Proximity Labeling of p53 Bound Enhancer Elements

An honors thesis presented to the
Department of Biological Sciences
University at Albany, State University of New York
In partial fulfillment of the requirements
for graduation with Honors in Biological Sciences
and
graduation from The Honors College

Daniel Koskas

Research Mentor: Morgan Sammons, Ph.D.

April 2022

Abstract

A cell's fate is determined by complex programs of gene expression. Growth and cell differentiation are influenced by regulatory elements like enhancers and promoters. Transcription Factors (TFs) bind to these regulatory elements to initiate and control the rate of transcription as well as recruit transcriptional machinery such as RNA Polymerase II to increase the likelihood that a particular gene is transcribed. The transcription factor p53 functions as a tumor suppressor in vertebrates and plays a key role in regulating the cell's response to various forms of stress. We will be creating a doxycycline-inducible APEX-p53 plasmid using PCR amplification and Gibson Assembly which will be used to proximity-label neighboring proteins when P53 is bound to DNA. Doing so allows us to identify the location of P53 at any point during its stress-mediated response cycle. We believe that APEX2-mediated biotinylation may allow for a simple and cost-effective way to proximity label proteins, RNA species, and identify DNA binding protein locations in the genome when compared to alternative methods like BioID and Gro-Seq.

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my advisor Dr. Morgan Sammons for allowing me to be a part of this amazing research opportunity. Dr. Sammons has always pushed me to do my best work and approach biology in a new light. I learned to embrace mistakes I made during my time in the lab because every mistake is an opportunity to grow and learn. Without his constant support and guidance, I would have never been able to be where I am today as a student and as a researcher. I could not have had a better mentor for my undergraduate research and for that I am grateful.

I would also like to extend my thanks to Serene Durham, who was the graduate student I had the pleasure of working with. No matter how many questions I had or issues with my experiments I had to fix, Serene was always there to help and guide me. I seriously could not have done this without her and I owe her the sincerest amount of gratitude.

I would also like to thanks my other fellow lab members: Jhos Dufflart, Dana Woodstock, Gaby Baniulyte, Abby McCann, and Kate Sazon. These individuals always made coming into the Sammons lab a fun and welcoming experience.

List of Figures

Table of Contents

Introduction

Cell differentiation and growth are determined by highly regulated and complex programs of gene expression. Genes are sequences of DNA of variable lengths ranging from a few hundred to over 2 million base pairs in length (Lowry et al. 2016). These sequences are made up of nucleotides that serve as the instructions to synthesize gene products, such as RNA and proteins that have the ability to control the traits and expression of an organism. Transcription is the process in which the information stored within these gene sequences is encoded into mRNA molecules (messenger RNA). These coding mRNA molecules are then processed and read by ribosomes to produce proteins. However, not all RNA is transcribed into a protein. These non-coding RNA segments (ncRNA) have the ability to regulate gene expression and transcription without being decoded into proteins (Mercer et al. 2009).

Highly regulated interactions between transcription factors and transcriptional regulatory elements like enhancers and promoters control the network of a cell's gene expression and growth. Recent genetic studies have discovered highly regulated programs of gene expression, which have been shown to be regulated by large interconnected regulatory networks (Davidson, 2010). Genes expression is regulated through the collaborative interactions of cis-regulatory elements like core promoters and promoter-proximal elements, in addition to enhancers that are located at sizable distances away from the transcriptional start sites (TSSs) (Spitz, 2012). Interactions between regulatory elements like enhancers and promoters are integral to understanding this complex network of regulation. In short, they are DNA encoded sequences that serve as scaffolds for transcription factors and control gene transcription in space and time.

Promoters are responsible for turning a gene on or off because the process of transcription is initiated at the promoter region. They are a sequence of DNA located directly upstream of a gene or the 5′ end of the transcription initiation site. Transcription factors and RNA polymerase bind to the promoter in order to initiate transcription of the gene. Enhancers are small sequences of DNA (around 50-1500 base pairs long) that transcription factors can bind to in order to recruit the appropriate transcriptional machinery, like RNA polymerase, that work in tandem with promoter sequences by increasing the likelihood that a transcription of a specific gene will occur (Pennacchio et al. 2013). Unlike promoters, enhancers are not located directly upstream of the target gene. Enhancers can be located at various distances relative to the target gene, even up to one mega base away (1,000,000 base pairs). These elements are able to function at such a distance by forming chromatin loops that serve to move the enhancer into closer proximity to the target gene. Researchers believe that DNA-binding transcription factors bound at promoters and enhancers are able to recruit looping factors that can mediate the long-range interactions (Pennacchio et al. 2013).

Transcription factors are proteins that bind to specific DNA regulatory regions, namely enhancer and promoter regions. They are able to initiate gene transcription (sometimes inhibit) and aid in the production of mRNA through their DNA interactions (Zaret, 2011). TFs are able to recognize small 6-12bp long DNA sequences. This low sequence specificity points to TF having much more complex rules for their interactions with DNA, aside from just a simple affinity to bind to DNA (Spitz, 2012). Transcription factors do not operate independently, instead, they form complex regulatory networks where several factors interact together to regulate gene transcription. The transactivation domain in transcription factors is responsible for their cooperative associations as it provides scaffolds for other proteins to bind such as transcriptional

coregulators (Papavassiliou, 1995). The mystery of how TFs are able to interpret specific areas of the genome and bind to the correct sequences at the "right" time has been intensely studied over the past few decades. Studies using methods such as Chip-Seq and Chip-Chip to study the binding sites of transcription factors have provided novel insights into the activity of TFs. The data suggest that TFs can vary widely in the number of genomic binding sites and exceed the known number of possible direct gene targets (MacQuarrie et al. 2011). The study of TF interactions cannot be reduced solely to their role in transcriptional regulation as they might have functionality independent of that.

The p53 family consists of three distinct genes, p53, p63, and p73. p53 evolved from the ancestral genes p63 and p73, which are genes found in almost all vertebrates (Sammons et al 2020). The origin of the p53 ancestor genes was first observed in single-cell choanoflagellates and metazoan sea anemones. In early organisms, the function of these genes was to protect their germ-line gametes from DNA damage, a function that is continued to be seen in vertebrates and humans (Belyi et. al 2009). These homologs share similar features with each other including structural organization and DNA binding sites. p73 and p63 have been shown to have the ability to transactivate targets of p53, such as P21 (the cell cycle inhibitor). p73 has also been found to activate other p53 promoter regions like *Bax*, IGF-BP3, and *cyclin G* (Di Como et al. 1999). While p63 and p73 bind to similar or identical DNA sequences as p53, there are differences between these family members that allow for them to take on specific and specialized roles within an organism. In a knockout experiment with mice, it was shown that p63 and p73 play important roles in the development of tissue in organisms. P63 knockout mice were observed to

be born without skin and appendages, suffering high mortality rates before birth (Yang et. al 1999). Similarly, p73 knockout mice often died in utero and had a high mortality rate (Yang et. al 2000).

The p53 transcription factor plays an integral role in maintaining cellular genomic integrity and controlling cell growth. It does so by binding to p53 responsive elements within the genome and activating transcription of the genes in proximity to p53 binding sites (Oren, 2003). The absence of p53 function can result in the abnormal growth of cells; the expression and activity of p53 are highly regulated as a result. In the cell, p53 is normally present at low levels due to its relatively short half-life, however, when the cell is exposed to stress, and DNA damage, p53 is stabilized and activated (Pietsch et al. 2008). p53 having such a short-half life prevents it from having extensive effects on cell phenotype, and similarly, the few p53 molecules that exist under low-stress conditions are ineffective as transcriptional activators (Oren, 2003). p53 also acts as a tumor suppressor and is mutated in nearly 50% of human cancers. Many tumors containing wild-type p53 are often deactivated or repressed by means of aberrant degradation, deregulation of activators, and repressors. (Sammons et al. 2020). Its translational function and regulatory network are able to determine whether cells should pause the cell cycle until the stress or damage is resolved, or if the cell should induce apoptosis and die (Sammons et al. 2020). p53 is unique in the fact that it can function as both a tumor suppressor and a tumor promoter.

When mutated, p53 can prevent wild-type p53 from binding to DNA and inhibit their stress response pathways. The most common alterations to p53 resulting in the majority of tumors come from a single amino acid substitution in the 393-amino-acid protein. Many of these mutations are localized in the central DNA binding domain and a number of hotspots have been

identified like R175, G245, R248, R349, R273, and R282 (Muller, 2013). Most cancers are able to avoid the pathway of p53, allowing them to remain undetected and continue growing.

Due to p53's importance in cancer and tumor development, p53 is one of the most studied proteins in biology and researchers have been intricately studying the stress response pathway for years. In the lab, our goal is to determine what proteins wild-type p53 interacts with when it is bound to DNA. Knowing what proteins p53 comes into contact with when active can provide us with insights into the stress response pathway. A question that arises though is how will we be able to know what proteins p53 comes into contact with?

Ascorbic acid peroxidase or APEX is an engineered fusion protein that allows for rapid proximity labeling. Initially designed as a way to enable the use of electron microscopy to resolve the sub mitochondrial localization of regulatory protein MICU1, it was found that APEX also acts as a promiscuous labeling enzyme for live-cell proteomics (Lam et. al 2015). In cells pre-incubated with biotin phenol (BP), APEX can catalyze the oxidation of biotin phenol to biotin phenoxyl when there is a presence of hydrogen peroxide. The resulting phenoxyl radicals react with proximal proteins within a radius of 1-10 nm from the APEX2-fusion. These biotin phenol radicals are short-lived and are unable to pass through the cellular membranes of intact cells and restrict the labeling to the subcellular compartment to which APEX was targeted (Hung et al 2016). Due to the powerful enzyme kinetics APEX allows for, it is capable of creating a snapshot of proximal proteins with a staggeringly quick labeling time (1 min). In prior experiments, APEX was bound to a mitochondrial matrix-targeting motif and was utilized to map the proteome of the mitochondrial matrix in mammalian cells, identifying 500 proteins and 30 novel protein candidates (Hyun-Woo et al. 2013). APEX can also be performed in *in vivo* model systems.

