

**IMPERIAL COLLEGE LONDON**

Faculty of Medicine, Department of Surgery and Cancer

**Genome Engineering in the Study of APOBEC3 Gene Function**

Thesis submitted for the degree of Doctor of Philosophy

Raed Farzan

Supervisors: Professor Laki Buluwela, Professor Simak Ali, and Dr Manikandan Periyasamy

July 2019

## Abstract

The “apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like” (APOBEC) gene family defines a set of related Cytidine Deaminases. The best characterised of these are ones that function in the process of mRNA editing. Members of the family are also known to use DNA as their preferred substrate, acting in a consensus sequence, and mediating Cytidine deamination leading to a C to T transition. Recently, this process has been recognised as a major mutational mechanism in several cancer types, where it has been further suggested that members of the APOBEC3 family, and potentially APOBEC3H, that cause this. In order to understand the role of APOBEC3H further, we have used CRISPR-Cas9 gene editing to show that APOBEC3H is a p53 regulated target gene. This has involved making a p53 knockout version of the MCF7 breast cancer cell line, using CRISPR-Cas9 to delete part of the third coding exon of the gene. With this line, it has been possible to confirm that APOBEC3H expression is dependent on p53. In further work, CRISPR-Cas9 was used to make an APOBEC3H knockout in the HCT116 colon cancer line. Having established two independent APOBEC3H knockout lines in HCT116, RNAseq analysis was carried out to compare gene expression of mutant and wild-type cells in both the control and p53 activated setting. Analysis of the RNAseq data has shown that knockout of APOBEC3H leads to an attenuated p53 response, suggesting the possibility that APOBEC3H is co-regulator of p53 regulated gene expression. This is further indicated by the demonstration that over-expression of APOBEC3H in HCT116 wild-type cells enhances the expression of p53 target genes following p53 activation. Taken together, these data identify APOBEC3H as a p53 regulated gene, with a potential role in modulating the p53 response, and potentially identify a new mechanism for APOBEC3H action in cancer.

### **Declaration of originality**

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy. I hereby declare that the work reported in this thesis is my own, or if not, it is clearly stated and fully acknowledged. Some of the findings presented in this thesis have been published in the following articles:

Manikandan Periyasamy, Anup K. Singh, Carolina Gemma, Christian Kranjec, **Raed Farzan**, Damien A. Leach, Naveenan Navaratnam, Hajnalka L. Pálincás, Beata G. Vértessy, Tim R. Fenton, John Doorbar, Frances Fuller-Pace, David W. Meek, R. Charles Coombes, Laki Buluwela, Simak Ali, “p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells.” *Nucleic acids research* vol. 45,19 (2017): 11056-11069. doi:10.1093/nar/gkx721

Linda Smith, **Raed Farzan**, Simak Ali, Laki Buluwela, Adrian T. Saurin & David W. Meek  
“Author Correction: The responses of cancer cells to PLK1 inhibitors reveal a novel protective role for p53 in maintaining centrosome separation.” *Scientific reports* vol. 8,1 5237. 22 Mar. 2018, doi:10.1038/s41598-018-23384-5

Raed Farzan

30<sup>th</sup> July 2019

## **Copyright declaration**

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non-Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

***“Verily, all things have We created in proportion and measure.”***

**Qur’an 54:49**

## Acknowledgements

I would firstly like to sincerely thank my supervisors Professor Laki Buluwela, Professor Simak Ali and Dr Manikandan Periyasamy, for all of their patience, enthusiasm and support. I have been extremely lucky to work with Professor Laki Buluwela who cared so much about my work, and who responded to my questions and queries so promptly.

I am grateful to our group members, past and present, Dr Manikandan Periyasamy, Dr Carolina Gemma, Dr Anup Singh, Miss Van Nguyen, Dr Hetal Dattani, Dr Joel Fulton, Dr Meng-Lay Lin, Dr Chun-Fui Lai, Miss Laura Ramos Garcia, Dr Kirsty Balachandran, Dr Georgina Sava, Miss Isabella Goldsbrough, Miss Hailing Fan and Dr Alison Harrod. It has been a pleasure to work in such a friendly and supportive environment.

For their contributions to this project, I would like to thank Dr Manikandan Periyasamy, who were there to help since the very first day until the day of submission. Miss Van Nguyen for her kind help in RNAseq. Analysis, Dr Anup Kumar Singh, who I am indebted to for several discussions that helped me in APOBEC work, Dr Alison Harrod for her help in cloning work, and Dr Rosemary Fisher for karyotyping of my engineered line.

I am also thankful to my home institution, King Saud University for sponsoring my PhD scholarship, and the Cultural Bureau at the Royal Embassy of Saudi Arabia in London for their unlimited support.

My deepest gratitude goes to my wife, Sumayah, for her love, belief and help in writing technical skills and to my brothers and sisters; Mohammed, Dr Ahlam, Amirah, Ebtehal, Atheer, Abdulrahman and Abdulaziz for their continuous encouragements.

Thanks to my friends who made things possible, particularly His Excellency Dr Yazeed Alsheikh, His Excellency Dr Abdulmajeed Albanyan, Dr Saad Alarifi, Mr Abdullah Aljumaah, Dr Bader Aldawaish, Dr Ali Alsiehem, Mr Saud Alwadani, Dr Habeeb Alhabeeb, Mr Yaala Assiri and Mr Hosam Alzahrani.

I wholeheartedly would like to thank my parents for not only bearing my absence for years but for their generous support during this time. Therefore, I would like to dedicate this thesis to my first teacher, my Mother, Turkiah Alghamdi, and my first friend, my Father, Ahmed Farzan.

## Table of Contents

Abstract .....	2
Declaration of originality .....	3
Copyright declaration.....	4
Acknowledgements.....	6
List of figures.....	12
List of tables.....	15
Abbreviations.....	16
1 - Introduction .....	19
1.1 – Signatures of mutations in human cancer .....	19
1.2 - The APOBEC/AID family of proteins .....	23
1.2.1 - APOBEC mediated mutation in cancer .....	24
1.3 - APOBEC3 .....	28
1.4 – APOBEC3B .....	31
1.5 – APOBEC3H.....	32
1.5.1 - APOBEC3H dimerisation by RNA .....	33
1.6 – p53.....	34
1.7 - APOBECs and cancer.....	40
1.8 - Uses of CRISPR/CAS9 as a gene editing tool in human cancer cell lines.....	42
1.8.1 - Reverse genetic approaches in human cell lines .....	42
1.9 – Aims .....	46
2 – Materials and methods .....	47
2.1 - Materials .....	47
2.1.1 - General reagents, materials and equipment.....	47
2.1.2 - Sundries .....	47
2.1.3 - General equipment.....	48
2.1.4 - Cell culture reagents .....	49
2.1.5 - Chemicals and antibiotics.....	49

2.1.6 - General stock solutions.....	49
2.1.7 - Microbiological reagents .....	51
2.2 – Methods.....	51
2.2.1 - Tissue culture.....	51
2.2.2 - Preparation and use of frozen cell stocks .....	52
2.2.3 - Cell counting and cell viability measurement .....	53
2.2.4 - Measurement of cell growth using the Sulforhodamine B (SRB) assay .....	53
2.2.5 - Cell cloning.....	54
2.2.6 - Nucleofection of cells.....	55
2.2.7 - Gibson assembly cloning of CRISPR sgRNA sequences .....	56
2.2.7.1 - Generation of the CRISPR guide RNAs fragment for cloning .....	56
2.2.7.1.1 - CRISPR template hybridisation.....	56
2.2.7.1.2 - CRISPR template end-repair .....	57
2.2.7.1.3 - PCR amplification and purification of double stranded CRISPR gRNA DNA template .....	57
2.2.7.2 - p41824 pgRNA-U6 vector preparation .....	59
2.2.7.3 - Cloning of double stranded CRISPR gRNA DNA sequences into the p41824 pgRNA-U6 vector using Gibson assembly.....	60
2.2.7.4 - Transformation of chemically competent cells with Gibson assembly reaction products .....	61
2.2.8 - Plasmid preparation .....	61
2.2.8.1 - Plasmid mini preparation.....	62
2.2.8.2 - Plasmid maxi preparation .....	62
2.2.9 - Polymerase chain reaction (PCR).....	63
2.2.10 - Genomic DNA preparation.....	64
2.2.11 - Agarose gel electrophoresis.....	65
2.2.12 - Extraction of total RNA from cultured cells .....	65
2.2.13 - Complementary DNA (cDNA) preparation by reverse transcription.....	66
2.2.14 - TaqMan quantitative real-time PCR (qRT-PCR) .....	67
2.2.15 - RNA-seq library construction.....	69
2.2.16 - RNA-sequencing analysis.....	69
2.2.17 - Protein lysate preparation .....	70
2.2.18 - Determination of protein concentration.....	70



2.2.19 - SDS-Polyacrylamide gel electrophoresis and Western blotting.....	71
2.2.20 - Sanger DNA sequencing .....	72
3 – Results 1: p53 regulation of APOBEC3B and APOBEC3H expression.....	72
3.1 - p53 regulation of the APOBEC3 locus.....	73
3.2 - Generation of p53 knock out MCF7 line.....	73
3.3 - Cloning and characterisation of p53 directed sgRNAs .....	76
3.4 - CRISPR mediated p53 coding exon 3 deletion in MCF7 Luc cells .....	76
.....	87
3.5 - p53 expression and functionality in the MCF7 Luc $\Delta$ p53 MA2 line .....	88
3.6 - Use of the MCF7 Luc $\Delta$ p53 MA2 line to study regulation of APOBEC3H .....	93
3.7 - Discussion.....	94
4 - Results 2: Generation of an APOBEC3H gene knockout in the HCT116 human cancer cell line.....	98
4.1 - APOBEC3H knockout in the HCT116 cell line.....	98
4.2 - CRISPR-Cas9 mediated knockout of APOBEC3H by indel mutation. ....	100
4.3 - Design of CRISPRs for making DNA deletions of APOBEC3H exon3.....	102
4.4 – Establishment of HCT116 APOBEC3H knockout lines .....	105
4.5 - Isolation and initial characterisation of HCT116 APOBEC3H knockout lines .....	108
4.6 - Characterisation of gene expression from the APOBEC3H locus in HCT116 cells and the HCT116- APOBEC3H exon3 deleted clones 2B3 and 2C3.....	114
4.7 - Sub-cloning of the HCT116- APOBEC3H exon3 deleted clones 2B3 and 2C3 and characterisation of the APOBEC3 locus in the APOBEC3H knockout lines.....	118
4.8 - Growth and sensitivity to Nutlin in the HCT116 2B3 and 2C3 subclones.....	121
4.9 - Discussion.....	123
5 - Results 3: RNAseq analysis of the APOBEC3H gene knockout constructed in the HCT116 human colon cancer cell line.....	127
5.1 - RNAseq analysis of HCT116 A3H knockout lines 2B3-2B2 and 2C3-1B2.....	127

5.2 - Overview of strategy .....	129
5.3 - RNAseq .....	138
5.3.1 - RNAseq – library construction and next generation sequencing .....	138
5.3.2 - RNAseq analysis.....	138
5.3.3 - Analysis of differential expression between HCT116 wild-type, HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cells.....	142
5.3.4 - Down-regulated gene expression in the APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 indicates a loss of the Y chromosome.....	146
5.3.5 - Up-regulated gene expression in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 is enriched for processed pseudogene transcription. ....	153
5.3.6 - Gene set enrichment analysis of differentially regulated gene expression in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines .....	158
5.3.7 - Evaluation of the Nutlin - p53 response in wild-type HCT116 cells and in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines.....	160
5.4 - Discussion.....	173
5.4.1 - Differential gene expression in the HCT116 APOBEC3H knockout lines and loss of the Y chromosome.....	173
5.4.2 - Differential gene expression in the HCT116 APOBEC3H knockout lines and processed pseudogene expression.....	175
5.4.3 - Differential gene expression in the HCT116 APOBEC3H knockout lines and p53 regulated gene expression. ....	176
6 – Discussion .....	178
6.1 - APOBEC3H and p53 .....	178
6.1.1 - APOBEC3H as a mediator of retroviral resistance .....	178
6.1.2 - APOBEC3H in the inhibition of retrotransposon mobilisation.....	180
6.1.3 - p53 induction of APOBEC3H as a mechanism to inhibit retroviral and retrotransposable propagation.....	181
6.2 - CRISPR-Cas9 genome engineering.....	185
6.3 - Summary and future work .....	187
References.....	190

Supplementary 1. p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells .....211

Supplementary 2. The responses of cancer cells to PLK1 inhibitors reveal a novel protective role for p53 in maintaining centrosome separation.....225

## List of figures

Figure 1.1. Somatic mutational signatures – scope and mechanisms .....	21
Figure 1.2. The 96-tumour mutation signature model adapted from Alexandrov <i>et al.</i> , (2013b) .....	22
Figure 1.3. Cytidine deamination and the human APOBEC family.....	26
Figure 1.4. Overview of human APOBEC cytidine deaminase function .....	27
Figure 1.5. The APOBEC3 gene locus .....	30
Figure 1.6. Key p53 target genes and pathways .....	37
Figure 1.7. p53 and gene regulatory networks.....	39
Figure 1.8. Overview of CRISPR-CAS9 system.....	45
Figure 3.1 TP53 gene structure and domains organization.....	75
Figure 3.2. TP53 coding exon 4 (coding exon 3) CRISPR guide RNA cloning .....	79
Figure 3.3. CRISPR 138076 -138077 / Cas9 mediated DNA deletion in TP53 exon 4 (coding exon 3) .....	80
Figure 3.4. PCR characterisation of MCF7 Luc clones following CRISPR138077/ CRISPR138076/ Cas9 co-transfection.....	84
Figure 3.5. Differential growth inhibition by Nutlin in wild-type and p53 knockout MCF7 Luc cells.....	85
Figure 3.6. TP53 exon 3 allele sequences from wild-type and TP53 knockout MCF7 Luc cells .....	87
Figure 3.7. Western blot analysis of p53 and p53 regulated expression in wild-type and p53 null, CRISPR-Cas9 engineered MCF7 Luc cells.....	89
Figure 3.8. qRT-PCR analysis of p53 regulated gene expression in MCF7 Luc wild-type and MCF7 Luc $\Delta$ p53 MA2 cells.....	92
Figure 4.1. APOBEC3H gene expression in MCF7 Luc and HCT116 cell lines.....	99
Figure 4.2. APOBEC3H targeted gene editing using CRIPSR 182078 .....	101
Figure 4.3. Design of CRISPR guide RNAs for APOBEC3H exon 3.....	104
Figure 4.4. Cloning of CRISPR guide RNAs for APOBEC 3H exon 3 into the U6 CRISPR sgRNA vector.....	106
Figure 4.5. Optimised plasmid nucleofection of HCT116 cells .....	107
Figure 4.6. PCR characterisation of HCT116 clones following CRISPR guide13H 3738/ guide23H 3940/ Cas9 co-transfection.....	110

Figure 4.7. DNA sequence characterisation of HCT116 APOBEC3H knockout clones 2B3 and 2C3 .....	111
Figure 4.8. Western blot analysis of HCT116 APOBEC3H knockout clones 2B3, 2C3 and 2A .....	113
Figure 4.9. APOBEC3H transcript variants.....	116
Figure 4.10. APOBEC3H transcript variants PCR analysis .....	117
Figure 4.11. Integrity of the APOBEC3 locus in the HCT116 and HCT116 APOBEC3H 2B3 and 2C3 knockout cells.....	120
Figure 4.12. Growth and Nutlin sensitivity of HCT116 wild-type and HCT116 APOBEC3H knockout cells .....	122
Figure 5.1. Culture morphology of wild-type HCT116 cells and APOBEC3H knockout HCT116 lines 2B3-2B2 and 2C3-1B2.....	128
Figure 5.2. Overview of sample generation for RNAseq analysis of HCT116 cells and HCT116 $\Delta$ A3H lines 2B3-2B2 and 2C3-1B2 .....	131
Figure 5.3. Overview of sample generation for RNAseq analysis of HCT116 cells and HCT116 $\Delta$ A3H lines 2B3-2B2 and 2C3-1B2 .....	132
Figure 5.4. p21 relative expression in HCT116 cells and APOBEC3H knockout HCT116 knockout cells .....	135
Figure 5.5. Assessment of genomic DNA contamination in RNA preparations .....	137
Figure 5.6. Overview of RNAseq analysis .....	140
Figure 5.7. Principal component analysis for control (DMSO) replicates in the two APOBEC3H KO lines compared to HCT116 WT cells .....	141
Figure 5.8. Differentially expressed genes shared by the APOBEC3H HCT116 knockout cell lines 2B3-2B2 and 2C3-1B2.....	144
Figure 5.9. Differentially expressed genes shared by the APOBEC3H HCT116 knockout cell lines 2B3-2B2 and 2C3-1B2.....	145
Figure 5.10. Y-Chromosome gene expression is down-regulated in the HCT116 APOBEC3H knockout cell lines .....	150
Figure 5.11. Down-regulated gene expression in the HCT116 APOBEC3H knockout cell lines.....	152
Figure 5.12. Up-regulated gene expression in the HCT116 APOBEC3H knockout cell lines .....	157
Figure 5.13. Gene set enrichment analysis of differentially regulated gene expression in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines .....	159

Figure 5.14. Evaluation of the Nutlin - p53 response in wild-type HCT116 cells and in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines.....	162
Figure 5.15. Evaluation of the Nutlin - p53 response in wild-type HCT116 cells and in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines – three-way analysis .....	163
Figure 5.16. The Nutlin - p53 gene signature is seen in HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines – three way and core p53 comparison.....	166
Figure 5.17. Comparative Nutlin regulated gene expression between HCT116 cells and in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines.....	169
Figure 5.18. Gene expression for p53 and p53 target genes in HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines .....	171
Figure 5.19. Nutlin induced gene expression following APOBEC3H over-expression in wild-type HCT116 cells .....	172

## List of tables

Table 2.1: General Stock Solution .....	50
Table 2.2: TaqMan Real-time PCR assay numbers (Applied Biosystems, Life Technologies, Paisley, UK).....	68
Table 5.1. Gene expression differences between wild-type HCT116 cells and the APOBEC3H HCT116 knockout lines	143
Table 5.2. Top 50 common down-regulated genes in the two APOBEC3H knockout cell lines .....	147
Table 5.3. Genotyping of HCT116 wild-type cells and the APOBEC3H HCT116 knockout cell lines 2B3-2B2 and 2C3-1B2 .....	151
Table 5.4. Top 49 common up-regulated genes in the two APOBEC3H knockout cell lines .....	154

## Abbreviations

APOBEC	Apolipoprotein B mRNA editing, enzyme catalytic
aa	Amino acid
AID	Activation induced cytidine deaminase
BSA	Bovine serum albumin
bp	Base pair
Ct	Cycle threshold
cDNA	Complementary DNA
CRISPR	Clustered regularly interspaced short palindromic repeats
Cas	CRISPR-associated
COSMIC	Catalogue of somatic mutations in cancer
CTD	C-terminal domain
CDK	Cyclin dependent kinase
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation sequencing
DNA	Deoxyribonucleic acid
dsDNA	Double strand DNA
dNTP	Deoxyribonucleotide triphosphate
DBD	DNA binding domain
DEG	Differential expressed gene
DMSO	Dimethyl sulfoxide
ddH <sub>2</sub> O	Double distilled water
DMEM	Dulbecco's Modified Eagle Medium
DNA-PK	Dependent protein kinase
DNA DNMT	DNA methyltransferase
DSB	Double strand DNA break



ER $\alpha$	Estrogen Receptor $\alpha$
EtOH	Ethanol
EDTA	Ethylenediaminetetraacetic acid
ER	Estrogen receptor
FCS	Fetal calf serum
gDNA	Genomic DNA
GFP	Green fluorescence protein
GADD45	Growth arrest and DNA-damage-inducible protein
HDR	Homology-directed repair
Indel	Insertion/deletion
Kb	Kilo-base pair
LBD	Ligand binding domain
mRNA	messenger RNA
MAPK	Mitogen activated protein kinase
mg	Milligram
ml	Millilitre
ug/ $\mu$ g	Microgram
5meC	5-methyl cytosine
5hmC	5-hydroxymethyl cytosine
5hmU	5-hydroxymethyl uridine
miRNA	MicroRNA
NR	Nuclear receptor
NHEJ	Non-homologous end joining
nt	Nucleotide
Luc	Luciferase

PCR	Polymerase chain reaction
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
pre-crRNA	Precursor-CRISPR-RNA
PSG	L-glutamine-penicillin-streptomycin
qPCR	Quantitative PCR
RIPA	Radioimmune precipitation buffer
RNA-seq	RNA-sequencing
RT	Reverse transcription
RT-PCR	Real-time PCR
RNA	Ribonucleic acid
RIPA	Radioimmune precipitation buffer
RFP	Red fluorescent Protein
ssDNA	Single strand DNA
sgRNA	Single guide RNA
SNP	Single nucleotide polymorphism
TSS	Transcription start site
tracrRNA	Transactivating crRNA
WT	Wilde-type
△	Mutant

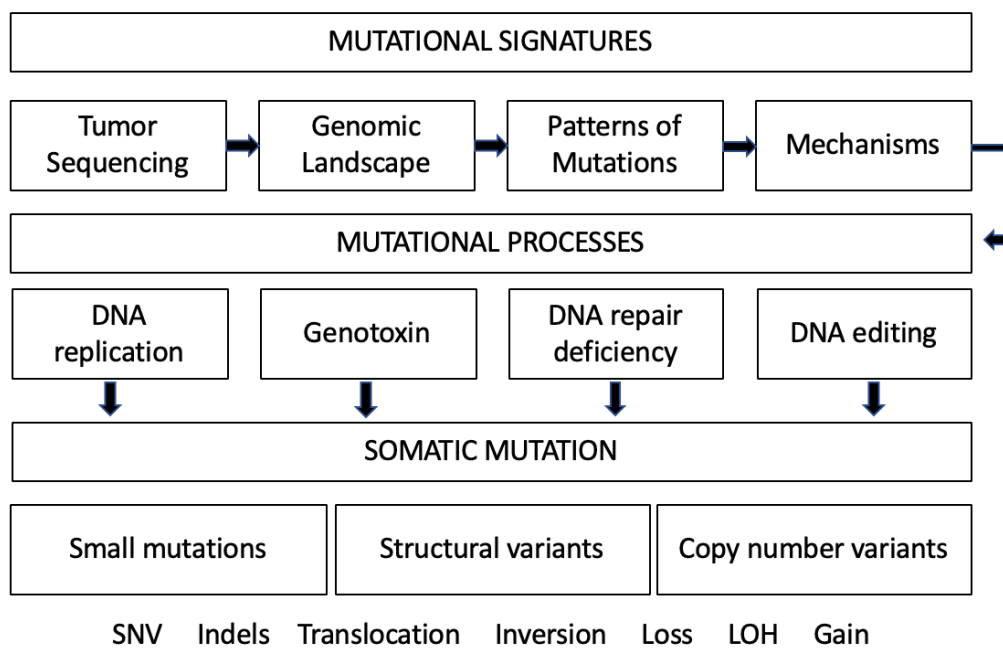
## **1 - Introduction**

### **1.1 – Signatures of mutations in human cancer**

The large-scale genome sequencing of patient cohorts has described a landscape of somatic mutations, and it is these mutations occurring during the lifetime of an individual, that cause a very significant percentage of cancers in humans (Forbes *et al.*, 2017). Mutation occurs in the DNA of the cell, such that its genome is imprinted with a corresponding mutational signature of the event (Stratton, 2013). The application of a mathematical approach to the analysis of large numbers of mutations has recently led to a cataloguing of somatic mutations from different types of cancers in humans (Alexandrov *et al.*, 2013), and has facilitated the identification of a set of mutational signatures that encompass ten types of genomic rearrangement and over 40 types of base substitutions (Alexandrov, 2018).

A “mutational signature” is a characteristic combination of different types of mutation resulting from specific processes of mutagenesis, including infidelity in DNA replication, exposure to endogenous and exogenous genotoxins, defective DNA repair pathways and enzymatic DNA editing. Interpreting a cancerous mutational signature helps gain an understanding of the biological mechanisms in carcinogenesis and normal somatic mutagenesis (Figure 1.1) (Forbes *et al.*, 2017). The analysis of mutational signatures for cancerous agents requires data from cancer genome sequencing along with the sequencing of paired-normal DNA, in order to build a catalogue of the types and frequency of mutations in specific tumours. Different mutation types, which can include structural variants, single nucleotide variants and indels, can be used either separately or in combination to model cancerous mutational signatures (Alexandrov *et al.*, 2013).

183,016 substitutions catalogued in 21 genomes of breast cancers show that mutational signatures exist in this tumour type, with this being carried out by means of a proof-of-principle exercise (Nik-Zainal *et al.*, 2012). How each substitution should be classified was decided with regard to the sequence 5' and 3' to each mutated base. As a result, there are six base substitution classes and, for each mutated base, 16 possible sequence contexts (A, C, G or T at both the 5' base and the 3' base). It follows that 96 possible mutated trinucleotides exist for each tumour (Figure 1.2A). The five signatures for substitutions in these tumours: A-E, now known as 1B, 2, 3, 8 and 13, were extracted mathematically (Nik-Zainal and Morganella, 2017). The following associations were observed to exist for these signatures – 1, 1B and 5: age of diagnosis, 2 and 13: APOBEC cytidine deaminases activity. 3: deficiency in *BRCA1/BRCA2*. 6, 20 and 26: mismatch repair deficiency. Higher levels of *BRCA1/BRCA2* loss have been recorded for tumours with Signature 8, though it was also found elsewhere at lower levels. No etiology has yet been established for Signatures 17, 18 and 30 (Nik-Zainal *et al.*, 2019). During tumour evolution, signatures are seen to combine, so as to contribute to the overall tumour signature (Figure 1.2B).

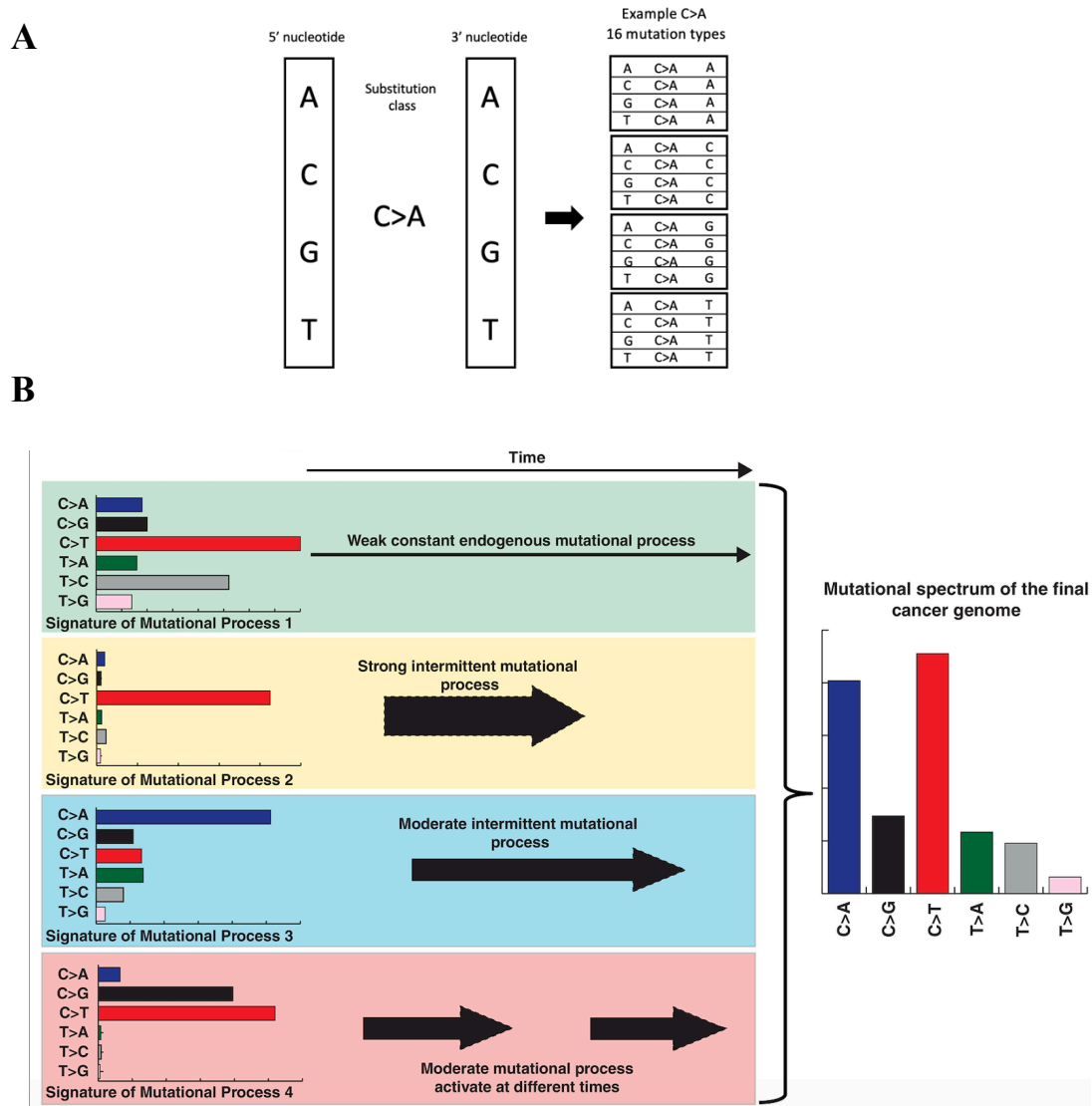


**Figure 1.1. Somatic mutational signatures – scope and mechanisms**

Tumour somatic landscapes are shaped by a range of mutagenesis processes, and interpreting the patterns underlying cancer mutations reveals how recurring mutational patterns are related to each other and makes it possible to infer their possible causes. The mechanisms involved in mutational signatures fall into the following categories:

- 1 - DNA replication infidelity: Nucleotides that have been wrongly incorporated are removed by DNA polymerase in an exonuclease enzymatic reaction – a process known as “DNA proofreading”. When DNA polymerase cannot correct replication errors, continuing cell mitosis causes mutations to accumulate progressively.
- 2 – Genotoxins: Endogenous cellular mutations include, for example, C>T transition caused by spontaneous 5-methylcytosine deamination. There are a number of types of exogenous carcinogens. They include: Ultraviolet (UVB) radiation which directly damages DNA and is known to increase the risk of melanoma and other skin cancers. Alkylating antineoplastic agents, a group of chemotherapy agents which cause DNA crosslinking and interfere with repair and replication of DNA through addition to DNA of the alkyl group. Their high mitosis rate makes cancer cells particularly vulnerable.  
Tobacco: Among the carcinogens harmful to DNA are acrolein, cyanide, nitrosamines and polycyclic aromatic hydrocarbons (see health effects of tobacco).
- 3 - DNA repair deficiency: HRD (homologous recombination deficiency) causes double-strand breaks in DNA. Accurate breakpoint repair requires a homologous recombination mechanism. MMR (DNA mismatch repair) deficiency, on the other hand, is a shortfall in the mechanism by which erroneous insertion, deletion or incorporation of base pairs is identified and repaired.
- 4 - Enzymatic DNA editing: Cytidine deaminase enzymes include cytidine deaminase/CDA, activation-induced cytidine deaminase and the APOBEC protein family) form part of the natural immune system. They help to control retroviruses and transposons elements including endogenous retroviruses. Cytidine deamination leading to C>T transition (genetics) mutations is caused by these enzymes.

Adapted from Helleday *et al.*, (2014) and Robert Weinberg, R. (2006). *The Biology of Cancer*. 2nd ed. New York: Garland Science



**Figure 1.2. The 96-tumour mutation signature model adapted from Alexandrov *et al.*, (2013b)**

By considering the 5' flanking base (A, C, G, T), the 6 substitution classes (C>A, C>G, C>T, T>A, T>C, T>G) and 3' flanking base (A, C, G, T) Alexandrov *et al* suggested 96 possible mutation types ( $4 \times 6 \times 4 = 96$ ). **A.** The 16 possible mutation types, arising from the substitution C>A are shown as an example. **B.** In this example, taken from (Alexandrov and Stratton, 2014), the six classes of somatic substitutions result in a unique mutational signature in a tumour. In the clonal progenitor of the tumour, initial mutations come from mutational process 1. As the tumour progresses, other mutational process engages, with the final tumour genome reflecting these mixed contributions (Alexandrov and Stratton 2014).

## 1.2 - The APOBEC/AID family of proteins

The APOBEC (Apolipoprotein B Editing Complex) family of proteins comprises eleven members that function as C-to-U nucleotide editing enzymes (Figure 1.3A). They are the key to the processes of deamination that results in somatic and germline DNA mutations (Wedekind *et al.*, 2003). Subgroupings of the eleven APOBECs are APOBEC1; APOBEC2; the APOBEC3 sub-family (APOBEC3-A, B, C, D, F, G, H); APOBEC4 and AID, the Activation-induced cytidine deaminase (Figure 1.3B). APOBECs are similar in structure to the zinc-dependent deaminases, as they share a catalytic backbone, with the zinc-dependent deaminases usually involved in purine and pyrimidine metabolism (Mitra *et al.*, 2015; Shi *et al.*, 2017). Overall, the AID/APOBEC protein structure comprises five  $\beta$  strands and  $\alpha$  helices, with a catalytic pocket defined by a single zinc atom co-ordinated with histidine and cysteines and by a water molecule, which is activated by the zinc ion and results in the nucleophilic attack of the C4 atom in the cytidine ring (Figure 1.3A) (Salter *et al.*, 2016). A conserved glutamine residue of the catalytic domain acts as a proton shuttle to aid in the transfer of a proton from water to the N3 nitrogen of the cytidine ring, donating a proton to the leaving ammonia group. The catalytic domain is encoded by the APOBEC proteins N-terminal domain, and in four of the eleven members, a pseudocatalytic domain is encoded by the C-terminal domain.

APOBEC genes are encoded on four human chromosomes, with APOBEC1 and AID genes on chromosome 12, APOBEC2 on chromosome 6 and APOBEC4 on chromosome 1 (Figure 1.3B). Among the functions of human APOBEC are editing RNA, somatic hypermutation during immunoglobulin heavy chain class switch recombination, retrovirus and endogenous

retroelements restriction, DNA induced inflammation suppression, and genomic DNA deamination (Figure 1.4A) (Adolph *et al.*, 2018).

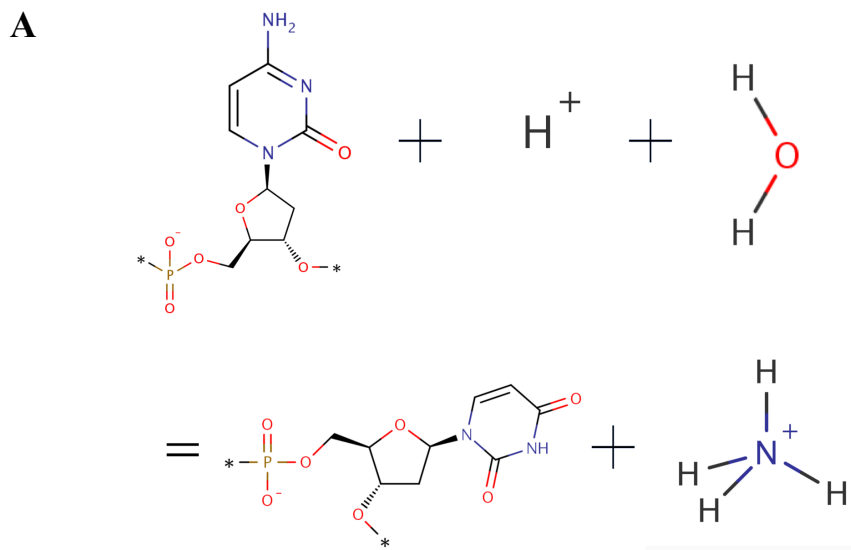
### **1.2.1 - APOBEC mediated mutation in cancer**

A recent study investigating single-base substitution mutation signatures in more than one thousand cancer cell lines analysed the APOBEC mutation signature by means of two further approaches (Jarvis *et al.*, 2018). In the majority of cases, more than one mutation signature was evident in a cell line. The “ageing” signature, for example, which is more frequently observed in primary tumours than any other signatures, is also seen in most cancer cell lines. Further, cancer cell lines from bladder, breast, cervical, lung and neck and head cancers showed enriched APOBEC mutation signatures, which also constituted a large percentage of all mutations (Jarvis *et al.*, 2018).

It is now recognised that APOBEC mediated enzymatic cytosine deaminase activity on DNA leads to mutations in different tumour types and are considered to be a major source of mutations (Roberts and Gordenin, 2014). As outlined above, mutational processes result from endogenous or exogenous processes, and occur with different “strengths” during the tumour process (Stratton, 2011). Broadly, these can be considered as “driver mutations” that participate in tumour formation, and other “passenger mutations”, which are considered to be neutral as regards tumour formation and progression, but which provide useful markers to study tumour evolution (de Bruin *et al.*, 2014). In this context, C > T is the major source of mutation in many cancers (Nik-Zainal *et al.*, 2014), and this signature corresponds to APOBEC activity on cytidine deamination. APOBEC cytosine deaminases have been suggested to enhance subclonal expansions and intratumoral heterogeneity, since APOBEC mutational signatures may be elevated in tumour subclones (Figure 1.4B) (Swanton *et al.*, 2015). However, more



recent studies suggest that APOBEC mediated mutations may occur in an “episodic” manner, but the triggers for this type of activity are unclear (Petljak *et al.*, 2019).



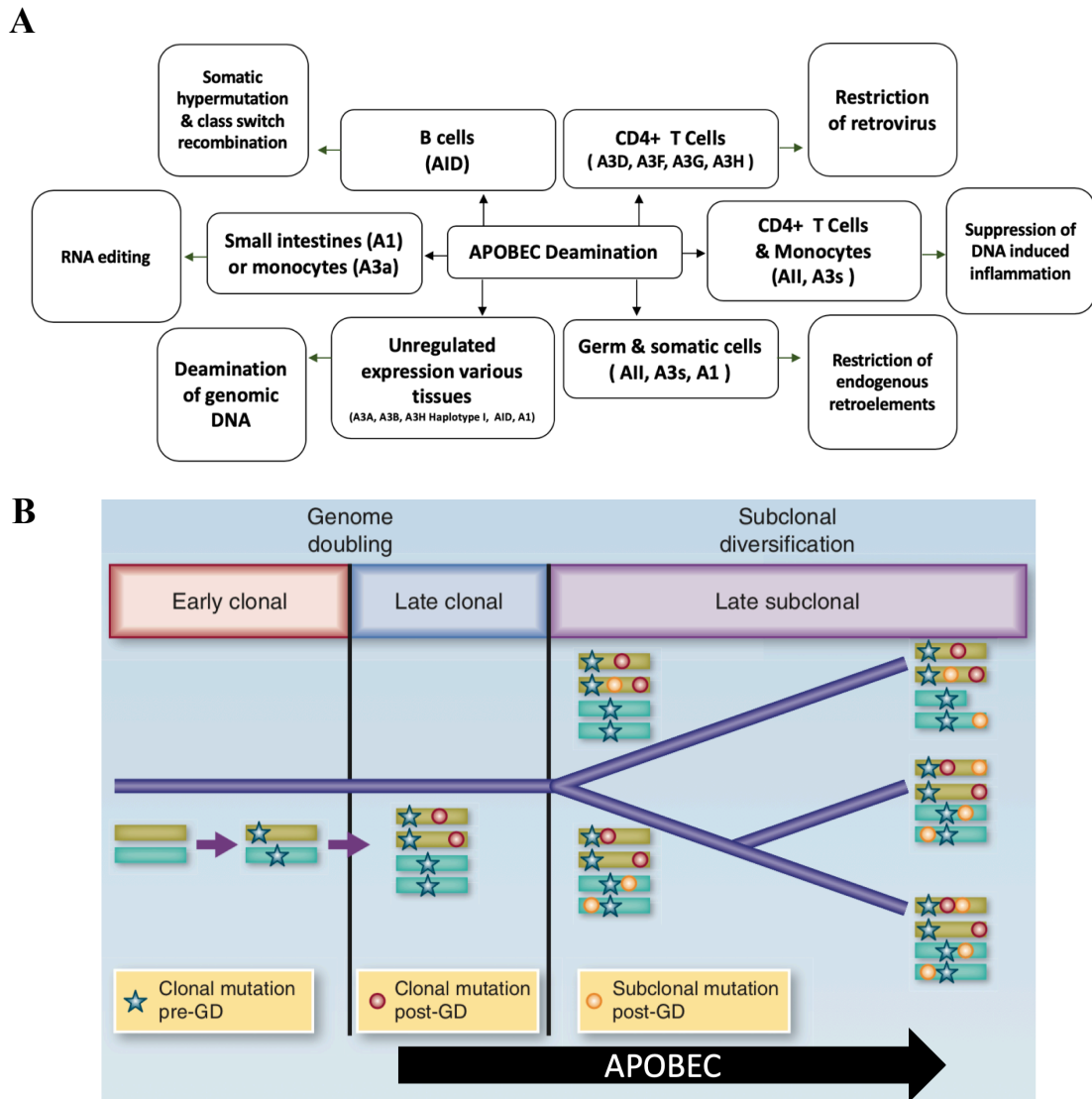
**B**

Gene location	APOBEC	Size (aa)	Protein Organisation	
			CD1	CD2
Chr 12	APOBEC1	236	Shaded red	
	AID	198	Shaded red	
Chr 6	APOBEC2	242	Shaded red	
	APOBEC3A	199	Shaded red	
Chr 22	APOBEC3B	382	Shaded red	Shaded red
	APOBEC3C	190	Shaded red	
	APOBEC3D/E	386	Shaded red	Shaded red
	APOBEC3F	373	Shaded red	Shaded red
	APOBEC3G	384	Shaded red	Shaded red
	APOBEC3H	200	Shaded red	
Chr 1	APOBEC4	367	Shaded red	

**Figure 1.3. Cytidine deamination and the human APOBEC family**

**A.** Cytidine deamination is the hydrolytic deamination of cytosine via the removal of the amino group at the fourth position of the pyrimidine ring, leading to the formation of uracil. The reaction occurs spontaneously, at a low rate, and is catalysed by members of the APOBEC family of cytosine deaminases, which encode up to two copies of a generic cytosine deaminase domain. Uniprot.org

**B.** The APOBEC gene family comprises 11 members, encoded by genes on four human chromosomes. The activation induced deaminase (AID) and APOBEC1 genes are located on chromosome 12 at two loci separated by ~1MB. The APOBEC2 gene is located on chromosome 2. The APOBEC3 genes are clustered in a 145 Kb region on Chromosome 22. APOBEC4 is encoded on chromosome 1. The sizes of the APOBEC proteins encoded by each gene is listed and their organisation shown as block diagrams. APOBEC proteins contain cytosine deaminase (CD) catalytic domains (shaded red), which are characterised by the conserved amino acid sequence motif “His-X-Glu-X23-28-Pro-Cys-X2-4-Cys”, where the cysteine and histidine residues coordinate with a single  $Zn^{2+}$  ion. APOBECs 1, 3A, 3C, and 3H each contain only one CD domain (CD1), while APOBEC3B, APOBEC3D/E, APOBEC3F and APOBEC3G each contain two CD domains (CD1 and CD2). Adapted from Goila-Gaur and Strebel, (2008).



**Figure 1.4. Overview of human APOBEC cytidine deaminase function**

A. APOBEC enzymes are widely expressed, with some family members showing both a tissue and cell specific manner. APOBEC enzymes can edit RNA in the small intestine (A1) as well as in monocytes (A3A). AID deaminations in B-cells initiate antibody diversification pathways for somatic hypermutation and class switch recombination. Replication of CD4+ T cells is restricted by retroviral ssDNA intermediates and A3 deamination, while foreign DNA can be restricted in the same cells and in monocytes by A3's leading to a reduced DNA-induced inflammatory response (Prohaska *et al.*, 2014). APOBEC enzymes can restrict retrotransposons in germ and somatic cells through RNA binding or by deaminating reverse transcripts. Unregulated APOBEC expression in various tissues can be the cause of unregulated deamination in genomic DNA leading to cell transformation or instability of the genome, ultimately leading to cancer. The figure above shows the APOBECs involved in each process (Prohaska *et al.*, 2014). Adapted from Adolph *et al.*, (2018) **B**. Using DNA sequencing to understand the temporal acquisition of DNA mutations during tumour evolution has led to the conclusion that APOBEC mediated mutations occur before subclonal diversification, but after the early clonal stage. Adapted from Swanton *et al* (2015).

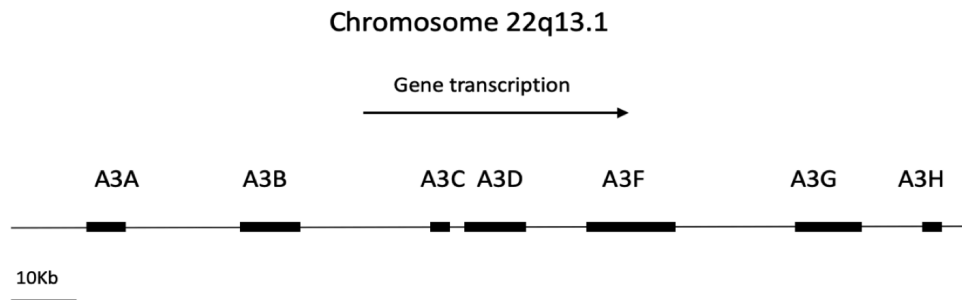
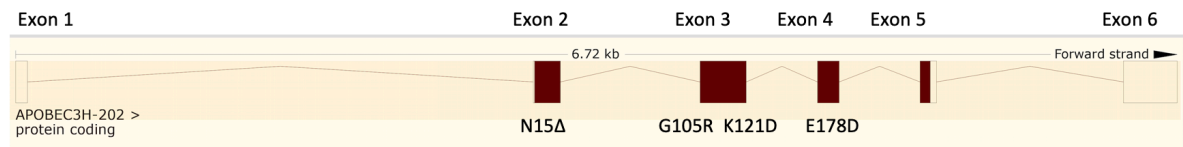
### 1.3 - APOBEC3

All the APOBEC3 genes are encoded in a head to tail cluster in a 145 Kb region of chromosome 22q13.1 (Figure 1.5A) (Conticello, 2008). These are expressed in most cell types and tissues, with being overexpressed in cancer cell lines and tissues (Conticello, 2008). APOBEC3A, APOBEC3C and APOBEC3H are distinguished by encoding single zinc-finger domains, while APOBEC3B, APOBEC3D, APOBEC3G and APOBEC3F encode two zinc-finger domains (Rebhandl *et al.*, 2015). APOBEC3 enzymes inhibit retroviral replication by deaminating retroviral DNA intermediates and, in so doing play a significant role in innate immunity (Harris and Liddament, 2004). Amongst APOBEC proteins, APOBEC3G was the first shown to have strong anti-HIV-1 activity (Sheehy *et al.*, 2002). APOBEC3G deaminates cytidine (dC) to uridine (dU) specifically in the retroviral minus strand, after reverse transcription to cDNA. As a result, this leads to base transitions of guanosine (G) to adenine (A) in the viral genome (Harris and Liddament, 2004; Mangeat *et al.*, 2003). However, to counteract this, HIV-1 is able to inhibit the anti-retroviral function of APOBEC3G through the production of the virally encoded virion infectivity factor (VIF) (Goila-Gaur and Strebel, 2008; Repke *et al.*, 1992). VIF causes APOBEC3G degradation through polyubiquitylation and also interferes with APOBEC3G protein translation (Marin *et al.*, 2003; Stopak *et al.*, 2003).

Other APOBEC3 family proteins are also found to inhibit retroviral infection strongly. For example, APOBEC3B and APOBEC3F can significantly reduce retroviral infection (Bishop *et al.*, 2004), with APOBEC3B being able to inhibit both wild-type and VIF deficient HIV-1 (Bishop *et al.*, 2004; Harris and Liddament, 2004). APOBEC3D, APOBEC3C and APOBEC3A are weak HIV-1 inhibitors and VIF sensitive (Bishop *et al.*, 2004; Bourara *et al.*, 2007). By contrast, while APOBEC3H is also seen to be effective against retroviral replication, the expression of this protein is found to be low in primates, where APOBEC mediated

retroviral protection is a significant mechanism (OhAinle *et al.*, 2006). Finally, while cytidine deamination was thought to be the main mechanism of antiviral activity for APOBEC proteins, more recent studies have suggested APOBEC mediated resistance in the absence of DNA editing (Holmes *et al.*, 2007).

APOBEC3 proteins also protect against other viruses including human T-cell lymphotropic virus, hepatitis B virus, hepatitis C virus (HCV), human papillomavirus (HPV) and human herpesviruses (LaRue *et al.*, 2008). APOBEC3 enzymes are also able to restrict the movement of non-LTR and LTR retrotransposable elements, including both LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements) (Chiu and Greene, 2008). APOBEC3D, APOBEC3F and APOBEC3G are all present in the cytoplasm, while APOBEC3A, APOBEC3C and APOBEC3H are in both cytoplasm and the nucleus (Kinomoto *et al.*, 2007).

**A****B**

**Figure 1.5. The APOBEC3 gene locus**

**A.** The APOBEC 3 gene locus encodes seven APOBEC3 genes that are clustered head to tail and span 83,126 bp (A-D, F-H). Adapted from Duggal, Fu *et al.*, (2013) **B.** The organization of the human APOBEC3H gene. The APOBEC3H gene is organised into six exons, the first of which is non-coding. The gene shows a high level of polymorphism, with five different single amino acid variants - N15Δ, R18L, G105R, K121D, and E178D. The exon location of the involved residues is as indicated. Adapted from

[https://www.ensembl.org/Homo\\_sapiens/Gene/Summary?db=core;g=ENSG00000100298;r=22:39097224-39104067](https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000100298;r=22:39097224-39104067) .

## 1.4 – APOBEC3B

Overexpression of APOBEC3B occurs in a number of human cancer types, correlating with APOBEC3B tumour mutation signatures (Burns *et al.*, 2013). Deletion of the APOBEC3B coding region in chromosome 22 has been observed to be a naturally occurring genetic variant (Kidd *et al.*, 2007), and results in the fusion of APOBEC3A and APOBEC3B genes, which replaces the 3' UTR region in APOBEC3A with that from APOBEC3B. The different stability of the resulting fusion gene RNA results in different modes of expression and regulation of an APOBEC3A/ APOBEC3B like protein (Mussil *et al.*, 2013; Shi *et al.*, 2017; Wijesinghe and Bhagwat, 2012). The APOBEC3B deletion polymorphism leads to an elevated risk for breast cancer (Cescon *et al.*, 2015; Gohler *et al.*, 2005; Nik-Zainal *et al.*, 2014), but shows no phenotypic difference in breast cancer cells (Cescon *et al.*, 2015). However, there is an association between poor patient survival in estrogen receptor positive breast cancer and high APOBEC3B expression, and this may be due to the demonstration that APOBEC3B can control the growth of breast cancer cells by the promotion of estrogen receptor transcriptional activity through protein interaction. Further, the study of Periyasamy *et al* (2015) showed that APOBEC3B can cause C-to-U mutations in the promoter regions of estrogen receptor target genes, where it causes local DNA repair. Repairing APOBEC3B-induced lesions potentially allows remodelling of chromatin and stimulates gene expression, making APOBEC mediated cytidine deamination a potentially important part of gene activation.

## 1.5 – APOBEC3H

The APOBEC3H gene is conserved throughout mammalian evolution, implying that APOBEC3H is likely to mediate an important function that predates primate evolution (Cascalho, 2004). Human APOBEC3H (A3H) belongs to the sub-group of the APOBEC3 cytidine deaminase family which actively restricts replication of HIV-1 (Shi *et al.*, 2017). APOBEC3H is also an antiviral effector for other retroviruses (OhAinle *et al.*, 2006) and is the most divergent member of the APOBEC3 family, with a Z3 type Zn-coordinating domain that is phylogenetically distinct from the Z1- and Z2 domains seen other APOBEC3 proteins (Shi *et al.*, 2017). APOBEC3H is distinguished amongst the APOBEC3 family, as the gene which has the highest number of SNP variants, resulting in, at least, five different single amino acid variants (N15Δ, R18L, G105R, K121D, and E178D), which collectively define seven haplotypes, of which three encode stable forms of APOBEC3H protein capable of suppressing HIV (Wang *et al.*, 2011). APOBEC3H haplotypes II, V and VII, share both N15Δ, and 105R and produce stable proteins, exerting strong *in vitro* activity against the replication of HIV-1 (Iseda *et al.*, 2016) Unstable APOBEC3H haplotypes, inactive against HIV, contain two single nucleotide polymorphisms that are found with other polymorphisms in different combinations (Akre *et al.*, 2016). As a result, APOBEC3H antiretroviral activity is influenced by polymorphisms and alternative splicing variants and has been associated with population enhanced resistance to HIV-1 virus, so that the G105R, K121E and E178D haplotypes are frequently observed in African populations but are considerably less common in Asian and European ones. It has also been suggested that APOEC3H has been key in preventing cross-species transmission of simian retroviruses to humans (Zhang *et al.*, 2017).



The APOBEC3H gene is organised into six exons, the first of which is non-coding (Figure 1.5B). The resultant alternative splicing within the gene allows the generation of four APOBEC3H coding variants, encoding 154, 182, 183, and 200 amino acid forms of APOBEC3H protein (Ebrahimi *et al.*, 2018). Further, these authors show through bioinformatics and functional analyses that the splice variant giving rise to the 200 amino acid form of the enzyme is encoded by APOBEC3H haplotype II, with this being the most potent human, antiviral form of the enzyme. The importance of this is made apparent by their observation that HIV-1 protease cleaves the 200 amino acid form into shorter, less active isoforms.

### **1.5.1 - APOBEC3H dimerisation by RNA**

Recently, several reports have identified another unique feature of APOBEC3H, namely the ability of the enzyme to dimerise through the binding of dsRNA molecules (Bohn *et al.*, 2017; Feng *et al.*, 2018; Ito *et al.*, 2018; Matsuoka *et al.*, 2018; Shaban *et al.*, 2018). While the mechanism of dimerisation is clear, there has been ambiguity as to which parts of APOBEC3H molecule are involved and whether the bound RNA acts as an allosteric, or direct inhibitor of the deaminase activity in the dimerised protein (Ito *et al.*, 2018; Shaban *et al.*, 2018) or, indeed, does inhibit deamination of DNA at all (Bohn *et al.*, 2017). There is also ambiguity as to whether monomers of APOBEC3H are active (Ito *et al.*, 2018). The most recent paper in this regard, from Feng *et al* (2018) examined the activity of recombinant APOBEC3H protein and concluded that the enzyme requires this dsRNA-mediated dimerisation for enzymatic activity, but not binding ssDNA.

## 1.6 – p53

The TP53 gene is located in humans on the short arm of chromosome 17 (17p13.1) (Matlashewski *et al.*, 1984). It spans 20 kb, has a non-coding exon 1 and a first intron that, at 10 kb, is very long. Five regions in the coding sequence show high conservation in vertebrates, mostly in exons 2, 5, 6, 7 and 8 (May and May, 1999). About 50% of the tumour suppressor gene p53 is mutated or lost in cancers (Olivier *et al.*, 2010) and this protein's role in monitoring and controlling cell metabolic disorders and genetics damage through cell cycle inhibition and programmed cell death is critical (Kubbutat *et al.*, 1997). MDM2 has a negative effect that degrades p53, which remains low in unstressed cells; the interaction between p53 and MDM2 de-links with an accumulation of p53 and leads to an increased half-life for p53 (Lavin and Gueven, 2006; Vousden, 2009).

A number of mechanisms are in play as part of p53's role in regulation or progression involving cell cycle, apoptosis, and genomic stability. It can activate repair proteins for damaged DNA and may, therefore, be an important factor in ageing (Olivier *et al.*, 2010). p53 can impede growth by keeping the cell cycle at the G1/S regulation point on DNA damage recognition. Holding the cell here long enough will give DNA repair proteins enough time to fix the damage, so that the cell cycle can continue. It can also initiate apoptosis, or programmed cell death if the damage is not repaired and is essential for the senescence response to short telomeres (Figure 1.6) (Gilbert, 2013).

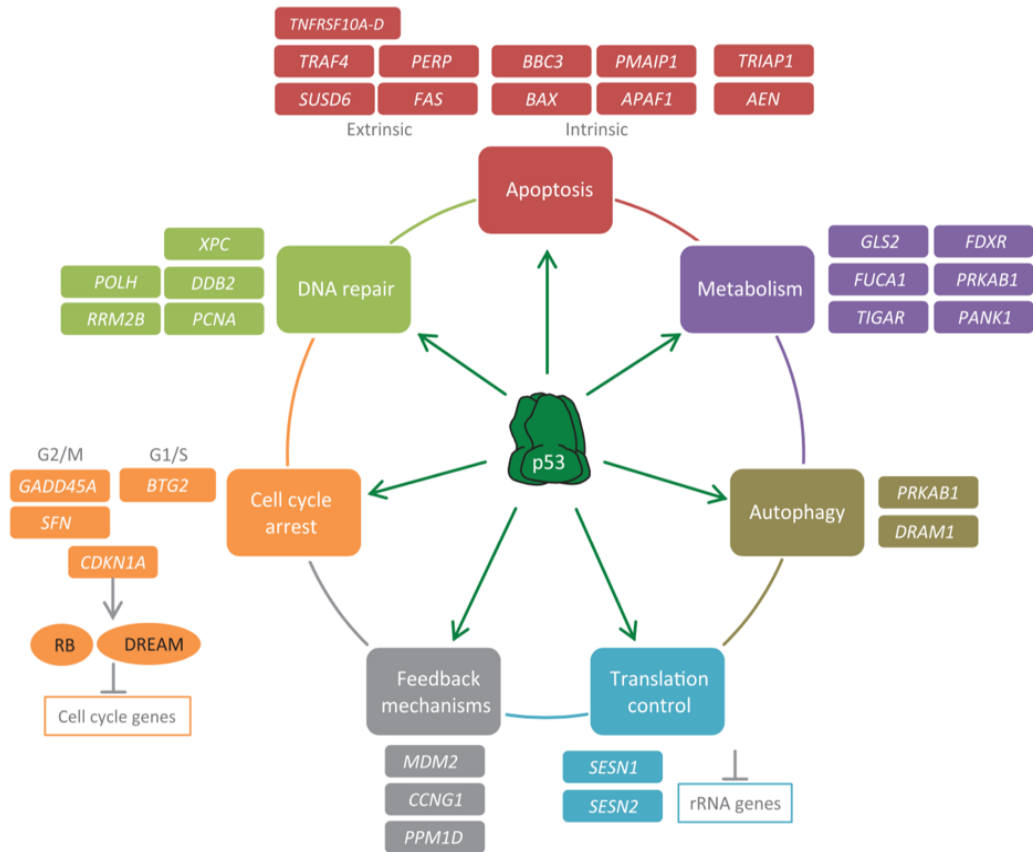
p53 typically binds in a sequence-specific way to DNA sites comprising two decameric motifs or half-sites taking the general form RRRCWWGYYY (R = A/G, W = A/T, Y = C/T) (el-Deiry *et al.*, 1992). There are three major functional domains to the p53 molecule: the N terminus containing the transactivation domain; the core domain containing the sequence-specific DNA

binding domain; and the C terminus containing oligomerisation and regulatory domains (Kitayner *et al.*, 2006). p53's main target for mutation is in the core domain, in which are to be found 80% to 90% of all missense mutations found in human tumours (Figure 3.1) (Rivlin *et al.*, 2011)

p21, the p53 direct transcriptional target is encoded by the CDKN1A gene and mediates cell cycle inhibition (el-Deiry *et al.*, 1993). This binds to, and inhibits cyclin-dependent kinases (CDKs) to stop cell cycle progression (Harper *et al.*, 1993). In the event of DNA damage, p21 is induced by p53 to arrest the G1 phase, so that cells can repair the damage (el-Deiry *et al.*, 1994). Depending on how serious the damage is, p53's ability to regulate the activity of the pro- and anti-apoptotic Bcl-2 family proteins, both directly and indirectly, means that it is also able to induce cell apoptosis (Hemann and Lowe, 2006). p53 directly regulates expression of two apoptotic proteins: Bax, and the p53-upregulated modulator of apoptosis (Puma). In both cases, localisation is to the mitochondria (Benchimol, 2001; Miyashita *et al.*, 1995). p53 has been shown to have transactivation-independent roles in inducing apoptosis. These include direct interaction with anti-apoptotic protein Bcl-2, inactivating Bcl-2 ( Figure 1.6) (Hemann and Lowe, 2006).

p53 inactivation in cancer is mainly through allelic loss and single-base changes, which are most often missense mutations. A central step in the adenoma-carcinoma transition is thought to be somatic mutation of a single copy of *TP53*, followed by loss of heterozygosity (LOH) affecting the short arm of chromosome 17, which contains *TP53* gene (Baker *et al.*, 1990; Olivier *et al.*, 2010). Missense mutations are mostly clustered around the p53 DNA binding domain, and frequent somatic mutation targets are codons 175, 245, 248, 273 and 282 in which mutations prevent p53's ability to transactivate. As the transcriptional activity of p53 requires

that the protein be tetramerised, mutant p53 has been suggested as exerting a dominant-negative effect on wild-type p53 in cells retaining one wild-type allele (Olivier *et al.*, 2010).

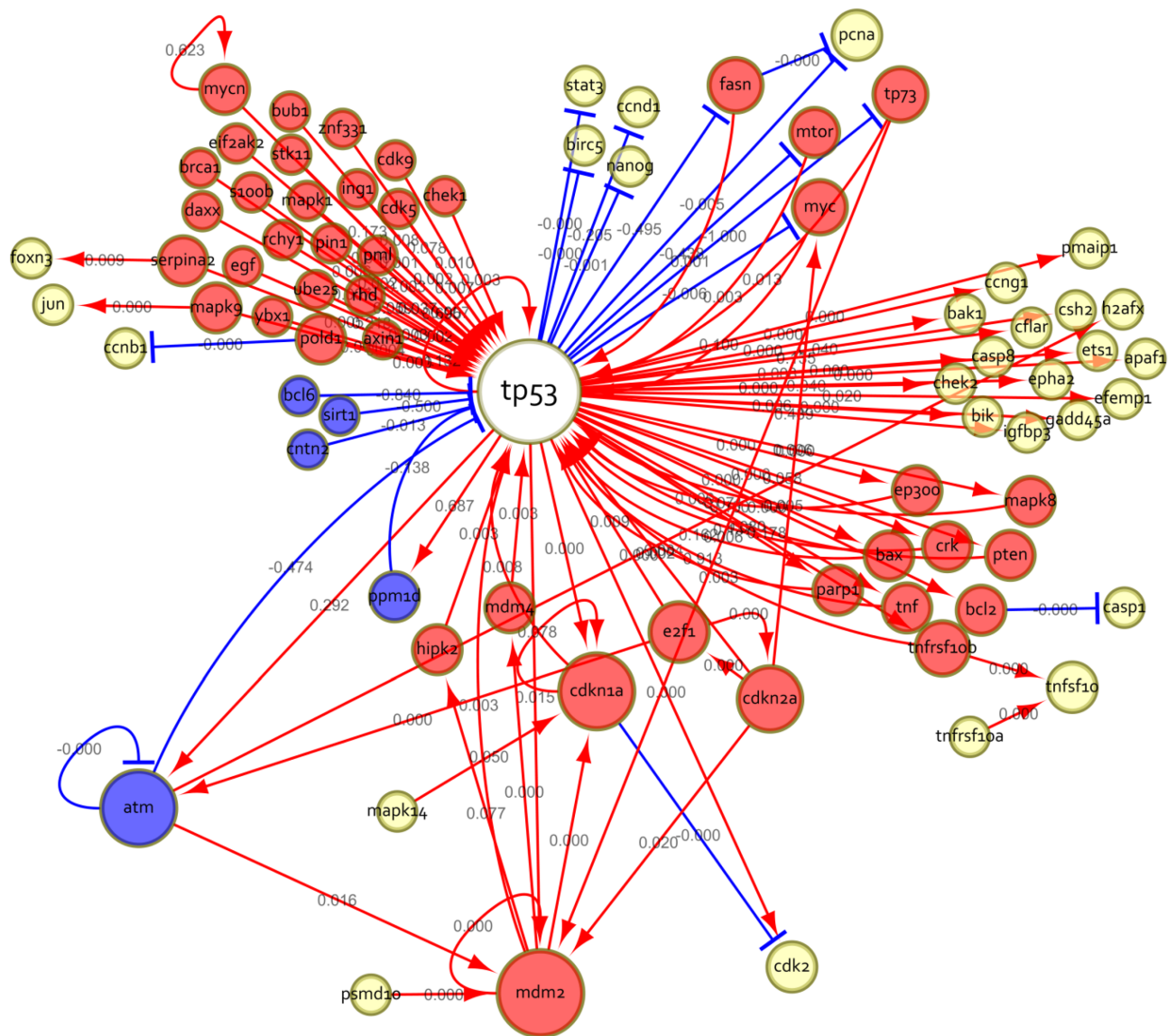


**Figure 1.6. Key p53 target genes and pathways**

The diagram shows important p53 target genes, associated with cell cycle arrest, DNA repair, apoptosis and metabolism, autophagy, translation control and feedback mechanisms. (Fischer, 2017).

In 319 studies published between 1992 and 2016 (Figure 1.7) p53 was reported to activate 246, with 91 repressed and 9 both activated and repressed. In at least six data sets, the top 116 genes were all p53-activated targets. As well as CDKN1A and RRM2B, well-known p53 target genes identified in most data sets included MDM2, GDF15, SUSP6, TMPS, DRAGO, KIAA0247, GADD45A, PLK3, BTG2, TIGAR (C12orf5), TNFRSF10B, PPM1D, BAX, AEN, PLK2, SESN1, FAS and KITLG (Fischer, 2017). p53 binds in different parts of the gene and at varying distances from the TSS, with multiple binding sites for p53 in some genes (Laptenko *et al.*, 2015) and binding within 1kb of the TSS in most p53 target genes (Fischer, 2017).

Activated or deleted, p53 integrates chromosomal stresses and immune responses by influencing APOBEC3 gene expression, and these key components of the innate immune system also influence the stability of the genome (Menendez *et al.*, 2017). p53 represses expression of APOBEC3B and cytosine deaminase activity in cancer cells, through a - mechanism dependent on p21, which reduces p53 activity through mutation and upregulates A3B (Periyasamy *et al.*, 2017). APOBEC3H has also been identified as a p53 target gene (Tebaldi *et al.*, 2015).



**Figure 1.7. p53 and gene regulatory networks**

Red nodes signify p53 induction or activation, blue indicates suppression or deactivation, and yellow indicates genes without direct action on p53. Red arrows show induction/activation, while blue arrows indicate inhibition. Node size corresponds to the number of connections to other nodes. To check their accuracy, each interaction was tested against source literature.

