena, J. (2018). A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nature medicine, 24*(8), 1216-1224.

Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: using Docker, GATK, and WDL in Terra*: O'Reilly Media.

Van Schaeybroeck, S., Karaiskou-McCaul, A., Kelly, D., Longley, D., Galligan, L., Van Cutsem, E., & Johnston, P. (2005). Epidermal growth factor receptor activity determines response of colorectal cancer cells to gefitinib alone and in combination with chemotherapy. *Clinical Cancer Research, 11*(20), 7480-7489.

Vascotto, C., Fantini, D., Romanello, M., Cesaratto, L., Deganuto, M., Leonardi, A., . . . Scaloni, A. (2009). APE1/Ref-1 interacts with NPM1 within nucleoli and plays a role in the rRNA quality control process. *Molecular and cellular biology, 29*(7), 1834-1854.

Vascotto, C., Lirussi, L., Poletto, M., Tiribelli, M., Damiani, D., Fabbro, D., . . . Tell, G. (2014). Functional regulation of the apurinic/apyrimidinic

endonuclease 1 by nucleophosmin: impact on tumor biology. *Oncogene, 33*(22), 2876-2887.

Venkatesan, S., Angelova, M., Puttick, C., Zhai, H., Caswell, D. R., Lu, W.-T., . . . Bellelli, R. (2021). Induction of APOBEC3 exacerbates DNA replication stress and chromosomal instability in early breast and lung cancer evolution. *Cancer discovery, 11*(10), 2456-2473.

Venkatesan, S., Rosenthal, R., Kanu, N., McGranahan, N., Bartek, J., Quezada, S., . . . Swanton, C. (2018). Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Annals of Oncology, 29*(3), 563-572.

Vieira, V. C., Leonard, B., White, E. A., Starrett, G. J., Temiz, N. A., Lorenz, L. D., . . . Howley, P. M. (2014). Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B. *MBio, 5*(6), e02234-02214.

Walker, B. A., Wardell, C. P., Murison, A., Boyle, E. M., Begum, D. B., Dahir, N. M., . . . Kaiser, M. F. (2015). APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nature communications, 6*(1), 1-11.

Wang, H.-F., Takenaka, K., Nakanishi, A., & Miki, Y. (2011). BRCA2 and nucleophosmin coregulate centrosome amplification and form a complex with the Rho effector kinase ROCK2. *Cancer research, 71*(1), 68-77.

Wang, S., Jia, M., He, Z., & Liu, X.-S. (2018). APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer. *Oncogene, 37*(29), 3924-3936.

Warrick, J. I., Walter, V., Yamashita, H., Chung, E., Shuman, L., Amponsa, V. O., . . . Yue, F. (2016). FOXA1, GATA3 and PPARγ cooperate to drive luminal subtype in bladder cancer: a molecular analysis of established human cell lines. *Scientific reports, 6*, 38531.

Waters, C. E., Saldivar, J. C., Amin, Z. A., Schrock, M. S., & Huebner, K. (2015). FHIT loss-induced DNA damage creates optimal APOBEC substrates: Insights into APOBEC-mediated mutagenesis. *Oncotarget, 6*(5), 3409.

Williams, S. V., Hurst, C. D., & Knowles, M. A. (2013). Oncogenic FGFR3 gene fusions in bladder cancer. *Human molecular genetics, 22*(4), 795-803.

Wu, J., Sun, L., Chen, X., Du, F., Shi, H., Chen, C., & Chen, Z. J. (2013). Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science, 339*(6121), 826-830.

Wu, Z.-H., Shi, Y., Tibbetts, R. S., & Miyamoto, S. (2006). Molecular linkage between the kinase ATM and NF-κB signaling in response to genotoxic stimuli. *Science, 311*(5764), 1141-1146.

Wu, Z. H., & Miyamoto, S. (2008). Induction of a pro‐apoptotic ATM‐NF‐κB pathway and its repression by ATR in response to replication stress. *The EMBO journal, 27*(14), 1963-1973.

Xiao, G., Harhaj, E. W., & Sun, S.-C. (2001). NF-κB-inducing kinase regulates the processing of NF-κB2 p100. *Molecular cell, 7*(2), 401-409.

Xiao, X., Yang, H., Arutiunian, V., Fang, Y., Besse, G., Morimoto, C., . . . Chen, X. S. (2017). Structural determinants of APOBEC3B non-catalytic domain for molecular assembly and catalytic regulation. *Nucleic acids research, 45*(12), 7494-7506.