APEX also has the capability to label RNA which makes it a powerful tool to study the dynamics of RNAs and label the organization of the transcriptome. Previous studies have demonstrated that APEX can be used to biotinylate nearby RNA *in vivo* as well as map the subcellular RNA localization in mammalian cells (Padrón et. al 2019). For our experimental purposes, we will be using the improved variant of APEX, called APEX2. In prior experimental trials, an issue that arose with the initial APEX variant was its low sensitivity. When APEX was expressed at low levels, its interactions with biotin-phenol became undetectable. The APEX2 variant allows for higher biotinylation activity at lower levels of cells expressing APEX2 as well as performing better during lentiviral transductions (Lam et al. 2015).

There are other methods of proximity labeling like BioID which is a system that allows for a screening of protein-protein interactions within a living cell. This technique utilizes biotin ligase fused to a protein of interest that when expressed in cells can be induced to biotinylate any interacting proteins (Sears et al 2019). There are a couple of disadvantages that come with using BioID. It has difficulty detecting transient interactions as well as weak interactions. Furthermore, these issues are exemplified due to the inability to apply stringent wash conditions which leads to a considerable background level of irrelevant proteins (Roux et. al 2013). Additionally, BioID can take hours to label neighboring proteins.

Contrary to BioID, APEX is unable to provide a report of PPIs over an extended period of time due to the short active window (1 second) of biotinylation. This can lead to APEX missing some of the transient protein associations during labeling like the ones that occur during the various stages of the cell cycle. However, the rapid kinetics of APEX allows for protein labeling at any moment and allows for other avenues of experiments to be explored. Comparative experiments such as inducing APEX at specific times before or after a

pharmaceutical drug is administered are now possible because of the ability to control when APEX biotinylation occurs (Dae In Kim, 2016).

Global run-on sequencing (GRO-seq) is another method commonly used to identify transcriptional regulatory elements and is one of the most widely used techniques. GRO-seq uses a Nuclear Run-On Assay, where nuclei from mammalian cells are isolated and then washed to remove free nucleotides and kept at freezing temperatures to halt transcription. The major limitations of GRO-seq are the difficulty of the technique as well as the required starting amount of cells (around 10^7). GRO-seq is also limited to cell cultures and *in vitro* models. APEX2 labeling can provide advantages in areas that Gro-seq cannot excel in. As mentioned previously, APEX2 labeling can be done in *in vivo* models which allows for more accurate evaluation of TF interactions in a complex model. APEX2 also provides a less complex procedure to execute and can be a more accessible method compared to Gro-seq.

Our goal is to create a p53-APEX2 doxycycline-inducible plasmid using cloning techniques such as PCR amplification and Gibson Assembly. APEX2 being directly attached to p53 allows us to follow p53 during its cell stress response pathway and biotinylate any proteins, RNA species, and identify DNA binding protein locations. We believe that APEX2 can serve as an alternative technique to well-known methods like Gro-seq and BioID because of its ability to provide a simple and cost-effective way to proximity label.

Methods

Designing the APEX-p53-TET-ONE Experimental Plasmid

Bacterial Transformation and Plasmid Extraction

We will be using the pLV-Tet-One plasmid purchased from Takara Bio USA. This plasmid features a Tet-On 3G transactivator Protein that has been evolved to have a higher sensitivity towards doxycycline induction and a P_{TRE3GS} inducible promoter that allows for low basal expression and high maximal expression after induction. The P_{TRE3GS} also lacks binding sites for native mammalian transcription factors, so it is virtually silent before induction. The plasmid also includes a tet-on system that allows for controlled induction of the plasmid through the use of doxycycline (citation here). The starting backbone, CMV-APEX2-p53, was taken from former graduate student Ally Catizone. The pLV-Tet-One plasmid and plasmid CMV-APEX2-p53 were transformed using E Coli., and plated on LB-Agar plates supplemented with ampicillin at 37°C o/n. The following day, 3 colonies from each plate were selected for propagation in liquid LB cultures overnight at 37°C. Following the incubation, the DNA was extracted from the cultures using the Monarch Genomic DNA Purification Kit.

Polymerase Chain Reaction Amplification

PCR Primers SL1355 and SL1356 (see **Figure 1.**) were designed using the SnapGene tool and ordered from IDT (Integrated DNA Technologies). SL1355 would be used as the forward primer and SL1356 would be the reverse primer to amplify a 2770 nucleotide sequence corresponding to a fusion between the coding sequence of APEX2 and full-length p53 cDNA sequence. Using CMV-APEX-p53 plasmid as the template, APEX-p53 was PCR amplified using the 2X NE High Fidelity Q5 Master Mix (MM) and the designed primers at a volume of 25ul.

Two negative controls were established for this experiment. One negative control, CMV-APEX-p53 (-MM), was done without using any master mix to see if any amplification occurred despite the lack of primers and polymerase present in the MM. The second negative control, CMV-APEX-p53 (-Primers), had no Gibson primers present. The samples were annealed at 72 degrees C. The cycling conditions of the PCR were 25 cycles at 98°C for 30 s, 98°C for 10s, 72°C for 30s, and 72°C for 45s.

Primer Name:	Primer Sequence:
SL1355 Forward	attcaaggcctctcgagcctGAGGAGCCGCAGTCAGATC
SL1356 Reverse	taatacgactcactatagttTCAGTCTGAGTCAGGCCC

Figure 1. *Primer Names and Sequences for PCR Amplification of APEX2-p53*

Gel Electrophoresis/Gel Imaging/Gel Purification

Products of the PCR amplification, CMV-APEX-p53 (+), and CMV-APEX-p53 (-Primers) were loaded and run on an agarose gel with ethidium bromide at 140V for 1-2 hours. The gel was prepared at a 1% concentration. The gel was then imaged using the Bio-Rad gel imaging system and the bands for CMV-APEX-p53 (+) were cut out. The sample was then gel extracted using the Monarch DNA Gel Extraction Kit and eluted with 12ul of elution buffer. We then took the DNA concentrations using the Nanodrop.

Gibson Assembly

Gibson assembly was performed on the PCR amplified APEX-p53 (insert) and the digested PLV-TET-ONE (vector) at a 0.2 picomolar concentration (2:1 vector-insert) using

2X NEB High Fidelity Assembly Master Mix to clone the TET-ONE-APEX-p53 plasmid. A negative control was established, (-Insert) TET-ONE-APEX-p53 to assure that amplification would not occur without the insert present. The Gibson products were plated on LB agar plates supplemented with 50 mg/ml of ampicillin and incubated at 37°C for 24 hours. The following day, 3 colonies from each plate were selected for propagation in liquid LB cultures overnight at 37°C. Following the incubation, the DNA was extracted from the cultures using the Monarch Genomic DNA Purification Kit.

Designing the Negative Control TET-ONE-APEX Plasmid (APEX Alone)

Enzyme Digestion of PLV-X Parent Plasmid

2ul of plasmid PLV-TET-ONE Puro was added to a 1.5ml tube at a concentration of 1000ng/ul. 1ul of restriction enzyme *EcoRI* was added to the tube. 1ul of restriction enzyme *BamHI* was added along with 1ul of a 10X Fast-Digest buffer (FD-BUFF). The remaining volume was filled with deionized water for a total volume of 20ul. The tubes were then incubated at 37 degrees C for 45 minutes. While the tubes were being incubated an agarose gel was prepared by adding 1 gram of agarose powder and 1ml of TAE buffer to an Erlenmeyer flask. The contents were heated to a boil. After cooling down to 32°C, 5ul/mL of Thermo Fisher Scientific ethidium bromide was added to the mixture. Following incubation, 5ul of 5X Purple Loading dye was added to each sample. The samples were then loaded onto a 0.5mg/mL ethidium bromide agarose gel alongside a 100bp ladder and run at 140V for an hour. The gel was then imaged and the digested bands were extracted using the Monarch DNA Gel Extraction Kit.

PCR Amplification of APEX Alone

PCR Primers SL1633 and SL1634 (see **Figure 2**) were designed using the SnapGene tool and ordered from IDT (Integrated DNA Technologies). SL1633 would be used as the forward primer and SL1634 would be the reverse primer to amplify a 777 nucleotide sequence corresponding to a fusion between the N-terminus of the coding sequence of APEX and the C-terminus. Using CMV-APEX-p53 plasmid as the template, p53 was PCR amplified using the 2X NE High Fidelity Q5 Master Mix (MM) and the designed primers at a volume of 25ul. Two negative controls were established for this experiment. One negative control, CMV-APEX (-MM), was done without using any master mix to see if any amplification occurred despite the lack of primers and polymerase present in the MM. The second negative control, CMV-APEX (-Primers), had no Gibson primers present. The cycling conditions of the PCR were 25 cycles at 98°C for 30 s, 98°C for 10s, 72°C for 30s, and 72°C for 45s.

Primer Name:	Primer Sequence:
SL1633 Forward	ctcgcaggggaggtggtctgttaGGCATCAGCAAACCCAAG
SL1634 Reverse	tgactcagacGGAAAGTCTTACCCAACTG

Figure 2. *Primer Names and Sequences for PCR Amplification of APEX2*

Gel Electrophoresis/Gel Imaging/Gel Purification

Products of the PCR amplification, CMV-APEX (+), and CMV-APEX (-Primers) were loaded and run on an agarose gel with ethidium bromide at 140V for 1-2 hours. The gel was prepared at a 1% concentration. The gel was then imaged and the bands for CMV-APEX (+) were cut out. The sample was then gel extracted using the Monarch DNA Gel Extraction Kit and eluted with 12ul of elution buffer.