The assessment included whether the sentence represented an association between the stated genes and whether the resulting sign (-/inhibits or +/stimulates) corresponded to the sentence statement. from (Chen *et al.*, 2014).

## 1.7 - APOBECs and cancer

Some 7000 human tumour samples have been found to feature mutations derived from APOBEC activity (Rebhandl *et al.*, 2015), as characterized by dC to dT transitions within a T-C-W (for APOBEC3) motif. These are seen to be significantly present in bladder, breast and thyroid cancers and lung adenocarcinoma (Alexandrov *et al.*, 2013; Nik-Zainal *et al.*, 2012; Roberts *et al.*, 2013; Swanton *et al.*, 2015). Localised hypermutations, known as ‘kataegis’ from the Greek for ‘shower’, have also been found in 21 breast cancer genomes, suggesting the specific involvement of APOBEC3 proteins in the generation of these mutation clusters (Nik-Zainal *et al.*, 2012). Further, elevated levels of genomic uracil corresponded with overexpression of APOBEC in breast cancer cells (Burns *et al.*, 2015; Swanton *et al.*, 2015). Collectively, these studies implicate APOBEC3 proteins in tumour DNA editing and mutation processes.

To induce kataegis requires a functional DNA repair machinery and the prevailing explanation is that during ssDNA resection, DNA repair generates ssDNA substrates for APOBEC proteins so that DNA lesions initiated through APOBEC enzymes or by other damaging agents such as UV-light, radiation and ROS induce DNA repair either by Base Excision Repair (BER) or by Mismatch Repair (MMR), both of which require a 5’-3’ resection of the damaged DNA strand and the generation of long stretches of ssDNA that can be attacked by APOBEC3 proteins. This mutagenesis, induced by DNA repair, leads to strand-coordinated mutations, with only dCs on the non-resected DNA strand deaminated, and to clusters of mutations focused on regions of a few kilobases, typically resected during DNA repair (Chen *et al.*, 2014).

Certain members from the APOBEC3 family have been found to be overexpressed by cancer cell lines and cancer tissues (Conticello, 2008). Overexpression of APOBEC3B in breast cancer cell lines and breast cancers is particularly worthy of note (Cescon *et al.*, 2015), as it



has recently been shown this is needed for ER-positive breast cancer cell growth, where APOBEC3B acts as an ER co-regulator (Periyasamy *et al.*, 2015). Further, by the use of Chromatin Immuno Precipitation (ChIP), this study showed that APOBEC3B was localised in the genome of ER-positive breast cancer cells at ER binding regions. This causes local DNA strand breaks, which result from C-to-U transitions catalysed by the enzyme's cytidine deaminase activity, which are, subsequently, subject to DNA repair. This provides a link between APOBEC3B action and chromatin remodelling, in a process aiding expression of ER target genes.

50% of breast tumours, and most breast cancer cell lines feature overexpression of APOBEC3B (Zhao *et al.*, 2018). APOBEC3B is constitutively nuclear and is the only DNA deaminase activity detected in breast cancer cell extracts. Overexpression of APOBEC3B also induces cytotoxicity and DNA damage, leading to multinucleated cells before cell death (Kuong and Loeb, 2013). Expression of APOBEC3B show a positive correlation with overall cytosine mutation loads in breast cancer (Burns *et al.*, 2013), but the corresponding mutational signature is still visible in APOBEC3B-null breast tumours and suggests that other APOBECs, possibly APOBEC3H (APOBEC3H-Haplotype I- see below) are responsible for APOBEC signature mutations when APOBEC3B is absent. There is also an association between APOBEC3H Haplotype I and clonal TCA/T-biased mutations in lung adenocarcinoma, which suggests that this enzyme's contribution to cancer mutagenesis may be both significant, and broad (Burns *et al.*, 2015).

APOBEC3H- Haplotype I has a strong DNA cytosine deaminase activity, with a clear TC dinucleotide substrates preference, which has been shown by enzyme activity assays and HIV-1 restriction and mutation experiments (Starrett *et al.*, 2016). However, after compensating for lower levels of protein expression, APOBEC3H-Haplotype I, and also APOBEC3H-

Haplotype II show similar enzymatic activities and local motif preferences. Subcellular localisation showed the APOBEC3H- Haplotype I protein to be significantly more nuclear than APOBEC3H- Haplotype II protein, suggesting possible disruption by Gly105 of the interaction with a cytoplasmic retaining factor. Taken together, these findings would suggest a plausible mechanism for APOBEC3H- Haplotype associations for somatic mutation in cancer (Burns *et al.*, 2015)

## **1.8 - Uses of CRISPR/CAS9 as a gene editing tool in human cancer cell lines**

### **1.8.1 - Reverse genetic approaches in human cell lines**

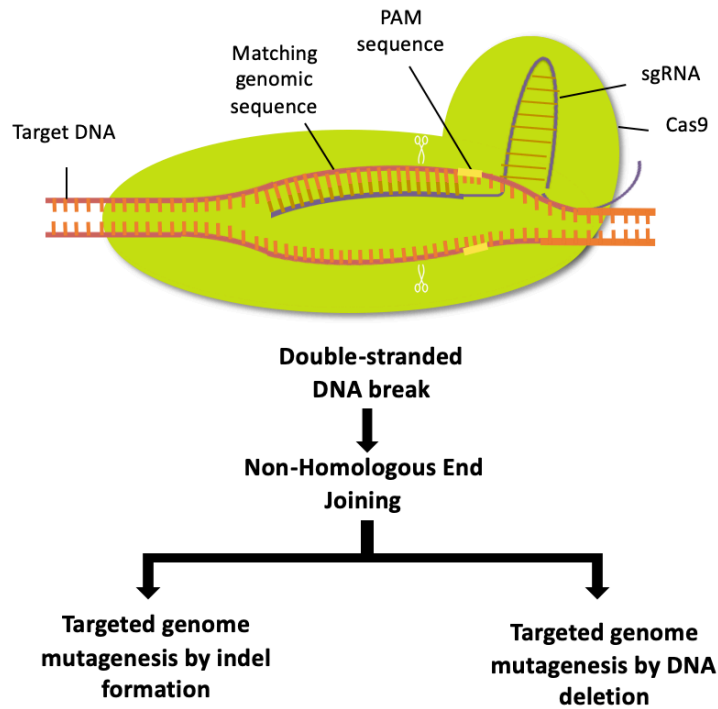
The availability of annotated genome sequences for humans, and a wide variety of other model organisms (Diehl and Boyle, 2016) has led to enormous advances in the understanding of gene function, and the definition of new mutations for many human diseases. This collective data essentially points to the use of “reverse genetics”, an approach where gene sequence is used to study gene function, primarily through the production of engineered genomic mutations. Reverse genetics has been successfully used in a variety of model systems (Hardy *et al.*, 2010). However, in humans, it is not possible to model such mutations by transgenic approaches and look at whole organism phenotypes. Hence, human systems have largely focussed on the use of cell lines, and methodologies to manipulate target genes in this context. To date, methods have included the use of RNA interference (Kim and Rossi, 2007), and genome engineering approaches with engineered zinc finger nucleases (Carroll, 2011), and engineered TAL nucleases (TALENs) (Joung and Sander, 2013). While these techniques are all of utility, they also have significant drawbacks in terms of sustained gene expression affects, off-target issues, complexity of use and cost (Boettcher and McManus, 2015). The newly developed clustered

regularly interspaced short palindromic repeats (CRISPR)-CRISPR associated proteins (CRISPR-Cas) system offers an alternative system for targeted genome editing, which is simple, low cost and allows the generation of null mutations, essentially by indel formation, as well as the ability to make knock in mutations, when used in conjunction with an appropriate template for homologous recombination (Boettcher and McManus, 2015).

CRISPRs are clustered, regularly interspaced short palindromic repeats that are able to integrate exogenous DNA sequences into the host bacterial genome. They have been found in a large number of bacteria and archaea, where they act as a defence mechanism, shielding cells from bacteriophages and conjugative plasmids (Marraffini and Sontheimer, 2010). In this manner CRISPRs act to give an adaptive, heritable record of past exposure and are used for the expression of CRISPR guide RNAs (gRNAs) that are able to recognise intrusive nucleic acids, and act in combination with a host nuclease, Cas, to bring about the targeted destruction of invading pathogen genomes (Marraffini and Sontheimer, 2010). CRISPR systems are classified into Type I, II, or III. The Type II system uses Cas9, a single nuclease enzyme, which binds CRISPR gRNAs, so as to allow targeting that leads to the specific cleavage of the targeted dsDNA sequences. Functionally, CRISPR target sequences are 20nt long, and need to be followed at the 3' end by the Cas9 licensing Protospacer Adjacent Motif (PAM) NGG. As the probability of a random match to a 20-nucleotide sequence (1 in  $4^{20}$ , or 1 in  $1 \times 10^{12}$  bases) is greater than the size of human genome ( $6 \times 10^9$  bases). CRISPR target sequences can be identified in complex genomes, so that they are essentially unique, providing little scope for off-target effects. (Jinek, 2012)

Following the discovery and understanding of the Type II CRISPR- Cas9 system, this has recently been adapted for genome engineering in eukaryotic systems, including human cells, by developing transfectable expression constructs for humanised, nuclear targeted Cas9 and

hybrid guide RNAs, or “sgRNAs”. These encode twenty nucleotide genome targeting sequences as part of a larger structural RNA capable of being bound by the Cas9 enzyme (Mali *et al.*, 2013). As fairly small structured RNAs, sgRNAs are best transcribed by RNA Polymerase III and expressed using the U6 RNA gene promoter (Sanchez-Rivera and Jacks, 2015; Yang *et al.*, 2014). This basic system allows CRISPR targeted cleavage of DNA. The cleaved DNA will be subject to DNA repair, which will regenerate the DNA site for cleavage, unless a fault has been made in repair, leading to insertion, or deletion, also known as “indel” mutation. In this way CRISPR-Cas9 can lead to targeted indel mutations, which then allows reverse genetic approaches to gene studies. In addition, the CRISPR-Cas system has been further modified to allow additional modes of genome engineering, and targeted gene regulation (Adli, 2018).



**Figure 1.8. Overview of CRISPR-CAS9 system**

CRISPR mediated genome editing requires the use of an sgRNA and Cas-9. When introduced into a cell these can cause a double stranded DNA break that can lead to indel mutation formation by Non-Homologous End Joining (NHEJ). NHEJ can lead to the efficient introduction of insertion/deletion mutations (indels) of various lengths, which can disrupt the translational reading frame of a coding sequence. Adapted from Addgene.org

## 1.9 – Aims

In evaluating APOBEC3 expression, our group have found that p53 activation, through the use of Nutlin, a compound which interferes with the interaction between p53 and Mdm2, resulted in an increase of APOBEC3H expression and reduced APOBEC3B expression in cells with wild-type p53. This finding suggested that APOBEC3B and APOBEC3H are potential p53 target genes and indicated the possibility that cytidine deaminase activity may be a part of the p53 gene response.

My proposed investigation is:

- To use CRISPR directed genome editing to investigate the role of APOBEC3B and APOBEC3H in the regulation of p53 and its function.
- To use CRISPR directed genome editing to develop a cancer cell line model in which expression of APOBEC3H is knocked out, so as to further define the role of APOBEC3H in cancer cells. This will be done using transcriptomic analysis of the knockout lines with Next Generation RNA sequencing (RNAseq).

## **2 – Materials and methods**

### **2.1 - Materials**

#### **2.1.1 - General reagents, materials and equipment**

Stock chemicals, and molecular biology grade, sterile water was obtained from Sigma-Aldrich (Gillingham, UK), unless otherwise stated. Standard molecular biology reagents, including DNA and protein size markers were obtained from ThermoFisher (Life Technologies, Paisley, UK) and New England Biolabs (NEB; Hitchin, UK). Single stranded oligonucleotide primers were synthesised by Invitrogen and were dissolved at a final concentration of 100µM in molecular biology grade sterile water. TaqMan® and SYBR® Green real-time PCR reagents were obtained from Applied Biosystems (Life Technologies, Paisley, UK). Plasmid vectors were obtained as bacterial “stab” cultures from Addgene ([/www.addgene.org](http://www.addgene.org)), under the cover of Material Transfer Agreements made on behalf of the recipient laboratory. Protease and phosphatase inhibitor cocktail tablets were obtained from Roche Diagnostics Ltd (Sussex, UK) and SuperSignal West Pico and Femto substrates were obtained from ThermoFisher (Life Technologies, Paisley, UK).

#### **2.1.2 - Sundries**

1.5ml and 2ml microcentrifuge tubes and Gilson pipette compatible tips were obtained from Starlab Ltd (Milton Keynes, UK). Thin walled -PCR tubes (0.2ml and 0.5ml) were obtained from Applied Biosystems (Life Technologies, Paisley, UK). Disposable 30ml high speed centrifuge tubes were obtained from Sarstedt Ltd (Leicester, UK) and 250ml Sorvall high speed centrifuge bottles from ThermoFisher (Life Technologies, Paisley, UK). Tissue culture and

microbiology grade plastics were obtained from Corning (Dow Corning, Dewsbury, UK) and Sterilin (Newport, UK). Nitrocellulose membrane (0.45µm), for Western blotting was obtained from Bio-Rad Laboratories Ltd (Hemel Hempstead, UK).

### **2.1.3 - General equipment**

Large volume (>30ml) high-speed centrifugation was performed using a Sorvall RC6 Plus centrifuge with Sorval SS34 and Fiberlite™ F14-6 x 250y Fixed-Angle Rotors (ThermoFisher, Life Technologies, Paisley, UK). Bench top harvesting of tissue culture cells was carried out using a Heraeus Megafuge 1 centrifuge (Cambridge, UK). Bench top Sorvall Pico™ microcentrifuges (Sorvall, Leicester, UK) were used for volumes less than 2ml. Incubations were carried out in a water bath (Grant SUB Aqua Pro, Grant Instruments Ltd, Shepreth, UK), incubator oven (LEEC, Nottingham, UK), shaking incubator (New Brunswick Scientific Company Incorporated, Edison, USA) or block heater (Grant-bio, Wolf Laboratories, York, UK). PCR was carried out using an Applied Biosystems Veriti™ 96 well thermal cycler (Life Technologies) and real-time quantitative PCR was carried out using an Applied Biosystems 7900HT Fast Real-Time PCR System (Life Technologies, Paisley, UK). Horizontal agarose gel electrophoresis tanks were purchased from Thistle Scientific (Glasgow, UK). Western blotting vertical gel electrophoresis tanks and transfer apparatus were from Hoeffer (GE Healthcare Life Science, Buckinghamshire, UK) and Bio-Rad. Power packs were also obtained from Bio-Rad. Routine DNA and RNA quantification was carried out using spectrophotometric measurements from a Nanodrop™ ND-1000 (Labtech International, UK). Samples were mixed using a Vortex-Genie 2 (Scientific Industries, London, UK).



#### **2.1.4 - Cell culture reagents**

Cells were maintained at 37°C in a humidified air atmosphere supplied with 5% CO<sub>2</sub> in RS Biotech Galaxy R+ CO<sub>2</sub> incubators (New Brunswick, Eppendorf, Cambridge, UK). Tissue culture work was carried out in Hepaire ducted BIOMAT 2 Microbiological Safety Cabinets (Contained Air Solutions Ltd, Manchester, UK). Cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM), supplemented with L-Glutamine-Penicillin-Streptomycin solution (200mM L-glutamine, 10,000 units/ml penicillin, 10mg/ml streptomycin in 0.9% sodium chloride) ("PSG"; Sigma-Aldrich, Gillingham, UK) and 10% heat treated Foetal Calf Serum (FCS; First Link Ltd, Birmingham, UK). Cell harvesting and passaging was carried out using 0.02% ethylenediaminetetraacetic acid (EDTA) solution and 10X trypsin solution supplied by Sigma Aldrich (Gillingham, UK). Cell counting was carried out using a Countess II automated cell counter and disposable microfluidic slides (ThermoFisher, Paisley, UK).

#### **2.1.5 - Chemicals and antibiotics**

Nutlin (Nutlin-3), G418 (Geneticin), Kanamycin and Ampicillin were obtained from Sigma-Aldrich, (Gillingham, UK).

#### **2.1.6 - General stock solutions**

Solutions were made in reverse osmosis purified water (Merk-Millipore, Watford, UK), autoclaved and stored at room temperature unless otherwise stated.

**Table 2.1: General stock solution**

Phosphate buffered saline (PBS) (10x)	137mM NaCl, 3mM KCl, 8mM Na <sub>2</sub> HPO <sub>4</sub> , 1.5mM KH <sub>2</sub> PO <sub>4</sub> , pH 7.4
Tris-EDTA (TE) buffer	10mM Tris-HCl pH 8.0, 1mM Na-EDTA
10% APS (Ammonium Persulfate)	1g in 10ml of molecular biology grade water, aliquoted and stored at -20°C

### **2.1.7 - Microbiological reagents**

Luria-Bertani (LB)-broth and LB-broth agar capsules were obtained from MP Biomedical, LLC (Illkirch, France). These were prepared according to the manufacturer's instructions using reverse osmosis purified water (Merk-Millipore, Watford, UK) and autoclaved, before being stored at room temperature. Selection antibiotics were added to media, prior to inoculation with bacteria. For bacterial plates, solidified agar was melted using a microwave oven, and then cooled to 50°C before the addition of antibiotics. Ampicillin sodium salt (Sigma Aldrich, Gillingham, UK), was dissolved in molecular biology grade, purified water at 100mg/ml and sterilised by filtration through a 0.45µm filter, and frozen at -20°C in 1ml aliquots. This was used at a final concentration of 100µg/ml. Similarly, Kanamycin sulphate (Sigma Aldrich, Gillingham, UK), was dissolved in molecular biology grade, purified water to 50mg/ml, sterilised by filtration through a 0.45µm filter, and frozen at -20°C in 1ml aliquots. Kanamycin was used at a final concentration of 50µg/ml.

## **2.2 – Methods**

### **2.2.1 - Tissue culture**

MCF7 Luc (Cambridge Bioscience, Cambridge, UK) are a variant of the MCF7 estrogen responsive breast cancer cell line (Soule *et al.*, 1973), which have been modified to incorporate a constitutively expressed fire-fly luciferase gene for tracking purposes. The line has been previously used in studies of estrogen responsive breast cancer cell growth (Harrod *et al* 2016). HCT116 (Brattain *et al.*, 1981) is a human colon cancer derived cell line (ATCC, LGC standards, Teddington, UK). Cells were routinely cultured in DMEM containing 10% FCS

(First Link Ltd, Birmingham, UK). and 1% PSG (Sigma Aldrich, Gillingham, UK). Cells were grown at 37°C in a 5% CO<sub>2</sub> humidified environment, and maintained in 75 cm<sup>2</sup>, or 150 cm<sup>2</sup> cell culture flasks (Dow Corning, Dewsbury, UK), where cells were allowed to reach 80-90% confluence before passaging. Cells were passaged using Trypsin digestion and harvesting. Cell monolayers were washed twice with 0.02% EDTA solution (Sigma Aldrich, Gillingham, UK) at 37°C, and then incubated with 5µl/cm<sup>2</sup> of 1x Trypsin diluted in 0.02% EDTA solution (Sigma Aldrich, Gillingham, UK). Trypsin digestion was maintained for 5 min at 37°C to detach cells. After trypsinisation, 10ml of pre-warmed culture medium was added to neutralize the trypsin activity, and the cell suspension subject to gentle titration to produce single cell suspensions. Cells were serially passaged into tissue flasks using seeding ratios of 1:4, 1:8, or 1:16. Cell lines were tested on a 3-monthly basis for Mycoplasma using MycoAlert™ Mycoplasma Detection Kit (Lonza, Cambridge, UK).

### **2.2.2 - Preparation and use of frozen cell stocks**

Cell lines were frozen for long-term storage in liquid nitrogen. Confluent, or near confluent, T150 flasks were harvested and the cells recovered by centrifugation (Heraeus Megafuge 1 centrifuge, Heraeus, Cambridge, UK) at 1200rpm (~300xg) for 5 min. Cell pellets were resuspended in 6 ml of freezing medium (90% FCS, 10% DMSO), previously filter-sterilised using a 0.22µm syringe filter. The resulting cell suspensions were aliquoted as 1ml volumes into 2ml cryogenic vials. Cell suspensions were then chilled to -80°C using isopropanol baths (“Mr Frosty” racks, Nalgene, ThermoFisher Scientific), and incubated for at least 24 h, before being transferred to liquid nitrogen for long-term storage. For recovery of frozen cells, vials

were warmed at 37°C for 5 min, and then dispensed, drop-wise into T75 flasks containing 30ml of warmed medium.

### **2.2.3 - Cell counting and cell viability measurement**

For experimental seeding, the number of cells in suspensions was determined by cell counting using the Countess™ II FL Automated Cell Counter (ThermoFisher, Life Technologies, Paisley, UK)) and disposable, micro-fluidic counting slides. For counting and assessing cell viability, cell suspensions were diluted with an equal volume of 0.4% trypan blue solution (ThermoFisher (Life Technologies, Paisley, UK)), and 50µl of the trypan blue stained cell suspension loaded onto the counting slide. The Countess™ II FL Automated Cell Counter analyses images of the suspended cell suspension to give cell counts and percentage cell viability. The live cell count was subsequently used for seeding calculations.

### **2.2.4 - Measurement of cell growth using the Sulforhodamine B (SRB) assay**

For experiments measuring cell growth using the SRB assay (Vichai and Kirtikara, 2006), cultures of replicate 96 well plates were made. These contained 3,000 cells per well, which were seeded using an eight channel, multichannel pipette, set to dispense 100µl volumes per tip, using cell suspensions made in full medium, and adjusted to 30,000 cells/ml. The medium was then changed every three days, with fresh medium supplemented with the appropriate treatment agents and plates fixed on days 0, 3, 6, 9 and 12, using 100µl 40% w/v trichloroacetic acid (TCA) in reverse osmosis purified water, at 4°C, for a minimum of 1 h. The fixed cells

were subsequently washed five times with reverse osmosis purified water and then stained with 0.4% w/v Sulforhodamine B (SRB, Sigma-Aldrich, Gillingham, UK) in 1% v/v acetic acid in reverse osmosis purified water, for 1 h at room temperature. Excess dye was removed following five washes in 1% acetic acid, and the stained cultures left to dry, inverted over tissue, at room temperature. The dye in each well was solubilised by the addition of 100µl of 10mM Tris-base to each well, and 10 min of incubation with shaking. SRB was measured by absorbance was determined using a Tecan Sunrise™ microplate reader (Tecan, Reading, UK).

### **2.2.5 - Cell cloning**

Cells were seeded at clonal densities (typically 3,000-5,000 cells in a 15cm dish) and allowed to grow as distinct clones, visible to the naked eye (typically 128 -256cells, or 7-8 doublings). Clones were picked using trypsin-soaked Scienceware® cloning discs (Sigma-Aldrich, Gillingham, UK). To do this, cells were washed twice in PBS, and after the second wash discs placed on-top of clones to be picked. Cells were returned to the tissue culture incubator for 5 min and the discs then individually transferred using fine needle forceps to individual wells of 24 well dishes, each containing 1.5ml of full medium. Cells were allowed to migrate off the paper discs used for cloning for 48 h and subsequently allowed to grow to confluence, with regular media changes. Once confluent, wells were harvested and 90% of the cells in the well-used to obtain cell pellets for analysis by PCR. The remaining cells were allowed to re-attach and grow again to confluence, at which point clones which were found to positive for DNA editing were moved to T25 flasks, and subsequently expanded further.

### 2.2.6 - Nucleofection of cells

DNA mixes containing 2µg of plasmid DNA in a volume of 5µl were used to nucleofect cells using an AMAXA® 4D Nucleofector device (Lonza, Slough, UK). Nucleofection protocols were adapted from those described by the manufacturer for MCF7 and HCT116 cell lines, respectively. Briefly, cells were harvested by trypsinisation and used to make single-cell suspensions by titration in full medium. Aliquots of  $2 \times 10^6$  cells in 1.5ml microcentrifuge tubes were then pelleted by centrifugation (200 rpm, 5 min Sorvall Pico Microfuge; Sorvall, Leicester, UK), washed once in PBS, and then resuspended in 100µl of Nucleofector® Solution V (AMAXA® Cell Line Nucleofector Kit V, Lonza, Slough, UK), together with 5µl of the DNA mix. The cell suspension was then transferred to a nucleofection cuvette, taking care not to trap air bubbles between the electrodes, and the cuvette loaded into sample carousel of the AMAXA® 4D Nucleofector device. Nucleofection was carried out using the optimised programmes recommended by the manufacturer, namely programme EN-130 for MCF7 and programme EN-113 for HCT116. Following nucleofection, cells were left to recover in the nucleofection reaction mix for exactly 5 min, after which the cells were resuspended in 1ml of culture media, pre-warmed to 37°C. Nucleofected cells were seeded diluted into 10ml of full medium and 5ml seeded into a T25 flask, with the remainder seeded at clonal densities into a 15cm<sup>2</sup> dishes. Cells seeded in the T25 flask were allowed to grow to confluence, so as to enable DNA preparation, and subsequent establishment of frozen stocks, while clones arising from seeding onto 15cm plates were picked using paper cloning discs and expanded for further characterisation.

## **2.2.7 - Gibson assembly cloning of CRISPR sgRNA sequences**

CRISPR guide RNAs were assembled through the synthesis of 100nt DNA fragments, encoding guide RNA sequences adjacent to a U6 promoter sequence. These were subsequently cloned, by Gibson Assembly, into a U6 expression plasmid vector, p41824 pgRNA-U6, which encodes the remainder of the synthetic guide RNA (sgRNA) backbone.

### **2.2.7.1 - Generation of the CRISPR guide RNAs fragment for cloning**

#### **2.2.7.1.1 - CRISPR template hybridisation**

CRISPR oligonucleotide pairs were synthesised as 60 base sequences, where the last 20 bases are complementary between the two sequences and correspond to the CRISPR guide RNA.

These were annealed in a hybridisation mix that consisted of:

8µl 5x Phusion® HF buffer (New England Biolabs, Hitchin, UK)

2µl oligonucleotide 1 (100µM),

2µl oligonucleotide2 (100µM)

38µl Molecular Biology grade water

To enable hybridisation, the tubes were floated in a beaker containing 2L of boiling water, which was allowed to cool gradually to room temperature over a period of 2 h. The annealed primer mix was then briefly centrifuged to recover the hybridisation volume.



#### **2.2.7.1.2 - CRISPR template end-repair**

The annealed primer products were converted to double stranded DNA, by using the high fidelity Phusion® DNA polymerase (New England Biolabs, Hitchin, UK) to fill in the single stranded, 40nt extensions flanking the annealed region. The reaction mixes for this consisted of:

20µl of the annealed oligonucleotide

4µl 5x Phusion® HF buffer (New England Biolabs, Hitchin, UK)

0.8µl 10mM dNTP mix (New England Biolabs, Hitchin, UK)

0.4µl Phusion® DNA polymerase (New England Biolabs, Hitchin, UK)

14.8µl Molecular Biology grade water

This reaction mix was then incubated on a programmable heating block (Applied Biosystems Veriti™ 96 well thermal cycler; Life Technologies, Paisley, UK) at 72°C for 10 min. The final reaction volume was recovered by brief centrifugation.

#### **2.2.7.1.3 - PCR amplification and purification of double stranded CRISPR gRNA DNA template**

Double stranded CRISPR gRNA DNA templates were amplified with the Polymerase Chain Reaction (PCR), again using Phusion® DNA polymerase (New England Biolabs, Hitchin, UK). For this, a PCR primer mix was used, consisting of:

5µl 100µM SA3984CRISPRf primer (TTTCTTGGCTTTATATATCT)

5µl 100µM SA3985CRSIPRr primer (GACTAGCCTTATTTTAACTT)

40µl Molecular Biology grade water

This mix was used to amplify the double stranded CRISPR gRNA DNA product as follows:

5µl 5x Phusion® HF buffer (New England Biolabs, Hitchin, UK)

1µl double stranded CRISPR gRNA DNA template,

5µl SA3984/SA3985 primer mix

1µl 10mM dNTP mix (New England Biolabs, Hitchin, UK),

0.5µl Phusion® DNA polymerase (New England Biolabs, Hitchin, UK)

37.5µl Molecular Biology grade water

The reaction was used to amplify DNA using the following cycling profile on an Applied Biosystems Veriti™ 96 well thermal cycler (Life Technologies, Paisley, UK):

98°C for 30 s,

Followed by 25 cycles of 98°C for 15 s, 50°C for 15 s and 72°C for 15 s,

Followed by an elongation step of 72°C for 2 min, and a hold step at 4°C.

One tenth of the reaction was then checked by agarose gel electrophoresis, and the remainder purified by spin column chromatography using the QIAquick PCR Purification Kit (Qiagen

Ltd., Manchester UK). The concentration of the eluted, purified DNA was determined using a Nanodrop™ spectrophotometer (ThermoFisher, Life Technologies, Paisley, UK) and diluted to a stock concentration of 1ng/μl.

#### **2.2.7.2 - p41824 pgRNA-U6 vector preparation**

The p41824 pgRNA-U6 vector (Addgene) encodes a single Afl II restriction enzyme site in the region between the U6 promoter and sgRNA backbone (<https://www.addgene.org/41824/>). To enable cloning into this site, the vector was linearised, by restriction enzyme digestion as follows:

1μl Afl II (20u/μl; (New England Biolabs, Hitchin, UK)

10μl 10x NEB buffer 4 (New England Biolabs, Hitchin, UK)

1μl 100x BSA (100μg/ml),

5μg of p41824 pgRNA-U6 vector DNA

Molecular Biology grade water to a total volume of 100μl.

Digestion was carried out in a 37°C oven for 4 h, followed by heat inactivation at 65°C for 20 min. Digest products from 1μl (50ng) of the reaction were resolved agarose gel electrophoresis, so as to confirm digest completion. The remaining, Afl II digested p41824 pgRNA-U6 vector DNA was purified by spin column chromatography using the QIAquick PCR Purification Kit (Qiagen Ltd., Manchester UK), and the concentration of the eluted, purified DNA determined using a Nanodrop™ spectrophotometer (ThermoFisher, Life Technologies, Paisley, UK).

### **2.2.7.3 - Cloning of double stranded CRISPR gRNA DNA sequences into the p41824 pgRNA-U6 vector using Gibson assembly**

Gibson Assembly (Gibson *et al.*, 2009) was originally described as a method for the simultaneous joining of multiple DNA fragments for the purposes of DNA cloning, and has since become a useful, efficient method for the cloning of synthetic DNA fragments, in an orientation specific manner. The Gibson reaction relies on the use of three enzymes. The first is a strand specific, 3' to 5' exonuclease that allows the formation of single stranded ends on double stranded DNA molecules, which then allows base pairing with shared, complementary sequences carried by a second DNA molecule to join with. The other two enzymes in the reaction are a DNA polymerase and DNA ligase, which act to repair DNA on the annealed DNA fragments. For the cloning of sgRNA sequences into p41824 pgRNA-U6 vector, the Gibson reaction is used the simplest configuration, where the vector and insert share regions of homology that allow pairing following exonuclease activity. The assembly reaction is further simplified, as all three enzymic components work efficiently in the same buffer, allowing the use of a Gibson Assembly master mix.

The Gibson Assembly reaction was set up as follows:

50ng of Afl II p41824 pgRNA-U6 linearized plasmid vector DNA

6.37ng of 100bp insert DNA

Molecular Biology grade water to a total volume of 10 $\mu$ l.

10 $\mu$ l of Gibson Assembly® Master Mix (New England Biolabs, Hitchin, UK)

The reaction was incubated on an Applied Biosystems Veriti™ 96 well thermal cycler (Life Technologies, Paisley, UK) at 50°C for 60 min, and then stored on ice, or frozen at -20°C, until used for bacterial transformation.

#### **2.2.7.4 - Transformation of chemically competent cells with Gibson assembly reaction products**

A 50µl aliquot of frozen, chemically competent NEB 5-alpha Competent E. coli cells (New England Biolabs, Hitchin, UK) were transferred from -80°C and thawed for 5 min on ice. 2µl of Gibson assembled DNA product was added and mixed gently by titration. The resulting bacterial-DNA suspension was incubated on ice for 30 min, and then heat shocked at 42°C for 30 s, before being placed on ice again for 2 min. To allow growth, and expression of kanamycin resistance, 950µl of SOC bacterial medium (New England Biolabs, Hitchin, UK) was added and the bacterial suspension transferred to a 7ml plastic bijou tube. Tubes were then incubated with shaking (250rpm) at 37°C for 60 min. 50µl and 100µl aliquots of transformed cells were finally spread onto warmed plates with Kanamycin at 50µg/ml, and the plates incubated, inverted, overnight at 37°C.

#### **2.2.8 - Plasmid preparation**

Plasmid DNAs were prepared using spin (mini preparation), and drip column chromatography (maxi preparation) (Qiagen Ltd., Manchester UK).

### **2.2.8.1 - Plasmid mini preparation**

Plasmid DNA was prepared from 1.5ml cultures containing the appropriate antibiotic, inoculated with single colonies from agar plates, and grown in 7ml plastic bijou tubes (Sterilin, Newport, UK) cultures grown overnight, with shaking at 200rpm, at 37°C. The resulting saturated cultures were transferred to 1.5ml microcentrifuge tubes and harvested by centrifugation (13,000 rpm, 5 min, Sorvall Pico™ microcentrifuge; Sorvall, Leicester, UK) and the pellets drained and used for plasmid preparation using QIAprep Spin miniprep kit (Qiagen Ltd., Manchester UK). The final DNA was primarily used as template in DNA sequence analysis and to facilitate expansion of positive clones, the original culture dregs were kept at 4°C.

### **2.2.8.2 - Plasmid maxi preparation**

For larger scale preparation of plasmid DNA, bacterial cultures were first grown as saturated 1.5ml starter cultures containing the appropriate antibiotic and inoculated with single colonies from agar plates, or from culture dregs following plasmid mini preparation. These were subsequently used to cultures grown overnight, with shaking at 200rpm, at 37°C. grown to saturation in 200ml of LB medium, supplemented with the appropriate antibiotic. Bacteria were harvested by high speed centrifugation in a Sorvall RC6 Plus centrifuge with Sorval Fiberlite™ F14-6 x 250y Fixed-Angle Rotor (ThermoFisher, Life Technologies, Paisley, UK) at 4,000g for 30 min. The resulting pellets were drained and used to make plasmid DNA using the Qiagen -tip 500 kit (Qiagen Ltd., Manchester UK). The protocol followed was essentially as described by the manufacturer, except the cleared bacterial lysates were passed through muslin onto the

purification columns, so as to remove particulate debris. The resulting eluted plasmid DNA was precipitated with isopropanol and harvested by high speed centrifugation in 30ml disposable centrifuge tubes (Sarstedt Ltd, Leicester, UK) using a Sorvall RC6 Plus centrifuge with Sorval SS34 Fixed-Angle Rotor (ThermoFisher, Life Technologies, Paisley, UK) at 10,000g for 30 min. Drained DNA pellets were dissolved in 400µl Tris-EDTA buffer (pH 8) and transferred to a 1.5ml microcentrifuge tube, where the sample was extracted with 400µl of Phenol Chloroform isoamyl alcohol 25:24:1 saturated with 10nM Tris, pH8, 1nM EDTA. The sample was vortexed, and the aqueous phase separated from the resulting emulsion by centrifugation (13,000 rpm, 5 min, Sorvall Pico™ microcentrifuge; Sorvall, Leicester, UK). The uppermost, aqueous phase was transferred to a fresh 1.5ml microcentrifuge tube, and the plasmid DNA precipitated by the addition of 40µL 3M Na Acetate (pH5.2) and 1ml 100% ethanol, inverted to mix and left in the - 80c freezer for 10-15 min. The precipitated DNA was pelleted by centrifugation, 13,000 rpm, 10 min, Sorvall Pico™ microcentrifuge; Sorvall, Leicester, UK), and the pellets washed twice in 70% ethanol, and the dried under vacuum. The dried DNA was resuspended in 300µl Tris-EDTA buffer (pH 8). The final DNA was primarily used as template in DNA sequence analysis, and inn cell transfection work.

### **2.2.9 - Polymerase chain reaction (PCR)**

PCR primers were designed using the on-line Primer3 design tool (Untergasser *et al.*, 2012), such that primer melting temperatures ( $T_m$ ) were specified as 60°C. This ensured all primers could be used under the same amplification conditions and allowed standard PCR assay conditions to be defined. PCR reactions were carried out using Reddymix 2x PCR enzyme master mix (ThermoFisher, Life Technologies, Paisley, UK). PCR reactions (25µl) consisted of 12.5µl Reddymix, typically 200 ng of genomic DNA template, 2.5µl of 10µM forward

primer, 10 $\mu$ M reverse primer and Molecular Biology grade water up to a final volume of 25 $\mu$ l. DNA amplification by PCR was carried out on a programmable heating block (Applied Biosystems Veriti™ 96 well thermal cycler; Life Technologies, Paisley, UK). Cycling conditions used were typically as follows: A pre-cycle of 95°C for 2 min, followed by 35 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s, followed by a final elongation cycle of 72°C for 5 min. samples were left on hold at 4°C.

Agarose gel electrophoresis was used to analyse products, and if used as template for DNA sequencing, PCR products were purified with either the QIAquick PCR Purification Kit (Qiagen Ltd., Manchester UK) or with Clean Sweep (ExoSAP PCR clean-up reagent; ThermoFisher, Life Technologies, Paisley, UK).

#### **2.2.10 - Genomic DNA preparation**

Tissue cultured cells were harvested by trypsinisation, and 2-4 million cells pelleted by centrifugation (13,000 rpm, 5 min, Sorvall Pico™ microcentrifuge; Sorvall, Leicester, UK). The resulting pellets were resuspended in 200 $\mu$ l PBS and genomic DNA extracted using the PureLink Genomic DNA Kit (Invitrogen; ThermoFisher, Life Technologies, Paisley, UK), according to manufacturer's instructions. Genomic DNA was eluted in 100 $\mu$ l PureLink Genomic Elution Buffer (10mM Tris-HCl pH 9.0, 0.1mM EDTA), and the DNA concentration and purity determined by measuring absorbance at 260 and 280nm using the NanoDrop ND-1000 spectrophotometer (ThermoFisher, Life Technologies, Paisley, UK).



### **2.2.11 - Agarose gel electrophoresis**

1.5-2% agarose gels (molecular biology grade; Sigma-Aldrich, Gillingham, UK) were made, by boiling until dissolved in 150ml 1x TBE buffer (Sigma-Aldrich, Gillingham, UK). For DNA visualisation, 15µl SYBR Safe DNA gel stain (ThermoFisher, Life Technologies, Paisley, UK) was added before the gel set. The gel was then run in TBE buffer at 150V, 100mA (constant power), in a horizontal gel Electrophoresis tank (24x24cm H3 tank; Thistle Scientific, Glasgow UK). Samples that did not already contain Reddymix were resuspended in 6x Mass Ruler loading dye solution (ThermoFisher, Life Technologies, Paisley, UK). The following DNA size markers were used to estimate fragment sizes, depending on expected size range: MassRuler Low Range DNA Ladder, MassRuler, High Range DNA Ladder (ThermoFisher, Life Technologies, Paisley, UK) and NEB PCR size markers (New England Biolabs, Hitchin, UK). A UVIpro Platinum Digital Gel Doc System ultraviolet transilluminator (UVItec, Cambridge, UK) was used to visualise DNA.

### **2.2.12 - Extraction of total RNA from cultured cells**

RNA was prepared from adherent cell cultures, lysed in-situ. To do this RNA was extracted and purified using the RNeasy Plus Mini Kit (Qiagen Ltd., Manchester UK). Briefly, cells were washed twice with PBS and collected by scraping into RLT cell lysis buffer (Qiagen Ltd., Manchester UK) containing 1% β-mercaptoethanol. Cells lysates were then homogenised by centrifugation through a QIAshredder spin column (Qiagen Ltd., Manchester UK) at 13,000 rpm, 5 min, Sorvall Pico™ microcentrifuge (Sorvall, Leicester, UK). Following the RNeasy Plus protocol, column bound RNA was eluted serially from columns, first in 30µl, and then in 15µl nuclease-free water, and the eluted volumes pooled. RNA concentration and purity were

determined by measuring absorbance at 260 and 280nm with the NanoDrop ND-1000 spectrophotometer. Additionally, for samples undergoing Next Generation RNA sequencing, RNA integrity (RIN) (Schroeder *et al.*, 2006) was determined using an Agilent 2100 Bioanalyser (Agilent Technologies UK Ltd, Cheadle, UK). This was accessed as part of service run by the Imperial College BRC Genomics Laboratory (<http://www1.imperial.ac.uk/genomicsfacility>).

### **2.2.13 - Complementary DNA (cDNA) preparation by reverse transcription**

cDNA was prepared from total RNA using a recombinant M-MuLV reverse transcriptase (RevertAid Reverse Transcriptase enzyme; ThermoFisher, Life Technologies, Paisley, UK), primed with random hexamer primers (pdN6). Reactions (20µl) were prepared in thin walled reaction tubes, as follows;

4µl of 5x RT Reaction buffer (ThermoFisher, Life Technologies, Paisley, UK)

2µl of dNTP mix (10mM each; ThermoFisher, Life Technologies, Paisley, UK)

1µl of pdN6 random hexamer primers (0.2µg/µl; ThermoFisher, Life Technologies, Paisley, UK)

2µg of RNA made up to 12µl with nuclease free water

1µl of RevertAid Reverse Transcriptase enzyme (200u; ThermoFisher, Life Technologies, Paisley, UK)

The reaction mixture was mixed by gentle titration and then incubated on a programmable heating block (Applied Biosystems Veriti™ 96 well thermal cycler; Life Technologies, Paisley, UK) at 42°C for 1 h, followed by 5 min at 95°C to heat inactivate enzyme and denature cDNA and RNA template. The resulting cDNA was then diluted to a total volume of 200µl with nuclease free water. Typically, 2µl of diluted cDNA was used as template in PCR reactions.

#### **2.2.14 - TaqMan quantitative real-time PCR (qRT-PCR)**

Gene expression analysis was carried out using TaqMan gene expression assays (Applied Biosystems, Life Technologies, Paisley, UK) used on an Applied Biosystems 7900HT Fast Real-Time PCR System. TaqMan<sup>gene</sup> expression assay numbers are listed in Table 2.2. Each sample was assayed using four technical replicates organised on Microamp Fast Optical 96-well Reaction Plate (Applied Biosystems, Life Technologies, Paisley, UK), sealed using Microamp Optical Adhesive Film (Applied Biosystems, Life Technologies, Paisley, UK). Reactions were prepared as follows: 10µl TaqMan Fast Universal PCR master mix (Applied Biosystems, Life Technologies, Paisley, UK), 1µl TaqMan assay probe, 2µl cDNA, 7µl Molecular Biology grade water. The thermal cycling profile used was as follows: one cycle of 95°C for 20 s, followed by 40 cycles of 95°C for 1 second and 60°C for 20 s. The expression of the gene of interest was normalised to the housekeeping genes GAPDH and TBP, and expressed relative to the appropriate control, if applicable, using the comparative Ct method as follows:  $2^{-(Ct \text{ target gene} - Ct \text{ GAPDH, or TBP})}$ .

**Table 2.2: TaqMan Real-time PCR assay numbers (Applied Biosystems, Life Technologies, Paisley, UK)**

<b>Gene</b>	<b>Primer Code</b>
CDKN1A (p21)	Hs00355782_m1
GAPDH	Hs99999905_m1
MDM2	Hs01066930_m1
APOBEC3B	Hs00358981_m1
APOBEC3H	Hs00419665_m1
GALNT14	Hs00226180_m1
SPARC1	Hs00234160_m1
PXDN	Hs01034602_m1
VCAN	Hs00171642_m1
EIF1AY	Hs01040047_m1
SLC03A1	Hs00203184_m1
LARGE1	Hs00893935_m1
TCF21	Hs00162646_m1
TBP	Hs00427620_m1
ANGPT2	Hs00169867_m1
HKDC1	Hs00228405_m1
UTY	Hs01076483_m1
CHEK1	Hs00967506_m1
CHEK2	Hs0020048_m1
TP53	Hs01034249_m1
GADD45A	Hs00169255_m1
Survivin (BIRC5)	Hs00153353_m1
BAX	Hs00180269_m1
BTG2	Hs00198887_m1

### **2.2.15 - RNA-seq library construction**

RNA-seq library preparation and subsequent Next Generation Sequencing (NGS) was carried out using the service provided by the service run by the Imperial College BRC Genomics Laboratory (<http://www1.imperial.ac.uk/genomicsfacility>). RNA of sufficient quality, as judged by RIN analysis, was used to make libraries using the NEBNext poly(A) mRNA magnetic isolation module, NEBNext Ultra II Directional RNA Library Prep Kit for Illumina and NEBNext Multiplex Oligos (New England Biolabs, Hitchin, UK) and library concentrations measured using the Qubit™ 2.0 fluorometer (ThermoFisher, Life Technologies, Paisley, UK). The resulting library quality was additionally assessed using an Agilent 2100 Bioanalyzer loaded with a DNA high sensitivity (HS) chip (Agilent Technologies UK Ltd, Cheshire, UK).

### **2.2.16 - RNA-sequencing analysis**

RNA-seq was carried out for 75bp paired-end sequencing reads using the Illumina HiSeq 4000 sequencing platform (Illumina Cambridge Limited, Cambridge, UK), with multiplexing - here, a pool of 36 samples per lane, replicated across four lanes. Raw sequence reads were deconvoluted and provided according to the individual replicate address code. Subsequent RNAseq analysis, leading to differentially expressed gene lists, was carried out with Ms Van Nguyen, Surgery & Cancer, Imperial College London. Briefly, reads were aligned to the gencode v19 annotation file using TopHat 2.0.14 (Kim *et al.*, 2015) and the aligned reads counted using HTSeq 0.6.1.(Anders *et al.*, 2015) The programme 'DESeq2' (Love *et al.*, 2014) was used to normalise read counts using a regularized log transformation (log) and shrunken log<sub>2</sub> fold changes were calculated to determine differentially expressed genes between sets.

These lists were further analysed for overlap using the on-line Venny 2.1 Venn diagram tool (<https://bioinfogp.cnb.csic.es/tools/venny/>).

### **2.2.17 - Protein lysate preparation**

Protein lysates were made in-situ from cells grown in six-well plates using RIPA extraction buffer. For this, the culture medium was aspirated off and the cell monolayer washed twice with ice cold PBS. The washed cells were lysed with 200µl of ice cold RIPA buffer (50mM Tris-HCl pH8.0, 150mM NaCl, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% SDS; Sigma-Aldrich, Gillingham, UK), supplemented with a protease and phosphatase inhibitor cocktail (Roche, Welwyn Garden City, UK). Lysate was harvested by scraping on ice and transferred to 1.5ml microcentrifuge tubes on ice, and incubated continued on ice for 30 min, with vortexing every five min to enhance lysis. Cell lysates were cleared by centrifugation (13,000 rpm, 5 min, Sorvall Pico™ microcentrifuge; Sorvall, Leicester, UK) at 4°C. The cleared supernatant, containing soluble protein extract, was aliquoted into to chilled 1.5ml microcentrifuge tubes and stored at -80°C.

### **2.2.18 - Determination of protein concentration**

The concentration of the protein lysates was determined using the Pierce® BCA Protein Assay Kit (ThermoFisher, Life Technologies, Paisley, UK). A standard curve was generated using serial dilutions (1-30µg) of bovine serum albumin (BSA). The BCA reagent A and B were diluted 50:1 and 200µl of diluted BCA reagent was mixed with 10µl of lysates and BSA

standards in a 96 well plate (Dow Corning, Dewsbury, UK)). The samples were mixed and incubated at room temperature for 30 min. The optical density (OD) was measured at 562 nm. The OD of the standard was plotted, and the protein concentration of the samples was determined from the standard curve.

### **2.2.19 - SDS-Polyacrylamide gel electrophoresis and Western blotting**

Cell lysates were denatured at 100°C for 5 min prior to gel loading. Twenty micrograms of RIPA protein lysate were resolved on a 15% SDS-polyacrylamide resolving gel, with a 5% stacking gel, using a protein mini-gel electrophoresis system (GE lifesciences, Amersham UK). Electrophoresis was carried out in Tris- Glycine running buffer, pH 8.3, at 60V for 30 min and then at 100V for 1 to 2 h until the dye front reached the bottom of the resolving gel. Separated proteins on resolving gels were transferred onto Hybond ECL Nitrocellulose membrane (GE lifesciences, Amersham UK) at 100V for 90 min. The membrane was then incubated in blocking buffer (0.1% Tween, 5% dry skimmed milk in PBS) for 1 h at room temperature to reduce non-specific binding of primary antibody. The primary antibodies were diluted in blocking buffer and incubated with the membrane overnight at 4°C with gentle shaking. The membrane was then washed three times in PBS/ 0.1% Tween (15 min each wash) on a rocking platform and then incubated with the HRP-conjugated secondary antibody diluted in blocking buffer for 1 h. The membrane was then washed again three times in PBS/ 0.1% Tween, with rocking agitation. For visualisation, SuperSignal West Pico Chemiluminescent Substrate (Perbio Science UK Ltd., Cramlington, UK) was added to the membrane followed by autoradiography using Hyperfilm ECL (GE lifesciences, Amersham UK). Films were developed using a Konica SRX-1001A X-ray developer (Konica- Minolta, London, UK).

### **2.2.20 - Sanger DNA sequencing**

Sanger big-dye, di-deoxy terminator DNA sequencing was carried out on purified DNA (plasmid templates, or ExoSAP enzymically cleaned PCR products) using the GENEWIZ

(formerly Beckman) DNA sequencing service (<https://www.genewiz.com>).

## **3 – Results 1: p53 regulation of APOBEC3B and APOBEC3H expression**



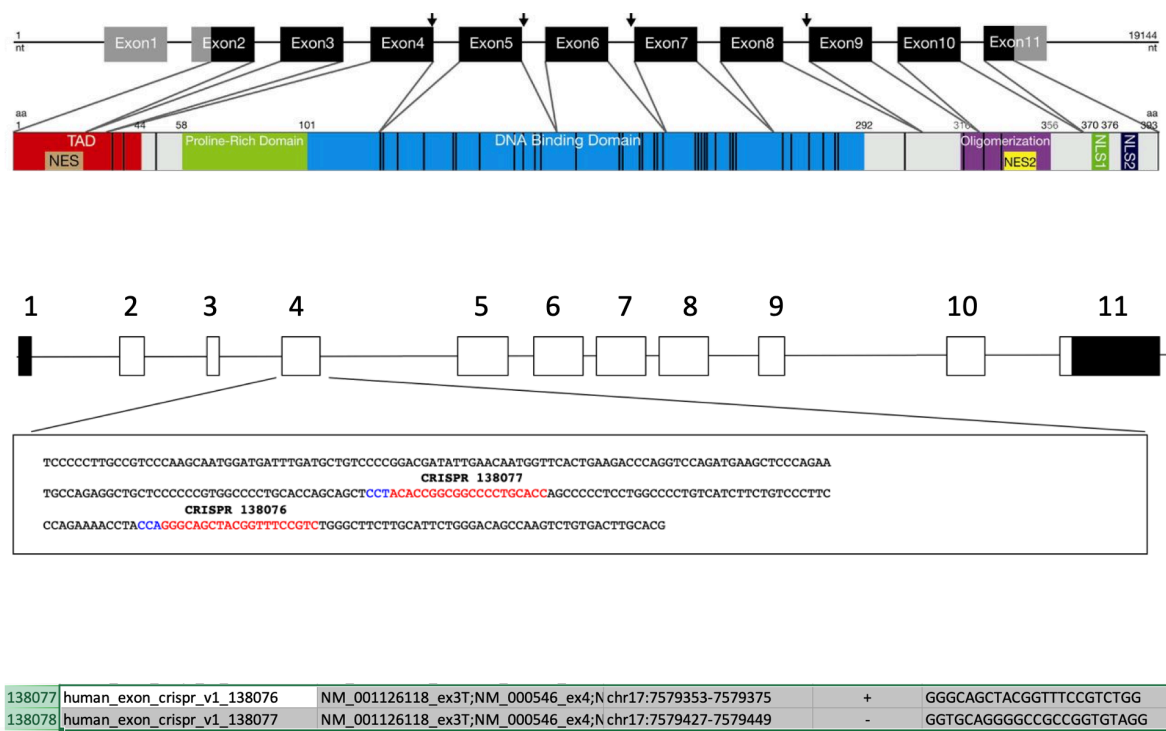
### 3.1 - p53 regulation of the APOBEC3 locus

Previous work to identify p53 regulated gene targets has suggested the possibility that APOBEC3 genes may be targets for p53 regulation (Periyasamy *et al.*, 2017). This regulation may be particularly crucial in estrogen receptor-positive breast cancer cells, as we have previously shown that APOBEC3B is a co-activator of the estrogen receptor (Periyasamy *et al.*, 2015), and APOBEC mediated mutational signatures (e.g. kataegis) (Seplyarskiy *et al.*, 2016) have been seen in the disease. In order to address the potential p53 regulation of APOBEC gene expression in this setting, we set out to make a p53 null derivative of the estrogen-responsive breast cancer cell line MCF7, which is known to have a wild-type p53 genotype. To do this, we used CRISPR/Cas9 mediated deletion within the p53 gene.

### 3.2 - Generation of p53 knock out MCF7 line

As shown in Figure 3.1, the p53 gene is comprised of 11 exons, with the first exon being non-coding. Inactivating mutations of p53 are frequently found in coding exon 3 (Exon 4) of the gene, which encodes part of the DNA binding domain of the protein. On this basis, we reasoned that deletions within coding exon 3, which would result in frameshift and premature termination of the protein, would be effective in obtaining a knockout. To make such a deletion, two CRISPRs, CRISPR138076 and CRISPR138077 were identified from the human exon targeted sgRNA database of (Mali *et al.*, 2013), which were localised to coding exon3 ( Figure 3.1). An important feature of the CRISPRs is that they start with a "G" base, making them amenable to efficient transcription by the U6 – RNA polymerase III promoter, which can be used to express sgRNAs. For the deletion, we used MCF-7 Luc cell line, which encodes a

constitutively expressed luciferase gene for use in cell tracking, and which is a line we have previously used in CRISPR/Cas9 mediated genome engineering.



**Figure 3.1 TP53 gene structure and domains organization**

The sequence for coding Exon 3, which encodes part of the p53 DNA binding domain is shown in the expanded panel (Dufour *et al.*, 2013), with the positions of CRISPR138077 (ACACCGGCGGCCCTGCACC) and CRISPR138076 (GGGCAGCTACGGTTTCCGTC) (Mali *et al.*, 2013) highlighted in red, and their PAM sequences highlighted in blue.

### **3.3 - Cloning and characterisation of p53 directed sgRNAs**

CRISPR138076 and CRISPR138077 were each obtained as two overlapping oligonucleotides, which were annealed and subsequently PCR amplified to produce DNA fragments suitable for DNA cloning. These fragments were subsequently cloned into the AflIII restriction enzyme site of the p41815 U6 promoter-sgRNA expression vector described by Mali *et al.*, (2013) and Sanger sequencing of ten plasmid mini-preparations for each construct used to identify recombinant clones. This resulted in the identification of two clones, plasmid clone two encoding CRISPR 138076 and plasmid clone ten encoding CRISPR 138077. These were subsequently expanded to make maxi-preparation DNA, which was further sequenced to confirm the cloning of the targeting CRISPR motifs (Figure 3.2).

### **3.4 - CRISPR mediated p53 coding exon 3 deletion in MCF7 Luc cells**

CRISPR mediated deletion of sequences in the third coding exon of TP53, the p53 gene, was carried out using an Amaxa Type II nucleofector (Lonza, Cologne, Germany) and the MCF7 transfection protocol recommended by the manufacturer. For the purposes of CRISPR mediated genome engineering, two sets of plasmids need to be co-transfected, these being the CRISPR expression plasmid and a Cas9 expression plasmid. In order to monitor for efficient co-transfection, the MCF7 Luc cells were co-transfected with the two TP53 coding exon3 CRISPR expression plasmids CRISPR 138076 and 138077, the hCas9 expression plasmid, and two additional fluorescent, reporter plasmids pMAX GFP and pM-Cherry expression plasmids. These cultures were then visualized 48 h following transfection in bright field and for GFP and RFP expression. As can be seen in Figure 3.3A, there is high concordance between the

transfected GFP and RFP expression, indicating efficient co-transfection, and by inference, efficient co-transfection of the CRISPR 138076 and 138077, the hCas9 expression plasmids.

A

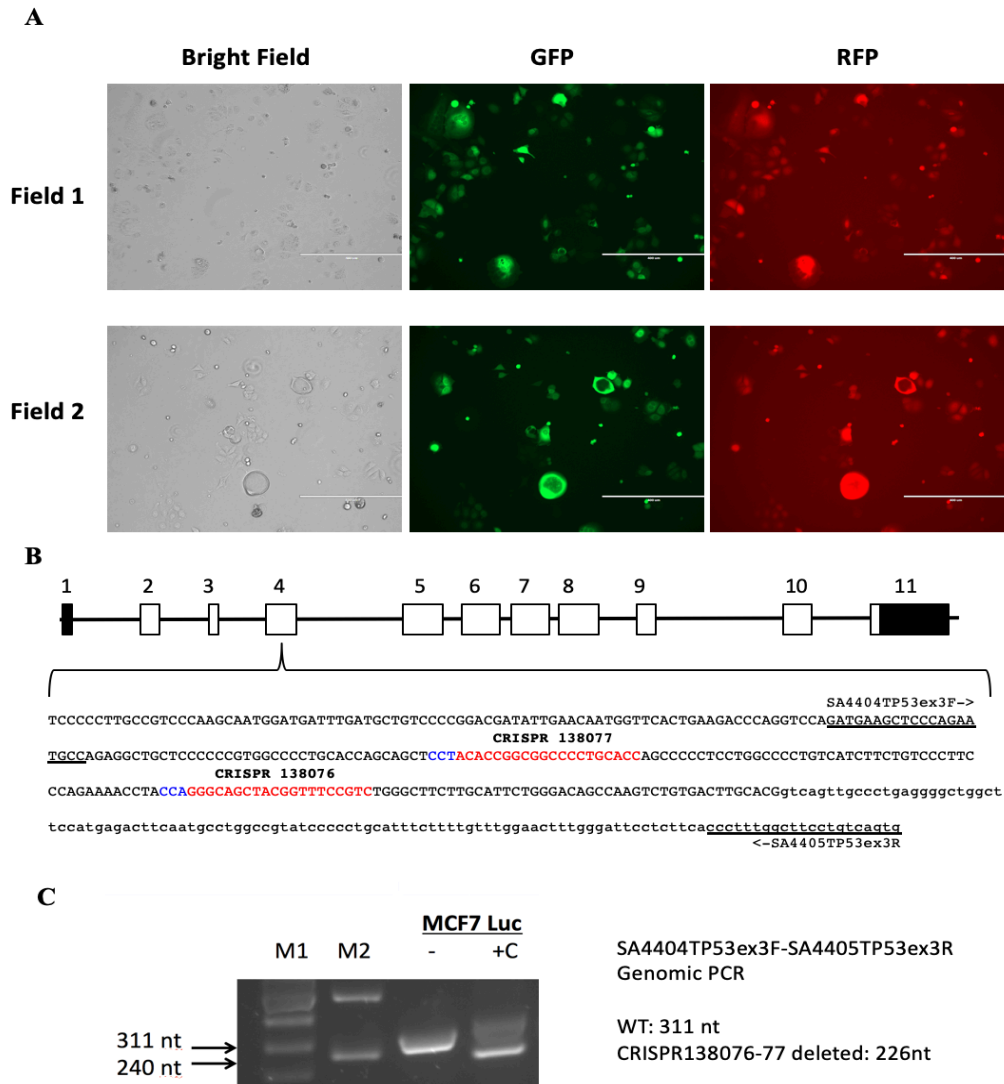


B



**Figure 3.2. TP53 coding exon 4 (coding exon 3) CRISPR guide RNA cloning**

**A.** Following Gibson cloning, mini- preparation for both CRISPRs was carried out and 10 samples prepared. Based on sgRNA transcription start sequence (i.e. Antisense for CRISPR 138076 GACGGAAACCGTAGCTGCCc and Antisense for CRISPR 138077 ACACCGGCGGCCCTGCACc). Clones 10 and 2 for CRISPR's 138076 and 138077 respectively were positive and used for expansion to maxi-prep. **B.** Maxi-prep sequences of both CRISPR's 138076 and 138077. For CRISPR 138076 one plasmid prepared while CRISPR 138077 two plasmids prepared both at higher concentrations. Both minipreps and maxipreps sequenced with 30 uM of our SA3118 primers. Positive clones are shown and CRISPR sequence highlighted.



**Figure 3.3. CRISPR 138076 -138077 / Cas9 mediated DNA deletion in TP53 exon 4 (coding exon 3)**

A. CRISPR mediated deletion of sequences in the third coding exon of TP53 was carried out using an Amaxa Type II nucleofector (Lonza, Cologne, Germany) and the MCF7 transfection protocol recommended by the manufacturer. For the purposes of CRISPR mediated genome engineering, plasmids need to be co-transfected, these being a CRISPR expression plasmids and a Cas9 expression plasmid. co- transfected with the pMAX GFP (Green Fluorescent Protein) and pM-Cherry (Red Fluorescent Protein-RFP) expression plasmids, using Nucleofection under conditions previously optimised for the MCF7 cell lines were visualised 48 h following transfection in bright field and for GFP expression (x10 magnification). There was a very high proportion of transfected cells (estimated >70%). These nucleofection conditions were used to co-transfect plasmid expression constructs for humanised Cas9, and the two TP53 Exon 4 (coding Exon3) CRISPR expression plasmids encoding CRISPR 138076 and 138077, as shown in B. Evidence for CRISPR mediated DNA deletion was carried out using a PCR based assay with primers SA4404TP53ex3F and. SA4405TP53ex3R. As can be seen in C, the co-transfected Cas9/CRISPR pool contain both full-length (311 nt) and CRISPR deleted Exon 4 (226 nt) sequences.



Following nucleofection, the cells were divided into two pools. The first was allowed to grow to confluence, passaged, and used to make genomic DNA for PCR analysis for evidence of DNA deletion within TP53 coding exon 3. As shown in Figure 3.3B, PCR primers (SA4404TP53ex3F and SA4405TP53ex3R) were designed to allow amplification of DNA spanning a region 63 bases upstream of the cleavage site for CRISPR138077 and 163nt downstream of CRISPR138076, within the intron. This assay will produce a 311nt product from wild-type MCF7 Luc cells, but in CRISPR deleted alleles will result in a 226nt product. As can be seen in Figure 3.3C, analysis of the nucleofected pool clearly shows the presence of a strong SA4404TP53ex3F - SA4405TP53ex3R CRISPR 138076 -138077 deleted PCR product, suggesting both efficient co-transfection and strong CRISPR activity in these experiments.

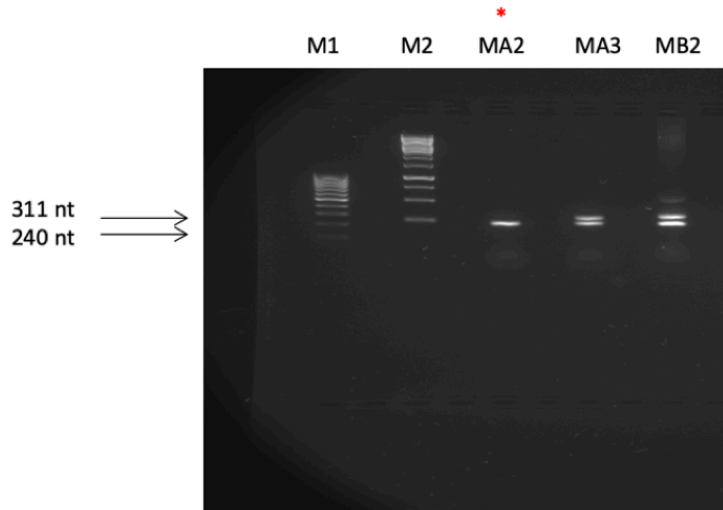
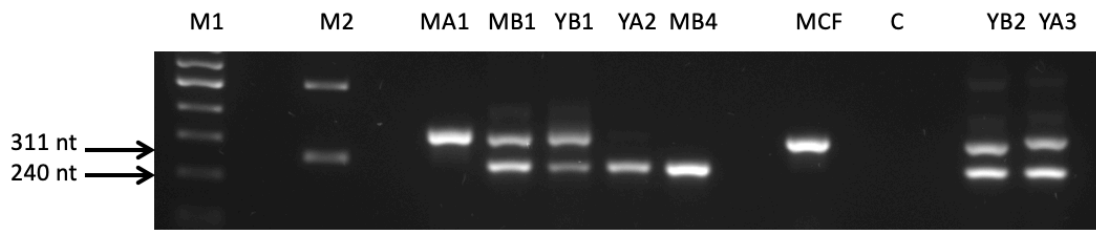
In order to further assess the effectiveness of the strategy used here to make p53 knockout cells, the sensitivity of the CRISPR 138076 -138077 / Cas9 treated nucleofected pool of MCF7 Luc cells to Nutlin was examined. As described previously, Nutlin inhibits the interaction between mdm2 and p53, blocking p53 activation. In the wild-type p53 context, Nutlin treatment is growth inhibitory and can lead to cell death (Shangary and Wang, 2009). As can be seen in Figure 3.5 following 6 days of nutlin treatment with 10 $\mu$ M, which is known to activate p53 in MCF7 cells (Periyasamy *et al.*, 2017), more cells survived in the CRISPR 138076 -138077 / Cas9 treated nucleofected pool, when compared to the pMAX GFP and pM-Cherry only nucleofection controls. These experiments indicated a marked, reduced Nutlin sensitivity in the CRISPR 138076 -138077 / Cas9 treated nucleofected pool, suggesting the presence of p53 knockout cells in this population.

Having established p53 knockout cells, single-cell cloning was carried out, in order to obtain knockout lines from the CRISPR 138076 -138077 / Cas9 treated nucleofected pool of MCF7 Luc cells. Cells were seeded at densities of 3,000 -10,000 cells in 15 cm dishes and left to establish clones of greater than 1,000 cells (9-10 doublings, equivalent to 3 weeks). Individual clones were then harvested *in-situ*, on to cloning discs and transferred to the wells of 24 well dishes, with 48 clones being isolated. Once confluent, these were passaged up to six-well plates and, when confluent, used to make genomic DNA for PCR analysis of 48 clones, using the SA4404TP53ex3F - SA4405TP53ex3R assay, which identified six clones with TP53 coding exon 3 deletions (Figure 3.4). Of these, four were heterozygote, showing both wild-type (311nt) and CRISPR 138076 -138077 deleted (240 nt) PCR products (MCF7 Luc  $\Delta$ p53 MA2 clones MB1, YB1, YB2 and YA3). A further two clones, MCF7 Luc  $\Delta$ p53 MA2 and MCF7 Luc  $\Delta$ p53 MB4, showed only a single product, indicative of the possibility of a homozygous CRISPR 138076 -138077 / Cas9 mediated deletion in TP53 coding exon 3. Assuming a diploid TP53 genotype, the cloning exercise potentially represents 88 non-deleted alleles and 8 deleted alleles, or efficiency of deletion of around 8.3%.