Yamazaki, H., Shirakawa, K., Matsumoto, T., Hirabayashi, S., Murakawa, Y., Kobayashi, M., . . . Maruyama, W. (2019). Endogenous APOBEC3B overexpression constitutively generates DNA substitutions and deletions in myeloma cells. *Scientific reports, 9*(1), 1-14.

Yamazaki, H., Shirakawa, K., Matsumoto, T., Kazuma, Y., Matsui, H., Horisawa, Y., . . . Shindo, K. (2020). APOBEC3B reporter myeloma cell lines identify DNA damage response pathways leading to APOBEC3B expression. *PloS one, 15*(1), e0223463.

Yan, S., He, F., Gao, B., Wu, H., Li, M., Huang, L., . . . Li, Y. (2016). Increased APOBEC3B predicts worse outcomes in lung cancer: a comprehensive retrospective study. *Journal of Cancer, 7*(6), 618.

Yu, Y., Lowy, M. M., & Elble, R. C. (2016). Tet-On lentiviral transductants lose inducibility when silenced for extended intervals in mammary epithelial cells. *Metabolic engineering communications, 3*, 64-67.

Zhang, J.-P., Li, X.-L., Li, G.-H., Chen, W., Arakaki, C., Botimer, G. D., . . . Fu, Y.-W. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome biology, 18*(1), 1-18.

Zhang, X., Shi, H., Wu, J., Zhang, X., Sun, L., Chen, C., & Chen, Z. J. (2013). Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Molecular cell, 51*(2), 226-235.

Appendix

Primers

Appendix Table 1 Primers used in this thesis. Primer annealing temperatures were calculated for the specific polymerase with the Tm Calculator v.1.12.0 (New England Biolabs, USA). All primers were ordered from Integrated DNA Technologies, USA.

ID	Primer name	Primer sequence 5' → 3' (extension in lowercase)	Information: constructs primers were used for in brackets.
1	pcDNA3.1(+)_EcoRI	CTGCAGATATCCAGCACAGT	Linearizes pcDNA3.1(+)_Hygro
2	pcDNA3.1(+)_HindIII	TAAGTTTAAACGCTAGCCAGCTT	
3	NA3Bm-mini_F	ctggctagcgtttaaacttaGCCACCATGTACCCGTATG	Kozak-3xHA-miniTurbo-linker fragment (pcDNA3.1(+)_3xHA-miniTurbo-A3B; pcDNA3.1(+)_3xHA-miniTurbo-A3B^{Split}) Kozak-3xHA-TurboID-linker fragment (pcDNA3.1(+)_3xHA-TurboID-A3B)
4	NA3Bm-mini_R	gtggattcatctcgagcggccgcgtacgcgtCTTTTCGGCAGACCGCAG	
5	NA3Bm-A3B_F	acgcgtacgcggccgctcgagATGAATCCACAGATCAGAAATC	A3B-stop fragment (pcDNA3.1(+)_3xHA-miniTurbo-A3B; pcDNA3.1(+)_3xHA-miniTurbo-A3B^{Split}; pcDNA3.1(+)_3xHA-TurboID-A3B)
6	NA3Bm-A3B_R	actgtgctggatatctgcagTCAGTTTCCCTGATTCTG	
7	CA3Bm-A3B_F	ctggctagcgtttaaacttaGCCACCATGAATCCACAGATCAG	Kozak-A3B-linker fragment (pcDNA3.1(+)_A3B-miniTurbo-V5; pcDNA3.1(+)_A3B^{Split}-miniTurbo-V5).
8	CA3Bm-A3B_R	gcagcgggatCTCGAGCGGCCGCGTACG	
9	CA3Bm-mini_F	gccgctcgagATCCCGCTGCTGAACGCTAAAC	miniTurbo-V5 fragment (pcDNA3.1(+)_A3B-miniTurbo-V5; pcDNA3.1(+)_A3B^{Split}-miniTurbo-V5)
10	CA3Bm-mini_R	actgtgctggatatctgcagTCATTAGGTGCTGTCCAGGC	
12	Ctrl_m-mini_R	gcgtacgcgtCTTTTCGGCAGACCGCAG	Kozak-3xHA-miniTurbo-linker fragment (pcDNA3.1(+)_3xHA-miniTurbo-GFP; pcDNA3.1(+)_3xHA-TurboID-GFP when used with 3)

