Targeted deletion of repeat sequences in the 3’UTR of MDM2 using CRISPR/Cas9.

by

Abraham Awada

A thesis submitted to the Faculty of Graduate and Postdoctoral affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Biology

Carleton University
Ottawa, Ontario

© 2023 Abraham Awada
Abstract

Mutations in the *TP53* tumour suppressor gene are found in over 50% of cancers. This p53 protein is of central importance to prevent cancer progression. The MDM2 oncoprotein is an E3 ubiquitin ligase responsible for negatively regulating p53 to prevent sustained ectopic expression of p53, and adverse consequences for cells. The *MDM2* 3’UTR is exceptionally long (up to 75% of the 7.5 kb mRNA sequence) and contains many putative *cis*-acting regulatory elements, including unique repeat sequences predicted to form complex secondary structures. I generated several clonal populations of HCT116 cells heterozygous for a deletion of this region using a tandem sgRNA CRISPR-Cas9 approach. No gross differences in MDM2 protein or mRNA expression were detected. The deletions appeared to be facilitated by regions of homology and cannot readily be explained with the current understanding of DNA repair pathways associated with CRISPR-Cas9 gene editing.
Acknowledgments

I would like to thank the Carleton University Department of Biology for providing me with the resources necessary to succeed throughout the completion of my Master’s Degree. I would like to especially thank Dr. Bruce McKay for his supervision and support over the past two years, his patience and intuition allowed me to not only effectively perform experiments but to see how my research may become meaningful. Many thanks are also extended to both Dr. Kyle Biggar and Dr. Marc Ekker for being members of my advisory committee and providing insight and guidance on which direction my research should proceed. I would also like to thank Dr. Erin Van Zyl for mentoring me and sharing her invaluable knowledge in the lab which was critical to my success. My appreciation for previous and current McKay lab members also cannot be understated, as I am thankful for the friendships and companionship that they provided me with.
Table of Contents ........................................................................................................ Page #

List of Illustrations ..................................................................................................... v

List of Tables .............................................................................................................. vi

1. Introduction .............................................................................................................. 1
   1.1 The p53 tumour suppressor .............................................................................. 1
   1.2 MDM2 regulates p53 expression ...................................................................... 5
   1.3 Transcriptional and post-transcriptional regulation of MDM2 ....................... 9
   1.4 CRISPR ............................................................................................................ 13
   1.5 Rationale .......................................................................................................... 16

2. Materials and Methods .......................................................................................... 19
   2.1 Cell Culture and Drug Treatment .................................................................... 19
   2.2 Reverse Transcription and real-time quantitative polymerase chain reaction (RT-qPCR) .................................................................................................................. 19
   2.3 Immunoblot Assays ......................................................................................... 22
   2.4 Tandem gRNA Transfection Strategy ............................................................. 22
   2.5 RT-qPCR Dilution Standards .......................................................................... 23
   2.6 Statistical Analyses of Clones ......................................................................... 24

3. Results .................................................................................................................... 26
   3.1 MDM2 Expression in HCT116 Cells ............................................................... 26
   3.2 Generations of deletions in the 3’UTR of MDM2 ............................................. 29
   3.3 MDM2 expression in clonal populations ....................................................... 38
   3.4 Deletions occurred between sites of microhomology .................................... 44

4. Discussion .............................................................................................................. 46
   4.1 Heterozygous deletion of the repeat region ................................................... 46
       yielded unclear 3’UTR processing
   4.2 3’UTR Deletions did not follow CRISPR-Cas9 DSB repair literature ............. 49

5. Conclusion ............................................................................................................. 54

6. References ............................................................................................................. 55

7. Supplementary Materials ...................................................................................... 59
   7.1 Sequencing results (FASTA) .......................................................................... 59
List of Illustrations ...................................................................................... Page#

Figure 1 The p53 response illustrating MDM2 and the nutlins .......................... 2
Figure 2 MDM2-null and p53-null mouse models and their corresponding viability .... 7
Figure 3 Predicted hairpin structure in the 3’UTR of MDM2 ............................. 12
Figure 4 CRISPR-Cas9 mechanism of action and associated DSB repair pathways ... 14
Figure 5 Tandem CRISPR gRNA strategy and DNA repair pathway ..................... 18
Figure 6 Assessment of the HCT116 cell line ................................................ 28
Figure 7 Primer and gRNA design relative to the repeat sequence ......................... 30
Figure 8 Assessment of the deletion in the pooled population ............................ 32
Figure 9 Deletion screening and analysis in clonal populations ............................ 35
Figure 10 Heterozygosity assessment in clones ................................................ 37
Figure 11 Immunoblot analysis of all cell lines ................................................... 39
Figure 12 3’UTR mRNA expression analysis between clones ............................ 41
Figure 13 Repeat region mRNA expression analysis in clones ............................ 43
Figure 14 CRISPR-induced microhomology site usage ....................................... 45
Figure 15 Actual microhomology site usage compared to predicted usage .......... 52
List of Tables ................................................................................................................................. Page#

Table 1 Primer and gRNA sequences ................................................................. 21
Table 2 Statistical analysis of clones ................................................................. 25
Table 3 Sequencing results from break spanning primers .......................... 34
1. Introduction

1.1 The p53 tumour suppressor:

Identifying and studying the role of oncoproteins and tumor suppressors associated with cancer progression is crucial to expanding our knowledge of cancer as they are critical determinants of carcinogenesis. Mutations in the TP53 gene are found in over 50% of cancers as its protein p53 is a tumour suppressor responsible for the transcriptional regulation of over 500 genes (Aubrey et al., 2018). Genes associated in pathways such as cell senescence, DNA repair, apoptosis, and cell-cycle arrest are activated directly or indirectly by p53 to prevent cell growth and cell cycle progression in the presence of DNA damage or a stressed cellular state to reduce the potential for malignant transformation (Aubrey et al., 2018). DNA damage or cellular stress may induce a p53-mediated response by activating the ataxia-telangiectasia mutated (ATM) and ataxia-telangiectasia and Rad3-related (ATR) kinases which orchestrate the DNA damage response through downstream phosphorylation and activation of p53 (Marechal & Zou, 2013).
Figure 1. **Schematic representation of the p53 response.** p53 levels are controlled by MDM2, a p53-regulated E3 ubiquitin ligase that targets p53 for nuclear export and proteasome-mediated degradation. During times of DNA damage and cellular stress, the transcriptional activation of p53 may induce the upregulation of genes in cellular pathways involved in cell cycle arrest, DNA repair, cellular senescence, and apoptosis. Nutlins can also activate p53 by inhibiting this negative feedback loop (BioRender).
Mitotic cell division (mitosis) requires orderly progression through the cell cycle which is the fundamental biological process in which cell division occurs, and consists of DNA replication, cell growth, and cell division stages (Senturk & Manfredi, 2013). The cell cycle consists of G\(_1\), S, G\(_2\), and M phases, and cycle progression towards mitosis is heavily regulated by cyclin-dependent kinases (CDKs) (Senturk & Manfredi, 2013). The G\(_1\) (Gap 1) phase of the cell cycle is a growth phase in which the cell size increases in response to the cell preparing for DNA replication (Barnum & O’Connell, 2014). The S or synthesis phase of the cell cycle is the DNA synthesis phase where the DNA content in the cell doubles (Barnum & O’Connell, 2014). G\(_2\) (Gap 2) is the second gap phase in which the cell prepares for the mitotic phase (Barnum & O’Connell, 2014). The M phase of the cycle involves the nuclear division of genetic material, and subsequent cytokinesis resulting in the separation of two daughter cells (Barnum & O’Connell, 2014).