The two putative deletion clones MCF7 Luc  $\Delta$ p53 MA2 and MCF7 Luc  $\Delta$ p53 MB4 were selected for further characterisation. To do this, the genomic SA4404TP53ex3F - SA4405TP53ex3R PCR product from the lines was sequenced from both ends. Figure 3.6 shows the result of this analysis for the MCF7 Luc  $\Delta$ p53 MA2 clone, and shows this is homozygous for the intended, engineered gene deletion. The resulting deleted TP53 gene encodes a wild-type p53 sequence up to codon 81, with the mutant exon 3 encoding a further nine amino acid sequence (SCILGQPSL) not found in p53, followed by a termination codon (TGA). This results in a truncated mutant TP53 that is only 90 amino acids long and would be expected to be non-functional (Figure 3.6B). The second clone, MCF7 Luc  $\Delta$ p53 MB4,

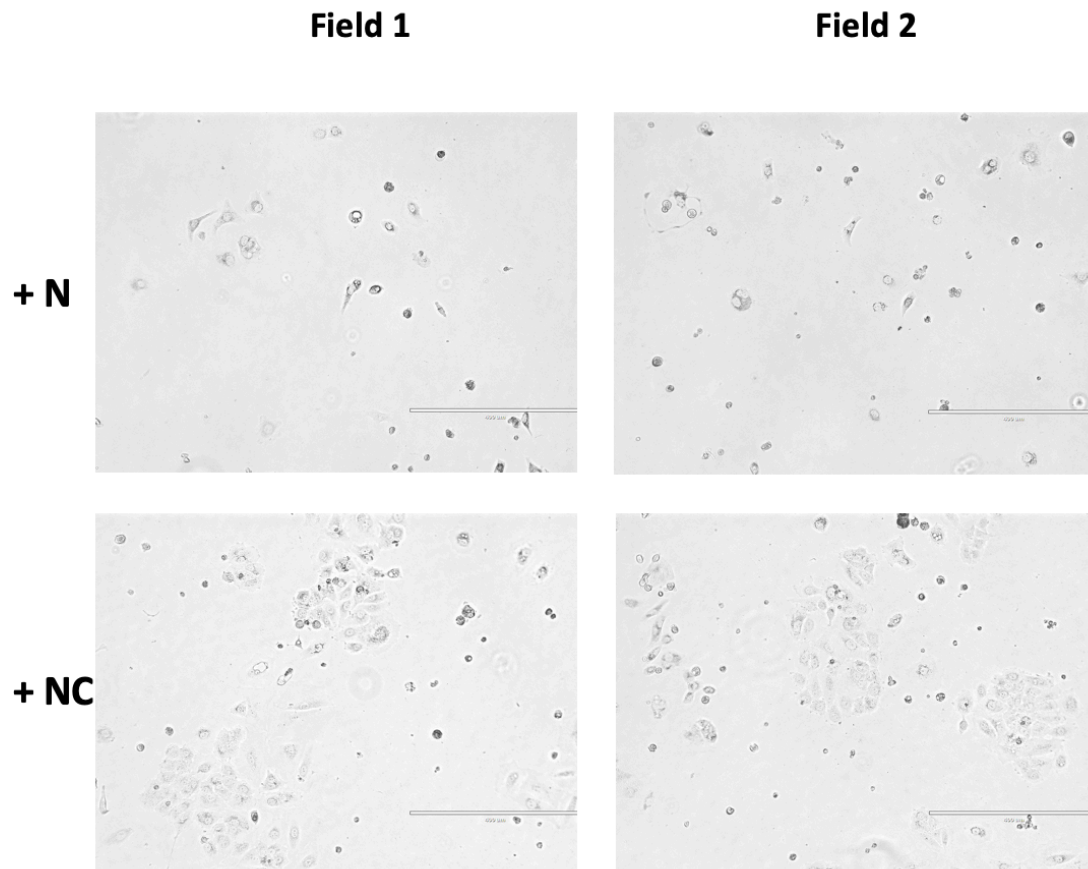
produced a mixed sequence chromatogram at the junction of the CRISPR 138076 -138077 fusion point (Figure 3.6C). Further characterisation of the two overlapping sequences showed that one came from an allele encoding a CRISPR 138076 -138077 mediated deletion, but incorporated an additional insertion of a “T” at the junction between the two CRISPR cut sites, creating an indel mutation (Figure 3.6C) This produces a mutant, that encodes an in-frame p53, which lacks the coding exon 3 deleted region, which incorporates part of the DNA binding domain The second allele encodes a combination of an insertion of seven bases and deletion of five. Overall, this allele encodes a truncated, 93 aa p53 protein that lacks the DNA binding domain in its entirety (Figure 3.6C).

Despite the fact that both the MCF7 Luc  $\Delta$ p53 MA2 and MB4 lines encode deletions in the TP53 gene, the presence of a near full-length coding allele in the MB4 line, with potential for some functionality, led us to not proceed further with this line. Hence further work was carried out only with the MA2 line.



**Figure 3.4. PCR characterisation of MCF7 Luc clones following CRISPR138077/CRISPR138076/ Cas9 co-transfection**

MCF7 Luc cells were co-transfected with CRISPR138077, CRISPR138076 and humanised Cas9 expression plasmids by nucleofection. Transfected cells were subsequently cloned and used to make genomic DNA. PCR analysis of TP53 Exon 4 was carried out on this DNA using the primers SA4404TP53ex3F (GATGAAGCTCCCAGAATGCC) SA4405TP53ex3R(cactgacaggaagccaaagg), which amplify a 311nt product from wild-type p53 cells. Digestion of Exon 4 by CRISPR 138076 and 138077 results in the excision of 71nt, and the generation of a 240nt PCR product. DNA size markers are shown in the first two lanes; M1 Low Range Mass Rule, M2 1 Kb ladder. MCF shows the Exon 4 PCR product from untransfected, control cells and C the negative, no template control.

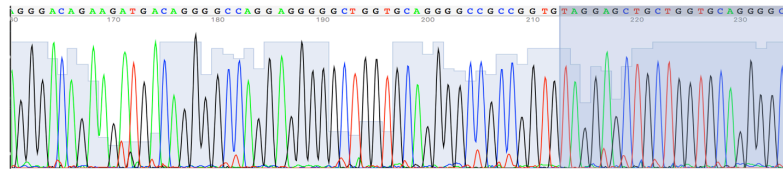


**Figure 3.5. Differential growth inhibition by Nutlin in wild-type and p53 knockout MCF7 Luc cells**

The growth response to Nutlin was assessed in MCF7 Luc cells following CRISPR mediated deletion of sequences in the third coding exon of the TP53 gene. Cells were treated with 10 $\mu$ M Nutlin, 72 h following nucleofection with plasmid expression constructs. For control cells, co-transfection was carried out with pMAX GFP (Green Fluorescent Protein) and pM-Cherry (Red Fluorescent Protein-RFP). For TP53 knockout, co-transfection was with expression plasmids for humanised Cas9, and the two TP53 exon 4 (coding exon 3) CRISPRs 138076 and 138077. Images were taken on Day 6 of Nutlin treatment, with two fields of view shown for the Nutlin treated, control, wild-type MCF7 Luc cells (+N) and CRISPR 138076/138077-Cas9 co-transfected cells (+NC), Magnification was x10, with scale bars of 400 $\mu$ M shown for each image. This analysis shows that growth is inhibited to a lesser extent by Nutlin in the CRISPR 138076/138077-Cas9 co-transfected cells.

**A**

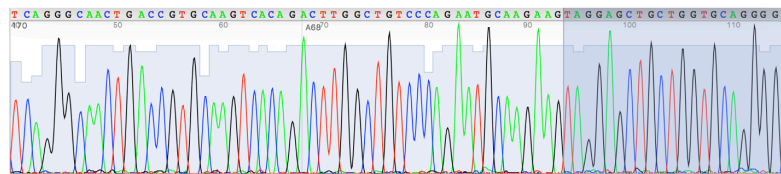
**MCF7 Luc TP53 coding Exon 3**



TCC CCC TTG CCG TCC CAA GCA ATG GAT GAT TTG ATG CTG TCC CCG GAC GAT ATT GAA CAA TGG TTC ACT GAA GAC CCA GGT CCA GAT GAA  
 S P L P S Q A M D D L M L S P D D I E Q W F T E D P G P D E  
 GCT CCC AGA ATG CCA GAG GCT GCT CCC CCC GTG GCC CCT GCA CCA GCA GCT CCT ACA CCG GCG GCC CCT GCA CCA GCC CCC TCC TGG CCC  
 A P R M P E A A P P V A P A P A A P T P A A P A P A P S W P  
 CTG TCA TCT TCT GTC CCT TCC CAG AAA ACC TAC CAG GGC AGC TAC GGT TTC CGT CTG GGC TTC TTG CAT TCT GGG ACA GCC AAG TCT GTG  
 L S S S V P S Q K T Y Q G S Y G F R L G F L H S G T A K S V  
 ACT TGC ACG  
 T C T

**B**

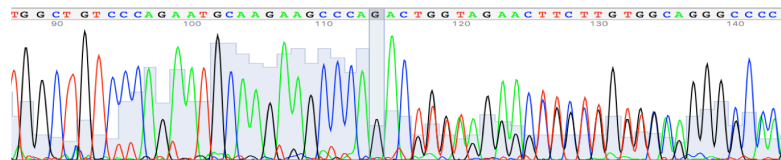
**MCF7 Luc Clone MA2 TP53 coding Exon 3**



TCC CCC TTG CCG TCC CAA GCA ATG GAT GAT TTG ATG CTG TCC CCG GAC GAT ATT GAA CAA TGG TTC ACT GAA GAC CCA GGT CCA GAT GAA  
 S P L P S Q A M D D L M L S P D D I E Q W F T E D P G P D E  
 GCT CCC AGA ATG CCA GAG GCT GCT CCC CCC GTG GCC CCT GCA CCA GCA GCT CCT ACT TCT TGC ATT CTG GGA CAG CCA AGT CTG TGA CTT  
 A P R M P E A A P P V A P A P A A P T S C I L G Q P S L \*

**C**

**MCF7 Luc Clone MB4 TP53 coding Exon 3**



**MCF7 Luc Clone MB4 TP53 coding Exon 3 – allele 1**

TCC CCC TTG CCG TCC CAA GCA ATG GAT GAT TTG ATG CTG TCC CCG GAC GAT ATT GAA CAA TGG TTC ACT GAA GAC CCA GGT CCA GAT GAA  
 S P L P S Q A M D D L M L S P D D I E Q W F T E D P G P D E  
 GCT CCC AGA ATG CCA GAG GCT GCT CCC CCC GTG GCC CCT GCA CCA GCA GCT CCT ACA AGT CTG GGC TTC TTG CAT TCT GGG ACA GCC AAG  
 A P R M P E A A P P V A P A P A A P T S L G F L H S G T A K  
 TCT GTG ACT TGC ACG  
 S V T C T

**MCF7 Luc Clone MB4 TP53 coding Exon 3-5 – allele 2**

tcc ccc ttg ccg tcc caa gca atg gat gat ttg atg ctg tcc ccg gac gat att gaa caa tgg ttc act gaa gac cca ggt cca gat gaa  
 s p l p s q a m d d l m l s p d d i e q w f t e d p g p d e  
 gct ccc aga atg cca gag gct gct ccc ccc gtg gcc ctt gca cca gca ctt ctt agc gtc tgg gct tct tgc att ctg tgg gct tct tgc  
 a p r m p e a a p p v a l a p a l l s v w a s c i l w a s c  
 att ctg gga cag cca agt ctg tga  
 i l g q p s l \*

**Figure 3.6. TP53 exon 3 allele sequences from wild-type and TP53 knockout MCF7 Luc cells**

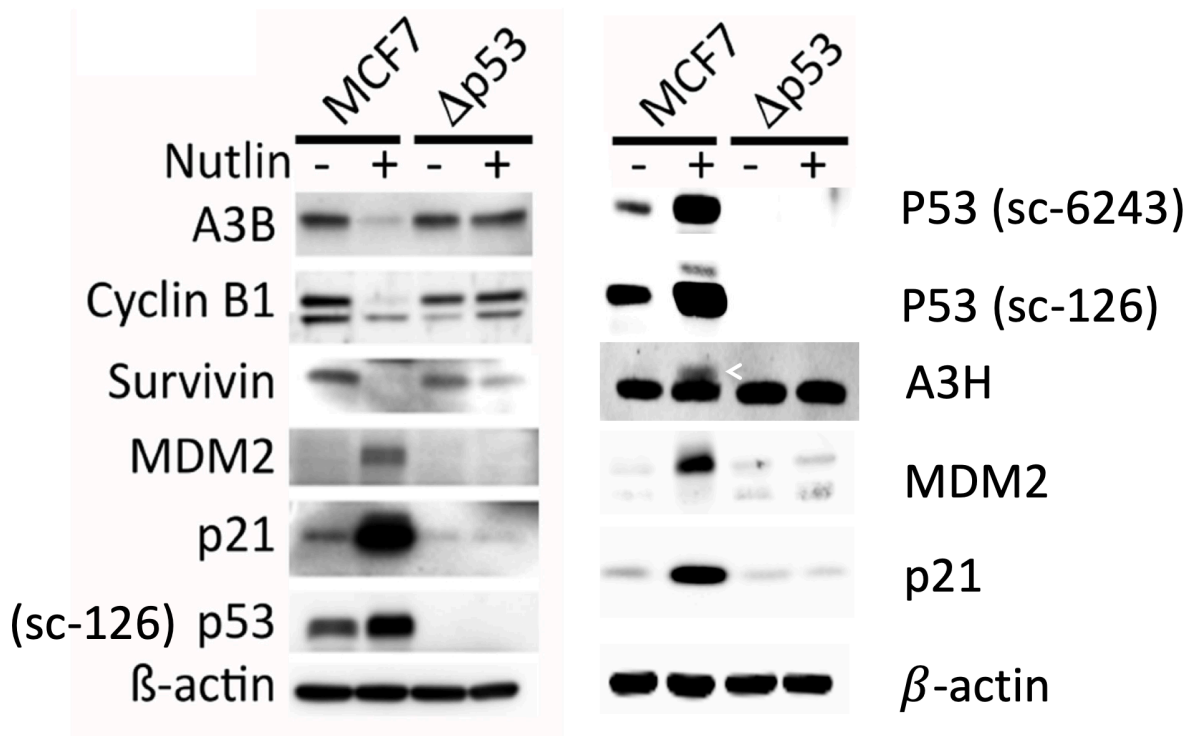
**A.** Wild-type MCF7 Luc TP53 Exon 3 sequence. The upper panel shows part of the sequence chromatogram for a genomic PCR product, in antisense orientation, for the wild-type exon. The lower panel shows the translated reading frame for Exon 3, used to encode p53. **B.** The MCF7 Luc Clone MA2 cell line encodes two allelic TP53 genes, each featuring a 94nt deletion spanning bases 147-240 of Exon 3. The upper panel shows part of the sequence chromatogram for a genomic PCR product from the clone for TP53 Exon 3, aligned to the wild-type sequence chromatogram in A (shaded region). The lower panel shows the translated reading frame encoding by the CRISPR 138076-138077 deleted p53 Exon 3. This deletion removes the CRISPR 138076 target and PAM sequence, while retaining the PAM (blue) and last two bases (red) of the CRISPR138077 target sequence. The resulting deleted Exon 3 encodes a wild-type p53 sequence up to codon 81. The mutant Exon3 encodes a further nine amino acid sequence (SCILGQPSL) not found in p53, followed by a termination codon (TGA) that results in a truncated mutant TP53 that is only 90 amino acids long. **C.** The MCF7 Luc Clone MB4 cell line encodes two allelic, mutated, TP53 genes, as seen in the sequence chromatogram of part of the TP53 Exon3 region in this line (upper panel). De-convolution of this chromatogram shows that one allele features a 93nt deletion spanning bases 147-240 of Exon 3, with an insertion of a single T at the junction. While deleting a region of coding Exon 3, this restores the reading frame, leading to a 366 amino acid protein with homology to p53 along its length. For the second allele, a deletion seen that removes the DNA between the CRISPR 138076 target and CRISPR138077 target sequences. The resulting deleted Exon 3 encodes a wild-type p53 sequence up to codon 74, followed by coding that leads to a further 25 amino acid sequence not found in p53, and a termination codon (TGA). This results in a truncated mutant TP53 that is only 99 amino acids long.

### 3.5 - p53 expression and functionality in the MCF7 Luc $\Delta$ p53 MA2 line

Initially, Western blotting analysis was used to assess p53 expression and activity in both wild-type MCF7 Luc and MCF7 Luc  $\Delta$ p53 MA2, using DMSO (control) and Nutlin treated cells. Following 24 h treatment with DMSO or Nutlin at 10 $\mu$ M, cells were used to make protein lysates, which were used in Western blotting. Figure 3.7 shows basal p53 expression in MCF7 Luc cells, which was significantly induced following Nutlin treatment (lanes 1 and 2). This expression was detected by two p53 antibodies, sc-6243 (rabbit polyclonal antiserum raised against full length recombinant human p53), and sc-126 (mouse monoclonal antibody raised against a 15aa peptide covering residues 11-25 of the protein). Probing lysates from the MCF7 Luc  $\Delta$ p53 MA2 line, with and without Nutlin, did not detect any p53 expression. Importantly, neither antibody detected a truncated p53 peptide, suggesting that the deleted allele produces an unstable protein. Probing for  $\beta$ -Actin and Lamin A/C (both control housekeeping proteins) shows difference were not due to sample loading.

Further blotting analyses were carried out for the proteins MDM2 and p21, both of which are known as p53 targets. In wild-type cells, both MDM2 and p21 showed weak, basal expression in control (DMSO) conditions, but were potently upregulated by Nutlin treatment (Lanes 1 and 2). In the MCF7 Luc  $\Delta$ p53 MA2 line, expression of both proteins were at a level equitable to the basal level seen in the wild-type line. However, no Nutlin stimulation of either protein was seen in this line.



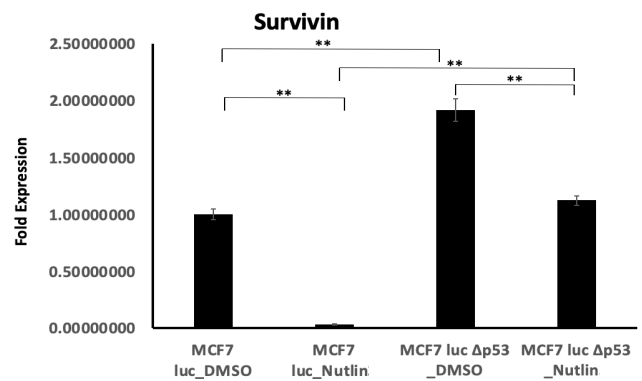
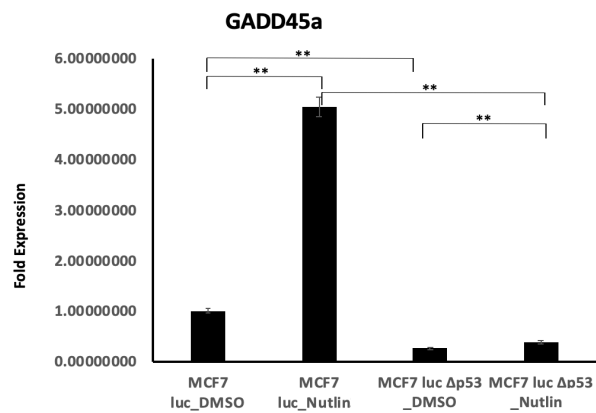
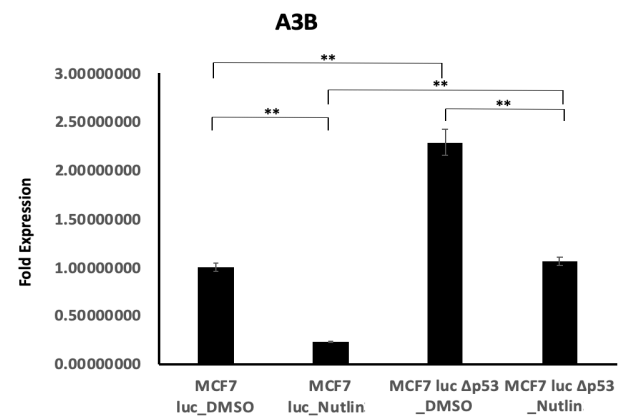
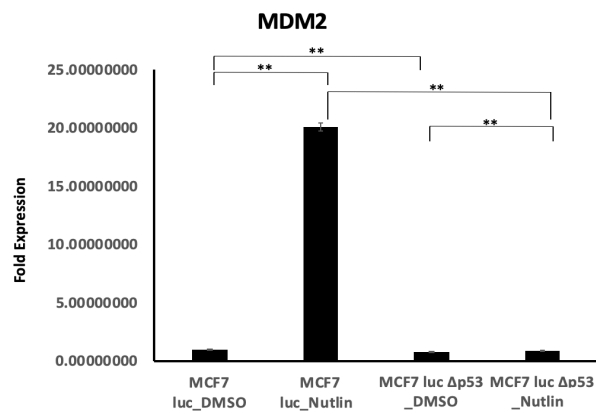
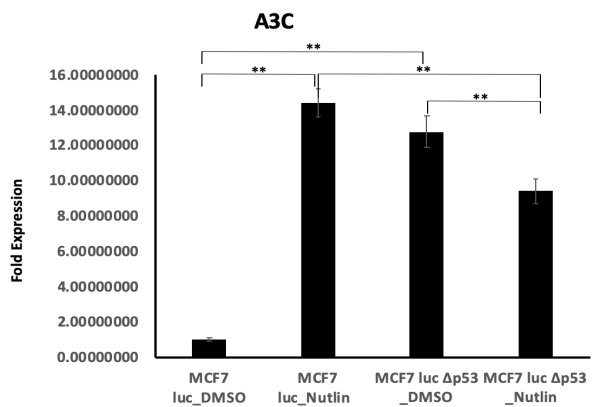
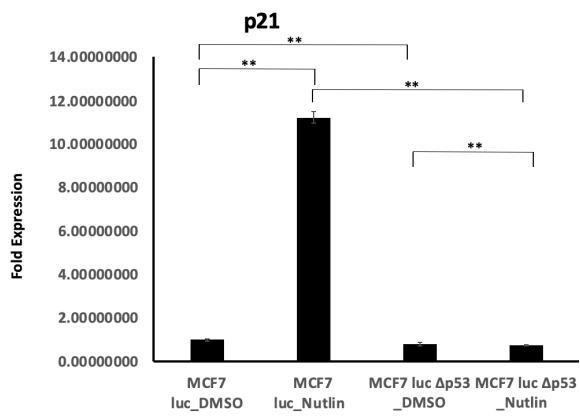
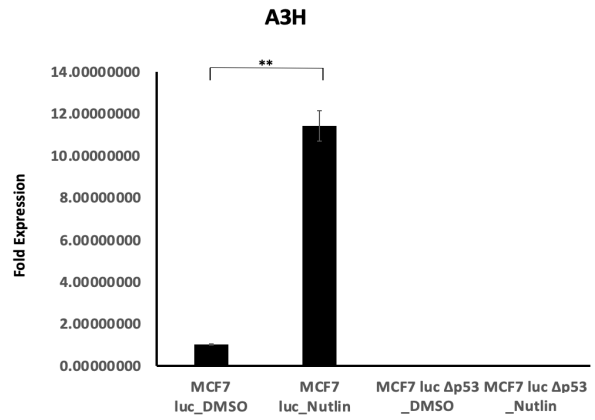
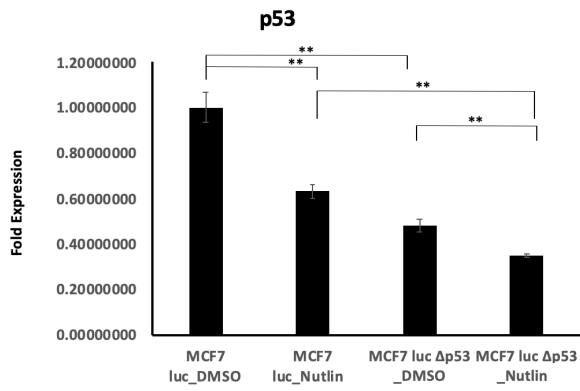


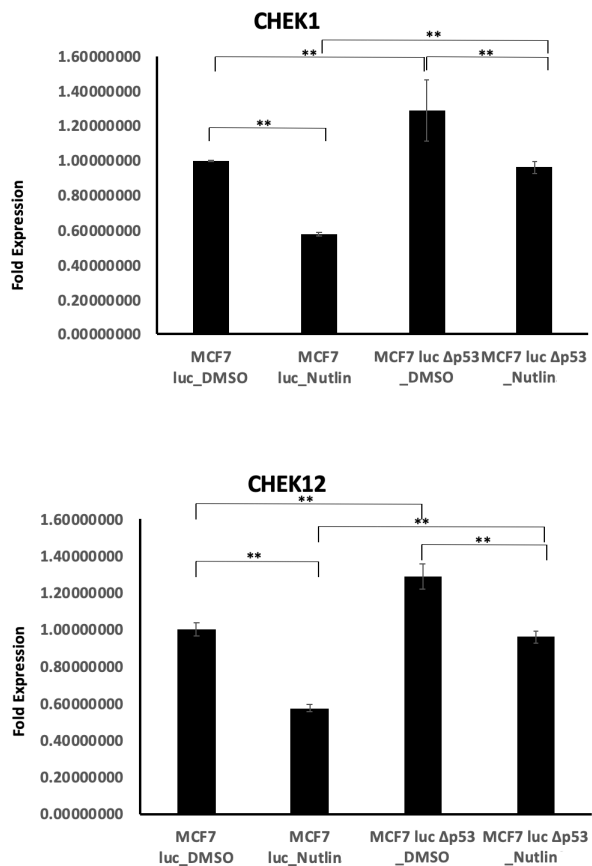
**Figure 3.7. Western blot analysis of p53 and p53 regulated expression in wild-type and p53 null, CRISPR-Cas9 engineered MCF7 Luc cells**

Western blot analysis was carried out for p53 and the p53 regulated proteins APOBEC3B (A3B), Cyclin B1, Survivin, MDM2 and p21 in the presence (+) and absence (-) of Nutlin, both in wild-type MCF7 Luc cells (MCF7), and the MCF7 Luc  $\Delta$ p53 MA2 line. These blots show no expression of p53 in the MCF7 Luc  $\Delta$ p53 line. Further, the p53 target proteins examined, failed to show Nutlin regulation in the p53 knockout line.

Expression of p53 and associated genes were also examined at the mRNA level using Q-RT-PCR analysis. The p53 TaqMan assay could detect both coding exon 3 mutant and wild-type transcripts. Figure 3.8 shows a decrease in p53 transcript following Nutlin treatment. In the MCF7 Luc  $\Delta$ p53 MA2 line, p53 transcript levels are lower, showing a similar fall in expression following Nutlin treatment. As expected, both, p21 and MDM2 expression are strongly stimulated by Nutlin treatment in wild-type MCF7 Luc cells. However, in the MCF7 Luc  $\Delta$ p53 MA2 line, expression is maintained at low, basal levels, following Nutlin treatment. A third p53 target, GADD45a, was included in this analysis, and Nutlin regulated expression of this confirmed the lack of a p53 response in the MCF7 Luc  $\Delta$ p53 MA2 line.

Taken together, both protein expression and gene expression data confirm that the MCF7 Luc  $\Delta$ p53 MA2 line behaves as a robust p53 knockout line.





**Figure 3.8. qRT-PCR analysis of p53 regulated gene expression in MCF7 Luc wild-type and MCF7 Luc Δp53 MA2 cells**

For qRT-PCR analysis, MCF7 Luc wild-type (MCF7 Luc) and MCF7 Luc Δp53 MA2 cells (MCF7 Luc Δp53) were grown in media supplemented with 10μM Nutlin, or vehicle (DMSO), and RNA prepared after 24 h. The RNA was subsequently used to make cDNA, and this used in qPCR, using TaqMan Gene Expression Assays. Gene expression was normalized with respect to GAPDH, house-keeping gene expression, and shown as fold relative to the vehicle treated MCF7-Luc control. This shows Nutlin regulated p53 gene expression is not a feature of the MCF7 Luc Δp53 MA2 line.

### 3.6 - Use of the MCF7 Luc $\Delta$ p53 MA2 line to study regulation of APOBEC3H

The MCF7 Luc  $\Delta$ p53 MA2 line, was used to evaluate APOBEC3B and APOBEC3H regulation by p53. To do this, MCF7 Luc  $\Delta$ p53 MA2 and parental p53 wild-type MCF7 cells were treated with and without Nutlin, and both RNA and protein lysates made. Protein lysates were used in Western blot analysis for APOBEC3H and APOBEC3B (Figure 3.7), using MDM2 and p21 induction as a measure of p53 activation. For APOBEC3H, it can be seen that Nutlin induced p53 activation in the MCF7 Luc  $\Delta$ p53 MA2 line does not result in the induction of either MDM2 or p21, which can be seen in control, wild-type cells. Further, APOBEC3H behaves in a similar way, showing Nutlin induced expression in the wild-type MCF7 cells, but not in the MCF7 Luc  $\Delta$ p53 MA2 line.

This conclusion was further investigated by examining gene expression, as shown in (Figure 3.8) Here it can be seen that in the wild-type MCF7 cells, APOBEC3H behaves as a strongly Nutlin induced, p53 activated gene, while in the MCF7 Luc  $\Delta$ p53 MA2 line, basal expression is lost, and Nutlin induced, p53 regulated expression is no longer seen.

By contrast to APOBEC3H, APOBEC3B is seen to be down-regulated following Nutlin treatment. To examine this further in MCF7, we have also examined the expression of several p53 repressed targets in wild-type and MCF7 Luc  $\Delta$ p53 MA2 cells, namely Survivin, CHECK1, CHECK2 and Cyclin B1. As can be seen in (Figure 3.8) at the gene expression level, Survivin, CHECK1 and CHECK2, are all up-regulated in MCF7 Luc  $\Delta$ p53 MA2 cells, when compared to MCF7 wild-type cells. As can be seen in Figure 3.7 the p53 mediated repression of Survivin was further demonstrated at the protein level in wild-type MCF7 cells but was found to be constitutively activated in the mutant MCF7 Luc  $\Delta$ p53 MA2 cells. This was also found to be the case for the p53 repressed target Cyclin B1. For APOBEC3B, expression

followed the pattern seen for the p53 repressed targets, namely increased basal expression in the MCF7 Luc  $\Delta$ p53 MA2 cells, when compared to MCF7 wild-type cells, with no Nutlin mediated repression in the MCF7 Luc  $\Delta$ p53 MA2 line.

### 3.7 - Discussion

In this chapter I have described a novel approach was used to create p53 null mutations, by the use of a pair of CRISPR guide RNAs to direct Cas9 to make a deletion within coding exon 3 (exon 4) of the TP53 gene which encodes part of the DNA binding domain of the p53 protein, and is a region which exhibits several p53 cancer mutations. This region was also chosen because of the identification of a suitable pair of CRISPR guide sequences, CRISPR138076 and CRISPR138077 from the Human exon targeted sgRNA database (Mali *et al.*, 2013) which is a collection of algorithm designed human exome specific guide RNAs, which conform to the requirement of starting with a guanine ("G"), thereby facilitating transcription using a U6 RNA polymerase III driven gene promoter. This gene promoter has been widely used to express small structured RNAs, including CRISPR guide RNAs, in a wide range of vertebrate target cell backgrounds, and was made available in a plasmid vector encoding both the U6 gene promoter and the framework for a Cas9 guide RNA without a CRISPR sequence. Gibson Assembly, an efficient and flexible *in-vitro* DNA recombination system for cloning DNA, carries out cloning of the appropriate CRISPR sequence into this. (Gibson *et al.*, 2010; Gibson *et al.*, 2009) This was further adapted, so as to incorporate steps to use annealed oligonucleotides (60mers), to make 100nt double-stranded DNA CRISPR encoding templates for Gibson cloning, and resulted in a streamlined, efficient cloning strategy, with plasmid cloning efficiencies of >90%.

The strategy to use a pair of CRISPR gRNAs to make an exon specific deletion has several advantages, in that it potentially offers the use of simple genomic PCR assays to detect the presence of CRISPR-Cas9 digested cells in pools of CRISPR gRNA, Cas9 co-transfected cells. This approach contrasts to other more problematic approaches, which have used DNA mismatch assays, such as the Surveyor nuclease assay (Qiu *et al.*, 2004), which are widely regarded as being of low sensitivity (<3%), and which are most often used to detect indel mutations generated at discrete, targeted sequences (Sentmanat *et al.*, 2018). Nevertheless, the approach of using CRISPR-Cas9 to make p53 knockout lines through indel mutation, as opposed to overt DNA deletion, has been successfully used in several model systems, including mouse (Walton *et al.*, 2016) and canine cells (Eun *et al.*, 2019).

An obvious limitation of using the paired CRISPR gRNA approach to obtaining DNA deletions *in-situ* is the requirement for cutting to occur simultaneously at both target sites, with the subsequent cleavage leading to non-homologous end-joining between the two cut sites during DNA repair, rather than repair through end-joining localised at the site of each of the two CRISPR-Cas9 cleavages. Indeed, in breast cancer cells, DNA repair by non-homologous end joining is low and comparatively inefficient (Mao *et al.*, 2009). Despite this, our experience with the use of CRISPR138076 and CRISPR138077, together with the Cas9 nuclease in MCF7 cells, shows that DNA deletion does occur frequently enough to be detected in both the original nucleofected cell pool and also in clones selected from this. However, in clones showing deletion, the predominant genotype was seen to be heterozygote, with deletion of one allele only, indicating that while simultaneous cleavage mediated by the two CRISPR gRNA is possible, this activity is often not sufficient to obtain deletion of all alleles present (Shadeo and Lam, 2006). This conclusion is not necessarily affected by TP53 copy number in MCF7, which

has been reported to be diploid for TP53 (Shadeo and Lam, 2006)), but clearly this could be compounded if deletions were attempted in lines with increased TP53 copy number.

The successful generation of the MCF7 Luc  $\Delta$ p53 MA2 line together with the matched wild-type MCF7 line provides a unique set of lines to study p53 in a breast cancer cell setting. Having developed these, we have gone on to use them to study p53 regulation of APOBEC3B. As reported in our paper (Periyasamy *et al.*, 2017) APOBEC3B expression is seen to be inversely related to p53 status in a series of different cancer cell lines, such that p53 represses expression of the gene. This was found to occur primarily through the induction of p21 (CDKN1A) and the recruitment of the repressive DREAM complex to the APOBEC3B gene promoter. Clearly, the loss of p53 through mutation would be expected to cause elevated APOBEC3B expression and cytosine deaminase activity in cancer cells, thereby potentiating a mechanism for cancer genome mutation.

The unique nature of the MCF7 Luc  $\Delta$ p53 MA2 line and the matched wild-type MCF7 line has since led to their use in other studies of p53 in breast cancer cells. This includes the recent work carried out in collaboration with the group of Dr. David Meek, University of Dundee, where the lines were used to investigate the response of cancer cell types to Polo-like Kinase 1 (PLK1) inhibitors, and showed that, paradoxically, functional p53 reduces the sensitivity to these by allowing centrosome separation to occur during mitosis (Smith *et al.*, 2018).

While conducting this work, Menendez *et al.*, 2017 published a paper investigating p53 regulation of APOBEC3 gene expression, and also identified APOBEC3B as a p53 repressed gene and APOBEC3H as a markedly p53 up-regulated gene, and clearly the most highly p53 regulated APOBEC3 gene. Together with our findings with the MCF7 Luc  $\Delta$ p53 MA2 line,

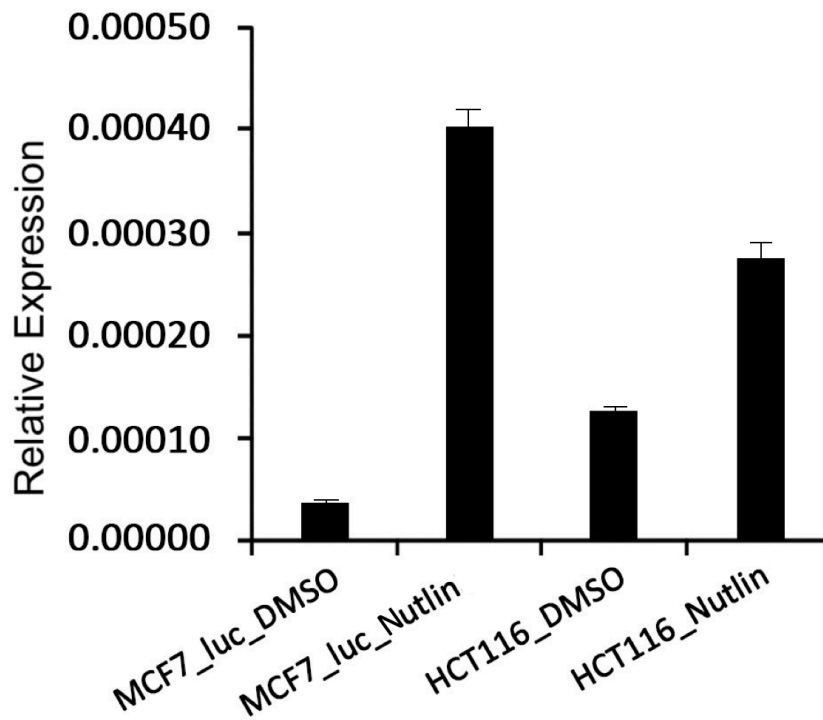


this raises the important question of the significance of the p53 regulation of APOBEC3H, and how this might be related to the down-regulation of APOBEC3B. In order to investigate this, we generated an APOBEC3H knockout line which has been characterized by RNAseq.

## **4 - Results 2: Generation of an APOBEC3H gene knockout in the HCT116 human cancer cell line**

### **4.1 - APOBEC3H knockout in the HCT116 cell line**

In order to further understand the link between p53 and APOBEC3H, we decided to knockout APOBEC3H in a cell line in which it is constitutively expressed APOBEC3H. To simplify analysis of the APOBEC3H knockout, as such an approach would not need the use of p53 activators such as Nutlin. Previous and current work in the group (Figure 4.1) had evaluated APOBEC3H expression in a broad selection of cancer cell lines and had identified a low level of constitutive APOBEC3H expression in the colon cancer cell line HCT116 (Brattain *et al.*, 1981; Periyasamy *et al.*, 2015; Periyasamy *et al.*, 2017). This line express wild-type p53 and has been widely used in investigations of p53 (Liu and Bodmer, 2006; Mashima *et al.*, 2005), where p53 knockout derivatives have been previously made (Bunz *et al.*, 1998). Furthermore, a previous study on APOBEC3 and p53 regulation used HCT116 (Menendez *et al.*, 2017).



**Figure 4.1. APOBEC3H gene expression in MCF7 Luc and HCT116 cell lines**

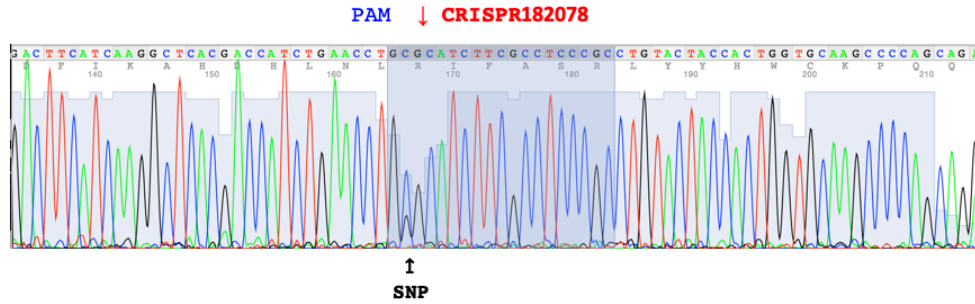
For qRT-PCR analysis, MCF7 Luc wild-type (MCF7 Luc) and HCT116 wild-type cells (HCT116) were grown in media supplemented with 10 $\mu$ M Nutlin, or vehicle (DMSO), and RNA prepared after 24 h. The RNA was subsequently used to make cDNA, and this used in qPCR. Gene expression is shown normalised with respect to GAPDH, house-keeping gene. This shows Nutlin induced APOBEC3H gene expression is comparable between the MCF7 Luc and HCT116 lines. However, basal APOBEC3H gene expression is considerably higher in HCT116.

## 4.2 - CRISPR-Cas9 mediated knockout of APOBEC3H by indel mutation.

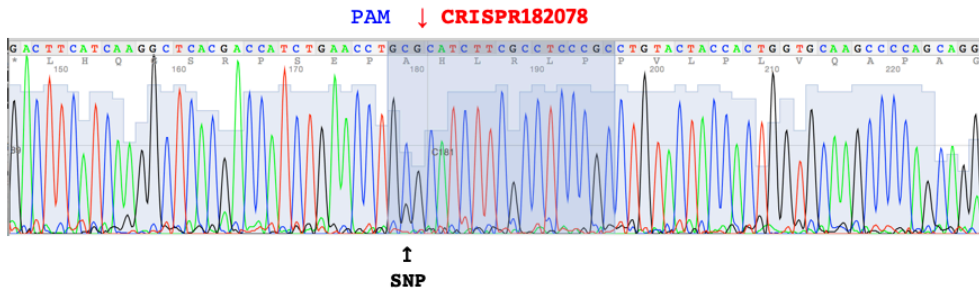
The APOBEC3H gene is comprised of five exons, with exons 2-4 being coding. Of these, exon 3 is the largest, encoding 44% of the protein, including the single cytidine deaminase domain (Dang *et al.*, 2008). CRISPR database (Mali *et al.*, 2013) used to identify TP53 Exon3 CRISPR, did not contain a pair of CRISPRs that could be used to similarly delete within coding exons of APOBEC3H. However, the database did contain a single CRISPR, CRISPR182078, which was potentially suitable for making indel mutations in APOBEC3H exon 3. Hence the expression plasmid for this was made and used in an initial attempt to make an APOBEC3H knockout.

CRISPR182078 was successfully cloned into the U6-CRISPR sgRNA plasmid expression vector, and this was nucleofected into HCT116 cells using conditions optimised for this cell line. Following this, genomic DNA was made from nucleofected cells, and this was used in a sequence analysis of DNA encoding the CRISPR182078 target sequence in APOBEC3H exon3. We have previously used this approach to identify active CRISPRs, since indel damage caused by their expression leads to heterogeneity in the sequence trace, precisely initiated at the CRISPR cleavage site (Harrod *et al.*, 2017). As shown in Figure 4.2, the comparison of the sequence chromatogram between untransfected and CRISPR182078- Cas9 co-transfected cells showed little evidence of extensive CRISPR activity. A second nucleofection experiment confirmed this conclusion. However, it was also noted in these analyses that the HCT116 line is heterozygous for a Single Nucleotide Polymorphism (SNP) which is encoded in the CRISPR target sequence, in close proximity to the cleavage site. As this mismatch would be expected to inhibit CRISPR182078- Cas9 cutting on the mismatched allele, the use of CRISPR182078 to obtain an APOBEC3H knockout in HCT116 cells was not explored further.

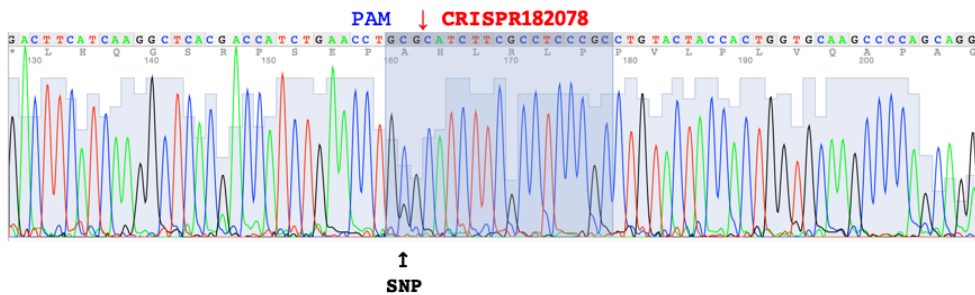
**WT HCT116**



**HCT 116: CRISPR 182078-1 transfected**



**HCT 116: CRISPR 78-2 transfected**



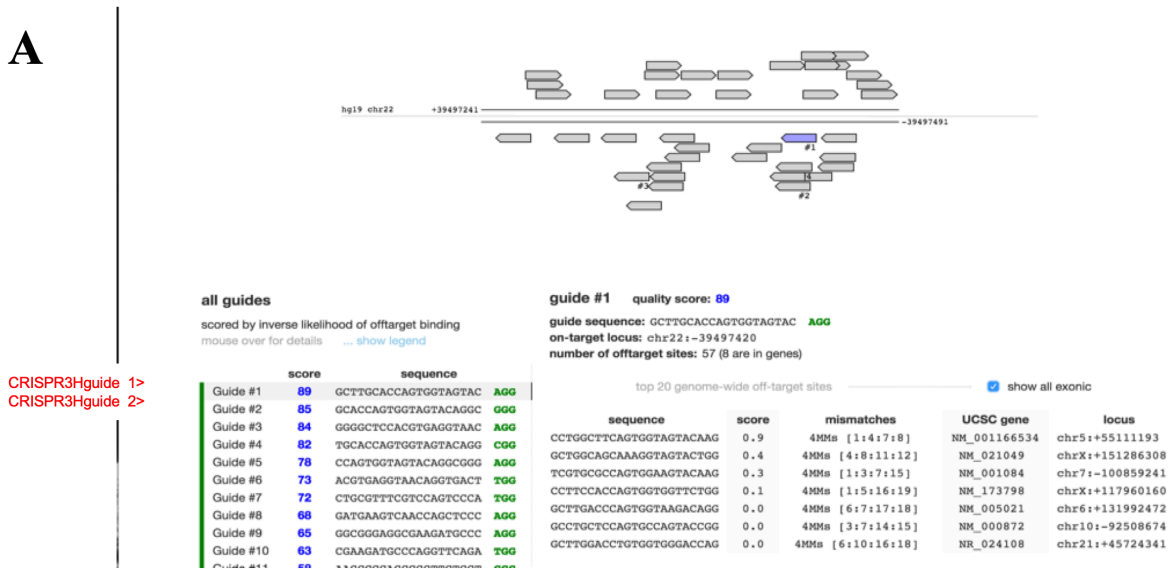
**Figure 4.2. APOBEC3H targeted gene editing using CRIPSR 182078**

The CRIPSR 182078 target sequence is located in exon3 of the APOBEC3H gene. CRIPSR 182078- Cas9 co-transfected HCT116 cells were used to prepare genomic DNA and used in a PCR assay with primers SA4416 (gactcctggcctctctcttc) -SA4417 (gcactttataactgcaaagcc), which generate a 355nt product. Sequencing of the resulting product shows little CRIPSR 182078 mediated indel formation, as compared to untransfected HCT116 cells (WT HCT116). The CRISPR is probably not very active, as no evidence of indel formation can be seen at the site of CRISPR 182078 directed cleavage (↓). The sequence differences on the variant allele (SNP) are found in the target sequence for CRISPR 182078, so are likely to prevent this from cutting the variant allele.

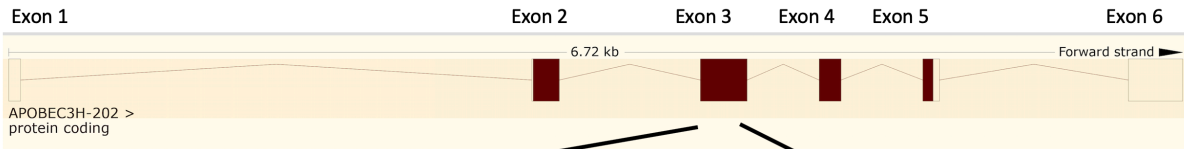
### 4.3 - Design of CRISPRs for making DNA deletions of APOBEC3H exon3

As our approach of making a p53 knockout cell line by using a pair of closely spaced CRISPRs to delete part of the p53 gene was successful, a similar strategy was employed for APOBEC3H. The on-line CRISPR design tool of the Zhang group at MIT (CRISPR.mit.edu) was used to design CRISPRs for APOBEC 3H exon 3. Briefly, this tool presents a ranked list of all possible guides in the query sequence, which is limited to 200nt and is ordered by the faithfulness of on-target activity calculated as 100% minus a weighted sum of off-target hit-scores in the entered specific genome. Guides having an aggregate score of greater than 50% are highlighted by green and are considered to be specific. As APOBEC3H exon3 is 268nt in size, the region was analysed as two overlapping sequences of 200nt. In addition to ranking, the resulting guides were also considered only if they started with a guanine (G), as this was required for transcription initiation by the U6 gene promoter. This analysis led to the design of two CRISPR guide RNAs; RNA 1 and 3, which have high aggregate scores (89 and 84) and are separated by 100nt in APOBEC3H exon3. Lastly, neither of these CRISPRs was found to include SNPs in HCT116, as judged by sequencing of APOBEC3H Exon3 in wild-type cells. The location of these CRISPRs in APOBEC3H exon3 is shown in Figure 4.3A and the identity of the predicted APOBEC3H exon3 deletion and encoded protein shown in Figure 4.3B. As can be seen, the CRISPR deleted allele can encode a truncated 129 aa protein, the first 78 aa of which can be aligned to Apobec3H protein. Further, the protein encoded by the APOBEC3H CRISPR exon3 deleted allele also lacks a complete cytidine deaminase domain. Taken together, these analyses indicate that the resulting mutant APOBEC3H protein would be inactive, even if this were produced.

**A**



CRISPR3Hguide 1>  
CRISPR3Hguide 2>



```

AAAAAGTCCATGCAGAAATTTGCTTTATTAACGAGATCAAGTCCATGGGACTGGACGAAACGCAGTCTACCAAGTCACCTGTACCTCAGTGGAGCC
CCTGCTCCTCCTGTGCCTGGGAGCTGGTTGACTTCATCAAGGCTCACGACCATCTGAACTGGGCATCTTCGCCTCCGCCTACTACCACTGGTGCAA
CCCCAGCAGAAAGGGCTGCGGCTTCTGTGTGGATCCAGGTCCCGGTGGAGGTCATGGGCTCCACAG
G C
  
```

CRISPR target sequences in red; PAM sequences in blue; HCT116 variant nucleotides (SNPs) in green

**B**

**APOBEC3H Wild-Type**

```

ATG GCT CTG TTA ACA GCC GAA ACA TTC CGC TTA CAG TTT AAC AAC AAG CGC CGC CTC AGA AGG CCT TAC TAC CCG AGG AAG GCC CTC TTG
M A L L T A E T F R L Q F N N K R R L R R P Y Y P R K A L L
TGT TAC CAG CTG ACG CCG CAG AAT GGC TCC ACG CCC ACG AGA GGC TAC TTT GAA AAC AAG aaa aag tgc cat gca gaa att tgc ttt att
C Y Q L T P Q N G S T P T R G Y F E N K K K C H A E I C F I
aac gag atc aag tcc atg gga ctg gac gaa acg cag tgc tac caa gtc acc tgt tac ctg acg tgg agc ccc tgc tcc tcc tgt gcc tgg
N E I K S M G L D E T Q C Y Q V T C Y L T W S P C S S C A W
gag ctg gtt gac ttc atc aag gct cac gac cat ctg aac ctg ggc atc ttc gcc tcc cgc ctg tac tac cac tgg tgc aag ccc cag cag
E L V D F I K A H D H L N L G I F A S R L Y Y H W C K P Q Q
aag ggg ctg cgg ctt ctg tgt gga tcc cag gtc cgg gtg gag gtc atg ggc ttc cca gag TTT GCT GAC TGC TGG GAA AAC TTT GTG GAC
K G L R L L C G S Q V P V E V M G F P E F A D C W E N F V D
CAC GAG AAA CCG CTT TCC TTC AAC CCC TAT AAG ATG TTA GAG GAG CTA GAT AAA AAC AGT CGA GCC ATA AAG CGA CCG CTT GAG AGG ATA
H E K P L S F N P Y K M L E E L D K N S R A I K R R R L E R I
AAG att cca ggg gta cgt gcg cag ggt cgt tac atg gat ata ttg tgt gat gct gag gtc tga
K I P G V R A Q G R Y M D I L C D A E V *
  
```

**APOBEC3H CRISPR Guide1-2 deleted**

```

ATG GCT CTG TTA ACA GCC GAA ACA TTC CGC TTA CAG TTT AAC AAC AAG CGC CGC CTC AGA AGG CCT TAC TAC CCG AGG AAG GCC CTC TTG
M A L L T A E T F R L Q F N N K R R L R R P Y Y P R K A L L
TGT TAC CAG CTG ACG CCG CAG AAT GGC TCC ACG CCC ACG AGA GGC TAC TTT GAA AAC AAG aaa aag tgc cat gca gaa att tgc ttt att
C Y Q L T P Q N G S T P T R G Y F E N K K K C H A E I C F I
aac gag atc aag tcc atg gga ctg gac gaa acg cag tgc tac caa gtc acc tgt tac ctg acg tgg agc ccc tgc tcc tcc tgt gcc tgg
N E I K S M G L D E T Q C Y Q V T C Y L T W S P C S S C A W
gag ctg gtt gac ttc atc aag gct cac gac cat ctg aac ctg ggc atc ttc gcc tcc cgc ctg tac tac cac tgg tgc aag ccc cag cag
E L V D F I K A H D H L N L G I F A S R L Y Y H W C K P Q Q
aag ggg ctg cgg ctt ctg tgt gga tcc cag gtc cgg gtg gag gtc atg ggc ttc cca gag TTT GCT GAC TGC TGG GAA AAC TTT GTG GAC
K G L R L L C G S Q V P V E V M G F P E F A D C W E N F V D
CAC GAG AAA CCG CTT TCC TTC AAC CCC TAT AAG ATG TTA GAG GAG CTA GAT AAA AAC AGT CGA GCC ATA AAG CGA CCG CTT GAG AGG ATA
H E K P L S F N P Y K M L E E L D K N S R A I K R R R L E R I
AAG att cca ggg gta cgt gcg cag ggt cgt tac atg gat ata ttg tgt gat gct gag gtc tga
ggg tac gtg cgc agg gtc gtt aca tgg ata
  
```

### Figure 4.3. Design of CRISPR guide RNAs for APOBEC3H exon 3

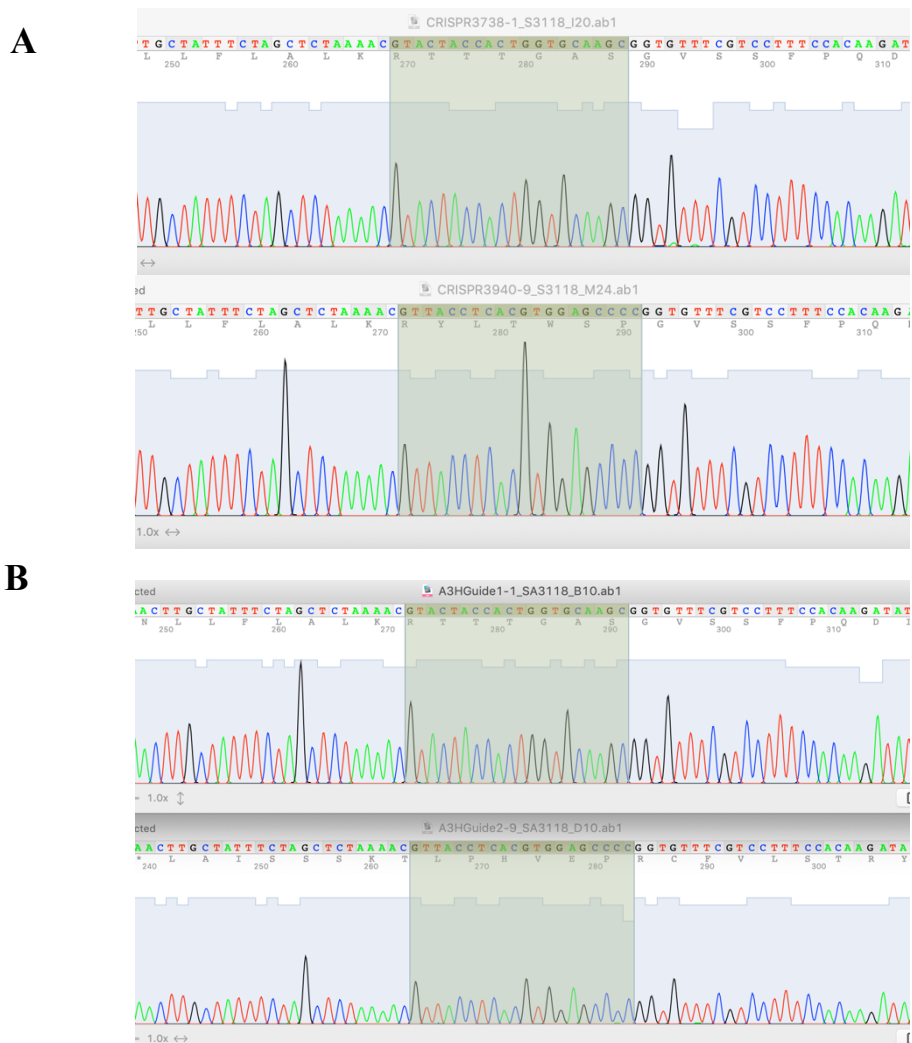
**A.** The on-line CRISPR design of the Zhang group at MIT (CRISPR.mit.edu) was used to design CRISPR guide RNAs for APOBEC 3H exon 3. This tool ranks all potential CRISPR guide RNAs in a 200nt target sequence using an aggregate score, that takes into account sequence characteristics and potential off-target matches. A ranking score of greater than 50% are coloured green and are considered to be specific. Guide RNA 1 and 3 have high scores (89 and 84) and are separated by 100nt at the coding sequences of A3H (exon 3). Neither of these CRISPRs lie on SNPs in HCT116. CRISPR3Hguide 1. PAM in upper case (blue) gcttgaccagtgtagtacAGG, CRISPR3Hguide 2. PAM in upper case (blue) ggggctccacgtgaggtaacAGG. **B.** Translation of APOBEC3H WT and CRISPR Guide 1-2 deleted open reading frames. Exon sequences are shown in alternating DNA case, with the encoded protein sequence shown below the coding region in the single letter amino acid code, for both the wild-type and engineered APOBEC3H mutant protein. The out of frame mutant sequence resulting from the CRISPR mediated deletion in APOBEC3H exon 3 is shown in magenta.



#### 4.4 – Establishment of HCT116 APOBEC3H knockout lines

The overall strategy used to make HCT116 APOBEC3H knockout lines was similar to the successful strategy used to make the MCF7 Luc  $\Delta$ p53 MA2 line, in that two CRISPR gRNAs were used to direct Cas9 to a specific APOBEC3H exon, in this case, exon 3, so as to obtain interstitial DNA deletions within the exon. For this, the two CRISPR.mit.edu algorithm designed guide RNAs, Guide 1 and Guide 3 were cloned into the previously used U6 promoter sgRNA expression plasmid (Figure 4.4). These were used to nucleofect HCT116 cells, either individually with the Cas9 expression plasmid or in combination. As can be seen in Figure 4.5 this resulted in efficient transfection (>80%), as judged by the expression of a GFP control expression plasmid used in parallel to the co-transfection of the two CRISPR sgRNA expression plasmids for Guides 1 and 3, and the Cas9 expression plasmid.

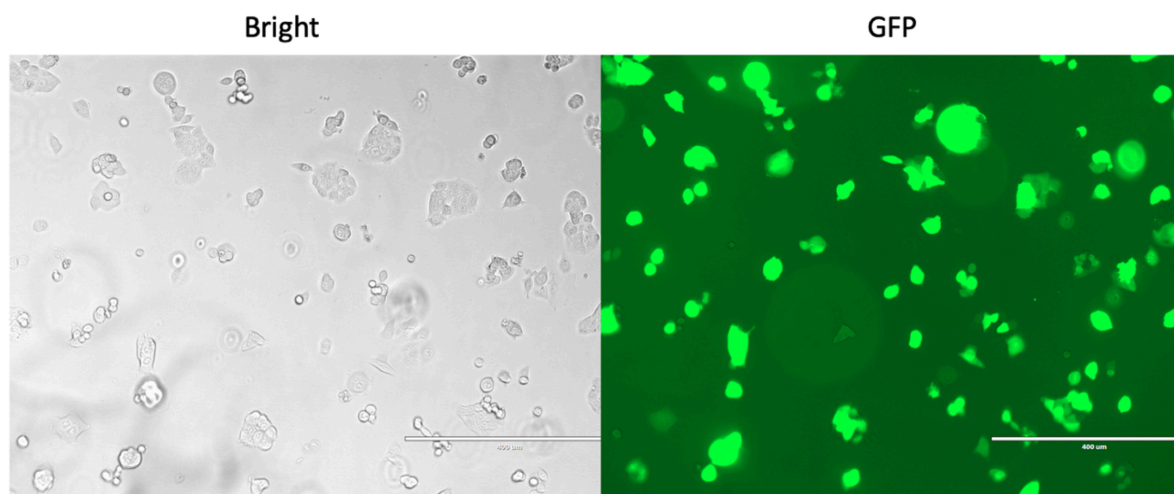
Following nucleofection, cells were seeded at clonal densities on 10cm dishes, and also in T25 flasks, to obtain confluent cultures to make DNA for PCR analysis. In this case, a PCR assay was developed for APOBEC3H exon 3, using the primers SA4416APOBEC3HEX3f1 (gactcctggcctctctcttc) SA4417APOBEC3HEX3r1 (gcactcttataactgcaaagcc). These generate a product size of 355nt from wild-type genomic DNA template. Digestion of APOBEC3H exon 3 using Cas9 programmed with CRISPR Guide1 and CRISPR Guide 3 would result in the excision of 100nt, and the generation of a 255nt PCR product.



**Figure 4.4. Cloning of CRISPR guide RNAs for APOBEC 3H exon 3 into the U6 CRISPR sgRNA vector**

**A.** Following Gibson cloning, mini- preparation for both CRISPRs Guide13H 3738 and Guide23H 3940 were carried out and 10 clones each prepared. Based on sgRNA transcription start sequence (i.e. Antisense for CRISPR Guide13H 3738 GTACTACCACTGGTGCAAGC and Antisense for CRISPR Guide23H 3940 GTTACCTCACGTGGAGCCCC). Cloning was effective and 7 out of 10 clones were positive for the sequence in CRISPR Guide13H 3738 and 6 out of 10 in CRISPR Guide23H 3940. In CRISPR Guide13H 3738, Clones 1 (conc. 34.4 ng/ul) and in CRISPR Guide23H 3940 clone 9 (49.4 ng/ul) were chosen to proceed maxi-prep.

**B.** Maxi-prep sequences of both CRISPR's Guide13H 3738 and Guide23H 3940. For CRISPR Guide13H 3738 and CRISPR Guide23H 3940 one plasmid was prepared for each at higher concentrations. Both minipreps and maxipreps were sequenced with 30 uM SA3118 primers.



**Figure 4.5. Optimised plasmid nucleofection of HCT116 cells**

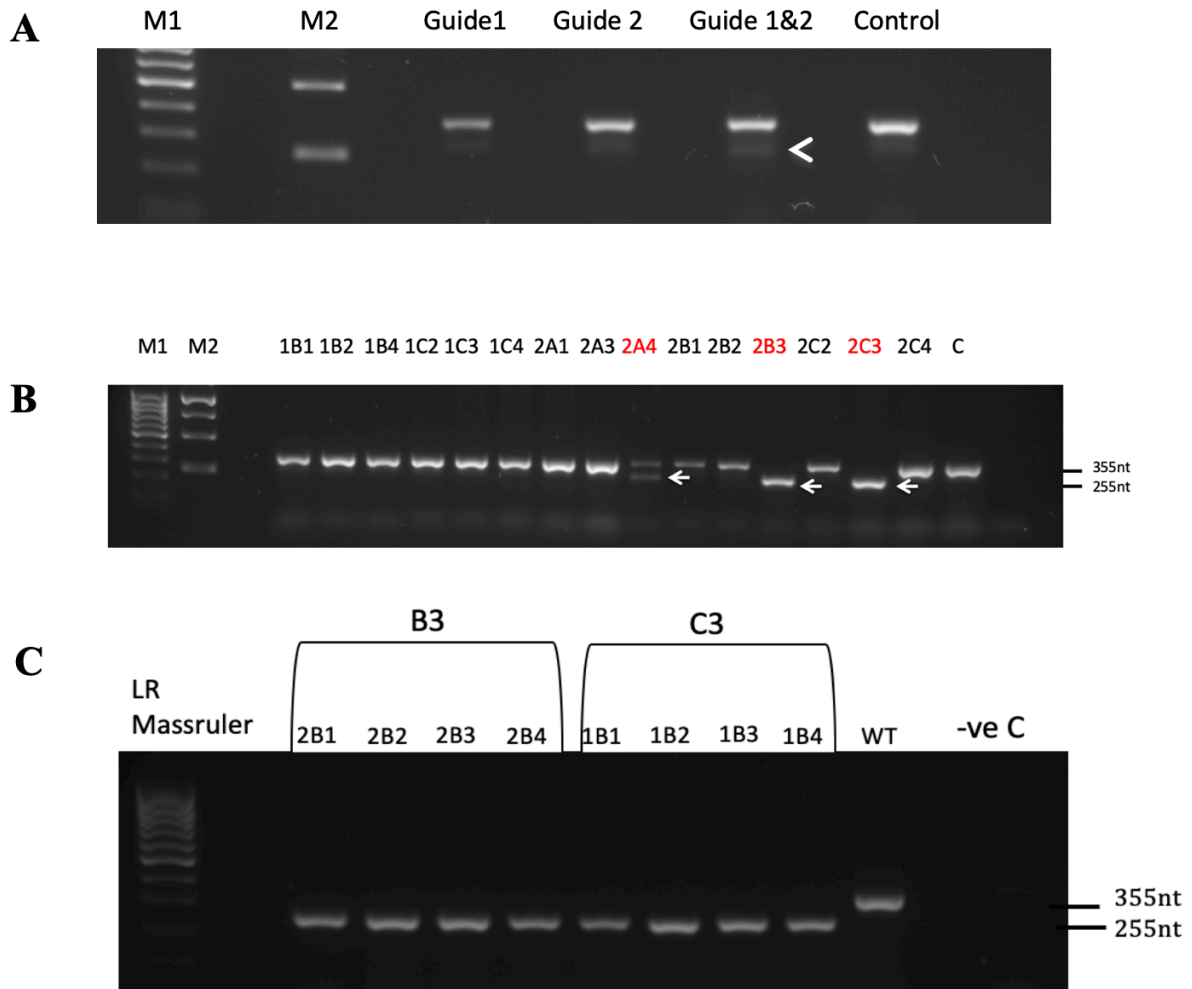
CRISPR mediated deletion of sequences in the third coding exon of APOBEC3H was carried out using an Amaxa Type II nucleofector and the HCT116 transfection protocol recommended by the manufacturer. HCT116 was transfected with the pMAX GFP, and visualised 48 h following transfection in bright field and for GFP expression (x10 magnification). This analysis shows nucleofection efficiencies  $> 70\%$ , and in line with the nucleofection efficiencies previously found for MCF7 Luc.