Gibson Assembly

Gibson assembly was performed on the PCR amplified TET-ONE-APEX segment (insert) and the digested PLVX backbone (vector) at a 0.2 picomolar concentration with a 2:1 ratio of insert to vector along with 10ul of NE High Fidelity Gibson Master Mix. Negative controls were established, (-Insert) APEX and (-MM) APEX.The Gibson products were plated on LB agar plates supplemented with 50 mg/ml of ampicillin and incubated at 37°C for 24 hours. The following day, 3 colonies from each plate were selected for propagation in liquid LB cultures overnight at 37°C. Following the incubation, the DNA was extracted from the cultures using the Monarch Genomic DNA Purification Kit.

Designing the Negative Control p53-TET-ONE Plasmid (p53 Alone)

PCR Amplification of p53 Alone

PCR Primers SL1635 and SL1636 (see **Figure 3**) were designed using the SnapGene tool and ordered from IDT (Integrated DNA Technologies). SL1635 would be used as the forward primer and SL1636 would be the reverse primer to amplify a 2000 nucleotide sequence corresponding to a fusion between the N-terminus of the coding sequence of p53 and the C-terminus. Using CMV-APEX-p53 plasmid as the template, p53 was PCR amplified using the 2X NE High Fidelity Q5 Master Mix (MM) and the designed primers at a volume of 25ul. Two negative controls were established for this experiment. One negative control, CMV-p53 (-MM), was done without using any master mix to see if any amplification occurred despite the lack of primers and polymerase present in the MM. The second negative control, CMV-p53 (-Primers),

had no Gibson primers present. The cycling conditions of the PCR were 25 cycles at 98°C for 30 s, 98°C for 10s, 72°C for 30s, and 72°C for 45s.

Primer Name:	Primer Sequence:
SL1635 Forward	aagactttccGTCTGAGTCAGGCCCTTC
SL1636 Reverse	cacttcctaccctcgtaaagatgGAGGAGCCGCAGTCAGATC

Figure 3. *Primer Names and Sequences for PCR Amplification of p53*

Gel Electrophoresis/Gel Imaging/Gel Purification

Products of the PCR amplification, CMV-P53 (+), and CMV-P53 (-Primers) were loaded and run on an agarose gel with 0.5mg/mL ethidium bromide at 140V for 1-2 hours. The gel was prepared at a 1% concentration. The gel was then imaged and the bands for CMV-APEX (+) were cut out. The sample was then gel extracted using the NEB Monarch DNA Gel Extraction Kit and eluted with 12ul of elution buffer.

Enzyme Digestion of PLV-TET-ONE Plasmid

4.5ul of plasmid PLV-TET-ONE Puro was added to a 1.5ml tube at a concentration of 1000ng/ul. 1ul of restriction enzyme *EcoRI* was added to the tube. 1ul of restriction enzyme *BamHI* was added along with 1ul of a 10X Fast-Digest buffer (FD-BUFF). The remaining volume was filled with deionized water for a total volume of 20ul. The tubes were then incubated at 37 degrees C for 45 minutes. While the tubes were being incubated an agarose gel was prepared by adding 1 gram of Fisher Scientific agarose powder and 1ml of TAE buffer to an Erlenmeyer flask. The contents were heated to a boil. After cooling down 5ul of

0.5mg/mL ethidium bromide was added to the mixture. Following incubation, 5ul of 5X Purple Loading dye was added to each sample. The samples were then loaded onto an ethidium bromide agarose gel alongside a 100bp ladder and run at 140V for an hour. The gel was then imaged and the appropriate bands were extracted using the NEB Monarch DNA Gel Extraction Kit.

Gibson Assembly

Gibson assembly was performed on the PCR amplified TET-ONE-p53 segment (insert) and the digested PLV-TET-ONE backbone (vector) at a 0.2 picomolar concentration with a 2:1 ratio of insert to vector along with 10ul of NE High Fidelity Gibson Master Mix. Negative controls were established, (-Insert) p53 and (-MM) p53. The Gibson products were plated on LB agar plates supplemented with 50 mg/ml of ampicillin and incubated at 37°C for 24 hours. The following day, 3 colonies from each plate were selected for propagation in liquid LB cultures overnight at 37°C. Following the incubation, the DNA was extracted from the cultures using the Monarch Genomic DNA Purification Kit.

Designing the p53-APEX-TET-ONE Plasmid

PCR Amplification of p53 and APEX

PCR Primers SL1635 and SL1636 (see Table XXX) were designed using the SnapGene tool and ordered from IDT (Integrated DNA Technologies). SL1635 would be used as the forward primer and SL1636 would be the reverse primer to amplify a 2000 nucleotide sequence corresponding to a fusion between the N-terminus of the coding sequence of p53 and the C-terminus. Using CMV-APEX-p53 plasmid as the template, p53 was PCR amplified using the 2X

NE High Fidelity Q5 Master Mix (MM) and the designed primers at a volume of 25ul. Two negative controls were established for this experiment. One negative control, CMV-p53 (-MM), was done without using any master mix to see if any amplification occurred despite the lack of primers and polymerase present in the MM. The second negative control, CMV-p53 (-Primers), had no Gibson primers present. The cycling conditions of the PCR were 25 cycles at 98°C for 30 s, 98°C for 10s, 72°C for 30s, and 72°C for 45s.

PCR Primers SL1633 and SL1634 (see Table XXX) were designed using the SnapGene tool and ordered from IDT (Integrated DNA Technologies). SL1635 would be used as the forward primer and SL1636 would be the reverse primer to amplify a 777 nucleotide sequence corresponding to a fusion between the N-terminus of the coding sequence of APEX and the C-terminus. Using CMV-APEX-p53 plasmid as the template, APEX was PCR amplified using the 2X NE High Fidelity Q5 Master Mix (MM) and the designed primers at a volume of 25ul. Two negative controls were established for this experiment. One negative control, CMV-APEX (-MM), was done without using any master mix to see if any amplification occurred despite the lack of primers and polymerase present in the MM. The second negative control, CMV-APEX (-Primers), had no Gibson primers present. The cycling conditions of the PCR were 25 cycles at 98°C for 30 s, 98°C for 10s, 72°C for 30s, and 72°C for 45s.

Gel Electrophoresis/Gel Imaging/Gel Purification

Products of the PCR amplification, CMV-P53 (+), and CMV-P53 (-Primers) were loaded and run on an agarose gel with 0.5mg/mL ethidium bromide at 140V for 1-2 hours. The gel was prepared at a 1% concentration. The gel was then imaged and the bands for CMV-APEX (+)

were cut out. The sample was then gel extracted using the NEB Monarch DNA Gel Extraction Kit and eluted with 12ul of elution buffer.

Products of the PCR amplification, CMV-APEX (+), and CMV-APEX (-Primers) were loaded and run on an agarose gel with 0.5mg/mL ethidium bromide at 140V for 1-2 hours. The gel was prepared at a 1% concentration. The gel was then imaged and the bands for CMV-APEX (+) were cut out. The sample was then gel extracted using the NEB Monarch DNA Gel Extraction Kit and eluted with 12ul of elution buffer.

Transient Transfection

HCT116 p53-/- (human colon carcinoma cells), acquired from the American Type Culture Collection (ATCC) were cultured, using Macoys's 5A media, to 60-80% confluency (500,000 cells) per well in two sets of a 6-well dish. In a 1.5ml tube, 2ug of each sequence-verified TET-ON plasmid was added (represents 10% of the total DNA per well) to 200ul of JETprime buffer and mixed by vortexing for 5 seconds. Contents of the tube were then briefly spun down in a microcentrifuge. Next, 4ul of JETprime was added and incubated at room temperature for 10 minutes. 200ul of the transfection mix was added to each well dropwise onto the cells and the cells were incubated at 37°C o/n.

	P53-APEX	P53-APEX (-DOX)	APEX-P53	APEX-P53 (-DOX)	P53 Alone	P53 Alone (-DOX)	APEX Alone	APEX Alone (-DOX)
JP Buff	200ul	200ul	200ul	200ul	200ul	200ul	200ul	200ul
JP Reagent	4ul	4ul	4ul	4ul	4ul	4ul	4ul	4ul
DNA	5.5ul	5.5ul	4.7ul	4.7ul	6.3ul	6.3ul	10.5ul	10.5ul

(2ug)								

Figure 4. *Measurements for Transient Transfection*

Doxycycline Induction

After a 24hr incubation, the media of the transfected HCT116 p53-Null Cells was replaced. One of the sets of 6-well plates received 10ug/mL of doxycycline in order to initiate the promoter of the TET-ON plasmid. The other 6-well plate received 2ul of deionized water. The cells were then incubated for an additional 24 hours in order to allow for the induction of the respective gene product

cDNA Preparation from RNA Extraction

Using the Applied Biosystems High Capacity cDNA kit, a master mix was prepared with RNA at 100ng/ul, 10x reverse transcriptase buffer, 25x dNTP mix, 10x random primers, MultiScribe reverse transcriptase, and nuclease-free water for a total volume of 20ul. A second master mix was prepared without reverse transcriptase as a negative control for future quantitative analysis. The samples were then placed in the thermocycler at 25°C for 10 min, 37°C for 120 min, 85°C for 5 min, and 4°C. Once completed, the concentrations of the DNA were determined via nanodrop.

Quantitative Polymerase Chain Reaction (qPCR)

A 10uM stock for each primer set was made, see **figure 5.** Next we made a 1:10 dilution of our cDNA reaction by taking 20uL of the cDNA reaction and adding 180uL of molecular-grade water. To make our standards for qPCR quantification, we used our cDNA reactions. 20uL of each diluted cDNA reaction was combined into a new tube. This represents the 1X standard.

Next we made a 1:10 dilution of our 1X standard to make our 0.1X standard. This was repeated for the 0.01X standard and the 0.001X standard. Our standards and cDNA primer stocks were then plated on the 384 well qPCR plate and loaded into the BioRad CFX 384 for analysis.