CDKs are a family of serine/threonine kinases that coordinate the cell cycle into its later stages by heterodimerizing with cyclins to regulate cell division (Borrero & El-Deiry, 2021). CDKs regulate the progression through the cell cycle from one phase to another by phosphorylating substrates that promote components of cell cycle progression such as DNA synthesis or cell growth (Barnum & O’Connell, 2014). One master regulator of the cell cycle that is targeted by CDKs is the retinoblastoma (Rb) protein (Vélez-Cruz & Johnson, 2017). Rb is a tumour suppressor that exerts control over the cell cycle by binding E2F transcription factors which promote the expression of genes involved in DNA replication and cell proliferation (Vélez-Cruz & Johnson, 2017). Rb bound to E2F transcription factors prevents the recruitment of activators and recruits co-repressors of gene expression, this interaction safeguards the G\(_1\) to S transition (Vélez-Cruz & Johnson, 2017). Rb is phosphorylated by
CDK2, CDK4, and CDK6 which results in its inactivation and dissociation from E2F transcription factors, resulting in cell-cycle progression due to the alleviation of transcriptional repression by Rb (Vélez-Cruz & Johnson, 2017).

The CDKN1A gene coding the p21 protein is induced by p53 and it is critical for p53-mediated cell-cycle arrest (Borrero & El-Deiry, 2021). p21 mediates cell-cycle arrest at different stages in the cell cycle by inhibiting cyclin-dependent kinases (CDKs) such as CDK1, CDK2, and CDK4 (Borrero & El-Deiry, 2021). CDK1 activity is highest in the progression from $G_2$ to M phase, CDK2 is primarily involved in the progression from $G_1$ to S phase, and CDK4 regulates entry into the cell cycle and its initiation in $G_1$, where its activity is highest at the $G_1$ to S transition (Liao et al., 2017; Ding et al., 2020). Inhibition of these CDKs by p21 results in an inability for the cell cycle to progress and cell cycle arrest (Borrero & El-Deiry, 2021). p53 therefore is responsible for primarily arresting cells at the $G_1$ to S transition, as well as the $G_2$ to M transition (Senturk & Manfredi, 2013).

Checkpoint proteins offer another level of cell cycle regulation and may arrest the cycle at different stages during times of genotoxic stress to ensure that mutations are not passed on to daughter cells to reduce the risk for tumorigenesis (Senturk & Manfredi, 2013). The ATM and ATR kinases are DNA damage sensors that may phosphorylate p53 directly or indirectly by phosphorylating cell cycle checkpoint kinases checkpoint protein 1 (Chk1) and checkpoint protein 2 (Chk2) which signal to CDKs (Marechal & Zou, 2013). These kinases are responsible for regulating cell cycle arrest in response to genotoxic stress through phosphorylation of p53 and downstream transient induction of p21 (Marechal & Zou, 2013). p53 also induces p21 expression in cells undergoing cellular senescence, which is an
irreversible growth arrested state that may be caused by different sources of cellular stress (Mijit et al., 2020).

Apoptosis is one programmed cell death pathway induced by a variety of cellular stresses including DNA damage and oncogenic stress to help avoid malignant transformation (Wang et al., 2007). p53 mediates apoptosis by inducing proteins such as PUMA, which inhibits cellular pro-survival proteins Bcl2 and Bcl-XL (Aubrey et al., 2017). This results in permeabilization of the outer mitochondrial membrane, cytochrome c release into the cytoplasm, and formation of the apoptosome complex by binding apoptotic protease activating factor-1 (Apaf-1) (Aubrey et al., 2017). Cytochrome c binding to Apaf-1 results in activation of caspase-9, which then activates other effector caspases to degrade substrates vital for cell viability (Aubrey et al., 2017).

p53 may induce cell-cycle arrest or cell senescence to permit time for repair and may also induce the expression of DNA repair proteins such as DDB2 involved in nucleotide excision repair (NER) responsible for repairing DNA damage caused by UV (Borrero & El-Deiry, 2021). p53 exerts systemic control over the DNA damage response through downstream interaction with cell cycle machinery to maintain cellular homeostasis and prevent malignant transformation (Figure 1) (Borrero & El-Deiry, 2021).

1.2 MDM2 regulates p53 expression:

p53 protein expression is very low in non-malignant unstressed cells however its mRNA is maintained at a relatively high level (Aubrey et al., 2017). p53 is continuously translated but it is rapidly targeted for proteasomal-mediated degradation by murine double minute 2 (MDM2) to keep p53 protein levels low in dividing cells (Figure 1) (Aubrey et al.,
MDM2 is an E3 ubiquitin ligase that is induced in a p53-dependent manner in response to p53 activation (Toledo & Wahl, 2007). The MDM2 protein forms part of a feedback loop that is responsible for negatively regulating p53 to prevent sustained ectopic expression, and other adverse consequences of unregulated p53 activity (Toledo & Wahl, 2007). In malignant cells, MDM2 overexpression is common and may result in reduced p53 expression (Hou et al., 2019). MDM2 overexpression is present in many cancer types, including lung cancer, breast cancer, and colorectal cancer (Hou et al., 2019).

The p53 protein possesses several functional domains including an N-terminus with two transactivation domains that interact with the transcriptional machinery of its target genes and pathways (Sullivan et al., 2018). MDM2 binds the first transactivation domain, DNA-binding domain, and the carboxy-terminal domain of p53 through its N-terminal domain and its acidic domain to repress its activity (Kung & Weber, 2022). MDM2 promotes p53’s ubiquitination, nuclear export, and disrupts its ability to serve as a transcription factor (Kung & Weber, 2022). Homozygous deletion of MDM2 is embryonic lethal in mice by day 8 of gestation (Jones et al., 1995). Compound knockout mice (p53null/ MDM2null) were fertile, developmentally sound, and phenotypically indistinguishable from wild-type mice, however were susceptible to early spontaneous tumour progression and died at a young age, this relationship is illustrated in Figure 2 (Jones et al., 1995). Therefore, the primary function of MDM2 is to keep p53 levels under control and deregulated p53 is inconsistent with mammalian development.
Figure 2. MDM2-null and p53-null mouse models and their corresponding viability. MDM2-null mice are subject to early embryonic lethality unless a homozygous knockout of p53 is also present (Jones et al., 1995; BioRender).
MDM2 also interacts with other proteins that play roles in regulating the MDM2/ p53 feedback loop (Pant et al., 2011). Alternate open reading frame (ARF) and murine double minute 4 (MDM4) are critical proteins MDM2 interacts with for regulation of p53 (Pant et al., 2011; Kung & Weber, 2022). ARF serves to activate p53 by physically inhibiting MDM2, while MDM4 heterodimerizes with MDM2 for stability so that MDM2 can direct its ubiquitin ligase activity towards p53 (Pant et al., 2011; Kung & Weber, 2022). ARF binds the acidic domain of MDM2 through its N-terminal domain causing the sequestering of the ARF-MDM2 complex in the nucleolus, preventing the nuclear export of p53 and consequently preserving its functionality (Kung & Weber, 2022). The heterodimerization between MDM2 and MDM4 is required for p53 regulation in early stages of embryogenesis (Pant et al., 2011). Complete homozygous deletion of either MDM2 or MDM4 is embryonic lethal and phenotypic rescue can be attained through concomitant deletion of p53 (Pant et al., 2011). Both MDM2 and MDM4 possess RING domains which serve as the source of their interaction, and homozygous deletion of the MDM4 RING domain results in embryonic lethality in a similar manner to homozygous MDM2 deletion (Pant et al. 2011).