Figure 4.6A shows the PCR analysis of genomic DNA from HCT116 cells transfected with Cas9 expression plasmid, together with CRISPR sgRNA expression plasmids for Guide 1, Guide 3 and the combination of Guide 1 and Guide 3. In all cases, a dominant, full-length, 355nt, wild-type PCR product was seen with all combinations of CRISPR guide RNAs, and with the control untransfected cells. A second, much weaker, slightly smaller product was also seen with all samples, which was presumably a non-specific PCR product. However, in the Cas9/Guide1/ Guide3 three-way co-transfection, there is a stronger PCR product of the expected size of 255nt, indicative of DNA deletion between the genomic target sequences for the two CRISPR gRNAs. However, as this was weak, when compared to the full-length product from the same sample, this suggests the resulting deletion occurs relatively inefficiently. There are several reasons why this might be, including low intrinsic activity for, one or both of the CRISPR guide RNAs used, and would require further investigation. Nevertheless, as the minor 255nt product was detected, analysis of clones seeded following nucleofection was continued.

#### **4.5 - Isolation and initial characterisation of HCT116 APOBEC3H knockout lines**

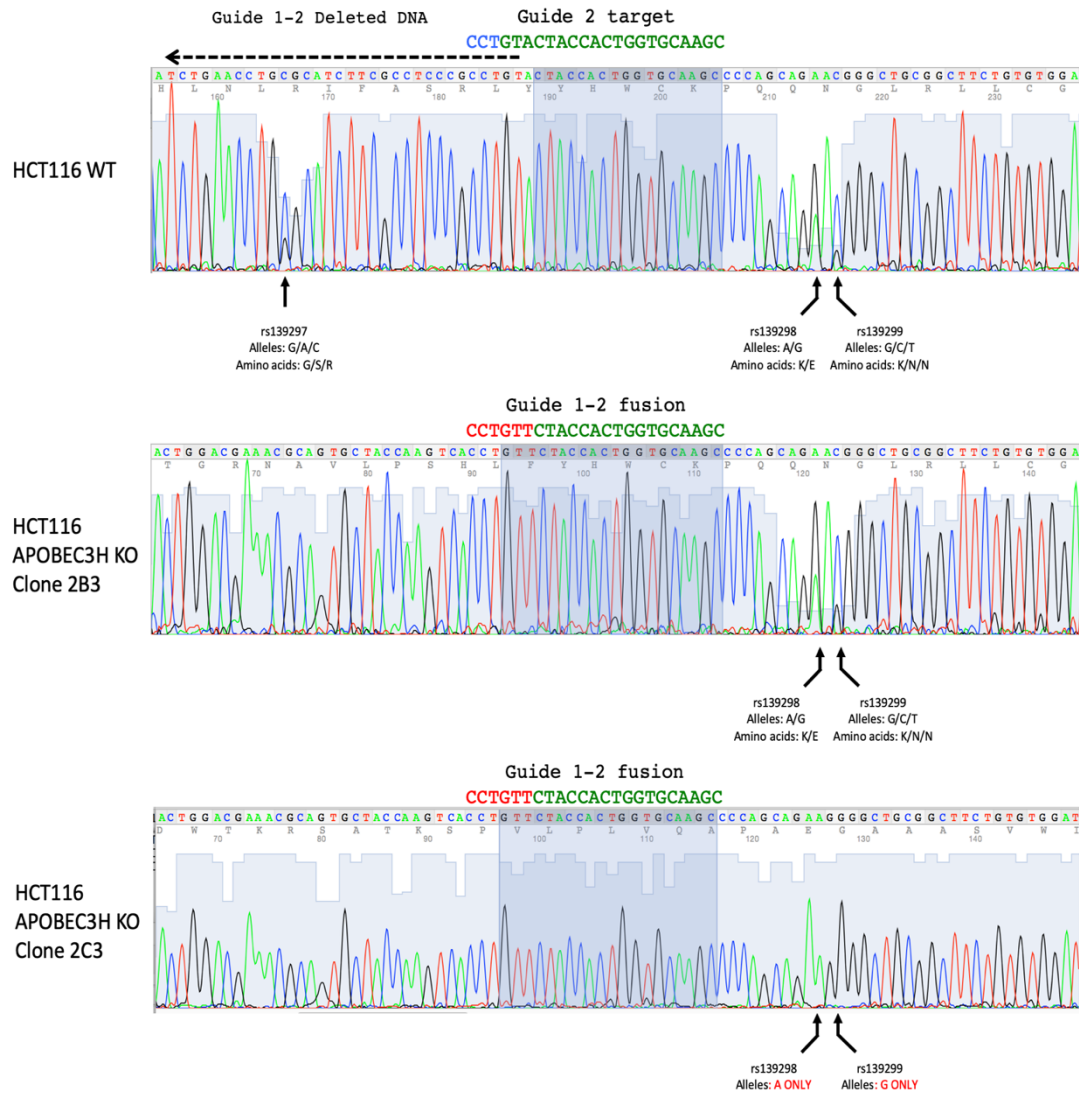
A total of 48 clones were isolated following the Cas9/Guide1/Guide3 three-way co-transfection of HCT116 cells. All but three lines showed the wild-type PCR product, although the possibility of indel mutation in these was not investigated further (Figure 4.6B) Of the remainder one clone, 2A4, showed a heterozygote genotype, with an engineered APOBEC3H deletion on one allele. DNA sequence analysis of this confirmed this genotype, showing the presence of a mutant, deleted and wild-type allele, but additionally a second allele with a wild-type sequence (Figure 4.6B). The remaining two clones, 2B3 and 2C3, produced a single 255nt

sized PCR product, indicative of DNA deletion on both alleles. The SA4416APOBEC3HEX3f1 - SA4417APOBEC3HEX3r1 PCR product for these two clones was subjected to DNA sequence analysis, as shown in Figure 4.7. From this, it is clear that both lines are homozygous for the predicted, engineered DNA deletion between CRISPR sgRNAs Guide 1 and Guide 3. However, as can be seen from the sequence chromatogram for the SA4416APOBEC3HEX3f1 - SA4417APOBEC3HEX3r1 PCR product from wild-type HCT116 cells (Figure 4.7), the line is heterozygous for two Single Nucleotide Polymorphisms (SNPs), namely SNP rs139928 (alleles A/G) and rs139299 (alleles G/C/T). While this is also the case for the 2B3 clone, for the 2C3 line these SNPs are homozygous, with SNP rs139928 being the A allele and rs139299 being the G allele, suggesting a haplotype of AG. Taken together, this data shows that the two clones, 2B3 and 2C3, are independent, as they have different genotypes. The analysis also suggests that the derivation of the 2C3 is more complicated than just the deletion of two alleles, with the possibility that the deletion on one allele may extend beyond the boundaries of the CRISPR target sites for sgRNAs 1 and 3. Taken together, the frequency of engineered, CRISPR Guide 1 -Guide 3 deleted alleles in this experiment was observed to be around 5.5% (i.e. five deleted and 91 wild-type alleles). While the occurrence of homozygous deletions was similar to that observed for the TP53 work described earlier, it was clear that the occurrence of heterozygote alleles was lower, an observation that, as stated previously may reflect the relative activity of the CRISPR sgRNAs used.



**Figure 4.6. PCR characterisation of HCT116 clones following CRISPR guide13H 3738/ guide23H 3940/ Cas9 co-transfection**

**A.** Transfected pool of cells with APOBEC 3H exon 3 Guide13H 3738 and Guide23H 3940 CRISPRs were subsequently cloned and used to make genomic DNA. PCR analysis of exon 3 was carried out on this DNA using the primers SA4416APOBEC3HEX3f1 (gactcctggcctctctcttc) SA4417APOBEC3HEX3r1 (gcactcttataactgcaaagcc) that generate a product size of 355 bp. Digestion of Exon 3 by CRISPR Guide13H 3738 and CRISPR Guide23H 3940 results in the excision of 100nt, and the generation of a 255nt PCR product. Cells were grown as a transfected pool and at clonal densities for clone selection. **B.** 15 clones growing from transfection. These have been analyzed by PCR and three deletion clones sent for sequencing. Clones 2B3, 2C3 are homozygous deleted for the CRISPR targeted region of APOBEC3H exon 3. DNA size markers are shown in the first two lanes; M1 Low Range Mass Rule, M2 1 Kb ladder. HCT116 shows the Exon 3 PCR product from untransfected, control cells and C the negative, no template control. All clones are HCT116 derived clones. Clones 2B3 and 2C3 both are Homozygous for deletion and clone 2A4 is Heterozygote for deletion. **C.** Further characterisation for the selected clones, 2B3 and 2C3. Subcloning the pools to grow as single colony and expanded, DNA extracted and PCR, sequenced, all have been done to ensure the purity of the deletion lines.

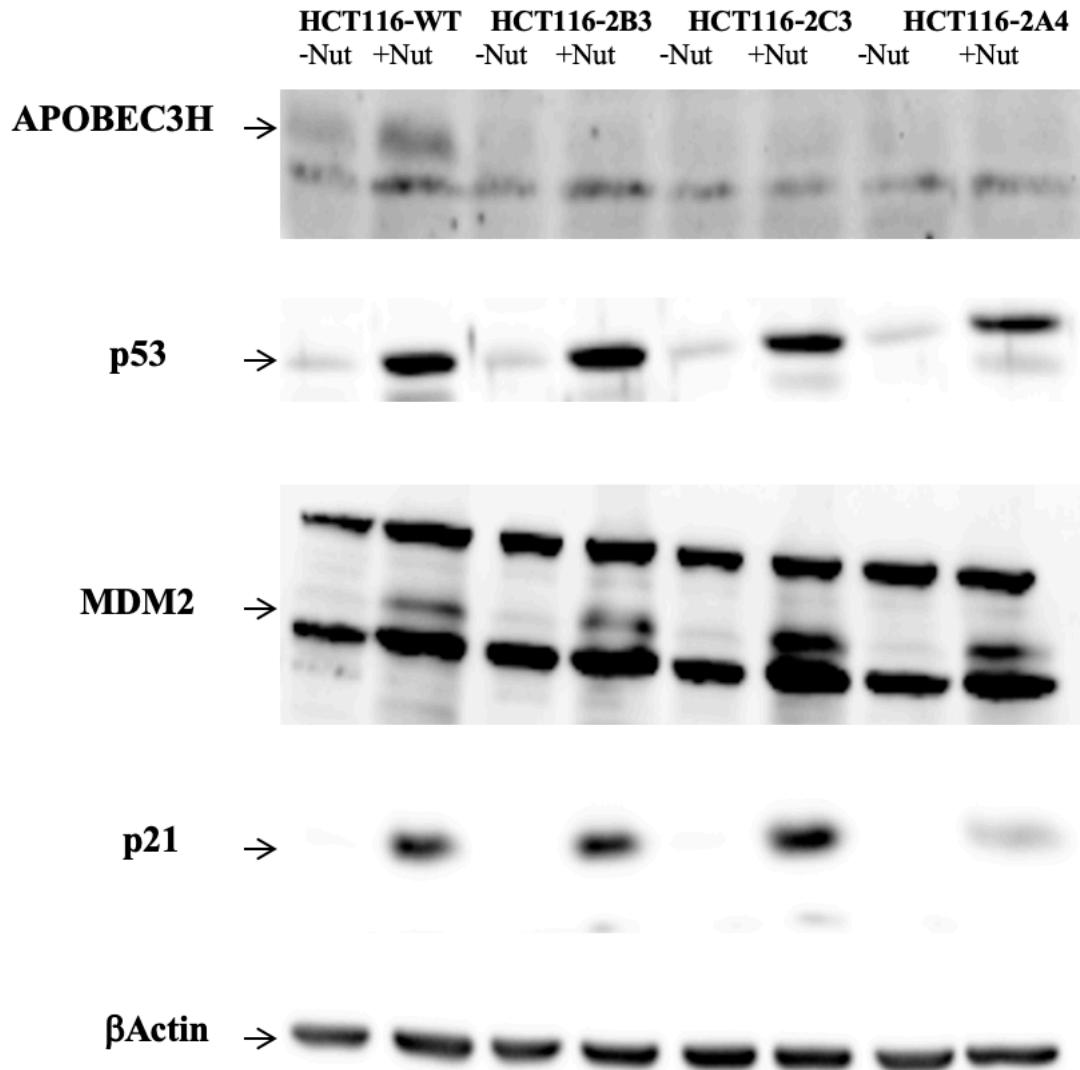


**Figure 4.7. DNA sequence characterisation of HCT116 APOBEC3H knockout clones 2B3 and 2C3**

Genomic DNA from HCT116 wild-type and HCT116 APOBEC3H Clones 2B3 and 2C3 was used as template for PCR with primers SA4416APOBEC3HEX3f1 - SA4417APOBEC3HEX3r1, which amplified APOBEC3H Exon3, and the product sequenced. The PCR product for clone 2B3 shows the predicted sequence for the deletion of DNA between Guide 1 and Guide 2 (Guide 1 and 3 in the designing step) CRISPR target motifs. The clone 2B3 sequence also shows the line is heterozygous for the two SNPs rs139298 (A/G) and rs139299 (G/C), confirming the HCT116 origin of this line. Clone 2C3 is homozygous for the two SNPs rs139298 (A) and rs139299 (G) and shows this is an independent line from clone 2B3.

Western blotting analysis was carried out to characterise clones HCT116 APOBEC3H 2A4, 2B3 and 2C3 (Figure 4.8). Nutlin treatment was used to induce APOBEC3H. In the wild-type HCT116 cells, weak expression of APOBEC3H can be seen in the absence of Nutlin treatment. Following Nutlin treatment, this is induced and follows the pattern of Nutlin induction seen for the positive control proteins p53, MDM2 and p21. By contrast, there is a very little basal expression of APOBEC3H in each of the three APOBEC3H CRISPR Guide1-Guide3 targeted clones. Further, these lines show no Nutlin induced APOBEC3H expression while continuing to demonstrate Nutlin induction of p53 and upregulation of the p53 targets MDM2 and p21. This demonstrates that the three HCT116 clones 2A4, 2B3 and 2C3, represent putative APOBEC3H knockout lines. Further, this suggests the possibility of indel mutation on the second, wild-type sized PCR product allele of the 2A4 line. However, as the HCT116 2B3 and 2C3 lines represent closely matched, independent knockout lines, where the APOBEC3H genotype was confirmed, and where the out of frame APOBEC3H protein product can be predicted, we choose to pursue these two lines further.





**Figure 4.8. Western blot analysis of HCT116 APOBEC3H knockout clones 2B3, 2C3 and 2A**

The CRISPR engineered HCT116 APOBEC3H knockout clones 2B3, 2C3 and 2A4 were treated with (+Nut) and without (-Nut) Nutlin for 24 h and used to make protein lysates. These were analysed by Western blotting for expression of APOBEC3H, p53 and the p53 targets MDM2 and p21.

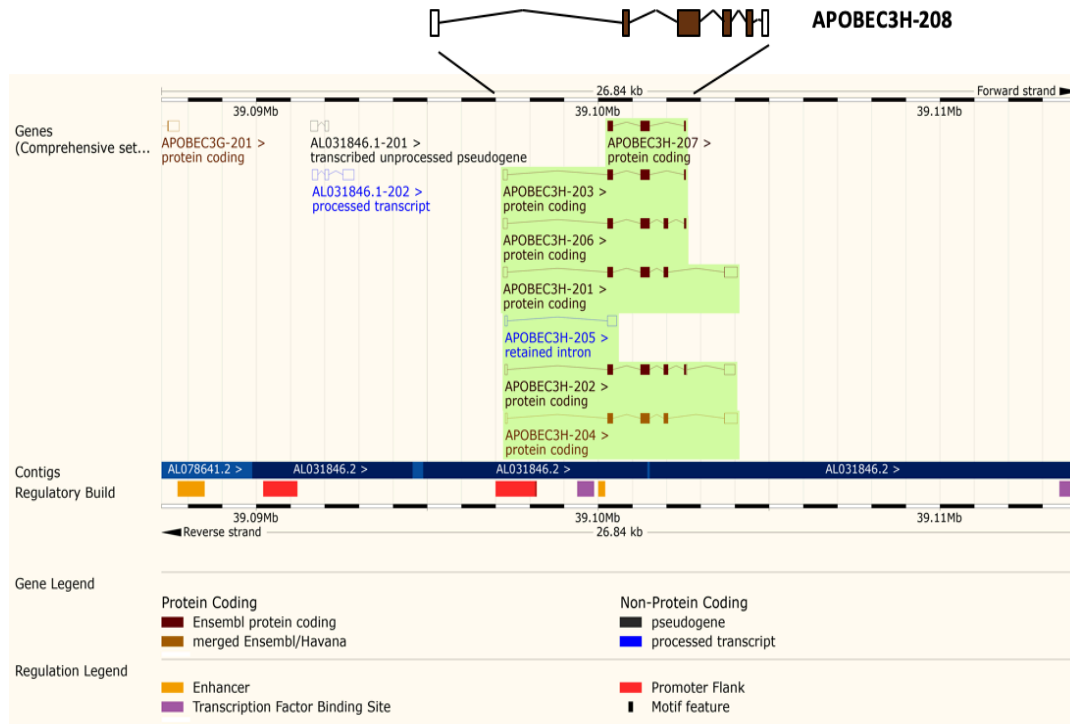
#### **4.6 - Characterisation of gene expression from the APOBEC3H locus in HCT116 cells and the HCT116- APOBEC3H exon3 deleted clones 2B3 and 2C3.**

Western blot analysis for APOBEC3H protein expression in the APOBEC3H exon 3 deleted clones 2B3 and 2C3 confirmed these did not express full-length APOBEC3H protein, a result that was corroborated by the genomic analyses of APOBEC3H exon3 in these lines. However, currently, there are seven alternative transcripts variants 201-207, that have been described for the APOBEC3H gene (Kersey *et al.*, 2016), of which six, 201-204, 206 and 207, are potentially protein-encoding, and all involve APOBEC3H exon3 (Figure 4.9 ). Of these, variants 202 and 206 have the most extended open reading frame and encode a 200aa protein that is considered to be full-length APOBEC3H (Refsland *et al.*, 2014).

In order to define APOBEC3H transcription in mutant and wild-type cells, RT-PCR analysis was carried out. This involved the use of a PCR assay using primers located in exon 2 (SA4446APOB3H RT-F2) and exon 4 (SA4447APOB3H RT-R2), which can detect variants 202, 203, 206 and 207. As can be seen in Figure 4.10, when used in end-point RT-PCR analysis, the SA4446APOB3H RT-F2 - SA4447APOB3H RT-R2 assay detected a single 275nt RT-PCR product, which contains the sequence junction arising from the CRISPR Guide 1-Guide 3 directed coding deletion. In the HCT116 APOBEC3H 2B3 line, this product was found to be strongly Nutlin regulated, while in the HCT116 APOBEC3H 2C3 line there appeared to be elevated basal expression in the absence of Nutlin, with some induction following Nutlin treatment.

By contrast to the single RT-PCR species seen for the APOBEC3H RNA expressed in the two exon 3 mutated lines, multiple RT-PCR products were seen in wild-type HCT116 cells, all of

which were Nutlin inducible. By size, two of these could be equated with the APOBEC3H variants 202/206 (RT-PCR product size of 500nt) and 203/207 (RT-PCR product size of 375 nt), with the variant 202/206 product being the strongest, and by inference, the most strongly Nutlin regulated species. In addition, the RT-PCR analysis identified a larger, Nutlin regulated species in the HCT116 wild-type cells. Further DNA sequencing of this RT-PCR product unexpectedly identified this as coming from a novel APOBEC3H transcript, which would be predicted to encode a 208aa APOBE3H variant, as shown in (Figure 4.10). Following the numbering of the other APOBEC3H variant transcripts, we have named this novel form Variant 208, and alignment of the predicted sequence for the variant transcript shows it is distinguished by the use of an alternative 3' exon (Figure 4.10). The significance of this new variant is unclear, although homology searches with the novel APOBEC3H Variant 208 exon suggest this is primate-specific and coming from a novel repetitive coding element.

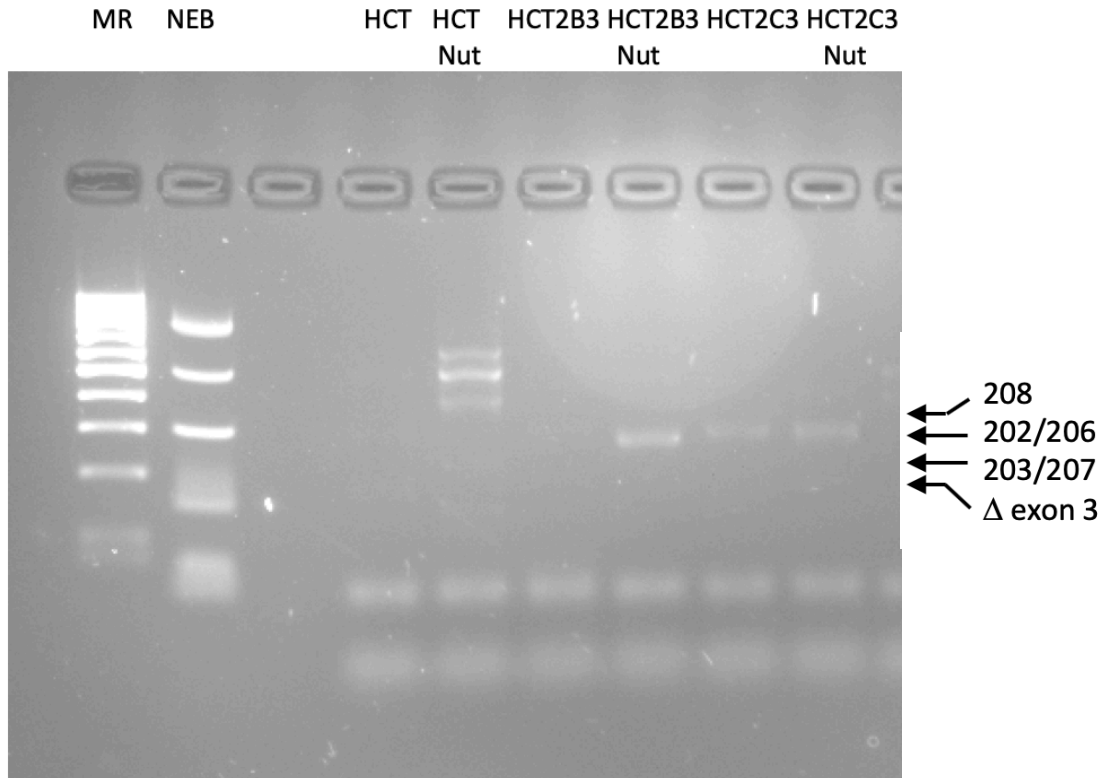


APOBEC3H variant	Encoded protein
201	182 aa protein
202	200 aa protein
203	154 aa protein
204	183 aa protein
205	No protein encoded
206	200 aa protein
207	154 aa protein
<b>208</b>	<b>208 aa protein</b>

Ensembl, ENSG00000100298

**Figure 4.9. APOBEC3H transcript variants**

There are seven known variant transcripts encoded by the A3H gene, variants 201-207. Six of these, 201-204, 206 and 207, are protein encoding, and all include APOBEC3H exon3. Of these, variations 202 and 206 have the previous longest ORF, and encode a 200aa protein that is considered as full length APOBEC3H. As shown in Figure 4.10, we have identified a new, minor variant form that we have named variant 208, that encodes a 200aa form of APOBEC3H.



```

V206 1 MALLTAETFRLOFNKRRLLRPYYPRKALLCYQLTPONGSTPTRGYFENKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDLNLFIFASRLYYHWCKPQKGLRL 125
      MALLTAETFRLOFNKRRLLRPYYPRKALLCYQLTPONGSTPTRGYFENKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDLNLFIFASRLYYHWCKPQKGLRL #
V208 1 MALLTAETFRLOFNKRRLLRPYYPRKALLCYQLTPONGSTPTRGYFENKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDLNLFIFASRLYYHWCKPQKGLRL 125
      * * 140 20 * * 160 40 * * 180 60 * * 200 80 * * 100 * * 120
V206 126 LCGSQVPVEVMGPFPEFADCWENFVDHEKPLSFNPKMLEELDKNRAIKRRLERIKIPGVRAQGRYMDILCDAEV*----- 201
      LCGSQVPVEVMGPFPEFADCWENFVDHEKPLSFNPKMLEELDKNRAIKRRLERIK V
V208 126 LCGSQVPVEVMGPFPEFADCWENFVDHEKPLSFNPKMLEELDKNRAIKRRLERIKTGSHSVVTQAGGQWCDLGSVQPLPPGFK* 209
      * * 140 * * 160 * * 180 * * 200

```

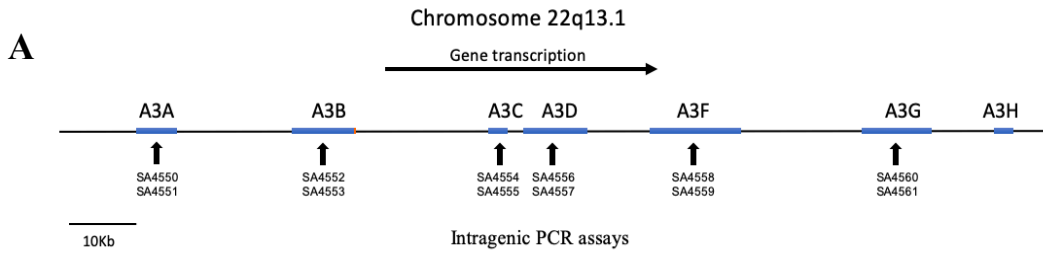
**Figure 4.10. APOBEC3H transcript variants PCR analysis**

To get further insight into the forms of APOBEC3H expressed in HCT116 following Nutlin treatment, Sequencing of the SA4446-SA4447 (which can detect variants 202, 203, 206, 207) End point RT-PCR product from Nutlin treated cells were used. This assay was able to identify a new A3H variant that encodes a 208 aa, the largest form of A3H so far defined (named 208). The variant uses a cryptic exon encoded in intron 4 (To be named Exon 4B). Exon 4B encodes a novel 27aa peptide with the sequence TGSHSVVTQAGGQWCDLGSVQPLPPGFK. Homology searches with this sequence of the non-redundant protein sequence database with BLAST-P (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) identifies this peptide as a novel repetitive coding element that is only seen in primate genomes.

#### **4.7 - Sub-cloning of the HCT116- APOBEC3H exon3 deleted clones 2B3 and 2C3 and characterisation of the APOBEC3 locus in the APOBEC3H knockout lines**

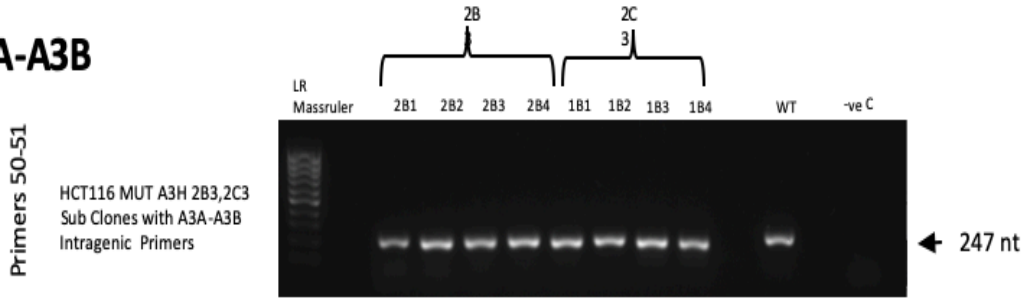
With a view to investigating the APOBEC3H exon3 deleted clones 2B3 and 2C3 in further detail, the lines were each sub-cloned, so as to ensure stocks were free of any contaminating wild-type cells. This exercise led to the isolation of four sub-clones for each line, which were initially verified by the use of the SA4416APOBEC3HEX3f1 - SA4417APOBEC3HEX3r1 genomic PCR assay, where it was seen that all eight clones had the APOBEC3H Exon 3 gene deletion (Figure 4.6C).

As described above, the PCR analysis of the APOBEC3H exon 3 region in the HCT116 2B3 and 2C3 clones showed that the two feature the intended engineered deletion in APOBEC3H exon3. However, SNP analysis suggested additional changes, possibly associated with more extensive DNA deletion in the 2C3 clone. Further, in an independent work carried out to knockout APOBEC3B, a single guide RNA was found to produce a complete deletion of the APOBEC3 locus (Dr Anup Singh, Imperial College, unpublished data). Hence, in order to evaluate the integrity of the APOBEC3 locus in the CRISPR engineered HCT116 lines, genomic PCR primers were designed to produce intragenic assays. As can be seen in Figure 4.11, these primers produced PCR products of the predicted expected size for all eight APOBEC3H Exon3 engineered sub-clones. The lack of detectable rearrangements and large-scale DNA deletions in the regions examined, suggested that in both sets of clones, the APOBEC3 locus was essentially intact.

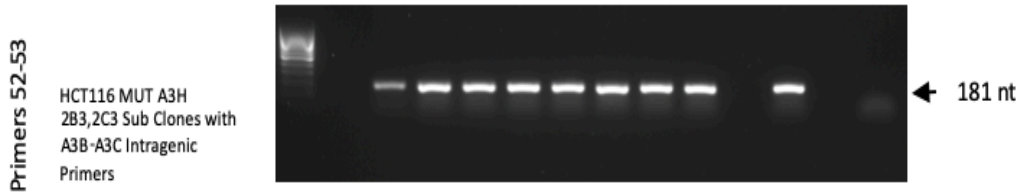


**B**

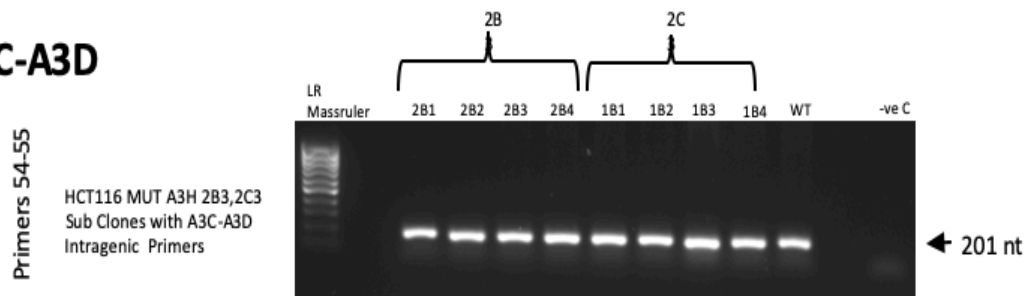
**A3A-A3B**



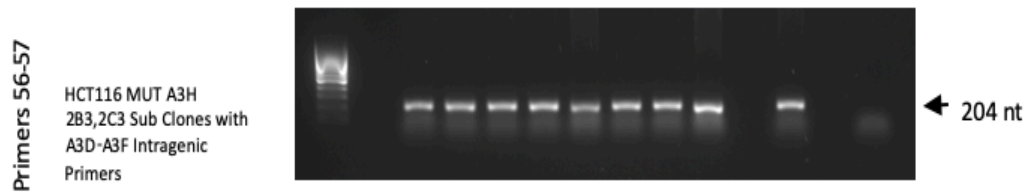
**A3B-A3C**

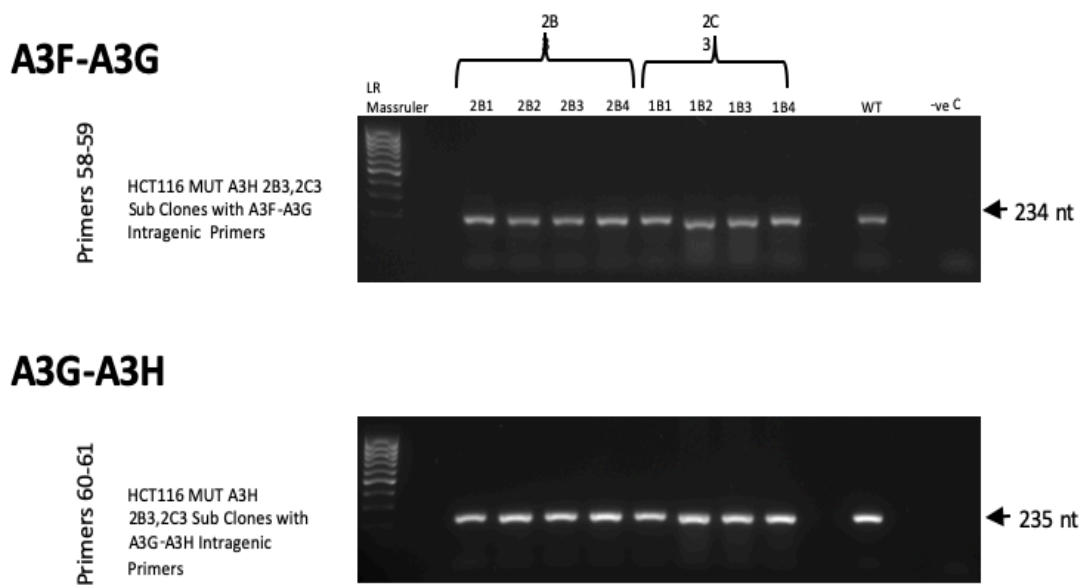


**A3C-A3D**



**A3D-A3F**





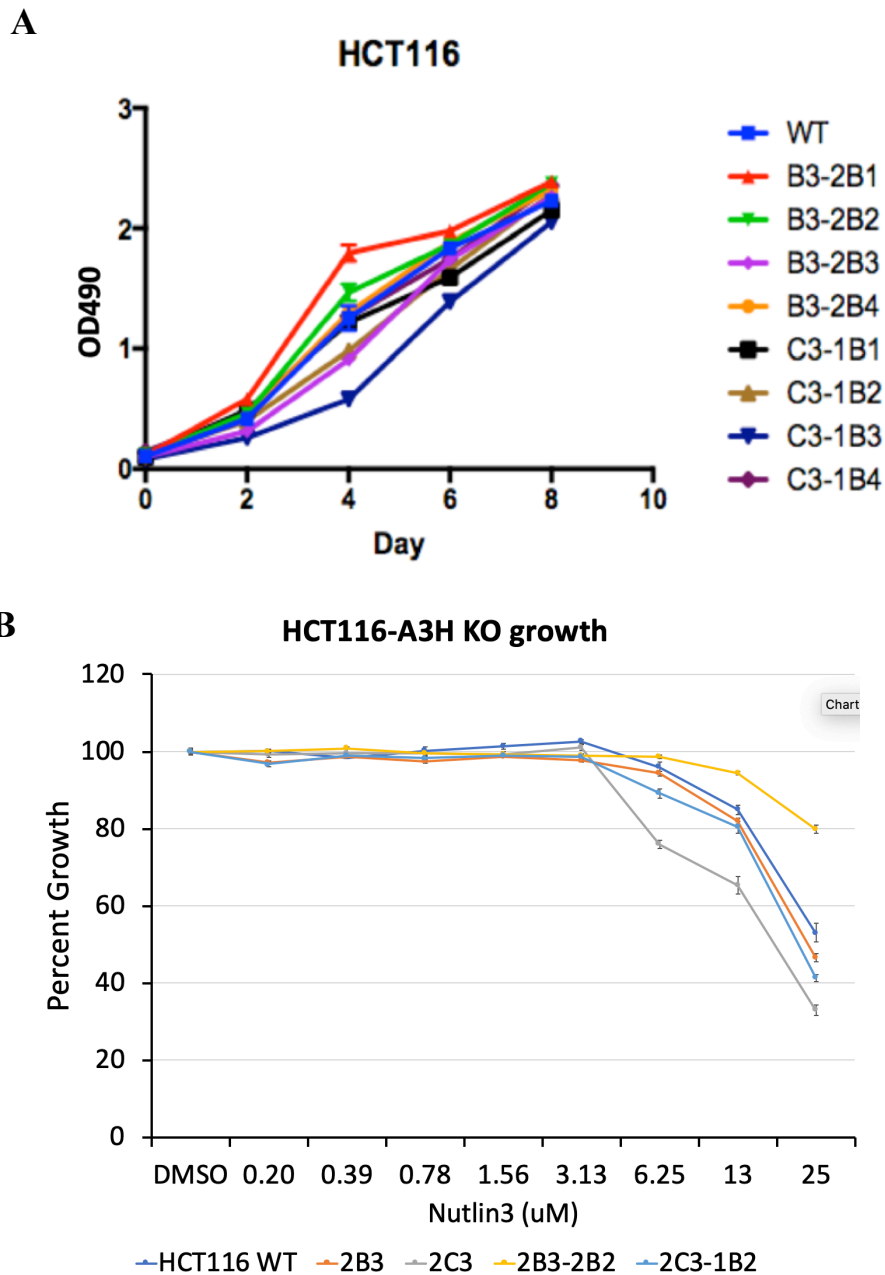
**Figure 4.11. Integrity of the APOBEC3 locus in the HCT116 and HCT116 APOBEC3H 2B3 and 2C3 knockout cells**

**A.** APOBEC3 Intrinsic PCR assays were designed and tested using HCT116 WT DNA, where all assays produce single PCR products of expected size. **B.** These intrinsic PCR assays used to assess the integrity of the APOBEC 3 locus in the CRISPR engineered knockout lines. The results show no evidence for gross –rearrangements, or extended DNA deletions within the region encoding APOBEC3A- 3F in the APOBEC3H knockout cells. **C.** The analysis also shows no evidence for gross–rearrangements, or extended DNA deletions within the region encoding APOBEC3A- 3H in the APOBEC3H knockout cells



#### **4.8 - Growth and sensitivity to Nutlin in the HCT116 2B3 and 2C3 subclones**

To further characterise the HCT116 APOBEC3H exon3 deleted cells, experiments were carried out to establish their growth characteristics. To do this, cells were seeded in 96 well dishes, and the growth followed in full medium using the SRB method. As can be seen in Figure 4.12A, the growth of the two sets of the HCT116 2B3 and 2C3 subclones is similar to that exhibited by HCT116 wild-type cells, with all cells exhibiting a doubling time of around 36 h. This result is not entirely unexpected, as, under normal conditions, the HCT116 cells express very little APOBEC3H. In order to see if growth in the presence of Nutlin is affected in these lines, growth was measured in a titration of Nutlin starting at 50 $\mu$ M, with two-fold dilutions over a range of eight dilutions. This allows the sensitivity to Nutlin to be seen, and the concentration of drug that causes a 50% inhibition of growth (GI50) to be determined. As can be seen in Figure 4.12B, for the parental knockout lines HCT116 2B3 and 2C3, and the two subclones (HCT116 2B3-2B2 and 2C3-1B2) there is no marked difference in Nutlin sensitivity, either between the different clones or between these and the HCT116 wild-type cells.



**Figure 4.12. Growth and Nutlin sensitivity of HCT116 wild-type and HCT116 APOBEC3H knockout cells**

**A.** Comparison under normal growth conditions shows that APOBEC3H deletion does not affect growth and cell doubling time. **B.** The sensitivity of HCT116 wild-type and APOBEC3H knockout clones to Nutlin was evaluated in a titration involving two-fold dilutions of the drug, starting at 50 $\mu$ M. This shows the IC50 for Nutlin in mutant and wild-type cells is very similar ( $\sim$ 25 $\mu$ M). Nutlin was subsequently used experimentally at 10 $\mu$ M.

## 4.9 - Discussion

In this chapter, we have described the derivation of two APOBEC3H knockout cell lines, HCT116 2B3 and 2C3, through CRISPR mediated DNA deletion directed at Exon 3 of the APOBEC3H gene. In getting to these lines, initial experiments were carried out with a single CRISPR gRNA, CRISPR182078, directed at the APOBEC3H gene and an approach to generate indel mutations. This proved unsuccessful, potentially because of the issue of sequence mismatch between the CRISPR guide RNA and the target sequence, which in this case incorporated a SNP for which the parental HCT116 cell line was heterozygous. Sequence heterogeneity in the guide RNA target clearly can be an issue, and in the case of APOBEC3H is compounded by a large number of coding SNPs found in the gene (Harari *et al.*, 2009).

As an alternative to using pre-defined CRISPR gRNA sequences, such as those produced by Mali *et al.*, (2013), it is possible to design CRISPR gRNAs, using web-based design tools. In essence, these carry out a simple series of steps, the first of which is to identify the 20 base sequences immediately upstream of PAM motifs (NGG) in the region of interest. This define a set of CRISPR gRNA that map within the target region. Once these are identified, the next stage in the analysis are homology searches for each CRISPR against the target genome, which in this case is human. This is an essential step in the CRISPR gRNA design, as those with homologous matches outside the target will potentially lead to off-target Cas9 mediated cutting, so will be scored poorly. However, as this step is very intensive with computer processing time, the web-based design is often batch run, and rarely interactive. For our work, we used the CRISPR design tool hosted by the group of Professor Feng Zhang at the Massachusetts Institute of Technology (CRISPR.MIT.EDU).

Following design, two APOBEC3H CRISPR gRNAs targeting exon 3 were constructed. These were found to be effective in obtaining the predicted DNA deletion when expressed with Cas9 in HCT116 cells. However, it was noted that the approach generated only one heterozygote deletion in the 48 clones examined, Given the small number of clones examined, it is difficult to conclude if this finding is significant, but it appears to be at variance with earlier findings with our work to make deletions within exon 3 of the TP53 gene in MCF7 cells. What is interesting is that in the single heterozygote deletion line, HCT116 APOBEC3H clone 2A4, there is no APOBEC3H expression, as judged by Western blot analysis, suggesting that the full-length allele may be mutated by indel due to CRISPR –Cas9 activity. This would need to be confirmed by further characterisation of the cell line, including DNA cloning and sequencing of APOBEC3H Exon 3 sequences.

The use of APOBEC3H CRISPR Guide 1 and Guide 3 RNAs with Cas9 in HCT116 cells led to two clones HCT116 APOBEC3H clone 2B3 and 2C3, both of which are homozygous for the expected DNA deletion in APOBEC3H exon 3. However, usefully, SNP analysis of the sequences flanking the region of DNA rearrangement clearly show these two clones are independent, as only clone 2B3 remains as heterozygous for two such SNPs. The reason why the second clone, HCT116 APOBEC3H clone 2C3 is homozygous for these SNPs is unclear but could involve additional DNA deletions, beyond the boundaries of the targets for APOBEC3H CRISPR Guide 1 and Guide 3 RNAs. If any additional deletion were found to be local to the proximity of APOBEC3H exon 3, then this would be unlikely to have any effects beyond those seen by the discrete deletion CRISPR-Cas9 engineered in exon 3. However, more substantial deletions or rearrangements following the DNA damage caused by CRISPR-Cas9 cutting may be more problematic and could affect other close by APOBEC3 genes. In order to assess this, a PCR based analysis of the APOBEC3 locus was carried out in the two knockout

lines. While not exhaustive, this analysis suggested that the locus was largely unaffected by the discreet deletions made in the APOBEC3H gene in the two lines.

As part of the generation of the HCT116 APOBEC3H knockout lines, we also characterised the haplotype and isoform expression for APOBEC3H in HCT116. Sequence analysis of APOBEC3H RT-PCR products showed that HCT116 has a Haplotype I (N15 R18 G105 K121 E178) and Haplotype II (N15 R18 R105 D121 D178) genotype. However, an unexpected finding from investigating Nutlin induced transcription of the APOBEC3H gene in the mutant and wild-type HCT116 cells was the identification of a novel variant form of APOBEC3H, variant 208. At 208 aa, this would encode the longest form of APOBEC3H. Previously, (Dang *et al.*, 2008) compared the amino acid sequences of primate APOBEC3H enzymes and concluded that the human APOBEC3H gene encodes a truncated version that arises from a premature stop codon in the last coding exon, removing 29 amino acids, when compared to the primate consensus sequence. Further, the human APOBEC3H “truncated” protein was found to be less active in retroviral restriction, when compared to the “full-length” forms, which included an engineered, extended form based on the human APOBEC3H sequence. This raises the possibility that variant 208, may have altered and perhaps enhanced anti-viral activity. While the function of this variant is unknown, it is apparent that the deletions made in exon 3 result in transcripts where the novel 3' exon encoding variant 208 is no longer used, as only a single RT-PCR product is seen in the knockout cells. Hence, there may be a regulatory association between variant 208 expression and sequences within exon 3. Further work on the function and regulation of APOBEC3H variant 208 is clearly required.

In order to work with homogenous clones for each of the two knockout lines, subcloning was carried out. This led to an assessment of growth in the two resulting sets of subclones for

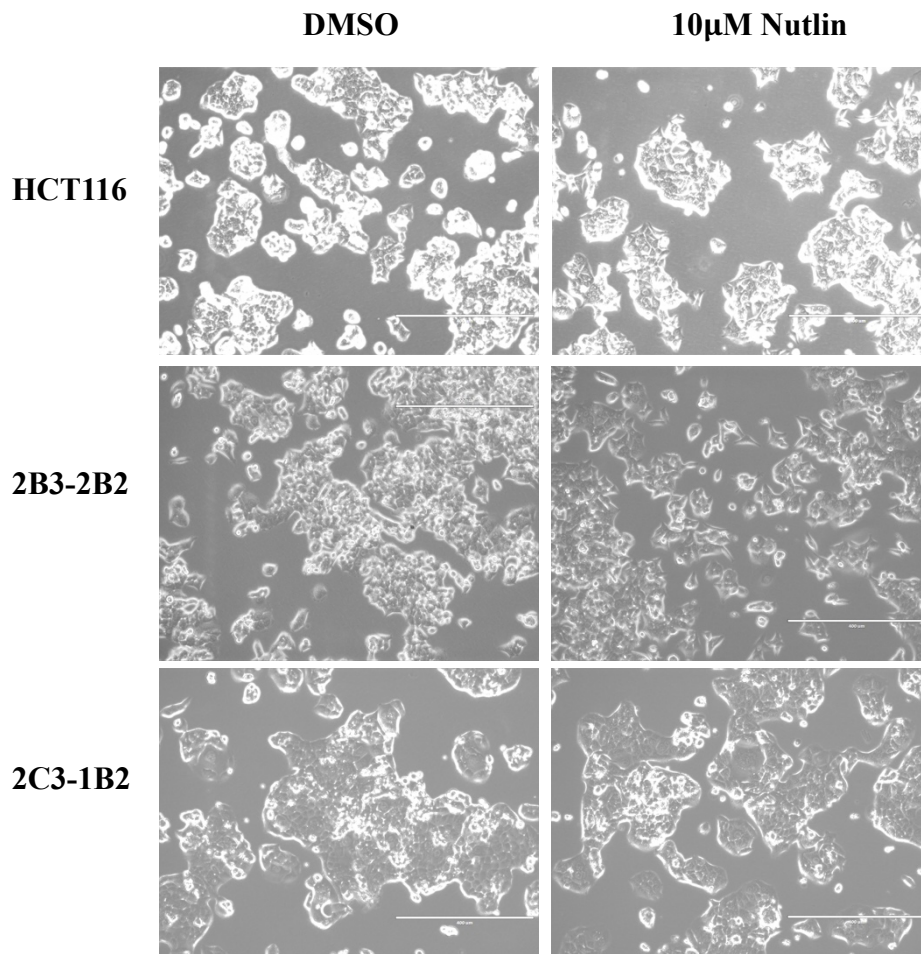
APOBEC3H clone 2B3 and 2C3, where it was found the knockout has no apparent effect on this. This is not surprising, as it is clear from other studies that basal APOBEC3H expression is low in many cell types, so it is unlikely to be required for growth. Indeed, while HCT116 cells express basal levels of APOBEC3H that can be detected by Western blotting, it is clear this is not required for growth and contrasts with APOBEC3B, where knockdown of expression inhibits growth (Periyasamy *et al.*, 2015).

From work presented here, we have seen APOBEC3H as a strongly p53 regulated gene, implying a role in the p53 response. From examining Nutlin sensitivity, it would appear this role is unlikely to be in the activation of p53 through MDM2, as the GI50 of the HCT116 wild-type and knockout cells is around 25 $\mu$ M, a concentration that is similar to the previously reported GI50 for Nutlin in HCT116 (Andrysik *et al.*, 2017). Hence, in order to gain further insight into the role of APOBEC3H, particularly as a p53 regulated gene, I used two knockout sub-clones, HCT116 APOBEC3H clone 2B3-2B2 and 2C3-1B2, in RNAseq analysis, in a Nutlin treated context.

## **5 - Results 3: RNAseq analysis of the APOBEC3H gene knockout constructed in the HCT116 human colon cancer cell line**

### **5.1 - RNAseq analysis of HCT116 A3H knockout lines 2B3-2B2 and 2C3-1B2**

As described above, we have established that APOBEC3H is a p53 regulated gene, suggesting a requirement for this Cytidine Deaminase enzyme in the p53 response. Further, the p53 response needs to be seen in the overall context of APOBEC3 gene regulation following p53 activation, where APOBEC3B is seen to be down regulated, suggesting that a shift from APOBEC3B to APOBEC3H might be important. The reason for this readdressing of APOBEC3 expression remains unclear, and, in order to gain further insight into this, we generated two independent HCT116 APOBEC3H knockout lines 2B3 and 2C3 and found that the knockout of APOBEC3H has little effect on cell growth (Figure 5.1), or sensitivity to Nutlin (Figure 4.12B), as an inhibitor of cell growth. To define the gene expression phenotype of HCT116 APOBEC3H knockout cells by RNAseq, the two knockout lines were sub-cloned to derive sets of clones, from which two, APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 were chosen for further study.



**Figure 5.1. Culture morphology of wild-type HCT116 cells and APOBEC3H knockout HCT116 lines 2B3-2B2 and 2C3-1B2**

Cells were seeded at a density  $5 \times 10^5$  cells in the wells of six well plates and allowed to adhere and spread for 24 h. Duplicate wells for wild-type HCT116 cells and APOBEC3H knockout HCT116 lines 2B3-2B2 and 2C3-1B2 were maintained in full medium, or in medium supplemented with 10 $\mu$ M Nutlin and viewed by bright field microscopy (x10 magnification). This shows treatment with Nutlin under these conditions does not cause marked changes in culture morphology, with little evidence for floating and dead cells.

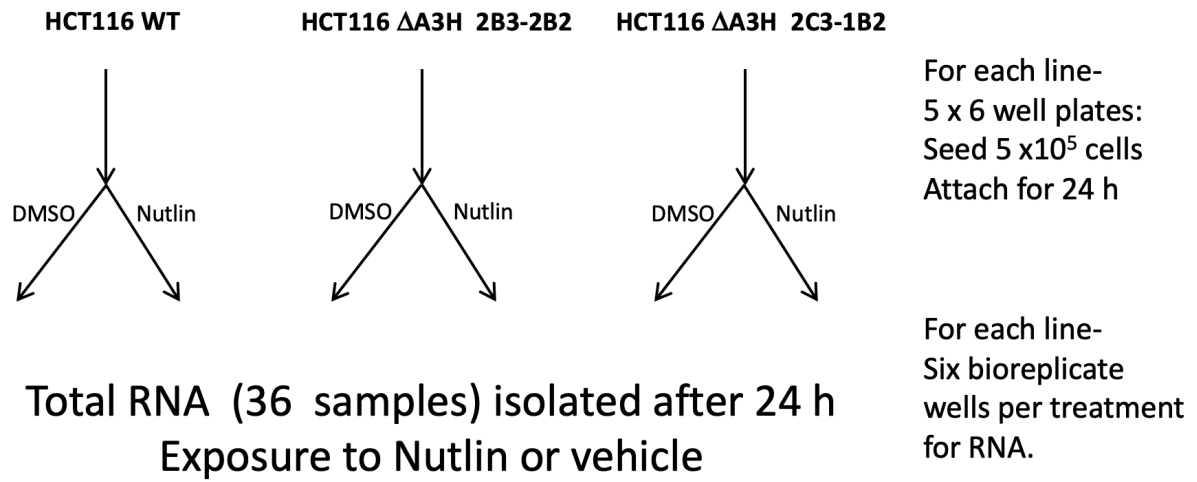


## 5.2 - Overview of strategy

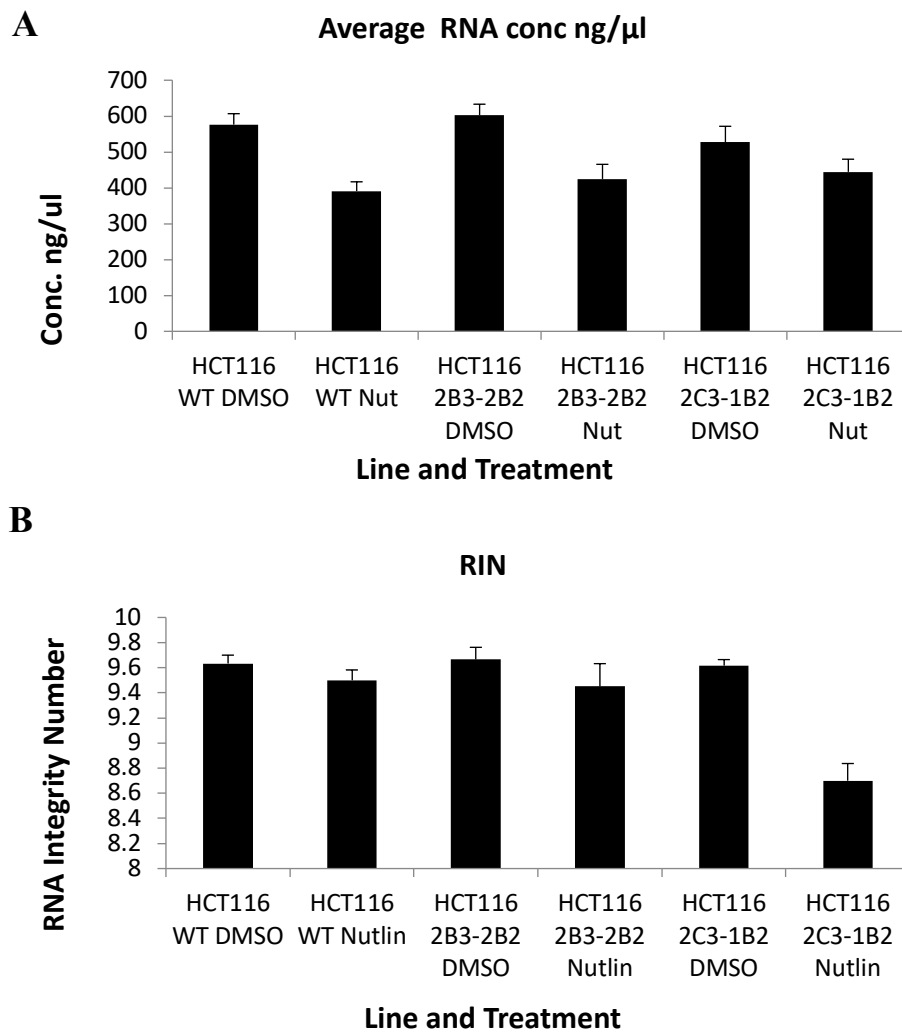
For the RNAseq analysis, gene expression profiles for the HCT116 APOBEC3H knockout cells, and also for these cells in the presence of activated p53, were determined. As HCT116 cells have a low basal level of APOBEC3H expression, this allows an assessment of the phenotype of the APOBEC3H knockout phenotype, both in control, and p53 activated setting. The overall strategy is shown in (Figure 5.2). In summary, cells for each APOBEC3H knockout line, and control HCT116 wild-type cells were seeded in twelve bioreplicate wells (2 x 6 well tissue culture plates) at a seeding density of  $5 \times 10^5$  cells per well. The cells were then left to attach and grow for 24 h, during which time they would be expected to double once. For each line, six of the twelve wells were then treated with 10 $\mu$ M Nutlin, by changing the medium for fresh medium containing 10mM Nutlin dissolved in DMSO, using 1 $\mu$ l /ml medium, while the remaining six wells were treated by changing for fresh medium containing an equivalent amount of DMSO (i.e. 1 $\mu$ l/ml medium), so as to provide a matched, vehicle-treated control. The cells were then left for a further 24 h, where the culture morphology of Nutlin treated cells and control cells was similar and showed little evidence of cytotoxicity in the treated cultures (Figure 5.1). These cultures were lysed *in-situ* and used to make RNA using a spin column purification method that has been shown to produce RNAseq quality RNA (Sellin Jeffries *et al.*, 2014), and which included a column step for the removal of genomic DNA (Qiagen RNAeasy plus). Both the concentration of Nutlin (10 $\mu$ M) and time of treatment are comparable to other studies using Nutlin treatment in HCT116 cells (Andrysik *et al.*, 2017). Overall this resulted in the production of 36 RNA samples; average yields shown in Figure 5.3A. As this shows, overall, RNA yields from the HCT116 wild-type cells and the two HCT116 APOBEC3H knockout lines APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 were comparable, and in each case resulted in lower RNA yields from Nutlin treated cells. As the

isolated RNA is predominantly ribosomal RNA (>90%), this suggests Nutlin affects ribosomal RNA levels in these cells. While it is clear that Nutlin causes p53 mediated effects on transcriptional and translational programmes (Loayza-Puch *et al.*, 2013). The relationship between Nutlin and ribosomal RNA levels is complex (Scala *et al.*, 2016).

Determining the RNA Integrity Number (RIN) initially assessed the quality of the isolated RNAs, Unlike with RNA yield, there was little effect of Nutlin on RIN, with the exception of the HCT116 APOBEC3H 2C3-1B2 line, where the Nutlin treated cells produced RNAs with the lowest RIN numbers, but still >8 (an average 8.7)( Figure 5.3B). This analysis showed that all samples were of high quality and above the cut-off required for RNAseq (>RIN 8) (Gallego Romero *et al.*, 2014).



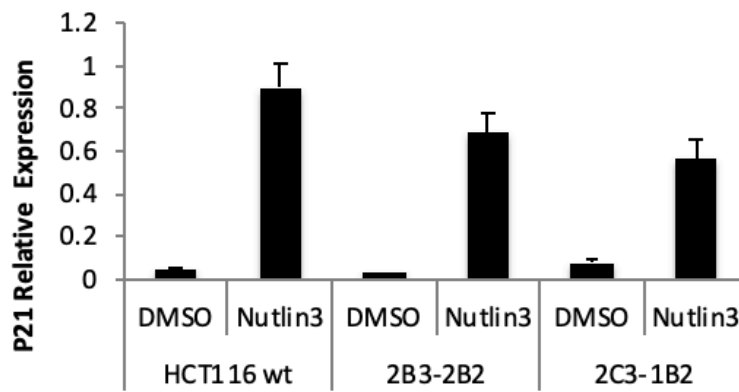
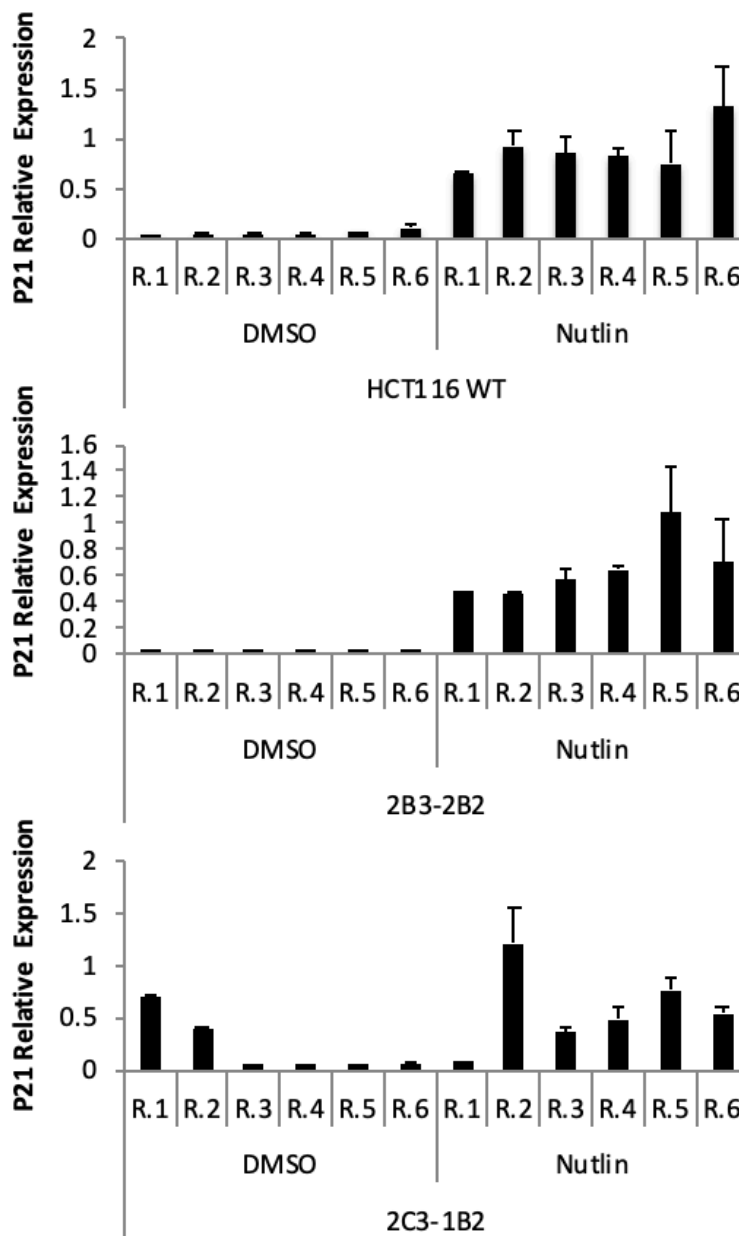
**Figure 5.2. Overview of sample generation for RNAseq analysis of HCT116 cells and HCT116  $\Delta$ A3H lines 2B3-2B2 and 2C3-1B2**



**Figure 5.3. Overview of sample generation for RNAseq analysis of HCT116 cells and HCT116  $\Delta$ A3H lines 2B3-2B2 and 2C3-1B2**

For RNAseq work, RNA was prepared from six bioreplicate cultures for each line and treatment (DMSO control and 10 $\mu$ M Nutlin). RNA samples were prepared in parallel using spin column chromatograph (RNeasy Plus- Qiagen) and eluted under the same conditions. **A.** The concentration of RNA for each replicate was determined and averaged for a set of six bioreplicates. The average RNA concentration for each replicate set is shown with standard errors of the mean. **B.** RNA samples for six bioreplicates for each cell line and treatment were analysed for integrity, and the RNA Integrity Number (RIN) determined for each sample. The average RIN for each replicate set is shown with standard errors of the mean.

cDNA was synthesised for each sample, and also, from pooled 2µg amounts of each RNA from a replicate set and using this to make cDNA representative of RNA for each cell line treated with Nutlin, and control DMSO vehicle. This was used in Q-RT-PCR to assess the expression of the p53 regulated genes p21 to confirm p53 activation by Nutlin treatment. As can be seen in Figure 5.4A, the multiple replicate RNAs show an appropriate induction of p21 gene expression by Nutlin in HCT116 wild-type cells and the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines. Further, the pooled samples, show a pattern of Nutlin induced p21 gene expression, as previously seen for the two HCT116 APOBEC3H knockout lines, together with the wild-type HCT116 cells (Figure 5.4B).

**A****B**

**Figure 5.4. p21 relative expression in HCT116 cells and APOBEC3H knockout HCT116 knockout cells**

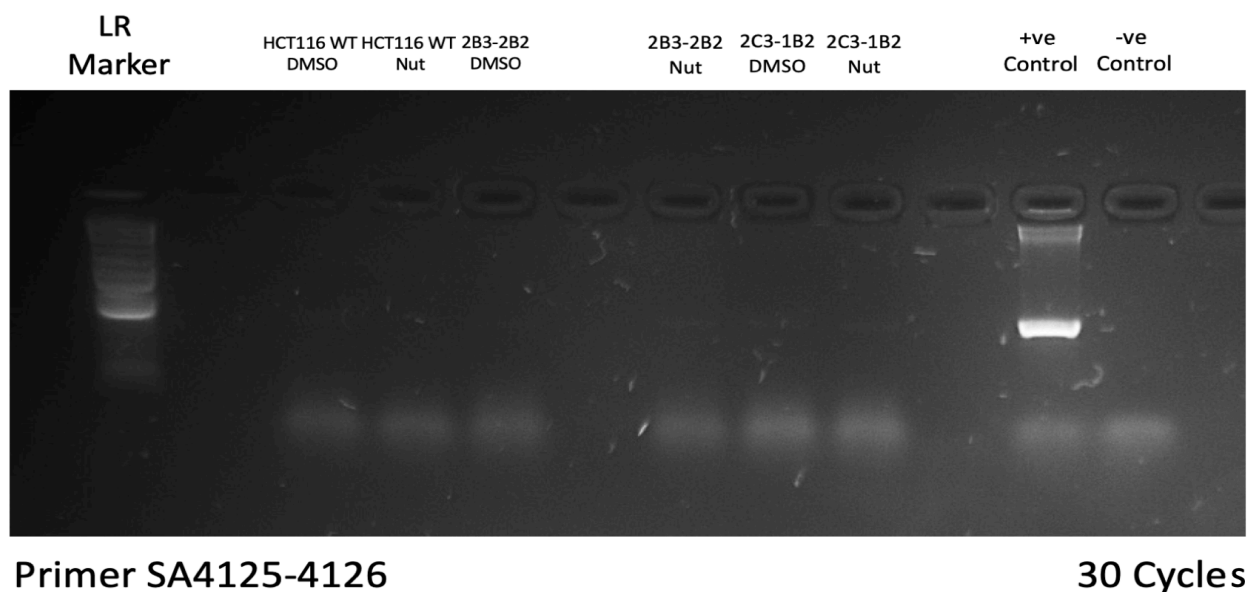
**A.** RNA prepared from six bioreplicate cultures of wild-type HCT116 cells and APOBEC3H knockout HCT116 lines 2B3-2B2 and 2C3-1B2 maintained in full medium, or in medium supplemented with 10 $\mu$ M Nutlin was pooled in equal amounts and used to make cDNA representative of each line and treatment. This cDNA was used to measure p21 gene expression, which was normalised to GAPDH housekeeping gene expression and is shown standard errors of the mean. **B.** p21 Expression was also measured for each bioreplicate RNA sample, which was normalised to GAPDH housekeeping gene expression, and is shown is shown with standard errors of the mean.