13	Ctrl_m-GFP_F	tgccgaaaagACGCGTACGCGGCCGCTC	GFP-stop fragment (pcDNA3.1(+)_3xHA-miniTurbo-GFP; pcDNA3.1(+)_3xHA-TurboID-GFP)
14	Ctrl_m-GFP_R	actgtgctggatatctgcagTTAAACTCTTTCTTCACCGGCATCTGCATCCG	
15	CA3BTur-A3B_R	tattgtctttCTCGAGCGGCCGCGTACG	Kozak-A3B-linker fragment (pcDNA3.1(+)_A3B-TurboID-V5 when used with 7)
16	CA3BTur-Turbo_F	gccgctcgagAAAGACAATACTGTGCCTCTG	TurboID-V5 fragment (pcDNA3.1(+)_A3B-TurboID-V5)
17	CA3BTur-Turbo_R	actgtgctggatatctgcagTCATTAGGTGCTGTCCAG	
18	pTRIPZ_R1_F	gccgaaaccggtGCCACCATGTACCCGTATGA	Sub-cloning 3xHA-miniTurbo-A3B into pTRIPZ (pTRIPZ_3xHA-miniTurbo-A3B; pTRIPZ_3xHA-TurboID-A3B) Leader sequence and AgeI/ClaI sequences
19	pTRIPZ_R1_R	gccgaaatcgatTCAGTTTCCCTGATTCTGGAGAATG	
20	pTRIPZ_R2_F	gccgaaaccggtGCCACCATGAATCCACAGAT	Sub-cloning A3B-miniTurbo-V5 into pTRIPZ (pTRIPZ_A3B-miniTurbo-V5; pTRIPZ_A3B-TurboID-V5) Leader sequence and AgeI/ClaI sequences
21	pTRIPZ_R2_R	gccgttatcgatTTAGGTGCTGTCCAGGCC	
22	pTRIPZ_GFP_R	gccgttatcgatTTAAACTCTTTCTTCACCGGCATCT	Sub-cloning 3xHA-miniTurbo-GFP into pTRIPZ (pTRIPZ_3xHA-miniTurbo-GFP; pTRIPZ_3xHA-TurboID-GFP); use with 18. Leader sequence and AgeI/ClaI sequences
23	pTRIPZA3A_F	CGATAGTTTGTTTGAATGAGGC	Linearizes pTRIPZ
24	pTRIPZA3A_R	TGCGATCTGACGGTTCAC	
25	N-termA3A_mini_F	tagtgaaccgtcagatcgcaGCCACCATGTACCCGTATG	Kozak-3xHA-miniTurbo-Linker fragment (pTRIPZ_3xHA-miniTurbo-A3A)
26	N-termA3A_mini_R	ctcgagcggccgcgtacgcgtCTTTTCGGCAGACCGCAG	
27	NtermA3A_A3A_F	acgcgtacgcggccgctcgagATGGAAGCCAGCCCAGCA	A3A-stop fragment (pTRIPZ_3xHA-miniTurbo-A3A)
28	NtermA3A_A3A_R	ctcattcaaacaaactatcgtcaGTTTCCCTGATTCTGGAGAATGGC	
29	C-termA3A_mini_F	gccgctcgagATCCCGCTGCTGAACGCT	Kozak-A3A-linker fragment (pTRIPZ_A3A-miniTurbo-V5)