Name:	Volume:	Concentration:
FWD Primer	5uL	100uM
REV Primer	5uL	100uM
Molecular Water	40uL	NA
Total:	50uL	

Figure 5. *qPCR Primer Stock Measurements*

Results

Design and Synthesis of a Doxycycline-inducible APEX2-p53 Fusion Protein

Bacterial Transformation and Plasmid Extraction

Our goal is to construct a series of APEX2 p53 fusion protein plasmids and controls in order to determine what enhancer elements p53 interacts with during its stress response pathway. Therefore, In order to produce sufficient plasmid concentrations to perform subsequent molecular cloning experiments, we transformed plasmids pLV-Tet-One and CMV-p53-APEX into bacteria using a standard heat shock-mediated transformation. After a 1-hour recovery step, the bacteria were plated on LB-Agar plates supplemented with ampicillin to select for bacterial colonies that took up the plasmid. We observed colonies for each of the transformed plasmids, which suggests that the bacteria were able to take up our plasmid and successfully replicated it. 3 colonies from each plate for propagation in liquid LB cultures overnight at 37°C. We then isolated the plasmids at concentrations of 130ng/ul, 115ng/ul, and 100ng/ul for pLV-Tet-One and 145ng/ul, 113ng/ul, and 68ng/ul for CMV-APEX-p53. These concentrations provide us with a sufficient amount of plasmid for subsequent cloning experiments like amplification of appropriate DNA sequences from these templates using polymerase chain reaction.

Polymerase Chain Reaction Amplification

In order to clone our APEX-p53 segment into the Plv-TET-ONE plasmid to create the APEX-p53-Tet-On plasmid, we used a polymerase chain reaction amplification. This allowed us to isolate and amplify our APEX-p53 segment to use for subsequent Gibson assembly. The

primers for the reaction were designed using SnapGene; one primer complementary to the sequence upstream of APEX2-p53 and the other 30 base pairs downstream of APEX2-p53 (see **figure 18**). After overnight incubation in the thermocycler, the amplified DNA was run through a 1% agarose gel with ethidium bromide to isolate the amplified sequence. Plasmid CMV-APEX2-p53 was amplified along with two negative controls, CMV-APEX2-p53 (-MM), and CMV-APEX2-p53 (-Primers) as shown in **Figure 6**. The PCR products were imaged using the Bio-Rad gel imaging system and we observed bands for the amplified sample at 2777 base pairs and noticed no bands present for both of the negative controls. The size of the observed band corresponded to the predicted size of the amplicon, and coupled with the lack of bands present in the negative control reactions, suggested that the PCR reaction was successful because we were able to amplify this sequence to sufficient quantities for future experiments.

	CMV-APEX-P53	(-MM) CMV-APEX-p53	(-PRIMER) CMV-APEX-p53
2X Master Mix	12.5ul	N/A	12.5ul
Gibson Primers	1.5ul	1.5ul	N/A
Template DNA	2ul	2ul	2ul
H20	9ul	21ul	10.5ul

Figure 6. *PCR of CMV-APEX-p53 Measurements*

Gibson Assembly APEX-p53-TET-ONE

To clone our PCR amplified APEX-p53 segment into the PLV-TET-ONE vector, Gibson assembly was performed on the PCR amplified APEX-p53 (template) and the digested PLV-X plasmid (vector) along with a negative control that contained no template DNA seen in **figure 7**. The experimental control and negative control were plated on agar plates supplemented with

ampicillin to select for bacterial colonies with ampicillin resistance. After a night of incubation, colonies were seen on the experimental agar plate, but there were no colonies present on the negative control. Colonies from the agar plate were selected for propagation in liquid LB cultures overnight at 37°C. We then used the Monarch DNA Purification Kit to extract the DNA and used a nanodrop to record the concentrations; Sample A (154ng/ul), Sample B (305ng/ul), Sample C (189ng/ul), Sample D (245ng/ul). To further verify if the newly ligated plasmid maintained the desired DNA sequence, samples were sent out for Sanger sequencing.

	(+)APEX2-p53-TET-ONE	(-Insert)APEX2-TET-ONE-p53
p53 Insert	1ul	0ul
TET-ONE Vector	6ul	6ul
2X MASTERMIX	10ul	10ul
Water	3ul	4ul

Figure 7. *Gibson Assembly APEX2-p53-TET-ONE Measurements*

Enzyme Digestion of PLV-X Parent Plasmid

We performed a restriction digest on the Plv-TET-ONE plasmid to linearize it and create compatible ends for our PCR products. *BamHI* and *EcoRI* were identified as suitable enzymes for digestion because of their proximity to the P_{TRE3GS} promoter. This allows us to insert the amplified APEX-p53 sequence directly upstream of the promoter which allows for the inducible expression of the coding sequence. A digestion mixture was made with the Plv-Tet-One plasmid at a concentration of 1000ng/ul, digestion enzymes ECOR1 and BAMH1, and NEB Fast-Digest buffer to catalyze the reaction. Furthermore, a control without restriction enzymes was established to ensure that the digestion was successful (**Figure 8**). After a 30-minute incubation

at 37°C, the samples were run through a 10% ethidium bromide agarose gel for 1 hour to separate DNA fragments properly. Following gel imaging, we observed bands at 9000 base pairs for the digested lane and 9100 base pairs for the control and indicated that the plasmid was properly digested. We then isolated the plasmid at a concentration of 10ng/ul. To further verify the digestion was successful, a diagnostic digest was performed using the restriction enzyme *XhoI* (**Figure 9**). *XhoI* cuts a ≋1000bp segment out of the APEX2-p53 plasmid but does not cut the PLV-TET-ONE plasmid because it lacks the restriction site present in the ligated plasmid. After gel imaging, we observed no band at 1000bp for the PLV-TET-ONE plasmid, but we did observe one for the APEX2-p53-TET-ONE plasmid.

	+BAM/ECO	-BAM/ECO
PLV-X (1000ng/ul)	4.5ul	4.5ul
ECOR1	1ul	0ul
BAMH1	1ul	0ul
Water	13ul	13.5ul

Figure 8. *Enzyme Digestion of PLV-X Backbone (Measurements)*

	Plv-Tet-One	APEX-P53-Tet-One
Plasmid (1000ng/ul)	3ul	2.5ul
XH0I	1ul	1ul
NEB Fast Digest Buffer	2ul	2ul
H_20	14ul	14.5ul

Figure 9. *Diagnostic Digest of APEX-p53-Tet-One (Measurements)*

Design and Synthesis of a Doxycycline-inducible APEX2 Fusion Protein (Negative Control)

Polymerase Chain Reaction Amplification

To isolate and amplify the APEX-p53 plasmid for subsequent cloning experiments, we performed polymerase chain amplification on the CMV-APEX-p53 plasmid. The primers for the reaction were designed using SnapGene; one primer complementary to the sequence upstream of APEX and the other 30 base pairs downstream of APEX (see **figure 19**). After overnight incubation in the thermocycler, the amplified DNA was run through a 1% agarose gel with ethidium bromide to isolate the amplified sequence. CMV-APEX was amplified along with two negative controls, CMV-APEX (-MM), and CMV-APEX (-Primers). The PCR products were imaged using a gel imager and observed bands for the amplified sample at 777 base pairs and noticed no bands present for both of the negative controls. The size of the observed band, coupled with the lack of bands present on the negative controls indicated that the PCR reaction was successful. Therefore, the product was used for subsequent Gibson assembly.

Gibson Assembly

We used Gibson assembly to join the digested Plv-Tet-On plasmid and the PCR amplified APEX discussed in the previous section. Doing so will allow us to create a circularized plasmid containing the features of our Tet-On plasmid and APEX2. The digested PLV-TET-ONE (vector) and APEX2 (insert) were used at a 2:1 ratio of vector to insert along with NEB 5x Gibson Mastermix. A negative control was established without any insert as seen in **figure 10**. The experimental control and negative control were transformed using E. Coli and plated onto LB-Agar plates supplemented with ampicillin to select for bacterial colonies that took up the plasmid. After a 24-hour incubation at 37°C, colonies were seen on the experimental agar plate suggesting that the cloning process was successful, but there were no colonies present on the negative control. To further verify if the ligated plasmid maintained the desired amino acid

sequence, samples were sent out for sequencing. Colonies from the agar plate were selected for propagation in liquid LB cultures overnight at 37°C. We then used the Monarch DNA Purification Kit to extract the DNA and used a nanodrop to record the concentrations; Sample A (169ng/ul), Sample B (254ng/ul), Sample C (209ng/ul), Sample D (245ng/ul). These samples will be used for future quantitative analysis and transfection.

	(+)Plv-Tet-On-APEX	(-)Primers
Gibson 5x Master Mix	10ul	10ul
APEX2	2.3ul	0ul
Plv-X	3.6ul	3.6ul
Water	4.1ul	6.4ul

Figure 10. *Gibson Assembly for Plv-Tet-On-APEX Measurements*

Designing the Negative Control p53-TET-ONE Plasmid (P53 Alone)

Polymerase Chain Reaction Amplification

In order to clone our p53 segment into the Plv-Tet-One plasmid to create the p53-Tet-On plasmid, we used a polymerase chain reaction amplification. This allowed us to isolate and amplify our desired segment to use for subsequent Gibson assembly. The primers for the reaction were designed using SnapGene; primer SL1635 that was complementary to the sequence upstream of p53 and the other 30 base pairs downstream of p53 (see **figure 20)**. After overnight incubation in the thermocycler, the amplified DNA was run through a 1% agarose gel with ethidium bromide to isolate the amplified sequence. Plasmid CMV-APEX-p53 was amplified along with two negative controls, CMV-APEX-p53 (-MM), and CMV-APEX-p53 (-Primers). The PCR products were imaged using a gel imager and observed bands for the amplified sample at 2000 base pairs and noticed no bands present for both of the negative

controls. The size of the observed band, coupled with the lack of bands present on the negative controls indicated that the PCR reaction was successful. This reaction product was then used for quantitative analysis and transfection.

Transient Transfection and Doxycycline Induction

In order to test the functionality of the newly designed plasmids, we used a transient transfection to insert our plasmids into HCT116 p53-/- (human colon carcinoma cells). We used HCT116 *TP53*$^{-/-}$cells because they do not express p53 mRNA or protein as the gene was knocked out via homologous recombination on both alleles. This allows us to know for sure that any p53 transcribed or translated is in fact from our plasmids and not inherently produced in the cell.

