University at Albany, State University of New York

Scholars Archive

Biological Sciences   Honors College

Spring 5-2022

Emerging Roles of Intragenic Enhancers in Gene Expression of GADD45A

Ekaterina Sazon

Emerging Roles of Intragenic Enhancers in Gene Expression of *GADD45A*

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

Ekaterina Sazon

Research Mentor: Morgan Sammons, Ph.D.
Second Reader: ChangHwan Lee, Ph.D.

May 2022

Abstract

Non-coding regulatory elements in the eukaryotic genomic DNA, like enhancers and promoters, control gene expression. Transcription factors recognize and bind to DNA sequences within these regulatory regions and affect said regulatory elements. The extent and specific circumstances regarding transcription factors are long-standing questions in the field. In this thesis, we wanted to investigate how transcription factors work in a particular case of enhancers, known as intragenic enhancers. We designed reporter gene assays that looked at the well-studied model transcription factor p53's regulation on its target gene *GADD45A* in basal and stress conditions. We determined that *GADD45A* is a p53-dependent gene, requiring the factor for maximum transcriptional output. Furthermore, we found that p53 requires additional transcription factors to attain the result, suggesting that intragenic enhancers behave similarly to traditional upstream enhancers. These novel observations highlight the diversity of regulatory elements within the genome and showcase the combinatory nature of transcription factors. Further work regarding the dynamics and function of transcription factors within intragenic enhancers will lead to a better understanding of gene regulatory biology.

Keywords: *Enhancer, Transcription factors, Cell stress, Regulatory gene expression*

Acknowledgments

I have had a great team behind me while researching and writing my thesis. I would first like to thank Dr. Morgan Sammons for allowing me to conduct research in his lab throughout my four years of college. With his patience, guidance, and steady support, I learned how to think critically and explore the immense research world. Dr. Morgan Sammons served far more than a research advisor and a mentor throughout my undergraduate education. He has been a significant influence on my continued pursuit of scholarly activities.

In addition to my advisor, I want to thank my colleagues within the Sammons lab, including Gaby, Baniulyte, Serene Durham, Dana Woodstock, Jhos Duffrt, Abby McCann, and Dan Koskas, as well as former members Ally Catizone, Aparna Prasad, and Zenab Minahas. They have always been there for me, including answering questions, challenging my thinking, and expanding my knowledge. I owe much of my success to their endless patience and encouragement.

And finally, thanks to my parents for 22 years of nurturing and having my back. Without them, I would have never enjoyed all the opportunities I have experienced.

List of Figures

List of Tables

Table of Contents

Introduction

Organisms are exposed to environmental stressors that threaten their cellular homeostasis: a steady internal state regarding physical and chemical conditions (Gordan et al., 2017). Stressors like mechanical damage, toxin exposure, and deprivation of crucial building blocks and nutrients, among others, evoke a cellular stress response. This response intricately regulates numerous biochemical processes to restore organismal and cellular homeostasis through repair and stabilization (Fulda et al., 2010). Alternatively, if the damage or cost to restore homeostasis is deemed too energetically costly, cells may undertake a programmed cell death. Both options aim for the same goal of returning an organism to its preferred homeostatic state. The homeostatic restoration option includes control of crucial cell cycle checkpoints, changes to energy utilization, and synthesis and capture of essential cellular building blocks, like nucleotides and amino acids. However, if the stressor causes damage that the cell cannot treat in a timely and energy-efficient matter, programmed cell death pathways eliminate the damaged cell. These pathways ensure that elimination is done orderly with defective cellular contents packed for safe degradation via immune cells (Vaux & Strasser, 1996). Unrepairable cells can also be placed into senescence, a form of permanent growth arrest via removal from the cell cycle to prevent the proliferation of defective cells. All options aim for the same goal of returning an organism to its preferred homeostatic state.
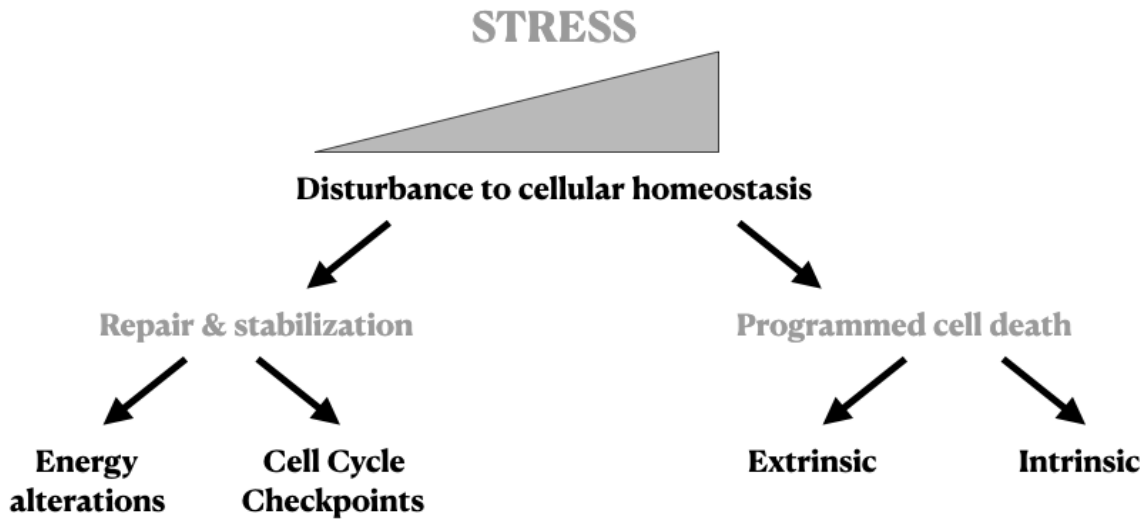
Figure 1. Generic chart depicting cellular stress responses

Note. Environmental stress causes a disturbance to cellular homeostasis, invoking a cellular stress response. The two major response types include repair/stabilization and programmed cell death. The former involves actions like energy alterations and cell cycle checkpoints, while the latter includes extrinsic and intrinsic cell death. Both aim to restore cellular homeostasis.

As noted previously, cellular stress can come in many forms, so it's no surprise that cells have evolved to have many different stress responses. Stress responses can be generic or tailored to combat a particular stress signal. In both cases, the cellular stress response often occurs at the level of gene regulation. The transcription and translation of specific genes are altered to aid in stress reduction while repressing genes that may inhibit the desired cellular or organismal outcome. Cells can also rewire their metabolic pathways to ensure adequate energy production needed for protein synthesis. Metabolic pathways can also downregulate specific processes to conserve energy required for more critical regions (Harding et al., 2003). Using this newly allotted energy, cells can start making necessary proteins and other building blocks. They need raw biochemical materials that they can acquire by rebalancing anabolic and catabolic pathways. Lastly, newly synthesized proteins need to be folded via chaperones. This process can be upregulated or downregulated to ensure a timely and efficient cellular stress response.

Cells are introduced to many different types of stress, ranging in severity levels. Each cell must decide how to deal with said stress with the end goal of returning to homeostasis or degrading down, with each necessitating different cellular strategies. These decisions are ultimately dictated by information encoded in the genome of that organism and based on the state of the cell during stress exposure. These differential life or death choices are due to differential molecular responses or cell states, such as differential gene regulation and processes that control it. For example, two cells with the same lineage/identity may differentially respond to an identical stress exposure based on the state of the cell cycle. This seemingly random but well-described phenomenon is still not well understood. Thus, research done in our lab focuses on how these differential stress responses may be dictated via stress-activated regulatory networks that share common cascades and components. These networks are engaged through

3

stress stimuli that activate transcription factors (TF), proteins that regulate gene expression via binding to DNA-encoded regulatory elements. Those genes can then influence downstream activities such as cell cycle arrest, programmed cell death, and previously mentioned pro-survival or pro-apoptotic activities.
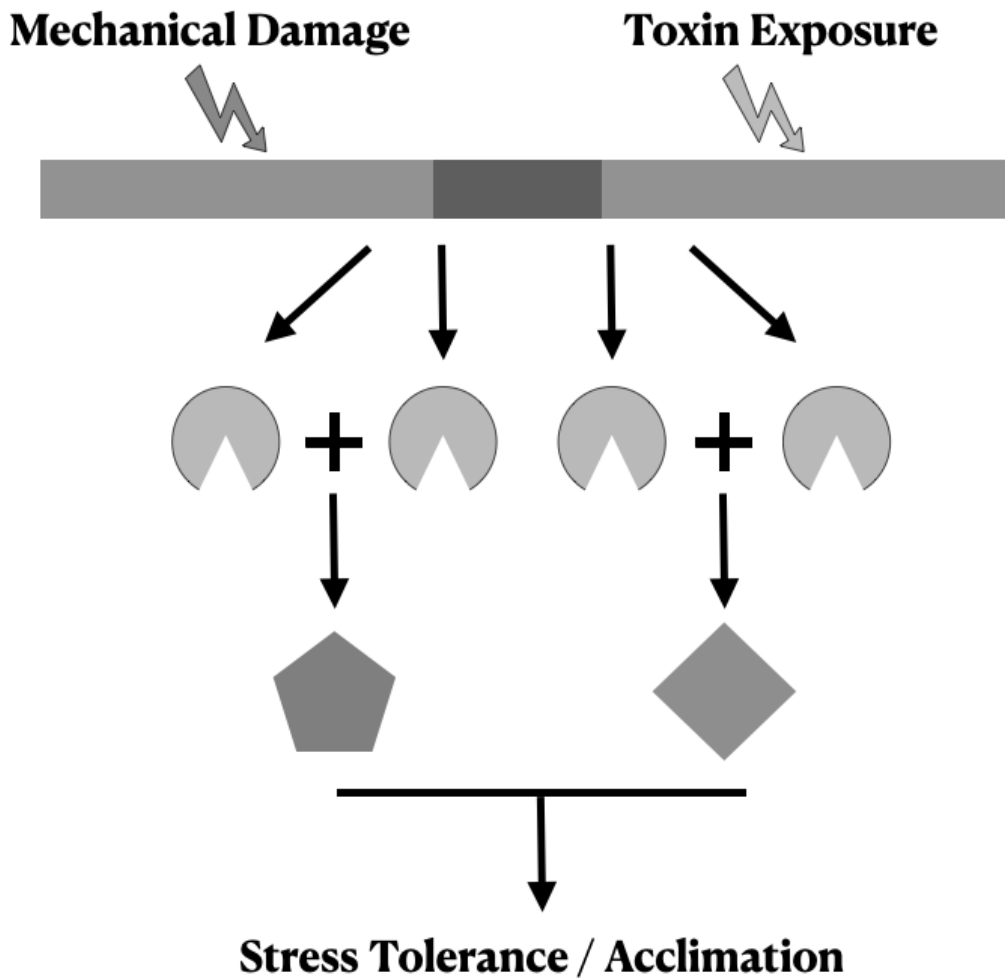
Figure 2. Graphic of stress regulatory pathways

Note. Stress regulatory pathways are activated through different environmental stressors, including mechanical damage and toxin exposure. They induce gene expression resulting in the end goal of stress tolerance and acclimation. It is believed that stress regulatory pathways share standard components and processes, resulting in similar end goals despite different stimuli.