30	C-termA3A_mini_R	ctcattcaaacaaactatcgTTAGGTGCTGTCCAGGCC	
31	CtermA3A_A3A_F	tagtgaaccgtcagatcgcagccaccATGGAAGCCAGCCCAGCATC	miniTurbo-V5 fragment (pTRIPZ_A3A-miniTurbo-V5)
32	CtermA3A_A3A_R	gcagcgggatCTCGAGCGGCCGCGTACG	
33	Colony_Screen_F	GGCTAACTAGAGAACCCACTGCTT	Colony PCR & sequencing (pcDNA3.1(+)_Hygro plasmids)
34	Colony_Screen_R	GGCACAGTCGAGGCTGAT	
35	Col_pTRIPZ_F	GAACGTATGTCGAGTTTATCCCT	Colony PCR & sequencing (pTRIPZ plasmids)
36	Col_pTRIPZ_R	AGCCTCATTCAAACAAACTATCGA	
37	Col_pTRIPZ_R2	TGCAGGCAACGATTCTGTAAAG	
38	Seq1_F	GAGATCCCGCTGCTGAAC	Sequencing (pTRIPZ_3xHA-miniTurbo-A3A)
39	Seq2_R	CAGGTACTGATTGGTGGAG	
40	Seq3_F	TTCGAGCAGGAAGGCCTG	Sequencing (pTRIPZ_A3A-miniTurbo-V5)
41	Seq4_R	TCAATCCCCCGGCTAATCC	
42	Stuffer_F	gccgtttgtacaTCGAGGTAACCGGATCCTG	Amplifying stuffer region from pTRIPZ (pTRIPZ-^{A3BSplit}-GFP) Contains leader sequence and BsrGI site.
43	Stuffer_R	gccgtttgtacaCAAACAACATCAAACAAACCAG	
44	A3BSplit_Seq_R	TGTTGATGGCGTGCAGGAA	Sequencing A3B^{Split} constructs
45	AAVS1_linear_F	CTGTGCCTTCTAGTTGCC	Linearizes AAVS1_Puro_Tet3G_3xFLAG_Twin_Strep (removes 3xFLAG_Twin_Strep)
46	AAVS1_linear_R	TTTACGAGGGTAGGAAGTGG	
47	R1_fwd	ccacttcctaccctcgtaaaGCCACCATGTACCCGTATG	

48	R1_rev	ctggcaactagaaggcacagTCAGTTTCCCTGATTCTGG	Sub-cloning (AAVS1-TetON_3xHA-miniTurbo-A3B^{Split}; AAVS1-TetON_3xHA-miniTurbo-A3A)
49	R2_fwd	ccacttcctaccctcgtaaaGCCACCATGAATCCACAG	Sub-cloning (AAVS1-TetON_A3B^{Split}-miniTurbo-V5)
50	R2_rev	ctggcaactagaaggcacagTTAGGTGCTGTCCAGGCC	
51	R5_rev	ctggcaactagaaggcacagTTAAACTCTTTCTTCACCGGC	Sub-cloning (AAVS1-TetON_3xHA-miniTurbo-GFP when used with 47)
52	R8_fwd	ccacttcctaccctcgtaaaGCCACCATGGAAGCCAGC	Sub-cloning (AAVS1-TetON_A3A-miniTurbo-V5 when used with 50)
53	AAVS1_Screen_F	CCGCTTTCGCACTTTAGCTG	Colony PCR (AAVS1-TetON plasmids)
54	AAVS1_Screen_R	GGCAAACAACAGATGGCTGG	
55	AAVS1_Seq_F	CCGCTTTCGCACTTTAGCTG	Sequencing (AAVS1-TetON plasmids)
56	AAVS1_Seq_R	GGCAAACAACAGATGGCTGG	
57	Seq1_A3B_F	GGGATTGATAAACAGGGAGC	Sequencing (AAVS1-TetON_3xHA-miniTurbo-A3B^{Split})
58	Seq2_A3B_R	GATGGTCAGGGTGACATTGG	
59	Seq3_A3B_F	AAATGCTGCGGGATGCTG	Sequencing (AAVS1-TetON_A3B^{Split}-miniTurbo-V5)
60	Seq4_A3B_R	CCTGCCTGCTGATATTCTGC	
61	A3A_SDM_F	aTGGAAGCCAGCCCAGCATC	Site-directed mutagenesis (AAVS1-TetON_3xHA-miniTurbo-A3A)
62	A3A_SDM_R	CTCGAGCGGCCGCGTA	
63	PAGE_F	GCAAGGACGCTGTAAGCAG	PCR amplification of genomic region surrounding guide target in exon 1.
64	PAGE_R	AAGAGCCTGACTGGGATTCGT	

65	Exon5_F	GGCTTTTGGTTTCCCCTGTC	PCR amplification of genomic region surrounding guide target in exon 5.
66	Exon5_R	GTCGGTCACCTCGTTGCATA	
67	HR-AAVS1_F	CTGCCGTCTCTCTCCTGAGT	Screening integration of transgene at genomic AAVS1integration locus.
68	HR-AAVS1_F2	CGGAACTCTGCCCTCTAACG	
69	HR-Puro_R2	GCTCGTAGAAGGGGAGGTTG	
70	HA-L_F	TTGCTCTCTGCTGTGTTGCT	PCR amplification and sequencing of genomic AAVS1 integration locus.
71	HA-L_R	GTGAGAGGTGACCCGAATCC	
72	HA-R_F	TGGTGACACACCCCCATTTC	
73	HA-R_R	GCTTGCCAAGGACTCAAACC	

Plasmids

Plasmid maps are included for key plasmids used in experimental work and to make other plasmids (Appendix Figure 1).