E3 ubiquitin ligases aside from MDM2 also directly and indirectly influence the MDM2/ p53 axis (Sane & Rezvani, 2017). There are currently dozens of known E3 ubiquitin ligases that directly target p53 for degradation, such as Pirh2, Cop1, and CHIP (Sane & Rezvani, 2017). E3 ubiquitin ligases can also indirectly regulate the MDM2/ p53 axis. Siva1 is an E3 ubiquitin ligase that degrades ARF, resulting in increased MDM2 expression (Sane & Rezvani, 2017). Regulation of the feedback loops between p53 and associated proteins by E3 ubiquitin ligases is critical in the cellular response to genotoxic stress (Sane & Rezvani, 2017).
1.3 Transcriptional and post-transcriptional regulation of MDM2:

The human MDM2 gene is very complex. It contains twelve exons subject to alternative splicing, produces mRNAs up to 7.5 kilobases (kb) in transcript length, and codes mRNA isoforms with exceptionally large 3’ untranslated regions (UTRs) that can make up over 75% of the sequence (Browning et al., 2020). 3’UTRs possess elements within their sequence that post-transcriptionally and translationally regulate mRNA and protein expression (Browning et al., 2020). The 3’UTR of MDM2 has increased in size throughout primate evolution from under 1kb to up to 5.7kb in length (Browning et al., 2020). Determinants of mRNA stability such as micro-RNA (miRNA) binding sites, alternative polyadenylation sites (APA), and AU-rich elements in the 3’UTR regulate its mRNA expression (Browning et al., 2020). The expansion of the MDM2 gene occurred through insertion of a variety of transposable elements including 5 Alu elements, and the inclusion of a processed RLP24 pseudogene downstream from the open reading frame (Browning et al., 2020). Transposable elements are sequences of DNA that possess the ability to relocate in the genome and may serve as regulators of gene expression (Drongeitis et al., 2019). Alu sequences are a type of short interspersed nuclear element (SINE) (repetitive non-coding DNA sequence) approximately 300 nucleotides in size and are a type of transposable element (Chen & Yang, 2017). Alu sequences regulate mRNA expression by forming secondary structures and introducing new cis-acting regulatory sites to 3’UTR sequences that subject them to increased levels of regulation (Chen & Yang, 2017). Pseudogenes are DNA sequences that have mutated throughout the course of evolution that resemble ancient genes however are inactive and often found in non-coding regions (Pink et al., 2011). The RLP24 pseudogene was inserted into the 3’UTR of MDM2 in a retrotransposition event that took
place millions of years ago and introduced cis-acting regulatory elements to the MDM2 mRNA transcript (Browning et al., 2020).

The 3’UTR of MDM2 has over 2000 predicted miRNA sites, of which 247 have been functionally characterized (Browning et al., 2020). miRNAs are short non-coding RNAs 22 nucleotides in size that regulate gene expression by binding the 3’UTR of target mRNAs and silencing them (Ardekani & Naeini, 2010). MiRNA are initially transcribed into primary transcripts known as pri-miRNA by RNA polymerase II or RNA polymerase III (Abdelfattah et al., 2014). pri-miRNA are then processed into the precursor hairpin molecule pre-miRNA by the DROSHA enzyme, this initiates the translocation to the cytoplasm of the pre-miRNA molecule by the Exportin-5 nuclear export factor for further processing (Abdelfattah et al., 2014). The enzyme Dicer then processes the pre-miRNA molecule and recruits other proteins to form the mature single-stranded miRNA, in which it is introduced into the RNA-induced silencing complex (RISC) (Abdelfattah et al., 2014). Mature miRNAs within the RISC may bind to complementary sequences within the 3’UTRs of target mRNAs and silence them by inhibiting translational machinery or degrading the mRNA transcript (Abdelfattah et al., 2014). Long 3’UTRs tend to have more potential miRNA binding sites because the number of sites present is proportional to 3’UTR sequence length (Cabrita et al., 2017).

The 3’UTR of MDM2 is subject to APA which influences isoform selection by producing mRNA variants that differ in 3’UTR length while leaving the coding region unaffected (Pereira-Castro & Moreira, 2021). APA subjects mRNA variants to differential regulation through the inclusion or exclusion of cis-acting regulatory elements in the 3’UTR (Pereira-Castro & Moreira, 2021). It is also salient that cancer cells often have shorter 3’UTRs through alterations in APA that avoid miRNA-mediated gene silencing (Pereira-
The 3’UTR of *MDM2* possesses 11 functional polyadenylation sites of which over 90% of the polyadenylation is reserved to just 3, the predominant polyadenylation site in *MDM2* is derived from the 3’ end of the RLP24 pseudogene (Browning *et al.*, 2020). In addition, our laboratory identified a previously uncharacterized cluster of DNA sequence repeats 44 nucleotides in length in the 3’UTR of *MDM2* just upstream of the major polyadenylation site (Browning *et al.*, 2020). The same repeats were found in a similar pattern in two other intergenic regions in the genome (Browning *et al.*, 2020). These repeat sequences in the 3’UTR of *MDM2* are predicted to form complex secondary structures, notably a large complex hairpin structure approximately 800 nucleotides in size upstream from the main polyadenylation site located in the mRNA transcript from nucleotides 5158-5944 (accession ID: NM_002392.6). It is known that secondary structure may act as a hindrance for polymerase efficiency and speed, as regions with complex secondary structure are known to pause and decrease the speed of RNA polymerase II (RNAPII) (Georgakopoulos-Soares *et al.*, 2022). RNAPII speed influences promoter-proximal pausing, exon recognition, splicing, and transcription termination (Georgakopoulos-Soares *et al.*, 2022). It is therefore possible that this repeat sequence in the 3’UTR, which is immediately upstream of the predominant polyadenylation site affects 3’end processing of the *MDM2* mRNA at the nearby site. We hypothesized that deletion of the repeat region would result in the selection of a more distal polyadenylation site and decrease mRNA stability, resulting in reduced MDM2 expression and potentially affecting the MDM2/p53 axis.
Figure 3. A representative RNA folding structure containing the repeat sequences (786 nucleotides) in the 3’UTR of MDM2. mFold was used to predict the secondary structure of the repeat region using the lowest conforming free energy (The UNAFOLD Web Server).
1.4 CRISPR:

The CRISPR/Cas9 system has been adapted from bacteria to create a sequence-directed gene editing tool that can be used to induce DNA double-stranded breaks (DSBs) in a defined sequence (Figure 4). CRISPR is used by bacteria and archaea to defend against viral infection by storing foreign viral genetic material in spacer sequences within their genome known as CRISPR arrays (Arroyo-Olarte et al., 2021). Should the bacteria or archaea re-encounter the virus, the spacer sequences derived from previously encountered viral genetic material are transcribed into crispr RNA (crRNA) and bound to a nuclease to target complementary sequences for excision (Arroyo-Olarte et al., 2021).

The adapted gene editing tool is comprised of an sgRNA which directs the endonuclease Cas9 to the desired sequence to induce a double-stranded break (Redman et al., 2016). The CRISPR sgRNA is a composite of two separate CRISPR-Cas9 RNAs, the crispr RNA (crRNA) and the tracrRNA (Redman et al., 2016). The crRNA component shares sequence complementarity with the sequence of interest and directs the complex, while the tracrRNA component forms the stem-loop structure to connect the crRNA to the Cas9 endonuclease (Redman et al., 2016). The protospacer adjacent motif (PAM) site is a three-nucleotide sequence located 2-5 nucleotides downstream from the 3’ end of the sgRNA that directs the Cas9 enzyme to induce a DSB (Redman et al., 2016).
Figure 4. **CRISPR-Cas9 mechanism of action and main associated single DSB repair pathways.** The CRISPR sgRNA guides the Cas9 endonuclease to the sequence of interest and creates a double-stranded DNA break. Non-homologous end joining and homology-directed repair are the two most common DNA repair pathways in the event of a single double-stranded break in the DNA sequence (BioRender).
DSBs including those induced by Cas9 are repaired by numerous DNA repair pathways with the general consensus being that the two most common pathways: non-homologous end joining (NHEJ) and homologous recombination (HR) are critical for the repair of CRISPR DSBs (Mao et al., 2008). CRISPR-mediated damage due to a single sgRNA-induced DSB is preferentially repaired by NHEJ (Song et al., 2021). NHEJ is more common than HR as it is faster, requires less cellular energy, and doesn’t require an available homologous sequence (Mao et al., 2008). The initial step of NHEJ is the recognition of the DSB by the Ku70/Ku80 heterodimer (Ku) which binds the DNA ends and recruits DNA dependent protein kinases (DNA-PKcs) to signal DNA damage (Brandsma & van Gent, 2012). The Artemis nuclease then binds the break and may trim DNA ends in the presence of nucleotide incompatibility, the ligation complex consisting of DNA ligase IV (LIG4), X-ray cross-complementation group 4 (XRCC4), and XRCC4 like factor (XLF) then ligates the break site (Brandsma & van Gent, 2012). The precise repair product is not entirely predictable (Mao et al., 2008).