While we suggest that these stress-activated networks are vital regulators of how cells mitigate stress, many questions remain, including what TFs are present, what roles they play, and their combination in response to the different stress pathways. These questions are the focus of this thesis.

Understanding these stress-activated networks starts with the most foundational unit: the deoxyribonucleic acid (DNA) molecules that make up an organism's genome. A DNA molecule contains sequences of nucleotides that act as a blueprint for all building blocks, including ones needed for stress response. As a result, the cell has many DNA molecules, collectively known as a genome. The average human genome is on the order of 3.2 billion individual nucleotides arranged specifically. It is no surprise that a cell would not want to comb through these sequences when it wants a specific protein. Thus, cells have evolved to quickly decode information embedded in the genome through TF proteins that "read" the DNA sequence. These factors survey the genome, looking for sequences they bind to when found. While the ribosome, the protein-making machinery, is not directly dependent on TFs, it utilizes TF's ability to decode DNA to allow RNA polymerase to make RNA. The ribosome can then use this RNA to make the protein.

TFs bind to three general DNA locations when controlling gene expression: promoters, enhancers, and silencers. All are considered cis-regulatory elements (CRE), which regulate transcription levels of a specific gene. CREs can regulate across large distances and simultaneously regulate one or many y genes. CREs can also be located within any genomic context, including introns, exons, and intergenic regions.

Promoters are CREs located directly upstream of the gene and are vital for transcriptional initiation. Promoters contain a transcriptional start site (TSS), binding motifs for general transcription factors, and RNA polymerase complexes' landing pad. RNA polymerase is the multi-subunit machine that reads the DNA sequence and synthesizes an RNA molecule complementary to the DNA sequence. As stated earlier, promoters contain multiple unique locations for TF binding. These factors can aid RNA polymerase recruitment, facilitate

permissive or repressive chromatin structures, or recruit other transcriptional regulator proteins and cofactors. As a result, TFs can broadly affect the downstream transcriptional activity of genes at various stages of the transcription cycle. Thus, mutations in these types of CREs can lead to hazardous consequences. Several studies have identified disease-causing regulatory mutations (Wray, 2007). Promoter mutations cause around 1% of all single-base substitution-caused disorders. Mutations in these regulations disrupt normal gene activation and transcription, resulting in either a decrease or increase in mRNA production (Cooper, 2002). Mutations can also remove the binding motifs required for protein factors like transcription factors (Cooper, 2002). Some known promoter-mutation-caused disorders include Bernard-Soulier syndrome, familial hypercholesterolemia, and hemophilia, leading to adverse health effects (Maston et al., 2006).

Enhancers are not a required element, with many genes not containing them at all. While enhancers are non-mandatory CREs; they can be required for certain types/modes of regulation. Promoter transcription is minimal without some contribution by these usually distal CREs. These position-independent CREs regulate transcription directly communicating with either singular or multiple promoters. This communication is due to mechanisms like chromatin looping, facilitating pre-initiation complex formation. Histone modification and other cofactors enable the construction of these loops by reorganizing the chromatin of these CREs. Looping also allows distant enhancers to regulate promoters hundreds of kilobases away. Mutations in enhancers are limited to *cis* effects on transcription (Sauna & Kimchi-Sarfaty, 2011). *Cis* effects are genetic expression effects caused by genetic factors found on the same DNA molecule as the target genes; a famous example is genetic variants found in promoters that affect TF binding (Signor &

Nuzhdin, 2018). Mutation's results include consequences such as downregulation or even unintended function gain.

Enhancers experience a higher regulation through secondary CREs known as long-range CREs. These CREs also have binding sites for TFs and are usually located distally from the promoter/gene location. They impact gene expression by inhibiting enhancers' ability to form loops (insulators) and recruitment of repressor TF (silencers). It is important to note that downregulation by these elements is a standard and typical gene process. It allows the cell to ensure that specific genes are active while others are dormant.
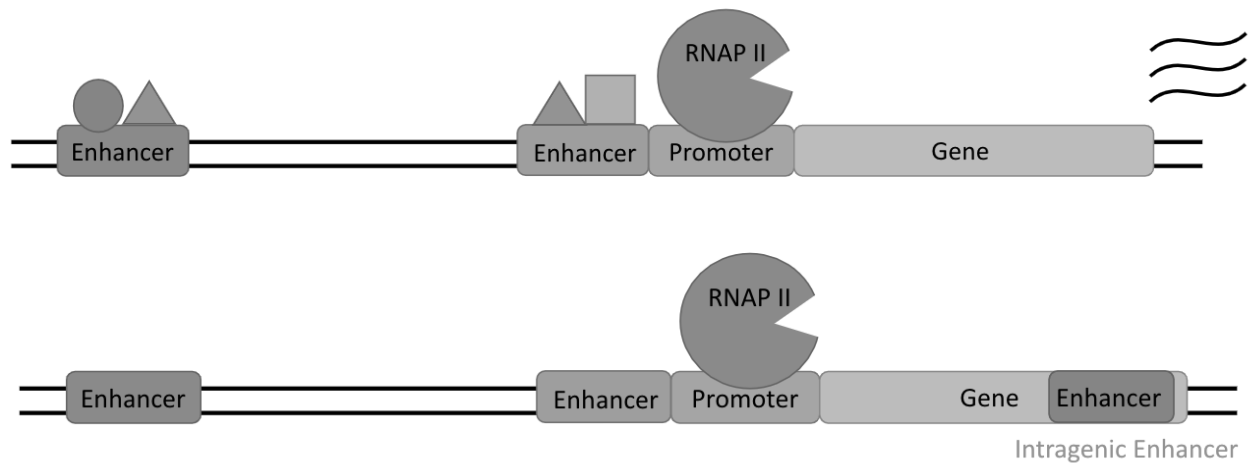
Figure 3. Graphic showcasing the difference between upstream enhancers and intragenic enhancers

Note. The genome structure comprises an upstream promoter where RNA polymerase II binds. While the promoter alone will cause transcription, enhancers are usually present to help increase transcription output. The top image shows a more typical genome structure with the enhancer being upstream, either near the gene or further out. In some cases, enhancers can be found within the gene bodies themselves. Their abilities and roles are still not well understood.

As stated previously, CREs like promoters and enhancers require TFs to bind to them to either promote or inhibit transcription. CREs act as scaffolds or landing pads for these factors, which perform the critical biochemical activities required for transcription. CREs can bind multiple factors simultaneously and factors containing various chemical properties. These factors can communicate and influence one another at these CREs, adding another layer to the complexity of regulation. Studies looking at cellular reprogramming and animal development have seen that both gene regulation and cell differentiation are dependent on multiple TFs working in a combinatory manner (Reiter et al., 2017). An example includes the *Drosophila* embryonic *even-skipped (eve)* muscle and heart enhancer, which requires the recruitment of mesodermal TFs Twist and Tinman (Halfon et al., 2000). In general, much of the research regarding this combinatory role was focused on the context of enhancers. Multiple studies have found enhancers that demonstrate this combinatory action, including the previously mentioned *eve* stripe 3 enhancer in *Drosophila*, the Oxt-a enhancer in *Ciona*, and the human interferon-beta enhancer (Reiter et al., 2017). These studies revealed that having the individual TFs is insufficient for maximum transcriptional activity, indicating the importance of cooperation between TFs (Yáñez-Cuna et al., 2014).

CREs themselves are regulated through mechanisms involving chromatin remodeling. Chromatin is a eukaryotic complex between DNA strands and histone proteins whose primary function is to package the long strands into a compact, more dense structures. This compact form prevents tangling, protection against DNA damage, and DNA replication/expression regulation. Chromatin remodeling is the modification of chromatin organization to allow transcriptional machinery to access the DNA. Enzymes modify the histone proteins through histone tail modifications that can tighten or loosen interactions with DNA. This dynamic process ensures

that DNA is only accessible at times of need, preventing unnecessary replication. In response to stress, chromatin relaxation is one of the earliest mechanisms initiated. Specific TFs known as pioneer factors can bind to the condensed chromatin structure, resulting in relaxation-activating modification (Zaret, 2020). Through relaxation and subsequent nucleosome remodeling, additional TFs access CREs and alter gene expression in favor of those needed for the current cellular environment.
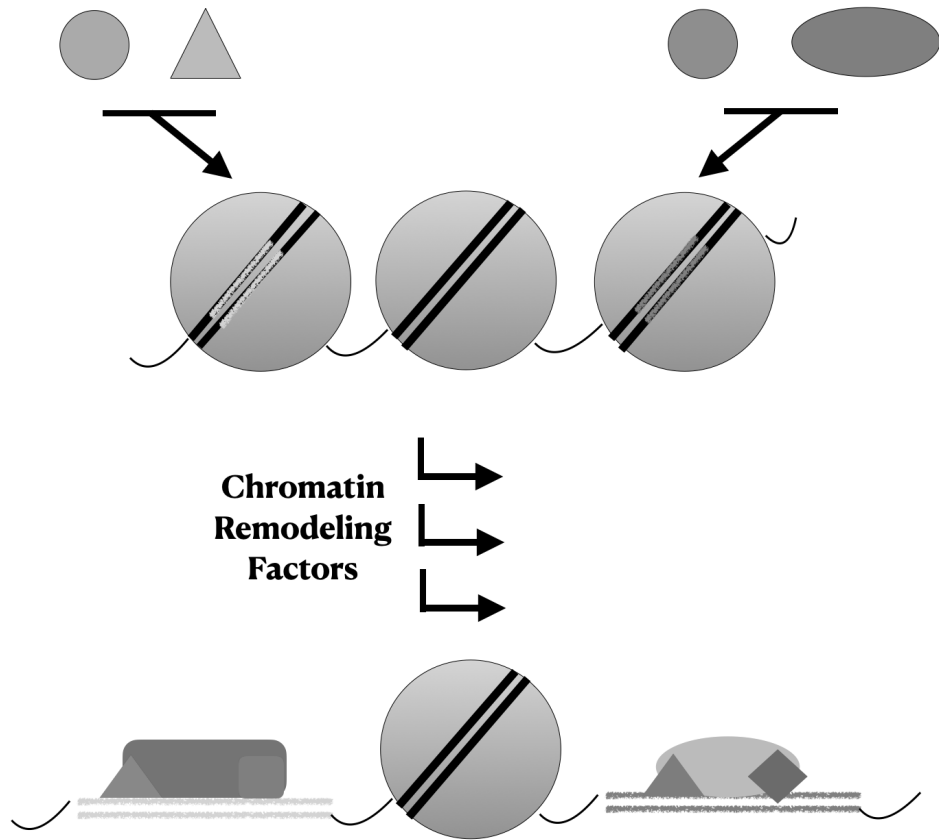
Figure 4. Graphic of the chromatin landscape