Appendix Table 2 Plasmid used and made in this thesis.

Plasmid Name	Source	Details	Used For
3xHA-TurboID-NLS_pCDNA3	Gift from Alice Ting (Addgene plasmid #107171 http://n2t.net/addgene:107171)	Expresses 3xHA-tagged TurboID in the mammalian nucleus (BirA mutant: Q65P, I87V, R118S, E140K, Q141R, S150G, L151P, V160A, T192A, K194I, M209V, M241T, S263P, I305V). 3xHA: YPYDVPDYAYPYDVPDYAYP YDVPDYA Vector backbone: pcDNA3 Promoter: CMV Bacterial resistance: Ampicillin Growth temperature: 37°C Growth strain: NEB 5-alpha Copy number: High Copy	PCR amplification of 3xHA-tagged TurboID for cloning.
3xHA-miniTurbo-NLS_pCDNA3	Gift from Alice Ting (Addgene plasmid #107172 http://n2t.net/addgene:107172)	Expresses 3xHA-tagged miniTurbo in the mammalian nucleus (BirA mutant: aa1-63 deleted; Q65P, I87V, R118S, E140K, Q141R, S150G, L151P, V160A, T192A, K194I, M209V, I305V). 3xHA: YPYDVPDYAYPYDVPDYAYP YDVPDYA Vector backbone: pcDNA3 Promoter: CMV Bacterial resistance: Ampicillin Growth temperature: 37°C Growth strain: NEB 5-alpha Copy number: High Copy	PCR amplification of 3xHA-tagged miniTurbo for cloning.
C1(1-29)-TurboID-V5_pCDNA3	Gift from Alice Ting (Addgene plasmid #107173 http://n2t.net/addgene:107173)	Expresses V5-tagged TurboID on the mammalian ER membrane (BirA mutant: Q65P, I87V, R118S, E140K, Q141R, S150G, L151P, V160A, T192A, K194I, M209V, M241T, S263P, I305V). V5: GKPIPNPLLGLDST	PCR amplification of V5-tagged TurboID for cloning.