NHEJ is an error-prone DNA repair pathway that frequently introduces small insertions or deletions known as indel mutations (Mao et al., 2008). Indel mutations may result in frameshift mutations and pre-mature stop codons that trigger the nonsense mediated decay (NMD) pathway (Mao et al., 2008). The NMD pathway is a proofreading measure that monitors the sequence quality of mRNA transcripts prior to translation (Nickless et al., 2017). mRNA transcripts with premature stop codons are identified by ribosomes during translation, and RNA decapping and deadenylation factors are recruited to expose transcript ends to cytoplasmic endonucleases for degradation (Nickless et al., 2017). NMD allows CRISPR-Cas9 to generate functionally null cell lines (Redman et al., 2016).
HR permits correction of mutations or introduction of specific mutations with CRISPR-Cas9 by providing a homologous template for repair and is initiated by 5’ - 3’ end resection of the break site by the MRN complex and CtIP (Ramsden et al., 2022). The 3’ ssDNA overhang is then coated with replication protein A (RPA) for the removal of secondary structure (Ramsden et al., 2022). BRCA2 displaces RPA so that strand invasion into the sister chromatid by the Rad51 protein may occur to search for a homologous sequence to be used as a template strand (Ramsden et al., 2022). The strand invasion event and use of a homologous sequence from the sister chromatid results in HR being an error-free mechanism of DSB repair (Ramsden et al., 2022).

1.5 Rationale:

We sought to delete repeat sequences in the 3’UTR of MDM2 to determine their function. Small indels are unlikely to disrupt untranslated regions efficiently. Transfection with a single sgRNA would not be effective in producing a large-scale deletion through either NHEJ or HR, we considered a tandem CRISPR gRNA strategy that can be used to delete the region between two sgRNA target sites. This strategy has been successfully used to generate an ATF4-deficient cell line in HCT116 cells (van Zyl et al., 2021). HCT116 are diploid colorectal carcinoma cells with a stable karyotype so they are often used for knockout studies (Bunz et al., 1998). These cells also express p53 and have an intact p53-MDM2 feedback loop because they exhibit a robust p53 response when treated with the MDM2 antagonist nutlin-3a (Shen & Maki, 2011). Nutlin-3a binds the p53-binding pocket of MDM2, preventing MDM2 from controlling p53 expression and activity (Aziz et al., 2011). Nutlin-3a is used as a non-genotoxic activator of p53 since it does not bind p53, directly interact with DNA or lead to genotoxic stress (Aziz et al., 2011). We predicted that this
method would be amenable to generating large deletions within the 3’UTR of MDM2 to test their function in untreated and nutlin-3a treated cells.
Figure 5. **Tandem CRISPR gRNA strategy and DNA repair pathway.** Dual sgRNAs located upstream and downstream of the repeat region induce DSB repair (BioRender). Loss of the intervening sequence is anticipated in some clonal populations.
2. Materials and Methods

2.1 Cell Culture and Drug Treatment:

HCT116 cells (Cat#: CCL-247, Manassas, VA) were obtained from the American Tissue Type Collection. HCT116 cells and associated cell lines were cultured in Mccoys media (Hyclone, San Angelo) and supplemented with a 12% mixture (3:1 ratio) of heat-inactivated NewBorn Calf Serum (Wisent BioProducts, Saint-Jean-Baptiste) to Fetal Bovine Calf Serum (Wisent), and 90 μg/mL penicillin, 90μg/mL streptomycin (Hyclone). Cells were seeded a minimum of 24 hours prior to being treated at a concentration of 0.5 × 10⁶ cells/mL in a 3, 6, or 10cm dish. Cells were treated in growth media with varying concentrations up to 20μM of Nutlin-3a (Sigma-Aldrich, St. Louis, MO) diluted in dimethyl sulfoxide (DMSO) (Calbiochem), or an equivalent volume of DMSO to serve as a vehicle control.

2.2 Reverse Transcription (RT) and real time quantitative polymerase chain reaction (RT-qPCR)

RNA was isolated using the BioBasics EZ-10 DNAaway RNA isolation kit (Cat#: BS88136), quantified using the DeNovix DS-11 Spectrophotometer, and proportional amounts were converted into complementary DNA (cDNA) using the High-Capacity cDNA reverse transcription kit (Cat#: 4368813, ThermoFisher Scientific). cDNA was diluted with 200μL of 0.1% diethyl pyrocarbonate (DEPC) (Cat#: D5758-25ML, Sigma-Aldrich) water. RT-qPCR was completed using the SensiFAST™ SYBR Hi-ROX Kit (Cat#: BIO-92005, BioLine) using the following primers on a StepOnePlus™ Real-Time PCR system (ThermoFisher Scientific): Coding Region F, R, UTR1 F, R, UTR2 F, R, UTR3 F, R, Deleted Region F, R, and GAPDH F, R, served as a loading control. These primers were
synthesized at Integrated DNA Technologies (Coralville, Ia). Sequence for all primers and
gRNAs used are found in Table 1.
Table 1. Primer and gRNA Sequences Used in Steps 2.2, 2.4, and 2.5.

<table>
<thead>
<tr>
<th>RNA Species</th>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>F Primer</td>
<td>Coding Region</td>
<td>CCATGATCTACAGGAACCTTGGAATAGTA</td>
</tr>
<tr>
<td>R Primer</td>
<td>Coding Region</td>
<td>TCACTACAGGATACCTTGAGTCC</td>
</tr>
<tr>
<td>F Primer</td>
<td>UTR1</td>
<td>AGTGAGAAATGCTCAATTCACA</td>
</tr>
<tr>
<td>R Primer</td>
<td>UTR1</td>
<td>GAGGTTGGTGAAGGAGGATGAGG</td>
</tr>
<tr>
<td>F Primer</td>
<td>UTR2</td>
<td>ATGACAGGGTCAGCATGTGG</td>
</tr>
<tr>
<td>R Primer</td>
<td>UTR2</td>
<td>CCTGCAACAGGGGTACAGAT</td>
</tr>
<tr>
<td>F Primer</td>
<td>UTR3</td>
<td>CCATGCCCTGCCACTTTAGA</td>
</tr>
<tr>
<td>R Primer</td>
<td>UTR3</td>
<td>CAGGCTGCCATGTGACCTAA</td>
</tr>
<tr>
<td>F Primer</td>
<td>Break Spanning</td>
<td>TAGTGGGAGGAGGAGGCGA</td>
</tr>
<tr>
<td>R Primer</td>
<td>Break Spanning</td>
<td>AATAGCTGTCAGTCCTCAGA</td>
</tr>
<tr>
<td>F Primer</td>
<td>Deleted Region</td>
<td>AGCAAAAACAGTTGAAAAATAACCCA</td>
</tr>
<tr>
<td>R Primer</td>
<td>Deleted Region</td>
<td>TGCTGTCAGAATTACCTCAGT</td>
</tr>
<tr>
<td>F Primer</td>
<td>GAPDH</td>
<td>AGCCACATCGCAGACAGACA</td>
</tr>
<tr>
<td>R Primer</td>
<td>GAPDH</td>
<td>GCCCAATACGCAACAAATC</td>
</tr>
<tr>
<td>gRNA</td>
<td>sgRNA 1</td>
<td>AATGCTTGGAGGCTCCTCCAC</td>
</tr>
<tr>
<td>gRNA</td>
<td>sgRNA 2</td>
<td>ATAGTACACTTGATATATGG</td>
</tr>
</tbody>
</table>
2.3 ImmunoBlotting

Cell monolayers were washed with PBS and then disrupted for protein collection using RIPA Buffer (Cat#: R0278, Sigma-Aldrich). Samples were sonicated for 10 seconds on ice. Protein quantification was performed using the Bio-Rad protein assay and equivalent amounts of protein were loaded onto a NuPage 4-12% Bis-Tris gel (Cat#: NP0336BOX, Invitrogen). Proteins were transferred to a nitrocellulose membrane and stained with Ponceau S (1.3 mg/mL in 1% Acetic Acid) to confirm equal loading and transfer. Membranes were blocked in 5% Milk in 1X TBST for one hour. Membranes were then incubated in the following primary antibodies overnight at 4°C: MDM2 (Santa-Cruz Biotechnology), p21 (Calbiochem), p53 (Santa-Cruz), and Actin (Sigma-Aldrich). Membranes were then washed 4 times for 5 minutes in 1X TBST pH 7.6 and incubated with either 1:10000 goat anti-mouse horseradish peroxidase (HRP) conjugated (Abcam) or 1:20000 goat anti-rabbit HRP conjugated (Abcam) secondary antibody for 2 hours at room temperature. Membranes were washed 4 more times for 5 minutes in 1X TBST pH 7.6 and then incubated in Clarity Western ECL substrate (Biorad) before being imaged in a Fusion FX Vilber Lourmat Imager.