The RNA prepared for RNAseq was assessed for contaminating genomic DNA. For this, a PCR assay designed to amplify genomic DNA from exon 8 of the human estrogen receptor – alpha gene (ESR1), was used. As can be seen in Figure 5.5 while a PCR product can be seen when control genomic template was used, no product was found when an equivalent amount of total RNA, from the RNA pools of six replicates, was used. This confirmed that there was no appreciable amount of genomic DNA contamination in the RNA isolated from RNAseq work.

Taken together, these data all suggested that the isolated RNA was of sufficient quality and showed marked sufficiently changes in gene expression following Nutlin treatment, for RNAseq analysis.



## Pooled cDNA (6 Replicates)



**Figure 5.5. Assessment of genomic DNA contamination in RNA preparations**

RNA prepared from six bioreplicate cultures of wild-type HCT116 cells and APOBEC3H knockout HCT116 lines 2B3-2B2 and 2C3-1B2 maintained in full medium, or in medium supplemented with 10 $\mu$ M Nutlin was pooled in equal amounts. 300ng of RNA from each pool was used in a genomic PCR reaction for ESR1 exon 8 (primers SA4125-4126 amplify a 462bp product from genomic DNA) and amplification carried out for 30 cycles. 300 ng of HCT116 cell genomic DNA was used as a positive control template (+ve). The negative control (-ve) was for a PCR assay with no template. The analysis shows evidence for little genomic DNA contamination in the HCT116 wild-type and APOBEC3H knockout isolated RNA samples.

## **5.3 - RNAseq**

### **5.3.1 - RNAseq – library construction and next generation sequencing**

The Imperial College BRC Genomics Facility (Imperial College London, 2019) carried out Next Generation RNAseq sequencing on the replicate RNA samples for HCT116 wild-type, HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines. Briefly, this involves three steps- (i) creation of a sequencing library, (ii) seeding and preparation of the flow cell and (iii) sequencing by synthesis. For library preparation, polyadenylated mRNA was isolated from the total RNA using NEBNext poly(A) mRNA magnetic isolation module (New England Biolabs, UK), and this was converted to cDNA and subsequently amplified using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina Sequencing and NEBNext Multiplex Oligos (New England Biolabs, UK). The resulting 36 libraries were then pooled in equimolar amounts and the final pool sequenced over four lanes of an Illumina HiSeq4000 NGS sequencer, set to read paired-end reads (2x75bp). This is designed to provide an expected coverage of approximately 35 million reads per RNA sample.

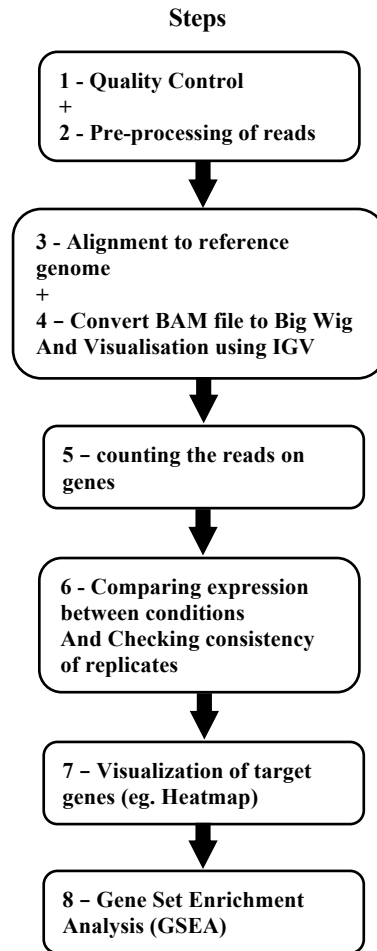
### **5.3.2 - RNAseq analysis**

Analysis of the RNAseq data was carried out with Ms Van Nguyen, Research Assistant, Department of Surgery & Cancer, Imperial College London. The workflow for this is shown in Figure 5.6 and involved a series of steps, starting with the raw reads generated by the Illumina HiSeq400 NGS sequencer, and alignment of these with a reference human genome sequence (hg19 human genome) downloaded from (HISAT2, 2017). The aligned reads are

used to obtain BAM files – Binary versions of Sequence Alignment Map (SAM) files, which are text-based, tab-delimited files that contain sequence alignment data. The BAM files allows analysis with Tophat (Trapnell *et al.*, 2009), a widely used ultrafast alignment software package that performs spliced alignments, and read-counts using the HT-seq package (Anders *et al.*, 2015). Differential gene expression was then obtained from read counts using the DESeq2 (Love *et al.*, 2014) Bioconductor package, which provides a method for differential analysis of count data.

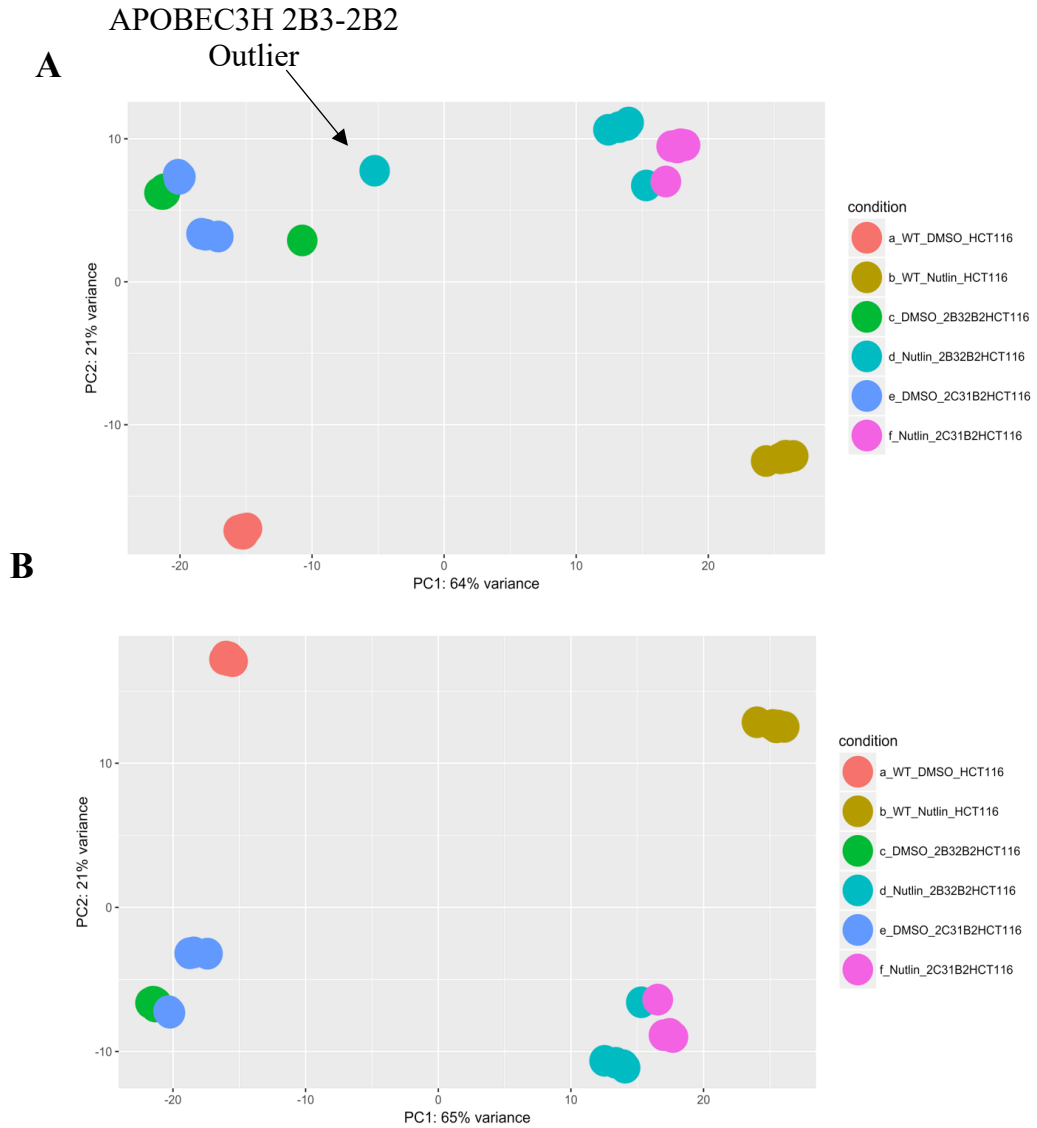
Once analysed with DESeq2, it is possible to determine the consistency of replicate sets of RNA, as defined by the gene expression they demonstrate, using Principal Component Analysis (PCA) (Ma and Dai, 2011). PCA is a statistical analysis method that leads to a dimensional reduction of large data sets, by reducing these to smaller ones, so that they are described by the dominant information in the large set. In this case, PCA analysis of each sample uses the gene expression data determined by the RNAseq for each sample, and results in assigned values, which when plotted in 2-D shows how samples are related, with the most alike lying close together. In this way, outlier samples in a replicate set can be identified.

The PCA analysis for the HCT116 wild-type cells and the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines, for both control DMSO and Nutlin treated replicate cultures is shown in Figure 5.7A. There are tight PCA clusters for all replicate sets, with the exception of one replicate from the HCT116 APOBEC3H 2B3-2B2 set. This sample was considered as an outlier and was removed from the analysis of this sample, and one sample from each of the five other replicate RNA sets. The PCA analysis for the edited sets is shown in Figure 5.7B. This produced tight PCA clustering for the replicates for each sample set.



**Figure 5.6. Overview of RNAseq analysis**

An overview and summary of the RNAseq analysis employed is shown:  
 Steps 1 and 2 are for quality control and pre-processing of reads and involve use of FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)  
 Steps 3 and 4 are for alignment to the reference genome (hg19 human genome; [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001405.13/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/)) and involve FastQC to generate BAM alignment files using Hisat2: <https://ccb.jhu.edu/software/hisat2/index.shtml>. BAM files can be converted to “Big Wig” files for data visualisation using Integrative Genomic Viewer (IGV) <https://software.broadinstitute.org/software/igv/>.  
 Step 5. The HT-seq package ([https://htseq.readthedocs.io/en/release\\_0.9.1/](https://htseq.readthedocs.io/en/release_0.9.1/)) is used for counting sequencing reads on genes.  
 Step 6 allows a comparison of gene expression between conditions: This is carried out using Deseq2 (<https://bioconductor.org/packages/release/bioc/html/DESeq2.html>).  
 Step 7 and 8. Separate R packages: heatmap2; ggplot2s can be used for visualisation of target genes (<https://www.bioconductor.org>) and GSEA ( gene set enrichment assay)



**Figure 5.7. Principal component analysis for control (DMSO) replicates in the two APOBEC3H KO lines compared to HCT116 WT cells**

Principal Component Analysis (PCA) was carried out on read counts generated for each sample using the HT-seq analysis package. **A.** Analysis of all six replicates per set showed good replicate clustering, with exception of HCT116 2B3-2B2 replicate 1 (arrowed). **B.** Sample HCT116 2B3-2B2 replicate 1 was removed as an outlier, and the PCA analysis repeated. This resulted in tighter PCA clustering across all replicate sets.

### **5.3.3 - Analysis of differential expression between HCT116 wild-type, HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cells**

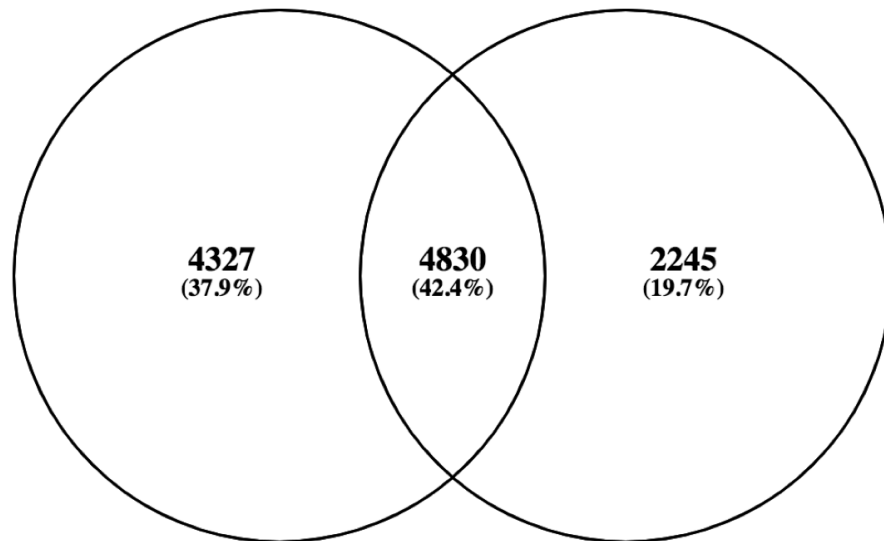
DESeq2 analyses was used to define gene expression differences between wild-type HCT116 cells and each of the APOBEC3H HCT116 knockout lines, APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2. Using a cut-off for the P value of  $<0.05$ , this identified 9,157 differentially expressed genes between wild-type HCT116 cells and APOBEC3H 2B3-2B2 cells and 7075 between wild-type HCT116 cells and APOBEC3H 2C3-1B2 cells (Table 5.1). Of these, 4830 (42.4%) differentially expressed genes were common to both APOBEC3H knockout lines (Figure 5.8) indicating a significant overlap in differentially expressed genes in the two cell lines. In order to further refine the differentially expressed genes into differentially up-regulated and differentially down-regulated, the P value  $<0.05$  gene list for each of the APOBEC3H knockout lines was further subject to a cut-off of fold change in gene expression of  $\text{Log}_2 \geq 0.5$  (1.4 fold). This identified 1509 down regulated and 999 up-regulated genes in the APOBEC3H 2B3-2B2 line and 1045 down regulated and 761 up-regulated genes in the APOBEC3H 2C3-1B2 line. The overlap between down-regulated genes between the two APOBEC3H lines was 554 (27.7%) genes (Figure 5.9A), while for the up-regulated genes, the overlap was 252 (14.6%) genes (Figure 5.9B).

**Table 5.1. Gene expression differences between wild-type HCT116 cells and the APOBEC3H HCT116 knockout lines**

<b>COMPARISON</b>	<b>P &lt; 0.05</b>	<b>Log2 FC &gt; 0.5</b>
<b>DMSO_ WT HCT116 vs 2B32B2</b>	9157	1509 DOWN & 999 UP
<b>DMSO_ WT HCT116 vs 2C31B2</b>	7075	1045 DOWN & 982 UP
<b>DMSO vs Nutlin_ WT116</b>	12681	3156 DOWN & 3721 UP
<b>DMSO vs Nutlin_ 2B32B2</b>	12352	2794 DOWN & 3433 UP
<b>DMSO vs Nutlin_ 2C31B2</b>	11144	2716 DOWN & 2954 UP
<b>Nutlin_ WT H116 vs 2B32B2</b>	8987	1599 DOWN & 1141 UP
<b>Nutlin_ WT H116 vs 2C31B2</b>	8736	1731 DOWN & 1113 UP
<b>Nutlin_ 2B32B2 vs 2C31B2</b>	9072	1435 DOWN & 1422 UP

**WT HCT116 vs 2B32B2**

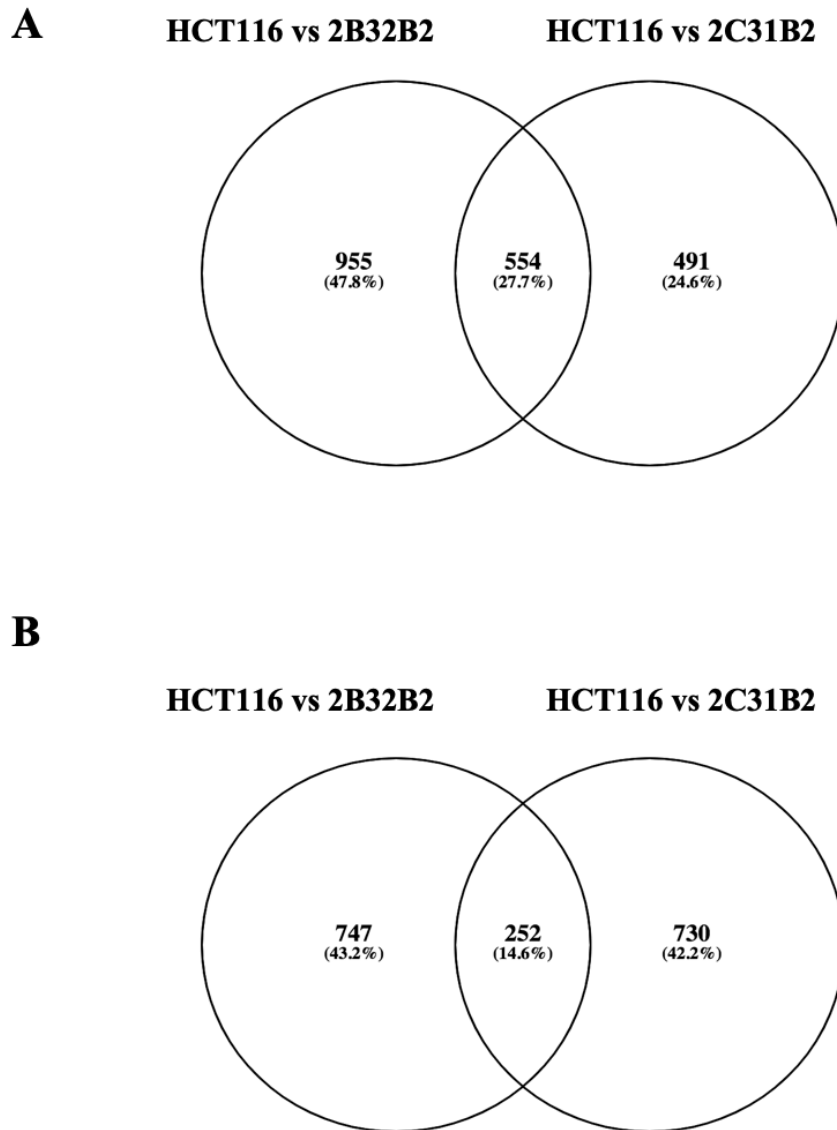
**WT HCT116 vs 2C31B2**



**Figure 5.8. Differentially expressed genes shared by the APOBEC3H HCT116 knockout cell lines 2B3-2B2 and 2C3-1B2**

Differentially expressed gene lists between the APOBEC3H HCT116 knockout cell lines 2B3-2B2 and 2C3-1B2 HCT116 wild-type cells generated using Dseq2 analysis, and ranked for P value <0.05, This identified 4830 differentially expressed genes that were common to the two knockout lines.





**Figure 5.9. Differentially expressed genes shared by the APOBEC3H HCT116 knockout cell lines 2B3-2B2 and 2C3-1B2**

Differentially expressed genes in the APOBEC3H HCT116 knockout cell lines 2B3-2B2 and 2C3-1B2, compared to gene expression in wild-type HCT116 cells, was analysed further by using cut-offs of  $P < 0.05$  and Log<sub>2</sub> fold expression of  $\geq 0.5$ . This identified 1509 down regulated and 999 up-regulated genes in the APOBEC3H 2B3-2B2 line and 1045 down regulated and 761 up-regulated genes in the APOBEC3H 2C3-1B2 line. **A.** 554 (27.7%) genes overlapped in the down-regulated gene lists between the two APOBEC3H knockout lines. **B.** For up-regulated genes, 252 (14.6%) overlapped between up-regulated gene lists between the two APOBEC3H knockout lines.

### **5.3.4 - Down-regulated gene expression in the APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 indicates a loss of the Y chromosome**

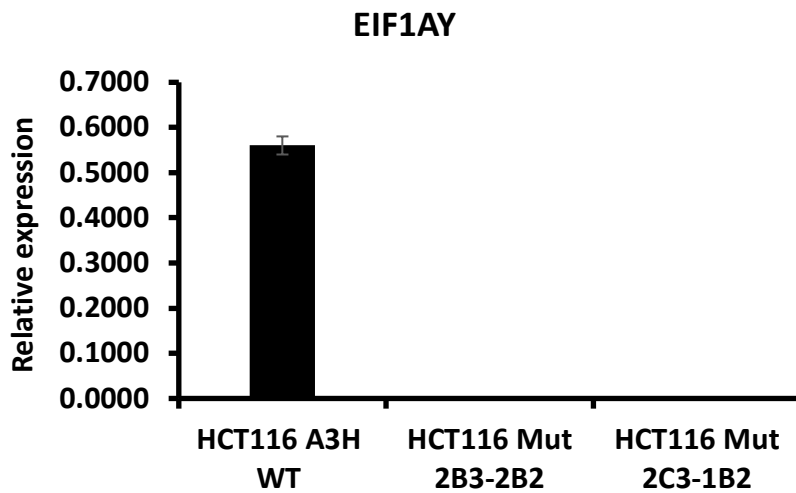
Table 5.2 shows the top 50 down-regulated genes in common in the APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines, when compared to wild-type HCT-116 cells. Analysis of this for gene location, highlights an enrichment for Y chromosome specific genes (highlighted in yellow). This down-regulation was further investigated using Q-RT-PCR for a set of Y-chromosomes encoded genes, as shown in (Figure 5.10), where it was apparent that Y-chromosome specific gene expression was lost in the two APOBEC3H knockout lines. A further analysis of the karyotype of the wild-type HCT116 cells and the APOBEC3H HCT116 knockout lines, APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2. was carried out by Dr R. Fisher, Department of Surgery & Cancer, Imperial College London, and this confirmed that both lines had lost the Y-chromosome (Table 5.3), and indicated that there was heterogeneity for the presence of the Y – chromosome in the wild-type HCT116 cells, as indicated by the dominance of the presence of the X-chromosome in this line. It was therefore likely that the APOBEC3H 2B3 and APOBEC3H 2C3 lines independently arose in cell clones that had lost the Y-chromosome. Nevertheless, the RNAseq analysis also identified genes whose expression was down-regulated in both lines, but not as a result of encoding on the Y chromosome, as exemplified by Q-RT-PCR analysis for PXDN (encoded on Chromosome 2p25.3), and GALNT14 (encoded on Chromosome 2p23.1), SPARC1 (encoded on Chromosome 5q33.1), SLCO3A1 (encoded on Chromosome 15q26.1), HKDC1 (encoded in chromosome 10q22.1), TCF21 (encoded on Chromosome 6q23.2 ) and LARGE1 (encoded on Chromosome 22q12.3) (Figure 5.11).

**Table 5.2. Top 50 common down-regulated genes in the two APOBEC3H knockout cell lines**

<b>Gene ID</b>	<b>Gene Name</b>	<b>Log2 fold differential gene expression</b>
ENSG00000198692.5	EIF1AY	-11.433215
ENSG00000131002.7	TXLNG2P	-11.390044
ENSG00000114374.8	USP9Y	-11.026585
ENSG00000067646.7	ZFY	-10.932125
ENSG00000129824.11	RPS4Y1	-10.750956
ENSG0000012817.11	KDM5D	-10.732626
ENSG00000130508.6	PXDN	-10.366693
ENSG00000067048.12	DDX3Y	-10.25939
ENSG00000183878.11	UTY	-9.6590723
ENSG00000154620.5	TMSB4Y	-9.0145996
ENSG00000233864.3	TTY15	-8.5843027
ENSG00000270038.1	RP11-1070N10.7	-8.1947981
ENSG00000118526.6	TCF21	-7.7557751
ENSG00000158089.10	GALNT14	-7.5928586
ENSG00000183960.4	KCNH8	-7.4340633
ENSG00000176728.3	TTY14	-7.2199497
ENSG00000242779.2	ZNF702P	-7.1883871
ENSG00000187098.10	MITF	-7.1601801
ENSG00000241859.2	KALP	-7.0196125
ENSG00000206538.3	VGLL3	-7.0142285
ENSG00000260197.1	RP11-424G14.1	-6.8582343

ENSG00000258927.1	RP11-1070N10.5	-6.7423485
ENSG00000099725.10	PRKY	-6.2658661
ENSG00000239893.1	ZNF736P9Y	-6.0565972
ENSG00000204394.8	VAR5	-5.9772222
ENSG00000250328.1	CTC-210G5.1	-5.9513541
ENSG00000133424.16	LARGE	-5.9399623
ENSG00000173391.4	OLR1	-5.937117
ENSG00000170011.9	MYRIP	-5.8557054
ENSG00000137337.10	MDC1	-5.8395002
ENSG00000070190.8	DAPP1	-5.8305833
ENSG00000160179.14	ABCG1	-5.8196934
ENSG00000169218.9	RSPO1	-5.6622851
ENSG00000226611.2	OFD1P2Y	-5.6515806
ENSG00000176463.9	SLCO3A1	-5.6293741
ENSG00000134198.5	TSPAN2	-5.5631366
ENSG00000102466.11	FGF14	-5.5290214
ENSG00000113140.6	SPARC	-5.4884616
ENSG00000077274.7	CAPN6	-5.4666759
ENSG00000122691.8	TWIST1	-5.3449798
ENSG00000204463.8	BAG6	-5.3237987
ENSG00000181143.11	MUC16	-5.154835
ENSG00000177614.5	PGBD5	-5.0441393
ENSG00000147862.10	NFIB	-4.9606331
ENSG00000227954.2	RP3-323P13.2	-4.8950826

ENSG00000089847.8	ANKRD24	-4.8389565
ENSG00000187094.7	CCK	-4.8130543
ENSG00000187621.10	TCL6	-4.7450475
ENSG00000269113.3	TRABD2B	-4.7398729



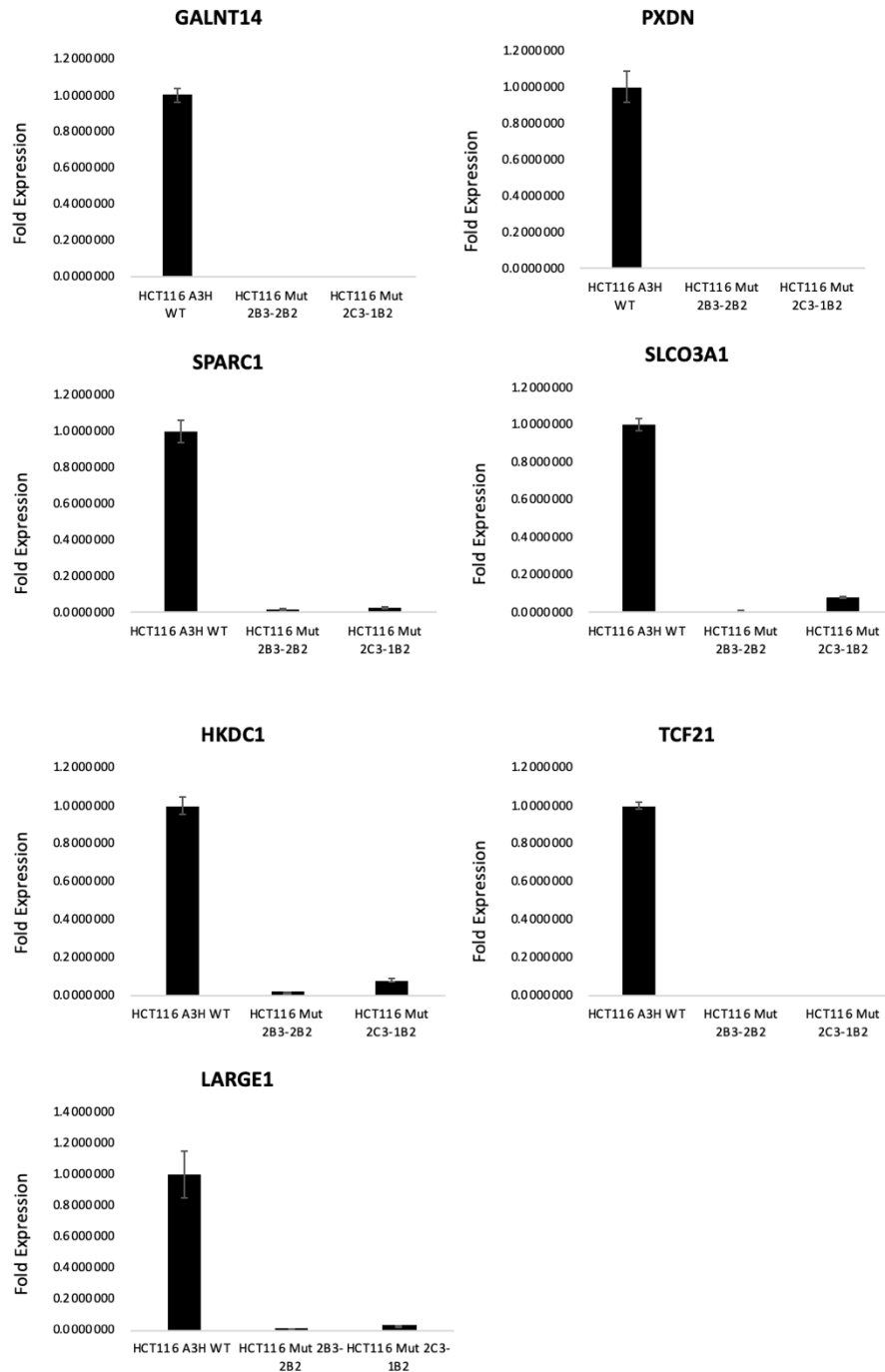
**Figure 5.10. Y-Chromosome gene expression is down-regulated in the HCT116 APOBEC3H knockout cell lines**

RNAseq analysis showed that genes encoded on the Y chromosome were lost in both HCT116 APOBEC3H knockout cell lines. This was further investigated by Q-RT-PCR analysis for the expression of genes on the Y chromosome, as exemplified here for the Y chromosome specific gene EIF1AY. Gene expression was measured using cDNA made to pooled RNA from the bioreplicate samples used for RNAseq, and normalised to expression of the control GAPDH housekeeping gene. Error bars are for standard error of the mean.

**Table 5.3. Genotyping of HCT116 wild-type cells and the APOBEC3H HCT116 knockout cell lines 2B3-2B2 and 2C3-1B2**

Line		ATCC	HCT116 WT	2B3-2B2	2C3-1B2
Locus	chr	Alleles			
D7S820	7	11 - 12	11 -12	11 -12	11 -12
CSF1PO	5	7 - 10	7 - 10	7 - 10	7 - 10
THO1	11	8 - 9	8 - 9	8 - 9	8 - 9
D13S317	13	10 - 12	10 - 12	10 - 13	10 - 13
D16S539	16	11 - 13	11 - 13 > 14	10 - 13 - 16	11 - 13 - 14
vWA	12	17 - 22	17 - 18 - 22 > 23	18 - 22	18 - 23
TPOX	2	8 - 9	8	8	8
AMEL	XY	XY	X >> Y	X	X
D5S818	5	10 - 11	10 - 11	10 - 11	11

American Type Culture Collection (ATCC) short-tandem repeat (STR) profiling was carried out using DNA extracted from HCT116 wild-type cells and the APOBEC3H HCT116 knockout cell lines 2B3-2B2 and 2C3-1B2. Allele variations between the lines and the ATCC reference genotype are highlighted. Overall the analysis confirms the lines as being HCT116 origin. The Amelogenin locus (AMEL), duplicated on X and Y, is seen in both ATCC the reference genome and our HCT116 wild-type stock. However, AMEL genotyping is consistent with the loss of the Y chromosome in both APOBEC3H HCT116 knockout cell lines 2B3-2B2 and 2C3-1B2, as indicated by the RNAseq analysis.



**Figure 5.11. Down-regulated gene expression in the HCT116 APOBEC3H knockout cell lines**

Down regulated gene expression, in addition to the loss of Y chromosome encoded gene expression in the APOBEC3H knockout cell lines through chromosome loss, was investigated. Q-RT-PCR analysis of candidate, strongly down-regulated genes common to both HCT116 APOBEC3H knockout cell lines is shown for the genes GALNT14, PXDN, SPARC1, SLCO3A1, HKDC1, TCF21 and LARGE1. Error bars are for standard error of the mean.



### **5.3.5 - Up-regulated gene expression in the HCT1116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 is enriched for processed pseudogene transcription.**

Table 5.4 shows the top 49 up-regulated genes, when compared to the wild-type HCT116 cells, which were shared by the APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines. What is striking is that 61% (30/49) of these correspond to processed pseudogenes (shaded in yellow). Of the remaining genes, six come from non-coding transcripts (shaded in red), with the remaining 13 being protein encoding genes. While this enrichment was clear, it was also apparent from the read counts, that the differential expression reflected changes for several at very low levels of expression. This, together with the fact that many of these processed pseudogenes differ from their active gene homologues by only a few nucleotides in sequence, all make the detection of pseudogenes very challenging. Therefore, it was not possible to validate the up regulation of these processed pseudogenes further.

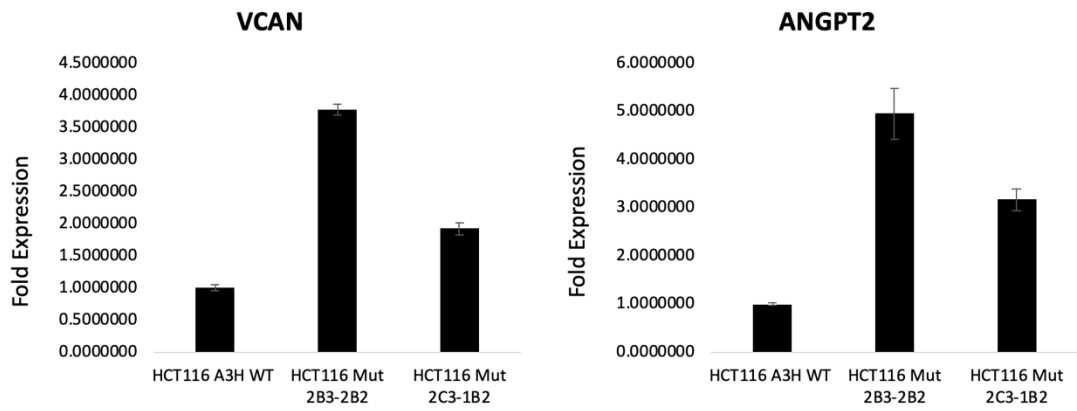
The RNAseq analysis also showed that, in addition to the putative up-regulation of a set of pseudogenes in the APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines, protein encoding genes were also found to be up-regulated. This is exemplified by the genes VCAN (encoded on chromosome 5q12-14) and ANGPT2 (encoded on chromosome 8p23.1), both well characterised genes which were seen to be up-regulated in both the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines (Figure 5.12).

**Table 5.4. Top 49 common up-regulated genes in the two APOBEC3H knockout cell lines**

Gene ID	Gene Name	Log2 Fold differential gene expression
ENSG00000214222.2	TUBBP2	10.72569374
ENSG00000231313.2	CLIC1P1	9.574021874
ENSG00000231500.2	RPS18	8.438432038
ENSG00000233383.1	RP11-324F21.1	8.209046778
ENSG00000236216.4	PPP1R11P1	7.721691549
ENSG00000230897.1	RPS18P12	5.45183181
ENSG00000181433.5	SAGE1	5.254063122
ENSG00000234335.1	RPS4XP11	5.098889575
ENSG00000212994.4	RPS26P6	5.072925533
ENSG00000219186.2	FTH1P19	4.590436388
ENSG00000213860.3	RPL21P75	4.487873731
ENSG00000244171.3	PBX2P1	4.465236868
ENSG00000235207.1	TUBBP6	4.083833273
ENSG00000234981.1	RP11-534L20.4	4.029565295
ENSG00000206995.1	Y_RNA	4.014961841
ENSG00000241612.1	RP13-585F24.1	4.006157548
ENSG00000261557.1	EEF1A1P38	3.909442745
ENSG00000215480.3	OR7E37P	3.838026835
ENSG00000224858.4	RPL29P11	3.77553042
ENSG00000244582.2	RPL21P120	3.745277588
ENSG00000243071.1	RP11-613M5.2	3.737421175
ENSG00000227008.2	RP3-417G15.1	3.653513862

ENSG00000203576.1	AC090519.7	3.622273629
ENSG00000269048.1	Z98049.1	3.621804124
ENSG00000206824.1	Y_RNA	3.526105086
ENSG00000239470.2	RP11-16F15.2	3.517956243
ENSG00000008517.12	IL32	3.442565277
ENSG00000177688.5	SUMO4	3.437155918
ENSG00000253570.1	RNF5P1	3.379153571
ENSG00000259264.1	RP11-60L3.1	3.343035919
ENSG00000234287.1	RP11-761N21.2	3.304968326
ENSG00000262874.1	CTD-2616J11.4	3.268021337
ENSG00000250762.1	RP1-313L4.4	3.254557576
ENSG00000249835.2	VCAN-AS1	3.212860361
ENSG00000125845.6	BMP2	3.18061873
ENSG00000250504.2	KRT18P51	3.161949985
ENSG00000248050.1	RP11-422N16.3	3.097281301
ENSG00000237296.5	SMG1P1	3.085689424
ENSG00000105131.3	EPHX3	3.054219153
ENSG00000229314.4	ORM1	3.009269966
ENSG00000243007.1	RP11-106M3.1	2.992788688
ENSG00000147041.7	SYTL5	2.894836678
ENSG00000261780.2	CTD-2354A18.1	2.803059456
ENSG00000228709.1	AP001065.15	2.80017747
ENSG00000230637.2	CTA-246H3.8	2.754723821
ENSG00000240540.2	RPL3P9	2.703807077

ENSG00000153993.9	SEMA3D	2.648090985
ENSG00000255735.1	AC110619.1	2.644545687
ENSG00000219932.5	RPL12P8	2.559027771













**Figure 5.12. Up-regulated gene expression in the HCT116 APOBEC3H knockout cell lines**

Up-regulated gene expression, in addition to processed pseudogene expression was investigated. Q-RT-PCR analysis of candidate, strongly up-regulated genes common to both HCT116 APOBEC3H knockout cell lines is shown for the genes VCAN (versican) and

### **5.3.6 - Gene set enrichment analysis of differentially regulated gene expression in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines**

Gene Set Enrichment Analysis (GSEA) is a method of statistical analysis that can be applied to gene lists coming from differential gene expression analysis, which aims to identify enrichment, or depletion for groups of genes that are related by being assigned to particular functional pathways and gene expression responses. Generally, the analysis can be carried out on-line using GSEA analysis programmes use databases of gene assignments to pathways, programmes, etc. In order to gain further insight into the phenotype of the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines.

GSEA analysis was carried out using the GSEA server hosted by the Broad Institute (<http://software.broadinstitute.org/gsea/index.jsp>), using the common differentially expressed gene list for the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines, filtered for Differential Gene Expression:  $p < 0.05$  and 2 fold change (2900 genes). As can be seen in (Figure 5.13), this analysis identifies “Genes involved in p53 pathways and networks” as being the third most significant enriched pathway. This implication is explored further, as detailed below.

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p-value	FDR q-value
HALLMARK_ESTROGEN_RESPONSE_EARLY [200]	Genes defining early response to estrogen.	55		2.89 e-22	7.22 e-21
HALLMARK_ESTROGEN_RESPONSE_LATE [200]	Genes defining late response to estrogen.	55		2.89 e-22	7.22 e-21
HALLMARK_P53_PATHWAY [200]	Genes involved in p53 pathways and networks.	53		1.05 e-20	1.75 e-19
HALLMARK_HYPOXIA [200]	Genes up-regulated in response to low oxygen levels (hypoxia).	52		6.1 e-20	7.63 e-19
HALLMARK_TNFA_SIGNALING_VIA_NFKB [200]	Genes regulated by NF-kB in response to TNF [GeneID=7124].	51		3.45 e-19	3.45 e-18
HALLMARK_MITOTIC_SPINDLE [200]	Genes important for mitotic spindle assembly.	50		1.9 e-18	1.59 e-17
HALLMARK_GLYCOLYSIS [200]	Genes encoding proteins involved in glycolysis and gluconeogenesis.	49		1.02 e-17	7.3 e-17
HALLMARK_G2M_CHECKPOINT [200]	Genes involved in the G2/M checkpoint, as in progression through the cell division cycle.	47		2.71 e-16	1.7 e-15
HALLMARK_INTERFERON_GAMMA_RESPONSE [200]	Genes up-regulated in response to IFNG [GeneID=3458].	46		1.34 e-15	7.45 e-15
HALLMARK_DNA_REPAIR [150]	Genes involved in DNA repair.	39		2.51 e-15	1.25 e-14

**Figure 5.13. Gene set enrichment analysis of differentially regulated gene expression in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 lines**

Differentially expressed genes common to both HCT116 APOBEC3H knockout cell lines, with a 2 fold, or greater differential expression (P value<0.05), were used in Gene Set Enrichment Analysis (GSEA) and the top ten enriched pathways identified. Note, the analysis identified s “p53 pathways and networks” in this list (third).

### **5.3.7 - Evaluation of the Nutlin - p53 response in wild-type HCT116 cells and in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines**

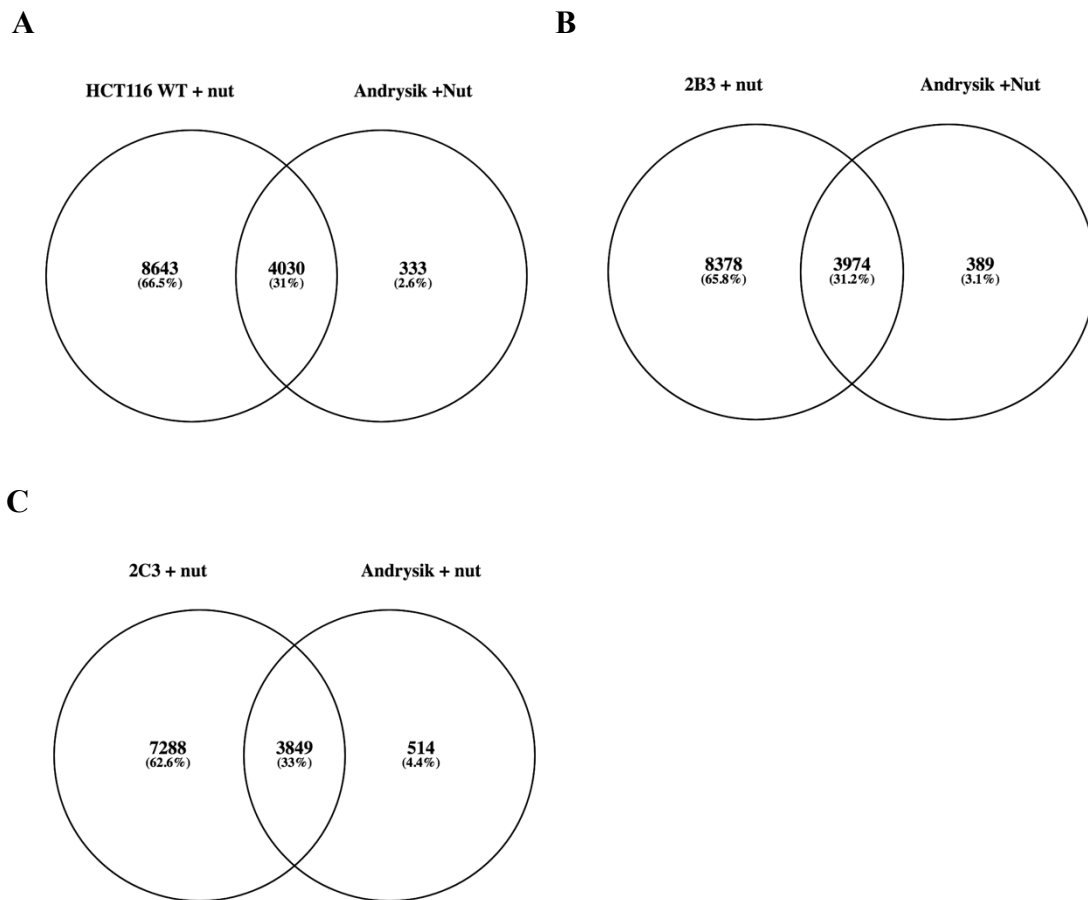
The Nutlin-p53 response has been studied in detail in a variety of cell lines, including in HCT116. A series of papers have been published from the group of Espinosa (Sullivan *et al.*, 2018), who have progressively refined the identity of genes that are differentially expressed when cells are treated with Nutlin. The most recent iteration of this work, published by Andrysik (Andrysik *et al.*, 2017) is a comprehensive assessment of p53 target genes, using 10 $\mu$ M Nutlin to activate in the lines HCT116, MCF7 and SJSA, for 12 h, with the subsequent analysis being carried by RNAseq, Chromatin Immunoprecipitation –sequencing (ChIP-seq) (Nakato and Shirahige, 2017) and Global Run On–sequencing (GRO-seq) (Gardini, 2017), and integrating this data to produce a core set of p53 regulated genes, that we refer to as the “Andrysik Core” list. We have used the RNAseq data presented in this paper, together with the Andrysik Core list, as a p53 gene response reference gene list to compare with.

RNAseq analysis was carried out on the Nutlin treated (10 $\mu$ M, 24 h), replicate samples for HCT116 cells and these used in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines. The PCA analysis and quality control of these samples has already been described above. Using DESeq analysis, we identified 12673 Nutlin regulated genes ( $p < 0.05$ ) in the HCT116 wild-type line. These overlapped with the Andrysik, Nutlin regulated gene list ( $p < 0.05$ ) of 4363 genes, to show 4030 genes in common (Figure 5.14A). This confirmed the integrity of the Nutlin response in the HCT116 wild-type line we have used in our studies. Further, a similar number of Nutlin regulated genes overlapping with the Andrysik, Nutlin regulated gene list ( $p < 0.05$ ) for the HCT116 APOBEC3H 2B3-2B2 line (overlap of 3974



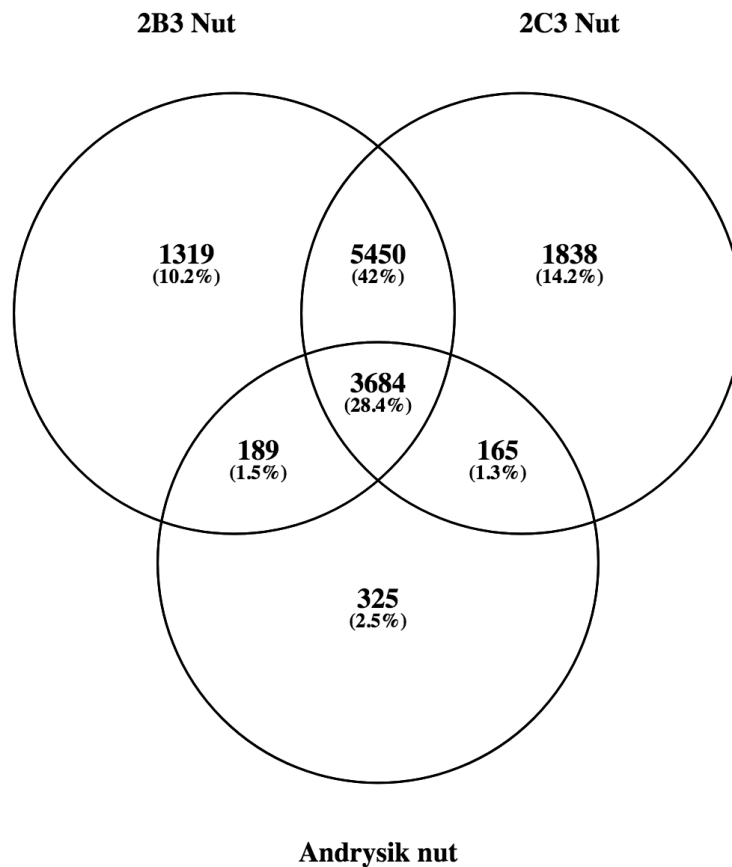
genes) (Figure 5.14B) and the HCT116 APOBEC3H 2C3-1B2 line (overlap of 3849 genes) (Figure 5.14C).

Taken together, these analyses suggest that the overall p53 response in the two APOBEC3H knockout cell lines is similar to that seen in the HCT116 wild-type cells. A three way comparison between the Nutlin regulated gene lists ( $p < 0.05$ ) for the two APOBEC3H knockout cell lines and the Andrysik, Nutlin regulated gene list ( $p < 0.05$ ), shows that in addition to those genes shared with the Andrysik, Nutlin regulated gene set, a further 5450 Nutlin regulated genes are shared between two APOBEC3H knockout cell lines, suggesting there are additional p53-Nutlin regulated genes in these cells, not accounted for in the Andrysik experiments (Figure 5.15).



**Figure 5.14. Evaluation of the Nutlin - p53 response in wild-type HCT116 cells and in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines**

The comprehensive study of Andrysik *et al.*, 2017, identified 4363 Nutlin regulated genes in HCT116 wild-type cells ( $p$  value  $<0.05$ ). We have used this gene list (Andrysik +Nut) to assess the overlap with differentially Nutlin regulated genes in our HCT116 cells and in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines. **A.** For HCT116 wild-type cells (HCT116Wt+Nut), 12673 Nutlin regulated genes were identified ( $p$  value  $<0.05$ ). Of these, 4030 (31%) overlap with the Andrysik HCT116 Nutlin regulated gene list. This overlap accounts for 92% of the Andrysik HCT116 Nutlin regulated gene set. The analysis also identifies an additional 8643 genes that are differentially regulated by Nutlin in HCT116 wild-type cells, which are not seen in the Andrysik *et al.*, 2017 Nutlin regulated gene set, as defined by a cut-off of  $p$  value  $< 0.05$ . **B.** A similar comparison of the Nutlin regulated genes ( $p$  value  $<0.05$ ) in the HCT116 APOBEC3H 2B3-2B2 identifies 3974 genes which overlap with the Andrysik HCT116 Nutlin regulated gene set, while for the APOBEC3H 2C3-1B2 line (2C3+nut), the number of overlapping genes is 3849 (**C**).

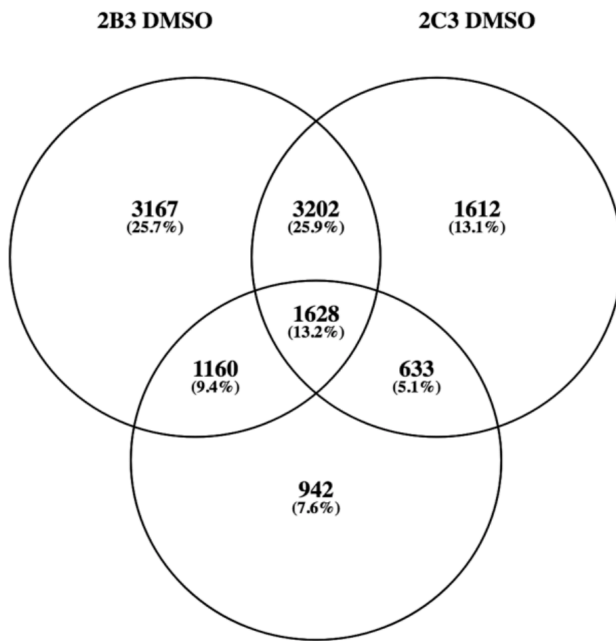


**Figure 5.15. Evaluation of the Nutlin - p53 response in wild-type HCT116 cells and in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines – three-way analysis**

A three-way comparison between the Nutlin regulated gene lists ( $p < 0.05$ ) for the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines and the Andrysik HCT116 Nutlin regulated gene set. 3684 genes in Andrysik study are also regulated in both HCT116 APOBEC3H knockout lines by Nutlin, with an additional 5450 genes that are Nutlin regulated, and shared between two mutant lines. Overall, 92.5% of the Andrysik HCT116 Nutlin regulated gene set are regulated by Nutlin in the both HCT116 APOBEC3H knockout lines. ( $p$  value  $< 0.05$ ).

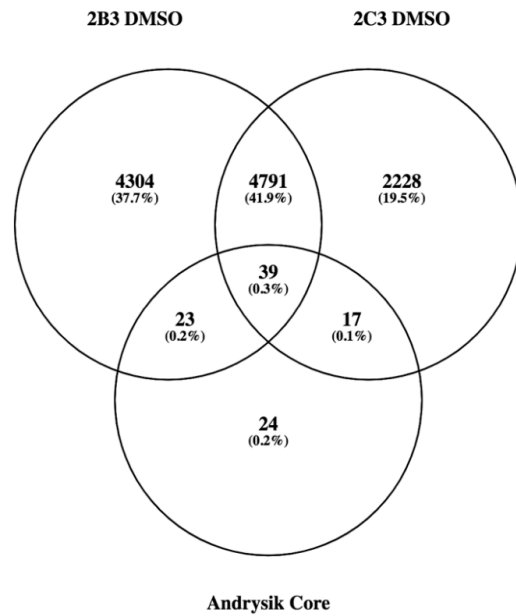
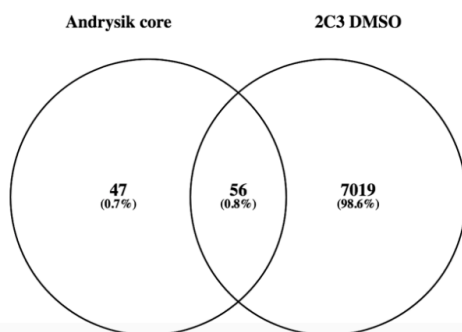
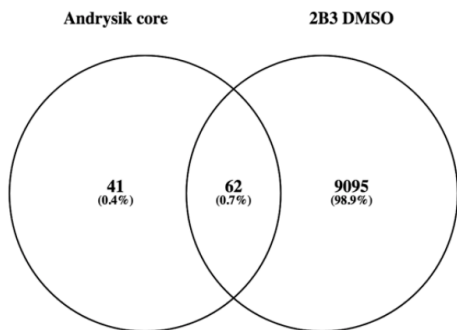
A three –way comparison of the overlap between the differential gene lists between the HCT116 APOBEC3H 2B3-2B2, HCT116 APOBEC3H 2C3-1B2 and HCT116 wild-type cells, shows that only 942 (7.6%) genes from the Andrysiak HCT116 – Nutlin gene set do not feature as differentially expressed genes in either of the two APOBEC3H knockout lines (Figure 5.16A). Similarly, only 24 genes out of the 103 Andrysiak Nutlin core regulated gene list, which represents a collated set of direct p53 targets, is not differentially expressed in either of the HCT116 APOBEC3H knockout lines (Figure 5.16 B).

**A**



**Andrysik Nut**

**B**



**Figure 5.16. The Nutlin - p53 gene signature is seen in HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines – three way and core p53 comparison**

**A.** Lists were compared for differentially expressed genes between wild-type HCT116 cells and each of the HCT116 APOBEC3H knockout lines (2B3 DMSO and 2C3 DMSO), and the Andrysik HCT116 Nutlin regulated gene set (Andrysik Nut). A total of 3420 (78 %) of the Andrysik HCT116 Nutlin regulated genes overlap with the differentially expressed genes between wild-type HCT116 cells and each of the HCT116 APOBEC3H knockout lines 2B3-2B2 and 2C3-1B2. Of these, 1628 genes (37%) of HCT116 Nutlin regulated genes overlap with the differential genes seen in both HCT116 APOBEC3H knockout lines 2B3-2B2 and 2C3-1B2. **B.** Andrysik *et al* (2017) also defined a 103-core gene list of Nutlin regulated genes, which were also collated for those which are direct targets for p53 binding by ChIPseq and show transcriptional activation by GROseq. Over half of the Andrysik core list are contained in HCT116 APOBEC3H knockout line differential gene expression lists (2B3 DMSO and 2C3 DMSO).

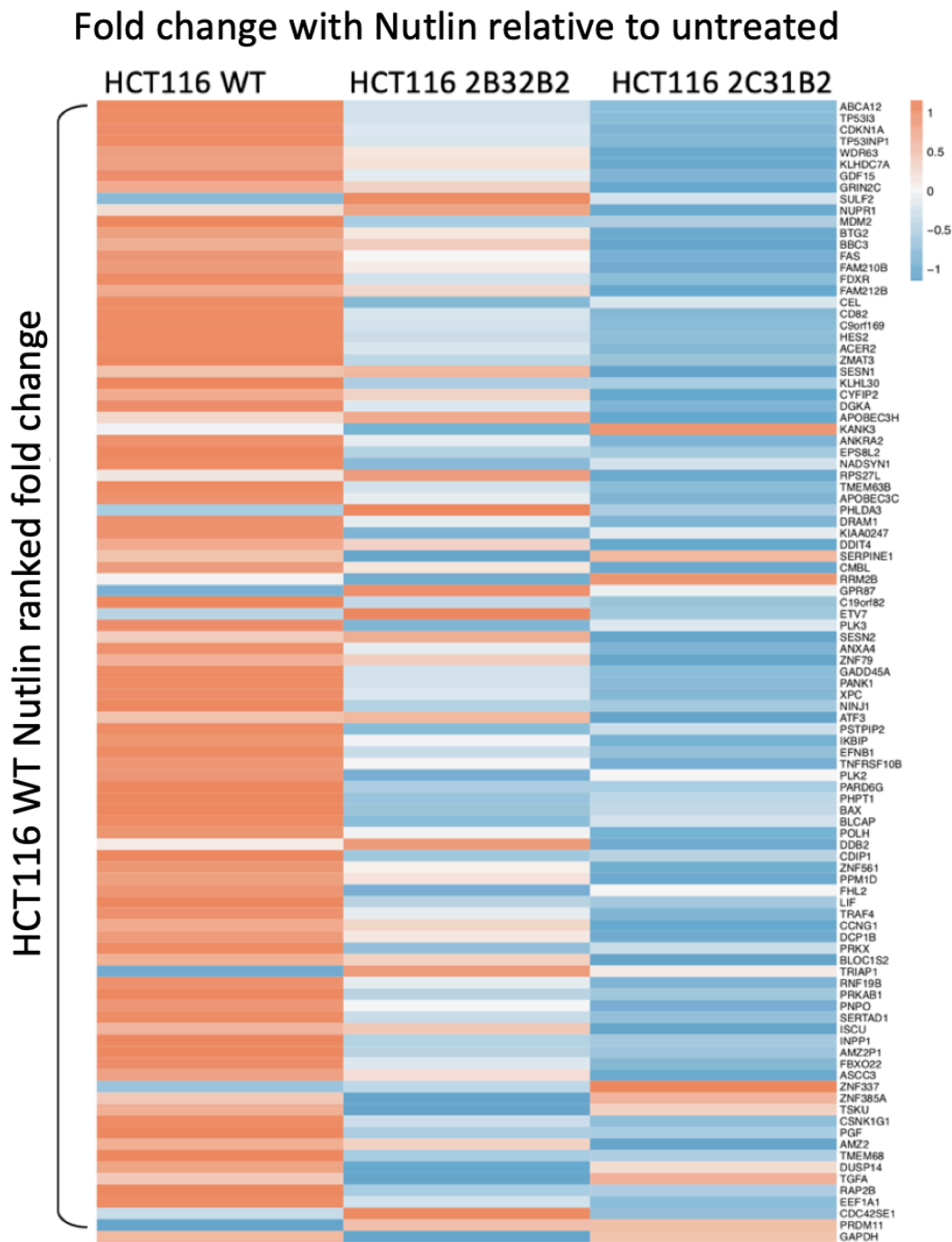
Taken together, these analyses suggest in the APOBEC3H knockout lines compared to wild-type cells many of the Nutlin modulated p53 responsive already show differential expression. In order to understand this further, we analysed the expression, as determined by read count from the RNAseq analysis, for the 103 Andrysiak core p53 regulated gene list in HCT116 wild-type and the HCT116 APOBEC3H 2B3-2B2, HCT116 APOBEC3H 2C3-1B2 lines. To do this, gene expression for each of the 103 genes in the Andrysiak core p53 regulated gene list was determined as a fold change with respect to the DMSO control, and then the genes ranked according to this for the Nutlin treatment in the HCT116 wild-type cells. A heat map was then generated to show the fold change following Nutlin treatment for the ranked genes, in the two APOBEC3H knockout cell lines. As can be seen in Figure 5.17, the heatmap analysis shows that the p53 response in the two APOBEC3H knockout lines is greatly attenuated, when compared to the Nutlin response seen in the HCT116 wild-type control cells. This conclusion was further validated using Q-RT-PCR for a panel of p53 regulated genes from the Andrysiak core gene set, and also for p53. As can be seen, in Figure 5.18, Nutlin treatment results in a decrease in p53 gene expression, which is seen to be significantly lower in the two APOBEC3H knockout lines, when compared to the level in Nutlin treated HCT116 wild-type cells. Further, for a set of six p53 target genes, including APOBEC3H, the Nutlin induced, p53 mediated activation of gene expression was seen to be attenuated in most.

These data suggest the possibility that APOBEC3H could potentially attenuate the Nutlin induced, p53 mediated gene response in HCT116 cells. In order to investigate this further, HCT116 wild-type cells were stably transfected with an APOBEC3H expression plasmid (gift of Dr Lai Chun Fui, Imperial College London) and the Nutlin induced gene expression response in these cells evaluated. As can be seen in Figure 5.19, when compared to control, empty vector HCT116 stably transfected cells, this results in a 10 fold increase in APOBEC3H

gene expression, which would be expected to mask the induction of the endogenous APOBEC3H gene by Nutlin in these cells. While this over-expression has little effect on the basal expression of the p53 regulated genes p21, MDM2 and BTG2, in each case there is a higher level of p53 induced gene expression in the APOBEC3H over-expressing cells. This effect is particularly marked for MDM2, where there is a greater than 60 fold induction of MDM2 over the basal levels seen in the APOBEC3H over-expressing cells.

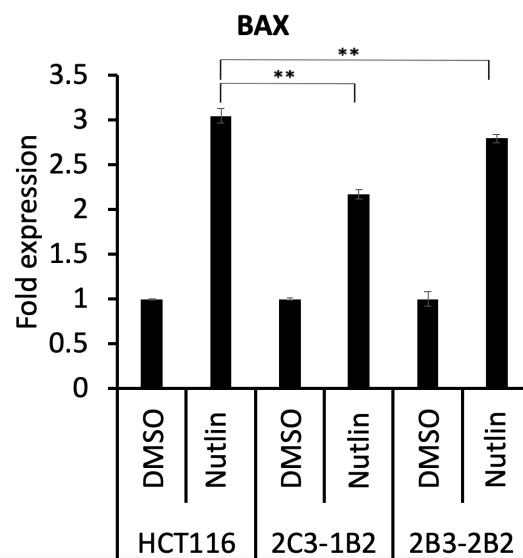
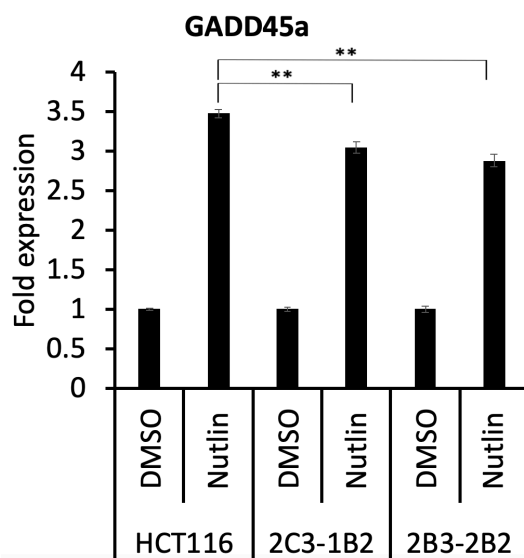
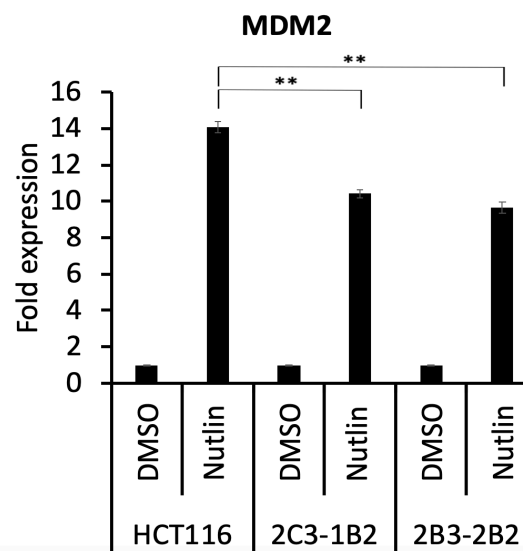
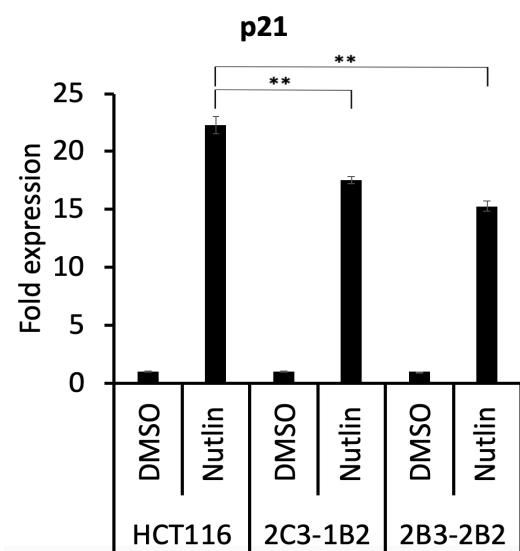
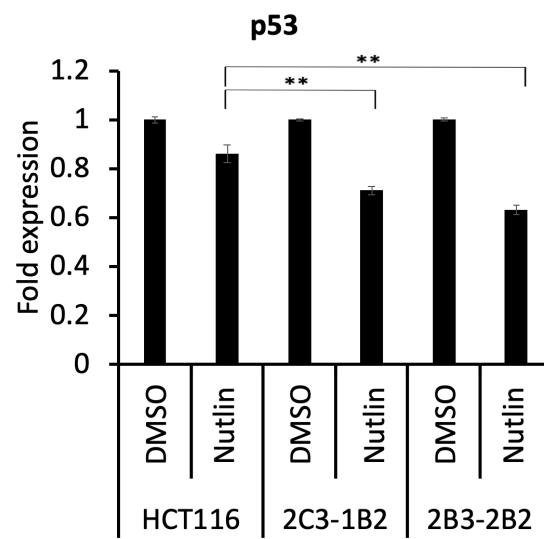
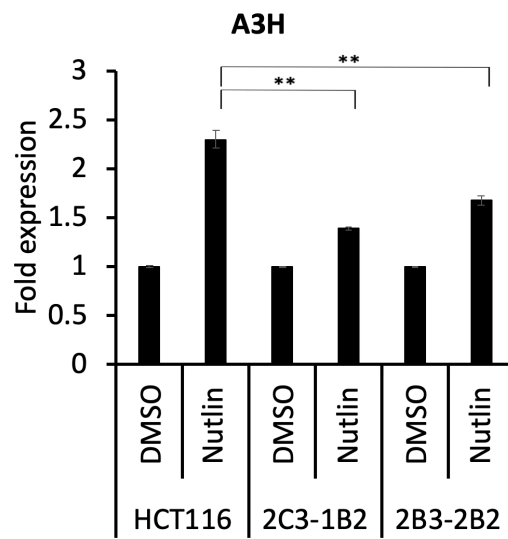
Taken together, these data suggest a role for APOBEC3H in mediating the strength of the p53 activated response and would define a new role for APOBEC3H.