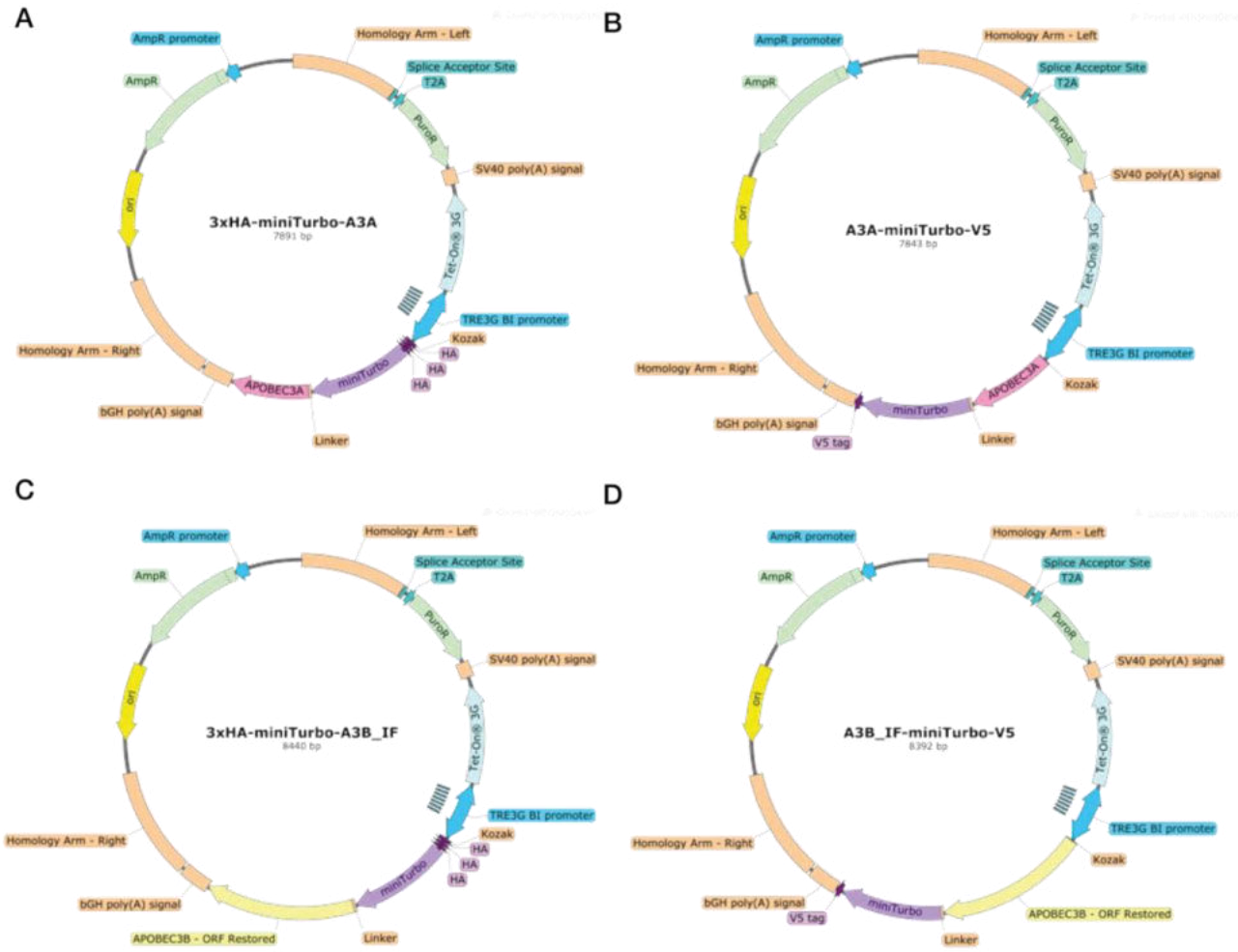
		Vector backbone: pcDNA3 Promoter: CMV Bacterial resistance: Ampicillin Growth temperature: 37°C Growth strain: NEB 5-alpha Copy number: High Copy	
C1(1-29)-miniTurbo-V5_pCDNA3	Gift from Alice Ting (Addgene plasmid #107174 http://n2t.net/addgene:107174)	Expresses V5-tagged miniTurbo on the mammalian ER membrane miniTurbo (BirA mutant: aa1-63 deleted; Q65P, I87V, R118S, E140K, Q141R, S150G, L151P, V160A, T192A, K194I, M209V, I305V). V5: GKPIPNPLLGLDST Vector backbone: pcDNA3 Promoter: CMV Bacterial resistance: Ampicillin Growth temperature: 37°C Growth strain: NEB 5-alpha Copy number: High Copy	PCR amplification of V5-tagged miniTurbo for cloning.
APOBEC3A (NM_145699) Human Tagged ORF Clone	Origene (CAT#: RC220995)	Expresses APOBEC3A (Myc-DDK-tagged)-Human apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A (APOBEC3A), transcript variant 1 Synonymous mutations (Arg123 and Ala139). Vector backbone: pCMV6-Entry Promoter: CMV Bacterial resistance: Kanamycin Mammalian selection: Neomycin	PCR amplification of A3A ORF for cloning.
AAVS1_Puro_Tet3G_3xFLAG_Twin_Strep	Gift from Yannick Doyon (Addgene plasmid #92099 http://n2t.net/addgene:92099)	Empty Backbone. Targets transgene to the *AAVS1* Genomic Safe Harbour Locus. Transgene expression is controlled by an Auto-Regulated Tet-On 3G System. Vector backbone: pUC19 Promoter: Tet-On 3G Bidirectional Mammalian selection: Puromycin Bacterial Resistance: Ampicillin Growth Temperature: 37°C Growth Strain: NEB Stable Copy number: High Copy	3xFLAG_Twin_Strep tag was removed and resulting backbone was used for cloning.