Note. DNA sequences are wrapped around histone proteins that make up nucleosomes. The DNA sequence is mainly inaccessible to other proteins. Specific TFs, known as pioneer factors, can open the DNA for chromatin remodeling factors. These factors allow DNA to be bound and the associated gene to be expressed.

To better understand the cellular stress response, we need to decode which conditions are triggered to respond to a cell's various stress conditions. As one can see, many different components govern gene regulation. Depending on when elements are activated (temporal), what features are activated (spatial), and at what quantity (quantitative), the cell can produce a wide range of different transcriptional and cellular states. Other genes are activated depending on the type of stress stimuli being experienced, but occasionally, induced genes overlap between the different stimuli. Thus, we can speculate that specific common TFs are involved in stress response. One such heavily studied TF is the p53 encoded by the *TP53* gene.

p53 is known as the "Guardian of the Genome" due to its potent tumor suppressor functions in response to DNA damage (Lane, 1992). This activity is due to p53's ability to regulate an extensive network of target genes downstream of stress. p53 is part of a family of TFs, known as the p53 family, that evolved from a common ancestral gene found in most invertebrates (Belyi et al., 2010). This ancestral gene was believed to protect the cell, both its integrity and germline, by inducing cell death in the presence of genome damage. Later, higher vertebrates divided the gene into three individual factors with diversified functions. p53 took on the role of tumor suppression, while the paralogs p63 and p73 controlled developmental parts. p63 is required to properly develop stratified epithelia found in the epidermis, urinary tract, and other internal cavities (Koster & Roop, 2003). A loss of p63 has been shown to lead to developmental failure due to benign and malignant tumors forming within the respiratory passage and the gastrointestinal tracts (Flores, 2007). p73, on the other hand, is required for the development process related to neurogenesis of neural structures and cerebrospinal fluid function maintenance (Yang & Bronson, 2000). Loss of p73 resulted in smaller numbers and poorer qualities of neural cells such as oligodendrocytes and neurospheres (Agostini et al., 2012). While

all three family members have been studied, p53 is often the focus of most research due to its role in human disease prevention.

p53 is a sequence-specific TF, activated in response to a wide range of stress-induced stimuli, including DNA damage. As mentioned previously, it plays a prominent role in conversing stability through the prevention of genome mutations. p53 primarily carries its role out by acting as a TF that induces the expression of downstream genes needed for cellular repair/cellular stress response (Cadwell & Zmabetti, 2001). It has been noted that over 50% of all human tumors contain p53 mutant variants suggesting p53's significant role in cancer prevention. Most p53 mutations occur through changes in the p53's DNA binding domain, resulting in the factor unable to bind to specific DNA elements. Loss of DNA binding also leads to loss of gene regulatory activity, underscoring the importance of the ability of p53 to activate target genes for its tumor suppressor role. p53 has shown remarkable flexibility regarding the variation of its binding site, with only a select nucleotide causing complete loss of function. p53's transcriptional output via the vast genes involved determines cellular fate. p53 has three significant roles in the cell: growth arrest, DNA repair, or apoptosis (Shaw, 1996). p53 involves many genes that share a commonality of returning a cell to a homeostatic state but do it in various ways and with other genes.

As stated previously, p53 activates many target genes that restore cellular and genomic homeostasis or prevent the spread of mutations. One such gene is *GADD45A*, activated in the growth arrest processes. *GADD45A*, an abbreviation of Growth Arrest and DNA Damage Inducible, is a member of a family of three proteins, Gadd45A, Gadd45B, and Gadd45G, which modulate response to genotoxic or physiological stress. Gadd45a, the first discovered family member, is activated by growth arrest conditions and mutagens, such as ionizing radiation

(Papathanasiou et al., 1991). Its activation is modulated by multiple TFs, including via p53-dependent and independent means. TFs bind to multiple putative regulatory elements within and near the *GADD45A* gene, including a canonical upstream enhancer, a proximal promoter, and even within an intragenic enhancer location. These specific binding sites are unique in comparing p53 and other TF-activated transcription in two different CREs. Previous research has found that intragenic enhancers play many notable roles, including working as alternative promoters (Kowalczyk et al., 2012), dampening gene expression, or contributing to transcription levels.
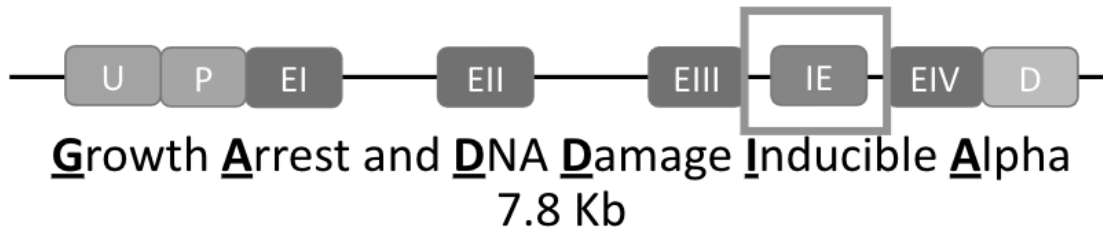
Figure 5. Graphic showing GADD45a genome structure

Note. GADD45A gene is in human chromosome 1 and is induced after growth arrest or DNA damage. One of the genes belonging to the *GADD45A* gene family shares similar stressor activators. *GADD45A* is a relatively short gene, being only 7.8 kilobases, with an upstream promoter and canonical enhancer. It contains an enhancer sequence within its third intron, highlighted in red. This enhancer sequence has shown transcription binding sequences, including the standard TF p53.

Our research goal was to further understand the complexity of the stress pathway by examining what role intragenic enhancers play in gene regulation, specifically in the case of *GADD45A*. Using the *GADD45A* gene as a model, we analyzed transcription rates at the basal level and after altering key TF binding locations within the intragenic enhancer. We used the gold-standard luciferase reporter gene assay in two different cell lines to determine the effect of the absence and presence of two TFs: p53 and AP1. Future directions will include altering promoter sequences, either through scrambling the canonical promoter or replacing it with other promoters, allowing us to see the relationships between a specific gene promoter and intragenic enhancer. Furthermore, we can answer evolutionary questions by changing the intragenic enhancer sequence by scrambling it, replacing the whole enhancer, and even replacing it with a different organism's enhancer.

The transcriptome is far more complicated than previously believed. Understanding how these pathways are regulated allows us to use this knowledge to develop tools and techniques to ensure these pathways are well maintained.

Materials and Methods

1. Cloning putative intragenic enhancers for Luciferase Reporter Gene Assays

The putative wild-type *GADD45A* intragenic reporter plasmid was constructed through Gibson assembly by adding compatible sequences to our amplicon, the *GADD45A* intragenic enhancer, for cloning into the *XhoI* site of Promega's pGL4.24[*luc2P/minP*] vector.

The sequence of the putative intragenic enhancer of *GADD45A* was retrieved using UCSC Genome Browser (genome.ucsc.edu) based on TF (transcription factor) ChIP-seq cluster data. A double-stranded DNA template gBlock (**Table 2)** containing that gene fragment was purchased from Integrated DNA Technologies (IDT). The gBlock was amplified using primers SL1273 and SL1274 with Q5 High-fidelity DNA polymerase via Polymerase Chain Reaction (PCR). The primers SL1273 and SL 1274 were designed to add approximately 445 bp of sequence homologous to the sequence flanking the *XHOI* site in pGL4.24 using New England BioLab's (NEB) website NEBBuilder. Gel electrophoresis verified the expected size of the amplicon representing the enhancer sequence by running a 2% agarose gel with an NEB 100bp DNA ladder. A band corresponding to the anticipated length of the enhancer (485 bp) was cut out from the gel for DNA extraction and purification using the NEB Monarch Gel Extraction Kit.

The pGL4.24 plasmid was restriction digested using the *XhoI* restriction enzyme site directly upstream of the minimal promoter sequence. The digested plasmid was verified by running a 1% agarose gel with NEB 1 kb DNA ladder. The 4411 bp band was extracted from the gel and purified using the Monarch Gel Extraction Kit.

Gibson assembly was used to fuse the enhancer region to the cut pGL4.24 plasmid using the HiFi Genome Assembly kit (NEB). Two negative controls were run: plasmid only and

plasmid plus the HiFi DNA Assembly Master Mix (and without PCR amplicon insert). Gibson reactions were run for 15 minutes at 50°C. Then 10% of the mixture was transformed into chemically competent *E. Coli* bacterial cells from New England BioLabs and plated into 15 mL LB Agar plates containing 1mL of a 1000mg/mL ampicillin stock. The plates were incubated at 37˚C for 12 hours. Individual colonies were then grown for 16 hours in 5mL of LB supplemented with 1 mL of 1000mg/mL ampicillin. Plasmid DNA was then isolated from these bacterial cultures using the E.Z.N.A plasmid DNA mini kit (Omega Bio-Tek). The plasmid was verified using Eurofins Genomics' Sanger sequencing using primer RV3 that had complementary binding to our inserted fragment. A pool of dsDNA fragments is measured and separated by the base and then graphed onto an electropherogram through Sanger sequencing. Using that electropherogram, we referred to our in-silico plasmid and checked the two matched sequences.