To measure the levels of transcription after doxycycline induction, we used qPCR analysis. We used primers SL67 and SL68 to track p53 transcriptional activity by annealing to the sequence of p53 mRNA, as plasmids p53-APEX2-TET-ONE, APEX2-p53-TET-ONE, and p53-TET-ONE should all exhibit an increase in p53 transcriptional activity when induced with doxycycline. Primers SL1781 and SL1782 were used to select for APEX2 transcriptional activity expected in plasmids p53-APEX2-TET-ONE, APEX2-p53-TET-ONE, and APEX2-TET-ONE. We also used primers SL57 and SL58 for P21, a p53 target gene, which should be more highly expressed if p53 levels increase. Lastly, primers SL29 and SL30 were used to measure levels of GAPDH to normalize the results (See **Figure 11.** for list of primer names and sequences).

The results show that when induced by doxycycline plasmid p53-APEX2-TET-ONE did not exhibit any notable levels of p53 activity or APEX2 activity, and similarly, no P21 activity was observed. This suggests that the positioning of p53 on the C-terminus of APEX2 might interfere with p53s binding ability or cause p53 to not be expressed at all. For plasmid APEX2-

P53-TET-ONE, there was a slight change seen in p53, APEX2, and P21 activity when doxycycline was added which is consistent with the specifications of the plasmid (see **figures 15, 16, and 17**). p53-TET-ONE showed p53 and P21 activity but also exhibited low levels of APEX2 activity when induced. This is not consistent with the plasmid as it does not contain APEX2. APEX2-TET-ONE demonstrated slight p53 activity and APEX2 activity, but a negligible amount of P21 activity. Comparing these results to the non-induced plasmids, there appears to be some inconsistency. In most cases, the plasmids that were not treated with doxycycline were shown to exhibit higher levels of transcription of p53, APEX2, and P21. This could suggest that the plasmids created do not function as intended and perhaps that a p53-APEX2 fusion would not work within a TET-ON inducible system, but evidence points to a different conclusion. Levels of APEX2 transcription were seen to be highest in the untreated p53-TET-ONE plasmid, however, this plasmid contains no coding region for APEX2 and the HCT116 cells are also void of any APEX2. Similarly, for the non-induced APEX2-TET-ONE plasmid, we observed large amounts of p53 transcriptional activity. With this in mind, it is likely that there was DNA contamination present in some of the samples used for qPCR analysis.

Primer Name:	Primer Sequence:
SL67 p53 fwd	CCAGAAAACCTACCAGGGCA
SL68 p53 rev	GAATGCAAGAAGCCCAGACG
SL1781 APEX2 fwd	CTTCCTCCCTGAGTTTGCAG
SL1782 APEX2 rev	GCTAGAGCCAACCTGACTGG
SL57 P21 fwd	GCAGAGGAAGACCATGTGGAC
SL58 P21 rev	GCGAGGCACAAGGGTACAAG
SL29 GAP fwd	TCCTGGGCTACACCGATGA

SL30 GAP rev	GGCGTCGAACACAGACGAAT

Figure 11. *Primer Names and Sequences for qPCR Analysis*

Discussion

We aimed to create a doxycycline-inducible p53-APEX2 fusion protein to identify p53-bound enhancer elements. In order to do so, we needed to find a way to ligate our p53 protein to APEX2 and furthermore, insert those ligated proteins into the PLV-TET-ONE vector to allow for controlled doxycycline induction of our protein fusion. Using laboratory techniques such as polymerase chain amplification and Gibson assembly, we were able to create two functional experimental plasmids (p53-APEX2-TET-ONE and APEX2-p53-TET-ONE) as well as two negative controls (APEX2-TET-ONE and p53-TET-ONE). To test whether or not these plasmids functioned as intended, we transfected them into HCT116 p53-Null Cells and induced them with doxycycline. We needed to ensure that the proper protein products were being produced by each of the corresponding plasmids, so we used a quantitative polymerase chain reaction to measure the levels of transcription after induction. It was not possible to make an assessment as to the functionality of the plasmids as they did not exhibit the patterns of transcription we expected, and therefore the data was inconclusive. Due to the time constraints created by the Covid-19 pandemic, we were not able to attempt to redo the qPCR experiments.

p53 binding events can be grouped into 3 distinct classes. The first class of p53 binding occurs near transcriptional start sites (TSS) and is the area where promoter-associated chromatin modification occurs. The second class of p53 binding occurs at promoter-distal enhancer elements that demonstrates p53s ability to initiate transcription and induce histone modification like acetylation. The third class of p53 binding sites is absent of normal chromatin modifications and falls within regions of inaccessible chromatin (Sammons et al.) Typically, in order for transcription factors to bind to their regulatory regions, nucleosomes need to be actively removed

and outcompeted by transcription factors. However, it has been hypothesized that special pioneer transcription factors have the ability to interact and bind to compacted and closed-off areas of chromatin. p53 is able to exhibit such behavior making it a solid candidate to be a pioneer transcription factor. p53 binding to chromatin has been extensively studied using chromatin immunoprecipitation-based techniques like Chip-PET, and CHIP-seq. These studies revealed that over 50% of p53 binding sites are on nucleosome-rich regions of DNA as well as demonstrated that p53 is able to override these areas of chromatin without the use of any auxiliary transcription factors (Sullivan et al. 2018).

The plasmids we created allow us to explore the dynamics of p53 binding and its unique pioneer factor activity. The experimental APEX2-p53-TET-ONE and p53-APEX2-TET-ONE plasmids give us a good foundation to further our research on p53 (see **figure 12**). The next step is to use the created pLV-Tet-One plasmid variants to produce a lentivirus that will allow us to make cell lines that stably integrate our gene array of interest. Therefore, these cells would then be able to produce the gene of interest (the respective p53 APEX fusions and + and - controls) upon the addition of doxycycline. We believe this will allow us to monitor what proteins p53 interacts with when bound to DNA by proximity labeling via APEX2. However, within the nucleus of the cell, there are hundreds if not thousands of proteins present and it is important to distinguish which proteins are associated with p53 and which are irrelevant to our studies. So how will we distinguish what is relevant and what is not? The negative controls we created allow us to remedy this problem.

The first negative control we designed was the TET-ON-APEX plasmid (APEX alone) depicted in **figure 13**. This plasmid has all the features of the TET-ONE plasmid and APEX is free of any target protein. We designed this plasmid to test APEX2's functionality when used in

a Tet-On inducible system and ensure that it properly biotinylates protein targets, but it also provides insights on p53. Once transfected into a cell, the TET-ON-APEX plasmid would only be present in the cytoplasm of the host cell, and without any protein fused to APEX2, it lacks the instructions to be directed to the nucleus. We can then induce APEX2 and biotinylate any proteins and RNA. This will demonstrate that APEX2 is capable of biotinylation as well as identify any proteins and RNA not relevant to the experiment. Most of the RNA and proteins that APEX2 alone would biotinylate would be localized in the cytoplasm, but there is a chance that APEX2 could be directed into the nucleus and label proteins there as well.

The next negative control we designed was TET-ONE-p53 (p53 alone) shown in **figure 14.** This plasmid allows us to ensure that p53 can be expressed at a normal level when induced by doxycycline in the Tet-On system. Furthermore, we need to confirm that the APEX2-p53 or p53-APEX2 fusions we created are acting as wild-type p53 does. We can do that via qPCR of known p53 transcriptional targets (P21) by expressing wild-type p53 and seeing how the APEX2 fusions compare. To explore this further, we can design another control using site-directed mutagenesis. As we know, p53's ability to bind to DNA is integral for its function in the stress response pathway and recruiting transcriptional machinery, but with a single amino acid substitution mutation to the central DNA-binding domain of p53, we can remove its ability to bind to DNA. We can then fuse it to APEX2 and clone it into our Tet-One plasmid. While this plasmid would be directed to the nucleus of the cell because of the inclusion of the p53 gene, it would not be able to function the same as the wild-type version of p53. It would lose the ability to bind to DNA and with that comes the loss of ability to recruit target genes and other transcriptional elements. Once induced by doxycycline, APEX2 will be able to tag proteins and RNA present in the nucleus, but any enhancer elements that are specific to p53 binding to DNA

will be excluded. Using this control allows us to narrow down what proteins are unique to wild-type DNA binding p53 and guide our future experiments. Furthermore, this control would provide us with potential p53 interactions that occur off of DNA but are still localized in the nucleus as well as general off-target biotinylation. Anything that p53 might interact with randomly within the nucleus will be tagged and coupled with the APEX2 alone control, while it may not always be present in the nucleus, will allow us to filter protein and RNA interactions that are novel to p53.

The results from the qPCR analysis provided us with data that was inconclusive as to the functionality of our plasmids. As discussed previously in the results section, it is clear that DNA contamination was responsible for the uncharacteristically high levels of expression seen for the non-induced plasmids like APEX2-TET-ONE, and p53-TET-ONE. For future iterations of this experiment, a larger focus should be placed on the DNA purification step during RNA extraction to ensure that DNA contamination is not a major factor in the overall results. However, the data gathered, despite its flaws, is not without merit. Our p53-APEX-TET-ONE plasmid displayed no levels of transcriptional activity for p53, P21, and APEX2, yet our APEX-p53-TET-ONE plasmid performed much better, demonstrating activity for all three. This suggests that when p53 is fused to the C-terminus of APEX, its ability to function properly and exhibit normal levels of transcription are inhibited and when fused to the N-terminus of APEX its functionality seems to be unchanged. This has to be further tested with more qPCR experimentation, but it is likely that p53s functionality might be negatively affected by the addition of APEX2.