pcDNA3.1(+)_Hygro	Gift from Dr Daniel Miller (ICR)	Empty backbone. Promoter: CMV Mammalian selection: Hygromycin Bacterial resistance: Ampicillin Growth temperature: 37°C Growth strain: NEB 5-alpha Copy number: High Copy	Backbone for cloning.
pTRIPZ (GE Healthcare, UK)	Gift from Dr Michael Walton (ICR)	Empty backbone. Promoter: CMV (TRE; tetracycline inducible) Mammalian selection: Puromycin Bacterial resistance: Ampicillin; Zeocin Contains TurboRFP reporter.	Backbone for cloning. Amplification of stuffer sequence.
pTRIPZ_A3B-GFP	Gift from Dr Michael Walton and Dr Chi Zhang (ICR).	Expresses A3B C-terminally tagged with TurboGFP. Vector backbone: pTRIPZ Linker sequence: ACGCGTACGCGGCCGCTCGAG (aa: TRTRPLE).	PCR amplification of A3B, linker sequence and TurboGFP.
pTRIPZ_A3B**-GFP	Gift from Dr Michael Walton and Dr Chi Zhang (ICR).	Expresses catalytically inactive A3B (A3B**; E68Q, E255Q) C-terminally tagged with TurboGFP. Vector backbone: pTRIPZ Linker sequence: ACGCGTACGCGGCCGCTCGAG (aa: TRTRPLE).	PCR amplification of A3B**, linker sequence and TurboGFP.
pTRIPZ_A3B^{Split}-GFP	This study	Contains A3B^{Split} (ORF disrupted A3B) C-terminally tagged with TurboGFP. Stuffer sequence introduces stop codon preventing expression of full length A3B. Vector backbone: pTRIPZ Linker sequence: ACGCGTACGCGGCCGCTCGAG (aa: TRTRPLE). Stuffer Sequence: CAAACAACATCAAACAAACCAGCAGACATATGCAACAAGACACGAGACGGAGTAATGGCCGGCCGCATTAGTCTTCCAATTGAAAAAAGTGATTTAATTTATACCATTTTAATTCAGCTTTGTAAAAATGTATCAAAGAGATAGCAAGGTATTCAGTTTTAGTAAACAAGATAATTGCTCCTAAAGTAGCCCCTTGAATTCTGATCAGGATCCGGTTACCTCGA	PCR amplification of A3B^{Split} sequence.
pcDNA3.1(+)_3xHA-miniTurbo-A3B^{Split}	This study	Contains A3B^{Split} N-terminally tagged with 3xHA-miniTurbo biotin ligase. Vector backbone: pcDNA3 Linker sequence: ACGCGTACGCGGCCGCTCGAG (aa: TRTRPLE).	Sub-cloning insert into AAVS1_TetON backbone.

pcDNA3.1(+)_A3B^{Split}-miniTurbo-V5	This study	Contains A3B^{Split} C-terminally tagged with miniTurbo-V5 biotin ligase. Vector backbone: pcDNA3 Linker sequence: ACGCGTACGCGGCCGCTCGAG (aa: TRTRPLE).	Sub-cloning insert into AAVS1_TetON backbone.
pcDNA3.1(+)_3xHA-miniTurbo-GFP	This study	Contains TurboGFP N-terminally tagged with 3xHA-miniTurbo biotin ligase. Vector backbone: pcDNA3 Linker sequence: ACGCGTACGCGGCCGCTCGAG (aa: TRTRPLE).	Sub-cloning insert into AAVS1_TetON backbone.
pTRIPZ_3xHA-miniTurbo-A3A	This study	Contains A3A N-terminally tagged with 3xHA-miniTurbo biotin ligase. Synonymous mutations (Arg123 and Ala139). Vector backbone: pTRIPZ Linker sequence: ACGCGTACGCGGCCGCTCGAG (aa: TRTRPLE).	Sub-cloning insert into AAVS1_TetON backbone.
pTRIPZ_A3A-miniTurbo-V5	This study	Contains A3A C-terminally tagged with miniTurbo-V5 biotin ligase. Synonymous mutations (Arg123 and Ala139). Vector backbone: pTRIPZ Linker sequence: ACGCGTACGCGGCCGCTCGAG (aa: TRTRPLE).	Sub-cloning insert into AAVS1_TetON backbone.
AAVS1-TetON_3xHA-miniTurbo-A3B^{Split}	This study	Contains A3B^{Split} N-terminally tagged with 3xHA-miniTurbo biotin ligase. Vector backbone: AAVS1_Puro_Tet3G	Production of AAVS1-TetON_3xHA-miniTurbo-A3B_IF
AAVS1-TetON_3xHA-miniTurbo-A3B_IF	This study	Contains A3B N-terminally tagged with 3xHA-miniTurbo biotin ligase. Digested with BsrGI and ligated to remove stuffer sequence and restore A3B ORF. Vector backbone: AAVS1_Puro_Tet3G	Transient transfection. Stable cell line generation using CRISPR/Cas9 integration into AAVS1 locus.
AAVS1-TetON_A3B^{Split}-miniTurbo-V5	This study	Contains A3B^{Split} C-terminally tagged with miniTurbo-V5 biotin ligase. Vector backbone: AAVS1_Puro_Tet3G	Production of AAVS1-TetON_A3B_IF-miniTurbo-V5
AAVS1-TetON_A3B_IF-miniTurbo-V5	This study	Contains A3B C-terminally tagged with miniTurbo-V5 biotin ligase. Digested with BsrGI and ligated to remove stuffer sequence and restore A3B ORF. Vector backbone: AAVS1_Puro_Tet3G	Transient transfection. Stable cell line generation using CRISPR/Cas9 integration into AAVS1 locus.