Figure 6. Gibson assembly graphic

Note. This figure shows the primary mechanism of how Gibson Assembly ligates two linear strands together. Primers and restriction digest produce ends that have anneal to one another.

Construction of mutant reporter plasmids was done through Gibson assembly by adding

compatible sequences to an amplicon, the *GADD45A* intragenic enhancer containing mutated

putative regions, for cloning into the *XhoI* site of Promega's pGL4.24[*luc2P/minP*] vector.

Figure 7. Changes to putative binding sites of TFs p53 and AP1

Note. This figure shows the changes made to the putative binding sites of TFs p53 and AP1. The bases that were changed are underlined and highlighted in red.

Using UCSC Genome Browser, putative p53 and AP1 binding sites in the intragenic enhancer sequence were identified based on their similarity to canonical motifs. The putative p53 binding site has the sequence AACATGTCTAAGCATGCT. We mutated the wild-type sequence to AA**G**AT**C**TCTAAG**G**AT**C**CT (**Table 1)**, which should ablate the binding of p53 to DNA based on previously reported literature (Yang et al., 1995). The putative AP1 binding site has the sequence GTGAGTCA, which we mutated to G**A**GA**TCT**A (**Table 1)** based on previous literature (Pearson et al., 2009), showing AP1 binding loss. A double-stranded DNA template gBlock (**Table 2)** containing that mutated gene fragment for each TF mutation was purchased from Integrated DNA Technologies (IDT).

The gBlock was amplified using primers SL1273 and SL1274 with Q5 High-fidelity DNA polymerase via Polymerase Chain Reaction (PCR). The primers SL1273 and SL 1274 were designed to add approximately 445 bp of sequence homologous to the sequence flanking the *XHOI* site in pGL4.24 using New England BioLab's (NEB) website NEBBuilder. Gel electrophoresis verified the expected size of the amplicon representing the mutated enhancer sequence by running a 2% agarose gel with an NEB 100bp DNA ladder. A band corresponding to the predicted size of the enhancer (485 bp) was cut out from the gel for DNA extraction and purification using the NEB Monarch Gel Extraction Kit.

The pGL4.24 plasmid was restriction digested using the *XhoI* restriction enzyme site directly upstream of the minimal promoter sequence. The digested plasmid was verified by running a 1% agarose gel with NEB 1 kb DNA ladder. The 4411 bp band was extracted from the gel and purified using the Monarch Gel Extraction Kit.

Gibson assembly was used to fuse the enhancer region to the cut pGL4.24 plasmid using the HiFi Genome Assembly kit (NEB). Two negative controls were run: plasmid only and plasmid plus the HiFi DNA Assembly Master Mix (and without PCR amplicon insert). Gibson reactions were run for 15 minutes at 50°C. Then 10% of the mixture was transformed into chemically competent *E. Coli* bacterial cells from New England BioLabs and plated into 15 mL LB Agar plates containing 1mL of a 1000mg/mL ampicillin stock. The plates were incubated at 37˚C for 12 hours. Individual colonies were then grown for 16 hours in 5mL of LB supplemented with 1 mL of 1000mg/mL ampicillin. Plasmid DNA was then isolated from these bacterial cultures using the E.Z.N.A plasmid DNA mini kit (Omega Bio-Tek). The plasmid was verified using Eurofins Genomics' Sanger sequencing using primer RV3 that had complementary binding to our inserted fragment. A pool of dsDNA fragments is measured and separated by the base and then graphed onto an electropherogram through Sanger sequencing. Using that electropherogram, we referred to our in-silico plasmid and checked the two matched sequences.

C. Construction of a double p53 binding site reporter plasmid was created based on Promega's pGL4.38 [*luc2P*/p53 RE/Hygro] vector, containing two copies of the p53 response element. The reporter plasmid was constructed by annealing oligos using T4 Polynucleotide Kinase (NEB) and then fusing it into a digested gap between XhoI and KpnI of Promega's pGL4.24[luc2P/minP] vector via a ligation reaction. Two oligos, one measuring 67 bp and the other measuring 75 bp, were ordered from IDT that contained the p53 response element (**Table 2)** and had complementary overhangs to the pGL4.24 vector at the *XhoI* and *KpnI* sites.

The pGL4.24 plasmid was restriction digested using the *XhoI* and *KpnI* restriction enzyme site directly upstream of the minimal promoter sequence. The digested plasmid was verified by running a 1% agarose gel with NEB 1 kb DNA ladder. The 4014 bp band was extracted from the gel and purified using the Monarch Gel Extraction Kit.

Ligation was used to fuse the RE to the cut pGL4.24 plasmid using T4 DNA ligase (NEB). Two negative controls were run: plasmid only and plasmid plus the T4 DNA ligase (and without oligo amplicon insert). Ligation reactions were run for 1 hour at room temperature. Then 30% of the mixture was transformed into chemically competent *E. Coli* bacterial cells from New England BioLabs and plated into 15 mL LB Agar plates containing 1mL of a 1000x ampicillin stock. Pure DNA was isolated from the bacteria using E.Z.N.A plasmid DNA mini kit (Omega Bio-Tek). The plasmid was verified using Eurofins Genomics' Sanger sequencing using primer RV3 that had complementary binding to our inserted fragment. Eurofins Genomics verified the plasmid sequence through Sanger Sequencing using primer RV3. A pool of dsDNA fragments is measured and separated by the base and then graphed onto an electropherogram through Sanger sequencing. Using that electropherogram, we referred to our in-silico plasmid and checked the two matched sequences.
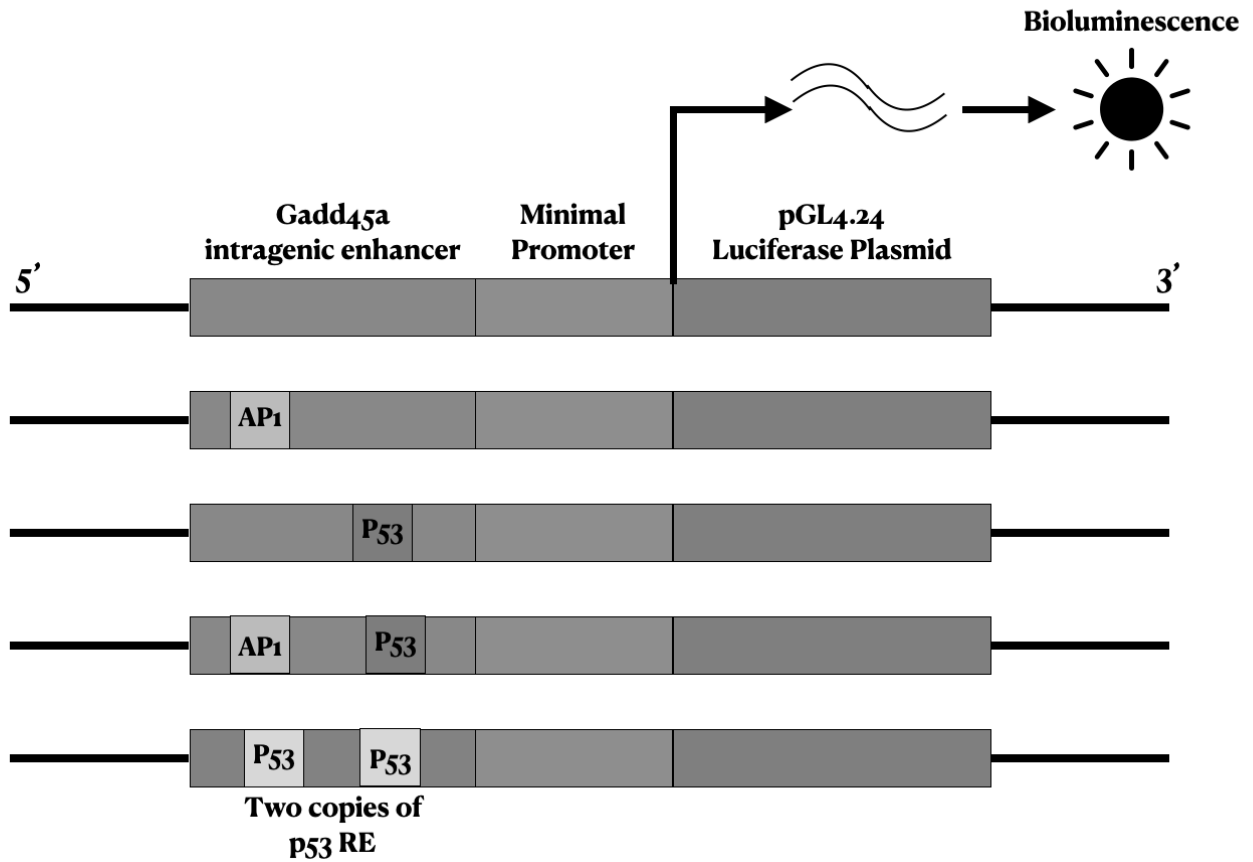
Figure 8. Constructs

Note. The following five plasmid constructs were made using Gibson assembly or ligation reaction between gBlock PCR products and restriction digested pGL4.24. The top plasmid is the wild-type enhancer, while the next three are the mutated versions. The bottom plasmid was made using the p53 response element from Promega's pGL4.38 [*luc2P*/p53 RE/Hygro] vector.

2. Luciferase Assay of constructed reporter plasmids

Two HCT116 human colorectal cancer cell lines, wildtype (WT) and *TP53* knockout (KO), were purchased from ATCC. Cells were cultured using McCoy's 5A, 1X (Iwakata & Grace Mod.) with L-glutamine media supplemented with 50 mL of Corning's Fetal Bovine Serum (REF 35-016-CV) and 10 mL of Invitrogen's 100x Penicillin-Streptomycin solution (REF # 15070-063). Cells were cultured in a 37-degree incubator. Corning 0.25% Trypsin, 2.21mM EDTA, and 1X [-] sodium bicarbonate were used to resuspend cells, and cells were counted using Life Technologies Countess II FL. Cells were seeded at 50,000 cells per well of a 96-well plate. Each cell line was transfected with 90 ng of either the wild-type enhancer, one of the mutated plasmids, or a no-enhancer plasmid negative control and ten ng of pGL4.74 [*hRluc/*CMV] Renilla using JetPrime transfection reagent.