2.4 Tandem gRNA Transfection Strategy

Tandem CRISPR single guide RNAs (sgRNA 1 and sgRNA 2) were designed using the IDT design too (Coralville, Ia), and their sequences can be found in Table 1 (https://www.idtdna.com/site/order/designtool/index/CRISPRSEQUENCE). HCT116 cells were seeded in a 96-well dish at 1.0 × 10^6 cells/ mL 24 hours prior to being transfected. The Cas9 ribonucleoprotein complexes were formed individually by mixing 2.4μg of each sgRNA, 250ng of Alt-R S.p Cas9 Nuclease V3 (IDT) and Cas9 Plus™ Reagent (IDT) in 150uL of serum and antibiotic free McCoy’s Media (Hyclone) and then allowing an
incubation period of 5 minutes at room temperature. 7.5μL of Lipofectamine CRISPRMAX™ reagent (Invitrogen) was diluted in 150μL of serum and antibiotic free media with serum in a separate tube. The tubes were then combined and incubated at room temperature for 20 minutes. Once the incubation period was complete, 100μL of HCT116 cells diluted to $0.4 \times 10^6$ cells/ mL was added to the now-combined transfection complexes. This final cocktail of cells and RNP complexes was then incubated for 48 hours at 37°C. Post 48-hour incubation, cells were seeded in 3 96-well plates at an average of 1 cell/ well. Clonal populations derived from single cells were isolated, released with 0.25% Trypsin-EDTA (Cat#: SH3004202, Fisher Scientific), and transferred to individual wells of 24-well dishes. The clonal populations were then expanded, and RNA was isolated for cDNA and RT-qPCR was performed isolated for screening using primers flanking the sgRNA sites found in Table 1 as Break Spanning F and R primers, respectively. PCR products were resolved on 2% agarose gels, visualized using a 1:10000 dilution of SYBR® Safe DNA gel stain (Cat#: S33103, Invitrogen), and imaged using the Fusion FX Vilber Lourmat Imager. Gel excision of PCR products of interest was then performed, and the PCR product was purified using the BioBasic EZ-10 Spin Column DNA Gel Extraction Kit (Cat#: BS654). These products were reamplified by RT-qPCR and then PCR products were purified using the BioBasic EZ-10 PCR Product Purification Kit (Cat#: BS364) and sent for sequencing at the Sequencing Facility of the Stemcore Laboratory at the Ottawa Hospital Research Institute. (Supplementary figures).

2.5 qPCR Dilution Standards

The relative efficiency of qPCR amplification was determined for each primer set by analysis of serial dilutions (1X, 10X, 100X, 1000X). Dilutions in 0.1% DEPC water (Sigma-
Aldrich) were made with cDNA from clonal populations and from parental cells to ensure that amplification was equally efficient using cDNA from deletion and control cells. RT-qPCR was then performed on a StepOnePlus™ Real-Time PCR system (ThermoFisher Scientific) using non-diluted and diluted samples. The relative Ct method was used to estimate relative concentration of input cDNA and a simple linear regression was performed (PRISM 9) to compare these values to relative input of dilution standards. The slope of the line represented the amplification efficiency of clones and parental cells. All PCR primer sets in Table 1 were tested in multiple clones. All primer sets approached 2-fold amplification per cycle (range of slopes: 0.9830 – 1.016).

2.6 Statistical Analyses of Clones

One-way and two-way ANOVA tests were performed to determine statistical significance among cell lines. One-way ANOVA statistical analysis employing the Dunnett multiple comparisons test was performed using the PRISM 10 software to analyze data in Figure 6. Nutlin-treated groups were analyzed relative to the DMSO vehicle control. Two-way ANOVA statistical analysis was performed employing the Dunnett multiple comparisons test using the PRISM 10 software to analyze data in figures 12 and 13. The row factor was nutlin-3a concentration, and the column factor was cell line (Table 2).
Table 2. Statistical Comparison of Clones

<table>
<thead>
<tr>
<th>Test</th>
<th>P-Value</th>
<th>Figure</th>
<th>Significant</th>
<th>Multiple Comparison Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-Way ANOVA</td>
<td>0.05</td>
<td>6</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>Two-Way ANOVA</td>
<td>0.05</td>
<td>12</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Two-Way ANOVA</td>
<td>0.05</td>
<td>13</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
3. Results

3.1 MDM2 expression in HCT116 cells

HCT116 cells are diploid human colorectal carcinoma cells that express wildtype p53 and exhibit a functional MDM2-mediated feedback loop that can be disrupted with nutlin-3a to induce a p53 response that includes increased MDM2 expression (Shen & Maki, 2011). To ensure that our HCT116 cells retained this loop, cells were treated with nutlin-3a for six hours and collected for western blot analysis of p53, p21, and MDM2 (Figure 6A). The p21 protein was included as another well-characterized target of p53 (Karimian et al., 2016). Actin was used as a loading control to monitor relative protein abundance between samples (Figure 6A). The expression of all 3 proteins increased in response to treatment with 10 and 20 µM nutlin-3a while actin actually decreased (Figure 6A). Therefore, the p53-MDM2 response to nutlin-3a appears to be intact.

The 3’UTR of MDM2 contains multiple alternative polyadenylation sites so it was important to ensure that MDM2 mRNAs contained the repeat sequence of interest. To do this, we designed PCR primers amplifying a variety of regions along the 3’UTR of MDM2 as well as the open reading frame to ensure that these sequences were represented in constitutively expressed and/or p53-induced MDM2 transcripts (Figure 6B). HCT116 cells treated with the MDM2 antagonist nutlin-3a exhibited a similar increase in PCR product at all MDM2 3’UTR positions examined (Figure 6C). Baseline MDM2 levels were about one tenth of that of GAPDH at UTR positions 1 and 2 and nutlin-3a increased MDM2 levels to be similar to GAPDH levels (Figure 6D). The RT-PCR signal at UTR3 was about 5-fold lower than more proximal primer sets before and after nutlin-3a exposure indicating that most MDM2 transcripts are polyadenylated between the UTR2 and UTR3 primer sites (Figure 6D), a
region containing the most commonly used poly-A signal sequence (position 6109 to 6114) (Browning et al., 2020). We have found that this polyadenylation site is readily detected by 3’RACE in HCT116 cells (Browning et al., 2023). This site is downstream of the repeat region of interest (Browning et al., 2020). Importantly, the ability to detect basal and p53-induced MDM2 mRNAs at UTR3 ensures that the repeat region is represented in MDM2 mRNAs and that the p53/MDM2 feedback loop was intact in these cells. The increase in MDM2 mRNA abundance relative to the vehicle control (DMSO) in parental HCT116 cells differed significantly by one-way ANOVA in many treated groups in Figure 6D.
Figure 6. The effect of nutlin-3a on MDM2 expression. A. HCT116 cells were treated with 10μM and 20μM nutlin-3a and the expression of p53, MDM2, and p21 was assessed by immunoblotting. Actin was used as an endogenous loading control. B. Schematic illustration of the position of primers relative to the open reading frame, polyadenylation sites and the hairpin sequence of interest (MDM2 variant 1: NM_002392.6). C. HCT116 cells were exposed to the indicated concentration of nutlin-3a for 6 hours. RNA was collected for qRT-PCR analysis with the primers in B. Data is expressed as the fold change in MDM2 expression detected with the indicate primer compared to no drug controls. D. The data in C is expressed as a ratio to GAPDH. Each value in C and D was determined from a minimum of 3 independent experiments and is expressed as the mean +/- SEM. ND: No Drug control. * p<0.05.
3.2 Generation of deletions in the 3’UTR of MDM2