AAVS1-TetON_3xHA-miniTurbo-A3A	This study	Contains A3A N-terminally tagged with 3xHA-miniTurbo biotin ligase. Synonymous mutations (Arg123 and Ala139). Vector backbone: AAVS1_Puro_Tet3G	Transient transfection. Stable cell line generation using CRISPR/Cas9 integration into AAVS1 locus.
AAVS1-TetON_A3A-miniTurbo-V5	This study	Contains A3A C-terminally tagged with miniTurbo-V5 biotin ligase. Synonymous mutations (Arg123 and Ala139). Vector backbone: AAVS1_Puro_Tet3G	Transient transfection. Stable cell line generation using CRISPR/Cas9 integration into AAVS1 locus.
AAVS1-TetON_3xHA-miniTurbo-GFP	This study	Contains TurboGFP N-terminally tagged with 3xHA-miniTurbo biotin ligase. Vector backbone: AAVS1_Puro_Tet3G	Transient transfection. Stable cell line generation using CRISPR/Cas9 integration into AAVS1 locus.
AAVS1_TetON_A3A-V5	This study	Contains A3A C-terminally tagged with V5 Tag. Puromycin resistance cassette switched for blasticidin cassette Vector backbone: AAVS1_Blasticidin_Tet3G	Transient transfection. Stable cell line generation using CRISPR/Cas9 integration into AAVS1 locus
AAVS1_TetON_A3A*-V5	This study	Contains A3A* (E72A mutant; catalytically inactive enzyme) C-terminally tagged with V5 Tag. Puromycin resistance cassette switched for blasticidin cassette Vector backbone: AAVS1_Blasticidin_Tet3G	Transient transfection. Stable cell line generation using CRISPR/Cas9 integration into AAVS1 locus
AAVS1_TetON_A3B^{Split}-V5	This study	Contains A3B C-terminally tagged with V5 Tag. Puromycin resistance cassette switched for blasticidin cassette Vector backbone: AAVS1_Blasticidin_Tet3G	Transient transfection. Stable cell line generation using CRISPR/Cas9 integration into AAVS1 locus
AAVS1-TetON_A3B_IF-V5	This study	Contains A3B C-terminally tagged with V5. Puromycin resistance cassette switched for blasticidin cassette Digested with BsrGI and ligated to remove stuffer sequence and restore	Transient transfection. Stable cell line generation using CRISPR/Cas9

		A3B ORF. Vector backbone: AAVS1_Blasticidin_Tet3G	integration into AAVS1 locus.
AAVS1_TetON_A3B**-V5	This study	Contains A3B** (E86Q, E255Q mutant; catalytically inactive enzyme) C-terminally tagged with V5 Tag. Puromycin resistance cassette switched for blasticidin cassette Vector backbone: AAVS1_Blasticidin_Tet3G	Transient transfection. Stable cell line generation using CRISPR/Cas9 integration into AAVS1 locus
AAVS1-PURO_TetON_A3A-V5	This study	Contains A3A C-terminally tagged with V5 Tag. Vector backbone: AAVS1_Puro_Tet3G	Stable cell line generation
AAVS1_PURO_TetON_A3A*-V5	This study	Contains A3A* (E72A mutant; catalytically inactive enzyme) C-terminally tagged with V5 Tag. Vector backbone: AAVS1_Puro_Tet3G	Stable cell line generation



Appendix Figure 1 BioID plasmid maps. (A) AAVS1-TetON_3xHA-miniTurbo-A3A. (B) AAVS1-TetON_A3A-miniTurbo-V5. (C) AAVS1-TetON_3xHA-miniTurbo-A3B_IF. (D) AAVS1-TetON_A3B_IF-miniTurbo-V5. The ORF of APOBEC3B has been restored through removal of stuffer fragment (C and D).