Samples were transfected in triplicate and three biological replicates. Samples were transfected into the cells after a 24-hour incubation alongside the JetPrime buffer. Each transfection component's concentration was done according to JetPrime's forward transfection protocol. After a 24 incubation, cells were treated with 100 uL of a 1000mg/mL DNA–damage-inducing etoposide stock with controls of RNase free H2O and DMSO accordingly, all concentrated at 1000mg/mL. Luciferase gene reporter assays were completed using the Promega Dual-Glo Assay kit. The BioTek Hi Synergy, the luminometer/plate reader, provided luminescent data for the firefly and Renilla enzymatic reactions. All data were analyzed using a two-way ANOVA, with normalization of firefly to Renilla, using GraphPad Prism software.

3. Qualitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR)

HCT 116 WT Cells were seeded at 500,000 cells per well in a six-well plate overnight and had a 6-hour treatment with cell stress inductor drug etoposide, with controls of RNase free H2O and DMSO accordingly, all concentrated at 1000mg/mL. Three different concentrations of etoposide were used: 100 mM, 0.01 mM, or 0 mM. RNA was extracted using the Quick-RNA Miniprep Kit (Zymo), following the recommended manufacturer's protocol. RNA concentrations were obtained using a Nanodrop Spectrophotometer. 1 μg of RNA was used to create cDNA using the High-Capacity cDNA Reverse Transcription kit (Thermofisher). RNA was added to both reverse transcriptase (RT) master mix and a no RT master mix and placed in a thermocycler for 10 minutes at 25 degrees, 120 minutes at 37 ˚C, and 5 minutes at 85 min. qPCR master mixes were created using the BioRad iTaq SYBR Green Supermix. The cDNA was plated with the master mix and standards onto a 384 well plate with *GADD45A* primers and a GAPDH control. The plate was read using the ABI 7900HT real-time PCR instrument and analyzed using a two-group *t*-test on GraphPad Prism software.

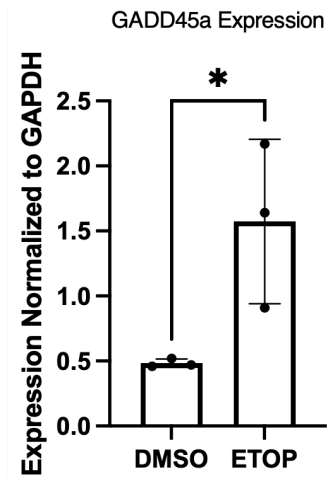1. Qualitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR)

Figure 9. RT-qPCR with GADD45a primers

Note. HCT 116 WT cells were treated with DMSO and etoposide at 1000mg/mL concentration for 15 hours. Expression was significantly increased when cells were treated with etoposide compared to baseline levels, as indicated by DMSO expression levels.

Previously literature had shown that transcription of *GADD45A* is induced during stress conditions, including DNA damage like the type generated by etoposide (Tamamori-Adachi et al., 2018). We sought to determine what effect stress-inducing drugs like etoposide would have on the rate of gene expression of the GADD45A gene. We performed reverse transcription-coupled quantitative PCR (RT-qPCR) to quantitatively determine the accumulation of *GADD45A* mRNA after treatment with etoposide or the negative vehicle control DMSO. After cell seeding, a six-hour treatment with 3 mL of 1000mg/mL of etoposide or DMSO in HCT116WT cells, the qPCR data were analyzed and normalized to the expression of hGAPDH mRNA. We expect GAPDH mRNA expression to be unchanged in response to either DMSO or etoposide, thus serving as a critical control for RNA isolation and reverse transcription reactions.

Compared to the vehicle DMSO control, we observed (**Figure 9)** that etoposide treatment increased the mRNA expression of *GADD45A*. These data were reproducible across three biological replicates, as demonstrated via ANOVA statistical test (**Figure 9).** These results suggest that treatment of 3 mL of 1000mg/mL etoposide likely induces DNA damage, which leads to the activation of *GADD45A* mRNA transcription. This is an important control to determine whether etoposide could act as an appropriate activator of DNA damage and, subsequently, *GADD45A* mRNA induction. Based on this data, we speculated that *GADD45A* was likely involved in the cellular stress responses, particularly those with DNA-damage-inducing stimuli. We, therefore, wanted to determine how *cis-regulatory elements (CRE) near and within the GADD45A* gene might be involved in basal (non-induced) and stress-activated (induced) transcription.

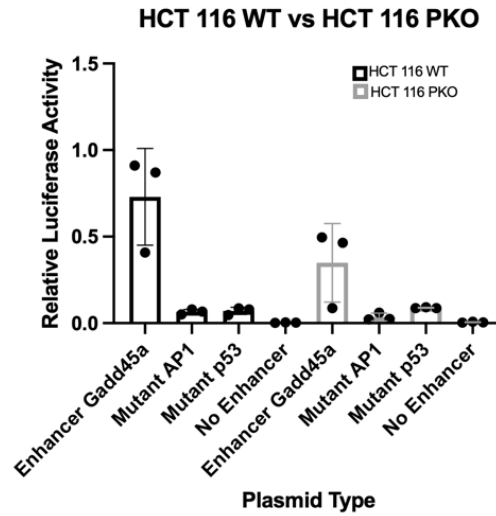2. Luciferase Assay of Reporter Plasmids

Figure 10. Luciferase Results with no treatment

Note. The following graph shows the luciferase activity of the wild-type enhancer, the AP1 mutated enhancer, the p53 mutated enhancer, and the no enhancer pGL4.24. Each construct was transfected into HCT 116 WT and HCT 116 p53 KO cell lines

31

We knew that etoposide increased *GADD45A* gene expression based on preliminary qPCR data. We wondered which CREs were involved in this stress-induced gene expression. Based on previous literature, *GADD45A* is regulated via a potential CRE within its third intron. This is considered an intragenic enhancer. As CREs act as scaffolds for TF (transcription factor) binding, it is essential to understand what TFs might bind at these enhancers and how that binding might change relative to changes in transcription. We used publicly available transcription factor motif predictions at the UCSC Genome Browser to identify a p53 binding site present within this putative intragenic enhancer. It is not surprising to find a p53 binding site here because previous data has shown p53's prominent role in many cellular stress responses, including in response to etoposide-induced DNA damage. Thus, we sought to determine whether p53 might bind to this location and control the expression of *GADD45A* in response to DNA damage. It is important to note that *GADD45A* is a well-studied p53 target gene and that we are attempting to determine the mechanism by which p53 serves this well-studied role as an activator of *GADD45A*.

Looking at the HCT 116 p53 KO cell line data in **Figure 10,** we notice a couple of interesting data points. HCT 116 p53 KO cells are cell lines with both alleles of the *TP53* gene knocked out. Thus, these cells do not contain p53 protein and cannot upregulate p53-dependent gene expression in response to DNA damage. So, while the WT enhancer does contain the binding site for p53, it is not bound by p53 as p53 protein is absent from these cells under all conditions. Comparing the WT enhancer in HCT116 WT and p53 KO shows a half-fold decrease in transcriptional activity. This indicates that the presence of p53 is required for maximum transcriptional activity. We might expect that removing the p53 binding site would not cause a change compared to the WT enhancer in KO cells since having no p53 around nor the binding

32

site would have a similar effect. However, we see an approximate 66% decrease. This might indicate that another cellular component is utilizing p53's binding site.

We know that TFs usually bind in combinatory manners, meaning that an enhancer can contain many different TF binding locations. We noticed the presence of a putative AP1 binding motif in the same intragenic enhancer containing a p53 motif within the *GADD45A* third intron. AP1 is a TF that plays a role in cellular processes involving cell growth and apoptosis. We hypothesized that it, too, played a role in Gadd45's cellular stress response (Daino et al., 2006).

To study the functions of the intragenic *GADD45A* gene and two of its possible regulating TFs, we needed to create a transcriptional activity reporter system. We used a gold-standard enhancer luciferase reporter assay and inserted our previously identified intragenic *GADD45A* enhancer upstream of a firefly luciferase gene. If the inserted sequence acts as a transcriptional enhancer, luciferase RNA would be transcribed and then translated into a functional enzyme that produces luminescence, which can then be measured. An increase in luciferase activity would correspond to an increase in transcription.

Luciferase assay reporters were created using primers that amplified dsDNA fragments containing desired genes. The resulting PCR products were used in Gibson assembly with pGL4.24. PGL4.24 was run alongside the reporters, acting as a negative control due to its lack of an enhancer. A total of three different reporters were constructed: the WT intragenic *GADD45A* enhancer, a mutated p53 binding site containing *GADD45A* intragenic enhancer, and a mutated AP1 binding site containing *GADD45A* intragenic enhancer. All four reporters were transfected into both HCT116 WT and HCT116 KO cells. The activity and effectiveness of each reporter's enhancer were measured by the normalized luciferase value, the ratio of normalized Renilla to luciferase activity.

As seen in **Figure 10**, the reporter with the highest luminescence detection was the WT intragenic enhancer in HCT116 WT cells, a 1.5-fold increase from the no enhancer pGL4.24 reporter. This fold increase indicates that *GADD45A*'s intragenic enhancer is acting as an enhancer since it caused a change in transcription levels. When we mutate either the AP1 or p53 binding sites in these intragenic enhancers, we see a considerable reduction in luminescence activity, indicating a less efficient transcriptional change. This finding implies that both TFs must have the maximum transcriptional change. However, neither TF binding site loss resulted in complete loss of enhancer abilities as evident that activity did not fall to no enhancer levels. This finding suggests that these TFs act additively, reminiscent of the Billboard model of enhancer activity (Catizone et al., 2020).