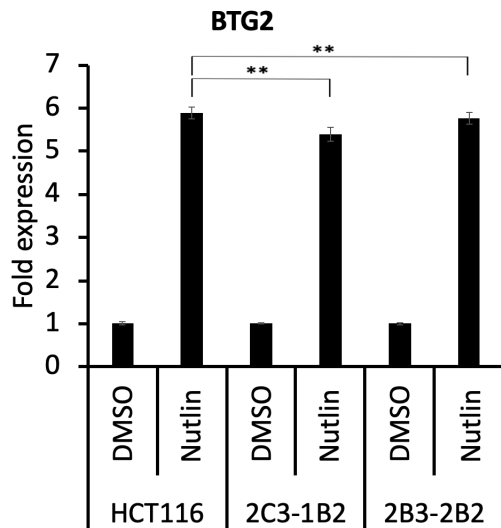




**Figure 5.17. Comparative Nutlin regulated gene expression between HCT116 cells and in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines**

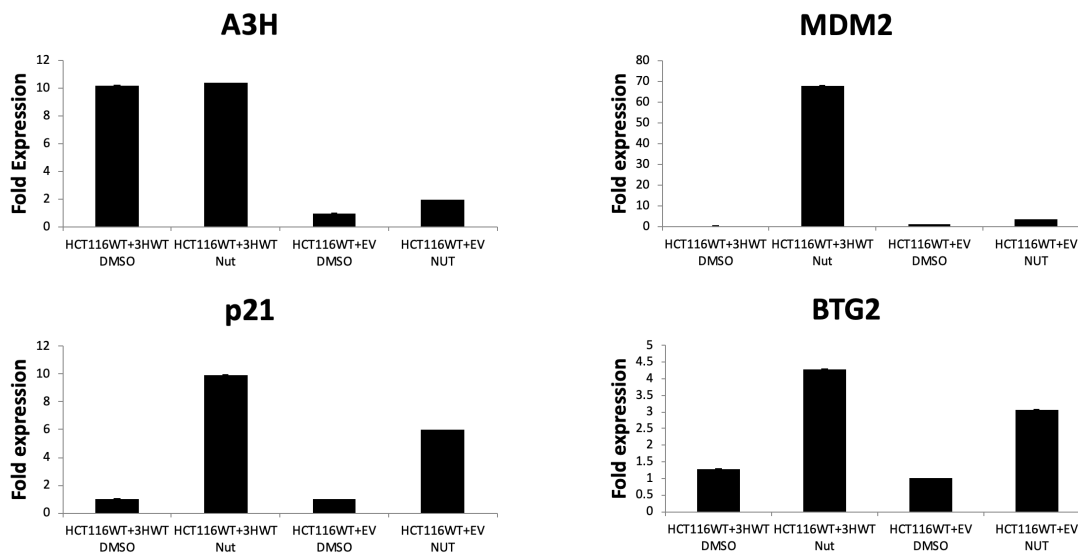
Heatmap analysis was used for the Andrysic core gene list (103 genes), in order to compare the fold change for these genes in wild-type HCT116 cells, and in the HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines, treated with Nutlin.





**Figure 5.18. Gene expression for p53 and p53 target genes in HCT116 APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 cell lines**

RNA prepared from six bioreplicate cultures of wild-type HCT116 cells and APOBEC3H knockout HCT116 lines 2B3-2B2 and 2C3-1B2 maintained in full medium, or in medium supplemented with 10 $\mu$ M Nutlin was pooled in equal amounts and used to make cDNA representative of each line and treatment. This cDNA was used to measure gene expression, which was normalised to GAPDH housekeeping gene expression and shown as fold expression relative to the matched control sample, with standard errors of the mean. Significant differences ( $P < 0.05$ ) in Nutlin induced gene expression, as measured by paired T-tests, are indicated by “\*\*”. Further, T-test is statistically significant between DMSO and Nutlin in each line. The analysed genes are p53 and the p53 target genes, A3H, p21, MDM2, GADD45a, BAX and BTG2. DMSO treated lines normalized to 1 fold.



**Figure 5.19. Nutlin induced gene expression following APOBEC3H over-expression in wild-type HCT116 cells**

HCT116 wild-type cells (HCT116 WT) were stably transfected with an APOBEC3H expression plasmid (3HWT; gift of Dr Lai Chun Fui, Imperial College London) and empty plasmid vector (EV) and the expression of Nutlin induced gene expression responses in these cells evaluated. Cells were nucleofected with plasmid DNA and selected with the cytotoxic antibiotic G418. The selected cultures were subsequently treated with 10 $\mu$ M Nutlin, or DMSO vehicle control, for 24 h and then used to extract RNA. This RNA was used in Q-RT-PCR analysis, with expression being normalized against GAPDH housekeeping gene expression and shown as fold against the DMSO treated, EV control. As can be seen, there is around ten-fold overexpression of APOBEC3H following stable transfection with the APOBEC3H expression plasmid. Further, over-expression of APOBEC3H in these cells results in stronger activation of the Nutlin/ p53 regulated genes BTG2, p21 and MDM2.

## 5.4 - Discussion

In order to gain further insight into the phenotype of HCT116 APOBEC3H knockout cells an RNAseq analysis of the two knockout lines APOBEC3H 2B3-2B2 and APOBEC3H 2C3-1B2 was performed. The recent studies by Lamarre *et al.*, (2018) and Liu *et al.*, (2014) have evaluated the impact of sequence depth and the number of RNAseq replicates used, and concluded that the number of replicates has a higher impact on the number of differentially expressed genes that are detected, than the overall sequencing depth. The use of six replicates for each cell line and treatment, together with a reading depth of 64.7 million reads per sample in our study, are both well above the minimal effective replicate numbers and depths cited in these studies. Principal Component Analysis of the resulting RNAseq read data identified only one sample, HCT116 2B3-2B2 replicate 1, as a potential outlier. As the studies by Lamarre *et al.*, (2018) and Liu *et al.*, (2014) suggested that reducing the number of replicates from six to five would not impact greatly on the detection of differentially expressed genes, our analysis proceeded by using the data from only five of the bioreplicates for each set, with HCT116 2B3-2B2 replicate 1 not used.

### 5.4.1 - Differential gene expression in the HCT116 APOBEC3H knockout lines and loss of the Y chromosome

The first part of the RNAseq analysis carried out in this study focussed on the identification of differentially expressed genes that result from the knockout of APOBEC3H in the HCT116 cell line. Overall, this identified a total of 4830 genes that were differentially expressed and shared in the two APOBEC3H knockout cell lines. Initially, this gene list was ranked according to gene expression, leading to the identification of strongly repressed and strongly activated genes in the APOBEC3H knockout lines. As can be seen in Table 5.2, the down-regulated

genes were dominated by genes encoded on the Y chromosome. Loss of the Y chromosome from human cancer cell lines has been noted previously (Honma *et al.*, 1996) , and has even been noted for stocks of HCT116. A more elaborate study of 600 Human cancer cell lines found that on prolonged passaging, many lines lost either the Y or the inactive X chromosome (Xu *et al.*, 2017). However, even though the Y chromosome encodes less than 30 genes (Helena Mangs and Morris, 2007), the loss of the Y chromosome may be expected to have a considerable impact on cellular phenotypes and gene expression. For example, the loss of the Y chromosome is a common feature of Renal Cell Carcinoma, with the loss of KDM5D and KDM6C, two Y chromosome encoded Lysine Methyl Transferases being of particular importance (Arseneault *et al.*, 2017). In this respect, it is not entirely clear which of the 4830 genes that were differentially expressed and shared in the two APOBEC3H knockout cell lines indirectly result from loss of the Y chromosome in these lines.

In addition to the loss of Y chromosome gene expression, other strongly down regulated genes were identified. These are dispersed throughout the genome, so were unlikely to result from further loss of genomic loci. While these genes were not analysed systematically, the magnitude of the changes predicted by RNAseq analysis were essentially validated by the Q-RT-PCR analysis. The validated genes included: PXDN, a gene associated with epithelial-to-mesenchymal transition; GALNT14 a gene involved in organ-specific metastasis (Song *et al.*, 2016); SPARC1, a gene encoding a extracellular protein involved in cancer and metastasis (Viloria *et al.*, 2016) ; HKDC1 a gene encoding a hexokinase that mediates metastasis (Chen *et al.*, 2019); TCF21 (Qin *et al.*, 2018) , a tumour suppressor.

Taken together, the RNAseq analysis indicates that there are substantial numbers of genes whose expression is lost / strongly down-regulated in the HCT116 APOBEC3H knockout cells,

some of which arise from loss of the Y chromosome, and others whose expression may require APOBEC3H activity.

#### **5.4.2 - Differential gene expression in the HCT116 APOBEC3H knockout lines and processed pseudogene expression**

Analysis of up-regulated gene expression in the HCT116 APOBEC3H knockout cells unexpectedly highlighted processed pseudogenes. Processed pseudogenes are non-autonomous, retrotransposed sequences, where cellular RNAs have been converted into cDNA through the reverse transcriptase produced by long interspersed elements (Esnault *et al.*, 2000). While long interspersed element mobilisation was thought to be restricted to early embryonic development (Klein and O'Neill, 2018), recent studies have suggested they may play an important role in generating insertional mutagenesis and genomic instability in cancer (Burns, 2017). Indeed, it has been recently observed that processed pseudogenes themselves can be acquired somatically, particularly during tumor evolution (Cooke *et al.*, 2014).

Transcribed processed pseudogenes represent a sub-class of pseudogenes, which as a result of their transcription may act functionally (Harrison *et al.*, 2005). However, as reverse transcribed, retrotransposed sequences, these types of pseudogenes are usually thought to be transcriptionally silent, unless they become incorporated into transcriptionally active loci. Given the apparent widespread distribution of processed pseudogenes, this possibility is more likely than previously thought (Torrents *et al.*, 2003).

The link between APOBEC3H and processed pseudogene expression is not clear. While APOBEC3H is known to restrict retrotransposable elements (OhAinle *et al.*, 2008) this activity

may not necessarily require the deaminase activity of the protein (Feng *et al.*, 2017). Further, RNAseq analyses, as carried out here, is not informative for retrotransposable element mobilisation, as these sequences do not feature in the reference genomes used for analysis. Hence, it is not possible to use the data to suggest that the knockout lines feature activation of retrotransposable element across the genome, which could result in the activation of transcribed pseudogene sequences, also integrated across the genome. Such studies are further complicated by the fact that, due to their homology to their progenitor genes, processed pseudogene expression is not easily validated. Current tools include the use of Next Generation Sequencing in conjunction with gene analysis methods (Zheng and Gerstein, 2006) and resources, such as the GENCODE pseudogene resource (Pei *et al.*, 2012).

#### **5.4.3 - Differential gene expression in the HCT116 APOBEC3H knockout lines and p53 regulated gene expression.**

In the first part of this thesis, we carried out work to show that APOBEC3H is a p53 regulated gene (Periyasamy *et al.*, 2017) The further development of the HCT116 APOBEC3H knockout cells and subsequent analysis by RNAseq identified an enrichment for Nutlin-p53 regulated genes in the set identified as differentially expressed between the knockout lines and wild-type HCT116. This suggested the possibility that the p53 response in the APOBEC3H knockout cells was modified, a conclusion that was supported by the RNAseq analysis of the Nutlin regulated- p53 gene regulated response in the HCT116 APOBEC3H knockout cells. Here it was seen that the Nutlin regulated- p53 gene regulated response was blunted, as confirmed by Q-RT-PCR validation experiments for a panel of p53 regulated genes. The role for



APOBEC3H in this phenotype was further indicated by the demonstration that over-expression of the gene in HCT116 cells enhanced p53 regulated gene expression.

The exact mechanism by which APOBEC3H can modulate p53 regulated gene expression is not clear. Previously, we have shown that APOBEC3B is a co-activator for the estrogen receptor, helping to activate estrogen regulated gene expression(Periyasamy *et al.*, 2015). On this basis, it may be the case that APOBEC3H is also a transcriptional co-activator, but of p53. In the case of APOBEC3B, the cytidine deaminase activity was shown to be important, as it allowed deamination at estrogen receptor responsive promoters, leading to localized DNA repair and, potentially, chromatin remodeling. In the same way, APOBEC3H, through targeted deamination could be involved in the activation of p53 regulated gene expression. The mechanistic involvement of APOBEC3H in gene regulation in this manner would also be potentially important in the generation of tumor mutations (Starrett *et al.*, 2016).

## 6 – Discussion

### 6.1 - APOBEC3H and p53

#### 6.1.1 - APOBEC3H as a mediator of retroviral resistance

APOBEC3H is a member of the APOBEC3 sub-family of cytidine deaminase enzymes. The APOBEC3 family, as a whole, are known to mediate antiviral activity, by causing lethal G-to-A hypermutation in retroviruses and retrotransposable elements through the conversion of deoxycytosine to deoxyuracil in newly reverse-transcribed, single-stranded cDNA. *In vitro* cell transfection experiments show anti-HIV-1 activity for all APOBEC3 proteins (Larue *et al.*, 2010). However, *in vivo*, when cellular expression and ability to encapsidate into viral particles is taken into account, only APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H show activity, and are thought to be key in HIV-1 retroviral resistance (Hultquist *et al.*, 2011).

Primates encode seven APOBEC3 proteins, with APOBEC3F, APOBEC3G, and APOBEC3H being antagonised by the HIV-1 encoded viral Vif accessory protein (Krisko *et al.*, 2016; Mariani *et al.*, 2003; OhAinle *et al.*, 2006; Virgen and Hatzioannou, 2007). APOBEC3H stands out from the APOBEC3F and APOBEC3G, in three ways. Firstly, the enzyme encodes a single Zn-coordinating domain, located in the deaminase catalytic centre, which is unlike APOBEC3D, APOBEC3F and APOBEC3G, which encode two single Zn-coordinating domains (Refsland and Harris, 2013; Xiao *et al.*, 2016). The second distinguishing feature for APOBEC3H, is the recent finding that the enzyme undergoes dimerization, mediated by the binding of small, double stranded RNAs which are derived from retroviral origin (Bohn *et al.*, 2017; Feng *et al.*, 2018; Ito *et al.*, 2018; Matsuoka *et al.*, 2018; Shaban *et al.*, 2018). The RNAs are GC-rich palindrome-like sequences of retroviral origin (Matsuoka *et al.*, 2018), and

it has been shown that RNA mediated dimerization is required for APOBEC3H cytoplasmic localisation (Shaban *et al.*, 2018). The dimerized structure is unusual as the two APOBEC3H molecules bind at the opposite ends of the seven-base-pair duplex RNA, interacting with both RNA strands, and essentially becoming tethered, but without necessarily making protein–protein contacts, as most clearly seen in the studies of Bohn *et al.*, (2017) and Feng *et al.*, (2018)

The third distinguishing feature of human APOBEC3H, compared to other APOBEC3 genes, is the large number of polymorphisms, and resulting haplotypes (Wang *et al.*, 2011). Among the seven APOBEC3H haplotypes, types III, IV and V result in unstable proteins, largely due to the absence of the N15 residue in the N15 $\Delta$  polymorphism (Li and Emerman, 2011; Wang *et al.*, 2011; Zhen *et al.*, 2012). Haplotypes I and II are common and have distinct sub-cellular localisations (Li and Emerman, 2011). Haplotype I protein is mostly found in the nucleus, while Haplotype II is mostly localised to the cytoplasm, with cytoplasmic localisation being dominant over nuclear localisation (Li and Emerman, 2011). In this study, Haplotype II was found to be actively retained in the cytoplasm, through interaction with specific host factors, while Haplotype I enters the nucleus by a passive mechanism. In this context, as dsRNA-mediated dimerisation of APOBEC3H also tightly regulates antiviral activity and localisation, both APOBEC3H haplotypes and effective RNA dimerisation are key factors in anti-viral activity. Conversely, haplotype I, non-RNA bound, monomer, APOBEC3H has been proposed to be a candidate for mutation of genomic DNA in cancer (Shaban *et al.*, 2018; Starrett *et al.*, 2016). Finally, of the two, haplotype I has been shown to be the most stable (OhAinle *et al.*, 2008). It is interesting to note this work has highlighted the C-terminal region of the protein as being important in stability and sub-cellular localisation, raising the possibility that the

APOBEC3H variant 208, the new splice variant found in our study, could potentially define a form with altered stability and localisation. Clearly, this requires further investigation.

### **6.1.2 - APOBEC3H in the inhibition of retrotransposon mobilisation**

Retrotransposons, or transposable elements are mobile DNA elements that account for 46% of the human genome (Bodak *et al.*, 2014). They are divided into two classes, based on their mechanism of transposition: 1. Class I retrotransposons, which like retroviruses, require an RNA intermediate and a ‘copy-and-paste’ mechanism. These represent about 42% of the human genome; 2. Class II DNA transposons, which move by a ‘cut-and-paste’ mechanism (Bodak *et al.*, 2014). The long-interspersed nuclear elements-1 (LINE-1) are the major type of Class I retrotransposable elements in the human genome, and which encode a 150 kDa protein reverse transcriptase required for converting the LINE-1 genomic RNA to a cDNA intermediate (Bodak *et al.*, 2014). As they constitute around 17% of the genome, LINE-1 elements have been found to make a major contribution to the overall transcriptome of somatic cells (Rangwala *et al.*, 2009).

Retrotransposons have been considered to encode junk DNA and, because of recombination insertional mutagenesis, are also considered to be harmful, parasitic elements, that can cause hereditary diseases (Mukherjee *et al.*, 2004), and act as a source of mutation in cancer (Rodic and Burns, 2013; Xiao-Jie *et al.*, 2016). For these reasons, and for the reason of genome evolution and gene regulation in the genomic landscape, retrotransposon expression is tightly regulated and predominantly repressed by DNA methylation (Beck *et al.*, 2010; Hata and Sakaki, 1997; Kazazian, 2000; Steinhoff and Schulz, 2003; Yoder *et al.*, 1997). As a result,

activation of LINE-1 expression has been seen in cancer, where hypomethylation is a feature of global epigenetic changes (Hagan and Rudin, 2002; Honda and Rahman, 2019; Kerachian and Kerachian, 2019; Torano *et al.*, 2012). It is likely that it is under these conditions that the ability for APOBEC3H to restrict LINE-1 mobilisation is important, although several studies have suggested this may not require deaminase activity (Feng *et al.*, 2017; Stenglein and Harris, 2006).

### **6.1.3 - p53 induction of APOBEC3H as a mechanism to inhibit retroviral and retrotransposable propagation**

Through the use of CRISPR mediated genome engineering, we have developed a p53 null derivative of MCF7, and used this to confirm the p53 regulation of APOBEC3H. We and others have also noted that APOBEC3H expression is strongly induced by the p53 activator Nutlin, in both MCF7 cells and in the colon cancer cell line HCT116 (Andrysik *et al.*, 2017). These results suggest the possibility that p53, through APOBEC3H activation, may be able to protect against retroviral and retrotransposable propagation. In considering this, the literature is more supportive of a role for p53 in regulating retrotransposition than inhibiting retrovirus replication. For antiviral activity, it has been found that cytokine signalling potentiates p53 responses, leading to synergy between cytokine and p53 signalling (Takaoka *et al.*, 2003), but p53 mediated retroviral restriction has not been noted. With regards to p53, it has been found that p53 does function to block retrotransposition (Wylie *et al.*, 2016), and LINE-1 methylation levels are seen to be affected by p53 mutation status (Shin *et al.*, 2019). However, the picture is more complicated, in that it has been noted that p53 binding sites can be found in transposable element sequences (Zemojtel and Vingron, 2012) and can lead to the increased

transcription of particular LINE-1 elements (Harris *et al.*, 2009). While sounding counter-intuitive, (Harris *et al.*, 2009) suggest this activation is protective, in that it also gives rise to elevated levels of DNA damage, through the increased nuclease activity encoded by the LINE-1 ORF1 gene, resulting in p53 mediated cell cycle arrest, and subsequent apoptosis. In addition to these mechanisms, it is now possible to propose that p53 induction of APOBEC3H also helps to protect against retrotransposition, by inhibiting LINE-1 replication. While we have not analysed the RNAseq data from the HCT116 APOBEC3H knockout lines for evidence of LINE-1 transcriptional activation, preliminary Q-RT-PCR experiments to compare LINE-1 expression in HCT116 wild-type and HCT116 APOBEC3H lines were attempted, but the results from these analyses were not clear enough to make conclusions. Despite this, it is also possible that the increased expression of processed pseudogenes in the HCT116 APOBEC3H knockout lines, as observed by the RNAseq analysis, may be indicative of transcription of normally silent regions of the genome, which would include activation of retrotransposon expression in the HCT116 APOBEC3H knockout lines. Further analysis of the expression of retrotransposons and processed pseudogenes in the HCT116 APOBEC3H knockout is clearly required and would be aided by new experiments to assess retrotransposon and processed pseudogene expression in the MCF7 Luc p53 knockout line.

In summary, the anti-retroviral and retrotransposable activities of APOBEC3H are well documented and the p53 induction of APOBEC3H may be indicative of a role for the enzyme in the p53 regulation of retrotransposable elements. Clearly, the p53 MCF7 luc knockout and HCT116 APOBEC3H knockout cell lines described here are new tools that could be used to further investigate this possibility.

#### **6.1.4 - APOBEC3H in the p53 response**

The RNAseq based differential gene expression analysis of the HCT116 APOBEC3H knockout cells, compared to wild-type HCT116 cells, identified a set of 4830 genes differentially expressed genes, with a cut-off of P value <0.05. This analysis highlighted a genotypic difference between the wild-type and APOBEC3H knockout cells, namely the loss of the Y- chromosome in both HCT116APOBEC3H knockout cell lines. Potentially, this difference complicates the interpretation of the two knockout lines. Nevertheless, the analysis has been used to validate the difference in gene expression between the wild-type and APOBEC3H knockout cells, which included a set of Y chromosome encoded genes and additional up, and down-regulated genes. The differential gene expression analysis was further refined by employing a cut-off for expression, which allowed a list of APOBEC3H knockout differentially expressed genes to be used in Gene Set Enrichment Analysis, or GSEA (Mathur *et al.*, 2018). This analysis identified p53 regulated genes as being enriched in the differentially expressed gene set and raised the possibility that p53 regulated gene expression is affected in the HCT116 APOBEC3H knockout cells. It was possible to further investigate this by RNAseq analysis to assess the Nutlin-p53 response in the HCT116 APOBEC3H cells.

A comparison of the Nutlin induced differential gene expression in the HCT116 wild-type and HCT116 APOBEC3H knockout cells by RNAseq showed that the gene response to Nutlin was very similar across all three lines, and also with the previously described Nutlin induced, differentially expressed gene set in HCT116 described by Andrysik *et al.* (2017) The analysis also suggested that loss of the Y chromosome did not overtly change the Nutlin-p53 response. However, the analysis did highlight that p53 regulated gene expression in the Nutlin treated HCT116 APOBEC3H cells was blunted. Further work showed that Nutlin induced p53 regulated gene expression was stimulated in HCT116 APOBEC3H over-expressing cells., indicating the possibility that APOBEC3H may have a role as a transcriptional co-activator in

p53 regulated gene expression. This would parallel earlier work from this group that showed APOBEC3B is a co-activator for Estrogen Receptor regulated gene expression in MCF7 breast cancer cells (Periyasamy *et al.*, 2015). This has led us to the model for APOBEC3H in p53 regulated gene expression. In this, p53 activation leads to APOBEC3H gene expression, which then provides a positive feedback, leading to enhanced p53 regulated gene expression. The induction of p53 can come from a variety of signals, including DNA damage caused by LINE-1 activation, as discussed above.

The possibility that p53 can be co-activated through a cytidine deaminase activity is novel and has not been indicated by previous work on p53 activation and co-activation. As described earlier, p53 is repressed by MDM2 which interacts with, and blocks the N-terminal region of the protein and also promotes degradation through ubiquitination (Sullivan *et al.*, 2018). Subsequently, the transcriptional activity of p53 is subject to modulation by a variety of post translational modifications, including phosphorylation and acetylation, some of which brings about target gene specificity (Beckerman and Prives, 2010). Once bound to DNA, p53 transcriptional activity is further dependent on the interaction with a wide variety of transcriptional co-activators, including Histone Acetyl Transferases and Histone Methyl Transferases (Liu and Chen, 2006). In recruiting proteins to the p53 transcriptional complex, some interactions are direct, while some are indirect through other components of the complex (Kim and Lozano, 2018; Sullivan *et al.*, 2018). For example, a search of the BioGRID protein-protein interaction database (Oughtred *et al.*, 2019) identified 1151 proteins that interact with p53, none of which were APOBECs, while a search for APOBEC3H only identified one interacting protein, the amyloid beta (A4) precursor protein. Collectively, these data suggest that in acting to co-activate p53 transcription, there may not need to be a direct interaction between p53 and APOBEC3H, which is different to the APOBEC3B co-activation of Estrogen



Receptor activity, where direct interaction is seen (Periyasamy *et al.*, 2015). Clearly, the role of APOBEC3H in the p53 response requires further investigation.

## **6.2 - CRISPR-Cas9 genome engineering**

The discovery of the CRISPR-Cas microbial adaptive immune system and its further development into a genome editing tool has been rapid. The discovery of CRISPR elements in bacterial genomes, and that these encoded sequences of bacteriophage origin was first reported by (Mojica *et al.*, (2005) and Pourcel *et al.*, (2005). This was closely followed by the identification of the Cas nucleases (Makarova *et al.*, 2006), and the subsequent demonstration that both RNAs transcribed from the CRISPR loci and the Cas nuclease was required for the adaptive phage immunity (Barrangou *et al.*, 2007; Brouns *et al.*, 2008). Until then, the CRISPR-Cas system was thought to target RNA. However, further characterisation demonstrated that the system targeted DNA (Marraffini and Sontheimer, 2008), with the exact mechanism of CRISPR-Cas DNA cleavage being shown in 2010 (Garneau *et al.*, 2010). The following year, Deltcheva *et al.*, (2011) identified the trans-activating CRISPR RNA (tracrRNA) as the final component of the CRISPR-Cas microbial adaptive immune system.

In the years that followed the CRISPR-Cas system was further developed by engineering. Firstly, In 2012, Gasiunas *et al.*, (2012) and Jinek *et al.*, (2012) reported that CRISPR guide RNAs and the tracrRNA could be fused together to create a functional, synthetic guide RNA (sgRNA). In 2013, Cong *et al.*, (2013) and Mali *et al.*, (2013) described the first use of the CRISPR-Cas9 system for genome editing in eukaryotic cells, for which humanised Cas9 (hCas9) and sgRNAs were used. The depositing of CRISPR sgRNA backbone and hCas9

expression plasmid vectors in the Addgene plasmid repository (Kamens, 2015) from the groups responsible for these two papers, allowed others to evaluate the technology. Our group to started using the technology soon afterwards, to make the p53 and APOBEC3H (Periyasamy *et al.*, 2017; Smith *et al.*, 2018) knockouts described here, as well as work leading to one of the earliest reports for making CRISPR-Cas9 mediated gene knockins in human breast cancer cells. In the case of the p53 and APOBEC3H knockouts, a strategy of using a pair of CRISPRs to bring about DNA deletion, as opposed to using single CRISPRs to cause indel mutations was employed, as this was primarily easier to monitor by using genomic PCR assays. In addition, our strategy has used transient transfection (nucleofection), as opposed to stable selection, which has the advantage of not producing off-target changes due to plasmid DNA integration. Having established the approach, laboratory have since been able to extend it to making gene knockouts for a series of other genes, including the retinoblastoma tumor suppressor gene, which is wild-type in MCF7 (Buluwela, Ali, and Sava unpublished data).

The CRISPR-Cas system has great utility. The use of the alternative type VI-D CRISPR-CasRx system allows the targeting of RNAs (Konermann *et al.*, 2018). Further the development of a nuclease deficient Cas9 (dCas9), and the fusion of this to different effector functions such as transcriptional activation, transcriptional repression and DNA editing activities have provided a useful “tool box” for the targeted regulation of genes, and additional approaches for gene editing (Patsali *et al.*, 2019). There have also been limited developments in enhancing the utility of CRISPR guide RNAs, where short sequence motifs that allow RNA-protein interactions have been incorporated into loops in the guide RNA structure to bring additional proteins to CRISPR-Cas complex (Konermann *et al.*, 2015). These and other developments in CRISPR-Cas technology are likely to be useful in the further understanding of p53, APOBEC3H and the p53-APOBEC3H pathway.

### 6.3 - Summary and future work

In this study, we have investigated whether the APOBEC3H cytidine deaminase gene is a p53 regulated target and found that knockout of APOBEC3H leads to an attenuation of the p53 gene regulated response. The use of CRISPR-Cas9 genome editing has been central to this. There are three areas of work that may be developed through this:

#### 1. The further characterisation of the p53 knockout phenotype in MCF7 cells.

The generation of the p53 knockout MCF7 Luc cells is novel, and potentially useful for studying the p53 response in breast cancer cells. On this basis, further characterisation of the line would be useful to the field. Ideally, this would include the application of bioinformatics, and various “omics” approaches, including RNAseq analysis. The line is currently subject to further work and has been shared with collaborators.

#### 2. Further investigation of the APOBEC3H knockout phenotype in HCT116 cells.

A more in-depth study of the consequence of the APOBEC3H knockout on the transcriptome of HCT116 cells requires an additional assessment of the consequence of the loss the Y chromosome that accompanied the loss of functional APOBEC3H in our two lines HCT116 APOBEC3H 2B3 and HCT116 APOBEC3H 2C3. To allow this to be carried out, we have recently cloned wild-type HCT116 cells that lack the Y chromosome. This would provide a more appropriate control to assess gene expression in the HCT116 APOBEC3H 2B3 and HCT116 APOBEC3H 2C3 lines. Work to carry out such analysis has been initiated. The resulting, revised RNAseq work would also include an analysis of LINE-1, retrotransposon activity, to ascertain if this has been affected by APOBEC3H knockout.

### 3. Further investigation of the p53 response in HCT116 APOBEC3H 2B3 and HCT116 APOBEC3H 2C3 lines

The study of the p53 response in the two APOBEC3H knockout lines, HCT116 APOBEC3H 2B3 and HCT116 APOBEC3H 2C3 needs to be further developed. so that an understanding of the roles for p53 and APOBEC3H in this can be better understood. Firstly, the p53 response in both wild-type and APOBEC3H knockout HCT116 cells needs to be studied in a time-course following Nutlin treatment, so as to determine if the blunting seen in the p53 response in the APOBEC3H knockout lines is a reflection of temporal changes. Next, APOBEC3H needs to be expressed in the APOBEC3H knockout lines, so as to determine if the effect on p53 mediated gene expression can be rescued. This has been attempted using transfection of APOBEC3H plasmid expression vectors, but with little success- over expression could only be achieved in the wild-type HCT116 cells. While the reason for this failure was not clear, it would be useful to try again, this time using lentiviral transduction approaches, which can potentially be more successful in gene transfection. In doing this, it would also be prudent to compare the activities of APOBEC3H haplotypes, and in particular Haplotype I and Haplotype II, which constitute the original HCT116 APOBEC3H genotype. Finally, the re-introduction of wild-type and mutant APOBEC3H into the HCT116 APOBEC3H knockout cells provides a useful background to study the possible functional interaction between APOBEC3H and p53. In these cells, studies using p53 reporter gene plasmids that parallel our earlier work with APOBEC3B and estrogen receptor would also be possible (Periyasamy *et al.*, 2015).

In summary, the work reported here identifies APOBEC3H as a p53 regulated cytidine deaminase, which itself may act as a co-activator of p53 regulated gene expression. This adds a new dimension to the accepted role for APOBEC3H, namely the inhibition of HIV

retroviruses and the mobilization of retrotransposable elements, and emphasizes the possibility that the p53 regulation of APOBEC3H may play a role in maintaining genome integrity by preventing retrotransposable mediated DNA damage and insertional mutagenesis.

## References

- Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nat Commun* 9, 1911.
- Adolph, M.B., Love, R.P., and Chelico, L. (2018). Biochemical Basis of APOBEC3 Deoxycytidine Deaminase Activity on Diverse DNA Substrates. *ACS Infect Dis* 4, 224-238.
- Akre, M.K., Starrett, G.J., Quist, J.S., Temiz, N.A., Carpenter, M.A., Tutt, A.N., Grigoriadis, A., and Harris, R.S. (2016). Mutation Processes in 293-Based Clones Overexpressing the DNA Cytosine Deaminase APOBEC3B. *PLoS One* 11, e0155391.
- Alexandrov, L. (2018). The Repertoire of Mutational Signatures in Human Cancer. *Environ Mol Mutagen* 59, 25-25.
- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L., *et al.* (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415-421.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166-169.
- Andrysiak, Z., Galbraith, M.D., Guarnieri, A.L., Zaccara, S., Sullivan, K.D., Pandey, A., MacBeth, M., Inga, A., and Espinosa, J.M. (2017). Identification of a core TP53 transcriptional program with highly distributed tumor suppressive activity. *Genome Res* 27, 1645-1657.
- Arseneault, M., Monlong, J., Vasudev, N.S., Laskar, R.S., Safisamghabadi, M., Harnden, P., Egevad, L., Nourbehesht, N., Panichnantakul, P., Holcatova, I., *et al.* (2017). Loss of chromosome Y leads to down regulation of KDM5D and KDM6C epigenetic modifiers in clear cell renal cell carcinoma. *Sci Rep* 7, 44876.
- Baker, S.J., Preisinger, A.C., Jessup, J.M., Paraskeva, C., Markowitz, S., Willson, J.K., Hamilton, S., and Vogelstein, B. (1990). p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res* 50, 7717-7722.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* *315*, 1709-1712.

Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. *Cell* *141*, 1159-1170.

Beckerman, R., and Prives, C. (2010). Transcriptional regulation by p53. *Cold Spring Harb Perspect Biol* *2*, a000935.

Benchimol, S. (2001). p53-dependent pathways of apoptosis. *Cell Death Differ* *8*, 1049-1051.

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

Bodak, M., Yu, J., and Ciaudo, C. (2014). Regulation of LINE-1 in mammals. *Biomol Concepts* *5*, 409-428.

Boettcher, M., and McManus, M.T. (2015). Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Mol Cell* *58*, 575-585.

Bohn, J.A., Thummar, K., York, A., Raymond, A., Brown, W.C., Bieniasz, P.D., Hatzioannou, T., and Smith, J.L. (2017). APOBEC3H structure reveals an unusual mechanism of interaction with duplex RNA. *Nat Commun* *8*, 1021.

Bourara, K., Liegler, T.J., and Grant, R.M. (2007). Target cell APOBEC3C can induce limited G-to-A mutation in HIV-1. *PLoS Pathog* *3*, 1477-1485.

Brattain, M.G., Fine, W.D., Khaled, F.M., Thompson, J., and Brattain, D.E. (1981). Heterogeneity of malignant cells from a human colonic carcinoma. *Cancer Res* *41*, 1751-1756.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960-964.

Bunz, F., Dutriaux, A., Lengauer, C., Waldman, T., Zhou, S., Brown, J.P., Sedivy, J.M., Kinzler, K.W., and Vogelstein, B. (1998). Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science* 282, 1497-1501.

Burns, K.H. (2017). Transposable elements in cancer. *Nat Rev Cancer* 17, 415-424.

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

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

Carroll, D. (2011). Genome Engineering With Zinc-Finger Nucleases. *Genetics* 188, 773-782.

Cascalho, M. (2004). Advantages and disadvantages of cytidine deamination. *J Immunol* 172, 6513-6518.

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

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

Chen, X., Lv, Y., Sun, Y., Zhang, H., Xie, W., Zhong, L., Chen, Q., Li, M., Li, L., Feng, J., *et al.* (2019). PGC1beta Regulates Breast Tumor Growth and Metastasis by SREBP1-Mediated HKDC1 Expression. *Front Oncol* 9, 290.



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

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, NY)* 339, 819-823.

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

Cooke, S.L., Shlien, A., Marshall, J., Pipinikas, C.P., Martincorena, I., Tubio, J.M., Li, Y., Menzies, A., Mudie, L., Ramakrishna, M., *et al.* (2014). Processed pseudogenes acquired somatically during cancer development. *Nat Commun* 5, 3644.

Dang, Y., Siew, L.M., Wang, X., Han, Y., Lampen, R., and Zheng, Y.H. (2008). Human cytidine deaminase APOBEC3H restricts HIV-1 replication. *J Biol Chem* 283, 11606-11614.

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

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602-607.

Diehl, A.G., and Boyle, A.P. (2016). Deciphering ENCODE. *Trends Genet* 32, 238-249.

Duggal, N.K., Fu, W., Akey, J.M., and Emerman, M. (2013). Identification and antiviral activity of common polymorphisms in the APOBEC3 locus in human populations. *Virology* 443, 329-337.

Ebrahimi, D., Richards, C.M., Carpenter, M.A., Wang, J., Ikeda, T., Becker, J.T., Cheng, A.Z., McCann, J.L., Shaban, N.M., Salamango, D.J., *et al.* (2018). Genetic and mechanistic basis for

APOBEC3H alternative splicing, retrovirus restriction, and counteraction by HIV-1 protease. *Nat Commun* 9, 4137.

el-Deiry, W.S., Harper, J.W., O'Connor, P.M., Velculescu, V.E., Canman, C.E., Jackman, J., Pietenpol, J.A., Burrell, M., Hill, D.E., Wang, Y., *et al.* (1994). WAF1/CIP1 is induced in p53-mediated G1 arrest and apoptosis. *Cancer Res* 54, 1169-1174.

el-Deiry, W.S., Kern, S.E., Pietenpol, J.A., Kinzler, K.W., and Vogelstein, B. (1992). Definition of a consensus binding site for p53. *Nat Genet* 1, 45-49.

el-Deiry, W.S., Tokino, T., Velculescu, V.E., Levy, D.B., Parsons, R., Trent, J.M., Lin, D., Mercer, W.E., Kinzler, K.W., and Vogelstein, B. (1993). WAF1, a potential mediator of p53 tumor suppression. *Cell* 75, 817-825.

Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nature Genetics* 24, 363-367.

Eun, K., Park, M.G., Jeong, Y.W., Jeong, Y.I., Hyun, S.H., Hwang, W.S., Kim, S.H., and Kim, H. (2019). Establishment of TP53-knockout canine cells using optimized CRIPSR/Cas9 vector system for canine cancer research. *BMC Biotechnol* 19, 1.

Feng, Y., Goubran, M.H., Follack, T.B., and Chelico, L. (2017). Deamination-independent restriction of LINE-1 retrotransposition by APOBEC3H. *Sci Rep* 7, 10881.

Feng, Y., Wong, L., Morse, M., Rouzina, I., Williams, M.C., and Chelico, L. (2018). RNA-Mediated Dimerization of the Human Deoxycytidine Deaminase APOBEC3H Influences Enzyme Activity and Interaction with Nucleic Acids. *J Mol Biol* 430, 4891-4907.

Fischer, M. (2017). Census and evaluation of p53 target genes. *Oncogene* 36, 3943-3956.

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

Gallego Romero, I., Pai, A.A., Tung, J., and Gilad, Y. (2014). RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol* 12, 42.

Gardini, A. (2017). Global Run-On Sequencing (GRO-Seq). *Methods Mol Biol* 1468, 111-120.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67-71.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109, E2579-2586.

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., *et al.* (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52-56.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343-345.

Gilbert, S.F. (2013). *Developmental Biology*, Vol 10th ed (Sunderland, MA USA: Sinauer Associates, Inc. Publishers.).

Gohler, T., Jager, S., Warnecke, G., Yasuda, H., Kim, E., and Deppert, W. (2005). Mutant p53 proteins bind DNA in a DNA structure-selective mode. *Nucleic Acids Res* 33, 1087-1100.

Goila-Gaur, R., and Strebel, K. (2008). HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology* 5, 51.

Hagan, C.R., and Rudin, C.M. (2002). Mobile genetic element activation and genotoxic cancer therapy: potential clinical implications. *Am J Pharmacogenomics* 2, 25-35.

Harari, A., Ooms, M., Mulder, L.C., and Simon, V. (2009). Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H. *J Virol* 83, 295-303.

Hardy, S., Legagneux, V., Audic, Y., and Paillard, L. (2010). Reverse genetics in eukaryotes. *Biol Cell* 102, 561-580.

Harper, J.W., Adami, G.R., Wei, N., Keyomarsi, K., and Elledge, S.J. (1993). The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* 75, 805-816.

Harris, C.R., Dewan, A., Zupnick, A., Normart, R., Gabriel, A., Prives, C., Levine, A.J., and Hoh, J. (2009). p53 responsive elements in human retrotransposons. *Oncogene* 28, 3857-3865.

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

Harrison, P.M., Zheng, D., Zhang, Z., Carriero, N., and Gerstein, M. (2005). Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res* 33, 2374-2383.

Harrod, A., Fulton, J., Nguyen, V.T.M., Periyasamy, M., Ramos-Garcia, L., Lai, C.F., Metodieva, G., de Giorgio, A., Williams, R.L., Santos, D.B., *et al.* (2017). Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer. *Oncogene* 36, 2286-2296.

Hata, K., and Sakaki, Y. (1997). Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* 189, 227-234.

Helena Mangs, A., and Morris, B.J. (2007). The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Curr Genomics* 8, 129-136.

Hemann, M.T., and Lowe, S.W. (2006). The p53-Bcl-2 connection. *Cell Death Differ* 13, 1256-1259.

HISAT2 (2017). graph-based alignment of next generation sequencing reads to a population of genomes.

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

Honda, T., and Rahman, M.A. (2019). Profiling of LINE-1-Related Genes in Hepatocellular Carcinoma. *Int J Mol Sci* 20.

Honma, M., Hayashi, M., Ohno, T., Mizusawa, H., Saijo, K., and Sofuni, T. (1996). Heterogeneity of the Y chromosome following long-term culture of the human lung cancer cell line A549. *In Vitro Cell Dev Biol Anim* 32, 262-264.

Hultquist, J.F., Lengyel, J.A., Refsland, E.W., LaRue, R.S., Lackey, L., Brown, W.L., and Harris, R.S. (2011). Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *J Virol* 85, 11220-11234.

Imperial College London (2019). Imperial BRC Genomics Facility.

Iseda, S., Takahashi, N., Poplimont, H., Nomura, T., Seki, S., Nakane, T., Nakamura, M., Shi, S., Ishii, H., Furukawa, S., *et al.* (2016). Biphasic CD8<sup>+</sup> T-Cell Defense in Simian Immunodeficiency Virus Control by Acute-Phase Passive Neutralizing Antibody Immunization. *J Virol* 90, 6276-6290.

Ito, F., Yang, H., Xiao, X., Li, S.X., Wolfe, A., Zirkle, B., Arutiunian, V., and Chen, X.S. (2018). Understanding the Structure, Multimerization, Subcellular Localization and mC Selectivity of a Genomic Mutator and Anti-HIV Factor APOBEC3H. *Sci Rep* 8, 3763.

Jarvis, M.C., Ebrahimi, D., Temiz, N.A., and Harris, R.S. (2018). Mutation Signatures Including APOBEC in Cancer Cell Lines. *JNCI Cancer Spectr* 2.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821.

Joung, J.K., and Sander, J.D. (2013). INNOVATION TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Bio* 14, 49-55.

Kamens, J. (2015). The Addgene repository: an international nonprofit plasmid and data resource. *Nucleic Acids Res* 43, D1152-1157.

Kazazian, H.H., Jr. (2000). Genetics. L1 retrotransposons shape the mammalian genome. *Science* 289, 1152-1153.

Kerachian, M.A., and Kerachian, M. (2019). Long interspersed nucleotide element-1 (LINE-1) methylation in colorectal cancer. *Clin Chim Acta* 488, 209-214.

Kersey, P.J., Allen, J.E., Armean, I., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C., *et al.* (2016). Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res* 44, D574-580.

Kidd, J.M., Newman, T.L., Tuzun, E., Kaul, R., and Eichler, E.E. (2007). Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet* 3, e63.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12, 357-360.

Kim, D.H., and Rossi, J.J. (2007). Strategies for silencing human disease using RNA interference. *Nature Reviews Genetics* 8, 173-184.

Kim, M.P., and Lozano, G. (2018). Mutant p53 partners in crime. *Cell Death Differ* 25, 161-168.

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

Kitayner, M., Rozenberg, H., Kessler, N., Rabinovich, D., Shaulov, L., Haran, T.E., and Shakked, Z. (2006). Structural basis of DNA recognition by p53 tetramers. *Mol Cell* 22, 741-753.

Klein, S.J., and O'Neill, R.J. (2018). Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res* 26, 5-23.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., *et al.* (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583-588.

Konermann, S., Lotfy, P., Brideau, N.J., Oki, J., Shokhirev, M.N., and Hsu, P.D. (2018). Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* 173, 665-676 e614.

Krisko, J.F., Begum, N., Baker, C.E., Foster, J.L., and Garcia, J.V. (2016). APOBEC3G and APOBEC3F Act in Concert To Extinguish HIV-1 Replication. *J Virol* 90, 4681-4695.

Kubbutat, M.H., Jones, S.N., and Vousden, K.H. (1997). Regulation of p53 stability by Mdm2. *Nature* 387, 299-303.

Kuong, K.J., and Loeb, L.A. (2013). APOBEC3B mutagenesis in cancer. *Nat Genet* 45, 964-965.

Lamarre, S., Frasse, P., Zouine, M., Labourdette, D., Sainderichin, E., Hu, G., Le Berre-Anton, V., Bouzayen, M., and Maza, E. (2018). Optimization of an RNA-Seq Differential Gene Expression Analysis Depending on Biological Replicate Number and Library Size. *Front Plant Sci* 9, 108.

Laptenko, O., Shiff, I., Freed-Pastor, W., Zupnick, A., Mattia, M., Freulich, E., Shamir, I., Kadouri, N., Kahan, T., Manfredi, J., *et al.* (2015). The p53 C terminus controls site-specific DNA binding and promotes structural changes within the central DNA binding domain. *Mol Cell* 57, 1034-1046.

LaRue, R.S., Jonsson, S.R., Silverstein, K.A., Lajoie, M., Bertrand, D., El-Mabrouk, N., Hotzel, I., Andresdottir, V., Smith, T.P., and 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 Mol Biol* 9, 104.

Larue, R.S., Lengyel, J., Jonsson, S.R., Andresdottir, V., and Harris, R.S. (2010). Lentiviral Vif degrades the APOBEC3Z3/APOBEC3H protein of its mammalian host and is capable of cross-species activity. *J Virol* *84*, 8193-8201.

Lavin, M.F., and Gueven, N. (2006). The complexity of p53 stabilization and activation. *Cell Death Differ* *13*, 941-950.

Li, M.M., and Emerman, M. (2011). Polymorphism in human APOBEC3H affects a phenotype dominant for subcellular localization and antiviral activity. *J Virol* *85*, 8197-8207.

Liu, G., and Chen, X. (2006). Regulation of the p53 transcriptional activity. *J Cell Biochem* *97*, 448-458.

Liu, Y., and Bodmer, W.F. (2006). Analysis of P53 mutations and their expression in 56 colorectal cancer cell lines. *P Natl Acad Sci USA* *103*, 976-981.

Liu, Y., Zhou, J., and White, K.P. (2014). RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* *30*, 301-304.

Loayza-Puch, F., Drost, J., Rooijers, K., Lopes, R., Elkon, R., and Agami, R. (2013). p53 induces transcriptional and translational programs to suppress cell proliferation and growth. *Genome Biol* *14*, R32.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* *15*, 550.

Ma, S., and Dai, Y. (2011). Principal component analysis based methods in bioinformatics studies. *Brief Bioinform* *12*, 714-722.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* *1*, 7.



Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823-826.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 424, 99-103.

Mao, Z., Jiang, Y., Liu, X., Seluanov, A., and Gorbunova, V. (2009). DNA repair by homologous recombination, but not by nonhomologous end joining, is elevated in breast cancer cells. *Neoplasia* 11, 683-691.

Mariani, R., Chen, D., Schrofelbauer, B., Navarro, F., Konig, R., Bollman, B., Munk, C., Nymark-McMahon, H., and Landau, N.R. (2003). Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* 114, 21-31.

Marin, M., Rose, K.M., Kozak, S.L., and Kabat, D. (2003). HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat Med* 9, 1398-1403.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843-1845.

Marraffini, L.A., and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* 11, 181-190.

Mashima, T., Oh-hara, T., Sato, S., Mochizuki, M., Sugimoto, Y., Yamazaki, K., Hamada, J., Tada, M., Moriuchi, T., Ishikawa, Y., *et al.* (2005). p53-defective tumors with a functional apoptosome-mediated pathway: a new therapeutic target. *J Natl Cancer Inst* 97, 765-777.

Mathur, R., Rotroff, D., Ma, J., Shojaie, A., and Motsinger-Reif, A. (2018). Gene set analysis methods: a systematic comparison. *BioData Min* 11, 8.

Matlashewski, G., Lamb, P., Pim, D., Peacock, J., Crawford, L., and Benchimol, S. (1984). Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene. *EMBO J* 3, 3257-3262.

Matsuoka, T., Nagae, T., Ode, H., Awazu, H., Kurosawa, T., Hamano, A., Matsuoka, K., Hachiya, A., Imahashi, M., Yokomaku, Y., *et al.* (2018). Structural basis of chimpanzee APOBEC3H dimerization stabilized by double-stranded RNA. *Nucleic Acids Res.*

May, P., and May, E. (1999). Twenty years of p53 research: structural and functional aspects of the p53 protein. *Oncogene 18*, 7621-7636.

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

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

Miyashita, T., Kitada, S., Krajewski, S., Horne, W.A., Delia, D., and Reed, J.C. (1995). Overexpression of the Bcl-2 protein increases the half-life of p21<sup>Bax</sup>. *J Biol Chem 270*, 26049-26052.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol 60*, 174-182.

Mukherjee, S., Mukhopadhyay, A., Banerjee, D., Chandak, G.R., and Ray, K. (2004). Molecular pathology of haemophilia B: identification of five novel mutations including a LINE 1 insertion in Indian patients. *Haemophilia 10*, 259-263.

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

Nakato, R., and Shirahige, K. (2017). Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Brief Bioinform 18*, 279-290.

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

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., *et al.* (2019). Author Correction: Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 566, E1.

Nik-Zainal, S., and Morganella, S. (2017). Mutational Signatures in Breast Cancer: The Problem at the DNA Level. *Clin Cancer Res* 23, 2617-2629.

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

OhAinle, M., Kerns, J.A., Li, M.M., Malik, H.S., and Emerman, M. (2008). Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. *Cell Host Microbe* 4, 249-259.

OhAinle, M., Kerns, J.A., Malik, H.S., and Emerman, M. (2006). Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. *J Virol* 80, 3853-3862.

Olivier, M., Hollstein, M., and Hainaut, P. (2010). TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol* 2, a001008.

Oughtred, R., Stark, C., Breitkreutz, B.J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., *et al.* (2019). The BioGRID interaction database: 2019 update. *Nucleic Acids Res* 47, D529-D541.

Patsali, P., Kleanthous, M., and Lederer, C.W. (2019). Disruptive Technology: CRISPR/Cas-Based Tools and Approaches. *Mol Diagn Ther* 23, 187-200.

Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X.J., Harte, R., Balasubramanian, S., Tanzer, A., Diekhans, M., *et al.* (2012). The GENCODE pseudogene resource. *Genome Biol* *13*, R51.

Periyasamy, M., Patel, H., Lai, C.F., Nguyen, V.T.M., Nevedomskaya, E., Harrod, A., Russell, R., Remenyi, J., Ochocka, A.M., Thomas, R.S., *et al.* (2015). APOBEC3B-Mediated Cytidine Deamination Is Required for Estrogen Receptor Action in Breast Cancer. *Cell Rep* *13*, 108-121.

Periyasamy, M., Singh, A.K., Gemma, C., Kranjec, C., Farzan, R., Leach, D.A., Navaratnam, N., Palinkas, H.L., Vertessy, B.G., Fenton, T.R., *et al.* (2017). p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells. *Nucleic Acids Res* *45*, 11056-11069.

Petljak, M., Alexandrov, L.B., Brammell, J.S., Price, S., Wedge, D.C., Grossmann, S., Dawson, K.J., Ju, Y.S., Iorio, F., Tubio, J.M.C., *et al.* (2019). Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell* *176*, 1282-1294 e1220.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* *151*, 653-663.

Qin, L., Yong, L., and Wei, D. (2018). A Novel Target of p53, TCF21, Can Respond to Hypoxia by MAPK Pathway Inactivation in Uterine Corpus Endometrial Carcinoma. *DNA and Cell Biology* *37*, 473-480.

Qiu, P., Shandilya, H., D'Alessio, J.M., O'Connor, K., Durocher, J., and Gerard, G.F. (2004). Mutation detection using Surveyor (TM) nuclease. *Biotechniques* *36*, 702-+.

Rangwala, S.H., Zhang, L., and Kazazian, H.H., Jr. (2009). Many LINE1 elements contribute to the transcriptome of human somatic cells. *Genome Biol* *10*, R100.

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

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

Refsland, E.W., Hultquist, J.F., Luengas, E.M., Ikeda, T., Shaban, N.M., Law, E.K., Brown, W.L., Reilly, C., Emerman, M., and Harris, R.S. (2014). Natural polymorphisms in human APOBEC3H and HIV-1 Vif combine in primary T lymphocytes to affect viral G-to-A mutation levels and infectivity. *PLoS Genet* 10, e1004761.

Repke, H., Gabuzda, D., Palu, G., Emmrich, F., and Sodroski, J. (1992). Effects of CD4 synthetic peptides on HIV type I envelope glycoprotein function. *J Immunol* 149, 1809-1816.

Rivlin, N., Brosh, R., Oren, M., and Rotter, V. (2011). Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. *Genes Cancer* 2, 466-474.

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

Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G., *et al.* (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* 45, 970-976.

Rodic, N., and Burns, K.H. (2013). Long interspersed element-1 (LINE-1): passenger or driver in human neoplasms? *PLoS Genet* 9, e1003402.

Salter, J.D., Bennett, R.P., and Smith, H.C. (2016). The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem Sci* 41, 578-594.

Sanchez-Rivera, F.J., and Jacks, T. (2015). Applications of the CRISPR-Cas9 system in cancer biology. *Nat Rev Cancer* 15, 387-395.

Scala, F., Brighenti, E., Govoni, M., Imbrogno, E., Fornari, F., Trere, D., Montanaro, L., and Derenzini, M. (2016). Direct relationship between the level of p53 stabilization induced by rRNA synthesis-inhibiting drugs and the cell ribosome biogenesis rate. *Oncogene* *35*, 977-989.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* *7*, 3.

Sellin Jeffries, M.K., Kiss, A.J., Smith, A.W., and Oris, J.T. (2014). A comparison of commercially-available automated and manual extraction kits for the isolation of total RNA from small tissue samples. *BMC Biotechnol* *14*, 94.

Sentmanat, M.F., Peters, S.T., Florian, C.P., Connelly, J.P., and Pruett-Miller, S.M. (2018). A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci Rep* *8*, 888.

Septyarskiy, V.B., Soldatov, R.A., Popadin, K.Y., Antonarakis, S.E., Bazykin, G.A., and Nikolaev, S.I. (2016). APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. *Genome Res* *26*, 174-182.

Shaban, N.M., Shi, K., Lauer, K.V., Carpenter, M.A., Richards, C.M., Salamango, D., Wang, J., Lopresti, M.W., Banerjee, S., Levin-Klein, R., *et al.* (2018). The Antiviral and Cancer Genomic DNA Deaminase APOBEC3H Is Regulated by an RNA-Mediated Dimerization Mechanism. *Mol Cell* *69*, 75-86 e79.

Shadeo, A., and Lam, W.L. (2006). Comprehensive copy number profiles of breast cancer cell model genomes. *Breast Cancer Res* *8*, R9.

Shangary, S., and Wang, S. (2009). Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. *Annu Rev Pharmacol Toxicol* *49*, 223-241.

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

Shi, K., Carpenter, M.A., Banerjee, S., Shaban, N.M., Kurahashi, K., Salamango, D.J., McCann, J.L., Starrett, G.J., Duffy, J.V., Demir, O., *et al.* (2017). Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat Struct Mol Biol* 24, 131-139.

Shin, Y.J., Kim, Y., Wen, X., Cho, N.Y., Lee, S., Kim, W.H., and Kang, G.H. (2019). Prognostic implications and interaction of L1 methylation and p53 expression statuses in advanced gastric cancer. *Clin Epigenetics* 11, 77.

Smith, L., Farzan, R., Ali, S., Buluwela, L., A, T.S., and Meek, D.W. (2018). Author Correction: The responses of cancer cells to PLK1 inhibitors reveal a novel protective role for p53 in maintaining centrosome separation. *Sci Rep* 8, 5237.

Song, K.H., Park, M.S., Nandu, T.S., Gadad, S., Kim, S.C., and Kim, M.Y. (2016). GALNT14 promotes lung-specific breast cancer metastasis by modulating self-renewal and interaction with the lung microenvironment. *Nat Commun* 7, 13796.

Soule, H.D., Vazquez, J., Long, A., Albert, S., and Brennan, M. (1973). A human cell line from a pleural effusion derived from a breast carcinoma. *J Natl Cancer Inst* 51, 1409-1416.

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

Steinhoff, C., and Schulz, W.A. (2003). Transcriptional regulation of the human LINE-1 retrotransposon L1.2B. *Mol Genet Genomics* 270, 394-402.

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

Stopak, K., de Noronha, C., Yonemoto, W., and Greene, W.C. (2003). HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol Cell* 12, 591-601.

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

Stratton, M.R. (2013). Journeys into the genome of cancer cells. *EMBO Mol Med* 5, 169-172.

Sullivan, K.D., Galbraith, M.D., Andrysiak, Z., and Espinosa, J.M. (2018). Mechanisms of transcriptional regulation by p53. *Cell Death Differ* 25, 133-143.

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

Takaoka, A., Hayakawa, S., Yanai, H., Stoiber, D., Negishi, H., Kikuchi, H., Sasaki, S., Imai, K., Shibue, T., Honda, K., *et al.* (2003). Integration of interferon-alpha/beta signalling to p53 responses in tumour suppression and antiviral defence. *Nature* 424, 516-523.

Tebaldi, T., Zaccara, S., Alessandrini, F., Bisio, A., Ciribilli, Y., and Inga, A. (2015). Whole-genome cartography of p53 response elements ranked on transactivation potential. *BMC Genomics* 16, 464.

Torano, E.G., Petrus, S., Fernandez, A.F., and Fraga, M.F. (2012). Global DNA hypomethylation in cancer: review of validated methods and clinical significance. *Clin Chem Lab Med* 50, 1733-1742.

Torrents, D., Suyama, M., Zdobnov, E., and Bork, P. (2003). A genome-wide survey of human pseudogenes. *Genome Res* 13, 2559-2567.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105-1111.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res* 40, e115.



Vichai, V., and Kirtikara, K. (2006). Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nat Protoc* 1, 1112-1116.

Viloria, K., Munasinghe, A., Asher, S., Bogeyere, R., Jones, L., and Hill, N.J. (2016). A holistic approach to dissecting SPARC family protein complexity reveals FSTL-1 as an inhibitor of pancreatic cancer cell growth. *Sci Rep* 6, 37839.

Virgen, C.A., and Hatzioannou, T. (2007). Antiretroviral activity and Vif sensitivity of rhesus macaque APOBEC3 proteins. *J Virol* 81, 13932-13937.

Vousden, K.H. (2009). Functions of p53 in metabolism and invasion. *Biochem Soc Trans* 37, 511-517.

Walton, J., Blagih, J., Ennis, D., Leung, E., Dowson, S., Farquharson, M., Tookman, L.A., Orange, C., Athineos, D., Mason, S., *et al.* (2016). CRISPR/Cas9-Mediated Trp53 and Brca2 Knockout to Generate Improved Murine Models of Ovarian High-Grade Serous Carcinoma. *Cancer Res* 76, 6118-6129.

Wang, X., Abudu, A., Son, S., Dang, Y., Venta, P.J., and Zheng, Y.H. (2011). Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity. *J Virol* 85, 3142-3152.

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

Wijesinghe, P., and Bhagwat, A.S. (2012). Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. *Nucleic Acids Res* 40, 9206-9217.

Wylie, A., Jones, A.E., D'Brot, A., Lu, W.J., Kurtz, P., Moran, J.V., Rakheja, D., Chen, K.S., Hammer, R.E., Comerford, S.A., *et al.* (2016). p53 genes function to restrain mobile elements. *Genes Dev* 30, 64-77.

Xiao, X., Li, S.X., Yang, H., and Chen, X.S. (2016). Crystal structures of APOBEC3G N-domain alone and its complex with DNA. *Nat Commun* 7, 12193.

Xiao-Jie, L., Hui-Ying, X., Qi, X., Jiang, X., and Shi-Jie, M. (2016). LINE-1 in cancer: multifaceted functions and potential clinical implications. *Genet Med* 18, 431-439.

Xu, J., Peng, X., Chen, Y., Zhang, Y., Ma, Q., Liang, L., Carter, A.C., Lu, X., and Wu, C.I. (2017). Free-living human cells reconfigure their chromosomes in the evolution back to unicellularity. *Elife* 6.

Yang, L., Mali, P., Kim-Kiselak, C., and Church, G. (2014). CRISPR-Cas-mediated targeted genome editing in human cells. *Methods Mol Biol* 1114, 245-267.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13, 335-340.

Zemojtel, T., and Vingron, M. (2012). P53 binding sites in transposons. *Front Genet* 3, 40.

Zhang, Z., Gu, Q., de Manuel Montero, M., Bravo, I.G., Marques-Bonet, T., Haussinger, D., and Munk, C. (2017). Stably expressed APOBEC3H forms a barrier for cross-species transmission of simian immunodeficiency virus of chimpanzee to humans. *PLoS Pathog* 13, e1006746.

Zhao, W., Peng, Y., Mills, G., and Peng, G. (2018). Abstract PD8-11: APOBEC3 contributes to mutational load in breast cancer. *Cancer Research* 78, PD8-11-PD18-11.

Zhen, A., Du, J., Zhou, X., Xiong, Y., and Yu, X.F. (2012). Reduced APOBEC3H variant antiviral activities are associated with altered RNA binding activities. *PLoS One* 7, e38771.

Zheng, D., and Gerstein, M.B. (2006). A computational approach for identifying pseudogenes in the ENCODE regions. *Genome Biol* 7 *Suppl 1*, S13 11-10.

Chen, G., Cairelli, M. J., Kilicoglu, H., Shin, D., & Rindfleisch, T. C. (2014). Augmenting Microarray Data with Literature-Based Knowledge to Enhance Gene Regulatory Network Inference. *PLoS Computational Biology*, 10(6). doi: 10.1371/journal.pcbi.1003666

# Supplementary 1. p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells

11056–11069 *Nucleic Acids Research*, 2017, Vol. 45, No. 19  
doi: 10.1093/nar/gkx721

Published online 16 August 2017

## p53 controls expression of the DNA deaminase APOBEC3B to limit its potential mutagenic activity in cancer cells

Manikandan Periyasamy<sup>1</sup>, Anup K. Singh<sup>1</sup>, Carolina Gemma<sup>1</sup>, Christian Kranjec<sup>2</sup>, Raed Farzan<sup>1</sup>, Damien A. Leach<sup>1</sup>, Naveenan Navaratnam<sup>3</sup>, Hajnalka L. Pálincás<sup>4,5</sup>, Beata G. Vértessy<sup>4,5</sup>, Tim R. Fenton<sup>6</sup>, John Doorbar<sup>2</sup>, Frances Fuller-Pace<sup>7</sup>, David W. Meek<sup>7</sup>, R. Charles Coombes<sup>1</sup>, Laki Buluwela<sup>1</sup> and Simak Ali<sup>1,\*</sup>

<sup>1</sup>Department of Surgery & Cancer, Imperial College London, Hammersmith Hospital Campus, London W12 0NN, UK, <sup>2</sup>Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK, <sup>3</sup>MRC London Institute of Medical Sciences, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London W12 0NN, UK, <sup>4</sup>Department of Applied Biotechnology and Food Science, Budapest University of Technology and Economics, Budapest 1111, Hungary, <sup>5</sup>Laboratory of Genome Metabolism and Repair, Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest 1117, Hungary, <sup>6</sup>School of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK and <sup>7</sup>Division of Cancer Research, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK

Received May 17, 2017; Revised July 30, 2017; Editorial Decision August 07, 2017; Accepted August 08, 2017

### ABSTRACT

Cancer genome sequencing has implicated the cytosine deaminase activity of apolipoprotein B mRNA editing enzyme catalytic polypeptide-like (APOBEC) genes as an important source of mutations in diverse cancers, with APOBEC3B (A3B) expression especially correlated with such cancer mutations. To better understand the processes directing A3B overexpression in cancer, and possible therapeutic avenues for targeting A3B, we have investigated the regulation of A3B gene expression. Here, we show that A3B expression is inversely related to p53 status in different cancer types and demonstrate that this is due to a direct and pivotal role for p53 in repressing A3B expression. This occurs through the induction of p21 (CDKN1A) and the recruitment of the repressive DREAM complex to the A3B gene promoter, such that loss of p53 through mutation, or human papilloma virus-mediated inhibition, prevents recruitment of the complex, thereby causing elevated A3B expression and cytosine deaminase activity in cancer cells. As p53 is frequently mutated in cancer, our findings provide a mechanism by which p53 loss can promote cancer mutagenesis.

### INTRODUCTION

The APOBEC3 family of cytosine deaminases are mediators of intrinsic immunity to retroviruses and endogenous retrotransposons, which act by causing cytosine-to-uracil (C-to-U) deamination in single-stranded DNA that is generated during reverse transcription (1,2), to promote deleterious mutations. The seven members of the APOBEC3 gene family (A3A/B/C/D/F/G/H), are related to the APOBEC1, APOBEC2, APOBEC4 and activation-induced deaminase (AID). APOBEC1 functions in RNA editing and DNA editing by AID is required for class-switch recombination and somatic hypermutation of immunoglobulin genes in B cells to augment antibody diversity (3). Additionally, apolipoprotein B mRNA editing enzyme catalytic polypeptide-like (APOBEC) genes, in particular AID, have been implicated in the epigenetic regulation of gene expression by directing the deamination of 5-hydroxymethylcytosine generated by TET enzyme conversion of 5-methylcytosine (for reviews see refs. (4–6)). Here, deamination by AID facilitates base excision repair, resulting in cytosine demethylation. Moreover, we recently reported a DNA methylation-independent role for A3B-mediated cytidine deamination and repair as a mechanism for chromatin remodelling that facilitates estrogen receptor (ER) target gene expression in breast cancer cells (7).