Once expression was confirmed at the protein and mRNA level, CRISPR sgRNAs flanking the hairpin sequence were designed using the IDT webtool (Figure 7). PCR primers were then designed to flank the sgRNAs upstream and downstream of the repeat region to distinguish wildtype and deleted sequences by gel electrophoresis of PCR products (Figure 7). This strategy was expected to reduce the size of the PCR product generated with flanking primers by approximately 800 nucleotides, making it easier to resolve the bands.
Figure 7. The position of PCR primers and sgRNA sites relative to the predicted hairpin sequence of interest. Forward primer 1 (F primer) and reverse primer 1 (R primer) were designed to flank sgRNAs 1 and 2. Block-like structures within the hairpin region represent four identical repeat sequences and one half-repeat within the 3’UTR. The position of the sgRNAs and PCR primers is relative to variant 1 of the MDM2 mRNA sequence. Black arrows represent relative orientation of the repeat sequences.
We used a tandem guide RNP strategy to deliver sgRNAs with recombinant Cas9 to target the 3’UTR of MDM2 in HCT116 cells. In an initial PCR test of pooled transfectants, we were able to detect the deletion product (Figure 8A). Individual bands were gel isolated and sent for sequencing using the F and R primers used in PCR amplification at StemCore labs at the Ottawa Hospital Research Institute. Forward and reverse sequence runs containing sequences 5’ and 3’ to the target sites were used to confirm the deletion. The 500 bp band was derived from the 3’UTR of MDM2 and a region of just over 900-nucleotides had been deleted (Figure 8). The forward and reverse sequences detected similar deletions but may indicate that there may have been two or more similar PCR products in the 500 bp band. The PCR amplification from the non-transfected parental cells were expected to generate a PCR product of 1.2kb but this falls within the high molecular weight smear and was not detected.
Figure 8. **Deletion detection in the pooled population.** A. PCR gel electrophoresis using cDNA isolated from the pooled population compared to cDNA used in the parental non-transfected population. Primers “F Primer” and “R Primer” depicted in Figure 7 were used to identify bands of interest. B. Schematic representing sanger sequencing results following gel excision and purification of the band of interest (BioRender).
Clonal populations of cells were obtained from the same transfection by serial dilution of cells from the pooled population. 48 hours following transfection twenty-eight separate clones were isolated and PCR analysis of cDNA from individual clones was performed to identify those carrying the deletion of interest. Twenty-two clones carried PCR products of the expected size and 4 of these were selected for sequencing and subsequent analysis (Figure 9B). Several extra bands were detected but sequence analysis indicated that were not generated from the $MDM2$ gene, so they appear to be non-specific PCR bands (Table 3).
### Table 3. All Sequencing Results from Targets of Break Spanning Primers

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Approximate PCR Fragment Size</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone 13</td>
<td>400</td>
<td>MDM2</td>
</tr>
<tr>
<td>Clone 23</td>
<td>300</td>
<td>N/A</td>
</tr>
<tr>
<td>Clone 23</td>
<td>500</td>
<td>MDM2</td>
</tr>
<tr>
<td>Clone 23</td>
<td>650</td>
<td>CFAP95</td>
</tr>
<tr>
<td>Clone 24</td>
<td>500</td>
<td>MDM2</td>
</tr>
<tr>
<td>Clone 24</td>
<td>650</td>
<td>N/A</td>
</tr>
<tr>
<td>Clone 28</td>
<td>500</td>
<td>MDM2</td>
</tr>
</tbody>
</table>
Figure 9. Identification of clones with deletion of the repeat sequence. A. A representative agarose gel used to screen for deletions is presented. Bands of interest are observable at approximately 500bp using the deletion spanning F Primer and R primer. B. Schematic illustration indicating the position of each deletion relative to sgRNAs, and the repeat sequence being targeted. Parentheses enclose bands of interest.
Sequence analysis of the 500 bp fragments indicated that all 4 clones contained deletions of the repeat region of interest (Figure 9B). The deletions were not identical, clones 23, 24 and 28 were very similar while clone 13 contained a larger deletion more similar to that detected in the mixed population in figure 8. To determine whether both alleles of MDM2 contained deletions, primers were designed within the deleted region. PCR analysis yielded the anticipated product in all clones and sequence analysis of gel excised bands confirmed that they were in the repeat region within the MDM2 3’UTR. We interpret this to indicate that all clones were heterozygous for the deleted allele (Figure 10A and B). Amplification of a PCR product of 132 base pairs would suggest that at least 1 allele retained the repeat region (Figure 10B).
Figure 10. **The clones are likely heterozygous.** A. PCR gel electrophoresis in clones 13, 23, 24, and 28 indicate that all clones contain wildtype sequence in the repeat region and are thus heterozygous. B. Schematic illustration of the location of the PCR primers used to test for the deletion (created with BioRender).
3.3 MDM2 expression in clonal populations

In order to determine if there were differences in expression of MDM2 across clones, protein was collected from control and nutlin-3a treated parental HCT116 cells and subclones carrying deletions in the 3’UTR of MDM2. The baseline level of p53 was similar among cell lines and p53 protein increased as expected following exposure to nutlin-3A (Figure 11). The baseline level of MDM2 expression was barely detectable before treatment but MDM2 increased in cell lines (Figure 11). Although there appears to be some variability in MDM2 expression, differences appear to be related to differences in loading, based on actin levels.
Figure 11. The effect of nutlin-3a on the expression of p53 and MDM2. Cells were treated with 10μM nutlin-3a to induce a p53 response and MDM2 and p53 proteins were visualized by western blotting. Actin was used as a loading control. wt: parental cells with wild-type MDM2. Similar results were obtained in 2 other independent experiments.
Similarly, RNA was isolated from untreated and nutlin-3a treated parental HCT116 cells and clones carrying deletions in the 3’UTR of MDM2 for qRT-PCR analysis of MDM2 expression using a variety of primers spanning the mRNA. Primers in the coding region detected an approximately 8-fold increase in MDM2 mRNA in all clones (Figure 12A). Similar increases in MDM2 expression were also detected with primers in the 3’UTR (Figure 12A). The fold increase in MDM2 expression did not differ significantly among clones (P=0.89, 0.50, 0.42 and 0.46, respectively by one way ANOVA). Therefore, the deletion in the MDM2 3’UTR did not detectably affect the p53-dependent induction of MDM2. The primary MDM2 transcript is subject to alternative polyadenylation and the primers used above were designed to detect a variety of mRNAs with different 3’ends (recall Figure 6B). The predominant polyadenylation site in MDM2 lies between PCR primer sets UTR2 and UTR3 (Browning et al, 2020). As expected, we detected a decrease in the amount of PCR product at the distal primer set (Figure 12B). Nonetheless, MDM2 was significantly increased at all positions (P<0.0001 for all 4 primer sets) but there were no significant differences among cell lines (P=0.57, 0.85, 0.27 and 0.68, respectively by two-way ANOVA) (Figure 12A). Therefore, we did not detect isoform-specific or cell line-specific differences in p53-induced MDM2 expression.
Figure 12. The expression of *MDM2* mRNA in clones containing deletions in the *MDM2* 3’UTR. A. Fold change mRNA expression in response to nutlin-3a. B. Absolute mRNA transcript abundance relative to GAPDH in all cell lines. Parental HCT116 and subclones with deletions in the 3’UTR of *MDM2* were treated with nutlin-3a or left untreated for 6 hours. RNA was collected for qRT-PCT analysis using the indicated primers in the coding region (ORF) and three sites within the 3’UTR (UTR1, 2 and 3) as outline in figure 6B. Each value represents the mean +/- SEM determined from between 3 and 6 independent experiments. P: Parental non-transfected HCT116 cells.
HCT116 cells are diploid and the four clones carrying deletions in the 3’UTR of MDM2 also carry wildtype sequence in the deleted region. Therefore, the clones appear to be heterozygous for the unedited allele. This complicates analysis of the qRT-PCR data above because qRT-PCR results represent the sum of PCR product generated from two different alleles. In an attempt to distinguish between the response of the deleted and non-deleted alleles, we used the primers amplifying within the deleted region (recall Figure 6B) and thus recognize only the unedited allele in each sample. We detected an MDM2 response in all clones comparable to that detected at all other positions within the MDM2 mRNA (Figure 13A). However, MDM2 levels, expressed as a ratio to GAPDH, were lower than expected for transcripts terminating around nucleotide 6120 (2nd PAS site), even in the parental cells. We suspect that the complex secondary structure in this area reduced the efficiency of cDNA synthesis and thus template available for amplification. No significant differences among cell lines were observed (P=0.71, 0.79, 0.99, and 0.99, respectively by two-way ANOVA).
Figure 13. *MDM2* expression detected using primers within the deleted repeat sequence. RNA was collected as described in Figure 12. Primers designed to amplify between repeat sequences (recall figure 10) were used in qRT-PCR analysis. Data is expressed as fold change in *MDM2* expression (A) and a ratio to GAPDH in (B). Each value represents the mean +/- SEM determined from at least 3 independent experiments.
3.4 Deletions occurred between sites of microhomology