We believe that the AP1 and p53 binding sites are necessary for maximum transcriptional activity driven by the intragenic enhancer, based on this luciferase data. The loss of either cause decreased transcriptional activity. However, this assay was run around normal non-stress conditions, which meant we looked at basal transcription levels. As stated previously, TFs are recruited based on certain stimuli. We need to induce a stressful situation in both cell lines to see the gene working under its intended conditions.
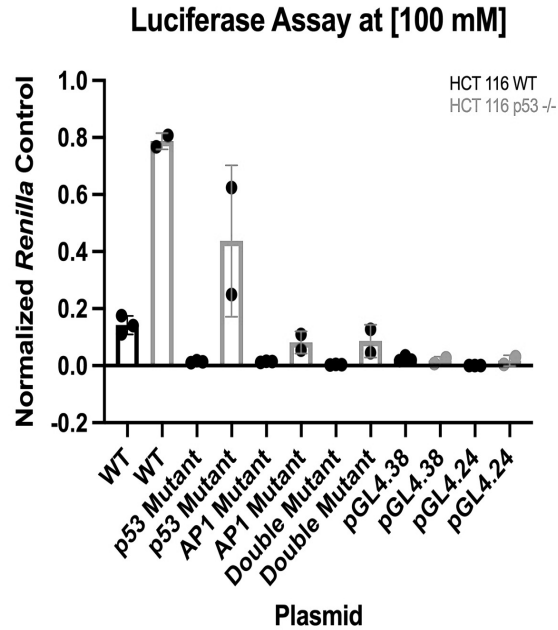
3. Luciferase Assay of Reporter Plasmids with Treatment

Luciferase Assay at [0 mM]

A.

Luciferase Assay at [0.01 mM]

B.

Luciferase Assay at [100 mM]
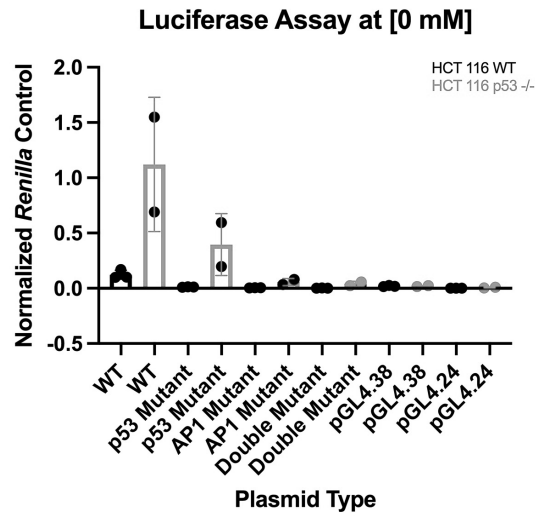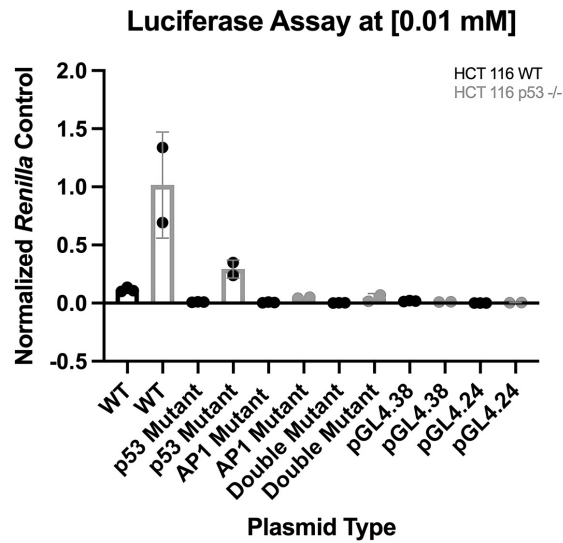
C.

Figure 11. Luciferase Assay in HCT 116 WT and p53 KO cells

Note. HCT 116 WT and p53 KO cells were transfected with six luciferase reporter plasmids. The plasmids included, as listed left from right, were a wild type of *GADD45A*s intragenic enhancer, a mutant containing an altered AP1 binding site in *GADD45A*'s intragenic enhancer, a mutant containing an altered p53 binding site in *GADD45A*a's intragenic enhancer, a double mutant containing both altered AP1 and p53 binding sites, a two copied p53 response element, and a minimal promoter. Each plasmid was treated with two etoposide conditions, 100 mM, or 0.01 mM, and a DMSO treatment (0 mM of etoposide).

Based on these preliminary luciferase reporter experiments, we made several alterations to the assay to better understand GADD*45A*'s intragenic enhancer's role in cellular stress response. We had initially only created luciferase assay reporter genes containing the WT intragenic enhancer, a mutated p53 version of said enhancer, and a mutated AP1 version of the said enhancer. The extent to which these two motifs work together or independently is unknown. Thus, we sought to test the effects of having both TF motifs mutated at once, so we needed to create one more reporter assay gene. We created this mutation using similar approaches outlined above (see *Materials and Methods)*. In brief, oligos containing the mutant sequences were

synthesized, annealed, and ligated into the pGL4.24 reporter plasmid to create a double mutant (p53 and AP1 motif mutants).

We decided to continue using a standard luciferase reporter assay as our transcriptional activity reporter system. However, we needed to add a day to our previous experiment to allow for an 18-hour cell treatment with different etoposide concentrations. We had decided to do three different etoposide concentrations: a high level (100 mM), a low level (0.01 mM), and a no-drug control. After HCT116 WT and p53 KO cells were seeded, transfection, and treated with etoposide, they were read using a luminometer/plate reader.

Looking at all three etoposide conditions, we notice that the double mutant plasmid contained a 2x fold decrease compared to the WT enhancer in HCT 116 cells. This suggests that the loss of both binding motifs results in a substantial reduction in transcriptional activity, indicating that both TFs are required for maximum transcriptional output. However, the loss of those two

motifs did not result in a complete loss of transcriptional activity. This observation suggests that other motifs are contributing to the transcriptional output. Future experiments would include looking at the presence of other TF motifs present and noting the effect of loss of said motif.

The goal of our initial experiment was to add DNA-damage-inducing etoposide onto both HCT116 cells (WT and p53 KO) to induce the WT enhancer, the p53 mutant, the AP1 mutant, and the p53/AP1 double mutant. We wanted to note the difference between the basal and induced activity. Initial trials could not produce such a difference, so we looked at the availability of similar plasmids online. When searching Promega's plasmid catalog, we noticed pGL4.38[*luc2P/*p53 RE/Hygro]. When looking at its protocol, we saw data suggesting that this plasmid could be induced via p53 due to the presence of a p53 regulatory element (**Table 2).**

We had expected that the presence of two binding sites would cause an increase in transcription that is greater when compared to the endogenous level. However, the data showed the opposite: the endogenous enhancer had relatively similar transcription levels to pGL4.38. This suggests that etoposide is not inducing transcription in pGL4.38 compared to Promega's data. Possible explanations for this inability to generate might be a biological problem occurring within the intragenic enhancer or a broader problem with our plasmid design. To perform a luciferase assay, we removed the enhancer from its native location within an intro and placed it upstream of a promoter. This change might have affected *GADD45A*'s transcriptional ability.

We also wanted to see the relative transcriptional activity of the endogenous *GADD45A* enhancer to a plasmid (pGL4.38) containing a synthetic enhancer with two p53 binding sites next to each other. Lastly, we wanted to mimic stress conditions in the cell to see their effect on the transcriptional activity of *GADD45A*, a gene supposedly heavily involved in stress response.

Looking at **Figure 11a**, we notice similar trends from the original luciferase assay experiment (**Figure 11**). This is expected since they are done under similar non-stress basal level conditions. Moving to **Figure 11b**, we notice similar trends to **Figure 11A,** which is expected since the change in etoposide concentration is not very large. One trend is the WT enhancer in the WT cells versus KO cells. The enhancer had a 1x fold increase in the KO relative to the WT cells. This is unexpected since previous luciferase data (see **Figure 10**) had shown the opposite effect. We see the same occurrence in the mutant p53 reporter in WT and KO cells between the luciferase experiments. Another data point of interest is that the WT enhancer in WT cells showed more significant transcriptional activity than all the mutations in WT cells. This supports previous findings that loss of either TF binding causes a reduction in the maximum

transcriptional activity. Furthermore, all three mutations contained the same transcription level, pointing to an additive relationship instead of a potentiative one.

Lastly, in **Figure 11C**, we noticed an interesting pattern of similar trends but quantitative lower transcription activity. For example, comparing the WT enhancer activity in KO cells treated with 100 mM of etoposide to the WT enhancer in KO cells treated with 0.01 mM, we see around a 2x fold decrease. This difference at high concentrations might indicate other pathways being activated in forgoing this one.

Discussion and Future Directions

In this work, we aimed to understand better how TFs (transcription factors) cooperate to drive the expression of the *GADD45A* gene at an intergenic enhancer found in the 3rd intron. Using luciferase reporter gene assays, we identified that both a putative p53 motif and a putative AP-1 family member motif were critical. Sequential luciferase assays were done with an etoposide treatment, a known DNA damage drug. The addition of this drug because of a qPCR experiment showed *GADD*45A mRNA levels increasing upon treatment with etoposide. Based on these three significant experiments, we further explored the role of TFs in the unique intragenic enhancer placement.

As stated previously, we had extracted RNA from HCT 116 WT cells after treatment with etoposide and a DMSO control. *GADD*45A is present in these cell lines, and we measured mRNA levels for both situations using qPCR primers. After analysis, we noted that cells treated with etoposide showed a 3x fold increase in GADD45 mRNA production relative to mRNA levels during only DMSO treatment. This increase supports previous findings that *GADD45A* is a gene that is involved during cellular stress, one that is caused by DNA damage. By showing that *GADD45A* is induced upon etoposide treatment, we were able to move on to experiments testing the effect of loss of specific TFs under cellular stress conditions. Some future experiments to expand on this data would test the *GADD45A* gene with other cellular stress response drugs like nutlin or tunicamycin. As stated previously, *GADD45A* is induced by additional cellular stress, including stimuli like ER stress and UV radiation. Based on the additional cellular stress, including triggers like ER stress and UV radiation. Based on the findings of those experiments, we could follow a similar format and perform a treated luciferase assay to see the effect of TF loss in intragenic enhancers during varied stress conditions.

Before *GADD45A*'s transcription rate under cellular stress, we wanted to see what

GADD45A's basal, unstimulated activity was. We also wanted to note what transcriptional

changes would occur to basal levels with loss of the p53 or the AP1 binding site. All three

scenarios were compared to the no-enhancer control pGL4.24. The assay was done in two cell

lines, HCT 116 WT and HCT 116 p53 knockout. The latter was utilized to see what the absence

of p53 would cause the transcriptional activity at the *GADD45A* intragenic enhancer.