The mutational capacity of APOBECs has led to the proposal that inappropriate and/or upregulated expression of these genes could promote mutations in genomic DNA, a

\*To whom correspondence should be addressed. Tel: +44 20 7594 2811; Email: [simak.ali@imperial.ac.uk](mailto:simak.ali@imperial.ac.uk)

© The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

possibility bolstered by the demonstration that AID can cause chromosomal mutations and rearrangements (8–10) and AID, as well as APOBEC1 promote tumorigenesis in transgenic mouse models (11–13). Ectopic expression studies in yeast and mammalian cells have shown that APOBEC3 enzymes can also promote mutations in genomic DNA (14–17). Importantly, sequencing of cancer genomes reveals that a large proportion of somatic mutations in diverse cancer types, including breast, ovarian, cervical, bladder, head and neck and lung cancer, are attributable to APOBEC activity (17–23). A3B is the only one of the 11 APOBEC genes that is consistently expressed at high levels in these cancer types and A3B expression correlates with the number of C-to-T and overall mutational load in cancer genomes (17,21,24). Genome sequencing of yeast cells expressing A3B identify kataeic mutational patterns similar to those that are observed in breast cancer genomes (25) and the incidence of C-to-T mutations in breast cancer cells is reduced by A3B knockdown (17). Together, these studies provide a compelling case for A3B as a driver of the mutational landscape and tumour evolution in many common cancers.

Gene expression analysis shows that A3B levels are low in normal tissues, but are elevated in many cancer types (7,17). Understanding the mechanisms that regulate A3B expression will provide important insights into the processes driving acquisition of cancer mutations and tumour evolution. Originally cloned on the basis of its induction by phorbol ester treatment of normal keratinocytes, A3B expression is stimulated by NF- $\kappa$ B activation by protein kinase C (26,27). Interestingly, mutational signatures associated with A3B activity are especially strong in cervical and head/neck cancers, in which human papillomaviruses (HPV) are important causative agents (20,21,28). Recently, the E6 and E7 viral oncogenes in high-risk HPVs were shown to promote A3B expression (29–32), highlighting a potential explanation for the A3B-associated mutator phenotype in HPV-positive cervical and head/neck cancer. As inhibition of p53 tumour suppressor activity/levels is a key function of E6 (33), we reasoned that p53 might be a direct regulator of A3B expression. Here, we show that p53 represses A3B expression and cytosine deaminase activity in cancer cells, through a p21-dependent mechanism and that loss of p53 activity through its mutation or HPV-16 E6/E7-mediated downregulation, causes A3B upregulation. Further, by assessing cellular cytosine deaminase activity and abasic site generation in genomic DNA, we show that loss of p53 activity through mutation or HPV-directed downregulation can promote increased mutagenic capacity of normal and cancer cells.

## MATERIALS AND METHODS

### Cell lines

Cell lines were obtained from ATCC (LGC Standards, UK) and cultured in Dulbecco's modified Eagle's medium (DMEM) containing 10% foetal calf serum (FCS). HCT116 p53<sup>-/-</sup> and HCT116 p21<sup>-/-</sup> cells were kindly provided by Dr B. Vogelstein (34,35). NIKS cell lines have been described previously (36), and were maintained at sub-confluence on  $\gamma$ -irradiated J2–3T3 feeder cells in com-

plete F medium, as described (37). Nutlin-3 (Bio-Techne Ltd, UK) was added to a final concentration of 10  $\mu$ M, unless otherwise stated. An equal volume of dimethylsulphoxide (DMSO) was added to the vehicle controls. HPV16 E6 and p53-interaction defective E6 mutant, Addgene ID #44152 and 44153, respectively, were a gift from P. Howley (38). The vector, MSCV-N- GFP (Addgene ID: 37855) was a gift from K. Munger (39). Retroviral constructs pLXSN, pLXSN HPV16 E6, pLXSN HPV16 E7 and pLXSN HPV16 E6/E7 were kindly provided by Dr Denise Galloway (40). For the generation of HPV-16 E6-SAT mutant, wild-type (WT) E6 was mutated using Pfu-Ultra DNA polymerase (Stratagene, London, UK) and the following primer pair: E6SAT\_forward

5'-GCAATGTTTCAGGACCCACAGGAGCGC  
CACAAAGTTACCACAGTTATGCACAGAGCTGC  
-3';

E6SAT\_reverse 5'-GCAGCTCTGTGCATAACTGTG  
GTAACCTTGTGGCGCTCTCTGTGGGTCTGA  
AACATTGC-3'.

### Breast cancer samples

The patients presented with primary, operable breast cancer to the Dundee Cancer Centre between 1997 and 2012 and provided written, informed consent for research use of their tissues. Use of the clinical material and data were approved by the Tayside Tissue Bank under delegated authority from the Tayside Local Research Ethics Committee. Total RNA was extracted from tumour samples using the Qiagen RNA extraction kit and cDNA was prepared using the High Capacity cDNA Reverse Transcription Kit and oligo-dT primers (Applied Biosystems), according to manufacturer's instructions. Polymerase chain reaction (PCR) with p53-specific primers (5'-TTCCACGACGGTGACAC GCT, 5'-CTTCTGACGCACACCTATTG) was used to amplify full length p53 cDNA, followed by Sanger sequencing to identify p53 mutations. Thirty of the 32 mutant samples encoded TP53 mutations/deletions described in the IARC TP53 database (41). Of the remaining samples, one case had a 3-nt deletion causing loss of Ala159 in exon 5, the second featured a single C insertion that would cause a frameshift after Pro309 in exon 9.

### Immunoblotting

Whole cell lysates were prepared in RIPA buffer supplemented with protease and phosphatase inhibitor cocktails (Roche, UK), and immunoblotted as described (7). The A3B antibody has been described previously (7). Antibodies for p21 (sc-397), p53 (sc-126), survivin (BIRC5; sc17779), cyclin B1 (CCNB1; sc-752), E2F4 (sc-866), LIN9 (sc-398234) and HPV16 E7 (sc-6981) were purchased from Santa Cruz Biotechnology Inc. (Germany).  $\beta$ -actin (ab6276) and MDM2 (ab16895) antibodies were obtained from Abcam plc, UK. The Lin54 (A303–799A-M) antibody was from Bethyl Laboratories Inc. (USA). HPV16 E6 antibody was from Arbor Vita corporation (CA, USA).

**Real-time RT-PCR (RT-qPCR) and ChIP-qPCR assay**

Total RNA was extracted from cells in culture, as described (7). Taqman Gene Expression Assays (Applied Biosystems, UK) were used for RT-qPCR on an ABI 7900HT machine and are detailed in Supplementary Data. Chromatin preparation and immunoprecipitation (ChIP) was performed exactly as described previously (7). Primer sequences for qPCR and ChIP antibodies are listed in Supplementary Data.

**RNA interference**

The reverse transfection method using Lipofectamine RNAiMAX (Invitrogen, ThermoFisher Scientific, UK) was used with double-stranded RNA oligonucleotides (siRNA), as described previously (7). Nutlin was added after 48 h and RNA and protein lysates were prepared a further 24 or 48 h later, as appropriate. ON-TARGETplus human siRNA for E2F4 (J-003471-12), Lin9 (L-018918-01), Lin54 (L-019325-01), p53 (L-003329-00) and p21 (J-003471-12) were purchased from Dharmacon, ThermoFisher Scientific, UK.

**CRISPR-Cas9 for disruption of the TP53 gene**

CRISPR-mediated deletion of sequences in the third coding exon of the p53 (TP53) gene was carried out using an Amara Type II nucleofector (Lonza, Cologne, Germany) and the MCF7 transfection protocol recommended by the manufacturer. This involved co-transfection of U6 promoter based expression plasmids for CRISPR138077 (5'-A CACCGGCGGCCCTGCACC-3') and CRISPR138076 (5'-GGGCAGCTACGGTTTCCGTC-3'), described previously (42) and a hCas9 expression plasmid (a gift from George Church; Addgene plasmid #41815). Following nucleofection, cells were allowed to grow in DMEM supplemented with 10% FCS, and established colonies were screened by PCR for the region encompassing the two CRISPR target sites in TP53 coding exon 3, using primers with the sequences 5'-GATGAAGCTCCCAGA ATGCC-3' and 5'-CACTGACAGGAAGCAAAGG-3', where the PCR product for WT and exon 3-deleted TP53 gene is 311 and 226 bp, respectively.

**Cytidine deaminase assay**

Cells were lysed in 25 mM HEPES (pH 7.4), 10% glycerol, 150 mM NaCl, 0.5% Triton X-100, 1 mM ethylenediaminetetraacetic acid, 1 mM MgCl<sub>2</sub> and 1 mM ZnCl<sub>2</sub>, supplemented with protease inhibitors. Lysates were incubated at 37°C for 15 min following addition of 2 µg RNase A (Qiagen). About 1 pmol ssDNA substrate 5' DY782-ATTATTATTATTATTATTATTTC ATTTATTTATTTATTTA-3' (Eurofins, UK) and 0.75U uracil-DNA glycosylase (NEB) were added to 10 µg protein lysate at 37°C for 1 h. A total of 10 µl 1N NaOH was added and samples incubated for 15 min at 37°C. Finally, 10 µl 1N HCl was added to neutralize the reaction and samples were separated by electrophoresis through 15% urea-polyacrylamide gel electrophoresis gels in Tris-borate-EDTA (1x) at 150V for 2–3 h.

**Apurinic/aprymidinic (AP) site assay**

Genomic DNA was prepared and apurinic/aprymidinic (AP) site determination was performed using the Oxiselect DNA damage ELISA kit (AP sites) (STA-324), according to manufacturer's protocol (Cell Biolabs Inc. San Diego, CA, USA), using the aldehyde reactive probe (ARP) DNA standards to quantify the number of genomic AP sites.

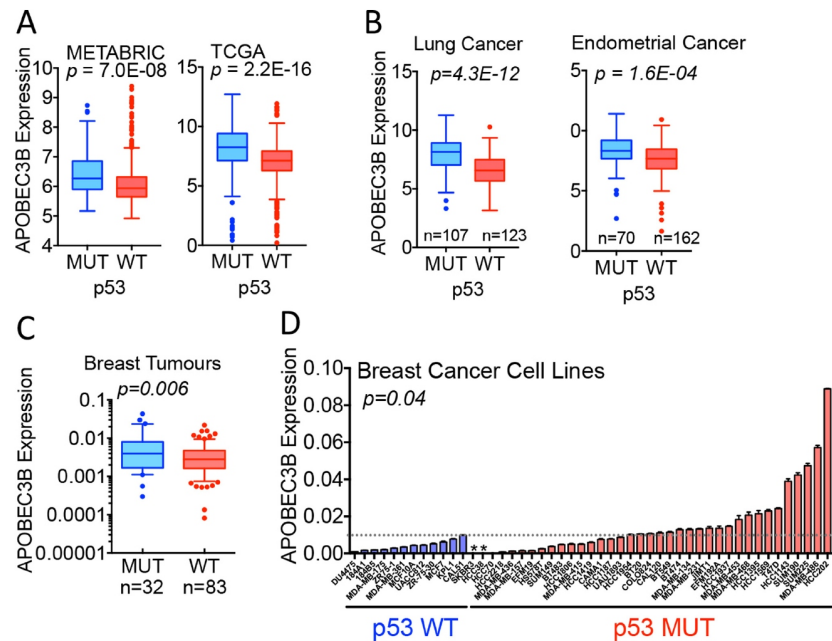
**Analysis of METABRIC breast cancers and TCGA datasets**

The TCGA pan-cancer level-3 somatic mutation and RNA expression data were downloaded from Synapse (<https://www.synapse.org/#!Synapse:syn300013>). Somatic mutation data for breast cancer (BRCA), lung adenocarcinoma (LUAD) and endometrial cancer (UCEC) from TCGA (43), were segregated according to p53 mutational status and mutational signatures 2 and 13, which have been ascribed to A3B (18), were determined for each tumour type. The Mann–Whitney–Wilcoxon statistical test was used to check the association between p53 mutation status and A3B gene expression and number of APOBEC mutational signatures 2 and 13. Association between p53 mutational status and A3B expression was similarly determined for the 2000 breast cancer samples in METABRIC (44).

**RESULTS****APOBEC3B expression is repressed by p53**

Previous analyses of gene expression datasets have indicated that A3B levels are elevated in breast cancers with somatic mutations in the p53 gene (TP53), compared with tumours with WT TP53 (17,45,46). In agreement with these findings, we observed a highly significant relationship between A3B expression and p53 mutational status in the METABRIC ( $P = 7.0 \times 10^{-8}$ ) and TCGA ( $P = 2.2 \times 10^{-16}$ ) breast cancer cohorts (Figure 1A). An association between A3B expression and p53 mutation was also seen in lung and endometrial cancer (Figure 1B). To further confirm this relationship by RT-qPCR, we analysed gene expression in RNA prepared from 115 primary breast cancers, collected prior to any therapy. A3B expression was significantly greater in p53 mutant tumours ( $P = 0.006$ ; Figure 1C) and was also higher in breast cancer cell lines with mutant, compared with WT p53 ( $P = 0.04$ ; Figure 1D).

In order to further understand the link between p53 status and A3B expression, we looked for evidence of A3B regulation by p53. Treatment with the p53 activator Nutlin-3 (hereafter referred to as Nutlin) was found to reduce A3B expression in breast cancer cell lines with WT p53 (Figure 2A and B). Nutlin did not affect A3B expression in breast cancer cells with p53 mutations. Specificity of Nutlin action on A3B was indicated by the fact that A3B repression occurred over a Nutlin dose range and time course that repressed other well-characterized p53-repressed genes (survivin, CHEK1, CHEK2) (47) and which induced expression of p53-activated genes (MDM2, p21) (Supplementary Figures S1 and 2). Expression of A3B and other p53-repressed genes was elevated and there was almost complete loss in expression of p53-activated genes in MCF7 cells in which the p53 gene was inactivated by CRISPR-Cas9 genome



**Figure 1.** APOBEC3B expression is elevated in breast cancer with mutated p53. (A) Analysis of A3B expression in METABRIC and TCGA breast cancer samples. For METABRIC, p53 mutational status was available for 820 samples ( $n = 99$  mutant p53) and 721 (WT p53). The TCGA dataset comprised 297 samples with mutant and 802 samples with wild-type (WT) p53. (B) A3B expression analysis for the TCGA gene expression datasets for lung and endometrial cancer according to p53 mutational status. (C) RT-qPCR shows that A3B expression is higher in p53 mutant breast cancers than in p53 WT tumours. A3B expression is shown relative to GAPDH levels. (D) RT-qPCR of 50 breast cancer cell lines shows that A3B expression is elevated in cell lines with p53 mutations. A3B expression is shown relative to expression of GAPDH. Asterisks show cell lines (SkBr3, HCC38) that encode the A3A-A3B variant and so do not express A3B.

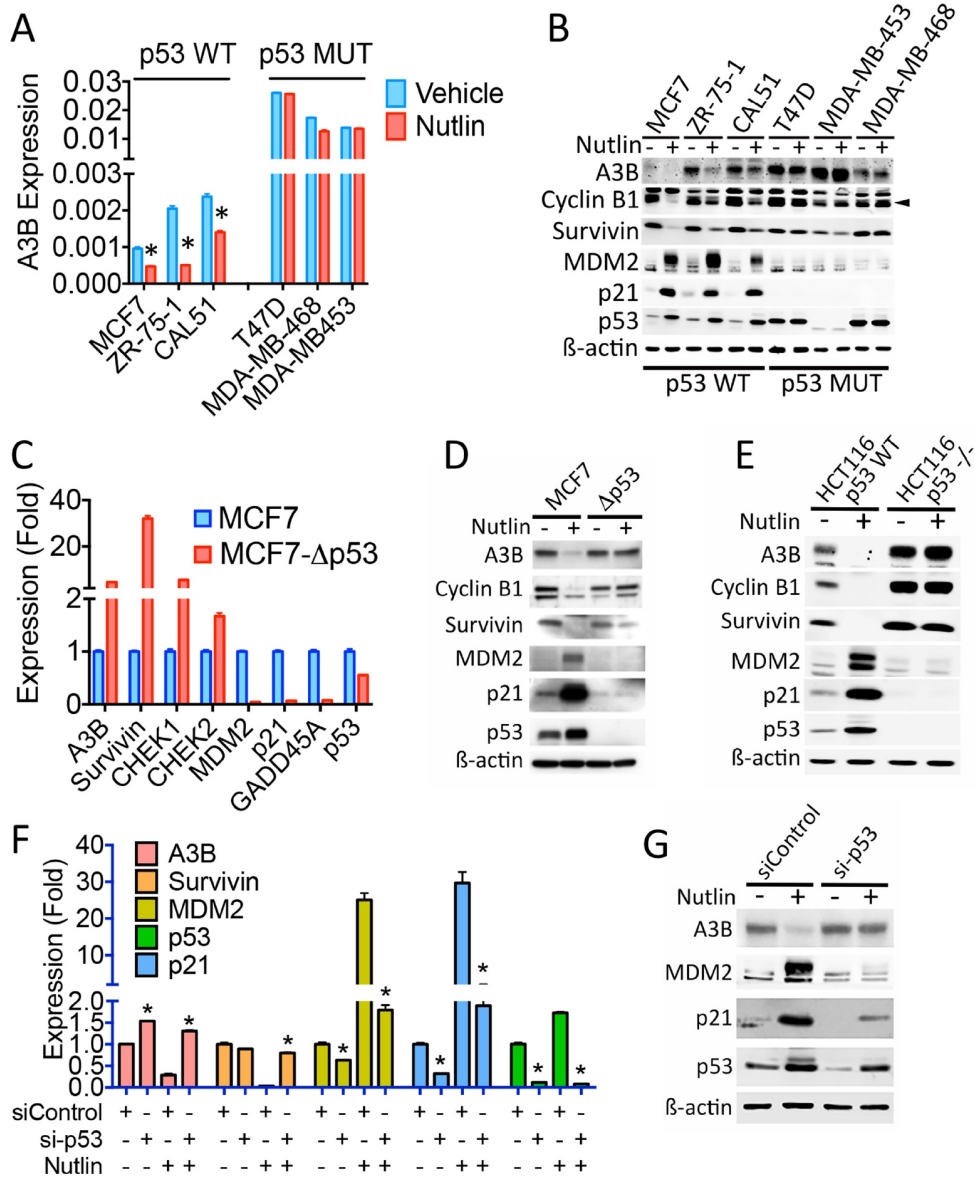
editing (Figure 2C and D). We note that despite appreciable levels of p53 mRNA, there was no detectable p53 protein in MCF7 cells following CRISPR-Cas9-mediated deletion of exon 3. A3B expression was also elevated in p53-null HCT116 colon cancer cells (HCT116-p53<sup>-/-</sup>), compared with the isogenic line expressing WT p53 (Figure 2E). Moreover, Nutlin did not inhibit A3B expression in p53-null HCT116 cells. Finally, siRNA-mediated p53 knockdown prevented A3B repression by Nutlin (Figure 2F and G). These experiments demonstrate clearly that A3B expression is regulated by p53 and indicate that loss of p53 activity due to its mutation, results in elevated expression of A3B.

#### p53 regulation of APOBEC3B expression is mediated by the E2F4/RB-containing DREAM repressive complex

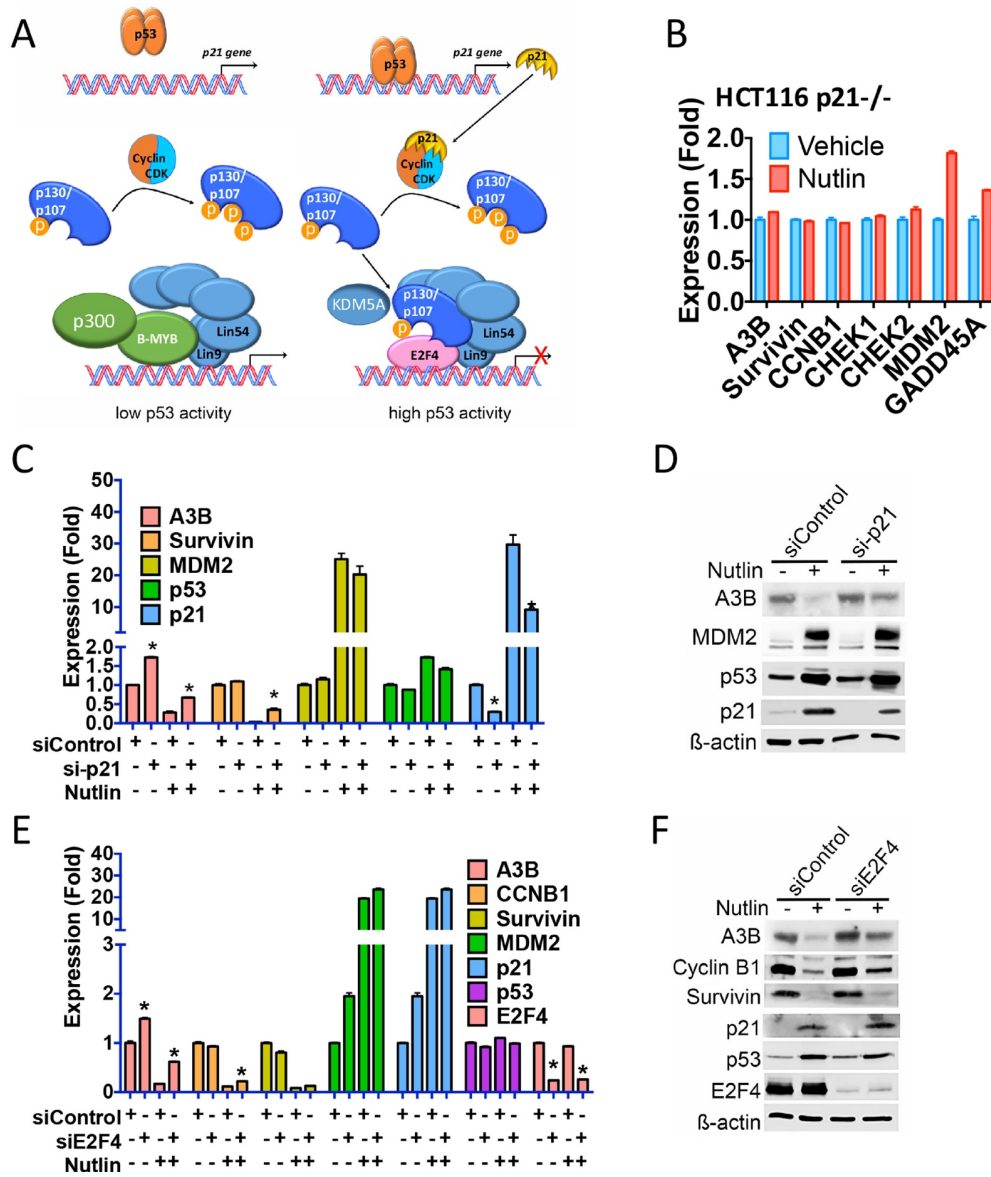
p53 ChIP following Nutlin treatment of HCT116 cells showed enrichment of p53 at the MDM2, GADD45A and CDKN1A (p21) gene promoters, but not at the A3B promoter (Supplementary Figure S3A). Analysis of available ChIP-seq datasets for several cell lines also failed to provide evidence for p53 binding within 20 kb of the A3B gene transcription start site, as exemplified for U2OS cells (48)

(Supplementary Figure S3B). There was also absence of p53 at promoters of other p53-repressed genes such as the survivin (BIRC5) gene. By contrast, Nutlin strongly promoted p53 recruitment to the MDM2 and p21 gene promoters.

Previous studies have shown that p53 induces expression of target genes by direct recruitment to gene promoters, whereas gene repression by p53 is indirect, frequently involving p21 (47). p21-directed inhibition of cyclin-dependent kinases prevents hyperphosphorylation of p107/p130 retinoblastoma (RB) proteins, promoting the conversion of the so-called multi-B-MYB-multi-val class B transcription activation complex, to the E2F4/p107/p130-containing DREAM (dimerization partner, RB-like, E2F and MuvB) repressive complex (Figure 3A) (49). In p21-null HCT116 cells, Nutlin did not repress expression of A3B or other p53-repressed genes (Figure 3B). Inhibition of A3B expression was also blunted by siRNA-mediated p21 knockdown (Figure 3C and D). Inhibition of A3B expression by Nutlin was alleviated by E2F4 knockdown (Figure 3E and F), and was also rescued by siRNA for Lin9 and Lin54, which are key components of the DREAM complex (Supplementary Figure S4).



**Figure 2.** p53 represses APOBEC3B expression in cancer cells. Nutlin (10  $\mu$ M) was added for 24 h in all experiments. (A) RT-qPCR of WT and mutant p53 breast cancer cell lines, treated with Nutlin ( $n = 3$ ). A3B expression is shown relative to GAPDH. Asterisks show significant ( $P < 0.05$ ) differences between vehicle and Nutlin-treated cells. (B) Immunoblotting of cell lysates following Nutlin treatment. The filled triangle shows position of Cyclin B1. (C) RT-qPCR of MCF7 cells in which exon 3 of the TP53 gene was targeted using CRISPR-Cas9 (MCF7- $\Delta$ p53). Expression of all examined genes was significantly ( $P < 0.05$ ) different between parental and  $\Delta$ p53 MCF7 cells. (D) Immunoblotting of protein lysates from MCF7 and MCF7- $\Delta$ p53 cells. (E) Protein lysates from HCT116 and p53-null HCT116 (HCT116-p53<sup>-/-</sup>) cells  $\pm$  Nutlin. (F) Twenty-four hours following transfection of HCT116 cells with siRNA for p53, Nutlin was added. RT-qPCR was performed using RNA prepared 48 h following addition of Nutlin. (G) Immunoblotting of HCT116 cells transfected with si-p53.



**Figure 3.** p53 regulation of APOBEC3B expression is mediated by p21 acting through the E2F4 DREAM transcriptional complex. (A) Shown is a model depicting the mechanism by which the DREAM complex regulates gene expression following p53 activation. (B) RT-qPCR for p21-null HCT116 cells treated with Nutlin for 24 h. (C) RT-qPCR of HCT116 cells transfected with p21 siRNA. Nutlin was added for 24 h. Significant ( $P < 0.05$ ;  $n = 3$ ) differences between vehicle and Nutlin-treated cells are highlighted by asterisks. (D) Immunoblotting of HCT116 cell lysates following p21 knockdown. (E and F) RT-qPCR and immunoblotting of HCT116 cells transfected with siRNA for E2F4.



ChIP, followed by real-time PCR (ChIP-qPCR), showed that Nutlin stimulates E2F4, p130 and LIN9 recruitment to the A3B gene promoter, with concomitant loss of the transcriptional activators B-MYB and p300 (Figure 4A–F; summarized in the heat map in Figure 4I). The Nutlin-stimulated transition from the activation complex to the repressive complex was accompanied by reduction in the histone marks, H3K9Ac and H3K4me3, both of which are high at the transcription start sites of active genes (Figure 4G and H; Supplementary Figure S3B). KDM5A/JARID1A, originally identified as a protein that binds to RB, contributes to the repression of E2F4 target genes by removing di- and tri-methyl groups from H3K4 (50,51). ChIP for KDM5A showed Nutlin stimulated KDM5A recruitment to the A3B promoter (Figure 4D), explaining the reduction in H3K4me3. Nutlin treatment similarly promoted E2F4, p130, LIN9 and KDM5A recruitment to the A3B promoter and reduced B-MYB, p300, H3K4me3 and H3K9Ac in HCT116 cells (Figure 4J and Supplementary Figure S5A). The dependence on p53 for the Nutlin stimulated gain of E2F4/p130/LIN9/KDM5A at the A3B gene promoter was confirmed in p53-null HCT116 cells (Figure 4J and Supplementary Figure S5B). Furthermore, there was no reduction in B-MYB or p300 enrichment at the A3B gene promoter in HCT116-p53<sup>-/-</sup> cells; nor was there a reduction in levels of histone H3 marks associated with active transcription. The importance of p21 in directing the recruitment of the DREAM complex was confirmed by the fact that its Nutlin promoted enrichment at the A3B gene promoter was prevented in p21-null HCT116 cells (Figure 4J and Supplementary Figure S5C).

#### **p53 inhibits the mutational capacity of cancer cells by repressing APOBEC3B expression**

The above results show that p53 represses A3B expression by directing the E2F4/p107/p130-containing DREAM complex to the A3B promoter and predict that p53 controls A3B expression to limit its mutagenic potential. Indeed, cytosine deaminase activity was generally higher in extracts from mutant p53 lines than in lysates from cells with WT p53 (Figure 5A and Supplementary Figure S6A), consistent with the elevated A3B expression levels in breast cancer cells with mutant p53. Importantly, Nutlin reduced cytosine deaminase activity in lysates from WT-p53 MCF7 and ZR-75-1 cells, but not in those from mutant p53 cells. The p53-dependence of Nutlin-mediated reduction in cytosine deaminase activity was confirmed in HCT116 and p53-null HCT116 cells (Figure 5B and Supplementary Figure S6B). Note that cytosine deaminase activity was strongly reduced by A3B siRNA (Figure 5C and Supplementary Figure S6C), consistent with the fact that A3B is the main APOBEC expressed in breast cancer cells (7,17).

Cytosine deamination by APOBECs generates U:G mismatches that are excised by uracil DNA glycosylases, to generate apurinic/abasic sites that can be processed by AP endonuclease (52). The aldehyde group on the open ring of AP sites can be labelled with an 'ARP' containing biotin, thus allowing detection and quantification of AP sites in the genome (53). Interestingly, Nutlin treatment led to reductions in genomic AP sites in cells with WT-p53, but not

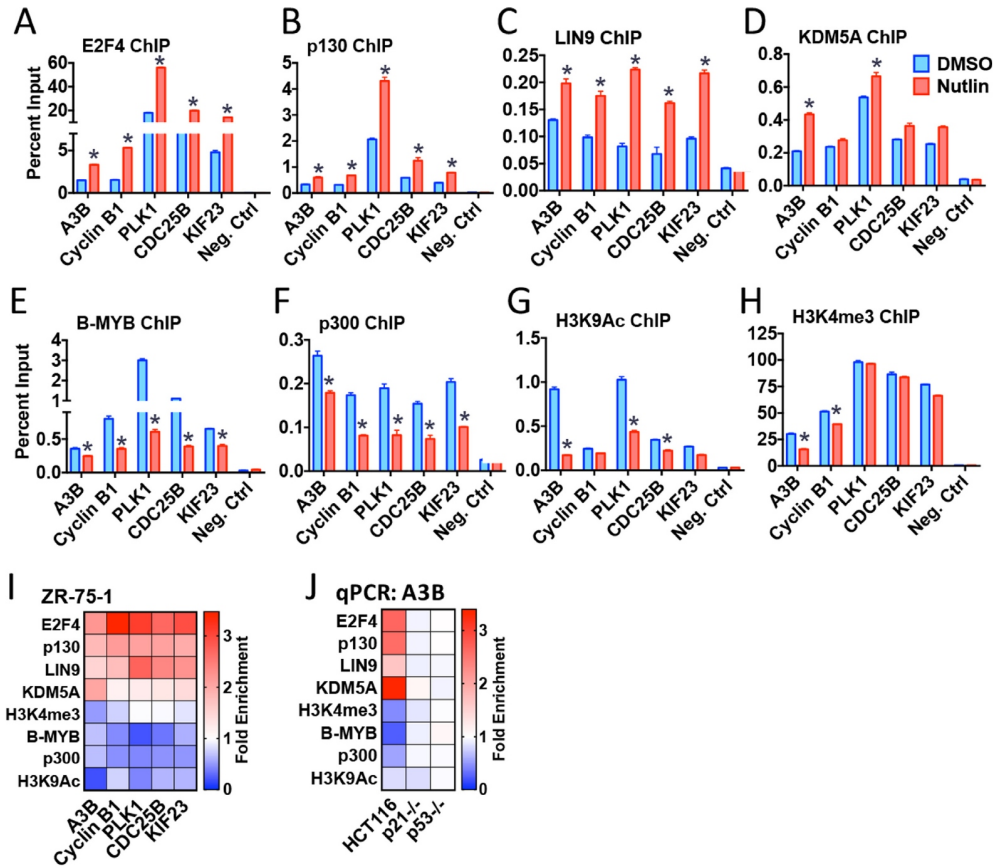
in cells with p53 mutations (Figure 5D). Nor did Nutlin affect AP sites in p53-null HCT116 cells (Figure 5E). Furthermore, in ZR-75-1 cells, A3B knockdown reduced AP sites and Nutlin did not further inhibit AP sites, indicating that the reduction in abasic sites is almost entirely mediated by the Nutlin/p53 regulation of A3B (Figure 5F).

Sequencing of tumour DNAs representing diverse cancer types, has identified patterns of DNA base alterations that are characteristic of the enzymatic activity of A3B (17–23). Interrogation of exome sequence datasets from the TCGA database showed that the frequency of mutations associated with the A3B mutational signatures was significantly higher for mutant p53 tumours compared with WT-p53 in breast ( $P = 2.07 \times 10^{-6}$ ), lung ( $P = 2.53 \times 10^{-8}$ ) and endometrial cancer ( $P = 0.048$ ) (Figure 5G). In conclusion, p53 restricts the mutagenic activity of A3B by repressing its expression, and loss of p53 in cancer results in elevated A3B expression, which may promote somatic mutagenesis in cancer.

#### **HPV E6 and E7 gene-mediated repression of p53 and the E2F4/p107/p130-containing DREAM complex promotes APOBEC3B expression and activity**

Given that p53 downregulation and Rb family inactivation are key functional targets of the high-risk HPV E6 and E7 genes (33), we ascertained whether the transcriptional mechanism for p53-mediated repression of A3B expression identified here is subverted by these viral oncogenes. Indeed, A3B expression was elevated in HPV16 E6 transfected HCT116 cells, but was unaffected by E6 in p53-null HCT116 cells (Figure 6A and B). Stimulation of A3B expression was prevented in cells transfected with an E6 mutant that is defective for interaction with p53 (54). Moreover, E6, but not mutant E6, abrogated repression of A3B by Nutlin. Consistent with these effects being mediated by inhibition of p53 action by E6, p21 expression was reduced by WT, but not mutant E6.

To further evaluate the mechanisms of E6 and E7 regulation of p53 activity and consequence for A3B expression, we repeated the analysis in human keratinocytes, the natural HPV target cell. For this purpose we used normal immortalized human keratinocytes (NIKS), a spontaneously immortalized but not transformed keratinocyte cell line (55), stably expressing E6, E7, E6 and E7 (E6/E7) (36). In agreement with the HCT116 results, p53 expression was abrogated in NIKS transduced with E6, but not mutant E6 (Figure 6C). This was accompanied by an increase in A3B and a reduction in expression of the p53-induced p21 and MDM2 genes (Figure 6C and D). E6 expression also prevented the Nutlin-mediated repression of A3B, survivin and cyclin B1, as well as induction of p21 and MDM2. In concordance with the described role for E7 in inhibiting the activity of Rb family members, expression of A3B and the other p53-repressed genes was elevated in E7-expressing NIKS, but E7 did not affect direct p53 targets (p21, MDM2). In this context, the inhibitory effect of HPV-16 E7 on p21 might contribute to the enhanced expression of p53-repressed genes (56,57). A recent meta-analysis of global HPV E7-regulated gene expression profiling datasets indicates that HPV-16 E7 can indeed activate expression of DREAM complex-repressed genes (58), including that of A3B (31,58), confirming our

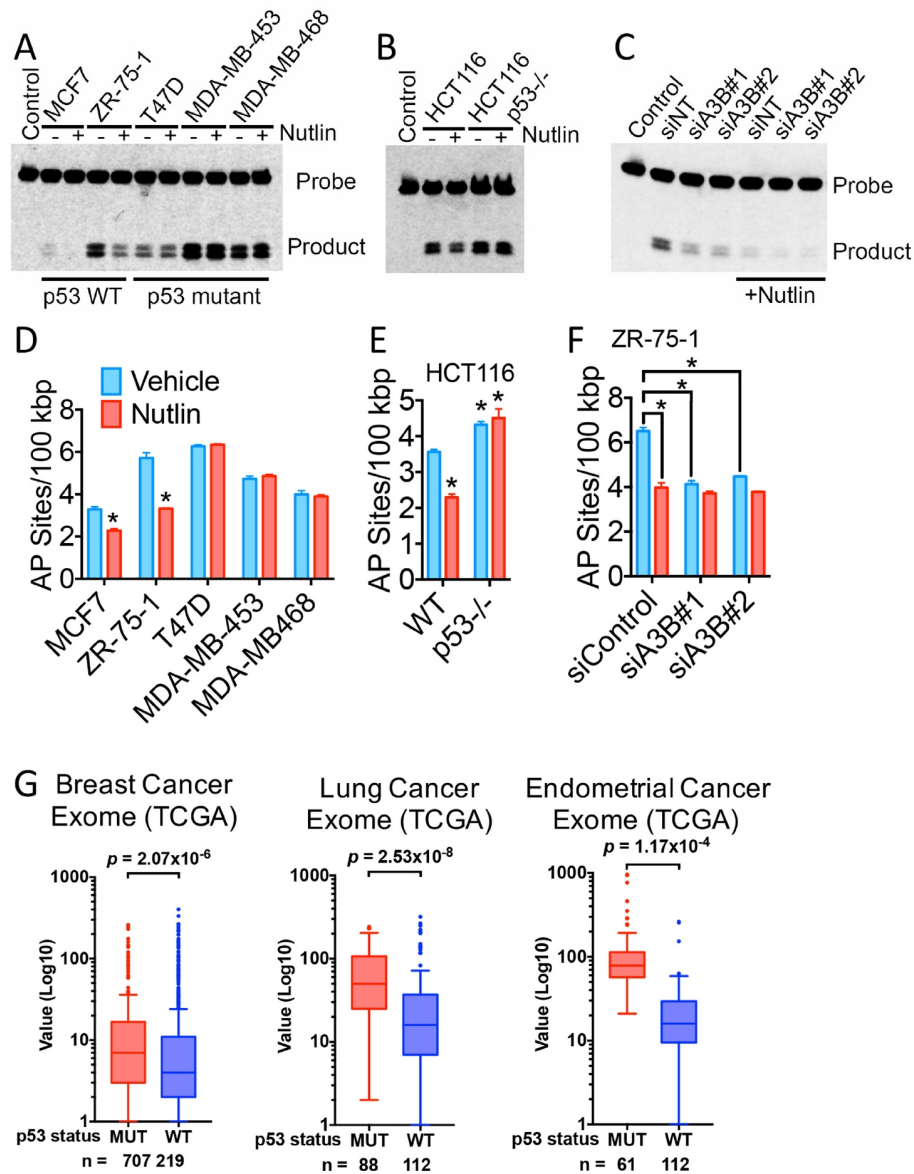


**Figure 4.** ChIP analysis of DREAM complex enrichment at the APOBEC3B gene promoter. (A–H) ZR-75–1 cells were treated with Nutlin (10  $\mu$ M, 24 h), followed by ChIP for transcription factors in the DREAM complex and for histone marks associated with active genes. Asterisks identify significant ( $P < 0.05$ ;  $n = 3$ ) differences in transcription factor recruitment and histone marks for the Nutlin-treated cells, relative to vehicle controls. (I) ChIP-qPCR for Nutlin-treated samples are shown, as fold enrichment relative to vehicle. (J) ChIP-qPCR for the A3B gene in HCT116 cells, in p21-null or in p53-null HCT116 cells. The heat map shows Nutlin-promoted changes in factor recruitment to the A3B gene, relative to vehicle controls. The full ChIP-qPCR data are shown in Supplementary Figure S5.

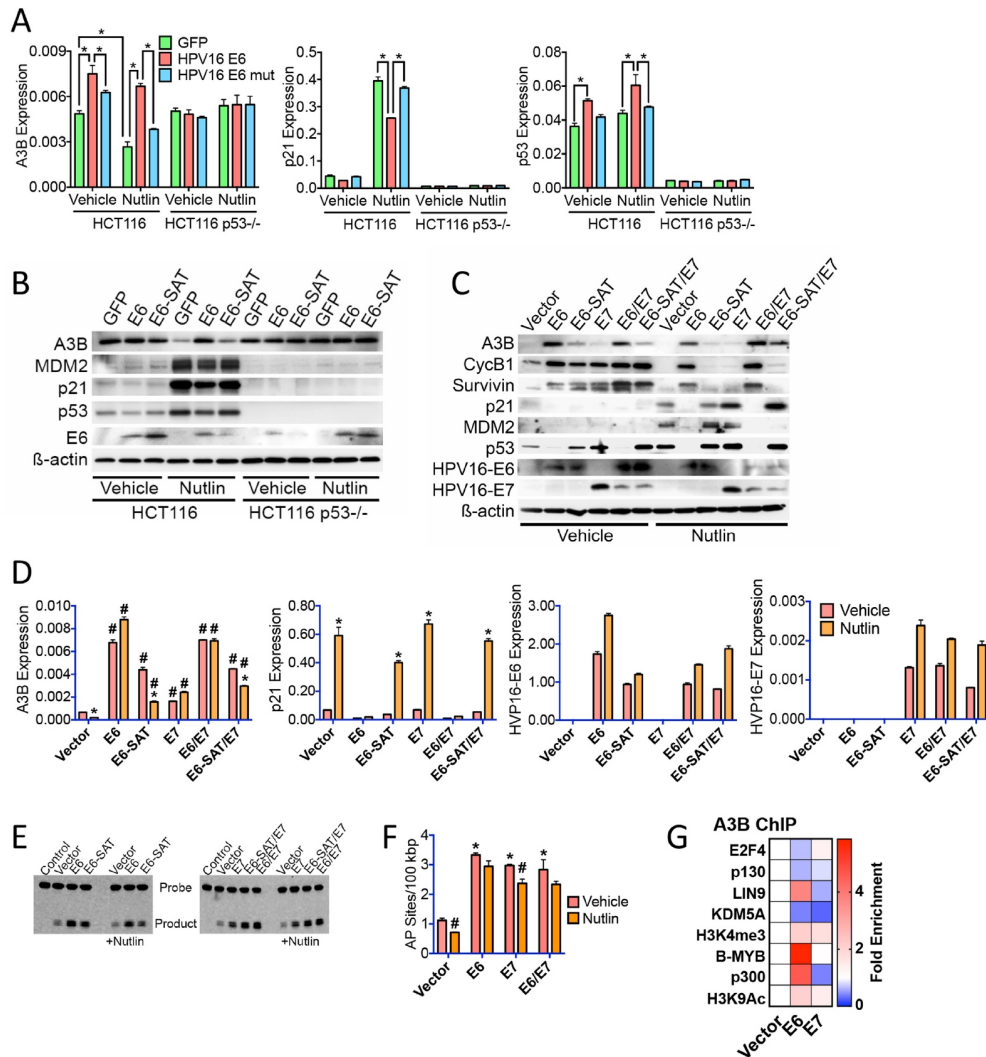
findings. However, E7 expression had a notably milder effect on A3B upregulation, compared to E6 (Figure 6C and D). This is consistent with previous studies in NIKS, where the HPV-16 E7-mediated inhibition of RB family members and effects on cell growth are alleviated by presence of growth factors in culture medium (36). Cytosine deaminase activity was elevated in NIKS expressing E6 and/or E7 (Figure 6E), as was the number of abasic sites detectable in genomic DNA (Figure 6F), consistent with elevated A3B expression and activity.

As observed in other cell types, p53 was not recruited to the A3B gene in NIKS (Supplementary Figure S7A). In agreement with its role in inhibiting p53, there was

duced p53 recruitment at the MDM2 and p21 promoters in E6 expressing NIKS, but not in cells expressing E7. B-MYB and p300 recruitment, as well as histone methylation and acetylation associated with gene activity, were elevated by E6 (Figure 6G). Conversely, recruitment of the repressive DREAM complex (E2F4, p130), as well as KDM5A, was reduced. Whilst B-MYB and p300 recruitments were not increased by E7, recruitment of the repressors p130 and KDM5A was reduced at the A3B promoter, accompanied by increased H3K4 tri-methylation. Similar results were obtained for the other p53-repressed genes analysed (Supplementary Figure S7). Taken together, these results demonstrate that stimulation of A3B expression by HPV



**Figure 5.** Inhibition of APOBEC3B expression by p53 controls mutagenesis in cancer cells. (A and B) Protein lysates prepared from Nutlin-treated cells were used in an *in vitro* cytidine deaminase assay. Positions of the substrate (probe) and the deamination product, are labeled. (C) Cytidine deamination with lysates prepared from ZR-75-1 cells, following transfection with two independent A3B siRNAs. (D-F) AP or abasic sites in genomic DNA were biotin labelled by conversion with an aldehyde reactive probe (ARP) and quantification of biotinylated DNA. Asterisks show significant ( $P < 0.05$ ;  $n = 3$ ) differences in AP sites for Nutlin-treated samples, relative to vehicle controls. (G) Analysis of A3B mutational signature for WT and mutant p53 in breast, lung and endometrial cancer from TCGA. Values on the y-axis represent the number of A3B mutational signatures 2 and 13 in each cancer.  $P$ -values were calculated using Mann-Whitney-Wilcoxon statistical test.



**Figure 6.** HPV16 E6 and E7 viral oncogenes promote APOBEC3B expression and activity by inhibiting the action of the E2F4/p107/p130 DREAM transcriptional repression complex. (A) HCT116 and p53-null HCT116 cells were transfected with HPV16 E6 or the E6-SAT mutant that does not interact with p53. RT-qPCR was performed using RNA prepared 24 h following addition of Nutlin (10  $\mu$ M). Asterisks show significant ( $P < 0.05$ ;  $n = 3$ ) differences in mRNA expression. (B) Cells treated as above, were processed for immunoblotting. (C) Normal immortalized human keratinocyte line (NIKS) stably transfected with E6, the E6-SAT mutant, E7, E6/E7, E7/mutant E6 or vector only, were treated with Nutlin (10  $\mu$ M) for 24 h, followed by preparation protein lysates for immunoblotting. (D) NIKS were treated as in C, followed by preparation of total RNA and RT-qPCR. Asterisks show significant ( $P < 0.05$ ) reduction in mRNA expression relative to the vehicle treated and vector transfected NIKS for each transfected line. Cross-hatches show significant ( $P < 0.05$ ) difference in mRNA expression relative to the vehicle treated and vector transfected NIKS. Results for three experiments are shown. (E) The *in vitro* cytidine deaminase assay was used to assess A3B activity in protein lysates prepared 24 h following addition of 10  $\mu$ M Nutlin to the E6 and E7 transfected NIKS cells. (F) Quantification of biotin-labelled ARP conversion of abasic sites in genomic DNA from E6 and E7 transfected NIKS cells. Asterisks show significant ( $P < 0.05$ ) differences in AP sites compared to vehicle-treated vector transfected NIKS; cross-hatches denote statistically significant reduction in AP sites by Nutlin treatment for each transfection. (G) ChIP-qPCR for vector, E6 or E7 transfected NIKS cells is shown as fold change in transcription factor/histone mark enrichment at the A3B gene promoter, relative to vehicle controls. The actual factor enrichment relative to input is shown in Supplementary Figure S7B.

E6 and E7 oncogenes is mediated by inhibition of the p53 directed transcriptional repression and the stimulation of transcriptional activation by the B-MYB/E2F4/p107/p130 DREAM complex.

## DISCUSSION

APOBEC3 genes play important roles in innate immunity and act by causing hypermutation of retroviral genomes and are implicated in reducing retrotransposon mobilization. Their mutational activities bring an inherent risk to the host genome, as shown by ectopic expression of APOBEC3 genes, which can demonstrably cause genomic mutations (14–17). Therefore, to safeguard genomic integrity, tight regulation of APOBEC expression is necessary in order to suppress their mutagenic potential. Indeed, there is a strong evidence that over-expression of A3B causes hypermutation in many cancer types. We observed that A3B expression is higher in breast cancers with mutant p53 than in those with WT p53, in agreement with previous reports (17,45,46). We have found that a similar relationship between A3B expression and p53 status can be extended to lung and endometrial cancer, cancer types in which the mutational landscape is marked by mutations consistent with A3B activity (20,21,23). These findings originally suggested that A3B might be involved in the mutation of the p53 gene. Indeed, A3B over-expression in breast cancer cell lines has been reported to promote cytosine deamination in the p53 gene (17). However, the majority of p53 mutations in cancer do not appear to conform to the TCW DNA motif targeted by A3B, we investigated whether A3B expression is regulated by p53 (ref. (59)), although as has been pointed out, p53 mutations detected in tumours will not simply reflect the exonic nucleotide sequence, but the selection imparted by the resultant amino acid change (60).

Notwithstanding, we show here that p53 controls APOBEC3B expression. Loss of p53 by siRNA-mediated knockdown, gene disruption or HPV E6 expression, increases A3B expression, while Nutlin activation of p53 represses A3B expression in multiple cell types, strong evidence that p53 is a negative regulator of A3B. Several mechanisms directing repression of p53 target genes have been reported, including direct binding of p53, or indirect recruitment via interaction with other transcription factors such as NF-Y (47,61,62). Despite the presence of potential p53 response element sequences in the A3B gene (48), analysis of diverse p53 ChIP-seq datasets provides little evidence for p53 recruitment to the A3B promoter. Indeed, in only one out of eight genome-wide p53 binding datasets analysed by Menendez *et al.* (48) was there evidence for p53 binding at the A3B promoter, who also reported weak Nutlin- and doxorubicin-induced p53 recruitment to the A3B gene (48). Moreover, we did not observe enrichment of p53 at the A3B promoter in ZR-75-1, HCT116 or NIKS, arguing against an important role for direct or indirect p53 recruitment to the A3B promoter. Furthermore, p53 activation did not repress A3B expression in p21-null HCT116 cells and p21 knockdown rescued A3B repression by Nutlin, demonstrating the importance of p21 in the repression of A3B by p53.

The MuvB complex, together with B-MYB drives expression of many cell-cycle regulated genes through the S phase and G2/M, recruitment to the cell cycle genes homology region (CHR) element being mediated by the LIN54 component of MuvB (49,63), whereas MuvB, together with E2F4 and p107/p130, known as the DREAM complex, represses expression of these genes. E2F4 and p107/p130 recruitment to this complex is promoted by p21-mediated inhibition of CDK directed p107/p130 hyperphosphorylation. ChIP experiments have demonstrated (i) the presence of B-MYB and the MuvB subunit LIN9 at the A3B promoter, together with the B-MYB associating p300 histone acetyltransferase, (ii) a reduction in B-MYB and p300 recruitment upon Nutlin treatment, accompanied by stimulation of E2F4 and p130 recruitment, as well as the RB-associated histone lysine demethylase KDM5A, (iii) together with a reduction in the levels of histone marks associated with active transcription. The role of the B-MYB/MuvB complex is confirmed by a recent study which showed that B-MYB is a transcriptional driver of A3B expression and that B-MYB and A3B expression is correlated in many cancer types (64). We further show that deletion of p53 or of p21 prevented Nutlin-mediated loss of B-MYB/p300 or gain of E2F4/p130/KDM5A. HPV16 E6 abrogated p21 expression, resulting in reduced E2F4/p130/KDM5A recruitment and promoting B-MYB and p300 binding and stimulation of activation histone marks at the A3B promoter. E7 also inhibited recruitment of p130 and KDM5A, consistent with action of E7 in inhibiting gene repression by RB proteins (33).

The presence of E2F4, LIN9, LIN54 and p130 at the A3B promoter are confirmed in global ChIP–chip experiments (49,65), and identify a region in the A3B promoter encoding a potential cell-cycle-dependent element and a CHR element, 5'-GGGAGGtcacTTAAG-3' ~50 bp upstream of the A3B transcription start site. Interestingly, expression of the TEA domain (TEAD) transcription factors and their co-activators YAP/TAZ is stimulated by HPV16 E6 directed degradation of p53 and YAP/TAZ/TEAD promotes A3B expression (32). Taken together, these findings further support a critical role for p53 in regulating A3B expression, with repression of A3B expression by the p53-mediated displacement of the MYB-MuvB transcription activation complex by the E2F4/p107/p130 DREAM transcriptional repression complex. We would speculate that p53 also displaces the YAP/TAZ/TEAD complex from the A3B promoter by repressing YAP, TAZ and TEAD expression.

Nutlin treatment results in high level expression of the direct p53 target gene, the MDM2 E3 ubiquitin ligase. MDM2 interacts with pRb to promote its ubiquitination and proteasomal degradation (66,67). Nutlin treatment also stimulates MDM2-dependent accumulation of hypophosphorylated pRb (68). Although MDM2 also interacts with p107 and p130, it does not promote their ubiquitination or proteasomal degradation (67). E2F1 also interacts with MDM2 and Nutlin regulates proteasomal degradation of E2F1 (69,70). However, we failed to observe an effect of Nutlin on E2F4 protein levels, arguing against a role for Nutlin and/or MDM2 driven proteasomal degradation of p107/p130 or E2F4 in p53-mediated repression of A3B expression.

It is nonetheless likely that other signals impact on A3B expression by acting to alter the expression and/or activities of one or more of the factors in the MuvB and/or DREAM complexes. For example, it has recently been shown that the PI3K/AKT pathway promotes gene expression by phosphorylating KDM5A, to relocalize it to the cytoplasm (71), which would provide a potential mechanism for the described high level A3B expression/activity associated with HER2 amplification and/or PTEN loss (72). Further, co-expression analysis of breast and other cancer types demonstrated was reported to show strong enrichment for mitosis and cell-cycle-associated functional ontology groups, with A3B expression (46), which is in keeping with the established role of MuvB and DREAM complexes in the regulating expression of cell cycle and mitosis genes (49,63).

In response to genotoxic and non-genotoxic insult that challenges genomic integrity, p53 mediates innate tumour suppression by altering expression of genes to favour biological functions such as cell cycle arrest, apoptosis and senescence. Our results indicate that p53 also protects the genome by limiting the mutational potential of genes whose primary cellular role is in mutagenic inactivation of foreign and mobilizable DNA. We have shown here that failure to suppress A3B expression following mutational or viral inactivation of p53, results in elevated A3B expression and activity, with attendant increase in potential genetic mutations, as demonstrated by the ability of activated WT, but not mutant, p53 to suppress abasic site generation in genomic DNA and through the demonstration that the frequency of mutations ascribed to A3B activity in diverse cancer types is elevated in p53 mutant tumours. Interestingly, in its recently identified role in regulating the expression of ER target genes in breast cancer cells (7), elevated A3B expression could promote transcriptional programs that aid breast cancer progression, which is consistent with association between high A3B expression and poor patient survival in ER-positive breast cancer (7,46,73) and response to the anti-estrogen tamoxifen (74). Thus, elevated A3B expression and activity due to p53 inactivation likely have important consequences for tumour development and tumour evolution, including response to therapies, both through its role in transcription and because of its potential for inflicting mutational damage.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

#### ACKNOWLEDGEMENTS

S.A. designed the study and compiled the manuscript with advise and help from all the authors. C.K., T.R.F. and J.D. led the HPV work. F.F-P. and D.W.M. provided clinical samples and data interpretation. N.N. provided reagents and advise on APOBEC3 genes. H.L.P. and B.G.V. provided advise and interpretation of the APOBEC3B activity assays. M.P., A.K.S., C.G., C.K. and R.F. carried out laboratory experiments and together with L.B. and R.C.C. interpreted the data.

#### FUNDING

Cancer Research UK [C37/A18784]; NIHR Biomedical Research Centre funding scheme; CRUK and Department of Health Imperial College Experimental Cancer Medicine Centre (ECMC); Rosetrees Trust [JS16/M229-CD1 to T.R.F.]; Breast Cancer Now [2010NovPR16 to F.F-P.]; Tayside Tissue Bank; National Research, Development and Innovation Office [K119493 to H.L.P., B.G.V.]. Funding for open access charge: Charity Open Access Fund (COAF) via the Imperial College Open Access Fund.  
*Conflict of interest statement.* None declared.

#### REFERENCES

- Chiu, Y.L. and Greene, W.C. (2008) The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu. Rev. Immunol.*, **26**, 317–353.
- Refsland, E.W. and Harris, R.S. (2013) The APOBEC3 family of retroelement restriction factors. *Curr. Top. Microbiol. Immunol.*, **371**, 1–27.
- Coticello, S.G. (2008) The AID/APOBEC family of nucleic acid mutators. *Genome Biol.*, **9**, 229.
- Franchini, D.M. and Petersen-Mahrt, S.K. (2014) AID and APOBEC deaminases: balancing DNA damage in epigenetics and immunity. *Epigenomics*, **6**, 427–443.
- Ramiro, A.R. and Barreto, V.M. (2015) Activation-induced cytidine deaminase and active cytidine demethylation. *Trends Biochem. Sci.*, **40**, 172–181.
- Bochtler, M., Kolano, A. and Xu, G.L. (2017) DNA demethylation pathways: additional players and regulators. *Bioessays*, **39**, 1–13.
- Periyasamy, M., Patel, H., Lai, C.F., Nguyen, V.T., Nevedomskaya, E., Harrod, A., Russell, R., Remenyi, J., Ochocka, A.M., Thomas, R.S. et al. (2015) APOBEC3B-mediated cytidine deamination is required for estrogen receptor action in breast cancer. *Cell Rep.*, **13**, 108–121.
- Di Noia, J.M. and Neuberger, M.S. (2007) Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.*, **76**, 1–22.
- Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K. et al. (2009) Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell*, **139**, 1069–1083.
- Robbiani, D.F. and Nussenzweig, M.C. (2013) Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. *Annu. Rev. Pathol.*, **8**, 79–103.
- Yamanaka, S., Balestra, M.E., Ferrell, L.D., Fan, J., Arnold, K.S., Taylor, S., Taylor, J.M. and Innerarity, T.L. (1995) Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 8483–8487.
- Okazaki, I.M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K. and Honjo, T. (2003) Constitutive expression of AID leads to tumorigenesis. *J. Exp. Med.*, **197**, 1173–1181.
- Ramiro, A.R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Honjo, T., Nussenzweig, A. and Nussenzweig, M.C. (2004) AID is required for c-myc/IgH chromosome translocations in vivo. *Cell*, **118**, 431–438.
- Nowarski, R., Wilner, O.I., Cheshin, O., Shahar, O.D., Kenig, E., Baraz, L., Britan-Rosich, E., Nagler, A., Harris, R.S., Goldberg, M. et al. (2012) APOBEC3G enhances lymphoma cell radioresistance by promoting cytidine deaminase-dependent DNA repair. *Blood*, **120**, 366–375.
- Landry, S., Narvaiza, I., Linfesty, D.C. and Weitzman, M.D. (2011) APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep.*, **12**, 444–450.
- Akre, M.K., Starrett, G.J., Quist, J.S., Temiz, N.A., Carpenter, M.A., Tutt, A.N., Grigoriadis, A. and Harris, R.S. (2016) Mutation processes in 293-based clones overexpressing the DNA cytosine deaminase APOBEC3B. *PLoS One*, **11**, e0155391.
- Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N.,

- Nikas, J.B. *et al.* (2013) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*, **494**, 366–370.
18. Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A. *et al.* (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell*, **149**, 979–993.
  19. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415–421.
  20. Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G. *et al.* (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.*, **45**, 970–976.
  21. Burns, M.B., Temiz, N.A. and Harris, R.S. (2013) Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.*, **45**, 977–983.
  22. Leonard, B., Hart, S.N., Burns, M.B., Carpenter, M.A., Temiz, N.A., Rathore, A., Vogel, R.I., Nikas, J.B., Law, E.K., Brown, W.L. *et al.* (2013) APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. *Cancer Res.*, **73**, 7222–7231.
  23. de Bruin, E.C., McGranahan, N., Mitter, R., Salm, M., Wedge, D.C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A.J. *et al.* (2014) Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*, **346**, 251–256.
  24. Swanton, C., McGranahan, N., Starrett, G.J. and Harris, R.S. (2015) APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discov.*, **5**, 704–712.
  25. Taylor, B.J., Nik-Zainal, S., Wu, Y.L., Stebbings, L.A., Raine, K., Campbell, P.J., Rada, C., Stratton, M.R. and Neuberger, M.S. (2013) DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *Elife*, **2**, e00534.
  26. Leonard, B., McCann, J.L., Starrett, G.J., Kosyakov, L., Luengas, E.M., Molan, A.M., Burns, M.B., McDougle, R.M., Parker, P.J., Brown, W.L. *et al.* (2015) The PKC/NF-kappaB signaling pathway induces APOBEC3B expression in multiple human cancers. *Cancer Res.*, **75**, 4538–4547.
  27. Madsen, P., Anant, S., Rasmussen, H.H., Gromov, P., Vorum, H., Dumanski, J.P., Tommerup, N., Collins, J.E., Wright, C.L., Dunham, I. *et al.* (1999) Psoriasis upregulated phorbol-1 shares structural but not functional similarity to the mRNA-editing protein apobec-1. *J. Invest. Dermatol.*, **113**, 162–169.
  28. Henderson, S., Chakravarthy, A., Su, X., Boshoff, C. and Fenton, T.R. (2014) APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell Rep.*, **7**, 1833–1841.
  29. Ohba, K., Ichihama, K., Yajima, M., Gemma, N., Nikaido, M., Wu, Q., Chong, P., Mori, S., Yamamoto, R., Wong, J.E. *et al.* (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**, e97787.
  30. Vieira, V.C., Leonard, B., White, E.A., Starrett, G.J., Temiz, N.A., Lorenz, L.D., Lee, D., Soares, M.A., Lambert, P.F., Howley, P.M. *et al.* (2014) Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B. *Mbio*, **5**, doi:10.1128/mBio.02234-14.
  31. Warren, C.J., Xu, T., Guo, K., Griffin, L.M., Westrich, J.A., Lee, D., Lambert, P.F., Santiago, M.L. and Poon, D. (2015) APOBEC3A functions as a restriction factor of human papillomavirus. *J. Virol.*, **89**, 688–702.
  32. Mori, S., Takeuchi, T., Ishii, Y., Yugawa, T., Kiyono, T., Nishina, H. and Kukimoto, I. (2017) Human papillomavirus 16 E6 upregulates APOBEC3B via the TEAD transcription factor. *J. Virol.*, **91**, doi:10.1128/JVI.02413-16.
  33. zur Hausen, H. (2002) Papillomaviruses and cancer: from basic studies to clinical application. *Nat. Rev. Cancer*, **2**, 342–350.
  34. Waldman, T., Kinzler, K.W. and Vogelstein, B. (1995) p21 is necessary for the p53-mediated G1 arrest in human cancer cells. *Cancer Res.*, **55**, 5187–5190.
  35. Bunz, F., Dutriaux, A., Lengauer, C., Waldman, T., Zhou, S., Brown, J.P., Sedivy, J.M., Kinzler, K.W. and Vogelstein, B. (1998) Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science*, **282**, 1497–1501.
  36. Kranjec, C., Holleywood, C., Libert, D., Griffin, H., Mahmood, R., Isaacson, E. and Doorbar, J. (2017) Modulation of basal cell fate during productive and transforming HPV16 infection is mediated by progressive E6-driven depletion of notch. *J. Pathol.*, **242**, 448–462.
  37. Lambert, P.F., Ozburn, M.A., Collins, A., Holmgren, S., Lee, D. and Nakahara, T. (2005) Using an immortalized cell line to study the HPV life cycle in organotypic “raft” cultures. *Methods Mol. Med.*, **119**, 141–155.
  38. White, E.A., Kramer, R.E., Tan, M.J., Hayes, S.D., Harper, J.W. and Howley, P.M. (2012) Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity. *J. Virol.*, **86**, 13174–13186.
  39. Rozenblatt-Rosen, O., Deo, R.C., Padi, M., Adelman, G., Calderwood, M.A., Rolland, T., Grace, M., Dricot, A., Askenazi, M., Tavares, M. *et al.* (2012) Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature*, **487**, 491–495.
  40. Halbert, C.L., Demers, G.W. and Galloway, D.A. (1992) The E6 and E7 genes of human papillomavirus type 6 have weak immortalizing activity in human epithelial cells. *J. Virol.*, **66**, 2125–2134.
  41. Bouaoun, L., Sonkin, D., Ardin, M., Hollstein, M., Byrnes, G., Zavadil, J. and Olivier, M. (2016) TP53 variations in human cancers: new lessons from the IARC TP53 database and genomics data. *Hum. Mutat.*, **37**, 865–876.
  42. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
  43. Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A. *et al.* (2013) Mutational landscape and significance across 12 major cancer types. *Nature*, **502**, 333–339.
  44. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y. *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**, 346–352.
  45. Silwal-Pandit, L., Vollen, H.K., Chin, S.F., Rueda, O.M., McKinney, S., Osako, T., Quigley, D.A., Kristensen, V.N., Aparicio, S., Borresen-Dale, A.L. *et al.* (2014) TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. *Clin. Cancer Res.*, **20**, 3569–3580.
  46. Cescon, D.W., Haibe-Kains, B. and Mak, T.W. (2015) APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 2841–2846.
  47. Fischer, M., Steiner, L. and Engeland, K. (2014) The transcription factor p53: not a repressor, solely an activator. *Cell Cycle*, **13**, 3037–3058.
  48. Menendez, D., Nguyen, T.A., Freudenberg, J.M., Mathew, V.J., Anderson, C.W., Jothi, R. and Resnick, M.A. (2013) Diverse stresses dramatically alter genome-wide p53 binding and reactivation landscape in human cancer cells. *Nucleic Acids Res.*, **41**, 7286–7301.
  49. Muller, G.A., Wintsche, A., Stangner, K., Prohaska, S.J., Stadler, P.F. and Engeland, K. (2014) The CHR site: definition and genome-wide identification of a cell cycle transcriptional element. *Nucleic Acids Res.*, **42**, 10331–10350.
  50. Lopez-Bigas, N., Kisiel, T.A., Dewaal, D.C., Holmes, K.B., Volkert, T.L., Gupta, S., Love, J., Murray, H.L., Young, R.A. and Benevolenskaya, E.V. (2008) Genome-wide analysis of the H3K4 histone demethylase RBP2 reveals a transcriptional program controlling differentiation. *Mol. Cell*, **31**, 520–530.
  51. Beshiri, M.L., Holmes, K.B., Richter, W.F., Hess, S., Islam, A.B., Yan, Q., Plante, L., Litovchick, L., Gevry, N., Lopez-Bigas, N. *et al.* (2012) Coordinated repression of cell cycle genes by KDM5A and E2F4 during differentiation. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 18499–18504.
  52. Helleday, T., Eshad, S. and Nik-Zainal, S. (2014) Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.*, **15**, 585–598.
  53. Kubo, K., Ide, H., Wallace, S.S. and Kow, Y.W. (1992) A novel, sensitive, and specific assay for abasic sites, the most commonly produced DNA lesion. *Biochemistry*, **31**, 3703–3708.
  54. Mietz, J.A., Unger, T., Huibregtse, J.M. and Howley, P.M. (1992) The transcriptional reactivation function of wild-type p53 is inhibited

- by SV40 large T-antigen and by HPV-16 E6 oncoprotein. *EMBO J.*, **11**, 5013–5020.
55. Allen-Hoffmann, B.L., Schlosser, S.J., Ivarie, C.A., Sattler, C.A., Meisner, L.F. and O'Connor, S.L. (2000) Normal growth and differentiation in a spontaneously immortalized near-diploid human keratinocyte cell line, NIKS. *J. Invest. Dermatol.*, **114**, 444–455.
  56. Helt, A.M., Funk, J.O. and Galloway, D.A. (2002) Inactivation of both the retinoblastoma tumor suppressor and p21 by the human papillomavirus type 16 E7 oncoprotein is necessary to inhibit cell cycle arrest in human epithelial cells. *J. Virol.*, **76**, 10559–10568.
  57. Shin, M.K., Balsitis, S., Brake, T. and Lambert, P.F. (2009) Human papillomavirus E7 oncoprotein overrides the tumor suppressor activity of p21Cip1 in cervical carcinogenesis. *Cancer Res.*, **69**, 5656–5663.
  58. Fischer, M., Uxa, S., Stanko, C., Magin, T.M. and Engeland, K. (2017) Human papilloma virus E7 oncoprotein abrogates the p53-p21-DREAM pathway. *Sci. Rep.*, **7**, 2603.
  59. Menzies, G.E., Reed, S.H., Brancale, A. and Lewis, P.D. (2015) Base damage, local sequence context and TP53 mutation hotspots: a molecular dynamics study of benzo[a]pyrene induced DNA distortion and mutability. *Nucleic Acids Res.*, **43**, 9133–9146.
  60. Leroy, B., Anderson, M. and Soussi, T. (2014) TP53 mutations in human cancer: database reassessment and prospects for the next decade. *Hum. Mutat.*, **35**, 672–688.
  61. Beckerman, R. and Prives, C. (2010) Transcriptional regulation by p53. *Cold Spring Harb. Perspect. Biol.*, **2**, a000935.
  62. Rinn, J.L. and Huarte, M. (2011) To repress or not to repress: this is the guardian's question. *Trends Cell Biol.*, **21**, 344–353.
  63. Sadasivam, S. and DeCaprio, J.A. (2013) The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat. Rev. Cancer*, **13**, 585–595.
  64. Chou, W.C., Chen, W.T., Hsiung, C.N., Hu, L.Y., Yu, J.C., Hsu, H.M. and Shen, C.Y. (2017) B-Myb induces APOBEC3B expression leading to somatic mutation in multiple cancers. *Sci. Rep.*, **7**, 44089.
  65. Litovchick, L., Sadasivam, S., Florens, L., Zhu, X., Swanson, S.K., Velmurugan, S., Chen, R., Washburn, M.P., Liu, X.S. and DeCaprio, J.A. (2007) Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. *Mol. Cell*, **26**, 539–551.
  66. Sdek, P., Ying, H., Chang, D.L., Qiu, W., Zheng, H., Touitou, R., Allday, M.J. and Xiao, Z.X. (2005) MDM2 promotes proteasome-dependent ubiquitin-independent degradation of retinoblastoma protein. *Mol. Cell*, **20**, 699–708.
  67. Uchida, C., Miwa, S., Kitagawa, K., Hattori, T., Isobe, T., Otani, S., Oda, T., Sugimura, H., Kamijo, T., Ookawa, K. *et al.* (2005) Enhanced Mdm2 activity inhibits pRB function via ubiquitin-dependent degradation. *EMBO J.*, **24**, 160–169.
  68. Du, W., Wu, J., Walsh, E.M., Zhang, Y., Chen, C.Y. and Xiao, Z.X. (2009) Nutlin-3 affects expression and function of retinoblastoma protein: role of retinoblastoma protein in cellular response to nutlin-3. *J. Biol. Chem.*, **284**, 26315–26321.
  69. Martin, K., Trouche, D., Hagemeier, C., Sorensen, T.S., La Thangue, N.B. and Kouzarides, T. (1995) Stimulation of E2F1/DP1 transcriptional activity by MDM2 oncoprotein. *Nature*, **375**, 691–694.
  70. Ambrosini, G., Sambol, E.B., Carvajal, D., Vassilev, L.T., Singer, S. and Schwartz, G.K. (2007) Mouse double minute antagonist Nutlin-3a enhances chemotherapy-induced apoptosis in cancer cells with mutant p53 by activating E2F1. *Oncogene*, **26**, 3473–3481.
  71. Spangle, J.M., Dreijerink, K.M., Groner, A.C., Cheng, H., Ohlson, C.E., Reyes, J., Lin, C.Y., Bradner, J., Zhao, J.J., Roberts, T.M. *et al.* (2016) PI3K/AKT signaling regulates H3K4 methylation in breast cancer. *Cell Rep.*, **15**, 2692–2704.
  72. Kanu, N., Cerone, M.A., Goh, G., Zalmas, L.P., Bartkova, J., Dietzen, M., McGranahan, N., Rogers, R., Law, E.K., Gromova, I. *et al.* (2016) DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. *Genome Biol.*, **17**, 185.
  73. Sieuwerts, A.M., Willis, S., Burns, M.B., Look, M.P., Gelder, M.E., Schlicker, A., Heideman, M.R., Jacobs, H., Wessels, L., Leyland-Jones, B. *et al.* (2014) Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Hormones Cancer*, **5**, 405–413.
  74. Law, E.K., Sieuwerts, A.M., LaPara, K., Leonard, B., Starrett, G.J., Molan, A.M., Temiz, N.A., Vogel, R.I., Meijer-van Gelder, M.E., Sweep, F.C. *et al.* (2016) The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci. Adv.*, **2**, e1601737.