APEX2 labeling is not without its share of shortcomings. A transcription factor such as p53 interacts with hundreds of different proteins and through the use of APEX2 we are able to know what proteins they are, but protein proximity labeling through APEX2 is unable to provide us with information as to where p53 is binding and where these protein interactions are occurring. This is a problem for understanding enhancer/promoter/transcriptional activation by p53, as TFs and cofactors often work together at particular places and not others. So if p53 is seen to interact with protein X and protein Y, it is impossible for us to determine whether they interact at the same place on DNA or at two different genes.

Techniques such as CASPEX can be used to answer the questions that APEX2 protein proximity labeling leaves us with. In short, CASPEX allows us to fuse a catalytically dead RNA-guided nuclease dCas9 to APEX2 and target specific genomic regions by using single guide RNAs (sgRNAs). This allows us to target APEX2 to known p53 binding locations and label nearby proteins. Compared to the APEX2-p53 fusion, results from CASPEX would be much more specialized because we would know that any biotinylated proteins came from the location determined by the guide RNA we put in. While this technique comes with advantages, such as knowing what elements p53 interacts with at a specific location, it is technically something not possible with a TF that binds many places like p53. APEX2-p53, while unable to determine the origin of biotinylated proteins, allows us to study the extensive network of p53 more efficiently and provide us with a larger data pool of novel p53 protein interactions.

APEX2-p53 can also be utilized to identify locations where p53 is bound in the genome. We would first induce APEX2 in the cell and biotinylate everything, including histones. Since the DNA wrapped around the nucleosome is protected from digestion by endo-exonucleases, we can use Micrococcal nuclease (MNase) to digest all DNA into mononucleosomes. We could then

immunoprecipitate the biotinylated mononucleosomes (pull them down) via streptavidin-coupled beads and treat them with proteinase which would digest all proteins, including histones, and leave behind just the DNA fragment. We could then sequence the remaining DNA, as it represents where the nucleosomes were in the genome, and any region that was biotinylated would indicate where p53 was.

We propose a new technique for studying the p53 stress response pathway and protein-protein interactions through the use of an APEX2-p53 fusion protein. We designed a novel plasmid that allows for doxycycline-inducible biotinylation of p53 enhancer elements. Our results demonstrate that our plasmid is capable of expressing p53 and APEX2 when transfected into HCT116 cells and induced with doxycycline. We believe that APEX2-mediated biotinylation may allow for a simple and cost-effective way to proximity label proteins, RNA species, and identify DNA binding protein locations in the genome when compared to alternative methods like BioID and Gro-Seq. Further experiments can and should be done to explore the capabilities of our APEX2-p53 fusion protein and the full extent of the effectiveness of this plasmid is yet to be seen. Essentially, this tool can be used to push forward our understanding of how the p53 network regulates gene transcription and the complex dynamics of the p53 stress response pathway.

References

Sears, R. M., May, D. G., & Roux, K. J. (2019). BioID as a Tool for Protein-Proximity Labeling in Living Cells. Methods in molecular biology (Clifton, N.J.), 2012, 299–313. https://doi.org/10.1007/978-1-4939-9546-2_15

Sammons, M. A., Nguyen, T.-A. T., McDade, S. S., & Fischer, M. (2020). Tumor suppressor p53: from engaging DNA to target gene regulation. Nucleic Acids Research, 48(16), 8848–8869. https://doi.org/10.1093/nar/gkaa666

Spitz, F., & Furlong, E. E. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, *13*(9), 613–626. https://doi.org/10.1038/nrg3207

Hung V, Udeshi ND, Lam SS, Loh KH, Cox KJ, Pedram K, Carr SA & Ting AY (2016) Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. Nat Protoc 11, 456–475

Kim, D. I., & Roux, K. J. (2016). Filling the void: Proximity-based labeling of proteins in living cells. Trends in Cell Biology, 26(11), 804–817. https://doi.org/10.1016/j.tcb.2016.09.004

Lam, S., Martell, J., Kamer, K. et al. Directed evolution of APEX2 for electron microscopy and proximity labeling. Nat Methods 12, 51–54 (2015). https://doi.org/10.1038/nmeth.3179

Di Como, C. J., Gaiddon, C., Prives, C. (1999). P73 function is inhibited by tumor-derived p53 mutants in mammalian cells. Molecular and Cellular Biology, 19(2), 1438–1449. https://doi.org/10.1128/mcb.19.2.1438

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: Five essential questions. Nature Reviews Genetics, 14(4), 288–295. https://doi.org/10.1038/nrg3458

Pietsch, E. C., Sykes, S. M., McMahon, S. B., & Murphy, M. E. (2008). The P53 family and programmed cell death. Oncogene, 27(50), 6507–6521. https://doi.org/10.1038/onc.2008.315

Roux, K. J., Kim, D. I., Burke, B. (2013). BioID: A screen for protein-protein interactions. Current Protocols in Protein Science, 74(1). https://doi.org/10.1002/0471140864.ps1923s74

Padrón, A., Iwasaki, S., & Ingolia, N. T. (2019). Proximity RNA labeling by Apex-Seq reveals the organization of translation initiation complexes and repressive RNA granules. Molecular Cell, 75(4). https://doi.org/10.1016/j.molcel.2019.07.030

Papavassiliou, A. G. (1995). Transcription factors. *New England Journal of Medicine*, *332*(1), 45–47. https://doi.org/10.1056/nejm199501053320108

Yang, A., Schweitzer, R., Sun, D., Kaghad, M., Walker, N., Bronson, R. T., Tabin, C., Sharpe, A., Caput, D., Crum, C., & McKeon, F. (1999). P63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. Nature, 398(6729), 714–718. https://doi.org/10.1038/19539

Yang, A., Walker, N., Bronson, R., Kaghad, M., Oosterwegel, M., Bonnin, J., Vagner, C., Bonnet, H., Dikkes, P., Sharpe, A., McKeon, F., & Caput, D. (2000). P73-deficient mice have neurological, pheromonal, and inflammatory defects but lack spontaneous tumours. Nature, 404(6773), 99–103. https://doi.org/10.1038/35003607

Kleinjan, D. A., & van Heyningen, V. (2005). Long-range control of gene expression: Emerging mechanisms and disruption in disease. *The American Journal of Human Genetics*, *76*(1), 8–32. https://doi.org/10.1086/426833

Belyi, V. A., Ak, P., Markert, E., Wang, H., Hu, W., Puzio-Kuter, A., & Levine, A. J. (2009). The origins and evolution of the p53 family of genes. Cold Spring Harbor Perspectives in Biology, 2(6). https://doi.org/10.1101/cshperspect.a001198

Muller, P. A., & Vousden, K. H. (2013). P53 mutations in cancer. *Nature Cell Biology*, *15*(1), 2–8. https://doi.org/10.1038/ncb2641

Spitz, F., & Furlong, E. E. (2012). Transcription factors: From enhancer binding to developmental control. Nature Reviews Genetics, 13(9), 613–626. https://doi.org/10.1038/nrg3207

Davidson, E. H. (2010). Emerging properties of animal gene regulatory networks. Nature, 468(7326), 911–920. https://doi.org/10.1038/nature09645

MacQuarrie, K. L., Fong, A. P., Morse, R. H., & Tapscott, S. J. (2011). Genome-wide transcription factor binding: Beyond Direct Target Regulation. Trends in Genetics, 27(4), 141–148. https://doi.org/10.1016/j.tig.2011.01.001

Oren, M. (2003). Decision making by p53: Life, death, and cancer. *Cell Death & Differentiation*, *10*(4), 431–442. https://doi.org/10.1038/sj.cdd.4401183

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2016). Breaking rad: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*(2), 142–152. https://doi.org/10.1111/1755-0998.12635

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: Insights into functions. Nature Reviews Genetics, 10(3), 155–159. https://doi.org/10.1038/nrg2521

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: Five essential questions. Nature Reviews Genetics, 14(4), 288–295. https://doi.org/10.1038/nrg3458

Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: Establishing competence for gene expression. Genes & Development, 25(21), 2227–2241. https://doi.org/10.1101/gad.176826.111

Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., & Ting, A. Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science, 339(6125), 1328–1331. https://doi.org/10.1126/science.1230593

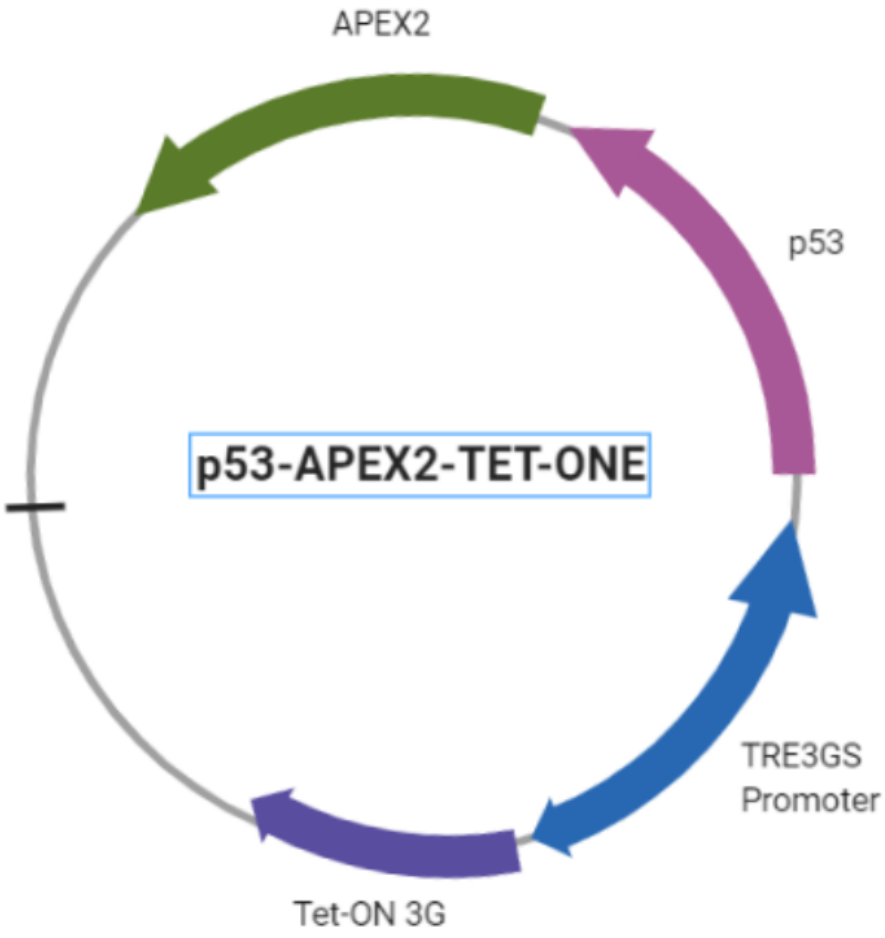

Figure 12. Visual representation of the p53-APEX2-TET-ONE plasmid including the TET-ON 3G regulation system and the TRE3G3 promoter