The first important piece of data to note was that in the WT cell line, we saw that the WT

GADD45A intragenic enhancer showed a 1.5x fold increase compared to the no enhancer

pGL4.24. This finding reveals that the *GADD45A* intragenic enhancer is an enhancer since it

affects transcriptional levels. While it was previously known that enhancers are present

throughout the genome, the ones located within introns were not focused heavily compared to

their more prototypical brethren. Our data shows that intragenic enhancers contribute to a gene's

total transcriptional output. A couple of papers have tried to decipher the role of intragenic

enhancers in a gene's transcription levels. One possibility is that it is added to the transcriptional

output, which is what our data support (Kowalczyk et al., 2012).

Another possibility was that it acted as an alternative promoter and produced unstable

RNA, known as eRNAs. We could create a luciferase plasmid containing the intragenic enhancer

with no promoter present to test this. We could also create a plasmid containing the canonical

GADD45A promoter and note the difference between the two. The plasmids could also be done

in a treated luciferase assay to see the effect during cellular stress conditions, allowing us to see

if environmental conditions impacted intragenic enhancer use. The completed and future

experiments would allow us to understand the role of intragenic enhancers further.

41

Another interesting finding from our first luciferase assay was that mutation of the putative p53 binding site resulted in a 66% decrease in transcription compared to the WT enhancer. This indicates that the *GADD45A* gene is p53-dependent and requires its presence to have maximum amounts of transcriptional output. The ability of p53 to bind to this intragenic enhancer leads us to believe that intragenic enhancers similarly interact with TFs as canonical enhancers. This adds to previous knowledge regarding the role of TFs to showcase their widespread influence on various CREs. Furthermore, we see that the mutated version is still relatively higher when comparing mutated p53 transcription levels to the no enhancer levels. This suggests that while p53 is needed for maximum outputs, it is not required to drive transcription via this enhancer element. To expand on this finding, a future experiment would include mutating particular segments of the p53 binding site to note which regions are necessary for binding. In addition, we can also replace the human p53 binding site with other species' p53, such as *Drosophila melanogaster,* to see if there is any conservation of binding sites in intragenic enhancers. Both experiments would further understand the complexity of p53 binding sites in the intragenic enhancers.

Alongside the p53 mutation, we also mutated the putative AP1 binding site. The mutation of the said binding site resulted in around a 66% decrease in transcription levels relative to the WT enhancer. Once again, this indicates to us that AP1 is required for maximum transcription output. However, much like p53, the mutated AP1 transcription was relatively higher than the no enhancer meaning that the AP1 motif is not a mandatory feature needed for transcriptional activation via this intragenic enhancer. While there has been substantial data regarding AP1's role in enhancers, there has been a knowledge gap regarding its role in intragenic enhancers. It is involved in gene regulation regarding cellular differentiation, apoptosis, and cell growth. While

it has also been shown to participate in cellular stress responses, much of its role remains unknown. Our data indicate that AP1 is an essential factor in *GADD45A*-inducible stress like DNA damage. Furthermore, the data also points out that AP1 can bind to intragenic enhancers showing a widespread role in the genome. A future experiment could look at the loss of AP1 under different stress conditions and note if its function is still important.

The mutation data from the first round of luciferase assay showed that mutation of putative motifs that are expected to prevent binding of the protein/TF did not result in total transcription loss. This observation leads us to believe that the two TFs have a relationship where they enhance one another but do not depend on each other. Future experiments could investigate if similar relationships occur with other speculated TF. Furthermore, we can continue exploring p53's and AP1's roles by changing dimensions like orientation/spacing/location. Literature has shown that TFs are rarely alone as enhancers and their orientation and combination affect transcription (Spitz & Furlong, 2012). There are three main enhancer grammar models of how enhancers encode function. The first one is the enhanceosome model, which is the most rigid of the three models. It states that TFs bind in an exact arrangement for transcription (Bazett-Jones et al., 1994). The billboard model is the opposite of the enhanceosome model and suggests that there are no constraints on where TFs can bind. Instead, the TFs simply need to be present for transcription to occur. (Kulkarni & Arnosti, 2003). The last is the TF-collective model, which states that multiple TFs bind at specific binding sites and interact with other bound TFs. This model highlights the importance of TF- TF interactions (Junion et al., 2012). Enhancers most likely incorporate aspects of all three models, with the first two models being the extremes.

Our data support the TF-collective model since we showed that loss of either TF resulted in a loss of transcriptional output. This indicates that the TFs interact in some way to modify

transcription. To further study the grammar of enhancers, we could move the TFs around in the gene and enhancer. This would allow us to see if the other grammar models occur and the limitations in TF interactions.

Our first luciferase assay included a condition where the plasmids were inserted into HCT 116 p53 KO cells. An interesting finding in that condition is that the WT enhancer in the KO cells had a 5x fold decrease compared to the same plasmid in the HCT 116 WT cells. This indicates that the protein p53 is required for maximum transcriptional output since we saw a reduction in transcription when it was not present. Interestingly, comparing the WT enhancer to the mutant p53 binding site in KO cells, we see that they are not identical. We see a 3x fold decrease in the mutant p53 plasmid compared to the WT enhancer. This suggests that if p53 is not present, something else can occupy its binding site and induce transcription. When we lose the binding site, whatever factor was binding there loses its ability to bind. Future experiments could look at genome data and see if previous literature has found any specific factors known to bind there. We could also note if those factors contain other binding sites within the intragenic enhancer.

We repeated the assay after our initial luciferase assay but added some new details. We added two new plasmids: a double mutant. Mutated p53 and AP1 binding sites and pGL4.38 included two p53 binding sites in its enhancer region. Furthermore, we treated both HCT 116 WT and KO cell lines with two different concentrations of etoposide, either a high dose (100 mM) or a low dose (0.01 mM). We ran a control treatment containing DMSO (0 mM etoposide). We decided to run the luciferase assay under etoposide to note how both cells' lines would react under high and low-stress levels. *GADD45A* is said to be induced by DNA damage, so applying DNA-damage-inducing drug-like etoposide should activate *GADD45A* transcription.

Before running the second luciferase assay, we wanted to confirm that etoposide would induce *GADD45A*. We did this by doing RT-qPCR using *GADD45A* primers that targeted *GADD45A*'s 3rd intron. We had extracted RNA from HCT 116 WT cells that were treated with 1000x etoposide or a DMSO control. If *GADD45A* were induced by DNA damage, we would expect to see increased mRNA levels of *GADD45A* in treated cells versus the control. Looking at **Figure 8,** you will notice a 3x fold increase of mRNA in cells under the etoposide condition compared to the control, meaning that etoposide is inducing *GADD45A* production.

As stated in the results, we ran our second luciferase assay based on qPCR data. We noted quite a few interesting observations from this experiment. First, we noticed that the WT enhancer in HCT 116 WT cells was considerably lower than in the KO cells across all three concentrations. Our past luciferase assay showed the opposite trend, where the WT enhancer was lower in the KO cells. A possible explanation is that in stress conditions, a loss of p53 invoked an increase in other TFs that, in combination, allowed for greater transcriptional output. To test if this was the case, we could vary our concentrations into smaller intervals and see if the trend appeared.

Another interesting observation is that we saw an increase in transcription levels of the WT enhancer in WT cells at higher levels of etoposide. This result makes sense since previous literature stated *GADD45A* was involved in the stress response when DNA damage occurs. So, increasing etoposide concentration induces more damage, thus requiring more of the gene to deal with stress management. We also saw an increase in transcription in KO cells for the mutated p53 plasmid, the mutated AP1 plasmid, and the double mutated plasmid. We can interpret this result as indicating that higher stress conditions result in more recruitment of TFs that, in

45

combination, can result in higher transcription even at the loss of two other TFs. Future experiments could include mutating other putative TF binding sites to confirm this interpretation.

Despite the experiments we ran, we did have some limitations regarding our reporter systems. It is important to note that the structure of the luciferase assay constructs is not directly comparable to the structure of the native gene. As stated earlier, we have been testing the activity of an intragenic enhancer found in the 3rd intron of *GADD45A*. Meanwhile, the canonical Luciferase reporter gene assay places the target enhancer upstream to a minimal promoter at a relatively close (proximal) distance. This setup is drastically different from the native structure, affecting and changing the actual transcriptional output. Changing the location of CRE can change chromatin structure, which affects transcription (Li et al., 2007).) This means that changing the orientation of regulatory elements could leave out valuable information. We would need to create a luciferase assay construct in which the proximal promoter is found upstream of the intragenic enhancer. This would result in a construct that better illustrates the native context of the gene.

We could take it a step further by studying the gene *in vivo*. This would allow for a more precise analysis of enhancer activity in the native genomic native context. We could do in vivo mutagenesis, using methods like CRISPR, to make changes within cells. Possible changes would include mutating the p53 and AP1 putative factors to see how they affect enhancer activity. We could also mutate the intragenic enhancer either by changing a select few nucleotides, doing a complete scramble, or when deleting the whole enhancer. While this method will offer a better sense of the native environment than doing Luciferase assays, it will not be the most native concept since we are not integrating our changes into the genomic DNA of cells.

Lastly, the previously mentioned experiments focused on only the result of transcription. This dramatically oversimplifies what is occurring in the cell. By only looking at the result, we miss the steps in between that lead to that result. Was the result due to increased transcription or a decrease in degradation of mRNA products? We would not be able to know this through the previous experiments. We can use MS2 tagging to examine the rate of transcription induced by the intragenic enhancer. We would do this by adding a gene sequence that naturally results in stem-loop formation. Furthermore, these MS2 stem-loops are recognized by the MCP stem-loop binding protein, which can be fused to GFP, resulting in fluorescence when the gene is transcribed, allowing for detection of the mRNA in living cells (Bertrand et al., 1998). This will enable us to understand better the proportion of transcription the intragenic enhancer contributes. Furthermore, it will let us see if the intragenic enhancers' products are being trafficked out of the nucleus and possibly translated if the stem-loop sequence is placed within an exon. This experiment will aid us to understand the mechanisms behind the intragenic enhancer even more.

Our preliminary experiments and literature review determined that *GADD45A* is a p53-dependent gene, meaning that p53 presence is required for maximum transcriptional output. However, p53 requires other TFs, such as AP1, to enact this maximum level highlighting the combinatorial and additive abilities of TFs. Canonical enhancers traditionally found upstream of the gene have also been shown to have TFs working in combinatorial roles. This similarity between the two enhancer types suggests that intragenic enhancers behave similarly to their more traditional brethren. This opens the possibility to study if intragenic enhancers share other characteristics such as looping mechanisms and silencer effects. These observations showcase the diversity of CREs and the complexity behind regulating them due to the combination of TFs

required. Future work regarding the mechanisms behind TFs that bind to intragenic enhancers

will lead to a better understanding of transcriptional biology and the genome itself.

References

Agostini, M., Rufini, A., Bampton, E. T. W., Bernassola, F., Melino, G., & Knight, R. A. (2013). The p53 family and stem cell biology. In P. Hainaut, M. Olivier, & K. Wiman (Eds), *p53 in the clinics* (pp. 65-76). Springer. https://doi.org/10.1007/978-1-4614-3676-8_4

Belyi, V., Ak, P., Markert, E., Wang, H., Hu, W., Puzio-Kuter, A., & Levine, A. (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

Cadwell, C., & Zambetti, G. (2001). The effects of wild-type p53 tumor suppressor activity and mutant p53 gain-of-function on cell growth. *Gene*, *277*(1-2), 15-30. https://doi.org/10.1016/s0378-1119(01)00696-5

Catizone, A., Uzunbas, G., Celadova, P., Kuang, S., Bose, D., & Sammons, M. (2020). Locally acting transcription factors regulate p53-dependent cis-regulatory element activity. *Nucleic Acids Research*, *48*(8), 4195-4213. https://doi.org/10.1093/nar/gkaa147

Cooper, D. (2002). Human gene mutation in pathology and evolution. *Journal of Inherited Metabolic Disease*, *25*(3), 157-182. https://doi.org/10.1023/a:1015621710660

Daino, K., Ichimura, S., & Nenoi, M. (2006). Both the basal transcriptional activity of the GADD45A gene and its enhancement after ionizing irradiation are mediated by AP-1 element. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1759*(10), 458-469. https://doi.org /10.1016/j.bbaexp.2006.09.005

Flores, E. (2007). The roles of p63 in cancer. *Cell Cycle*, *6*(3), 300-304. https://doi.org/
10.4161/cc.6.3.3793

Fulda, S., Gorman, A., Hori, O., & Samali, A. (2010). Cellular stress responses: Cell survival
and cell death. *International Journal of Cell Biology*, *2010*, 1-23.
https://doi.org/10.1155/2010/214074

Halfon, M., Carmena, A., Gisselbrecht, S., Sackerson, C., Jiménez, F., Baylies, M., &
Michelson, A. (2000). Ras pathway specificity is determined by the integration of
multiple signal-activated and tissue-restricted transcription factors. *Cell*, *103*(1), 63-74.
https://doi.org/10.1016/s0092-8674(00)00105-7

Harding, H., Zhang, Y., Zeng, H., Novoa, I., Lu, P., Calfon, M., Sadri, N., Yun, C., Popko, B.,
Paules, R., Stojdl, D., Bell, J., Hettmann, T., & Leiden, J. (2003). An integrated
stress response regulates amino acid metabolism and resistance to oxidative
stress. *Molecular Cell*, *11*(3), 619-633. https://doi.org/10.1016/s1097-2765(03)00105-9

Koster, M., & Roop, D. (2004). The role of p63 in development and differentiation of the
epidermis. *Journal of Dermatological Science*, *34*(1), 3-9.
https://doi.org/10.1016/j.jdermsci.2003.10.003

Lane, D. (1992). p53, guardian of the genome. *Nature*, *358*(6381), 15-16.
https://doi.org/10.1038/358015a0

Maston, G., Evans, S., & Green, M. (2006). Transcriptional regulatory elements in the human
genome. *Annual Review of Genomics and Human Genetics*, *7*(1), 29-59.
https://doi.org/10.1146/annurev.genom.7.080505.115623

Papathanasiou, M., Kerr, N., Robbins, J., McBride, O., Alamo, I., Barrett, S., Hickson, I., &

    Fornance, A. (1991). Induction by ionizing radiation of the gadd45 gene in cultured

    human cells: Lack of mediation by protein kinase C. *Molecular and Cellular*

    Biology, *11*(2), 1009-1016. https://doi.org/10.1128/mcb.11.2.1009-1016.1991

Pearson, J., Juarez, M., Kim, M., Drivenes, Ø., & McGinnis, W. (2009). Multiple transcription

    factor codes activate epidermal wound–response genes in *Drosophila*. *Proceedings of the*

    National Academy of Sciences, *106*(7), 2224-2229.

    https://doi.org/10.1073/pnas.0810219106

Reiter, F., Wienerroither, S., & Stark, A. (2017). Combinatorial function of transcription factors

    and cofactors. *Current Opinion in Genetics & Development*, *43*, 73-81.

    https://doi.org/10.1016/j.gde.2016.12.007

Sauna, Z., & Kimchi-Sarfaty, C. (2011). Understanding the contribution of synonymous

    mutations to human disease. *Nature Reviews Genetics*, *12*(10), 683-691.

    https://doi.org/10.1038/nrg3051

Shaw, P. (1996). The role of p53 in cell cycle regulation. *Pathology - Research and*

    Practice, *192*(7), 669-675. https://doi.org/10.1016/s0344-0338(96)80088-4

Tamamori-Adachi, M., Koga, A., Susa, T., Fujii, H., Tsuchiya, M., Okinaga, H., Hisaki, H.,

    Iizuka, M., Kitajima, S., & Okazaki, T. (2018). DNA damage response induced by

    Etoposide promotes steroidogenesis via GADD45A in cultured adrenal cells. *Scientific*

    Reports, *8*(1). https://doi.org/10.1038/s41598-018-27938-5

Vaux, D., & Strasser, A. (1996). The molecular biology of apoptosis. *Proceedings of the*

    National Academy of Sciences, *93*(6), 2239-2244. https://doi.org/10.1073/pnas.93.6.2239

Wang, Y., Schwedes, J., Parks, D., Mann, K., & Tegtmeyer, P. (1995). Interaction of p53 with its

      consensus DNA-binding site. *Molecular And Cellular Biology*, *15*(4), 2157-2165.

      https://doi.org/ 10.1128/mcb.15.4.2157

Wray, G. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews*

      Genetics, *8*(3), 206-216. https://doi.org/10.1038/nrg2063

Yáñez-Cuna, J., Arnold, C., Stampfel, G., Boryń, Ł., Gerlach, D., Rath, M., & Stark, A. (2014).

      Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat

      motifs as general enhancer features. *Genome Research*, *24*(7), 1147-1156.

      https://doi.org/10.1101/gr.169243.113

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

Zaret, K. (2020). Pioneer transcription factors initiating gene network changes. *Annual*

      Review of Genetics, *54*(1), 367-385. https://doi.org/10.1146/annurev-genet-030220-

      015007

Appendix

Table 1. List of DNA Oligonucleotide Primers: Primers ordered and used in various experiments

Primer Name	Sequence	Application
SL1273	FWD: GGTACCTGAGCTCGCTAGCCGTAAGGGACTGGGGGACTG	Gibson Assembly
SL1274	REV: GAGGCCAGATCTTGATATCCCCTCTTAATGTATTTCTAGGGAATTAACAAG	Gibson Assembly
RV3 Forward	CTAGCAAAATAGGCTGTCCC	Sequencing

* All primers ordered from Integrated DNA Technologies (IDT) or are standard primers
* FWD: forward primer; REV: reverse primer
* All primer sequences run 5' - 3'

Table 2. List of Sequences: All sequences, both native and modified that were used in plasmids

Name	Sequence	Application
Native *GADD45A* intragenic enhancer	GTAAGGGACTGGGGGAC TGCAGCCTGCAGGGTAG AGCCCCGGAAGGACGGG AGTCAGGGCTGGGTTGC CTGATTGTGGATCTGTGG TAGGTGGGGGTCAGGAG GGTGGCTGCCTTTGTCCG ACTAGAGTGTGGCTGGA CTTTCAGCCGAGATGTGC TAGTTTCATCACCAGGAT TTTCTGTGGTACAGAACA TGTCTAAGCATGCTGGG GACTGCCAGCAGCGGAA GAGATCCCTGTGAGTCA GCAGTCAGCCCAGCTAC TCCCTACCTACATCTGCA CTGCCTCCCGTGACTAAT TCCTTTAGCAGGGCAGA TTAGATAAAGCCAAATG AATTCCTGGCTCACCCCT CATTAAGGAGTCAGCTT CATTCTCTGCCAGTCAGA GCTAAAAATAGAAATTG TGTAGGAGACAAACCTT GTTAATTCCCTAGAAATA CATTAAGAGG	WT enhancer in pGL4.24
WT p53 binding motif	AACATGTCTAAGCATGC T	*Present in WT enhancer
WT AP1 binding motif	GTGAGTCA	*Present in WT enhancer
GADD45A intragenic enhancer containing both mutated p53 and AP1 binding sites	GTAAGGGACTGGGGGAC TGCAGCCTGCAGGGTAG AGCCCCGGAAGGACGGG AGTCAGGGCTGGGTTGC CTGATTGTGGATCTGTGG TAGGTGGGGGTCAGGAG GGTGGCTGCCTTTGTCCG ACTAGAGTGTGGCTGGA CTTTCAGCCGAGATGTGC TAGTTTCATCACCAGGAT	Double mutant *GADD45A* intragenic enhancer in pGL4.24

	TTTCTGTGGTACAGAAG ATCTCTAAGGATCCTGG GGACTGCCAGCAGCGGA AGAGATCCCTGAGATCT AGCAGTCAGCCCAGCTA CTCCCTACCTACATCTGC ACTGCCTCCCGTGACTA ATTCCTTTAGCAGGGCA GATTAGATAAAGCCAAA TGAATTCCTGGCTCACCC CTCATTAAGGAGTCAGC TTCATTCTCTGCCAGTCA GAGCTAAAAATAGAAAT TGTGTAGGAGACAAACC TTGTTAATTCCCTAGAAA TACATTAAGAGG	
Mutated p53 binding motif	AAGATCTCTAAGGATCC T	p53 mutant in intragenic enhancer *GADD45A* in pGL4.24
Mutated AP1 binding motif	GAGATCTA	AP1 mutant in intragenic enhancer *GADD45A* in pGL4.24
pGL4.38 p53 regulatory element	TACAGAACATGTCTAAG CATGCTGTGCCTTGCCTG GACTTGCCTGGCCTTGCC TT	pGL4.38 p53 RE in pGL4.23