# SCIENTIFIC REPORTS



Correction: Author Correction

OPEN

## The responses of cancer cells to PLK1 inhibitors reveal a novel protective role for p53 in maintaining centrosome separation

Received: 20 June 2017  
Accepted: 10 November 2017  
Published online: 23 November 2017

Linda Smith<sup>1</sup>, Raed Farzan<sup>2</sup>, Simak Ali<sup>1</sup>, Laki Buluwela<sup>2</sup>, Adrian T. Saurin<sup>1</sup> & David W. Meek<sup>1</sup>

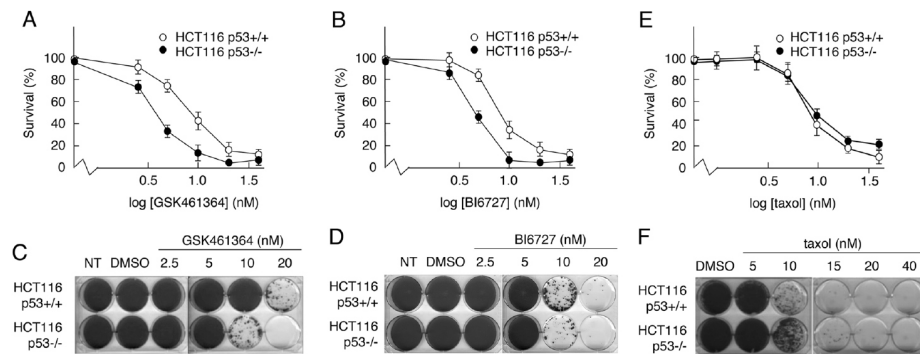
Polo-like kinase-1 (PLK1) plays a major role in driving mitotic events, including centrosome disjunction and separation, and is frequently over-expressed in human cancers. PLK1 inhibition is a promising therapeutic strategy and works by arresting cells in mitosis due to monopolar spindles. The p53 tumour suppressor protein is a short-lived transcription factor that can inhibit the growth, or stimulate the death, of developing cancer cells. Curiously, although p53 normally acts in an anti-cancer capacity, it can offer significant protection against inhibitors of PLK1, but the events underpinning this effect are not known. Here, we show that functional p53 reduces the sensitivity to PLK1 inhibitors by permitting centrosome separation to occur, allowing cells to traverse mitosis and re-enter cycle with a normal complement of 2N chromosomes. Protection entails the activation of p53 through the DNA damage-response enzymes, ATM and ATR, and requires the phosphorylation of p53 at the key regulatory site, Ser15. These data highlight a previously unrecognised link between p53, PLK1 and centrosome separation that has therapeutic implications for the use of PLK1 inhibitors in the clinic.

p53 is a short-lived transcription factor that is activated and stabilized in response to a range of cancer-relevant stress stimuli including DNA damage, hyper-proliferation, and hypoxia<sup>1-3</sup>. Activated/induced p53 orchestrates changes in gene expression leading to tumour suppressive outcomes of growth arrest (transient or permanent) or programmed cell death. Importantly, p53 also has homeostatic functions, such as control of stem cell renewal and regulation of intermediary metabolism, which may also contribute to tumour suppression<sup>3,4</sup>.

Cells experiencing impairment of the mitotic machinery can undergo apoptosis within mitosis (resulting from mitotic “catastrophe”), while others eventually escape the spindle assembly checkpoint, abort mitosis, and re-enter G1 with abnormal ploidy<sup>5</sup>. No direct role for p53 has been identified within mitosis itself. However, it is clear that p53 can respond to disruption to mitotic integrity following mitotic exit, at which point it can stimulate cell death or senescence as a means of preventing the survival of cells with chromosomal instability<sup>5,6</sup>. Cells failing to undergo normal mitotic progression accumulate *de novo* DNA damage, leading to activation of the protein kinases ATM (ataxia-telangiectasia mutated) and ATR (ATM- and Rad3-related) and, consequently, post-mitotic phosphorylation and activation of p53<sup>6-11</sup>. Cells encountering centrosomal impairment can also undergo delays in mitosis, with similar abortive outcomes<sup>12</sup>. Additionally, p53 controls the levels of Aurora A, an upstream component of the protein kinase cascades responsible for the timely disjunction and bidirectional movement of the centrosomes<sup>13,14</sup>.

PLK1 is a member of the polo-like kinase (PLK) family that mediates several key functions throughout mitosis including centrosome disjunction and movement, activation of cyclin B/CDK1, spindle assembly, and cytokinesis<sup>15,16</sup>. Consistent with these roles, inhibition of PLK1 arrests cells in early mitosis with a characteristic polo “ring” of chromosomes undergoing monopolar attachment to duplicated but unseparated centrosomes. More recently, PLK1 has also been linked to roles in DNA replication<sup>17,18</sup>. PLK1 levels are tightly regulated over the course of the cell cycle<sup>19-21</sup> and its protein kinase activity is activated through phosphorylation by Aurora A<sup>22,23</sup>.

<sup>1</sup>Division of Cancer Research, Medical Research Institute, Ninewells Hospital and Medical School, The University of Dundee, Dundee, DD1 9, SY, United Kingdom. <sup>2</sup>Department of Surgery and Cancer, Imperial College London, Hammersmith Hospital Campus, London, W12 0NN, United Kingdom. Correspondence and requests for materials should be addressed to D.W.M. (email: [d.w.meek@dundee.ac.uk](mailto:d.w.meek@dundee.ac.uk))



**Figure 1.** Cells expressing wild type p53 show reduced sensitivity to inhibitors, GSK461364 and BI6727. (A,B,E) Parental HCT116 (p53+/+) cells and a derivative line lacking p53 expression (p53-/-) were treated for 72 h with increasing concentrations of the PLK1 inhibitors, GSK461364 (A) or BI6727 (B), or with taxol (E). Cell viability was then measured using an MTS assay. (C,D,F) HCT116-p53+/+ or -p53-/- cells treated for 72 h with increasing concentrations of GSK461364 (C), BI6727 (D) or taxol (F). Following removal of drugs, cells were grown for a further 12 days. Surviving colonies were detected by staining with crystal violet. Data are representative of three independent experiments each conducted in duplicate.

PLK1 expression is down-regulated by p53 as part of the G2/M checkpoint<sup>24–26</sup> and its levels are elevated in a range of different tumour types, especially where p53 function has been lost<sup>27</sup>.

PLK1 is considered to be a highly promising cancer therapeutic target and several PLK1 inhibitors have shown promising results in clinical trials to date<sup>20,28–30</sup>. Several laboratories have reported that cancer cells lacking wild type p53 are significantly more sensitive to PLK1 inhibition as compared with cells retaining wild type p53 function<sup>26,31–35</sup>, suggesting that p53 can offer protection against PLK1 inhibitors. Importantly, this outcome has been established in a variety of cellular backgrounds<sup>32,35</sup>, and raises the possibility, from a therapeutic perspective, that cancers retaining wild type p53 may be less responsive to agents targeting PLK1. However, the mechanism(s) underpinning this apparent protective role of p53 remains unclear.

In the present study we show that, following treatment with either of two independent PLK1 inhibitors, GSK461364<sup>36</sup> and BI6727 (volasertib)<sup>37</sup>, p53-competent cells, but not p53-null cells, can survive and re-enter cell cycle with a normal complement of 2N chromosomes. Underpinning this effect, we find that the early mitotic delay induced by PLK1 inhibitors is significantly less in cells expressing wild type p53 which, unlike p53-null cells, are able to maintain the integrity of centrosome movement. These results highlight a novel p53-mediated compensatory pathway that can maintain cell integrity by overcoming impairment of mechanisms underpinning early mitosis, but which may work adversely from a therapeutic perspective.

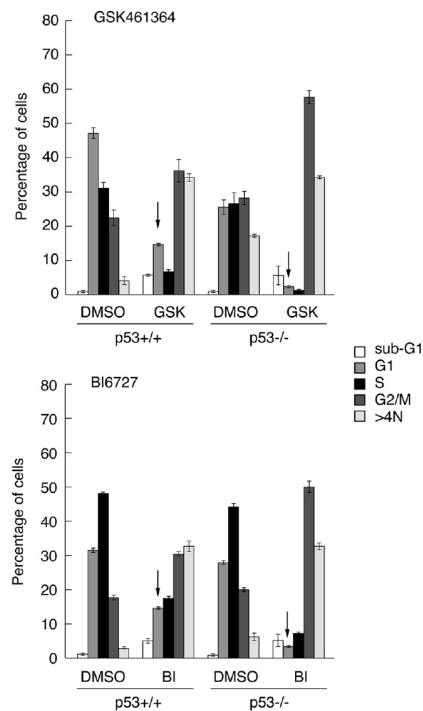
## Results

**Wild type p53 protects cells from death induced by PLK1-targeted inhibitors.** Several reports have suggested that PLK1-targeted drugs may be less effective towards cancer cells that retain p53 function<sup>26,31–35</sup>. To confirm these observations, the effects of two independently developed, commercially available PLK1 inhibitors, GSK461364 and BI6727, were measured in cell viability (MTS) assays using HCT116 cells (which express wild type p53) and a derivative line with a targeted deletion of the *TP53* gene<sup>38</sup>. The data (Fig. 1A,B) confirmed that, while both GSK461364 and BI6727 reduced the viability of cells in a dose-dependent manner, in each case cells expressing wild type p53 were significantly less sensitive to the drug treatment. These data are consistent with previous observations<sup>26,31–35</sup>. Given that the drugs were developed independently by different companies, it is unlikely that the outcomes of treatment resulted from off-target effects.

Responses were also measured using a colony-survival assay, which was conducted in a dose-dependent manner to provide a comparative indication of the minimum concentration required in each case to give rise to colonies and an assessment of the decrease in the number of colonies as the drug level increases incrementally. The data from these analyses confirmed that the cells were sensitive to GSK461364 and BI6727 in a dose-dependent manner (Fig. 1, panels C and D respectively). As with the MTS assay, cells expressing wild type p53 appeared to be less sensitive to drug-induced loss of survival, again highlighting a possible protective function for wild type p53.

To determine whether the differential sensitivity of the p53+/+ and p53-/- cells is a general feature of mitotic inhibitors, or whether it is specific to PLK1 inhibitors, cells were treated with increasing concentrations of taxol. Both MTS assays (Fig. 1E) and colony formation assays (Fig. 1F) indicated that p53 did not contribute significantly to HCT116 cell viability following treatment with taxol.

**Cells with a normal complement of 2N DNA are detectable following treatment with PLK1 inhibitors, but only when wild type p53 is present.** To determine how p53 can mediate a differential response to PLK1 inhibitors, HCT116-p53+/+ and -p53-/- cells were treated with 20 nM GSK461364 or 10 nM BI6727 for 24 h and analyzed by flow cytometry (Fig. 2). Untreated cells showed a standard cell cycle distribution



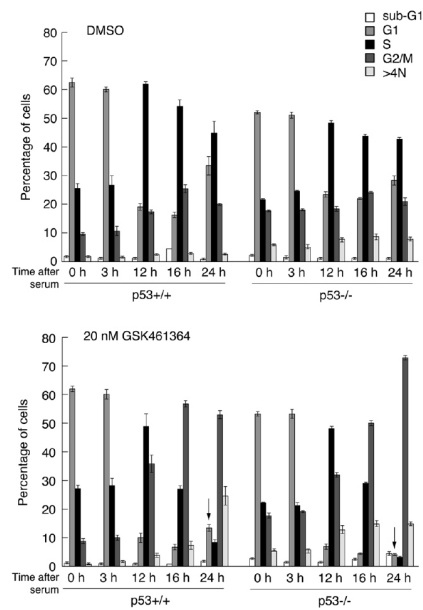
**Figure 2.** Cells expressing wild type p53 show a partial G1 arrest in response to GSK461364 or BI6727. HCT116-p53<sup>+/+</sup> or -p53<sup>-/-</sup> cells were treated for 24 h with GSK461364 (20 nM) or BI6727 (10 nM) or with DMSO as control. Cells were subsequently harvested and analyzed by flow cytometry. Arrows highlight increased number of p53<sup>+/+</sup> cells in G1 following drug treatment as compared with the p53<sup>-/-</sup> cells. The data are representative of two independent experiments, each done in triplicate. Error bars represent the standard deviation of the mean.

that was unaffected following treatment with DMSO alone. Upon treatment with GSK461364 the largest proportion of p53<sup>-/-</sup> cells was in G2/M (4N) with significant fractions at 8N (suggesting endo-reduplication) and sub-G1 (<2N, suggesting apoptosis) (Fig. 2). While these same fractions were also observed in the parental (p53<sup>+/+</sup>) cells, there was an additional population of cells at 2N representative of cells in G0/G1. Similar data were obtained when the cells had been treated with BI6727 in place of GSK461364 (Fig. 2).

The analysis was also conducted in MCF7 and U2OS cells following silencing or mock-silencing of p53. The data confirmed that the presence of cells in G1 was dependent on p53 and that it was not a cell line-specific response (Supplementary Fig. 1). (It should also be note that endo-reduplication was observed only in the HCT116 cells).

**The presence of 2N G1 cells after treatment occurs post-mitotically.** The presence of cells with 2N DNA following treatment with the PLK1 inhibitors could result from the arrest of cells in G1 phase at the time of treatment or, alternatively, from cells entering mitosis and undergoing cytokinesis. To distinguish between these two possibilities, HCT116 (p53<sup>+/+</sup> and p53<sup>-/-</sup>) cells were arrested in G0 by serum withdrawal for 24 h, then stimulated to re-enter cycle synchronously by the re-addition of serum. 1 h after serum addition the cells were treated with 20 nM GSK461364 and analyzed by flow cytometry over a 24 h period. The profiles of p53<sup>+/+</sup> and p53<sup>-/-</sup> cells were essentially indistinguishable up until 16 h post-treatment at which point the cells were almost exclusively 4N (Fig. 3). At 24 h post-treatment, however, a significant proportion of p53<sup>+/+</sup>, but not p53<sup>-/-</sup>, cells had re-entered cycle with the normal complement of 2N DNA. These data suggest that p53 permits at least a proportion of treated cells to transit mitosis, undergo cytokinesis, and re-enter G1 phase.

**PLK1 inhibitors induce phosphorylation and activation of p53 via enzymes in the DNA damage pathways.** Western blotting analysis of GSK461364- or BI6727-treated parental HCT116 cells and p53<sup>-/-</sup> derivatives indicated that the levels of PLK1 were significantly elevated at 24 h post-treatment in both lines (Supplementary Fig. 2A–C), consistent with arrest at G2/M, at which stage PLK1 levels are naturally elevated<sup>21</sup>.



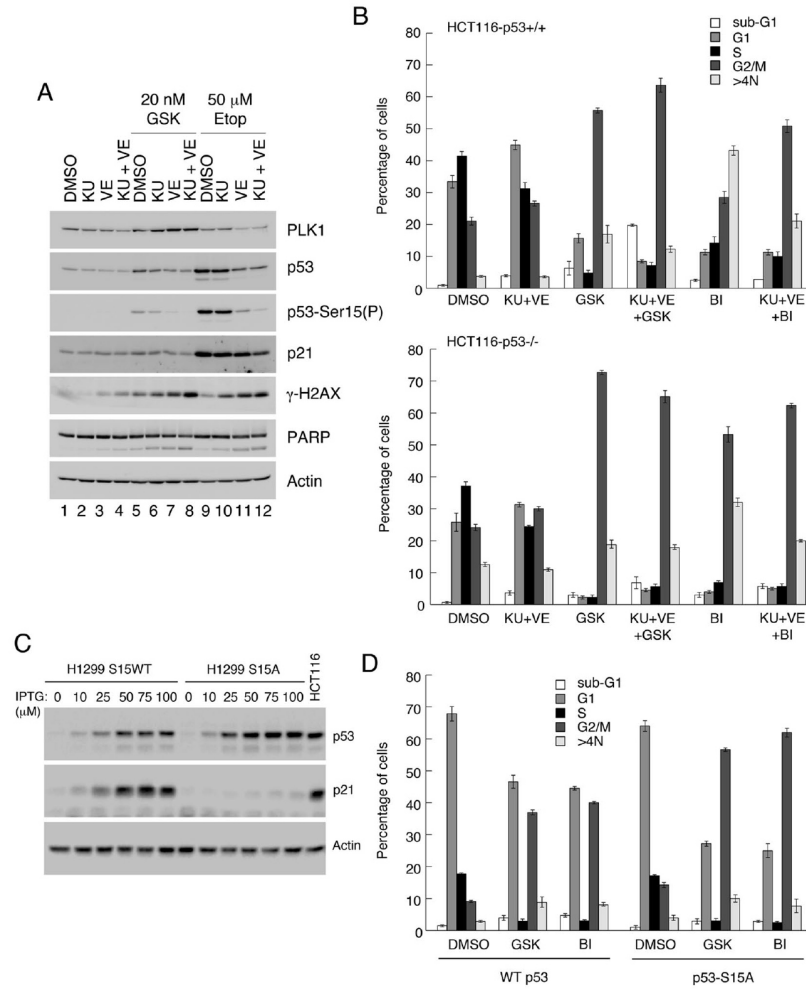
**Figure 3.** Treatment of synchronized HCT116 cells with GSK461364 permits re-entry into G1 with 2N DNA only if p53 is present. HCT116-p53+/+ or -p53-/- cells were arrested in 0.1% serum then stimulated to re-enter cell cycle by addition of 10% FBS. Treatments with DMSO (as control, upper panel) or GSK461364 (20 nM, lower panel) were initiated 1 h after serum stimulation. Cells were harvested for flow cytometry analysis at indicated times. Arrows (bottom panel) highlight the re-appearance of cells in G1 in the p53+/+ cells, but not in the p53-/- cells. The data are representative of two independent experiments, each done in triplicate. Error bars represent the standard deviation of the mean.

The coincident appearance of phosphorylated histone H3 confirmed arrest in M phase and not in G2. At 48 and 72 h PLK1 levels had declined suggesting that the cells had largely aborted mitotic arrest. The changes in PLK1 protein levels were independent of the presence of p53. Similar observations were made using MCF7 and U2OS cells (Supplementary Fig. 2D and E respectively) indicating that these effects were not cell line-dependent. Silencing of p53 in the U2OS cells confirmed that changes in PLK1 levels were independent of the presence of p53 (Supplementary Fig. 2E). Treatment of the parental HCT116, MCF7 or U2OS cells with either of the PLK1 inhibitors led to the induction of p53, p21 (a marker for p53 activity), and phosphorylated H2AX ( $\gamma$ -H2AX) (all evident by 24 h: Supplementary Fig. 2), suggesting that p53 may be induced through the DNA damage response. The late induction of these proteins fits with the well-established model that p53 responds to impairment of mitotic events after mitosis has been aborted<sup>6</sup>.

To explore whether protection against PLK1 inhibitors also required activation of p53 through DNA damage signaling, HCT116-p53+/+ cells were treated with GSK461364 for 24 h. Western blotting showed that phosphorylation of Ser15, a marker for p53 induction by DNA damage, increased over-and-above the relative increase in p53 protein levels (Fig. 4A: compare lanes 1 [mock-treated] and 5 [GSK461364-treated]). As a positive control the cellular response to etoposide was included for comparison (lane 9). To determine whether GSK461364-induced Ser15 phosphorylation was dependent on ATM and ATR, which phosphorylate p53-Ser15 in response to DNA strand breaks<sup>39</sup>, cells were pretreated either singly or in combination with inhibitors of the two kinases, KU55933 and VE821 respectively, prior to PLK1 inhibition (lanes 6–8). The data indicate that, while inhibition of either ATM or ATR impaired Ser15 phosphorylation, the combined use of both inhibitors eliminated this post-translational modification, decreased p53 induction and decreased p21 expression. (A similar effect was seen following etoposide treatment, as a positive control). These data support the idea that PLK1 inhibition activates p53 through the DNA damage pathways.

Further analysis by flow cytometry revealed that the fraction of 2N G1 cells that arose following PLK1 inhibition only in p53-competent cells, could also be eliminated through pretreatment with the ATM and ATR inhibitors (Fig. 4B). These observations support the idea that p53 induced by PLK1-mediated activation of DNA damage pathways permits cells to traverse cytokinesis and re-enter cycle with 2N DNA.

To confirm the role of DNA damage signaling pathways in activating p53 following PLK1 inhibition, the responses of clones of H1299 (p53-null) cells inducibly-expressing homeostatic levels of wild type p53 or a



**Figure 4.** Induction of p53 and appearance of post-mitotic G1 (2N) cells by PLK1 inhibitors occurs through the DNA damage response pathways. (A) HCT116-p53<sup>+/+</sup> cells were pre-treated for 1 h with 10 μM KU55933 and/or 10 μM VE821, and subsequently treated for 24 h with 20 nM GSK461364 or, as control, 50 μM etoposide. Cell extracts were analyzed by western blotting. (B) HCT116-p53<sup>+/+</sup> and -p53<sup>-/-</sup> cells were pre-treated for 1 h with 10 μM KU and 10 μM VE821 then further treated for 24 h with 20 nM GSK461364 or 10 nM BI6727, or with DMSO as control. Cells were then harvested and analyzed by flow cytometry. (C) H1299 (endogenous p53-null) cells ectopically expressing wild type p53, or a S15A substitution mutant of p53, via the LacSwitch II system (Stratagene) were treated for 16 h with increasing levels of the inducer, IPTG. Cell extracts were analyzed by western blotting as indicated. (D) H1299-WTp53 or H1299-S15A-p53 cells were treated for 16 h with 100 μM IPTG, followed by treatment for 24 h with 20 nM GSK461364 or 10 nM BI6727, or DMSO as control. Cells were harvested and analyzed by flow cytometry. Panels A and C show cropped western blots; full length gels including molecular weight markers are provided in the Supplementary Information. The data in panels B and D are each representative of two independent experiments, each done in triplicate. Error bars represent the standard deviation of the mean.

p53-S15A substitution mutant were examined<sup>40,41</sup>. Loss of the Ser15 phosphorylation site, which is required to induce p53-target genes<sup>40</sup> (Fig. 4C), led to a significant reduction in the appearance of cells with 2N DNA following impairment of PLK1 activity (Fig. 4D).

Activation of p53 following treatment with PLK1 inhibitors could occur as: (i) a direct result of DNA damage caused by inhibition of PLK1<sup>18,42</sup>; (ii) an off-target effect of the drugs; or (iii) could simply reflect cell death induced by these compounds. Silencing *PLK1* expression confirmed that loss of PLK1 gave rise to a  $\gamma$ -H2AX signal (Supplementary Fig. 3), indicating that the DNA damage was unlikely to be an off-target effect. Moreover, pre-treatment with the pan-caspase inhibitor, Z-VAD-FMK, for 1 h prior to addition of PLK1 inhibitors, effectively blocked apoptosis but did not impair the DNA damage signal induced by the PLK1 inhibitors (Supplementary Fig. 4A,B). Taken together, these data strongly suggest that the occurrence of DNA damage is a direct effect of PLK1 inhibition and is responsible for activating p53.

**p53 reduces mitotic delay following treatment with PLK1 inhibitors, but not taxol.** To explore further how p53 might permit PLK1-inhibited cells to re-enter cycle with 2N DNA the abilities of p53-competent and -null cells to traverse mitosis following PLK1 inhibition was compared by time-lapse microscopy. As a control, the responses to mitotic arrest with taxol were assessed. Treatment with increasing concentrations of GSK461364 led to progressively longer delays of the p53-competent cells in traversing mitosis (Fig. 5A). While qualitatively similar, the drug was significantly more potent in delaying mitotic progression of p53-null cells. These data suggest that p53 can mitigate mitotic arrest arising from PLK1 inhibition. When cells were treated with taxol in place of GSK461364, the dose-dependent delay of p53<sup>+/+</sup> and p53<sup>-/-</sup> cells in traversing mitosis was essentially indistinguishable (Fig. 5B), indicating that the p53 response is selective to PLK1 inhibition. When MCF7 cells, and a derivative line in which p53 had been deleted using CRISPR, were treated with the PLK1 inhibitors and analyzed in a manner similar to HCT116 cells, cells lacking p53 were again delayed significantly longer in mitosis than parental cells, confirming that the response was not cell line-dependent (Supplementary Fig. 5).

To determine whether the mitotic delay caused by PLK1 inhibition was associated with DNA damage pathways, the time-lapse analysis with the HCT116 cells was repeated in the presence of the ATM and ATR inhibitors (KU55933 and VE821 respectively). These drugs significantly extended mitotic delay in p53-competent cells but not in p53-null cells (Fig. 5C), suggesting that the ability to limit PLK1 inhibitor-dependent mitotic delay is dependent upon activation of p53 by DNA damage pathways.

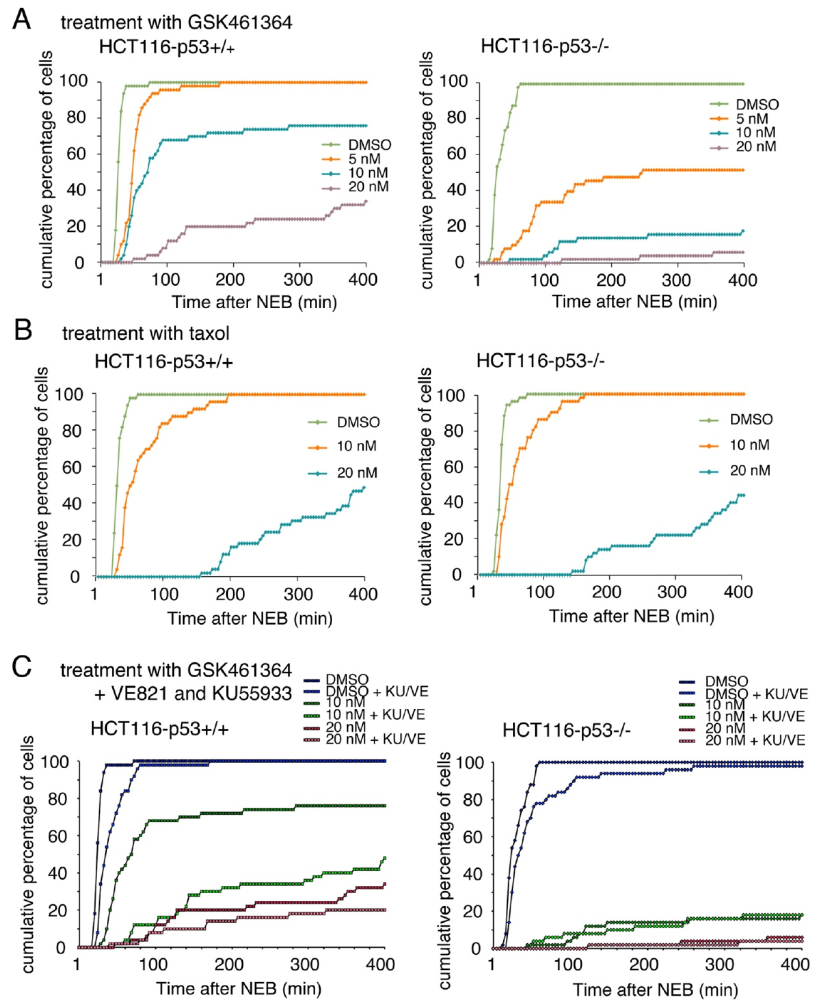
**p53 protects cells from PLK1 inhibition by maintaining centrosome separation.** To investigate how p53 can overcome arrest induced by PLK1 inhibitors, further time lapse analysis was conducted using SiR-DNA (SiR-Hoechst\*) to visualize the chromosomes (Supplementary Movies 1–4). Representative images from these analyses are shown in Fig. 6A. These data show that all untreated cells underwent a normal mitosis and cytokinesis, irrespective of p53 status. Following treatment with 20 nM GSK461364, however, 67% of p53-competent cells entering mitosis could still achieve cytokinesis as compared with only 22% of p53-null cells, most of which failed to achieve metaphase (Supplementary Movies 3 and 4 respectively; Fig. 6A). Moreover, quantitative analysis (Fig. 6B) showed that, while the median amount of time spent by untreated p53<sup>+/+</sup> and p53<sup>-/-</sup> cells in mitosis was indistinguishable, p53<sup>-/-</sup> cells showed a significantly greater delay in mitosis (median time 280 min) than p53<sup>+/+</sup> cells (median time 100 min) after GSK461364 treatment. Phospho-histone staining (Fig. 6C) confirmed that an increasingly larger proportion of p53<sup>-/-</sup> cells were in mitosis as compared with p53<sup>+/+</sup> parental cells, and that this was not the consequence of a differential rate at which cells enter mitosis (Fig. 6D), thereby ruling out a role for p53 in restricting mitotic entry. Taken together, these data suggest that p53 may respond to, and/or protect against, failure of a PLK1-mediated event(s) in early mitosis prior to the formation of a mitotic spindle.

PLK1 is a major upstream component of the protein kinase cascades that trigger uncoupling of the duplicated centrosomes and activation of Eg5, the principal kinesin motor that drives centrosomes to opposite poles in the cell<sup>48</sup> (Fig. 7A). To determine whether p53 has any impact on the impairment of these processes following PLK1 inhibition, HCT116-p53<sup>+/+</sup> and -p53<sup>-/-</sup> cells were treated with GSK461364 followed by fixing and staining with an antibody to gamma-tubulin to visualize the centrosomes. Individual cells were then examined by microscopy to determine whether centrosomal segregation had occurred. Untreated p53-competent and p53-null cells were indistinguishable, showing very little segregation failure. Following GSK461364 treatment, however, only 16% of the p53<sup>+/+</sup> cells failed to segregate their centrosomes as compared with greater than 60% of the p53<sup>-/-</sup> cells at the same level of drug exposure (Fig. 7B and Supplementary Fig. 6). Taken together, these data suggest that p53 may protect cells by ensuring fidelity of centrosome disjunction and/or bidirectional movement.

**p53 safeguards the fidelity of centrosome migration.** To determine how p53 might impact on the processes of centrosome separation, HCT116-p53<sup>+/+</sup> and -p53<sup>-/-</sup> cells were treated with increasing concentrations of the Eg5 inhibitor, S-Trityl-L-cysteine (STLC). Measurement of the time spent in mitosis showed that the p53-null cells were significantly more sensitive to the Eg5 inhibitor as compared with the p53-competent cells (Fig. 7C), thereby phenocopying the response to PLK1 inhibitors. Similarly, in an MTS (viability) assay, p53-null cells were significantly more sensitive to STLC than p53-competent cells (Fig. 7D). In contrast, as shown above, the sensitivity of the two lines to taxol was identical in each assay (Figs 1E,F and 5B respectively). These data strongly suggest that p53 can protect against impairment of centrosome separation, but has no effect once spindle formation begins. Moreover, given that separation of the centrosomes must occur *after* their disjunction, it is highly likely that p53 has little impact on the disjunction mechanism itself.

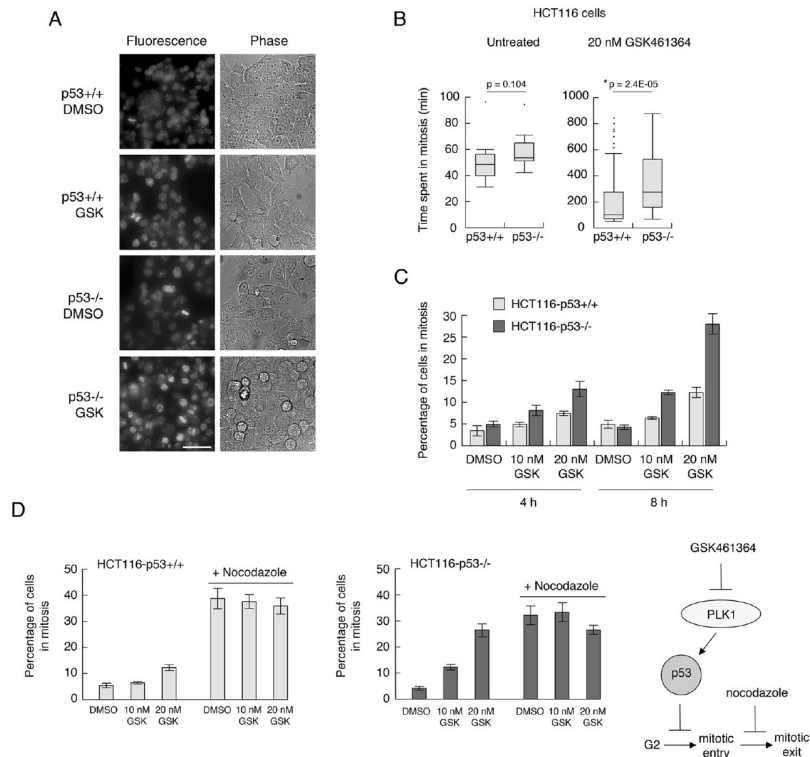
## Discussion

In the present study, we confirm that p53 can offer a selective advantage to cancer cells treated with PLK1 inhibitors. Our analysis indicates that, in response to the PLK1 inhibitors, GSK461364 or BI6727, a proportion of p53-competent, but not p53-null, cells can traverse mitosis and cytokinesis, in a p53- and ATM/ATR-dependent manner, and re-enter cycle with a normal complement of chromosomes.



**Figure 5.** p53 reduces the delay in mitosis resulting from inhibition of PLK1. HCT116-p53+/+ (left panels) and -p53-/- cells (right panels) were treated with DMSO, 5, 10 or 20 nM GSK461364 (A) or DMSO, 10 or 20 nM Taxol (B). Time-lapse microscopy analysis was then used to determine the duration of mitosis. (C) HCT116-p53+/+ and -p53-/- cells were pretreated for 1 h with DMSO or 10  $\mu$ M KU-55933 and 10  $\mu$ M VE-82, then subsequently treated with DMSO, 10 or 20 nM GSK461364. Time-lapse analysis was used to determine the duration of mitosis. Graphs show cumulative data from 50 cells for each treatment and are representative of three replicates.

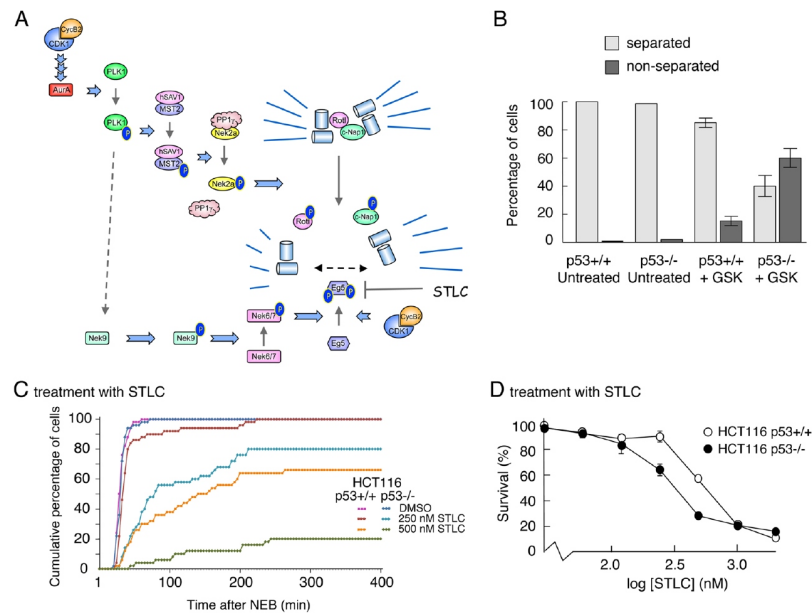
Our data are consistent with the findings of several laboratories, showing that cancer cells lacking wild type p53 show increased sensitivity to PLK1 inhibition as compared with cells retaining wild type p53 function<sup>26,31–35</sup>. Contrary to these studies, Sanhaji *et al.* have reported that p53 is not directly relevant to the outcome of treating cells with PLK1 inhibitors<sup>44,45</sup>. Although our conclusions regarding biological outcome differ from Sanhaji and colleagues, we do share a number of observations that highlight important differences in the response to PLK1 inhibition arising from the presence or absence of p53. These include: (i) cells lacking p53 show an increased proportion of mono-polar spindles following treatment with BI6727 (and BI2536) as compared with the p53-competent cells<sup>44</sup>, although quantitatively we see a much bigger effect (Fig. 7B, Supplementary Fig. 6); (ii) there is an increased proportion of cells in G1 after PLK1 inhibition when p53 is present<sup>44</sup> (Figs 2–4); and (iii)



**Figure 6.** Cells lacking p53 are delayed in mitosis following treatment with GSK461364 (A and B) HCT116-p53<sup>+/+</sup> and -p53<sup>-/-</sup> cells were stained with SiR-DNA (SiR-Hoechst\*) and treated with DMSO or 20 nM GSK461364. Cells were mounted on the Deltavision Elite live cell microscope and imaged using the Cy5 (far red) channel and in phase contrast every 4 minutes over 250 time points. (A) representative images of the final time points (16.6 h post-treatment). The scale bar (shown in the bottom left hand image) represents 50 microns. (B) left hand plot: the lengths of time spent in mitosis for 22 HCT116-p53<sup>+/+</sup> and 22 HCT116-p53<sup>-/-</sup> cells are presented as box plots (showing median and quartiles). The data from 77 HCT116-p53<sup>+/+</sup> and 77 HCT116-p53<sup>-/-</sup> GSK461364-treated cells is shown in the right hand plot. (C) HCT116-p53<sup>+/+</sup> and -p53<sup>-/-</sup> cells were treated with DMSO, 10 nM or 20 nM GSK461364 for 4 and 8 h. Cells were fixed with paraformaldehyde, stained with phosphorylated histone H3 (serine 10) antibody and propidium iodide. Cells were analyzed by flow cytometry, and the percentage of cells positive for phosphorylated histone H3 (serine 10) was plotted as a bar graph. (D) HCT116-p53<sup>+/+</sup> and -p53<sup>-/-</sup> cells were treated with DMSO, 10 nM GSK461364, or 20 nM GSK461364, in the absence of presence of 3.3  $\mu$ M Nocodazole. Cells were subsequently fixed and stained as in panel C. Schematic: by blocking mitotic exit the number of cells in mitosis is a measure of rate of entry into mitosis. The data in panels C and D are representative of two independent experiments. Error bars represent the standard deviation of the mean.

there is more DNA damage (as measured by the appearance of  $\gamma$ -H2AX) in the PLK1 inhibitor-treated p53<sup>-/-</sup> cells as compared with the p53<sup>+/+</sup> cells; in our own case we provide evidence this is NOT a direct result of apoptosis [Sup Fig. 4]). Clearly, therefore, while these similarities are evident, it is difficult to be certain regarding the basis on which we reach different conclusions concerning the effects of p53 on cellular fate following treatment with PLK1 inhibitors. One important consideration is perhaps the fact that antimitotic inhibitors can lead stochastically to multiple outcomes within the same population of cells<sup>46</sup>. In the case of PLK1 inhibitors, such a heterogeneous response can be observed from cell cycle profiles (e.g. see Fig. 2) where many cells arrest in M phase, while some undergo apoptosis, some endo-reduplicate (i.e. abort mitosis and re-enter cycle without undergoing cytokinesis leading either to permanent arrest in a G1-like state but with 4N DNA or being able to complete another cycle) and, when p53 is present, some undergo cytokinesis and re-enter cycle with 2N DNA. This heterogeneity is perhaps not surprising given that PLK1 is involved in several key events throughout mitosis. We suggest that such a heterogeneous response may impact upon the interpretation of data. Thus, while largely similar





**Figure 7.** p53 protects cells from inhibition of PLK1 by maintaining centrosome separation. (A) Schematic showing the protein kinase cascade responsible for regulating centrosome uncoupling and separation. (B) HCT116-p53<sup>+/+</sup> and -p53<sup>-/-</sup> cells were treated with 20 nM GSK461364 or DMSO for 8 h, and subsequently stained for  $\gamma$ -tubulin antibody (a centrosome marker). The number of mitotic cells showing a normal bipolar spindle (separated centrosomes) or monopolar spindle (non-separated centrosomes) was determined using fluorescence microscopy. Over 100 cells were counted for each condition (scored by two independent individuals) and plotted as percentages of separated v non-separated centrosomes (bar graph). (Counting was based on the detection of mitotic cells by use of DAPI, and then switching channels to look at gamma tubulin to assess separation of centrosomes in the mitotic cells). Data are representative of two independent experiments and error bars represent the standard deviation of the mean. (C) HCT116-p53<sup>+/+</sup> and -p53<sup>-/-</sup> cells were treated with DMSO, 250 or 500 nM STLC. Time-lapse analysis was then used to determine the duration of mitosis. The graph represents the cumulative data from 50 cells for each treatment and are representative of three replicates. (D) HCT116 p53<sup>+/+</sup> and p53<sup>-/-</sup> cells were treated for 72 h with increasing concentrations of STLC. Cell viability was measured using an MTS assay. Data are representative of three independent experiments each conducted in triplicate. Error bars represent the standard deviation of the mean.

(heterogeneous) outcomes may be observed following drug treatment regardless of the presence of p53, we suggest that protection conferred by p53 may allow a small proportion of cells to survive and it is these that underpin the observed differences in cellular fate between the p53<sup>+/+</sup> and p53<sup>-/-</sup> cells. It is thus entirely plausible that differences between various studies in growth conditions, drug treatment regimes (e.g. concentrations, exposure times etc.) or the types of assay used may either mask or reveal the behaviour of this surviving proportion of cells.

In addition to its well characterized roles in responding to failed mitoses<sup>5,6</sup>, p53 has an earlier role upon entering mitosis in maintaining a balance of the activities in the protein kinase cascade initiated by cyclin B2/CDK1 (Fig. 7A) that is responsible for the timely disjunction and separation of the centrosomes, through regulating the levels of Aurora A<sup>13,14</sup>. We propose here an *additional* involvement for p53 in this pathway in which p53 senses and responds to failure to drive centrosomes to opposite poles of the cell, as indicated by our finding that p53 is required to overcome defects in Eg5 motor activity but has no impact on the subsequent steps of microtubule dynamics. Given that transcription essentially shuts down as cells enter mitosis, this would suggest that either p53 engages in some novel, as yet undefined, transcription-independent function in prometaphase or, more plausibly, that in the presence of p53, a compensatory pathway(s) is already in place before cells enter mitosis, that allows at least a proportion of cells to passage through mitosis and undergo cytokinesis. The observation that p53-competent cells enter and traverse mitosis within a very short time after drug treatment (Supplementary Movie 3), suggests that the p53 “response” to PLK1 inhibitors is relatively fast, and therefore favors the idea that a safety net is already available to ensure proper centrosome dynamics.

The finding that direct inhibition of Eg5 phenocopies the outcomes of treating p53-competent and p53-null cells with PLK1 inhibitors (Fig. 7) suggests that p53 senses impairment of Eg5 activity or some downstream event

prior to establishing the spindle. Examination of the levels of Eg5 and Kif15 (motor proteins that are responsible for driving the uncoupled centrosomes to opposite poles of the cell<sup>47–50</sup>) in p53+/+ and p53–/– cells reveals no differences in the homeostatic levels of these proteins (Supplementary Fig. 7), or following stimulation with Nutlin (Supplementary Fig. 8), etoposide (Supplementary Fig. 9), PLK1 inhibitors (Supplementary Fig. 10A–D), or the Eg5 inhibitor, STLC (Supplementary Fig. 10E,F). However, it is possible that p53 may be able to compensate by modulating the levels of Eg5 regulators or partners. While the nature of this compensatory mechanism(s) remains unclear it should be a key focus in developing our understanding of the impact of p53 on maintaining mitotic integrity.

Our data also confirmed previous observations that inhibition of PLK1 can give rise to DNA damage<sup>18,42</sup> and suggest the possibility that this may be the activating event for p53-mediated protection. Alternatively, however, it is plausible that protection may be provided through basal, unstimulated levels of p53, given that p53-Ser15 phosphorylation, mediated in part by ATM/ATR, can occur homeostatically in the absence of DNA damage<sup>40</sup>.

Finally, our data raise the possibility that cancers retaining wild type p53 may be less responsive clinically to agents targeting PLK1. Given that, from a therapeutic perspective, this could ultimately impact on treatment, it will be important to establish an understanding of the relationship between *TP53* status and the sensitivity of human cancers to PLK1-directed drugs.

## Materials and Methods

**Cell culture.** Cells were routinely maintained at 37 °C/5% CO<sub>2</sub> in Dulbecco's modified Eagle's medium (Gibco) supplemented with 10% foetal bovine serum (Biosera) and 2 mM Glutamine (Gibco). MCF7 cells were obtained from ATCC (LGC Standards, UK). HCT116 (p53+/+ and p53–/–) colon carcinoma cells, U2OS osteosarcoma cells, H1299 non-small cell lung carcinoma cells were obtained from the Cancer Research UK Biorepository. Cells were expanded, aliquoted and stored at early passage after purchase/receipt, and were validated for p53 status. Cells were routinely checked for mycoplasma (Mycopalert™, Lonza) and were discarded after passage 16 from thawing.

**Antibodies and reagent.** Antibodies were as follows: p53 (DO-1; Moravian Biotechnology), p21 (H-164; sc-756; SantaCruz Biotechnology), PLK1 (208G4; Cell Signaling Technology), actin (A2066; Sigma), Poly (ADP-Ribose) Polymerase-1 (PARP) (9542; Cell Signaling Technology),  $\gamma$ -H2AX (Phospho-S139) (ab11174; AbCam), p53 (Phospho-S15) (#9284; Cell Signaling Technology), MDM2 (4B2; Moravian Biotechnology), Gamma-tubulin (T65557; Sigma), Histone H3 ((D1H2): 4499; Cell Signaling Technology), Histone H3 (Phospho-S10) (06–570; Millipore), BrdU Pure ((B44): 347580; Becton Dickinson).

PLK1 inhibitors, GSK461364 and BI6727, ATM and ATR inhibitors, KU-55933 and VE-821, and etoposide were from Selleckchem. The specificity and efficacy of GSK461364 and BI6727 have been described previously<sup>36,37</sup>. S-Trityl-L-cysteine (STLC) was from Sigma. Taxol was from LC Labs. Nocodazole was from Millipore.

**Western blot analysis.** SDS-PAGE and western blotting was carried out using standard conditions. Images were gathered using a BioRad ChemiDoc™ MP Imaging System using Image Lab Software 4.1. Minor adjustments only were made for background and contrast in Photoshop.

**MTS cell viability assay.**  $2.5 \times 10^3$  cells/well were seeded in triplicate in 96-well plates for each condition. The following day cells were left untreated, treated with DMSO (vehicle control) or with drugs as indicated in figure legends. After 68 hours of exposure to the drug, MTS assay (Promega) was performed as directed by the manufacturer.

**Colony formation assay.**  $5 \times 10^4$  cells were seeded in 6-well plates. The following day cells were either left untreated (fresh medium added), treated with DMSO (vehicle control), or with drugs as indicated in figure legends.

**Flow cytometry.** Cells were fixed in ice-cold ethanol, washed in PBS-1% FBS and treated with 2 M HCl for 20 min at 37 °C, and subsequently with 50  $\mu$ g/ml Propidium Iodide/200  $\mu$ g/ml RNase A in PBS. Analysis was performed using a Becton Dickinson FACS Canto Flow Cytometer and FlowJo software.

To discriminate between cells in G2 and M, cells were fixed in 0.5% (w/v) para-formaldehyde for 20 min at 37 °C and incubated with 100  $\mu$ l of 10  $\mu$ g/ml anti-phospho-(S10)-histone H3 antibody for 1 h, followed by washing in PBS-1% FBS and incubation with 12.5  $\mu$ g/ml FITC-labeled goat anti-rabbit IgG for 30 min. Cells were subsequently washed, stained with propidium iodide and analyzed as above.

**Gene silencing.** PLK1 and p53 siRNA transfections were performed in 6-well plates using  $2.5 \times 10^5$  cells and 10 nM siRNA per well, and reverse transfection procedure with Lipofectamine® RNAiMAX (Invitrogen) according to manufacturer's instructions. Cells were incubated for 24–48 h before harvesting or further treatment.

siRNA oligonucleotides were from Thermo Scientific: p53 (exon 7): 5'-GACUCCAGUGUAAUCUACUU-3', (OSLR-001137) PLK1:5'-GCACAUACCGCCUGAGUCU-3'; 5'-CCACCAAGGUUUUCGAUUG-3'; 5'-GCUCUCAAUGACUCAACA-3'; 5'-UCUCAAGGCCUCCUAAUAG-3', (L-003290-00) Non-silencing: 5'-CAGUCGCGUUUGCGACUGGU-3', (OSLR-001139).

**Live cell imaging.** Cells were seeded in 24-well plates to achieve 80% confluency at the time of treatment. After drug treatment imaging was performed in a heated chamber (37 °C, 5% CO<sub>2</sub>) using a 10X/03 objective on a Zeiss Axiovert 200 M microscope controlled by Micro-Manager software. Images were taken every 4 min (20 seconds exposure, 2  $\times$  2 binning) using a C4742-80-12AG camera (Hamamatsu), with 250 images being recorded for each position. Image J software (National Institutes of Health) was used to manually calculate time in mitosis, with start time being the rounding of the cell and end time being the decision of cellular fate (i.e. aborting mitosis or undergoing cytokinesis). For each condition 50 cells were analyzed.

**Live cell imaging (fluorescence).** Cells were seeded in  $\mu$ -slide 8-well chambers (Ibidi) and left overnight. The following day, the medium was removed and replaced with carbon dioxide-independent Leibovitz's L-15 medium (Gibco) supplemented with 10% (v/v) FBS and 100 units/ml of penicillin and 100  $\mu$ g/ml of streptomycin (Gibco). After 4 h, the medium was removed and 100 nM SiR-DNA (tebu-bio), diluted in L-15, was added to the cells for 15 min. Cells were then washed with PBS before addition of L-15 containing the appropriate drugs and mounting on the incubator chamber of a Deltavision Elite microscope fitted with a 40x/1.30NA U Plan FLN oil objective lens, and operated using SoftWoRx software. Images were captured using a CoolSNAP HQ2 camera (Photometrics) (250 time points, 4 min intervals,  $4 \times 4$  binning,  $256 \times 256$  image size, 4 optical sections (5  $\mu$ m optical spacing)).

**Immunofluorescence.** Cells were seeded on poly-L-lysine coated coverslips in 24-well plates 24 h prior to drug treatments. Cells were subsequently washed in PBS then fixed with 4% (w/v) para-formaldehyde solution for 15 min. After further PBS washes, cells were permeabilized in 0.1% (v/v) Triton X-100 in PBS for five min. After further PBS washes coverslips were soaked in 5% (w/v) BSA, 0.1% (v/v) Triton X-100 in TBS (blocking solution) for 15 min. Coverslips were then incubated in  $\gamma$ -tubulin antibody diluted in blocking solution for 1 h, followed by washing in PBS, incubation in Alexa Fluor<sup>®</sup> secondary antibody for 1 h and further washing in PBS. Coverslips were then incubated with DAPI for five min, washed in PBS, and mounted onto slides using ProLong<sup>®</sup> Gold Antifade Mountant (Life Technologies). Slides were left for 24 h at room temperature and subsequently stored at 4°C. Samples were examined using a Deltavision Elite microscope fitted with a 40x/1.30NA U Plan FLN or 100X/1.4, UPLS Apo oil objective lens (Olympus), and operated using SoftWoRx software. To measure centrosome separation, mitotic cells were detected using DAPI. The  $\gamma$ -tubulin staining was then used to determine visually whether the centrosomes has separated (i.e formed a spindle) or remained in tight proximity at the periphery of the cell. All of these measurements were made independently by two individuals. Representative images were acquired using a Leica DFC420 camera.

## References

- Bieging, K. T., Mello, S. S. & Attardi, L. D. Unravelling mechanisms of p53-mediated tumour suppression. *Nat. Rev. Cancer* **14**, 359–370 (2014).
- Vousden, K. H. & Lane, D. P. p53 in health and disease. *Nat. Rev. Mol. Cell Biol.* **8**, 275–283 (2007).
- Vousden, K. H. & Prives, C. Blinded by the Light: The Growing Complexity of p53. *Cell* **137**, 413–431 (2009).
- Li, T. *et al.* Tumor suppression in the absence of p53-mediated cell-cycle arrest, apoptosis, and senescence. *Cell* **149**, 1269–1283 (2012).
- Vitale, I., Galluzzi, L., Castedo, M. & Kroemer, G. Mitotic catastrophe: a mechanism for avoiding genomic instability. *Nat. Rev. Mol. Cell Biol.* **12**, 385–392 (2011).
- Hayashi, M. T. & Karlseder, J. DNA damage associated with mitosis and cytokinesis failure. *Oncogene* **32**, 4593–4601 (2013).
- Dalton, W. B., Yu, B. & Yang, V. W. p53 suppresses structural chromosome instability after mitotic arrest in human cells. *Oncogene* **29**, 1929–1940 (2010).
- Hayashi, M. T., Cesare, A. J., Fitzpatrick, J. A., Lazzarini-Denchi, E. & Karlseder, J. A telomere-dependent DNA damage checkpoint induced by prolonged mitotic arrest. *Nat. Struct. Mol. Biol.* **19**, 387–394 (2012).
- Janssen, A., van der Burg, M., Szuhai, K., Kops, G. J. & Medema, R. H. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science* **333**, 1895–1898 (2011).
- Quignon, F. *et al.* Sustained mitotic block elicits DNA breaks: one-step alteration of ploidy and chromosome integrity in mammalian cells. *Oncogene* **26**, 165–172 (2007).
- Wong, C. & Stearns, T. Mammalian cells lack checkpoints for tetraploidy, aberrant centrosome number, and cytokinesis failure. *BMC Cell Biol.* **6**, 6 (2005).
- Vitre, B. D. & Cleveland, D. W. Centrosomes, chromosome instability (CIN) and aneuploidy. *Curr. Op. Cell Biol.* **24**, 809–815 (2012).
- Nam, H. J. & van Deursen, J. M. Cyclin B2 and p53 control proper timing of centrosome separation. *Nat. Cell Biol.* **16**, 538–549 (2014).
- de Carcer, G. & Malumbres, M. A centrosomal route for cancer genome instability. *Nat. Cell Biol.* **16**, 504–506 (2014).
- Barr, F. A., Sillje, H. H. & Nigg, E. A. Polo-like kinases and the orchestration of cell division. *Nat. Rev. Mol. Cell Biol.* **5**, 429–440 (2004).
- de Carcer, G., Manning, G. & Malumbres, M. From Plk1 to Plk5: functional evolution of polo-like kinases. *Cell cycle* **10**, 2255–2262 (2011).
- Shen, M. *et al.* Centrosomal protein FOR20 is essential for S-phase progression by recruiting Plk1 to centrosomes. *Cell Res.* **23**, 1284–1295 (2013).
- Yim, H. & Erikson, R. L. Polo-like kinase 1 depletion induces DNA damage in early S prior to caspase activation. *Mol. Cell Biol.* **29**, 2609–2621 (2009).
- Archambault, V. & Glover, D. M. Polo-like kinases: conservation and divergence in their functions and regulation. *Nat. Rev. Mol. Cell Biol.* **10**, 265–275 (2009).
- Strebhardt, K. Multifaceted polo-like kinases: drug targets and antitargets for cancer therapy. *Nat. Rev. Drug Discov.* **9**, 643–660 (2010).
- Strebhardt, K. & Ullrich, A. Targeting polo-like kinase 1 for cancer therapy. *Nat. Rev. Cancer* **6**, 321–330 (2006).
- Seki, A., Coppinger, J. A., Jang, C. Y., Yates, J. R. & Fang, G. Bora and the kinase Aurora a cooperatively activate the kinase Plk1 and control mitotic entry. *Science* **320**, 1655–1658 (2008).
- Macurek, L. *et al.* Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery. *Nature* **455**, 119–123 (2008).
- Fischer, M., Quaa, M., Nickel, A. & Engeland, K. Indirect p53-dependent transcriptional repression of Survivin, CDC25C, and PLK1 genes requires the cyclin-dependent kinase inhibitor p21/CDKN1A and CDE/CHR promoter sites binding the DREAM complex. *Oncotarget* **6**, 41402–41417 (2015).
- Fischer, M., Quaa, M., Steiner, L. & Engeland, K. The p53-p21-DREAM-CDE/CHR pathway regulates G2/M cell cycle genes. *Nucleic Acids Res.* **44**, 164–174 (2016).
- McKenzie, L. *et al.* p53-dependent repression of polo-like kinase-1 (PLK1). *Cell cycle* **9**, 4200–4212 (2010).
- King, S. L. *et al.* Immunohistochemical detection of Polo-like kinase-1 (PLK1) in primary breast cancer is associated with TP53 mutation and poor clinical outcome. *Breast Cancer Res.* **14**, R40 (2012).
- Brandwein, J. M. Targeting polo-like kinase 1 in acute myeloid leukemia. *Ther. Adv. Hematol.* **6**, 80–87 (2015).
- Lee, K. S., Burke, T. R. Jr., Park, J. E., Bang, J. K. & Lee, E. Recent Advances and New Strategies in Targeting Plk1 for Anticancer Therapy. *Trends Pharmacol. Sci.* **36**, 858–877 (2015).

30. Medema, R. H., Lin, C. C. & Yang, J. C. Polo-like kinase 1 inhibitors and their potential role in anticancer therapy, with a focus on NSCLC. *Clin. Cancer Res.* **17**, 6459–6466 (2011).
31. Danovi, D. *et al.* A high-content small molecule screen identifies sensitivity of glioblastoma stem cells to inhibition of polo-like kinase 1. *PLoS One* **8**, e77053 (2013).
32. Degenhardt, Y. *et al.* Sensitivity of cancer cells to Plk1 inhibitor GSK461364A is associated with loss of p53 function and chromosome instability. *Mol. Cancer Therap.* **9**, 2079–2089 (2010).
33. Guan, R. *et al.* Small interfering RNA-mediated Polo-like kinase 1 depletion preferentially reduces the survival of p53-defective, oncogenic transformed cells and inhibits tumor growth in animals. *Cancer Res.* **65**, 2698–2704 (2005).
34. Liu, X., Lei, M. & Erikson, R. L. Normal cells, but not cancer cells, survive severe Plk1 depletion. *Mol. Cell. Biol.* **26**, 2093–2108 (2006).
35. Sur, S. *et al.* A panel of isogenic human cancer cells suggests a therapeutic approach for cancers with inactivated p53. *Proc. Natl. Acad. Sci. USA* **106**, 3964–3969 (2009).
36. Gilmartin, A. G. *et al.* Distinct concentration-dependent effects of the polo-like kinase 1-specific inhibitor GSK461364A, including differential effect on apoptosis. *Cancer Res.* **69**, 6969–6977 (2009).
37. Steegmaier, M. *et al.* BI 2536, a potent and selective inhibitor of polo-like kinase 1, inhibits tumor growth *in vivo*. *Curr. Biol.* **17**, 316–322 (2007).
38. Bunz, F. *et al.* Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science* **282**, 1497–1501 (1998).
39. Appella, E. & Anderson, C. W. Post-translational modifications and activation of p53 by genotoxic stresses. *Eur. J. Biochem.* **268**, 2764–2772 (2001).
40. Loughery, J., Cox, M., Smith, L. M. & Meek, D. W. Critical role for p53-serine 15 phosphorylation in stimulating transactivation at p53-responsive promoters. *Nucleic Acids Res.* **42**, 7666–7680 (2014).
41. Rocha, S., Martin, A. M., Meek, D. W. & Perkins, N. D. p53 represses cyclin D1 transcription through down-regulation of Bcl-3 and inducing increased association of the p52 NFκB subunit with histone deacetylase 1. *Mol. Cell. Biol.* **23**, 4713–4727 (2003).
42. Driscoll, D. L. *et al.* Plk1 inhibition causes post-mitotic DNA damage and senescence in a range of human tumor cell lines. *PLoS One* **9**, e111060 (2014).
43. Bruinsma, W., Raaijmakers, J. A. & Medema, R. H. Switching Polo-like kinase-1 on and off in time and space. *Trends Biochem. Sci.* **37**, 534–542 (2012).
44. Sanhaji, M. *et al.* p53 is not directly relevant to the response of Polo-like kinase 1 inhibitors. *Cell cycle* **11**, 543–553 (2012).
45. Sanhaji, M. *et al.* Polo-like kinase 1 inhibitors, mitotic stress and the tumor suppressor p53. *Cell cycle* **12**, 1340–1351 (2013).
46. Gascoigne, K. E. & Taylor, S. S. Cancer cells display profound intra- and interline variation following prolonged exposure to antimetabolic drugs. *Cancer Cell* **14**, 111–122 (2008).
47. Tanenbaum, M. E. & Medema, R. H. Mechanisms of centrosome separation and bipolar spindle assembly. *Dev. Cell* **19**, 797–806 (2010).
48. Tanenbaum, M. E. *et al.* Kif15 cooperates with eg5 to promote bipolar spindle assembly. *Curr. Biol.* **19**, 1703–1711 (2009).
49. Vanneste, D., Takagi, M., Imamoto, N. & Vernos, I. The role of Hk1p2 in the stabilization and maintenance of spindle bipolarity. *Curr. Biol.* **19**, 1712–1717 (2009).
50. van Heesbeen, R. G., Tanenbaum, M. E. & Medema, R. H. Balanced activity of three mitotic motors is required for bipolar spindle assembly and chromosome segregation. *Cell Rep.* **8**, 948–956 (2014).

#### Acknowledgements

This work was supported by the Medical Research Council (UK) Integrative Toxicology Training Partnership; Cancer Research UK Programme Foundation Award C47320/A21229; and Cancer Research UK grant number C37/A18784. RF is supported by a PhD scholarship from King Saud University, Riyadh, Kingdom of Saudi Arabia.

#### Author Contributions

L.S. performed the analyses and co-designed the experiments. R.F. generated and validated the CRISPR p53 knock-outs. D.W.M. and A.S. supervised the project and co-designed the experiments. S.A. and L.B. contributed to the experimental strategy and provided intellectual input. All authors reviewed the manuscript.

#### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-16394-2>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017