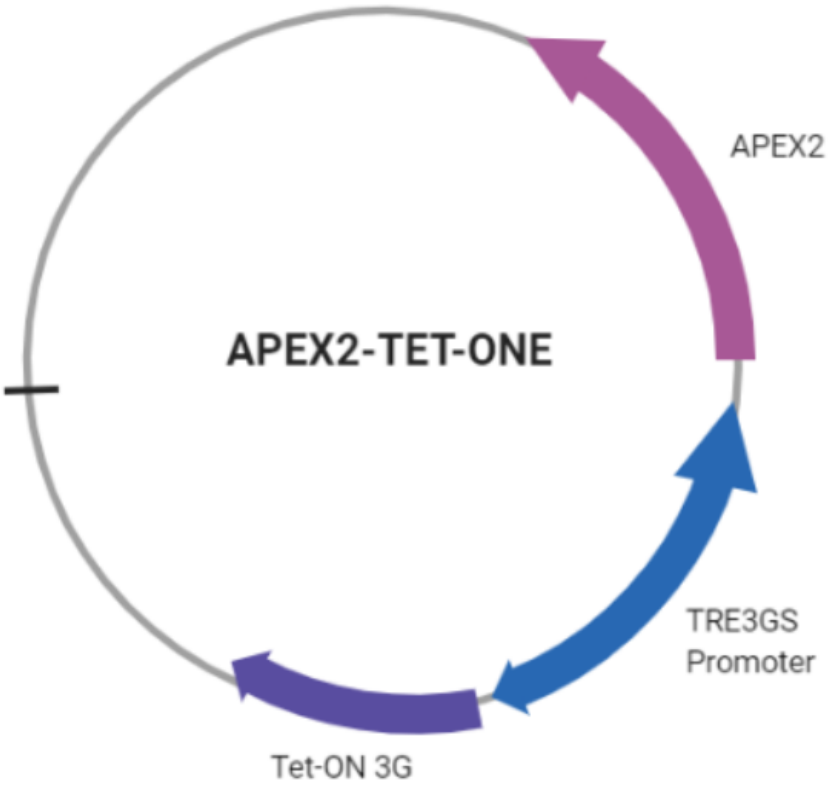

Figure 13. Visual representation of the APEX2-TET-ONE plasmid (APEX Alone) including the TET-ON 3G regulation system and the TRE3G3 promoter

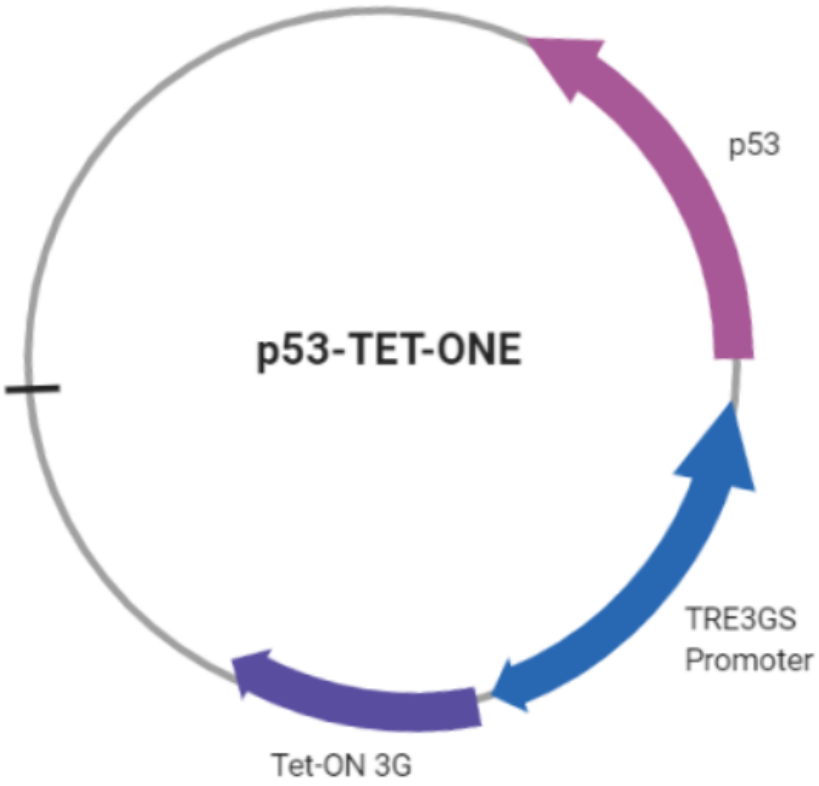

Figure 14. Visual representation of the p53-TET-ONE plasmid (p53 Alone) including the TET-ON 3G regulation system and the TRE3G3 promoter

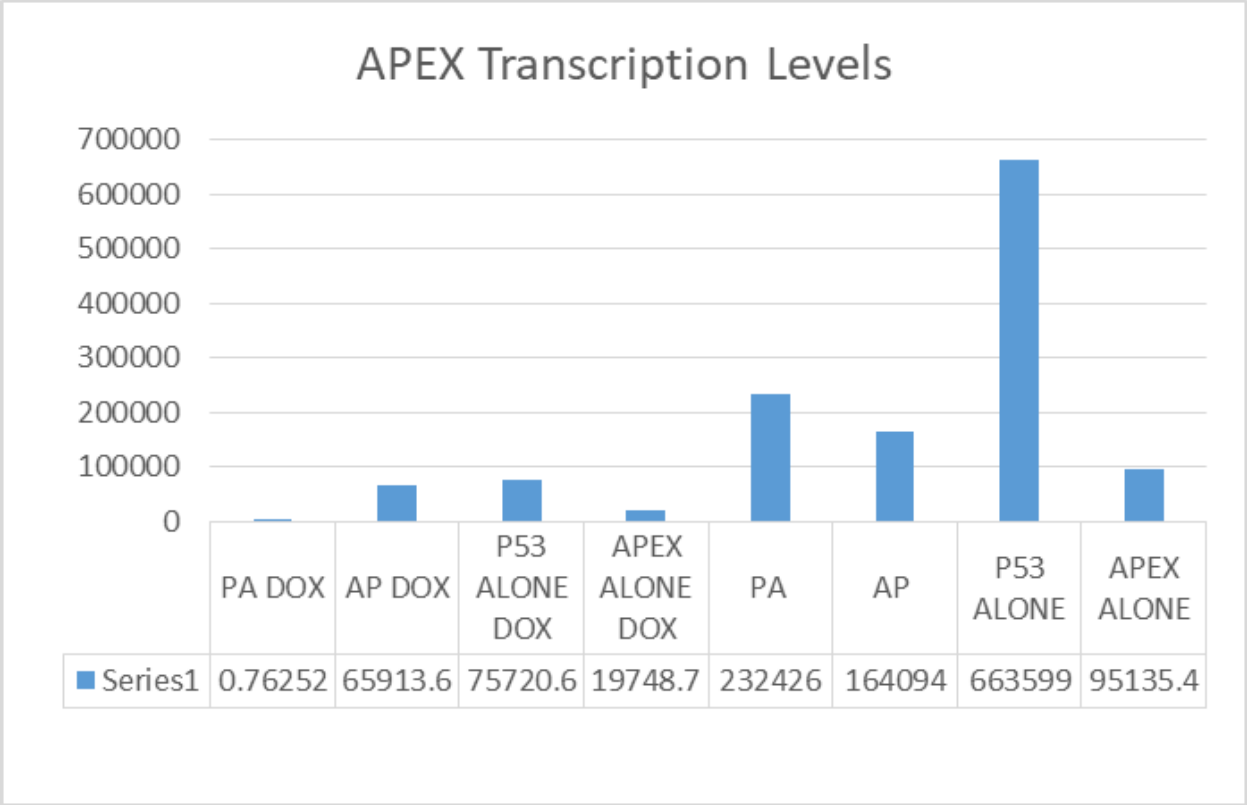

Figure 15. qPCR of p53-APEX2, APEX2-p53, p53 Alone, and APEX Alone
qPCR was performed on the designed plasmid to track for transcription levels of APEX. We compared the rates of doxycycline induced plasmids to non-induced plasmids to test plasmid functionality.