As previously indicated, the specific location of each deletion was determined by direct sequencing. It was immediately obvious that the deletions were not in or adjacent to the sgRNA binding sites (or PAM sequences). The deletion in clones 23, 24 and 28 spanned from approximately 45 nucleotides downstream of sgRNA1 site to 80 nucleotides upstream of the sgRNA2 target site (recall Figure 9B). Similarly, the deletion in clone 13 was located from about 120 nucleotides upstream of the sgRNA1 site to around 60 nucleotides upstream of the sgRNA2 site (recall Figure 9B).

We analyzed the sequence in cis orientation at the DNA repair sites required to generate the deletion. Remarkably, all deletions occurred between repeat sequences (Figure 14A and B). This observation, coupled with the distance from the sites of DNA CRISPR-mediated DNA strand breaks, can’t easily be explained by NHEJ or HR, the two pathways thought to contribute to the repair of CRISPR-induced DNA strand breaks. Our data suggests that these deletions originated through an alternative DNA repair pathway that uses microhomology and can occur at a distance from DSBs.
Figure 14. Deletions in the 3’UTR of MDM2 occurred between regions of microhomology. Schematic representation of sequence alignments across repair events illustrating that deletions occurred between sites with microhomology. A. Deletions in clones 23, 24 and 28 occurred between identical AGTTC sequences. Clone 23 differed from 24 and 28 by one nucleotide, retaining a T instead of an A immediately downstream of the deletion. This T was retained from position 5162 of the upstream sequence instead of the A at position 5904 of the downstream sequence. B. The regions of homology in clone 13 differed by 1 nucleotide (CTAGAAACCA vs CTAGAAGCA). The deleted allele retains the C from the downstream sequence (Figures were generated with BioRender).
4. Discussion

4.1 Heterozygous deletion of the repeat region did not appear to significantly affect 3’UTR processing or protein expression

The 3’UTR of MDM2 contains numerous putative cis-acting regulatory elements that serve as determinants of its mRNA stability (Browning et al., 2020). The 3’UTR of MDM2 has expanded throughout primate evolution and produces 3’UTR isoforms up to 5.7kb containing many regulatory sites (Browning et al., 2020). It was initially hypothesized that deletion of potential regulatory sequences, specifically four and a half identical repeat sequences predicted to form complex secondary structure immediately upstream of the main polyadenylation site would affect 3’end processing by selection of longer transcripts. Heterozygous deletion of these sequences in four independent clones did not appear to result in any significant change in mRNA expression or protein expression. However, this is complicated by the fact that the deletion was only present in one allele. The contribution of the individual alleles to MDM2 mRNA and protein expression could be different without the ability to distinguish them.

Fold change in MDM2 expression was measured at a variety of positions between nucleotides 612 and 6405 of the MDM2 mRNA in response to nutlin-3a treatment in all 4 clones and their parental cells line. There were no significant differences in nutlin-3a-induced MDM2 expression among cell lines at any position suggesting that the heterozygous deletion strains had a p53-MDM2 response equivalent to that of the parental cells. MDM2 expression relative to GAPDH levels appeared more variable towards the 3’end, both before and after exposure nutlin-3a, but again there was no significant difference among cell lines. There may
be insufficient power for statistical analysis or the presence of two alleles may affect our ability to detect alterations in expression.

One could have anticipated differences in expression due to shortening the 3’UTR by approximately 800 nucleotides, which would remove several cis-acting regulatory sites or secondary structure in nascent RNA that could affect RNA polymerase II efficiency (Mayr & Bartel, 2009). Although we did not detect differences among clones, we detected significant differences in relative MDM2 expression at different sites along the mRNA. As outlined in the material and methods, the primer sets were equally efficient, amplifying 2-fold per cycle. Therefore, the ratio of MDM2 to GAPDH at each position within the MDM2 mRNA could be used to infer differences in the expression of variants within the population.

The terminal exon of MDM2 contains the 3’UTR and all variant forms of MDM2 include the proximal end of this exon however alternative polyadenylation affects the final length of the 3’UTR and the presence of distal primer sequences in the mature mRNA and thus the cDNA population (Browning et al, 2020). In our analysis here, the UTR1 and UTR2 primer sets yielded the most qRT-PCR product. More distal 3’UTR sequences (UTR3) were under-represented, likely because the majority of the polyadenylation still falls in between the UTR2 and UTR3 primer sets, suggesting that the main polyadenylation site (Figure 6A) is still being heavily favoured in the presence of the deletion. The main APA site yields an MDM2 mRNA isoform of 4.3kb and deletion of the sequence of interest reduces the main mRNA transcript to 3.6kb or 3.4kb (clone-dependent) given that the deletion did not cause a shift in polyadenylation site preference (Browning et al., 2020). Additional experiments such as 3’Rapid Amplification of cDNA Ends (3’RACE) may be performed to confirm APA site usage.
We also detected less *MDM2* using primers within the coding region near the 5’end of the *MDM2* mRNA. *MDM2* has twelve exons, and it is subjected to alternative splicing so the reduced signal in the coding region is consistent with alternative splicing (Browning *et al.*, 2020). The coding region PCR primers were designed using the sequence of isoform 1 of *MDM2* but the same sequences are present in most isoforms. Importantly, isoform 2 (encoding MDM2 isoform h) doesn’t contain the region recognized by the reverse primer based on BLAST analysis. It is possible that variant 2 makes up a significant fraction of the total *MDM2* mRNA in these cells. The encoded protein isoform doesn’t contain the epitope used to generate the monoclonal antibody (amino acids 154-167). Isoform h contains all the domains of MDM2 required to regulate p53 but lacks a spacer region with no known function between the p53-binding domain and the nuclear localization signal. However, we don’t know if the protein is expressed.