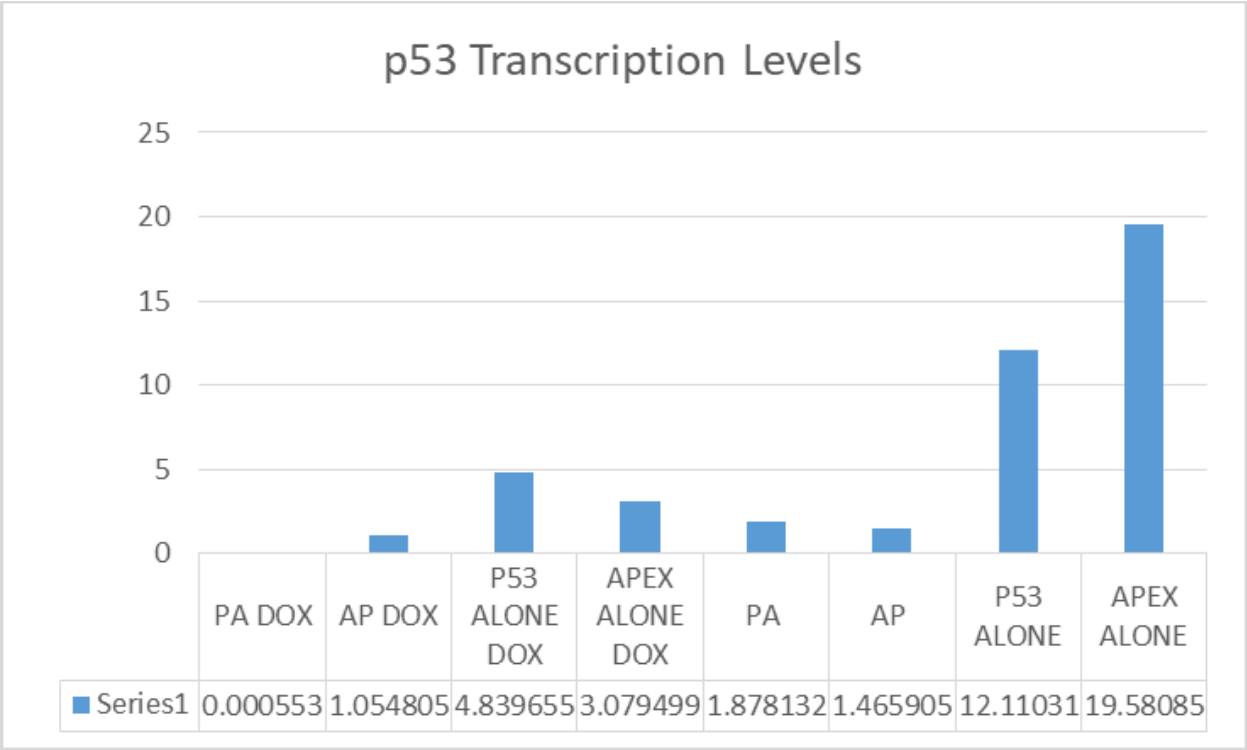

Figure 16. qPCR of p53-APEX2, APEX2-p53, p53 Alone, and APEX Alone
qPCR was performed on the designed plasmid to track for transcription levels of p53. We compared the rates of doxycycline induced plasmids to non-induced plasmids to test plasmid functionality.

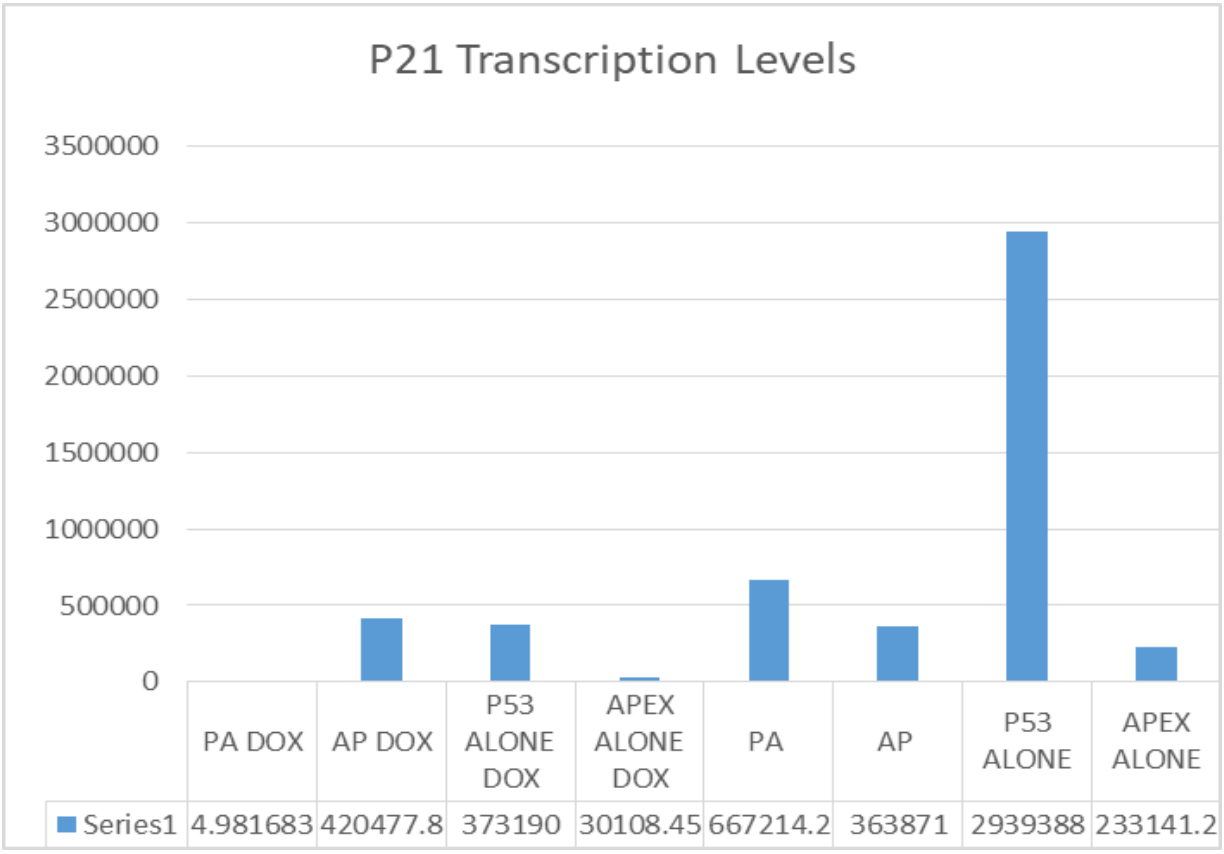

Figure 17. qPCR of p53-APEX2, APEX2-p53, p53 Alone, and APEX Alone
qPCR was performed on the designed plasmid to track for transcription levels of P21, a p53 target gene. We compared the rates of doxycycline induced plasmids to non-induced plasmids to test plasmid functionality.

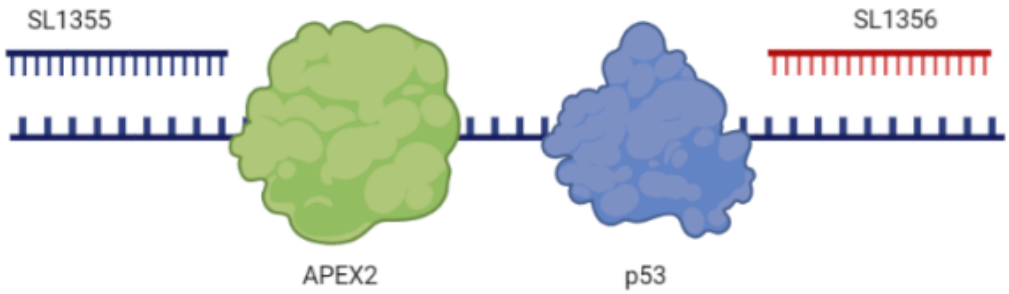

Figure 18. Visualization of PCR amplification of the APEX2-p53 segment using designed primers SL1355 and SL1356

The primers for the reaction were designed using SnapGene; primer SL1355 complementary to the sequence upstream of APEX-p53 and primer SL1356 30 base pairs downstream of APEX-p53

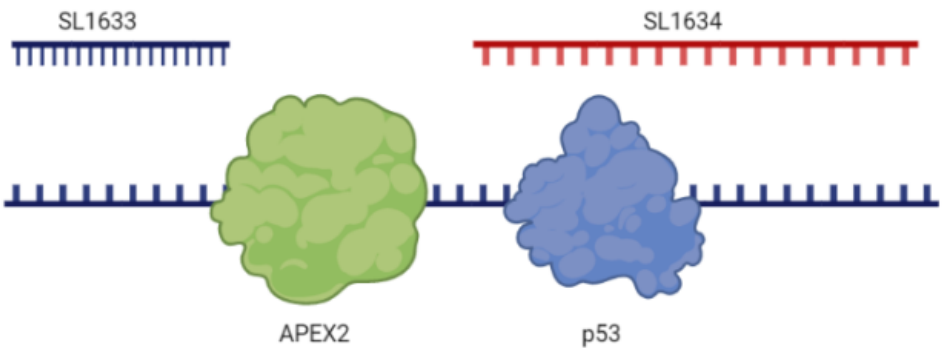

Figure 19. Visualization of PCR amplification of the APEX2 segment using designed primers SL1633 and SL1634

The primers for the reaction were designed using SnapGene; primer SL1633 complementary to the sequence upstream of APEX and primer SL1634 30 base pairs downstream of APEX

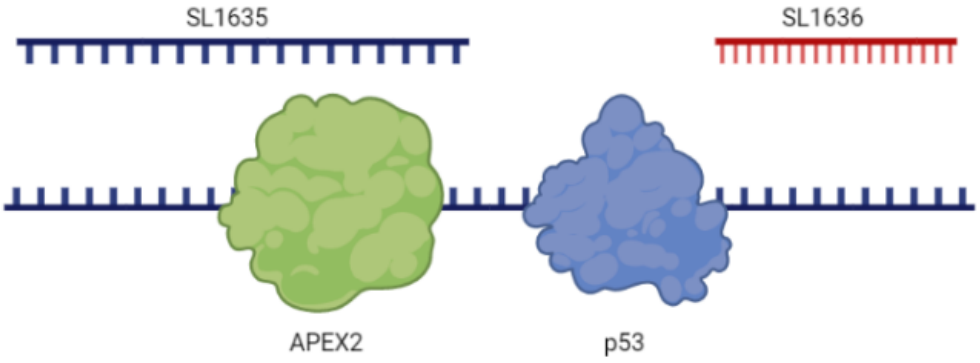

Figure 20. Visualization of PCR amplification of the p53 segment using designed primers SL1635 and SL1636

The primers for the reaction were designed using SnapGene; primer SL1635 that was complementary to the sequence upstream of p53 and primer SL1636 30 base pairs downstream of p53.