A consequence of a large-scale sequence deletion in the 3’UTR could affect mRNA half-life. The transcriptome average half-life of an mRNA is approximately 6 hours while the half-life of *MDM2* is less than an hour (Browning *et al.*, 2020). mRNA decay rate is directly related to 3’UTR length, as longer 3’UTRs are associated with a faster decay rate due to possessing an increased amount of *cis*-acting regulatory elements such as AU-rich elements (AREs), GU-rich elements (GREs), and miRNA binding sites (Mayr & Bartel, 2009). AREs and GREs are short mRNA sequences composed of adenosine and uracil, or guanine and uracil nucleotides that promote mRNA decay by recruiting RNA-binding proteins involved with deadenylation, decapping, and exonucleolytic degradation (Stoecklin *et al.*, 2006; Mayr & Bartel, 2009). The large 3’UTR of *MDM2* contains many of these regulatory sites that act as determinants of mRNA stability and contribute to the half-life of *MDM2* in the cell (Browning *et al.*, 2020).
mRNA half-life experiments may be performed as the large-scale deletion removed putative cis-acting regulatory elements from the transcript. Deletion of regulatory sequences such as miRNA binding sites or other determinants of mRNA stability may increase the mRNA half-life of *MDM2* in the cell, resulting in alterations to the p53 response. The deleted region of clone 13 possessed 55 predicted miRNA sites (miRDB), while the deleted region of clone 23, 24, and 28 possessed 43 predicted miRNA sites (miRDB). Though the deleted region of clone 13 began over 150 nucleotides upstream from the deleted region of the other clones and included more regulatory sequences, there was no significant difference in mRNA expression among cell lines (Figure 9B; Figure 12). Though all clones slightly displayed variability in both mRNA and protein expression upon induction with nutlin-3a, clone 24 displayed high basal mRNA and protein expression levels. This was consistent with high non-treated protein expression of p53 as well. This could be due to a mutation in *MDM2* or in *p53*, however further studies are required to confirm this as it was not detected in other clones with the same deletion. Immunoblot analysis revealed that MDM2 and p53 protein expression increased in all cell lines in the presence of nutlin-3a (Figure 12).

4.2 The locations of 3’UTR Deletions were unexpected based on the CRISPR-Cas9 literature

In this CRISPR-Cas9 gene editing strategy, sequencing results indicated that the deletions were not close to the sgRNA target sites or PAM sequences. The deletions occurred between regions of microhomology that were not in close proximity of the Cas9-target sites. Assuming that DNA strand breaks were induced at the Cas9-target sites, the DSB repair events are not consistent with either NHEJ or HR, the two DNA repair pathways implicated
in CRISPR-Cas9-mediated gene editing. They were most likely repaired by a microhomology-mediated process. One reported repair process yielding microhomology at the break site is microhomology-mediated end joining (MMEJ). Interestingly, a re-evaluation of deletions we reported in the *ATF4* gene using a similar tandem guide strategy in our recent report, indicates that this deletion occurred between identical 3 nucleotide sequences so this pattern may be common for tandem sgRNA CRISPR strategies (van Zyl *et al.*, 2021)

There is conflicting literature whether MMEJ is a version of NHEJ or if it is a distinct separate pathway. However, there is evidence that MMEJ is independent of NHEJ proteins Ku, XRCC4, and LIG4 required for NHEJ (Ramsden *et al.*, 2022). MMEJ also shares similarities with HR as both share a common intermediate, the end resection step (Ramsden *et al.*, 2022). The MMEJ response begins with the detection of a DSB by poly(ADP-ribose) polymerase-1 (PARP-1), which catalyzes end resection (Vu *et al.*, 2021). End resection involves 5’ nucleolytic processing of the DSB by the nuclease enzymes CtBP-interacting protein (CtIP) and the MRE11-RAD50-NBS1 (MRN) complex resulting in single stranded 3’ end overhangs near the break site (Ramsden *et al.*, 2022). Proteins such as PARP-1 and Fanconi anaemia group D2 protein (FANCD2) recruit DNA polymerase theta (Polθ) to ssDNA overhangs bound by replication protein A (RPA) (Ramsden *et al.*, 2022). Hydrolysis of ATP through the amino-terminal superfamiliy 2 helicase-like domain (HelD) of Polθ removes RPA from the resected 3’ overhangs so that the carboxy-terminal A-family polymerase domain (PolD) of Polθ can synthesize DNA at sites containing microhomology (Ramsden *et al.*, 2022). Microhomology annealing results in an unpaired 3’ flap which may be excised by nucleases such as xeroderma pigmentosum group F protein (XPF) and excision repair cross-complementing 1 protein (ERCC1), and flap endonuclease 1 (FEN1) (Ramsden
et al., 2022). DNA ligation is mediated by DNA ligase 3 (LIG3) and DNA ligase 1 (LIG1) to a lesser extent (Ramsden et al., 2022).

The relative position of the microhomology sites in our clonal populations to the predicted CRISPR induced cut sites in Figure 9B indicates that conventional MMEJ was not used, as microhomology sites were found within the deleted region. Current MMEJ literature suggests that microhomology sites upstream from sgRNA 1 and downstream from sgRNA 2 should have been used in 5’ end resection (Figure 14) (Ramsden et al., 2022). Instead, 7 of 8 repair sites fall between sgRNA target regions (within the anticipated deletion, recall Figure 9). Clones 23, 24, and 28 displayed a 5 nucleotide microhomology site [AGTTC] yielding a deletion of approximately 750 nucleotides, while the region of microhomology in clone 13 was 9 nucleotides in length [CTAGAACCA] and yielded a 900-nucleotide deletion. The mechanism explaining the selection of these microhomology sites is not known; however, the central domain of Polθ is known to interact with the Rad51 protein to antagonize HR (Zhou et al., 2021). The Rad51 protein binds at 3 different sites of the Polθ gene, and acts as a strand invasion protein in HR to repair DNA damage by invading and using the other allele as a template strand (Kruchinin & Makarova, 2023; Ceccaldi et al., 2015). Zhou et al. demonstrated that Polθ expression is synthetic lethal in HR-deficient ovarian cancer cells and that subsequent inhibition of Polθ induces cell death (Zhou et al., 2021). This interaction does not explain the selection of microhomology sites however a strand invasion event could explain the microhomology site usage in clones given that more research needs to be done to understand the full extent of the interaction between Polθ and Rad51.
Figure 15. Microhomology site usage did not follow conventional MMEJ. MH sites were found within the deleted region in both positions of clones 23, 24, and 28, and in one position in clone 13. Positioning of MH site usage supports a strand invasion hypothesis (BioRender).
The microhomology sites used in clones 23, 24, and 28 were within the deleted region and therefore should not exist in the edited allele, however the sites were present in the wild-type allele supporting the hypothesis of a strand invasion event (Figure 15). Dual-guide experiments in different cell lines should be performed in the future to monitor this repair mechanism as HCT116 cells possess mutated mismatch repair (MMR) protein human *MutL* homolog (MLH-1) which may potentially disrupt the overall balance between repair pathways (Imesh *et al.*, 2012). The main function of MLH-1 in MMR is to coordinate downstream processes by interacting with mismatch recognition proteins, and to recruit proteins such as exonuclease 1 (EXO1) to excise mismatched nucleotides (Martin-Lopez & Fishel, 2013). MLH-1 depletion has been shown to exhaust RPA caused by excess production of ssDNA overhangs due to EXO1 hyperactivity, therefore potentially affecting multiple DNA repair pathways (Guan *et al.*, 2021). Regardless of mechanism, this observation has important implications in understanding CRISPR gene editing outcomes.
5. Conclusion

We hypothesized that CRISPR-mediated deletion of repeat sequences immediately upstream of the main alternative polyadenylation site formed predicted complex secondary structures in the 3’UTR of MDM2 and would affect 3’end processing and 3’UTR expression. The extent to which deletion of the repeat region affects expression remains unclear. qPCR and immunoblot analysis revealed very little variation in mRNA and protein expression among clones and parental cells (Figure 11, Figure 12). However, caution should be exercised in interpreting these findings as we were only able to generate heterozygotes. The most surprising finding was the clear preference for deletions to occur between sites of microhomology and that these could arise from sequences between CRISPR target sites (i.e. within the intended deletion). Further studies using both clonal cell lines and non-cancer cell lines are required for understanding the mechanism surrounding the production of CRISPR-induced microhomology sites.
6. References:


Browning JWL, Rambo TME, McKay BC. Comparative genomic analysis of the 3’ UTR of human MDM2 identifies multiple transposable elements, an RLP24 pseudogene and a cluster of novel repeat sequences that arose during Primate evolution. Gene. 2020;741:144557.


7. Supplementary Materials:

7.1 Sequencing Results (Accession ID: NM_002392.6):

Relative mRNA position: 4875

agccgaaataagggttgcctgaaataaagtgcactatataattttgtttggcaaatactaagtctctgctgttagttgtaaacaacagttattttttttttttttttttaagccacacaattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt