New Targets of p53 Regulation:
From Aging to Immune Response

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Robyn Leary

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Anindya Bagchi PhD

September 2013

© Robyn L. Leary 2013

Acknowledgements

The day I began writing this thesis I opened up a fortune cookie which read: the real meaning of enlightenment is to gaze with undimmed eyes on all darkness. I want to thank everyone who has taught me how to challenge my perceptions of this world in a scientific manner. First and foremost, I wish to thank my friend and advisor, Dr. Anindya Bagchi. Early in my scientific career you listened to my nascent ideas. Sometimes they were vague or incorrect, but you gave me the freedom you gave me to pursue them. I will be ever grateful for the opportunities that you have given me to develop myself as a scientist.

I cannot thank Anindya without also thanking my fellow lab members. Mooney Tseng, I have enjoyed sharing this journey with you. I am proud to have worked with such a talented scientist. Thank you for the many intellectual contributions you have made to this work. I wish you the very best in all your future pursuits. Also, a special thank you to our lab technicians Li Zheng and Jon Zettervall as well as the army of undergraduates over the years in the Bagchi lab. I have been truly blessed to have such kind colleagues. When I reflect on my time spent at the University of Minnesota it makes leaving all the more bittersweet.

Through this process, I have had the great fortune to work with multiple outstanding collaborators. I wish to express my thanks to Dr. Aaron Sarver. Your guidance over the past few years has made me a more mature and confident

scientist. My most productive research was in done in collaboration with you. I especially thank you for your guidance with during the writing process. Next I wish to thank Dr. Yasu Kawakami. Your contributions to my scientific development are invaluable. You are the most through and technically gifted scientist I have ever met. I aspire to reach your level someday. Thank you for taking the time to train me and your many contributions to my experimental design. Also, thank Dr. Branden Moriarity and Surbhi Singhal who contributed greatly to the completion of this work. I look forward to following your future successes. I am grateful to have been surrounded by such outstanding scientists.

I would not be where I am now without the help of my many professional mentors at the University of Minnesota. Dr. David Largaespada and the members of the Largaespada, you have been the biggest supports of the Bagchi Lab from its inception. Thank you David for your open door policy. I admire your knowledge, and appreciate your support in every aspect of my career. I which to express my gratitude to Dr. Reuben Harris and Dr. Louis Mansky at the Institute for Molecular Virology and Dr. Dan Voytas at the Center for Genome Engineering. I am indebted to you for providing me with such spectacular opportunities to present my work and become a member of the scientific community. Thank you Dr. York Marahrens and Dr. Naoko Shima and members of the Marahrens and Shima labs for the many happy memories of our Friday afternoon lab meetings. I enjoyed getting to know you all on a more personal

level. Furthermore, I would also like to thank the rest of my doctoral committee, Dr. Michael O'Connor, Dr. Tim Hallstrom and Dr. Tim Starr. I sincerely appreciate the time you have spent encouraging me to question my work. You have all helped me to become a more rigorous scientist.

Finally, I wish to thank my family and friends who has been strong supporters of all my academic endeavors. Thank you, Stephen, Michele Leary, for being the most loving parents a child could wish for. I wish to especially thank my sister, Lesley Mazzarella. You motivate me to live up to your image of me. I have always admired you for your brilliance. I cannot wait to congratulate you on your doctoral dissertation as well. I also need to express my gratitude to my dear friend, Jessica Gavzer. Your practical advice is always refreshing and helps me keep things in perspective. Finally, a very special thanks to my partner Seth Dauphin. I can't think of anyone that I would rather have shared my time in Minnesota with. I hope to share many happy years together with you all, and I wish you a very happy and prosperous future!

Dedication

I dedicate this thesis to my mother, Michele Leary. You sparked my passion for science as a young girl. I fondly remember your stories about working in a lab. Without you exposure to plants and animals I would have never developed an interest in biology and chemistry. I am here today because of you.

Abstract

These studies combined a variety of molecular biology approaches to further identify novel functions of the universal stress sensor, p53. The first study describes a novel genetic model system to address the effects of p53 dosage in an in vivo context. By developing an allelic series through multiple genetic crosses, I obtained mice with incrementally increasing levels of constitutively active p53. I describe how p53 dosage results in differential physiological outcomes with respect to lifespan, fertility, cell morphology and chromatin structure. Furthermore, I identified novel transcripts of p53 that would encode for isoforms with the potential to alter p53 stability and localization. These findings lead me to propose that p53 has the ability to integrate mechanical stimuli into changes in gene expression. In response to mechanical stimuli, p53 may alter components of the extracellular matrix and chromatin. This change in the spatial organization of the nucleus may be an additional mechanism by which p53 promotes changes in gene expression patterns.

In the second study, I identified novel components of a tumor suppressor pathway with an alternative approach to analyze RNA sequencing data called Annotation Independent Transcriptome Assembly. Using this approach to interrogate the transcriptome of p53+/+ and p53-/- mouse embryonic fibroblasts (MEFs), I identified Mmergln-int among the top genomic loci differentially expressed between p53+/+ and p53 -/- cells. Mmergln-int is a murine

endogenous retrovirus that is transcriptionally upregulated in response to cellular stress and regulated by a highly conserved p53 response element in the LTR. When overexpressed, the envelope of Mmergln-int is sufficient to decrease cell viability. Interestingly, this envelope is present in the genomes of multiple vertebrate species and shares a high degree of homology with the exogenous virus HTLV-1. I observe the upregulation of human endogenous retroviral envelopes in response to p53 activation. I hypothesize that the expression of endogenous retroviruses may serve as a marker of oncogenic stress in prostate cancer with diagnostic potential. Furthermore, endogenous retroviruses may serve as a link between cancer initiation and the activation of an immune response to eliminate precancerous cells.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Transcriptional Regulation

The central dogma of molecular biology states genetic information flows from DNA to RNA to protein (1). It is estimated that the human genome contains between 20,000–25,000 protein-coding genes, however only a fraction of these are transcribed at a given time in each cell (2). Transcription, the synthesis of mRNA from a DNA template, is under tight regulation to ensure correct expression of genes during each phase of the cell cycle, stage of differentiation and environmental condition that is imposed on the cell. Transcriptional regulation occurs at three interconnected levels: the DNA sequence, chromatin structure and through the spatial organization of the nucleus.

Transcriptional Regulation at the Level of the DNA Sequence

The best-studied mode of transcriptional regulation occurs at the level of the DNA molecule. DNA is bound in a sequence specific manner by transcription factors. Transcription factors are multimeric proteins that contain a DNA binding domain. In eukaryotes, a variety of motifs recognize DNA, such as homeodomains, zinc fingers, helix-loop-helix motif or leucine zipper motifs. These protein domains physically interact with the DNA by inserting an alpha helix in the major groove of the DNA. The sequence that is bound by the transcription factor is referred to as the response element. Although the activator and repressor regions of transcription factors lack a characteristic structure, they

regulate gene expression by enhancing or repressing the binding of the transcriptional machinery and chromatin remodelers (3).

Transcriptional Regulation at the Level of Chromatin

The cells that compose multicellular organisms vary, despite the fact that each cell contains a copy of the entire genome within its nucleus. The DNA sequence remains relatively stable throughout the life of the cell with the exception of mutation, recombination and transposition. This indicates that the DNA sequence is not the sole regulator of gene expression. There must an additional mechanism that provides a memory of how a gene is transcriptionally regulated long after the signal which initiated this event has disappeared. The form of gene regulation is epigenetic, and occurs at the level of chromatin (Figure 1) (4).

The accessibility of DNA to serve as a template for transcription provides another level to regulate gene expression. The DNA molecule is not naked within the nucleus. DNA associates with proteins to create a compact, higher order three-dimensional structure. If one were to unravel the DNA of a human cell, its length would measure about 2 meters. The DNA must fit into a nucleus with a diameter of approximately 10μm. Chromatin, the complex of DNA and proteins makes this extraordinary feat possible (3).

The structure of chromatin is compacted highly compact. The basic unit of chromatin, the nucleosome is analogous to a spool of thread. An octamer

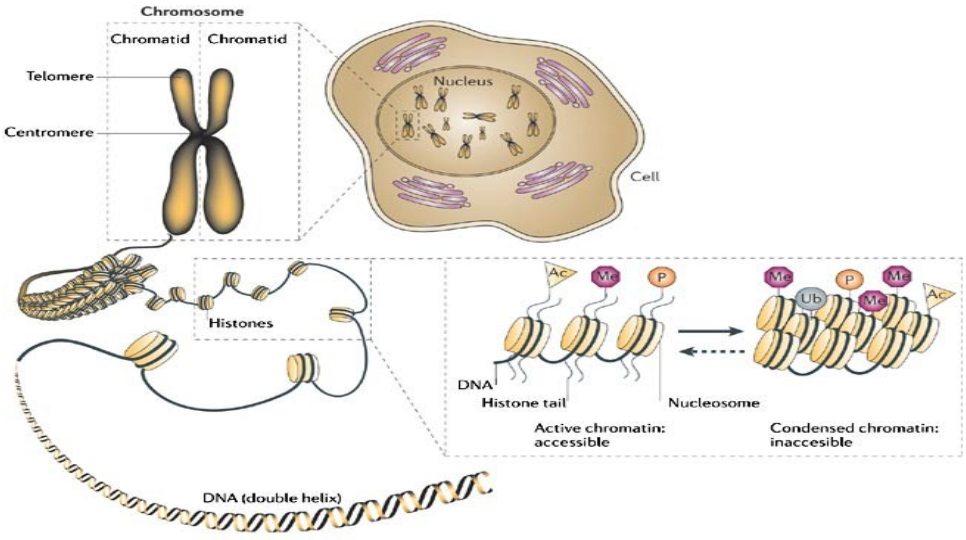


Figure 1. Changes in chromatin structure alter gene expression
The core histones form nucleosomes, which serve as a 'spool' for the DNA 'thread'. Histones undergo posttranslational modifications on tails and globular domains, which regulate chromatin condensation. Condensed chromatin is associated with transcriptional silencing. (from 4)

typically containing two of each histo ne protein H2A, H2B, H3 and H4 form the cylindrical spool-like structure. At the first level of compaction, precisely 146 basepairs of the DNA thread winds around the nucleosome 1.65 times. The tails formed by the amino terminal extensions of each core histone protein protrude to stabilize the DNA wrapped around the octamer. Electron microscopy reveals a structure 10 nm in diameter with the appearance of beads on a string. The 10 nm fiber can also undergo further compaction by the linker histone H1. Histone H1 locks the DNA in place and determines the angles at which the DNA enters and leaves the nucleosome. This creates a fiber with a 30 nm diameter (3). The 30 nm fiber forms large loops which are hypothesized to be organized into distinct chromatin domains (5).

The structure of chromatin directly correlates to its function. The actively transcribed regions of the genome are composed of euchromatin. Euchromatin is formed by the 10 and 30 nm chromatin fibers. Their de-condensed structure promotes transcription by allowing RNA polymerase and transcription factors access to the DNA. On the contrary, the highly compact heterochromatin is associated with low levels of transcription. These regions of the genome are deemed transcriptionally silent, because their structure of does not allow DNA to be exposed to the factors essential for transcription. Thus, the regulation of chromatin structure is essential to the regulation of transcription. Mechanisms to modify the structure of chromatin include methylation of DNA and posttranslational modifications of histones (6, 7).

The posttranslational modification of histones regulates the degree to which chromatin is condensed (8, 9). Modifications at amino- and carboxy-terminal histone tails and histone globular domains include methylation, acetylation, phosphorylation, ubiquitylation, sumoylation and ADP-ribosylation. Posttranslational modifications target specific histones at distinct amino acids. For example the lysine 9 of histone H3 can undergo methylation. Furthermore, the degree of methylation can also vary. Histone 3 can be mono-, di- or trimethylated at lysine 9 (9).

Genome wide studies of histone modifications revealed distinct patterns of specific modifications at genomic regions such as promoters, enhancers and transcribed genes. Posttranslational modifications appear to be associated with specific transcriptional states. For example, the methylation of histone 3 lysine 9, histone 4 lysine 20 and histone 3 lysine 27 are implicated in heterochromatin silencing and tend to be localized in transcriptionally inactive regions of the genome. On the contrary, the methylation of histone 3 lysine 4 and histone acetylation are highly enriched in active promoters and transcribed genes (10).

Many models have been proposed to suggest how posttranslational modifications alter the condensation of chromatin. It has been hypothesized that histone modifications change the net charge of the histones which alters their affinity to DNA. Weak DNA-histone interactions promote transcription. Additionally, the stability of the condensed chromatin structure may be altered by posttranslational modifications that interfere with interactions between adjacent histones proteins (10).

The posttranslational modifications of histones and DNA methylation compose part of a cellular memory system and prevent changes in cell identity. Transcriptional activation and repression through the biochemical modification of histones or DNA is an epigenetic phenomenon. As development proceeds or environmental conditions are altered, patterns of gene expression change and the information provided by the initial stimuli may be lost. However, the transcriptional state of the gene is maintained despite the loss of the initiating signal, due to the epigenetic modifications of DNA and histones (11).

Transcriptional Regulation at the Level of Nuclear Organization

The nucleus does not simply store the genetic information of the cell. It is a highly organized organelle containing numerous subnuclear structures. Many of these nuclear structures are implicated in transcriptional regulation, however the molecular mechanisms regulated by these structures remain unknown. Numerous studies suggest the higher order organization of the contents of the nucleus are nonrandom and it appears the spatial organization of nucleus is an additional mechanism by which gene expression can be regulated (12, 13).

Multiple studies suggest chromosomes inhabit distinct domains within the nucleus referred to as chromosome territories. Early research illustrated that the centromeres and telomeres of plant chromosomes are located at opposite sides of the nucleus (14). Furthermore, ultraviolet irradiation of tiny regions of nucleus illustrated that damage to a single site in the interphase nucleus does not result in damage throughout the cell. The DNA damage was localized discrete region of

the metaphase chromosome, thus chromosomes must be highly compact and restricted to distinct domains within the nucleus and this organization is somehow maintained among cells within a tissue (15). Finally, studies in *Drosophila* mapping the topography of multiple genomic loci revealed that the positions of genes within the nucleus are not random (16). Together, these studies lead to the concept of the chromosome territory.

Technological advances in studying the spatial organization of the nucleus is beginning to provide us with clues as to how the nuclear structure influences gene function (5). It is now possible to generate an “interactome” with technologies such as chromosome conformation capture (3C) or DNA adenine methylation identification (DamID) (17). Furthermore, live cell imaging has demonstrated genes looping out of their chromosome territory when located in domains of constitutively high expression, or upon activation of transcription (18). This looping out of chromatin facilitates transcription at distinct sites of the nucleus.

Although we often think of transcription in a linear fashion in which the RNA polymerase finds the promoter of a single gene, this is likely not the case. Transcription is highly compartmentalized. It can be visualized as discrete foci dispersed throughout the nucleus with bromo-UTP, or with antibody staining against the RNA polymerases. These hot spots of transcriptional activity are deemed transcription factories. It is hypothesized one function of transcription factory is to provide an efficient means to transcribe multiple genes at once (19). However, it has also been proposed that these transcription factories promote

chromosomal interactions through the co-transcription of genes (5). Although the precise mechanisms by which this occurs remain poorly understood, interactome studies are revealing that the sites chromosome territory intermingling occur between DNA sequences that encode the active transcription unit and the binding sites for transcriptional activator or repressor. Thus, it appears the chromatin loop is comprised of an active transcription unit and its respective regulatory motif (17).

It is not only active transcription that is compartmentalized; the nucleus maintains a clear structure for transcriptional repression. First, the orientation of a DNA sequence within the chromosome territory relates to its transcriptional activity. The nucleus is segregation in gene-rich and gene-poor chromatin domains. Transcriptionally silent genomic loci, such as telomeres and centromeres, reside at the periphery of the nucleus at the nuclear lamina (5). These regions of the genome are composed of repetitive DNA sequences which are maintained in a heterochromatic state (20).

Repetitive Elements

Repetitive elements compose a significant portion of the mammalian genome (21). These genomic regions were once considered 'junk DNA' due to the fact that they are extremely gene poor. Today it is evident that some repetitive elements are actively transcribed noncoding RNAs, however the function of most of these transcripts remain largely unknown. Distinct classes of repetitive elements are localized to specific regions of the chromosome (Figure

2) (20). Although repetitive elements are enriched at centromeres and telomeres, they are also distributed throughout the entire mammalian genome, and can be visualized as G and R bands on condensed chromosomes (22). Transposons such as LTR retrotransposons, LINE elements and SINE elements are classified as interspersed repeats. The centromeric region of the chromosome are formed by satellite DNA (20).

Repetitive Elements and Stress

The majority of repetitive sequences in a genome are classified as transposable elements. Transposons are present in every domain of life; they are present in the genomes of bacteria, archaea and eukarya (23). A transposable element is a DNA sequence that can insert itself or copies of itself into a new location of the genome. DNA transposons, such as Tc1/mariner, Mutator, piggyBac , Maverick, Sleeping Beauty, move through the genome through a cut and paste mechanism and make up 2.8% of human genome.

The retrotransposons, which mobilize through a copy and paste mechanism via an RNA intermediate are far more abundant. A conservative calculation indicates the comprise 42.2% of human genome and include families of repeats such as LINEs, SINEs and endogenous retroviruses (24).

Transposable elements are ubiquitous feature of a genome, which are hypothesized to generate genetic diversity in response to stress. Barbra McClintock discovered the transposon through her work on Ac/Ds elements in maize in during the 1940's (25). In her 1984 Nobel Prize lecture she

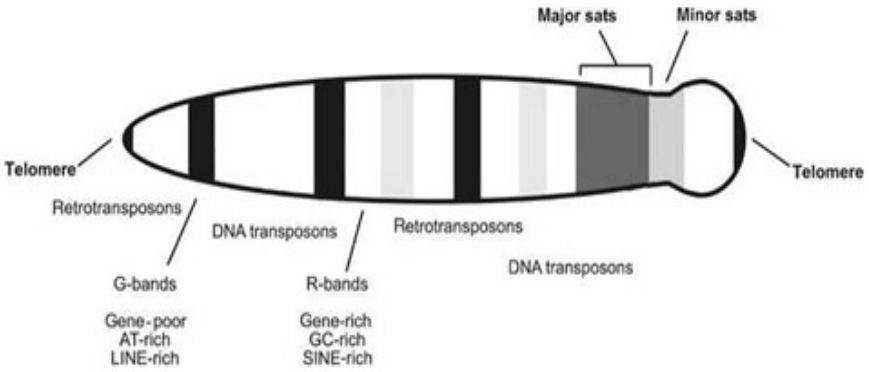


Figure 2. Repetitive elements in the mammalian genome
A schematic representation of a mouse chromosome shows repetitive DNA sequences are distributed throughout the entire mammalian genome. DNA transposons and subtypes of retro- or RNA transposons such as LTR transposons, LINE elements and SINE elements are classified as interspersed repeats. Heterochromatin composed of major and minor satellites compose the pericentromeric and centromeric regions of the chromosome, respectively. (from 20)

hypothesized that the function of the transposable element is to enhance fitness of the organism (26). Since this speech, numerous examples of stress responsive transposition have been reported in bacteria and yeast. For example, in bacteria, depletion of metabolic small molecules like cAMP lead to Group II intron mobilization (27). Furthermore, in yeast a DNA transposon mediates mating type switch in response to nutrient depletion. In yeast approximately 30 DNA damage checkpoint response proteins stimulate Ty1 retrotransposition (23).

In vertebrates our less is known about the role of stress in transposable elements. The mobilization of transposons is highly restricted in higher eukaryotic organisms (28). Nonetheless, expression of these elements is frequently reported. For example, LINEs are transcriptionally upregulated in response to ultraviolet irradiation, gamma irradiation and by selected drugs in human and rat (29, 30). Furthermore, numerous studies report that expression of endogenous retroviruses, the LTR retrotransposons, in disease states such as cancer, MS and lupus (31). The lack of understanding of the function of these elements has lead to repetitive sequences being referred to as the dark matter of the genome (32).

p53: The Universal Stress Sensor

The universal stress sensor belongs to a group of highly conserved transcription factors of p53/p63/p73 gene family (33). An ancestral gene of p53/p63/p73 family first makes an appearance in the early metazoans, specifically the sea anemone. Interestingly, the primary function of this ancestral

gene is to protect the germ line from genomic instability in response to cellular stress. From insects, worms, clams, vertebrates, and humans this function is maintained over one billion years of evolution. A common ancestral gene to most similar to p63/p73 is present in the genomes of almost all invertebrates. In cartilaginous fish, a duplication of the gene p63/p73-like gene occurs which produces a gene closely related to the p53. By the time the bony fish emerge p53, p63, and p73 exist in the genome, and in higher vertebrates diversification of these three transcription factors occurs. It is during this evolutionary period that somatic stem cells with the ability to regenerate adult tissues emerge. It is p53 that takes on the function of protecting these somatic stem cells and progenitor cells from genotoxic stress. Thus, it is first in higher vertebrates that p53 acquires structure and the tumor suppressor function for which it is widely known today.

The human p53 gene is located at chromosome 17p13.1. It is 19,198 nucleotides in length from exon 1 to exon 11. Transcription of the full length isoform begins in the second exon and ends in the 11th exon. This encodes a 393 amino acid protein with a well characterized structure (Figure 3). From residues one to 61 contain starting from the amino terminus, is the transactivation domain (TAD) that is further divided into TAD and TAD2. Residues 22 and 23 critical in this domain because negative regulator of p53, MDM-2, binds there to block transcription and promote the ubiquiniation of the carboxy-terminal domain to promote its degdration (34) and ubiquitinates lysine residues in the carboxy-terminal domain region (35). The TAD is followed by a

Figure 3. The domains of the p53 protein
The tumor suppressor, p53 contains multiple domains with well characterized functions. In murine p53, two amino terminal transactivation domains (TAD 1 and TAD 2) are followed by a proline rich region (PxxP). This region, as well as the carboxyl-terminal domain is critical for the negative regulation of p53. The carboxyl-terminal domain is composed by nuclear localization signal (NLS), oligomerization domain (OD) and basic region (BR). The best established function of p53 is its function as a transcription factor. The oligomerization domain promotes the formation of tetrameric structure which can bind DNA is a sequence specific manner via the DNA binding domain.

proline- rich domain (PRR) at residues 61–94, and this region of the protein is involved in apoptosis (36). Residues 102–292 are critical to the function of p53 as a transcription factor. This region comprises the DNA-binding domain, and is a hotspot for mutations in cancer. The DNA-binding domain recognizes the p53 response (p53 RE), a specific sequence in or near introns of genes which are regulated by p53 (37). This general sequence of the p53 RE is RRRCWWGYYY, where R=A or G, W= A or T and Y=C or T. This RRRCWWGYYY sequence is repeated twice; the prototype p53 RE is comprised of two decameric motifs referred to as half sites. The half sites are separated by a spacer of 0-13 nucleotides.

Like many transcription factors, the p53 protein functions as a multimeric protein. An oligomerization domain (OD) spans amino acid residues 324–355 to enable the formation of a tetramer, the structure which has the capacity to bind DNA in a sequence specific manner (38). The OD is followed by the carboxy-terminal domain (CTD) at residues 356–393. This CTD regulates DNA binding activity and protein stability (39). The amino-terminus and carboxy-terminus of p53 are alternatively spliced to produce distinct isoforms with a diverse functions (40).

As previously mentioned, the best established role of p53 is its function as a transcription factor. In response to stress, the p53 protein is stabilized by a variety of posttranslational modifications that can influence the specificity of the promoter bound by the transcription factor (41). Direct targets of p53 regulation are identified by the presence a p53 RE, which is bound by the p53 tetramer

(37). Subtle changes to this sequence alter the affinity of p53 binding, which influences the expression of each individual p53 target gene. Furthermore, the subset of p53 target genes that are activated during a given stress stimulus are also influenced by tissue specific co-activators and co-repressors recruited by p53 (42). These proteins work together to alter the epigenetic landscape of the target gene promoter and recruit the transcription machinery. Decades of extensive research provide insight in how these factors converge to promote discrete cellular outcomes in order to maintain the integrity of the genome.

Since p53 acts as a transcriptional activator of genes in response to cellular stress, it is intriguing to speculate if this master regulator potentially controls the expression of repetitive elements. Furthermore, is it possible these genomic elements could contribute to p53's tumor suppressor function. Although the identification of p53 target genes remains a mainstay in cancer research, the role of p53 in regulating repetitive regions of the genome remains unknown. Advances in high throughput sequencing technology are just now allowing us to interrogate the expression of repetitive regions the regions of the genome.

Technological Advances in Detecting Expression of Repetitive Elements

Recent technological advances in gene expression profiling are providing researchers with unprecedented opportunities to interrogate genome functions. The development of massively parallel cDNA sequencing, or RNA seq provides a more precise knowledge of a transcriptome over traditional hybridization techniques such as microarrays. RNA Seq provides researchers single base pair

resolution of transcripts with a higher dynamic range of to detect differential gene expression. This is achieved through the elimination of background caused by non-specific hybridization of probes. But the overall goal of transcriptomics is not only to accurately determine differential gene expression, it is also to identify all the different of transcripts present, including non-coding RNAs, small RNA, isoforms generated by alternative splicing and novel transcripts. This is of particular interest when studying cancer transcriptomes, because gene fusion transcripts can be generated by chromosomal abnormalities such as translocations, insertions, deletions or even chromothripsis, the shattering of chromosomes.

The identification of novel transcripts is difficult with conventional RNA Seq analysis methods, because they rely on mapping transcripts to a reference genome. Transcripts that do not map uniquely, because they are map to multiple genomic loci or are not present in the reference genome, may be randomly misaligned to an arbitrary genomic locus or remain unidentified. Thus, a transcript which is not present in a reference genome is essentially equivalent to a probe missing on a microarray. Therefore, the development of user-friendly tools for annotation independent transcriptome assembly is imperative for a better understanding of human disease states such as cancer or aging.

Chapter 2: Statement of Thesis

This thesis combines variety of molecular biology approaches to further identify novel functions of the universal stress sensor, p53. The first study describes a novel genetic model system to address the effects of p53 dosage in an in vivo context. By developing an allelic series through a series of genetic crosses, which obtained mice with incrementally increasing levels of constitutively active p53. We describe how p53 dosage results in differential physiological outcomes with respect to lifespan, fertility, cell morphology and chromatin structure. We identify novel transcripts of p53 which would encode for isoforms with the potential to alter p53 stability and localization. Our findings lead us to propose that p53 has the ability to integrate mechanical stimuli into changes in gene expression. In response to mechanical stimuli, p53 may alter components of the extracellular matrix and chromatin. This change in the spatial organization of the nucleus may be an additional mechanism by which p53 promotes changes in gene expression patterns.

In the second study, I demonstrate that it is possible to identify novel components of a tumor suppressor pathway with an alternative approach to analyze RNA sequencing data called Annotation Independent Transcriptome Assembly (AITA). Using this approach to interrogate the transcriptome of $p53^{+/+}$ and $p53^{-/-}$ mouse embryonic fibroblasts (MEFs), I identified Mmergln-int among the top genomic loci differentially expressed between $p53^{+/+}$ and $p53^{-/-}$ cells. Mmergln-int is a murine endogenous retrovirus of the family Gammaretroviridae. I show how p53 regulates the expression of Mmergln-int, and its expression

mediates cell viability. Notably, I identify p53 responsive endogenous retroviral components with a striking degree of homology to Mmergln-int in the genomes of numerous organisms from zebrafish to humans. In this study I propose a potentially a novel and conserved mechanism by that endogenous retroviruses contribute to p53 mediated tumor suppression.

In the chapter Future Directions, I outline a detailed proposal to further investigate the role of endogenous retroviruses in tumor suppression. This chapter was submitted to the National Cancer Institute as a F32 proposal for the PA-11-113, Ruth L. Kirschstein National Research Service Awards for Individual Postdoctoral Fellows with a submission date of August 8th, 2013.

Chapter 3: An In Vivo Allelic Series to Study the Effects of p53 Dosage

Introduction

Disruption of the p53 signaling pathway is the most commonly observed genetic defect in human cancers (43). P53 functions primarily as a tumor suppressor. A plethora of cellular stressors, like DNA damage or oncogene induced hyperproliferation, lead to the stabilization the p53 protein (44). P53, in turn, acts as a transcription factor to induce the expression of hundreds of downstream genes. The expression of these genes suppress tumorigenesis by inducing diverse signaling cascades to maintain the integrity of the genome. This results in distinct cellular outcomes such as apoptosis, senescence, DNA repair or cell cycle arrest. Thus, the concept of restoring p53 activity to tumors with the goal of correcting aberrations in these intricate processes is an attractive therapeutic option to treat human cancer (45, 46). However, it is not well understood how the precise level of p53 expression is essential to its function.

Some evidence suggests that differential levels of p53 expression result in varying physiological outcomes. One such study knocked down p53 with varying degrees of efficiency in hematopoietic stem cells derived from an Eμ Myc mouse model, then adoptively transferred the stem cells after sublethal irradiation. Interestingly, lymphoma with an increasing degree of severity resulted as the efficiently of the p53 knockdown increased (47). A similar effect is also observed among the human population. Numerous p53 germline polymorphisms have been documented in the human population (48). The clinical outcome of these

polymorphisms varies due to the degree of p53 activity that is lost. Whereas some mutations result in strong cancer predisposition phenotypes due to a non-function p53 protein, others result in proteins with limited functional alterations and low tumor penetrance (49). Thus, when p53 activity is differentially reduced, dramatically different consequences are observed in vivo.

Clearly p53 is a powerful tumor suppressor, and loss of p53 function has detrimental effects due to the inability of the cell to induce apoptosis, senescence and cell cycle arrest in response to a stress stimulus. On the other hand, the physiological effects of enhance p53 activity are perplexing. Mouse models with accelerated aging phenotypes, such as $Terc^{-/-}$, $Ku80^{-/-}$ express high levels of p53, the aging phenotypes are rescued by a $p53^{-/-}$ background (50-52). Furthermore, two p53 hypermorphic mouse models that express either a truncated p53 allele or naturally occurring p53 isoform exhibit an accelerated aging phenotype (53, 54). On the contrary, additional transgenic copies of the p53 gene do not result in significant changes in lifespan (55). Thus, a better understanding of the effect of differential levels of endogenous p53 activity may provide us with additional insight into the function this essential transcription factor.

To better understand the physiological outcome of p53 dosage, we generated an allelic series that results in incremental increases of the p53 transcript and protein in vivo. Here we report that increased p53 activity correlates with distinct physiological changes, decrease in lifespan and loss of fertility. As p53 levels further increase, we see distinct cellular changes associated with progeroid cells and the expression of novel p53 isoforms.

Results

In vivo modulation of p53 activity through an allelic series

To investigate the physiological consequences of varying levels of p53 in vivo, we crossed mice harboring a duplication of a 4.3 Mb allele (dp/+) on chromosome 4 with mice containing p53 loss of function allele as previously described (3, 56). The dp/+ mice bear a duplication on chromosome 4 that is syntenous to human 1p36 that results in the constitutive activation of p53 through the p16/p19 signaling pathway (Figure 1A) (Bagchi) In order to further manipulate p53 levels, mice with and without the dp/+ allele were bred to carry the wildtype p53 allele ($p53^{+/+}$) or be heterozygous ($p53^{+/lsl}$) or homozygous ($p53^{lsl/lsl}$) for an allele which results in the loss of p53 function through in insertion of a stop codon in exon 1 of the p53 gene (Figure 1B) (45). As expected, E18.5 dp/+ $p53^{lsl/lsl}$ mice did not express the p53 transcript (Figure 2A) or protein (Figure 2B). The p53 expression level increased in E18.5 dp/+ $p53^{+/lsl}$ and dp/+ mice in comparison to $p53^{+/+}$mice, respectively.

Next, we wished to determine if the presence of the dp/+ allele leads to the constitutive activation of p53. To this end, we investigated if an increase in p53 protein level correlates with an increase in p53 target gene activation. We evaluated the transcript levels of Mdm2, p21 and Bax in mouse embryonic fibroblasts (MEFs) generated from $p53^{lsl/lsl}$, $p53^{+/lsl}$, $p53^{+/+}$, dp/+ $p53^{+/lsl}$, dp/+ $p53^{+/lsl}$ and dp/+ embryos (Figure 3A). We observe a significant upregulation of these genes in the dp/+ $p53^{+/lsl}$ and dp/+ in comparison to the $p53^{+/+}$ state. To

further validate the constitutive activation of p53 in presence of the dp/+ allele, we investigated the stability of Mdm2. Mdm2 is an E3 ubiquitin ligase responsible for the degradation of p53 through an autoregulatory feedback loop. The phosphorylation of Ser166 stabilizes Mdm2 to regulate p53 activity. In dp/+, $p53^{+/+}$ and dp/+,$p53^{+/lsl}$ mice, we observe higher levels of Mdm2 phosphorlyated at Ser166 comparison to $p53^{+/+}$ MEFs. After observing high levels p21, we evaluated the transcript levels of p16 and p19 in MEFs generated from $p53^{lsl/lsl}$, $p53^{+/lsl}$, $p53^{+/+}$, dp/+ $p53^{+/lsl}$, dp/+ $p53^{+/lsl}$ and dp/+ embryos to confirm the senescent phenotype (Figure 3A).

Since a major function attributed to p53 is cellular senescence, therefore we wished to determine if increasing levels of p53 corresponds with an increase p21 expression level and decreases in cellular proliferation. In our p53 allelic series, we observed that as p53 protein level increased, p21 protein levels demonstrated a similar trend (Figure 3C). Furthermore, we observe a distinct decrease in cellular proliferation with increasing levels of p53 ex vivo (Figure 4A). Thus, with incremental increases in p53, we observed increasing levels of p53 target gene expression and enhanced p53 function with respect to cellular proliferation.

The role of p53 as a central regulator of lifespan remains controversial, whereas some models with enhanced p53 activity exhibit accelerated aging phenotypes other show no significant decrease in lifespan. To better understand how p53 contributes to lifespan, we generated Kaplan-Meier survival curves for mice of the p53 allelic series which were of the same genetic background and

were maintained under identical conditions. We previously reported that dp/+ mice exhibit perinatal lethality on p1 (Bagchi). Mice with the dp/+ p53 $^{+/lsl}$ genotype demonstrate a reduced median lifespan (125.5 days) compared to p53$^{+/+}$ mice (365 days), p53 $^{+/lsl}$ (214 days) dp/+ p53 $^{lsl/lsl}$ (150 days) and p53 $^{lsl/lsl}$ genotype (Figure 4B).

Constitutive activation of p53 results in age-related histopathological lesions in vivo

Although the dp/+ p53 $^{+/lsl}$ mice are viable, the exhibit striking histopathological features associated with a premature aging. The most prominent trait of the dp/+ p53$^{+/lsl}$ mice is an exaggerated rounding of the back that is not observed in other mice of the p53 allelic series. A radiograph detects kyphosis by an abnormal convex curvature of the spine of dp/+ p53$^{+/lsl}$ mice (Figure 5A, left) but not in p53$^{+/lsl}$ mice (Figure 5A, right). Histopathological examination reveals the abnormalities of the thoracic vertebrate of dp/+ p53$^{+/lsl}$ mice. Intervertebral discs protrude ventrally and dorsally to focally compress the spinal cord (Figure 5B, left) which results in the extreme curvature of the spine seen by radiograph. A closer examination of the Intervertebral discs of the dp/+ p53$^{+/lsl}$ mice indicates disc degeneration as seen by chondroid metaplasia of the nucleus pulposus, fragmentation of disc matrix and necrotic chondrocytes (Figure 5C, middle, left). Interestingly, we observe an approximate 3-fold upregulation transcriptional of mmp9, a matrix metalloproteinase in dp/+ MEFs, a marker of early onset disc degeneration in humans (57).

In addition to the pronounced curvature of the back, dp/+ $p53^{+/lsl}$ mice display precocious changes in tissues associated with aging. By 2 months of age, we observe alopecia and loss of whiskers in the dp/+ $p53^{+/lsl}$ mice, but not the other mice of the p53 allelic series. This is accompanied by epidermal hyperplasia. Furthermore, by histopathological analysis we observe atrophy of the male and female reproductive organs. Although mice reach sexual maturity by 8 weeks and all dp/+ $p53^{+/lsl}$ mice studied lived past this age, we never observed offspring from dp/+ $p53^{+/lsl}$ despite the presence of sperm in the vas deferens and ova in the ovaries.

Constitutive activation of p53 induces changes in nuclear morphology associated with progeroid cells

With the goal of further characterizing the effects of enhanced p53 activity, we investigated the cellular morphology of MEFS derived from embryos of the p53 allelic series. After observing a marked decrease in lifespan and early onset of histopathological changes associated with aging, we hypothesized that we may observe nuclear abnormalities associated with progeroid syndromes. To observe the gross nuclear morphology, we visualized the nuclear lamina of p53+/+ and dp/+ MEFs using an antibody against lamin A/C. Cross sections of the dp/+ nuclei reveal nuclear blebbing (Figure 6A). One feature of progeroid cells are changes in the distribution of trimethylated H3K9 hetrochromatin at the nuclear periphery. To asses the distribution of this heterochromatic marker, we co-stained the MEFs with H3K9me3 and lamin A/C. In dp/+ MEFs we see de-

locatization of trimethylated H3K9 heterochromatin in comparison to +/+ MEFs (Figure 6B).

To quantitatively assess the p53 role in the distribution of chromatin at the nucleus, we used transmission electron microscopy to measure the average number of foci and average thickness of the heterochromatin at the nuclear periphery (Figure 7A,B). The heterochromatin at the nuclear periphery of $p53^{+/+}$, $p53^{+/lsl}$ and $p53^{lsl/lsl}$ MEFs dose not demonstrate any statically significant changes in thickness with an average thickness of 107.93 ±25.34 nm, 104.71 ±20.64 nm, 96.82 ±23.64 nm, respectively. Furthermore, the number of heterochromatin foci does not vary in number in $p53^{+/+}$, $p53^{+/lsl}$ and $p53^{lsl/lsl}$ MEFs with an average of 28.6 ±9.74, 24.9 ±8.60, 27.2 ±6.99 foci per cell, respectively. However in dp/+ MEFs we observed profound changes in nuclear architecture. We observed a complete loss of heterochromatin foci with only 0.2 ±0.42 foci per cell counted ($p= 3.12616x10^{-08}$). Compared to $p53^{+/+}$ MEFs, we observed a significant reduction of heterochromatin at the nuclear periphery with an average thickness of 20.81 ±9.96 nm comparison to $p53^{+/+}$ MEFs ($p= 7.44x\ 10^{-9}$). Interestingly, this defect in nuclear structure is rescued in the absence of p53. We counted an average of 28.0±10.88 foci per cell in dp/+ $p53^{lsl/lsl}$ MEFs and obtained an average thickness of 106.14 ±23.59 at the nuclear periphery. An intermediate phenotype was observed in dp/+ $p53^{lsl/lsl}$ MEFs. The average thickness measured 35.91 nm which was a significant reduction when compared to $p53^{+/lsl}$ MEFs ($p= 6.33x10^{-08}$). The average number of foci counted per cell, 11.3 ±5.29, was also significant ($p= 4.70x\ 10^{-4}$). These results taken together demonstrate

that as p53 activity increases, loss of heterochromatin at the nuclear periphery also increases.

dp/+ MEFs express novel isoforms of p53

Global increases in the expression isoforms has been observed in aging tissues of humans and rodents (58, 59), and p53 isoforms are expressed in aging cells (54). Therefore, we wished to investigate what p53 transcripts are expressed in dp/+ MEFs that could potentially contribute to the progeroid phenotype. To this end, we generated cDNA from dp/+ and $p53^{+/+}$ MEFs for nested RT-PCR. We designed a forward PRC primer in exon 11 which was used for each reaction and staggered the reverse primers to amplify exons 1, 4, 8, and and 11 (Figure 8A). In dp/+ MEFs using primers spanning exons 4-11, 8-11 and 10-11 and 11, we detected PCR products of the expected size were generated from +/+ (data not shown) and dp/+ MEFs (Figure 8B, lanes 1-7, top bands lanes 3 and 6). We subsequently cloned and sequenced the PCR products and determined they correspond to the full length p53 transcript. In dp/+ MEFs, we detected PCR products smaller then the expected size using primers to span exons 8-11 and 11. The cloning and sequencing of the lower PCR bands revealed a 721 bp transcript in 18 out of 18 clones (Figure 8B, lanes 3 and 6, lower bands). We refer to this transcript as dp/+ p53 isoform 1.

The mechanisms which generate p53 isoforms include alternative splicing, usage of an internal promoters and alternative initiation of translation (60). Based on strategic primer design we determined that the transcript of dp/+ p53

isoform 1 initiates in exon 8 at nucleotide position 950. To confirm the transcriptional start site, we generated a primer in exon 8 upstream of nucleotide 950 at position 959-970. No lower band was detected suggesting that the initiation of transcription occurs via an alternative promoter residing within exon 8 (Figure 8B, lane 4). Additionally, primers in exons 1 and 4 do not amplify the isoform (Figure 8B, lanes 1, 2, and 7). To further validate the presence of this transcript, we designed a sequencing primer specific to dp/+ p53 isoform 1. The 721 bp transcript was detected in cDNA from dp/+ MEFs but not $p53^{+/+}$ MEFs. To confirm our sequencing results which reveal the transcript is alternatively spliced at nucleotide positions 1265-1486, we designed a primer which resides within the spliced region at position 1270-1292 in exon 11. This primer does not amplify the isoform (Figure 8B, lane 4). The translation of this isoform would result in an N-terminal truncation of the p53 protein with alternative CTD (Figure 9A,B).

Next, we used a similar nested PCR strategy as described above, but staggered the reverse primers to amplify exons 1, 5, 6, and 11 (Figure 10A). Using primers spanning exons 1-11, 6-11, 5-11 and 11, we generated PCR products of the expected side that were confirmed to contain full length p53 by cloned and sequencing (Figure 10B lanes 1-7 upper band). Additional products were obtained using primers spanning 5-11 and 6-11 and 1-11 (Figure 10B, lanes 3-7, lower bands). From 14 out of 14 clones sequenced, we detected an additional p53 isoform we refer to as dp/+ p53 isoform 2. This transcript is identical to full length p53 until exon 6 at nucleotide position 708. The isoform is

then alternatively spliced at adding an additional 76 bp of sequence, then transcribed in frame at nucleotide position 709-735. The transcript is spliced again at nucleotide position 1412 in exon 11, yielding a 1088 bp transcript we named dp/+ p53 isoform 2. The translation of this isoform would create a C-terminal truncation of p53 to yield an isoform with an alternative CTD.

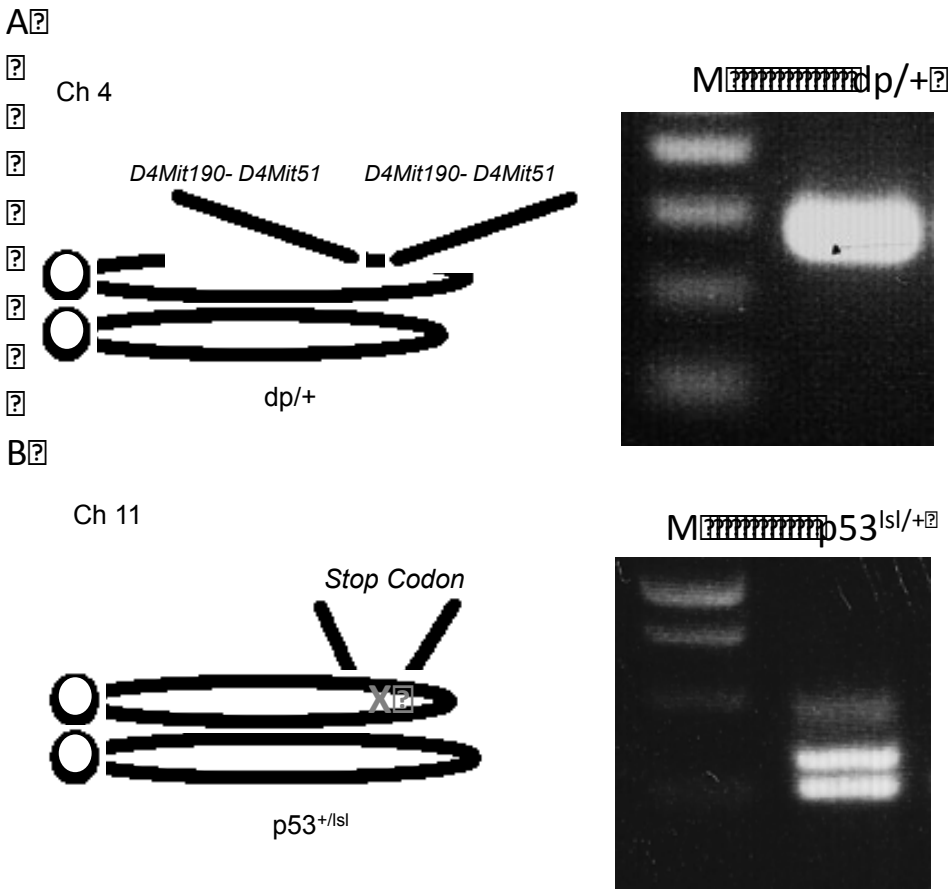


Figure 4. Modified alleles to modulate p53 dosage

(A) Schematic representation of the duplication (dp/+) and p53 loss of function allele. The presence of the dp/+ allele yields a 279 bp band (top left).
(B) The p53 status was determined by the presence of a 364 bp p53 wildtype allele (bottom, top band) and a 278 bp lox stop lox allele (lsl).

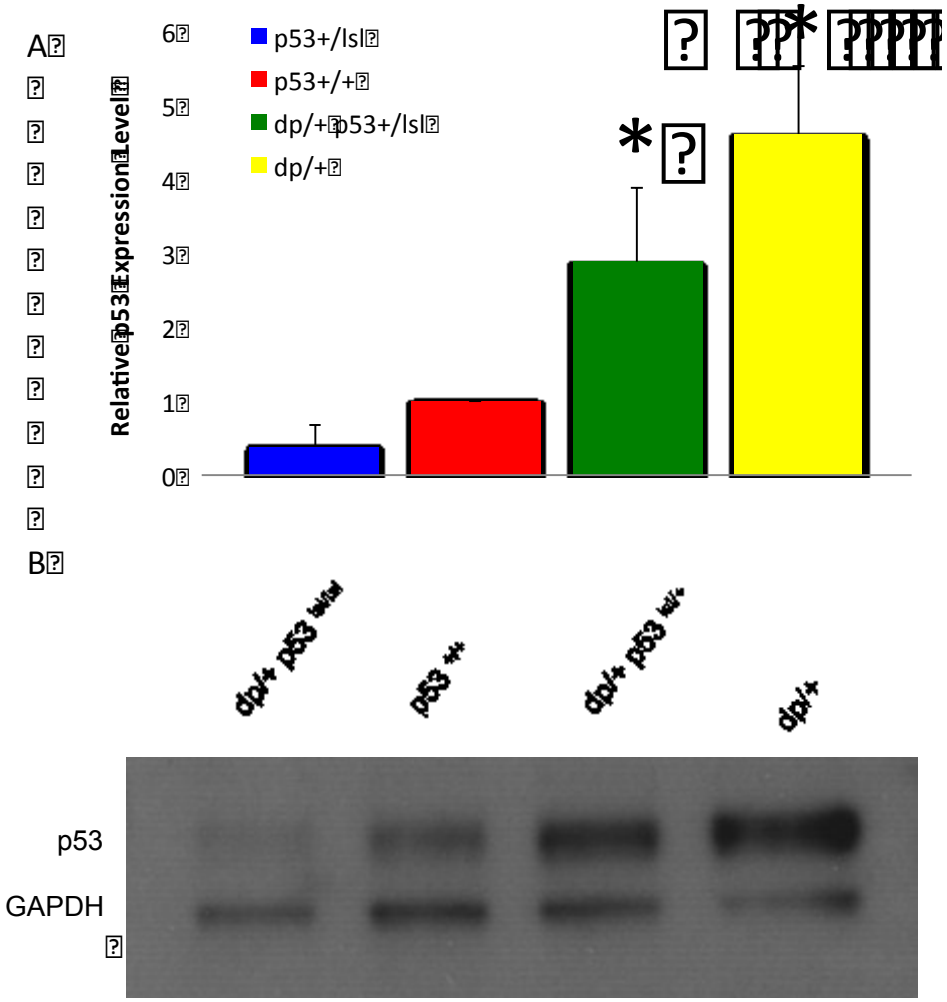


Figure 5. An in vivo murine allelic series of p53 dosage

(A) p53 transcript level was determined in MEFs of the indicated genotypes in bar graph. An astrix denotes p>0.005.
(B) p53 protein level was determined in E18.5 mice of the indicated genotypes by western blot.

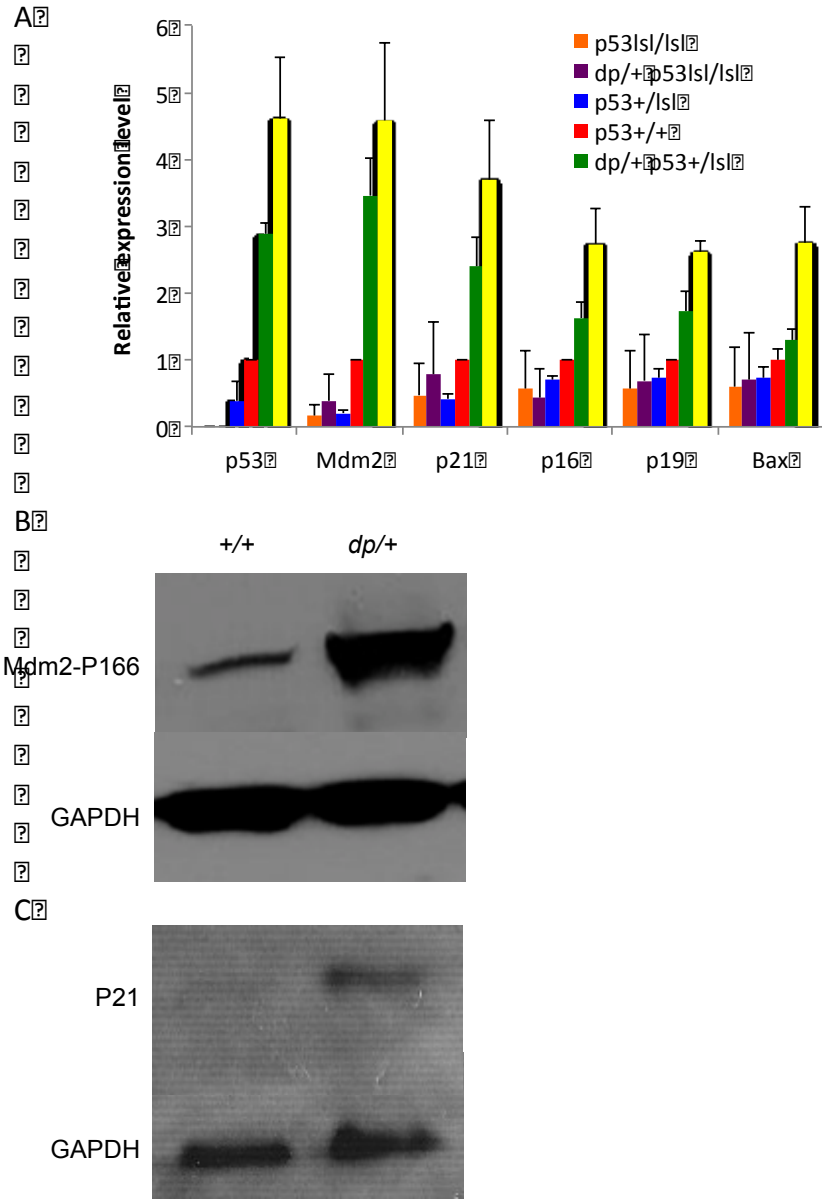


Figure 6. Increasing p53 levels correlates with increasing levels of p53 target gene activation and markers of senescence

(A) p53 target gene transcript levels were determined in MEFs of the indicated genotypes in bar graph.
(B) Mdm2 phosphoserine 166 level was determined in E18.5 mice of the indicated genotypes by western blot.
(C) p21 expression level was determined in E18.5 mice of the indicated genotypes by western blot.

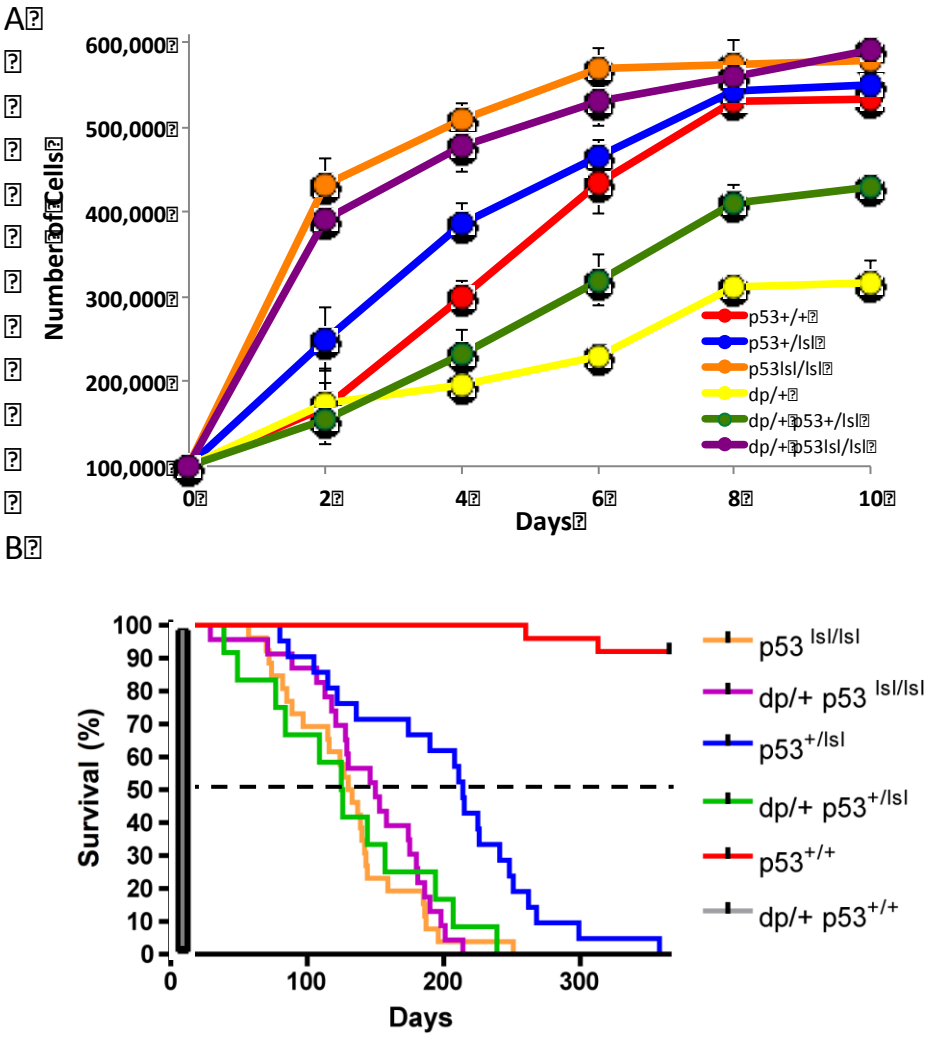


Figure 7. Increasing p53 levels correspond with decreases in cellular proliferation and lifespan

(A) MEFs rates of cellular proliferation
(B) Kaplan Meier survival curve of mice of the p53 allelic series. Dashed line at 50% is the median survival

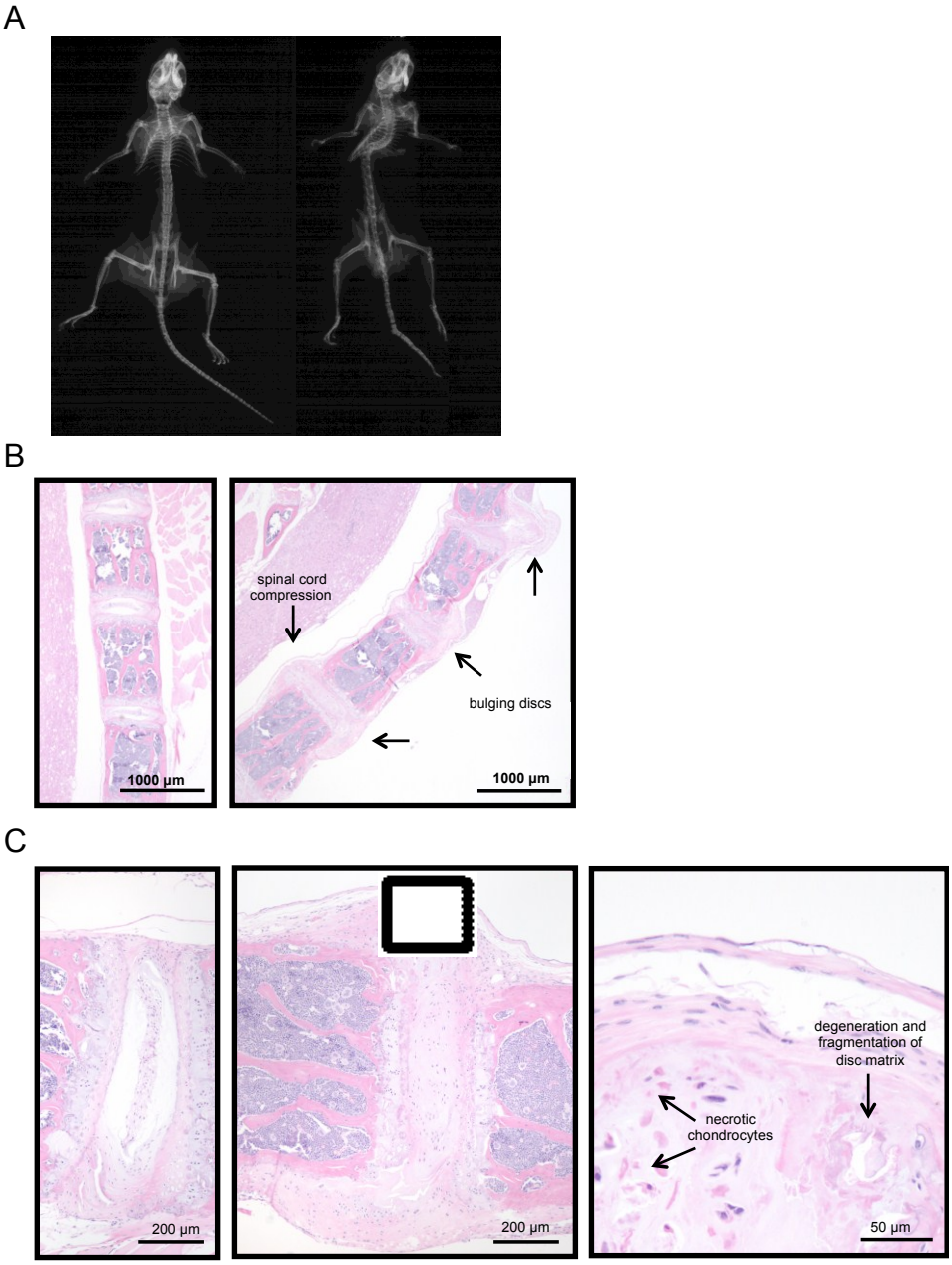


Figure 8. dp/+ p53 $^{lsl/+}$ mice exhibit kyphosis and intervertebral disc degeneration

(A) Radiograph images of p53lsl/+ (left) and dp/+ p53l$^{sl/lsl}$ (right) mice.
(B) Thoracic vertebrae of p53lsl/+ (left) and and dp/+ p53l$^{sl/lsl}$ (right) mice.
(C) Intervertebral discs of p53$^{lsl/+}$ (left) and and dp/+ p53$^{lsl/lsl}$ (middle) mice. Box with dotted lines depicts enlarged image (left).

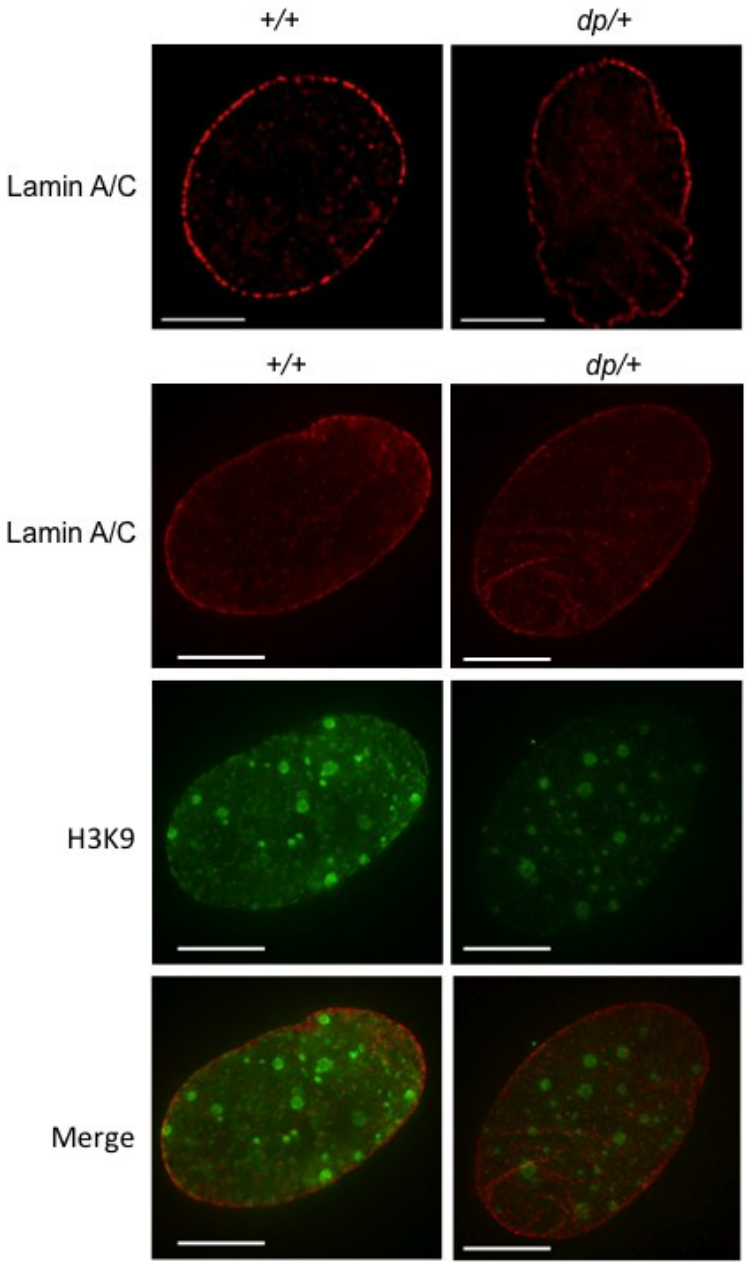


Figure 9. Constitutive activation of p53 results in nuclear abnormalities associated with progeroid syndromes

(A) Midline Z-stack of nuclear envelope of p53+/+ and dp/+ MEFs
(B) Localization of trimethylated H3K9 in p53+/+ and dp/+ MEFs

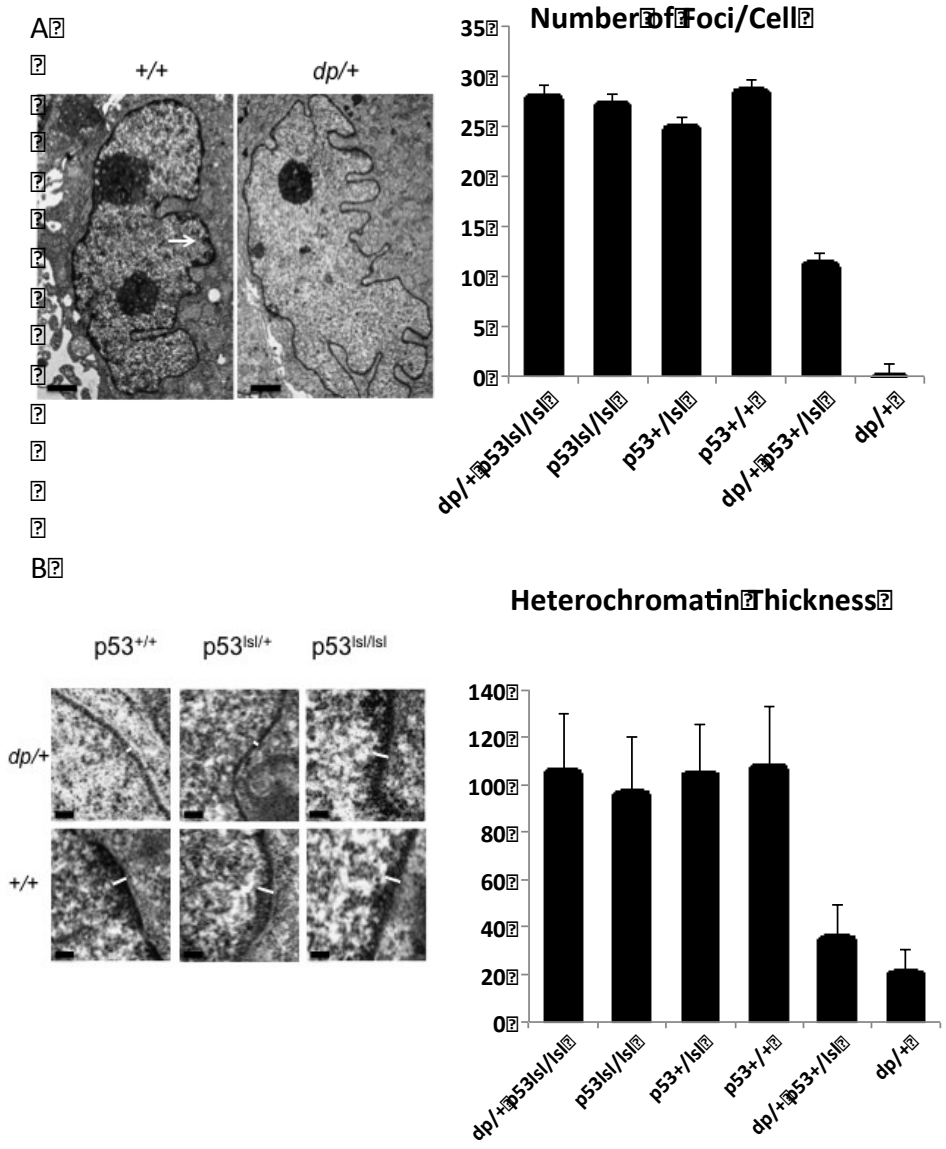
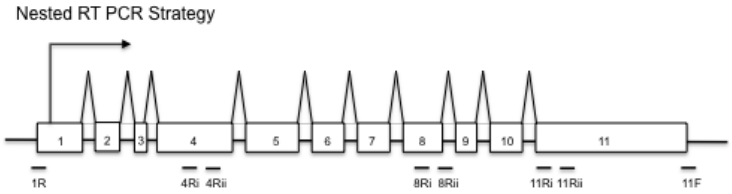


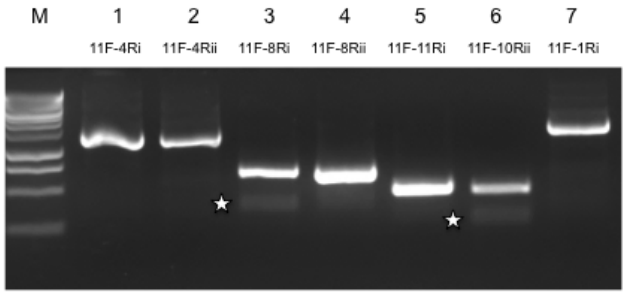
Figure 10. Constitutive activation of p53 results in loss of heterochromatin

(A) TEM images of p53+/+ and dp/+ MEFs for quantification of hetrochromatic foci
(B) TEM images and quantification of peripheral heterochromatin in p53+/+ and dp/+ MEFs

A.


Nested RT PCR Strategy
1
2
3
4
5
6
7
8
9
10
11
1R
4Ri
4Rii
8Ri
8Rii
11Ri
11Rii
11F

B.


M
1
2
3
4
5
6
7
11F-4Ri
11F-4Rii
11F-8Ri
11F-8Rii
11F-11Ri
11F-10Rii
11F-1Ri

C.

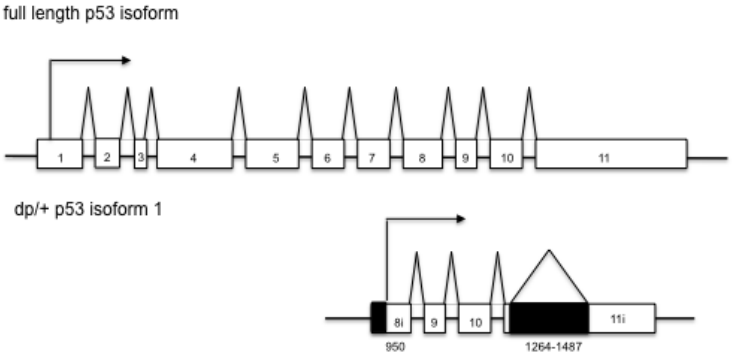

full length p53 isoform
1
2
3
4
5
6
7
8
9
10
11
dp/+ p53 isoform 1
8i
9
10
11i
950
1264-1487

Figure 11. Nested RT PCR Detects an N-Terminal Truncated Isoform of p53

(A) Nested PCR strategy with to enrich for p53 isoforms. Primers flanking exons 1, 4, 8 and 11 were designed to identify isoforms of p53
(B) Nested PCR detects the expression of an alternatively spliced isoform of murine p53. Primers spanning exons 8-11 at nucleotide positions 948-1673 (lane 3, bottom band) and 10-11 (lane 6, bottom band) enrich for a transcript which originates at position 950. This transcript is not detected by using upstream primers in exon 8 (lane 4) at nucleotide position 959-970, exon 4 (lanes 1 and 2) or exon 1. The transcript is not amplified using primer in exon 11 at nucleotide position 1270-1292 (lane 5) but can be amplified with an upstream primer at in exon 10 at nucleotide position 1246-1267. The bands and upper bands in lanes 1-7 correspond to the full length p53 transcript.
(C) The *dp/+* p53 isoform 1 originates from an internal promoter in exon 8 and is alternatively spliced in exon 11. The transcript of *dp/+* p53 isoform 1 initiates at an alternative promoter within exon 8 which results in a 19 bp 5' truncation of exon 8. Exon 11 is alternatively spliced at nucleotide positions 1264-1487 resulting in a 223 bp truncation of exon 11

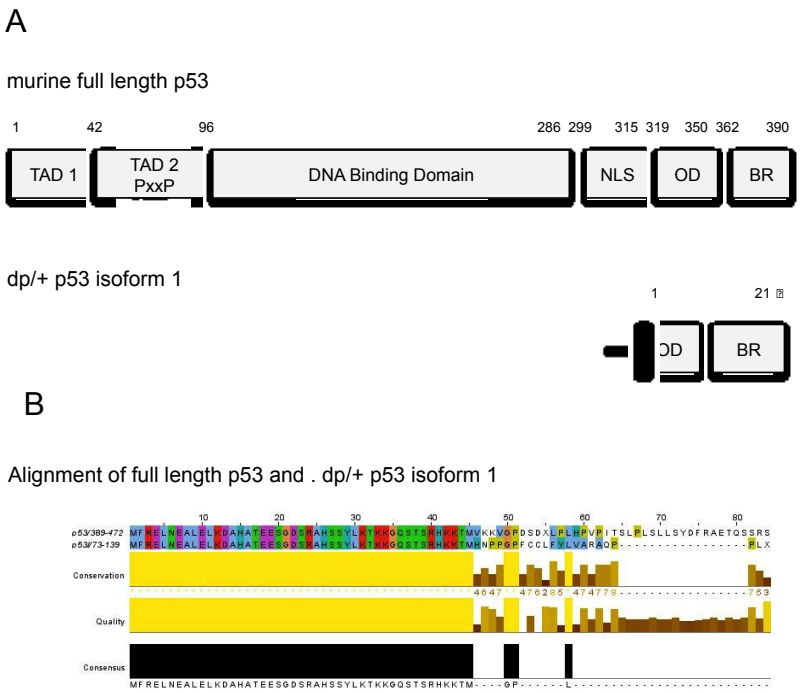


Figure 12. N-terminal truncation of dp/+ p53 isoform 1 yields protein with alternative CTD

(A) Full length p53 contains two transactivation domains (TAD 1, TAD 2) one of which contains a proline rich domain (PxxP), a DNA binding domain, nuclear localization domain (NLS), oligomerization domain (OD) and basic region (BR). The translation of dp/+ p53 isoform 1 results in a protein with a truncated amino terminus that contains only the oligomerization domain and the basic region

(B) The translation of dp/+ p53 isoform 1 generates a protein with an alternative CTD

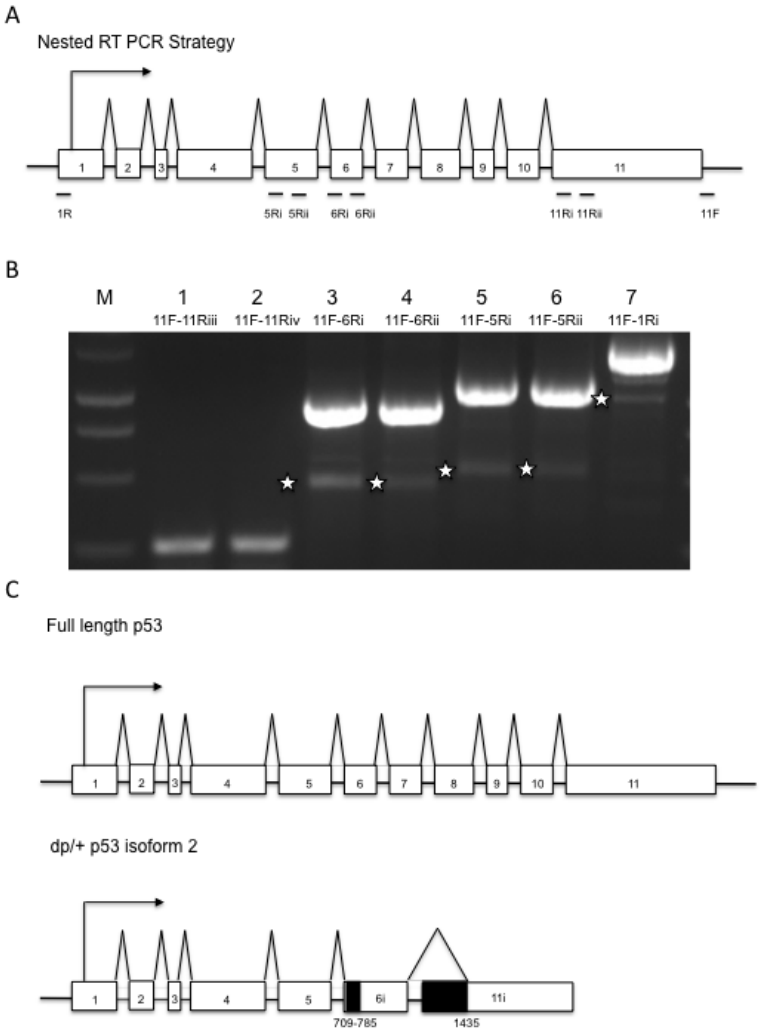

A
Nested RT PCR Strategy
1
2
3
4
5
6
7
8
9
10
11
1R
5Ri 5Rii
6Ri 6Rii
11Ri 11Rii
11F
B
M
1
11F-11Riii
2
11F-11Riv
3
11F-6Ri
4
11F-6Rii
5
11F-5Ri
6
11F-5Rii
7
11F-1Ri
C
Full length p53
1
2
3
4
5
6
7
8
9
10
11
dp/+ p53 isoform 2
1
2
3
4
5
6i
11i
709-785
1435

Figure 13. Nested RT-PCR Detects a C-Terminal Truncated Isoform of Murine p53

(A) Nested PCR strategy with to enrich for p53 isoforms. Primers flanking exons 1, 5, 6 and 11 were designed to identify isoforms of p53

(B) Nested PCR detects an additional alternatively spliced isoform of murine p53. Primers spanning exons 5-11 (lane 3 and 4, bottom bands) and 6-11 (lane 5 and 6, bottom bands) and exons 1-11 (lane 7, bottom band) enrich for a transcript that is not detected by using primers enriching for the 3' end of exon 11 at nucleotide positions 1433-1454 and 1469-1489 (lanes 1 and 2). The bands and upper bands in lanes 1-7 correspond to the full length p53 transcript

(C) The *dp/+* p53 isoform 2 is alternatively spliced in exon 6 to 11. The transcript of *dp/+* p53 isoform 1 is alternatively spliced at the 3' end of exon 6 at nucleotide positions 708 which adds 76 nucleotides of intronic DNA. The transcript is transcribed from nucleotide 709- 735 then spliced again to exon 11 at nucleotide position 1412

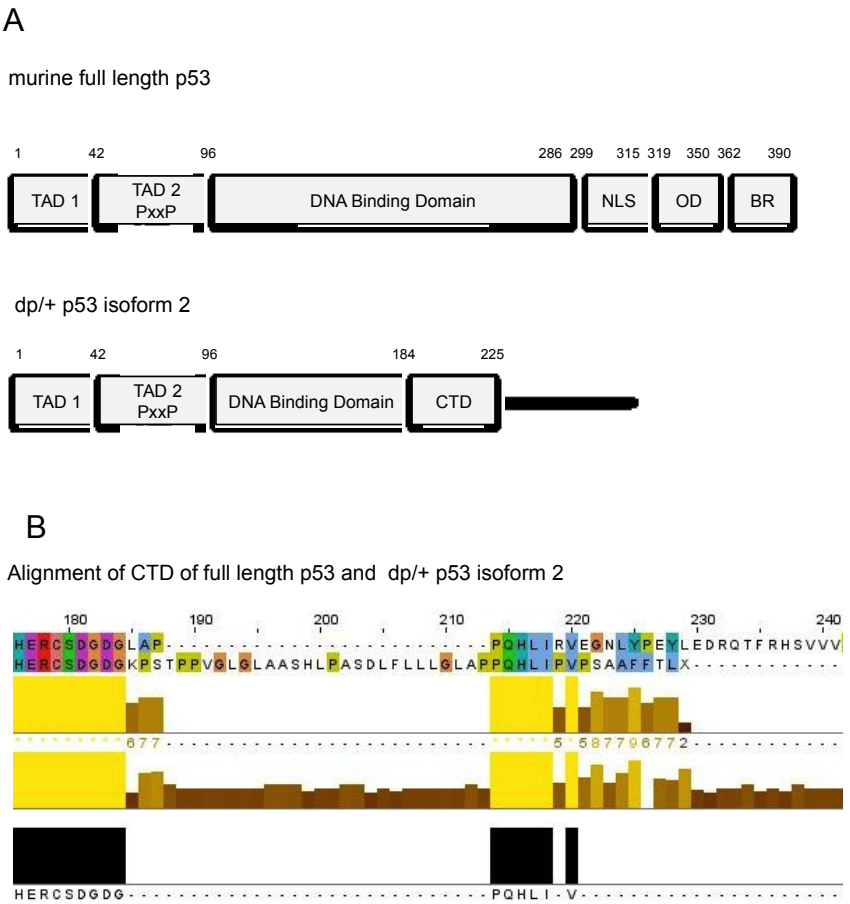


Figure 14. C-terminal truncation of dp/+ p53 isoform 2 yield alternative CTD

(A) Full length p53 contains two transactivation domains (TAD 1, TAD 2) one of which contains a proline rich domain (PxxP), a DNA binding domain, nuclear localization domain (NLS), oligomerization domain (OD) and basic region (BR). The translation of dp/+ p53 isoform 2 creates in a protein with a truncated carboxyl terminus that contains only half of a DNA binding domain region and a deleted NLS, OD and BR
(B) The translation of dp/+ p53 isoform 1 generates a protein with an alternative CTD

Primer	Position	Sequence
1Ri	**158-179**	**ATGACTGCCATGGAGGAGTCA**
5Ri	**641-661**	**AAGTCACAGCACATGACGGAG**
5Rii	**684-702**	**ATGAGCGCTGCTCCGATGG**
6Ri	**713-733**	**GCTCCTCCCCAGCATCTTATC**
6Rii	**764-785**	**CTGGAAGACAGGCAGACTTTTC**
11Riii	**1433-1454**	**TTACCTTGTAGCTAGGGCTCAG**
11Rviii	**1469-1489**	**AGTGGTTCCTGGCCCAAGTT**
11F	**1637-1659**	**AATGGAAGGAAAGTAGGCCCTG**

Table 1. PCR primers for RT Nested PCR to identify *dp/+* p53 isoform 1

Primer	Position	Sequence
1Ri	158-179	ATGACTGCCATGGAGGAGTCA
5Ri	641-661	AAGTCACAGCACATGACGGAG
5Rii	684-702	ATGAGCGCTGCTCCGATGG
6Ri	713-733	GCTCCTCCCCAGCATCTTATC
6Rii	764-785	CTGGAAGACAGGCAGACTTTTC
11Riii	1433-1454	TTACCTTGTAGCTAGGGCTCAG
11Rviii	1469-1489	AGTGGTTCCTGGCCCAAGTT
11F	1637-1659	AATGGAAGGAAAGTAGGCCCTG

Table 2. PCR primers for RT Nested PCR to identify *dp/+* p53 isoform 2

Discussion

In this article, we report a novel model system to study the effects of incremental increases of p53 activity. We hypothesized that with varying degrees of p53 expression, we would see distinct phenotypes attributed to differential p53 dosage. We previously reported that MEFs harboring a duplication of an allele syntenous to human 1p36 exhibited elevated levels of p53 (56). Through a series of genetic crosses, we obtained mice which harbored the duplication allele in combination with either wildtype p53 or heterozygous or homozygous an allele which results in p53 loss of function (45). First, we confirmed that the dp/+ $p53^{+/+}$ and dp/+ $p53^{lsl/+}$ embryos express increasing levels of the p53 protein and transcript, respectively. The presence of the dp/+ allele correlated with robust upregulation of p21 and Mdm2 phosphorylation of serine 166 in suggesting the dp/+ allele results in the constitutive activation of p53.

Next, we wished to determine if we could identify functions of p53 which behave in a dosage dependent manner. To this end, we examined the expression of p53 target genes associated with senescence and apoptosis. We evaluated the transcript levels of established p53 target genes by p21, Mdm2, Bax in MEFs generated from the p53 allelic series. As p53 levels increased, we observed increases in p21, Mdm2 and Bax. Dosage dependent behavior of apoptosis genes Puma and Noxa was not observed. Our finding may suggest that some p53 functions, like senescence, are sensitive to p53 dosage but others, such as apoptosis, may function in an on or off manner. To further test the consequences of p53 dosage with respect to cellular senescence, we

measured rates of cellular proliferation in early passage MEFs. As p53 levels increased, rates of cellular proliferation decreased.

After establishing that incremental increases in p53 dosage correspond with an increasing grade of cellular senescence ex vivo, we set out to investigate the in vivo repercussions. We hypothesize that one molecular basis of organismal aging stems from the accumulation of lesions in the genome. The aftermath of these lesions is an increasing basal level of p53. As p53 escalates, so does senescence. Stem cells pools in tissues decline and degenerate tissue cannot be repaired which eventually leads to the demise of the organism. This concept places p53 as a central regulator of aging.

Our data supports the hypothesis that p53 activity leads to decline a reduction in lifespan and tissue homeostasis. The lifespan of dp/+ $p53^{lsl/lsl}$ mice does not differ from $p53^{lsl/lsl}$ mice. However, the lifespan of dp/+ $p53^{lsl/+}$ mice is significantly shorter than wildtype mice, suggesting that elevated p53 indeed correlates with a shorter lifespan. Furthermore, we observe that elevated p53 is also associated with tissue degeneration. Although dp/+ mice die on postnatal day 1, perinatal lethality is rescued in dp/+ $p53^{lsl/lsl}$ and dp/+ $p53^{+/lsl}$ mice. Whole body histopathology of sex matched littermates subjected to identical environmental conditions reveals tissue degeneration in dp/+ $p53^{+/lsl}$ mice but not $p53^{+/lsl}$ controls.

We observe tissue degeneration at multiple sites of the body. The most significant lesions are detected in the intervertebral discs. A few recent studies support our findings (61, 62). In a progeroid mouse model deficient for the DNA

repair endonuclease, ERCC1 defects of the intervertebral discs were observed as a result of genotoxic stress post irradiation. A reduction of proteoglycans in the intervertebral discs due to upregulation of matrix degrading proteases as well as a decrease in the production of new disc matrix contributed to this phenotype. It is likely p53 is constitutively activated in this mouse model, and that p53 contributes to this phenotype. Furthermore, degeneration of male and female reproductive tissues and loss of fertility were seen in all dp/+ $p53^{lsl/+}$ mice. A body of literature supports our data demonstrating p53 levels influence reproduction (63, 64).

The decrease in lifespan and early onset of age related tissue degeneration observed in the dp/+ $p53^{lsl/+}$ mouse is characteristic of a segmental progeriod phenotype. One of the most striking cellular phenotypes is observed in progeroid syndromes such as Hutchinson-Gilford progeria syndrome, Emery–Dreifuss muscular dystrophy (65). Progeroid fibroblasts exhibit nuclear blebbing and gross changes in heterochromatin structure (66, 67). These changes include loss or delocalization of tri-methyated H3K9 heterochromatin (68). Therefore, we decided to investigate the effects of p53 activity on nuclear morphology. Single z-stacks through the midline of dp/+ MEFs imaged with an antibody against lamin A/C demonstrate abnormal nuclear morphology reminiscent of blebbing seen as in some laminopathies. We observe delocalization of tri-methyated H3K9 heterochromatin.

To quantitatively assess how p53 dosage may contribute to the structure of perinuclear heterochromatin, we determined the average thickness and

number of foci at the nuclear periphery. In dp/+ MEFs we see a profound reduction in thickness and essentially no foci are observed. The dp/+ $p53^{lsl/+}$ MEFs show an intermediate phenotype which is completely rescued in dp/+ p53 $p53^{lsl/lsl}$ MEFs. Thus increasing levels of constitutively activate p53 result in nuclear defects of heterochromatin attributed with progeroid syndromes. We hypothesized the p53, a key cell cycle regulator may regulate the structure of nuclear lamina by inducing its disassembly, and that this may be an underlying cause of the nuclear deformities. However, a change in lamin a/c levels or phosphorylated lamin a/c was not detected by western blot. The underlying mechanism by which p53 dose influences global chromatin structure are yet to be determined.

We know that chromatin changes during aging, and it is intriguing to imagine how transcription factors like p53 could contribute to this. Aged cells show loss of heterochromatin repressive histone marks and even nucleosome free regions of DNA, and the composition of core histones and histone variants change (66, 69-71). Although the exact molecular mechanisms that lead to these changes are unknown, our epigenetic data from the allelic series suggests p53 may be involved. We know that at a local level, p53 directly interacts with chromatin modifying proteins (72). At the global level, less in known about how p53 enforces higher-order chromatin structure and nuclear architecture as a cell ages. In response to genotoxic stress, p53 unwinds chromatin to repair DNA lesions (73). Also, DNA damage triggers redistribution of chromatin remodelers like histone deacetylase SIRT1 (74). It is possible that too much p53 may directly

lead to abnormal patterns of histone modifications and changes in higher-order chromatin structure.

Finally, we wished to determine what p53 transcripts generated in dp/+ MEFs. Genome wide transcriptome studies report global changes in transcription of aging cells due to defective splicing. Furthermore, numerous isoforms of p53 have been detected, and naturally occurring isoforms of p53 are associated with an accelerated aging phenotype (54). Therefore, we used a nested PCR approach to determine what isoforms of p53 are expressed in dp/+ MEFs. With this method, we detected two novel isoforms of the p53 transcript. By looking at what domains are modified we can speculate at the potential function of these isoforms.

The translation of the p53 isoforms in dp/+ MEFs could potentially contribute to the observed phenotypes. If dp/+ p53 isoform 1 is translated, it may have enhanced stability compared to the full length isoform of p53. Translation of dp/+ p53 isoform 1 would generate an amino terminal truncated protein with a CTD that differs in from full length p53 in length and amino acid composition (Figure 3A). Both the carboxy and amino terminus of p53 are essential in regulating p53 stability by Mdm2. The stability of p53 occurs through the binding of Mdm2 in the transactivation domain which leads to the ubiquinitation of the CTD. If dp/+ p53 isoform 1 maintains the ability to interact with full length p53 in a tetramer, it could prevent Mdm2 mediated degradation.

The expression of the isoforms could also alter the cellular localization of p53. The alternative splicing of exon 6 to 11 of dp/+ p53 isoform 2 would result in

a protein with an alterative carboxyl terminal domain. This isoform would also contain a truncated DNA binding domain and also lacks a NLS and oligomerization domain. The lack of NLS in the dp/+ p53 isoform 2 suggests cytoplasmic localization. Cytoplasmic p53 has the ability to interact with actin, and these interactions increase during cellular stress.

Together, these findings that with increasing dose of p53 we see alterations in the extracellular matrix and nuclear envelope leads us to ponder the role of p53 as a sensor of mechanical stress. The cytoskeleton is coupled to the nuclear envelope and chromatin to provide a link for the transmission of mechanical signals in response to changes in the extracellular matrix (75). Mechanical forces can even directly regulate gene expression directly by altering the transport of transcription factors into the nucleus. Furthermore, nuclear envelope proteins, such as lamin A, bind to transcription factors. Since there is an interconnection between these structural proteins (76, 77), any forces that are transferred through these components may potentially directly alter gene expression (78, 79).

Interestingly, transcriptional programs which demonstrate mechanical stress mediated control of transcription are linked with p53 activity. Increasing evidence demonstrates that mechanical control of transcription in involved in maintenance of pluripotency, cell fate determination, pattern formation and organogenesis (75). Decades of research suggest p53 is intricately involved in these functions. Thus, a more complete understanding of how p53 could act as a mechanosensitive sensor and transcriptional programs induced under its control

may contribute to basic research in fields such as tissue engineering and regenerative medicine.

Each study has limitations. One central caveat of this study stems from the use of the dp/+ allele. The dp/+ allele encompasses a 4.3 Mb region on mouse chromosome 4 which is syntenous to human 1p36. Although spontaneous somatic deletions have been observed in numerous types of cancers, to our knowledge no duplication of this locus exists. Thus, the allele we generated is artificial, and the consequences of this genetic aberration in humans is unknown. Furthermore, we do not have a clear understand of what genes in this locus contribute to the observed phenotypes. We must further address how the dp/+ allele fully contributes to the activation of p53. Bagchi et al. demonstrated that a key gene which contributes to the cancer phenotype when deleted is Chd5 (56). This chromatin remodeling protein is upstream of p53, and it modulates its expression through the p16/p19 pathway. Currently, the role of Chd5 in progeroid syndromes remains unexplored.

A recent study hints at a Chd5-p53 connection between chromatin reorganization during cellular senescence (80). Using a cutting edge proteomics approach, Liu et al. measured the influx of proteins to the nucleolus upon induction of senescence. Similar to the nuclear envelope, the nucleolus is associated with morphological changes during aging. Interestingly, cellular senescence was associated with an influx of some Chd family members, including Chd5, and extensive chromatin re-organization of the nucleolus was observed. ARF, as well as other p53 binding proteins localized to the nucleolus,

suggesting a compartmentalization of p53 and this chromatin remodeling family during cellular senescence. Therefore, an investigation of how the genes of the dp/+ allele contribute to enhanced p53 activity is warranted. This may provide new insight as to how p53 contributes to chromatin structure globally.

Studying the pathogenesis of age-related diseases is of major importance in a society with growing elderly population, and identifying an appropriate model system remains a challenge for the research community. The development of animal models for human disease aids our understanding of the underlying mechanisms that contribute to disease. This mechanistic insight is invaluable, because it enables the development of preventative and therapeutic strategies to combat disease. Ideally, we wish to a disease state in humans, however studies involving humans are often not feasible due the sample size necessary to overcome genetic and environmental variables or length of the lifespan. Thus, the use of animal models, such as genetically modified mouse models, with controlled genetic backgrounds and environment conditions are essential tools for basic research in the field of aging. This allelic series of p53 may be a powerful tool in elucidating mechanisms which contribute to age related diseases such as musculoskeletal degeneration as well as fundamental biological concepts pertaining to the how the spatial organization of the nucleus influences gene expression. The characterization of this model provides a foundation on which future work on the role of p53 in organismal aging can be built.

Materials and Methods

Cell Culture

MEFs were derived from 12.5 dpc embryos with a mixed B6/129 background. The embryos were homogenized and plated on 10 cm tissue culture dishes. When cells reached confluency, they were passaged on three 10 cm dishes (p1), then frozen down into three vials per plate. The p1 vials of MEFs were thawed and the cells were expanded to passage 3 for experimental use. All experiments were conducted with early passage cells (p3-p6). MEFs were cultured in DMEM supplemented with 10% FBS and penicillin-streptomycin (100 units/ml).

Microscopy

Immunoflourence was performed according to Spector and Smith using the abcam lamin A/C and H3K9me3 antibody (Abcam) Cells were imaged at 40x magnification with the delta vision or transmission electron microscopy.

Statistical Analysis

Statistical analyses were performed using Microsoft Excel and Prism. The differences between two samples were analyzed using Student's t test. The comparison of survival between mice shown in the Kaplan-Meier survival curves was performed using the Log Rank test.

Western Blot

Protein was harvested with SDS sample buffer, run on 10% SDS-polyacrylamide gels, and transferred to a PVDF membrane overnight at 20 V. The membranes were blocked with 5% milk in TBS-T, incubated with the primary antibodies and secondary antibodies diluted in 5% milk in TBS-T. For a loading control GAPDH Cell Signaling Technology (14C10, #2118) was used at a concentration of 1:2,000.

Nested RT PCR and cDNA Library Generation

RNA was isolated from MEFs three dp/+ and +/+ mice using Qiagen's RNeasy Mini kit according to the manufacturer's instructions. The RNA was run on a 2% agarose gel to check the integrity, and treated with DNaseI (Invitrogen) before cDNA synthesis. cDNA was synthesized with Promega's GoScript Reverse Transcription System. Transcripts of p53 were amplified from +/+ and dp/+ cDNA with the following primers in table 1. PCR products were purified from agarose gels using Wizard SV Gel and PCR Clean-Up System according to manufacturers instructions and cloned into TOPO pcDNA3.1 vector. Clones were sequenced with the CMV forward primer provided with the TOPO TA Cloning kit.

Bioinfomatics Analysis

The full length p53 nucleotide and protein sequences were retrieved from NCBI (Accession number NM 011649), and the domains of the protein were

determined with Uniprot. NCBI's BLAST was used to align transcript and protein sequences, and SIB's ExPASy was used for the translation of the DNA sequences to a protein sequences.

Sequences

Nucleotide sequence of full length p53

TTTCCCCTCCCACGTGCTCACCCTGGCTAAAGTTCTGTAGCTTCAGTTCATT
GGGACCATCCTGGCTGTAGGTAGCGACTACAGTTAGGGGGCACCTAGCATT
CAGGCCCTCATCCTCCTCCTTCCCAGCAGGGTGTCACGCTTCTCCGAAGAC
TGGATGACTGCCATGGAGGAGTCACAGTCGGATATCAGCCTCGAGCTCCCT
CTGAGCCAGGAGACATTTTCAGGCTTATGGAAACTACTTCCTCCAGAAGATA
TCCTGCCATCACCTCACTGCATGGACGATCTGTTGCTGCCCCAGGATGTTG
AGGAGTTTTTTGAAGGCCCAAGTGAAGCCCTCCGAGTGTCAGGAGCTCCTG
CAGCACAGGACCCTGTCACCGAGACCCCTGGGCCAGTGGCCCCTGCCCCA
GCCACTCCATGGCCCCTGTCATCTTTTGTCCCTTCTCAAAAAACTTACCAGG
GCAACTATGGCTTCCACCTGGGCTTCCTGCAGTCTGGGACAGCCAAGTCTG
TTATGTGCACGTACTCTCCTCCCCTCAATAAGCTATTCTGCCAGCTGGCGAA
GACGTGCCCTGTGCAGTTGTGGGTCAGCGCCACACCTCCAGCTGGGAGCC
GTGTCCGCGCCATGGCCATCTACAAGAAGTCACAGCACATGACGGAGGTC
GTGAGACGCTGCCCCCACCATGAGCGCTGCTCCGATGGTGATGGCCTGGC
TCCTCCCCAGCATCTTATCCGGGTGGAAGGAAATTTGTATCCCGAGTATCTG
GAAGACAGGCAGACTTTTCGCCACAGCGTGGTGGTACCTTATGAGCCACCC
GAGGCCGGCTCTGAGTATACCACCATCCACTACAAGTACATGTGTAATAGC
TCCTGCATGGGGGGCATGAACCGCCGACCTATCCTTACCATCATCACACTG
GAAGACTCCAGTGGGAACCTTCTGGGACGGGACAGCTTTGAGGTTCGTGTT
TGTGCCTGCCCTGGGAGAGACCGCCGTACAGAAGAAGAAAATTTCCGCAAA
AAGGAAGTCCTTTGCCCTGAACTGCCCCCAGGGAGCGCAAAGAGAGCGCT
GCCCACCTGCACAAGCGCCTCTCCCCCGCAAAAGAAAAAACCACTTGATGG
AGAGTATTTCACCCTCAAGATCCGCGGGCGTAAACGCTTCGAGATGTTCCG
GGAGCTGAATGAGGCCTTAGAGTTAAAGGATGCCCATGCTACAGAGGAGTC
TGGAGACAGCAGGGCTCACTCCAGCTACCTGAAGACCAAGAAGGGCCAGT
CTACTTCCCGCCATAAAAAAACAATGGTCAAGAAAGTGGGGCCTGACTCAG
ACTGACTGCCTCTGCATCCCGTCCCCATCACCAGCCTCCCCCTCTCCTTGC
TGTCTTATGACTTCAGGGCTGAGACACAATCCTCCCGGTCCCTTCTGCTGC
CTTTTTTACCTTGTAGCTAGGGCTCAGCCCCCTCTCTGAGTAGTGGTTCCTG
GCCCAAGTTGGGGAATAGGTTGATAGTTGTCAGGTCTCTGCTGGCCCAGCG
AAATTCTATCCAGCCAGTTGTTGGACCCTGGCACCTACAATGAAATCTCACC
CTACCCCACACCCTGTAAGATTCTATCTTGGGCCCTCATAGGGTCCATATCC
TCCAGGGCCTACTTTCCTTCCATTCTGCAAAGCCTGTCTGCATTTATCCACC
CCCCACCCTGTCTCCCTCTTTTTTTTTTTTTTACCCCTTTTTATATATCAATTT
CCTATTTTACAATAAAATTTTGTTATCACT

Nucleotide sequence of dp/+ p53 isoform 1

GACAGCTTTGAGGTTCGTGTTTGTGCCTGCCCTGGGAGAGACCGCCGTACA
GAAGAAGAAAATTTCCGCAAAAAGGAAGTCCTTTGCCCTGAACTGCCCCCA
GGGAGCGCAAAGAGAGCGCTGCCCACCTGCACAAGCGCCTCTCCCCCGCA
AAAGAAAAAACCACTTGATGGAGAGTATTTCACCCTCAAGATCCGCGGGCG
TAAACGCTTCGAGATGTTCCGGGAGCTGAATGAGGCCTTAGAGTTAAAGGA
TGCCCATGCTACAGAGGAGTCTGGAGACAGCAGGGCTCACTCCAGCTACCT
GAAGACCAAGAAGGGCCAGTCTACTTCCCGCCATAAAAAAACAATGCACAA
TCCTCCCGGTCCCTTCTGCTGCCTTTTTTACCTTGTAGCTAGGGCTCAGCCC
CCTCTCTGAGTAGTGGTTCCTGGCCCAAGTTGGGGAATAGGTTGATAGTTG
TCAGGTCTCTGCTGGCCCAGCGAAATTCTATCCAGCCAGTTGTTGGACCCT
GGCACCTACAATGAAATCTCACCCTACCCCACACCCTGTAAGATTCTATCTT
GGGCCCTCATAGGGTCCATATCCTCCAGGGCCTACTTTCCTTCCATTCTGCA
AAGCCTGTCTGCATTTATCCACCCCCCACCCTGTCTCCCTCTTTTTTTTTTTT
TTACCCCTTTTTATATATCAATTTCCTATTTTACAATAAAATTTTGTTATCACT

Nucleotide sequence of dp/+ p53 isoform 2

CTGCCATGGAGGAGTCACAGTCGGATATCAGCCTCGAGCTCCCTCTGAGCC
AGGAGACATTTTCAGGCTTATGGAAACTACTTCCTCCAGAAGATATCCTGCC
ATCACCTCACTGCATGGACGATCTGTTGCTGCCCCAGGATGTTGAGGAGTT
TTTTGAAGGCCCAAGTGAAGCCCTCCGAGTGTCAGGAGCTCCTGCAGCACA
GGACCCTGTCACCGAGACCCCTGGGCCAGTGGCCCCTGCCCCAGCCACTC
CATGGCCCCTGTCATCTTTTGTCCCTTCTCAAAAAACTTACCAGGGCAACTA
TGGCTTCCACCTGGGCTTCCTGCAGTCTGGGACAGCCAAGTCTGTTATGTG
CACGTACTCTCCTCCCCTCAATAAGCTATTCTGCCAGCTGGCGAAGACGTG
CCCTGTGCAGTTGTGGGTCAGCGCCACACCTCCAGCTGGGAGCCGTGTCC
GCGCCATGGCCATCTACAAGAAGTCACAGCACATGACGGAGGTCGTGAGA
CGCTGCCCCCACCATGAGCGCTGCTCCGATGGTGATGGTAAGCCCTCAAC
ACCGCCTGTGGGGTTAGGACTGGCAGCCTCCCATCTCCCGGCTTCTGACTT
ATTCTTGCTCTTAGGCCTGGCTCCTCCCCAGCATCTTATCCCGGTCCCTTCT
GCTGCCTTTTTTACCTTGTAGCTAGGGCTCAGCCCCCTCTCTGAGTAGTGGT
TCCTGGCCCAAGTTGGGGAATAGGTTGATAGTTGTCACGTCTCTGCTGGGC
CCAGCGAAATTCTATCCAGCCAGTGTGGACCCTGGCACCTACAATGAAATC
TCACCCTACCCACCACCCTGTAGATTCTATCTGGGGCCCTCATAGGGTCATA
TCTCAGGGCTTACTTTCTCATAAGGCGATCAGCACACTGGCGCCGTACTAAT
GATCGAGCTCGTACCAGCTTGCGTACCATGTCATAGCTGTTCTGTGTGATCT
GTAATCGGTTCACAGTCTACAGACATCGTGATCGTGGA
TATGTCGATAAACGT

Full length p53 exon nucleotide position

p53 Exon	Nucleotide Position in Transcript (nt)	Length of Exon (nt)
Exon 1	1-157	157
Exon 2	158-240	82
Exon 3	241-262	21
Exon 4	263-523	260
Exon 5	524-707	183
Exon 6	708-820	112
Exon 7	821-930	109
Exon 8	931-1067	136
Exon 9	1068-1141	73
Exon 10	1142-1248	106
Exon 11	1249-1772	523

dp/+-p53 isoform 1 exon nucleotide position

p53 Exon	Nucleotides	Length of Exon
Exon 8i	950-1067	117
Exon 9	1068-1141	73
Exon 10	1142-1248	106
Exon 11i	1249-1264, 1487-1772	300

dp/+-p53 isoform 2 exon nucleotide position

p53 Exon	Nucleotides	Length of Exon
Exon 1	1-157	157
Exon 2	158-240	82
Exon 3	241-262	21
Exon 4	263-523	260
Exon 5	524-707	183
Exon 6i	708, +76, 709-735	106
Exon 11i	1412-1772	360

Translation of full length p53

F P S H V L T L A K V L **Stop** L Q F I G T I L A V G S D Y S **Stop** G A P S I Q A
L I L L L P S R V S R F S E D W **Met** T A **Met** E E S Q S D I S L E L P L S Q E T
F S G L W K L L P P E D I L P S P H C **Met** D D L L L P Q D V E E F F E G P S
E A L R V S G A P A A Q D P V T E T P G P V A P A P A T P W P L S S F V P S
Q K T Y Q G N Y G F H L G F L Q S G T A K S V **Met** C T Y S P P L N K L F C
Q L A K T C P V Q L W V S A T P P A G S R V R A **Met** A I Y K K S Q H **Met** T
E V V R R C P H H E R C S D G D G L A P P Q H L I R V E G N L Y P E Y L E D
R Q T F R H S V V V P Y E P P E A G S E Y T T I H Y K Y **Met** C N S S C **Met** G
G **Met** N R R P I L T I I T L E D S S G N L L G R D S F E V R V C A C P G R D
R R T E E E N F R K K E V L C P E L P P G S A K R A L P T C T S A S P P Q K
K K P L D G E Y F T L K I R G R K R F E **Met** F R E L N E A L E L K D A H A T
E E S G D S R A H S S Y L K T K K G Q S T S R H K K T **Met** V K K V G P D S
D **Stop** L P L H P V P I T S L P L S L L S Y D F R A E T Q S S R S L L L P F L P
C S **Stop** G S A P S L S S G S W P K L G N R L I V V R S L L A Q R N S I Q P
V V G P W H L Q **Stop** N L T L P H T L **Stop** D S I L G P H R V H I L Q G L L S
F H S A K P V C I Y P P P T L S P S F F F F Y P F L Y I N F L F Y N K I L L S

Translation of dp/+ p53 isoform 1

D S F E V R V C A C P G R D R R T E E E N F R K K E V L C P E L P P G S A
K R A L P T C T S A S P P Q K K K P L D G E Y F T L K I R G R K R F E **Met** F
R E L N E A L E L K D A H A T E E S G D S R A H S S Y L K T K K G Q S T S R
H K K T **Met** H N P P G P F C C L F Y L V A R A Q P P L **Stop** V V V P G P S W
G I G **Stop** **Stop** L S G L C W P S E I L S S Q L L D P G T Y N E I S P Y P T P
C K I L S W A L I G S I S S R A Y F P S I L Q S L S A F I H P P P C L P L F F F
F T P F Y I S I S Y F T I K F C Y H

Translation of dp/+ p53 isoform 2

L **Stop** L Q F I G T I L A V G S D Y S **Stop** G A P S I Q A L I L L L P S R V S R
F S E D W **Met** T A **Met** E E S Q S D I S L E L P L S Q E T F S G L W K L L P P
E D I L P S P H C **Met** D D L L L P Q D V E E F F E G P S E A L R V S G A P A
A Q D P V T E T P G P V A P A P A T P W P L S S F V P S Q K T Y Q G N Y G
F H L G F L Q S G T A K S V **Met** C T Y S P P L N K L F C Q L A K T C P V Q L
W V S A T P P A G S R V R A **Met** A I Y K K S Q H **Met** T E V V R R C P H H E
R C S D G D G K P S T P P V G L G L A A S H L P A S D L F L L L G L A P P Q
H L I P V P S A A F F T L **Stop** L G L S P L S E **Stop** W F L A Q V G E **Stop** V D
S C H V S A G P S E I L S S Q C G P W H L Q **Stop** N L T L P T T L **Stop** I L S
G A L I G S Y L R A Y F L I R R S A H W R R T N D R A R T S L R T **Met** S
Stop L F C V I C N R F T V Y R H R D R G Y V D K R

Chapter 4: The Domestication of Endogenous Retroviruses by p53 to Mediate Cell Viability

Introduction

A substantial proportion of the mammalian genome is comprised of repetitive sequences. Initial analyses of genome sequencing projects revealed that repetitive elements make up 30-50% of mammalian genomes (25, 81), but recent computational analysis suggests that this may be a gross underestimate, demonstrating that 66-69% of the human genome may be repetitive in nature (82). The major sources of these repetitive sequences are transposons. Transposons are a universal feature of genomes; they are found in bacteria, archaea and eukarya which encompasses every domain of life (23). Sometimes referred to as 'jumping genes' or 'mobile elements', transposons are DNA sequences that insert themselves into new locations of the genome (28). Whereas the DNA transposons mobilize through a cut and paste mechanism, the retrotransposons copy and paste themselves via an RNA intermediate. Although transposons are broadly classified by their mode of mobilization, transposition remains under tight epigenetic repression to maintain the fidelity of the genome (83).

Recent advances in transcriptome analysis illustrate differential expression of distinct families of transposons among different cell and tissue types in mice and humans, but the specific functions of these transcripts have yet to be revealed (25). Barbara McClintock hypothesized that stress activates the expression of transposable elements to generate genetic diversity. Indeed,

numerous studies demonstrate the activation of transposons in response to various forms of stress from bacteria to humans (84). Their transposition can dramatically alter the landscape of the genome by promoting epigenetic changes, generating genomic rearrangements, or by scattering alternative regulatory sequences adjacent to genes (85). Furthermore, transposons can add transcribed, as well as functional, sequences to a genome (86).

The consequences of stress activation of transposons widely researched in bacteria, yeast, and plants however in mammals much remains unknown (28). In order to elucidate mammalian cell stress response mechanisms, this study attempted to determine which transposons are regulated by the universal stress sensor, p53. It was hypothesized that distinct transposons are transcriptionally regulated by p53. To this end, the transcriptome analysis was performed using an analysis to enrich for highly abundant sequences in normal cells and cells that were genetically modified to carry a null allele for p53 by RNA Sequencing (RNA Seq). Here we report the identification of stress responsive endogenous retroviruses (ERVs) of the HTLV-1 like Heptad Repeat 1-Heptad Repeat 2 family.

Results

In order to better understand the transcriptional role of p53, we utilized RNA Seq of early passage Mouse Embryonic Fibroblast generated from littermates in order to compare wildtype p53 (p53+/+) to single copy (p53-/+) and homozygous null allele of p53 (p53-/-). In parallel, traditional microarray experiments were also carried out to ensure our sequence processing pipeline

maintained data integrity. The array data comparing the p53+/+ to p53-/- state clearly identifies canonical p53 target genes whose transcript level changes when p53 is deleted. For example, Ccng1 (Cyclin G1), CCdkn1a (p21) levels are greatly decreased in the p53-/- MEFS. These canonical p53 target genes are upregulated in response to distinct stimuli such as genotoxic stress (44). In the RNA Seq data analyzed using Cuffdiff, we observe a loss of transcription of canonical p53 targets in the p53-/- MEFs. Statistical analyses of the array results identified over 50 genes and Cuffdiff analysis of the RNA Seq results identifies over 100 genes to be differentially expressed. 24 of these genes were identified by both methods. While not always significant, when a gene was present in both array and RNA Seq datasets the RNA Seq data showed changes of direction consistent with the array data. The RNA Seq further identified many genes that were not observed by array.

Next, we looked at the p53 transcript levels in the datasets. We observed that p53 was reduced by ~3 fold on the array and ~30 fold by sequencing. It is commonly known that for many genes array data underestimates the overall fold change. To systematically determine whether the transcriptional response was indicative of a decrease in p53, we submitted the responsive gene lists and their direction of change to IPA analyses. As expected, targets of p53 were highly significantly enriched in the datasets (Array 4.36e-13, RNA Seq-13). Additionally, the observed direction of change can be used to predict that Trp53 activity is significantly reduced in the p53-/- MEFs by Z-score (Figure 15) (Array -3, SEQ -4N). These results indicate that our analyses methods are consistent between

methods and that more complete information can be obtained by further analysis of the sequencing data.

Due to the fact that the RNA Seq summarization based on the Cuffdiff method requires genome annotation, we hypothesized that additional targets of p53 may exist that were not present within the current annotation. Furthermore, just as controlling the stringency of an interaction is essential to array hybridization, controlling the stringency of mapping is highly relevant to the analyses of RNA Seq data specifically when repetitive elements are the subject of study. To respond to these two concerns, we developed a strategy to utilize highly stringent mapping coupled with a method to identify differential transcription without regard to prior annotation using code originally designed to identify genomic locations overrepresented in a data set. We call this new bioinformatics tool Annotation Independent Transcriptome Assembly (AITA). The most overrepresented locations identified by AITA with a 5-fold cutoff included Cyclin G1, p21 and other p53 target genes (Table 3), indicating that our method would identify regions in a similar fashion as those identified using annotation (Figure 15). Intriguingly, additional regions were identified that were not annotated in the mm9 annotation in a similar fashion as Cyclin G1 and p21. Additionally these unclassified regions showed reduced expression in p53+/- MEFs (Figure 15).

The chromosomal location to which this region of the genome maps contains many repetitive elements. By stitching together the 100 base pair reads, we narrowed down the localization of this transcriptional peak to a region

approximately 10 kb in length. BLAT analyses of the highly abundant 7560 bp sequence found at the chromosome 8 region identified 13 copies of the transcript to be present at ≥ 99.0% within the mouse genome with additional copies on almost every chromosome in mouse. This sequence contains open reading frames (ORFs) encoding for gag, pro-pol, and env proteins and is classified as a Long Terminal Repeat (LTR) Retrotransposon (Figure 16). Using Repbase, we identify this sequence as Mmergln-int (87). In order to validate the RNA Seq results for the chr8 region containing the LTR Retrotransposon we designed RT-PCR primers to the gag and env regions. PCR products with the correct size are generated from cDNA from MEFS with p53 and are not found in libraries from M (Figure 4A). An additional retrovirus was observed on chromosome 2, however this transcript no longer contains ORFs. We also found p53 dependent expression of SINES (Table 4).

Because p53 is stabilized in response to stress, and stabilization increased the levels of p21 and Cyclin G1 we hypothesized that stabilization of p53 would also lead to increases in the transcript level of Mmergln-int. Exposure to ultraviolet light resulted in an ~ 2 fold increase of the Mmergln-int transcript. Additionally, qRT-PCR results show that p53 is necessary for basal transcription and in the absence of p53 no increase in retrovirus is observed. (Figure 17).

Previous work has identified a specific sequence, the p53 response element, in proximity to the transcriptional start site that is required for p53 dependent transactivation (37). Therefore, we looked for the p53 response element within the LTR of Mmergln-int. A highly conserved p53 response

element resides within the LTR (Figure 18). To demonstrate that p53 directly interacts with the LTR we conducted chromatin immunoprecipitation with an antibody against p53 (Figure 18). As a control, we are clearly able to observe an interaction between p53 and the p21 and Bax promoter that is not present in the absence of p53. Likewise, in the retroviral sequence we are able to clearly see that p53 interacts directly with the response element in the retroviral LTR in p53+/+ MEFs, but not in p53-/- MEFs.

In order to demonstrate that the LTR is sufficient to drive expression of Mmergln-int, we cloned the 430 bp LTR in front of a luciferase reporter plasmid. Transient transfection of the LTR reporter construct into p53+/+ MEFs cells leads to luciferase activity similar to a similar construct generated from the p21 promoter. Mutation of the p53 response element in the retroviral sequence abrogates the reporter activity in the retroviral reporter in a similar fashion to removal of the element in the p21 reporter.

We have shown that p53 is necessary for the transcription of Mmergln-int in response to stress. To determine whether the Mmergln-int has a function in classic p53 responsive pathways, we investigated cell viability via a MTT assay. We generated cell lines with a stably integrated doxycycline inducible Mmergln-int transgene or GFP (Figure 19). The overexpression of Mmergln-int leads to decreases in cell viability in p53+/+ MEFs (Figure 20). Since p53 stabilization increases Mmergln-int expression, we hypothesized that either 1) p53 would not be necessary for the decreases in cell viability indicative of a linear signaling mechanism activated by p53 or 2) p53 would be required for decreases in cell

viability, indicative of a feed forward interaction potentially leading to signal amplification. In p53 null MEFS, overexpression of Mmergln-int does not lead to the changes in cell viability observed in p53+/+ MEFs (20).

Interestingly, upon overexpression of Mmergln-int we noticed the appearance of multinucleated cells which do not occur following overexpression of GFP (Figure 21). This lead us to wonder if full length ERV was responsible for the decrease in cell viability, or if the Gag, Pro-Pol or Env proteins are sufficient to recapitulate the decrease in cell viability. Whereas transient overexpression of the Gag, Pol, or a control vector did not result in any significant change, overexpression of only the Env protein led to also lead to a decrease in cell viability similar to the overexpression of full length Mmergln-int (Figure 21).

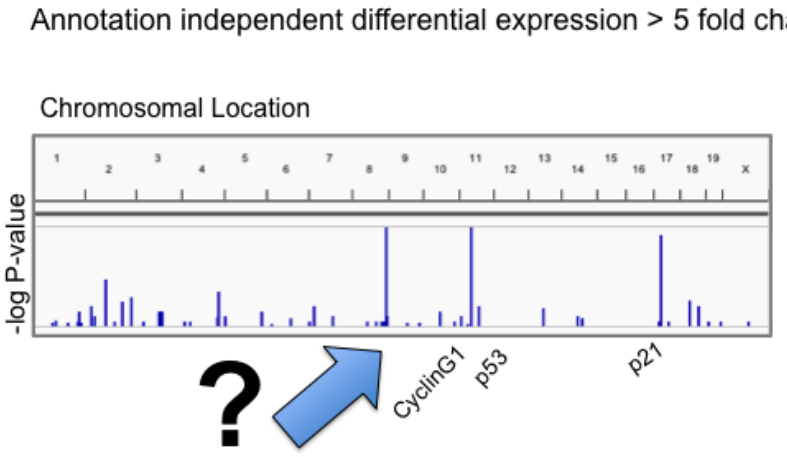
Further examination of the retroviral Env protein shows that it is a member of the HTLV-1 like Heptad Repeat 1-Heptad Repeat 2 family based on the domains this protein contains. Using the Conserved Domain Database, we discovered that this family of proteins is highly conserved and is present in modern infectious viruses (Figure 22) and is also found in many vertebrate sequences (Figure 23) including humans (88). Interestingly, human ERVs are upregulated in many cancers and autoimmune diseases (89).

To determine whether p53-mediated upregulation in response to genotoxic stress is a conserved feature of other ERVs, we designed qRT-PCR primers to specifically recognize the intact open reading frames of the human ERV envelopes (Table 3). Following UV treatment of MCF7 cells, qRT-PCR results show an upregulation of envR, env1 and envW similar to what was

observed with by Mmergln-int. Other HERVS were not responsive > 1.5 fold. To show that protein levels increased with increases in transcript, western blots using an envW antibody show increased protein levels 8 and 24 hours after UV exposure relative to non-treated MCF7cells at the appropriate molecular weight.

One difficulty which arises from studying repetitive elements is the difficulty distinguishing between highly similar sequences. Zebrafish have a single ERV, ZFERV which is limited to only a few copies in the genome. Also, this ERV is basally expressed only in the thymus and its mRNA can be visualized by in situ hybridization. To study the effect of genotoxic stress in the context of a tissue, we exposed 5 day old larva to UV and gamma irradiation (IR). We observed an increase in the intensity of the ZFERV probe in UV treated fish. We validated our findings by qRT-PCR by comparing expression levels of p53, p21 and ZFERV following UV and IR exposure to the basal expression level. We observe an increase in the transcripts of p53, p21 and ZFERV (Figure 24).

A



B

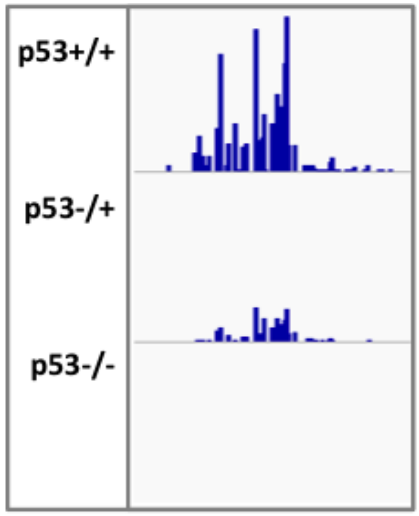
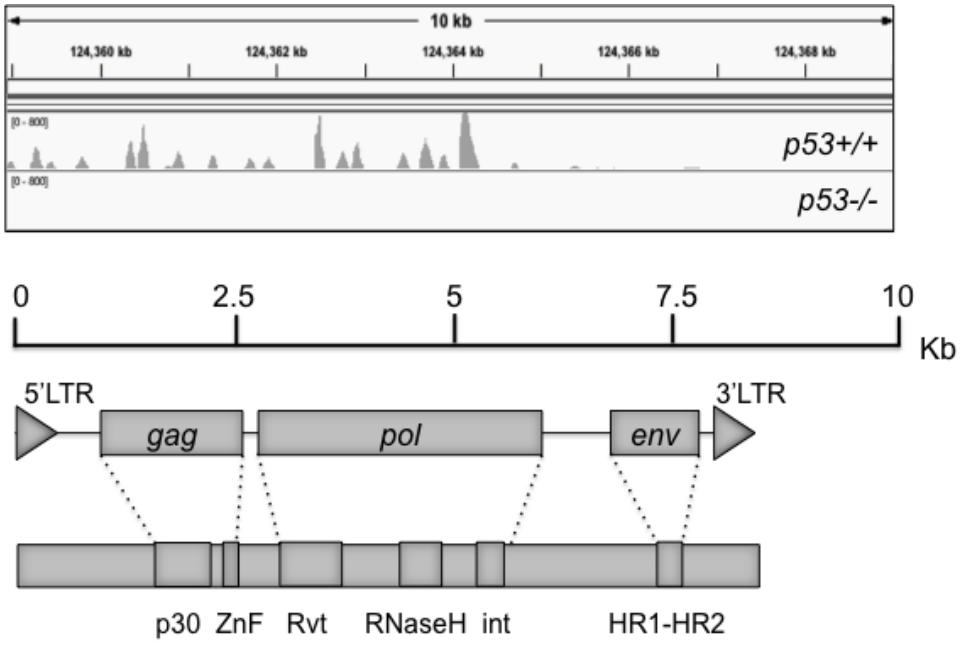


Figure 15. AITA identifies a non-annotated transcript arising from chromosome 8

(A) A highly expressed transcript is identified by AITA, blue bars represent sequencing reads

(B) The expression of the transcript is lost in p53-/- MEFs, and demonstrates reduced expression in the p53+/+ state

A



B

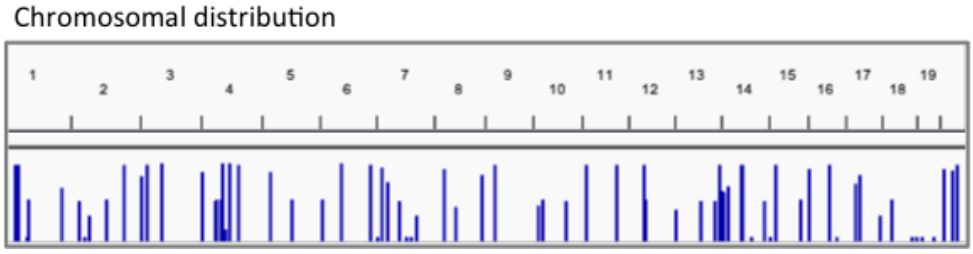


Figure 16. MmergIn-int is a highly abundant LTR retrotransposon

(A) MEFs and loss of expression in p53-/- MEFs is an ~ 8.4 kb LTR Retrotransposon named MmergIn-int. MmergIn-int is classified as an endogenous retrovirus and had retained open reading frames which encode for *gag*, *pro-pol* and *env* proteins

(B) The non-annotated transcript is present on every chromosome in mouse

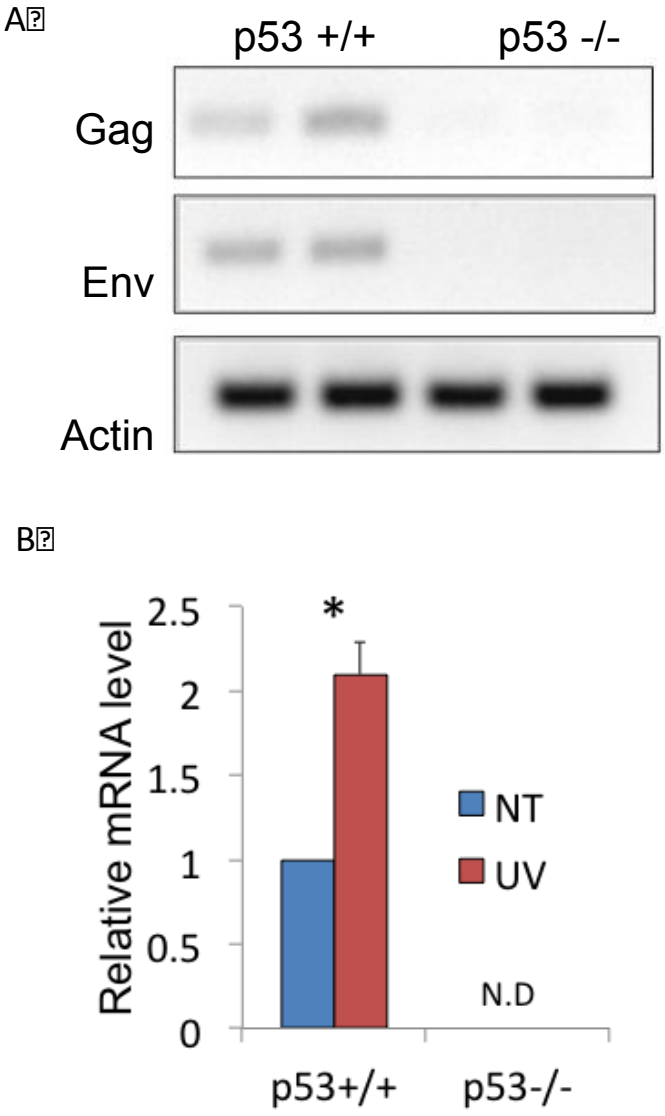


Figure 17. MmergIn-int is expressed in p53+/+ MEFs and transcriptionally upregulated in response to genotoxic stress

(A) RT-PCR analysis of MmergIn-int transcription in p53+/+, and p53-/- MEFs

(B) qRT-PCR analysis of MmergIn-int transcription in response to UV damage in p53+/+, and p53-/- MEFs

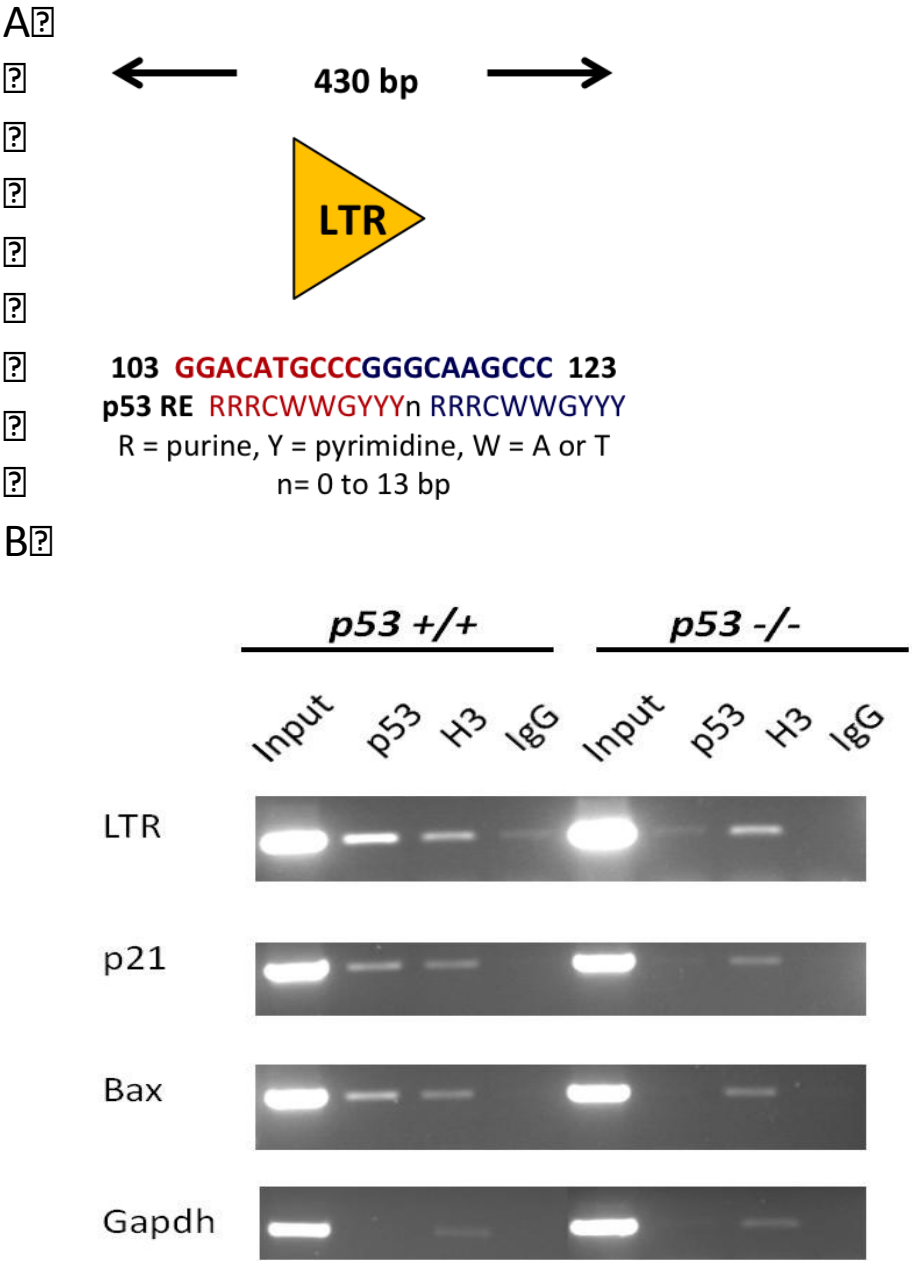


Figure 18. p53 regulates the expression of MmergIn-int by binding a p53 Response Element in the LTR
(A) The LTR of MmergIn-int contains a p53 response element at nucleotide position 103-123
(B) Chromatin immunoprecipitation with p53, H3 and IgG antibodies

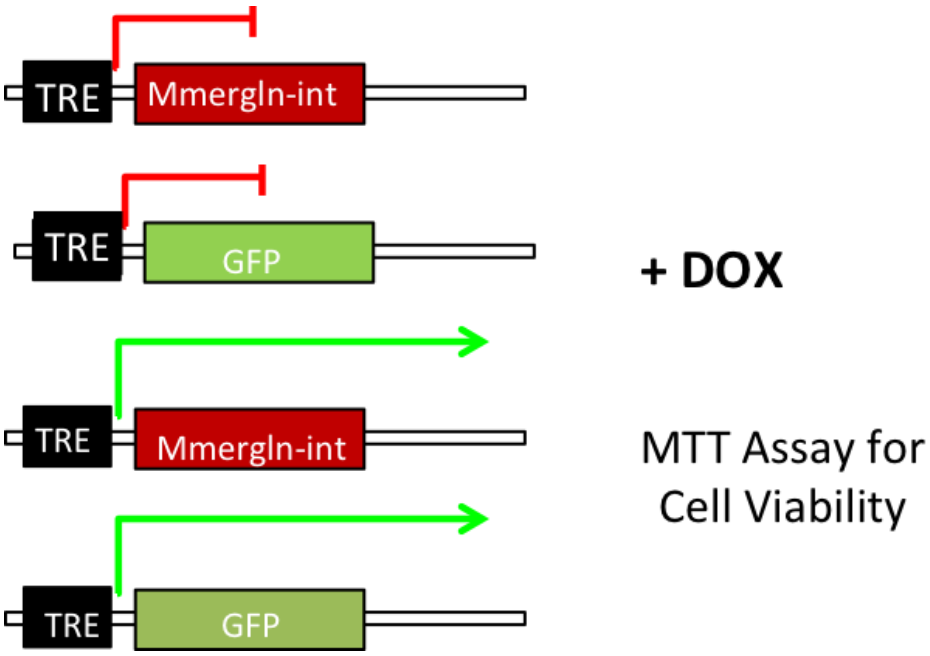


Figure 19. An inducible system to study the function of MmergIn-int

The expression of MmergIn-int or GFP is induced with doxycycline

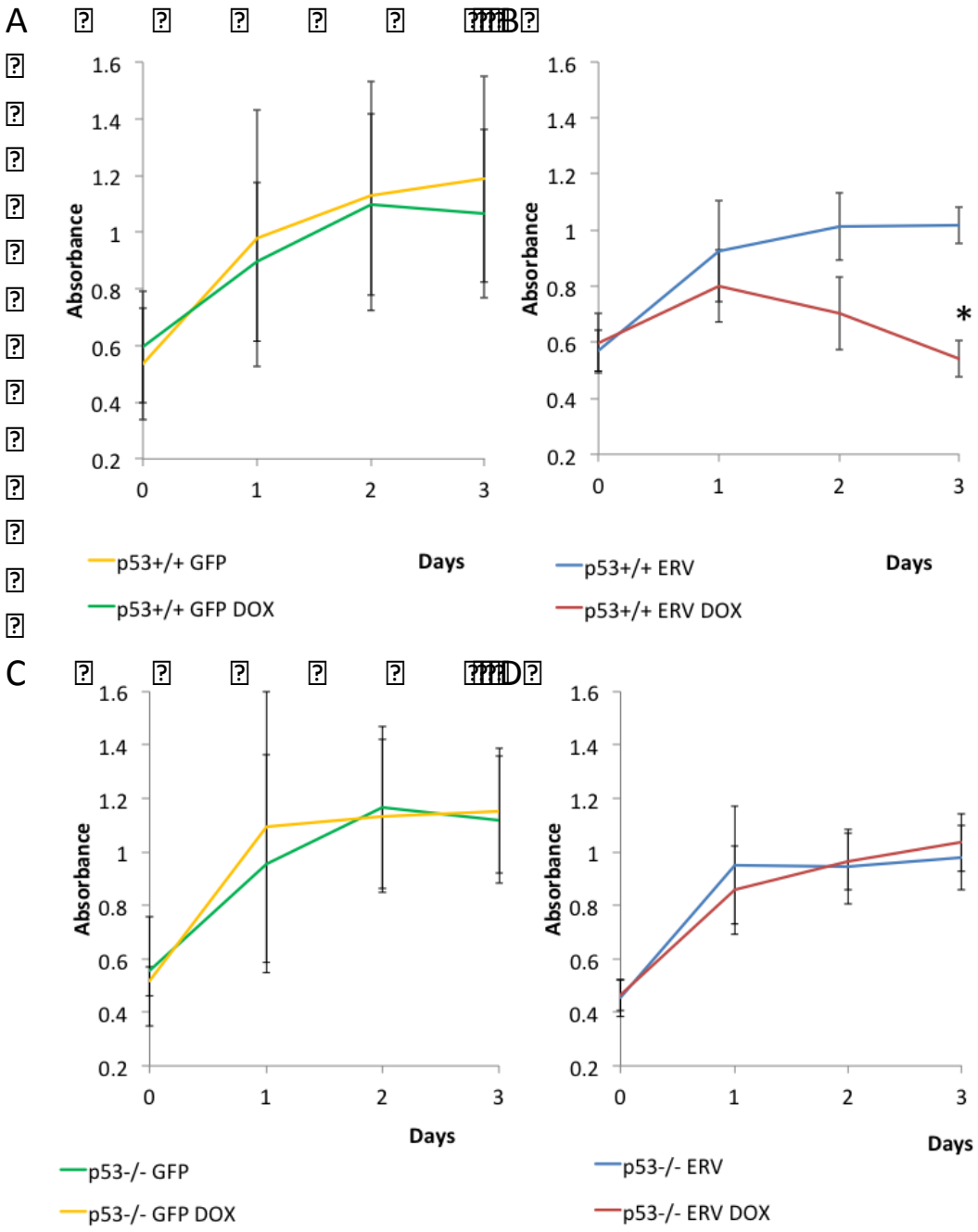


Figure 20. MmergIn-int over expression results in a decrease in p53+/+ MEFs, but not p53-/- MEFs

(A) Cell viability assay of p53+/+ MEFs induced to express GFP
(B) Cell viability assay of p53+/+ MEFs induced to express MmergIn-int, Astrix represents $p>0.05$
(C) Cell viability assay of p53-/- MEFs induced to express GFP

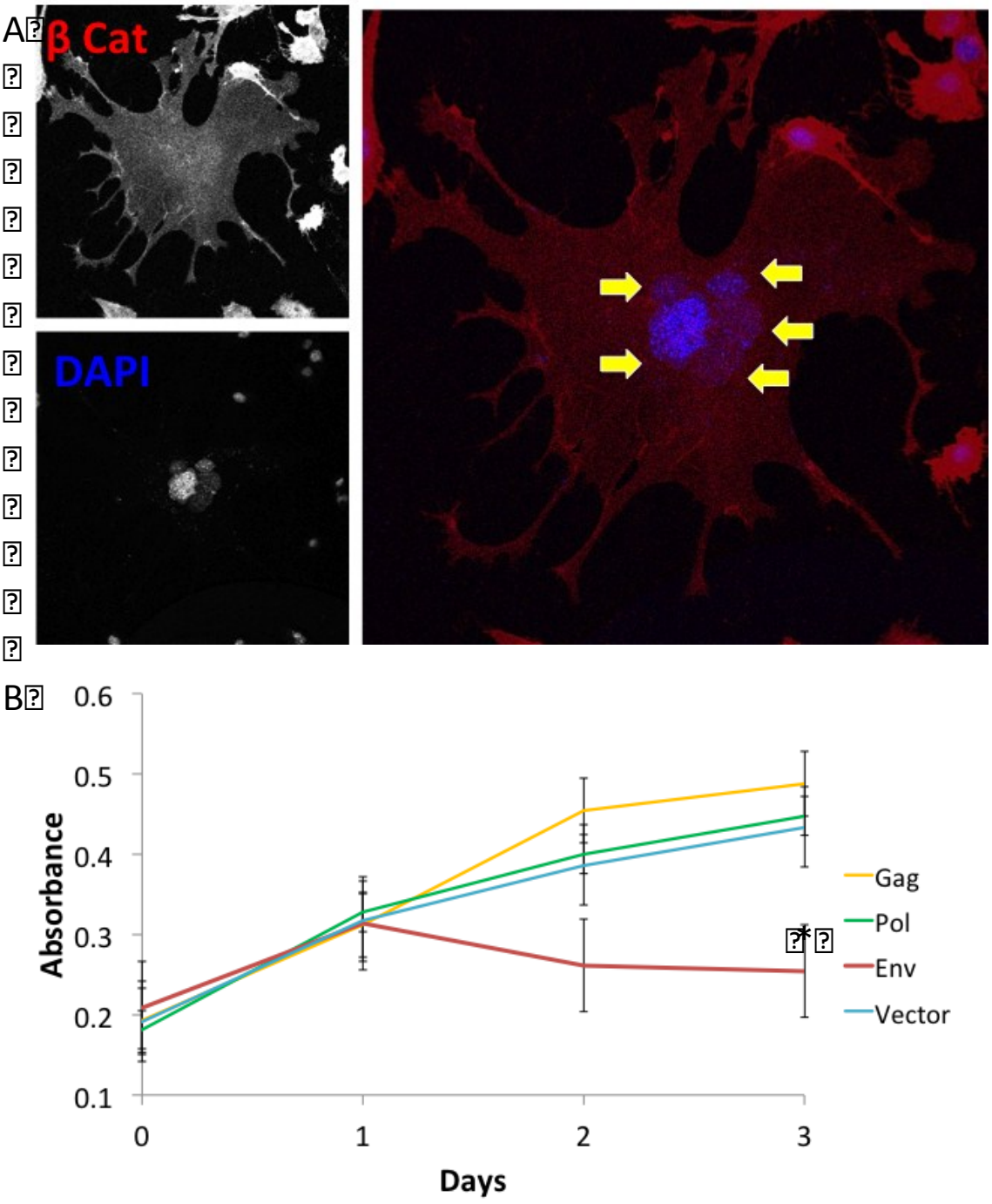


Figure 21. MmergIn-int overexpression results in cells fusion and the envelope is sufficient to induce a decrease in cell viability

(A) Confocal microscopy of MEFs 48hr after MmergIn-int induction depicts cell membrane with β Catenin and nuclei with DAPI. Yellow arrows indicate multiple nuclei within cell membrane.
(B) Cell viability assay after transient overexpression of Gag, Pol and Env

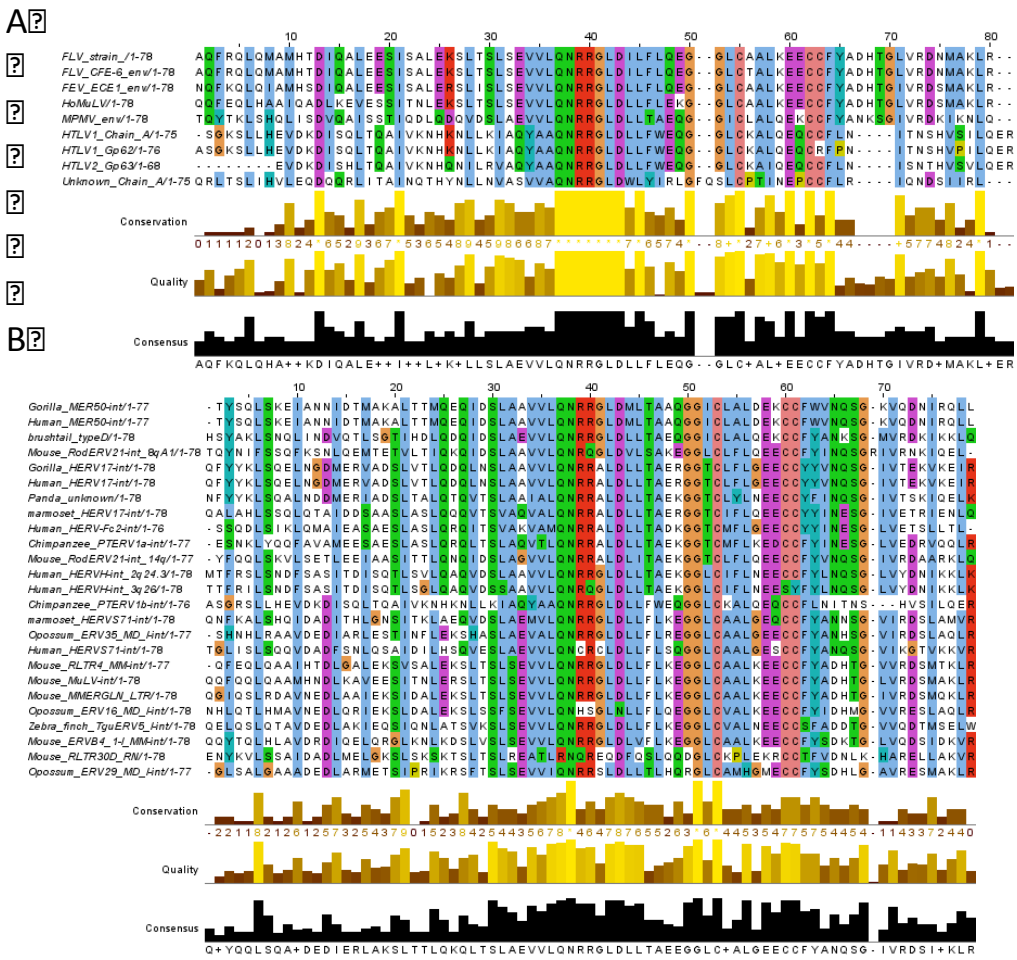
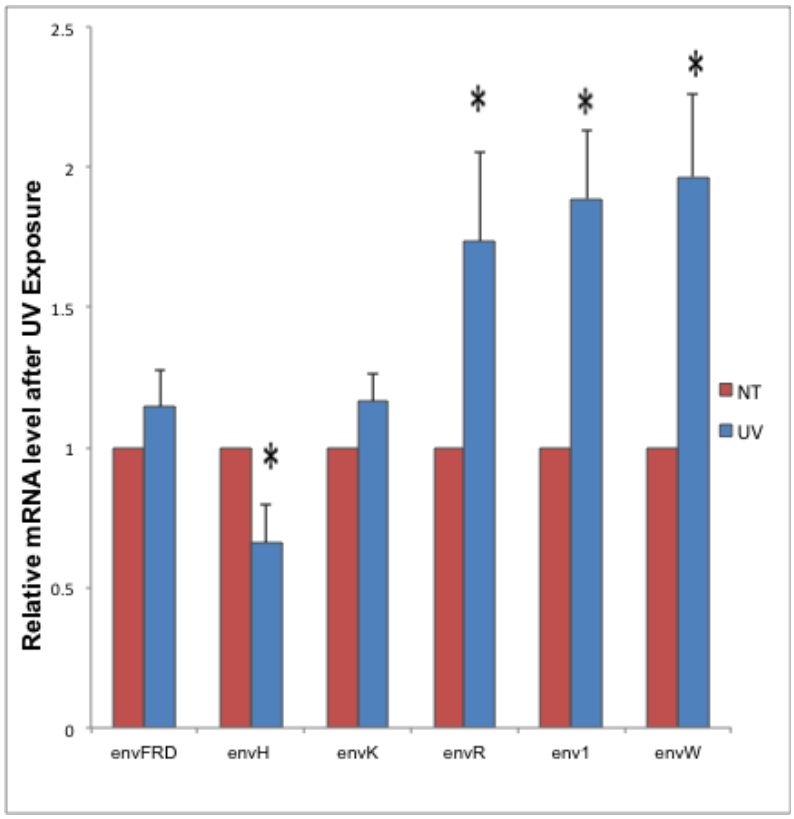


Figure 22. Conservation of the HTLV-1-like HR1 HR2 protein domain

(A) HTLV-1-like HR1 HR2 protein domain is found in the present day viruses HTLV 1, HTLV 2, FLV, FEV and HoMuLV

(B) The HTLV-1-like HR1 HR2 protein domain is endogenized in the genomes of multiple vertebrate species such as human, gorilla, chimpanzee, mouse, panda, marmoset, opossum, brushtail and zebrafinch

A



B

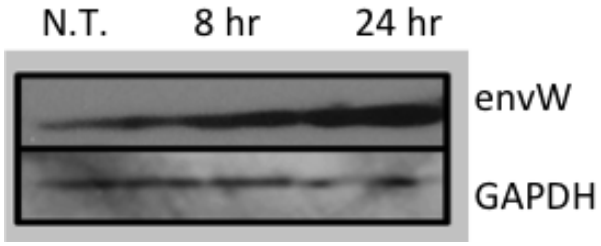


Figure 23. Human endogenous retroviruses are upregulated in response to genotoxic stress
(A) Transcriptional levels of human endogenous retroviral envelopes in response to UV stress in Mcf7 cells
(B) envW expression in Mcf7 cells after UV stress

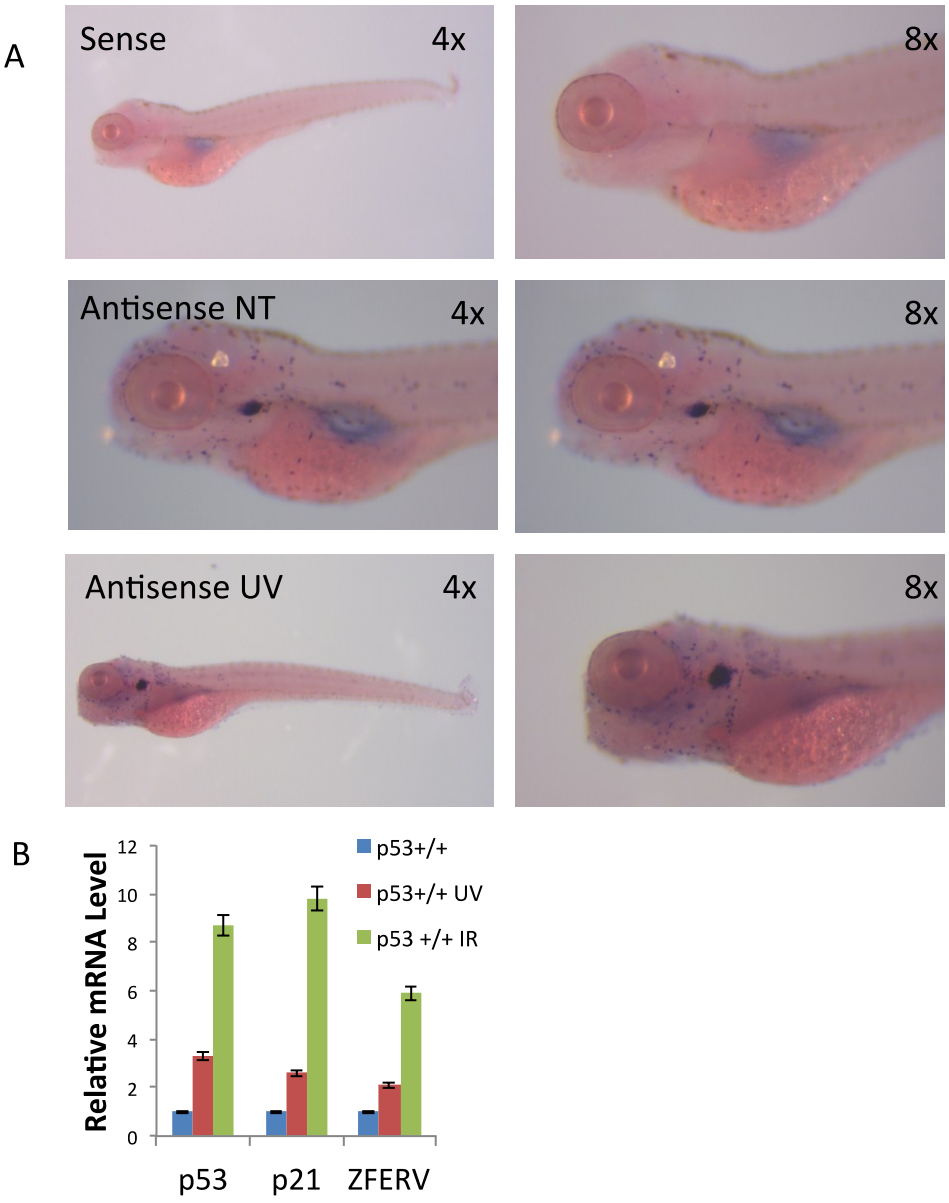


Figure 24. ZFERV is transcriptionally upregulated in response to genotoxic stress
(A) Sense and antisense probes generated to detect the envelope of ZFERV
(B) qRT-PCR analysis of p53, p21 and ZFERV

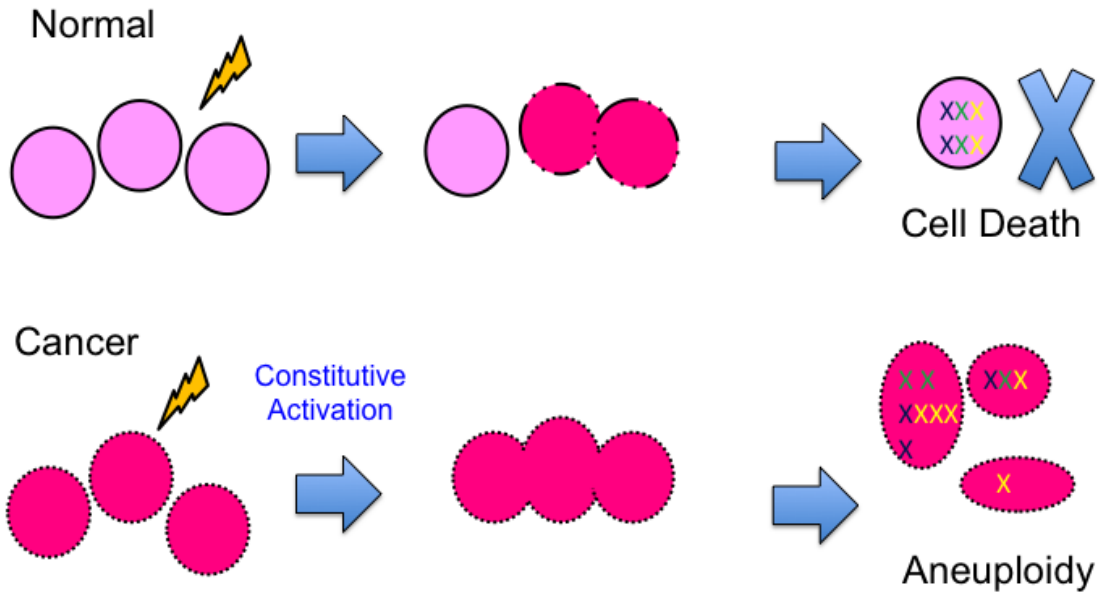


Figure 25. Endogenous retroviruses may contribute to p53 mediated tumor suppression via cell fusion, enhancing p53 signaling through a feed forward signal amplification loop and other mechanisms. However aberrant expression for fusogenic endogenous retroviral envelopes may promote cancer.

Position	Gene?
Chr 11: 40562100-40572100	Ccng1
Chr8: 124359000-124369000	No gene within 200bp
Chr17: 29228000-29238000	Cdkn1
Chr2: 7546700-75481700	No gene within 200bp
Chr4: 136436200-136456200	C1qc
Chr2: 165899700-165919700	Sulf2
Chr18: 35021200-35026200	Egr1
Chr2: 132692900-132712900	Lrrn4
Chr 11: 69403100-69408100	p53

Table 3. Differentially expressed genes identified by AITA

Name	Class
Mmergln-int	LTR Retrotransposon
MMVL30-int	LTR Retrotransposon
B1 Mm	SINE
RSINE1	SINE
B3	SINE
B1 Mur1	SINE
B1F2	SINE

Table 4. AITA detects p53 dependent expression of repetitive elements

Name	Location
envR	7q21.2
envW	7q21.2
env1	19q13.41
env2	19q13.41
envFc1	Xq21.33
envFRD	6p24.1
envRb	3p24.3
envH	3q26
envH	2q24.3
envK1	1q23.3
envK2	12q14.1
envT	19q13.11

Table 5. Human endogenous retroviral envelopes for qRT-PCR screen

Discussion

It is now appreciated that a large percent of the mammalian genome is comprised of repetitive elements, however the function of these sequences remains elusive. Most repetitive sequences are classified as transposons, mobile elements that are through to enhance the fitness of an organism by increasing genetic diversity in response to cellular stress (28). However, in humans and many other mammals, active mobilization of transposons is tightly restricted, and the expression of these sequences is concomitant with disease (89). The fact that transposons are no longer mobile in many species, but they remain preserved in the genomes of all species, suggests they have functions no longer dependent on mobilization. Using a genetic approach, we set out to identify stress responsive repetitive elements with the goal of learning more about how they contribute to the stress response of the HTLV-1 like Heptad Repeat 1-Heptad Repeat 2 family.

We developed a novel bioinformatics approach to analyze the expression of repetitive sequences, and we identified repetitive elements which were expressed only in cells with competent p53. Our approach, named Annotation Independent Transcriptome Assembly (AITA) is unlike traditional RNA Seq analysis tools. AITA allows us to easily play with the stringency of the transcript mapping of repetitive sequences. This methodology prevents misaligned sequences from being discarded from the transcriptome, which permits the analysis of their expression. By interrogating the transcriptomes of p53+/+ and p53-/- MEFs we identified the expression of distinct LTR Retrotransposons and

SINEs was lost in p53-/- MEFs. Interestingly, p53 dependent expression of SINEs has been previously reported. Furthermore, a bioinformatics analysis demonstrated that over 1,500 perfect p53 RE are found with in the LTRs of human ERV which strongly suggests a central role for p53 in their transcriptional regulation.

We demonstrate that p53 regulates the expression of the ERV Mmergln-int, by binding a conserved p53 response element in the LTR. The expression of Mmergln-int is detected in p53+/+ MEFs, and in response to genotoxic stress expression levels increase. The lack of Mmergln-int expression in p53-/- MEFs as well site directed mutagenesis experiments abrogating the function of the LTR by single mutations in the p53 regulatory sequence further indicate that p53 is the transcriptional regulator of this ERV.

After demonstrating the p53 regulates the expression of Mmergln-int, we set out to address its potential function. The tumor suppressor, p53 is a key regulator of cell viability through the induction genes involved in processes such as apoptosis, necrosis or senescence. Thus, we wished to determine the influence of Mmergln-int expression on viability in comparison to a GFP control. We observed distinct changes in cell viability by MTT assay in p53+/+ MEFs, but not p53-/- MEFs, suggesting a feed forward mechanism which could enhance p53 function. When overexpressing Mmergln-int, we noticed a change in cellular morphology 24 hours upon inducing the expression of the ERV. Some cells formed multinucleated syncytia, a phenotype associated with viral transduction.

The observation of cellular fusion made us question what components of the ERV contribute to the decrease in cellular viability. By overexpressing the gag, pro-pol and env individually, we determined that overexpression of Mmergln-int's envelope was sufficient to induce significant decreases in cell viability. Previously studies have observed that while the gag and pro-pol genes accumulate mutations that render them inactive, there is selective pressure to maintain the env suggesting it is functional (91).

Interestingly, the envelope of Mmergln-int shares a striking degree of homology to the envelope of exogenous virus, HTLV-1. The envelopes of Mmergln-int and HTLV-1 are classified by its highly conserved Heptad Repeat 1-Heptad Repeat 2 (HR1-HR2) domain (88). The HTLV-1 like Heptad Repeat 1-Heptad Repeat 2 family encompasses exogenous viruses such as HTLV-1, HTLV -2, Moloney murine leukemia virus and simian T-cell lymphotropic virus (88). The molecular features of this protein domain subfamily include the HR1 and HR2 coiled-coil forming helices which undergoes a conformation change to mediate cell fusion permitting viral entry (88).

Notably, we identify members of the HTLV-1 like Heptad Repeat 1-Heptad Repeat 2 family endogenized in the genomes of many species, including humans. This lead is to hypothesize that decreases in cell viability upon expression of envelopes of the maybe be a conserved function of these endogenized viruses. We evaluated the expression levels in response to genotoxic stress. Our data indicate transcriptional upregulation of three human ERV envelopes. Although few antibodies against human ERVs, we can validate

upregulation of envW in Mcf7 cells. Our in silico analysis detects putative p53 response elements residing in the regulatory region of envW.

Interestingly, envW is one of the human ERVs for which a function is known. An alternative name for envW is Syncytin. The Syncytin gene encodes for a fusogenic membrane glycoprotein which is involved in trophoblast fusion in the placenta (92). Syncytin's gag and pro-pol genes have undergone mutations which render them inactive, but the env remains preserved. The role of Syncytin expression in tissues besides the placenta remain poorly understood. Aberrant expression of Syncytin evokes an endoplasmic reticulum stress response as well as inflammation in the brain during multiple sclerosis (93). However, Syncytin has also been used to fight disease. It has been employed as an oncolytic agent which when overexpressed fuses cells to shrink tumors in vivo through the induction of apoptosis (94).

This leads us to ponder the implications of ERV envelope expression in response to cellular stress. On one hand, stress signals which activate p53 may lead to upregulation of fusogenic ERV envelopes which promote the elimination of damaged cells. We hypothesize this mechanism may be driven by cellular fusion as well as the unknown mechanisms which enhance p53's downstream pathways. On the other hand, abberent expression of fusogenic ERV envelopes may promote cancer. Multinucleated cells are hallmark of human cancer and have even been established as a prognostic factor to evaluate cancer initiation and progression (95, 96). An extensive body of work in the cancer field convincingly shows that multinucleated cells are the precursor to chromosomal

instability and aneuploidy, known drivers of cancer progression (97). It is even hypothesized that cellular fusion is the origin of cancer stem cells (98). Thus, future studies with to elucidate the physiological consequences of ERV expression in response to stress are warranted.

Materials and Methods

Cell Culture

MEFs were derived from 12.5 dpc embryos with a mixed B6/129 background. The embryos were homogenized and plated on 10 cm tissue culture dishes. When cells reached confluency, they were passaged on three 10 cm dishes (p1), then frozen down into three vials per plate. The p1 vials of MEFs were thawed and the cells were expanded to passage 3 for experimental use. All experiments were conducted with early passage cells (p3-p6). MEFs were cultured in DMEM supplemented with 10% FBS and penicillin-streptomycin (100 units/ml). MCF 7 cells were cultured in DMEM supplemented with 10% FBS, penicillin-streptomycin (100 units/ml) and insulin (0.01 mg/ml). In induce cellular stress, cells were exposed to UV light (50 J/m^2) or doxorubicin (50 mg/ml).

MEFs were transiently transfected using Invitrogen's Neon Transfection System or Promega's FuGENE HD transfection reagent. Unsing the Neon Transfection System, MEFs were pulsed once with a pulse voltage of 1,350 V and pulse width of 30 ms at a density of 5x 10^6/ ml cells. To generate stable cell lines, MEFs were selected with puromycin. The expression of the tet-on constructs was induced using 50 ng/ml of doxycycline.

Generation of Constructs

Mmergln-int was PCR amplified from BAC RP23-8H9 using Takara's LA TAQ DNA Polymerase and the following primers:

Mmergln int F:GGTTCCAATGCGCATTTGGAGGTCCCAGCGAGAT
Mmergln int R: GGTTCCAAACTAGTTTCCCCTCTTCTTCTGTTTAGAC.
Thermocycler conditions 94º 1 min, 94º 30 sec/58º 1 min/ 72º 9 min for 30 cycles/ 72º 10 min. were used.

The gag, pol and env were PCR amplified from BAC RP23-8H9 using Takara's LA TAQ DNA Polymerase and the following primers:
Pol F: GGTTCCAAAAGCTTATGCCTTTATTGGGGAGAGACTT,
Pol R: GGTTCCAAGCGGCCGCTCATTGAGTACCTCCCACGTTTG
Env F:GGTTCCAAAAGCTTATGATGAGTGGACTTTGGAGAA
Env R: GGTTCCAAGCGGCCGCTTAGAGGTCGGTGTCCCTTAAC
Gag F: GGTTCCAAAAGCTTATGGGACAGACCGTGTCTACTC
Gag R: GGTTCCAAGCGGCCGCTAGTCTTCATCTTCTCCAAGAG

The following thermocycler conditions were used for to amplify the pol gene: 94º 1 min, 94º 30 sec/58º 1 min/ 72º 3 min 30 sec for 30 cycles/ 72º 10 min. The gag and env genes were amplified with the following conditions: : 94º 1 min, 94º 30 sec/58º 1 min/ 72º 2 min for 30 cycles/ 72º 10 min. The PCR products were purified from an agarose gel using Wizard SV Gel and PCR Clean-Up System cloned into Invitrogen's pcDNA 3.1(+) expression vector.

Cell Viability

In order to determine cell viability, cells were seeded in a 96 well tissue culture dish at a concentration of x cells/well. Cell viability was measured using Promega's CellTiter 96 AQueous One Solution Cell Proliferation Assay (MTS). Measurements were taken at 12 hr, 36 hr, 60 and 84 hr after plating the cells. The MTT reagent was incubated at incubation at 37ºC, and the absorbance was measured with a microplate reader at 490 nm and 650 nm.

Transcriptome Analysis

RNA was isolated from cells and tissues using Qiagen's RNeasy Mini kit according to the manufacturer's instructions, and treated with DNase I (Invitrogen) before cDNA synthesis.

To detect the transcript of Mmergln cDNA was synthesized using Promega's GoScript Reverse Transcription System. Multiple PCR primers were designed to span the transcript of Mmergln-int. The following primer pairs were used for RT PCR:

GGACCAA GGA GAC CCAGAGAAG Ch8 Fi
CCCGTG TCAAC TAGAAAGTC Ch8 Ri
TCCGAGCCAATAACCCACAGG Ch8 Fii
TTTAGCAGAGGCCCGATACC Ch8 Rii
CCACGCATCCCACGTTAAGAG Ch8 Fiii
AGGCCACCAAGTCCACAGAG Ch8 Riii
CCAATTACTTTAACCCTGGCTGCCh8 Fiv
GTCGCTAGGTCTTCATTGACAGCh8 Riv

Promega's GoTaq was used to amplify the cDNA. The following thermocycler conditions were used to amplify the transcript of Mmergln-int: 95°C 2 min/95°C 30 sec, 54° C 15 sec, 72°C 15 sec for 22 cycles/72°C 5 min. Beta Actin loading control was run with the previous thermocycler conditions PCR products were run of 2% TBE agarose gels.

Real time quantitative PCR primers were designed to specifically detect to the envelope genes with a complete ORF. Primer sequences were designed using NCBI's Primer Blast or taken from (90) RNA was extracted from cells with Qiagen's RNeasy Mini kit according to the manufacturer's instructions. cDNA was synthesized using Promega's GoScript Reverse Transcription System. Real time quantitative PCR was performed using 25 µl of Promega's GoTaq qPCR Master Mix. The data was collected with the Eppendorf Realplex 2.2 Thermocycler with the following program: 50°C 2 min, 95°C 10 min, 40x 95°C 15

sec/ 60°C 1 min, 95°C 15 sec, 60°C 15 sec, increase to 95°C over 20 min for melting curve analysis.

Beta actin or GAPDH was used as an internal control to calculate differences in the amount of total RNA added in each individual reaction. Experiments with variation of the internal control less than a factor of 1were considered valid. A student's T test was performed to determine statistical significance. A p-value of p=0.05 was considered significant.

Primer sequences for internal controls :
human GAPDH F: TGCACCACCAACTGCTTAGC
human GAPDH R: GGCATGGACTGTGGTCATGAG
murine beta actin F: GTGGTTTTGATTCTCCTGTGTGC
murine beta actin R: GCCTTGTACCCATCAGGGA

Promoter Assay

The LTR of Mmergln-int PCR amplified from BAC RP 23-2D22 cloned into the promoterless pGL3 basic vector with the following primers:
LTR F: GGTTCCAACTCGAGTGAAAGGAAATA
LTR R: GGTTCCAAAAGCTTTGAAAGAACTCA.
The p21 and p21 mutant constructs are described in [Genes Dev. 1995 Apr 15;9(8):935-44.] Mutations to the Mmergln-LTR were generated using QuikChange II Site-Directed Mutagenesis Kit. A deletion was made at bases104-123 with the following primers:
del104-123 F: atagcagaacagaccaatcgcctccctagctc
del104-123 R: gagctagggaggcgattggtctgttctgctat.
A single cytosine to adenine transversion mutations were made with the following primers
c107a F: agaacagaccaggaaatgcccgggcaagc
c107a R: gcttgcccgggcatttcctggtctgttct
c117a F: gacatgcccgggaaagcccatcgcc

c117a R: ggcgatgggctttcccgggcatgtc.
Two cytosine to adenine transversion mutations were made with primers:
c107a c117a F: acagaccaggaaatgcccgggaaagcccatcgc
c107a c117a R: gcgatgggctttcccgggcatttcctggtctgt.
To generate the single base substitutions the following thermocycler conditions were used: 95º C 30 sec/ 95º C 30 sec, 55º 1 min, 68º 5 min 30 sec for 16 cycles and 18 cycles to generate the deletion and double base substitutions. p53 +/+ and p53 -/- MEFs were plated in 48 well tissue culture plates at a density of $3X10^4$ per well. MEFs were co-transfected with the 50 ng/ well of the pRL-TK Renilla and 200ng/ well of the pGL3 basic experimental constructs. Cells were lysed and luciferase activity was measured 48 hours post transfection using Promega's Dual-Luciferase Reporter Assay System according to manufacturer's instructions.
(mutations were based off papers: Transcriptional control of human p53 regulated genes (Levine, 2008) and Algorithm for prediction of tumor suppressor p53 affinity for binding sites in DNA (Veprintsev, 2008)

Western Blots

Protein was harvested with SDS sample buffer, run on 10% SDS-polyacrylamide gels, and transferred to a PVDF membrane overnight at 20 V. The membranes were blocked with 5% milk in TBS-T, incubated with the primary antibodies and secondary antibodies diluted in 5% milk in TBS-T. Protein was visualized using Genemate's (check exact name) Chemiluminescent HRP substrate. The abcam anti-hERVantibody (ab71115) was used at a concentration of 1:1,000 to detect Syncytin 1 and anti-HERV-FRD (ab90733) was used at a concentration of 1:100 to detect Syncytin 2 protein levels. For a loading control GAPDH Cell Signaling Technology (14C10, #2118) was used at a concentration of 1:2,000.

Chromatin Immunopercipitation

Milipore's Magna ChIP A - Chromatin Immunoprecipitation Kit was used according to manufactures instructions with a p53 antibody (1C12, Mouse mAb #2524 Cell Signaling Technology)

Primers to detect LTR:
ChIP LTR Fi GCTGAGAACATAGCAGAACAGACC
ChIP LTR Ri GCACCCAAGAATCACGAATAGAAC
ChIP LTR Rii AACAGGAGACAGTGGATTCGACC

Primers used as a positive control:
p21-F chip: CCTTTCTATCAGCCCCAGAGGATACC
p21-R chip: GGGACGTCCTTAATTATCTGG GGTC
Bax-F chip: GATGTTGTAGCCACCGCGTACAGCC
Bax-R chip: TTCATGGTAGAGAGCACTAAGGAGG

Immunofluorescence and Microscopy

Immunofluorescence was performed according to Spector, D.L. and H.C. Smith. 1986. Exp. Cell Res. 163, 87-94. Imaging was performed with a Zeiss Observer confocal microscope.

Bioinfomatics

mRNA expression analysis

mRNA transcript levels were detected by hybridization to illumina bead arrays. Each dataset was separately assessed for signal quality, quantile normalized and then probe sets mapping to the same gene were averaged.

RNA Seq

The RNA integrity was verified by quantification with a Ribo Green Assay and by running the samples on an Agilent Nano chip. Samples with a RIN number of 8

or above were used for mRNA seq library preparation. To prepare the library, mRNA was purified and fragmented. Next, cDNA was generated, end repair was performed, the 3' ends were adenylated and the DNA fragments were ligated to adaptors. Ligation products were purified from an agarose gel. The library was quantified and validated using an Agilent High Sensitivity chip, Pico Green assay and KAPA qPCR.

The characterization of repetitive elements was performed using the Table Browser function of the UCSC Genome Browser, and the classification of the repetitive elements was determined using BLAT and the Repeat Masker function. Identification of sequences of homology of to Mmergln-int were identified using the BLAT feature of the UCSC Genome Browser. Open reading frames were determined using NCBI's Open Reading Frame Finder. NCBI's Conserved Domain Database was used to investigate protein domains.

The consensus sequence of Mmergln-int

tttggaggccccagcgagatctgcgtgacacccaggaaccccgaaggacccctggaggtgcgtttgttt
gtgtgagtcttgttatgttgtctgttgtctaagtgtctaagtgtggcactgctgaatttgtgtcttagtt
tttcagttctgagattgtgggtttgagccccacctgtgttaccagttctggtattctgtattctggcagc
tgccactgcgttccgtaaggaccctagtggctgtgggaagacgacgatctatttccccacaggctgcacc
cttggaagacattccgagggagaccctggagtgcccggggtacggaacagtcaggaggacctggctgttg
cctggcagagtgaagaagagtgagtgctcttcctgccagaggagtggagcggaatcccactccatcagag
gtagcgtttggctggttgtgtaagtccagacgcagatgagtgtgcttggatgtcttagtattttccgtct
ctgtcattgtgttgtgtttactcttattcttcactatgggacagaccgtgtctactcctttatctttgac
taaggaccattggacggacattagggctagaggacaaaatttgtcagtaaaagtgaagaaaaagccatgg
atgactttctgttcctcagaatggcctgtttttggagtaggttggccagcagaaggaacttttttacttac
ccaccataagggctgtgaaggccattgtttttcaggaagggccagggtcgcatccagaccaacaaccgta
cttcatggtatgggaggacttggcacgctacccacccccgtgggttcacccattcctcccgcctctccac
cctggcaccaagattctagccatccgagaaaatggtgagaaagagaaaccaaaaccaccgctcgggagag
atgatgatcacagcacaccagtgatgaaaccccccaagatctatccagagattgaaaccccctgagtggc
ccaacacccctcaaccccaccgtatgctccccagccccaaccttcagctccctcaggaccccctgcctca
ggccccggccggaggaggggtccctccacaggaacaaggagccggcgaggagtcacccctgaggggcct
gcggattcaaccgtggcgctccccctcagggctattggggctccccctgccgatccaaatagtctacagc
ccctacagtattggccttttttcctcttctgacctctataactggaaagctaatcacccccctttttcag
aaaaccctgcaggactcactgggttggttgaatcattaatgtattcacaccagccgacctgggatgactg
ccagcagcttctgcagactctattcacaactgaggagagagagaggattctcctcgaggctcggaaaaat
gtccaagacgaggctgggcgccctgtccaaactccagctgagatagatgaaggatttccgctaacccggc
cccgatgggattataatacggcatcaggtagggaacgactgtccaattatcgccgggtcctagtggcggg
tctcagaggtgctgcccggcagcccacgaatctggccaaggtaagagaggttatgcagggagcgactgag
ccccccctcagtcttccttgaaaggctcatggaggcttataggagatataccccattcgaccccacgtctg
agggtcaaagggcctcagtaattatggccttcattggccagtcggctcctgacattaggaagaagttaca
gcgaattgagggcttgcaggattacaccataagggatgtagttagagaggcagagaaagtgtatcatagg
agagaaacagaagatgaaaagttagagagagagaaaagagagaaaagagaagatgaggataggagagaca
ggaggcaagaaaaggttttgactaggatcctggctgcagtaggagaaagagataatggaagaagaggtag
acagtcagggaacctgggagacaaaagacagcagggaccaaggagacccagagaaggcgggcagcgcctg

gagaggaaccaatgtgcatattgcaaggaaatgggccactggaagagcaactgtccggaaaaaaaaacaa
gaggtaaaggtgctttctcttggagaagatgaagactagggggaacggggcttgacccacctccccgagc
ctagggtaactttagaagtggaggggtcccctgtggactttctagttgacacgggagccgaattttcagt
actcaaaacacctctaggaaaagtgaagaaaaatgaaaaaaccttggtgatcggggccacgggacaaaaa
tcgtatccatggaccacatcccgagtagtagacatagggcgaaatcgagtaactcattcgtttctagtca
ttccagagtgtcctattccttattggggagagacttactaaccaagttaaaagcacaaataactttcac
ctctcatcgaccggaggttttctggggaataaaagcgccccagactctagagctgtctttacaactaggg
gaggaatatcgactttaccaaaataaagtaaagccccctgagggattacaggactggttgaatcgatacc
ctcaggcatgggcagagacgggaggagtggggatggcaaaactggtcccccccccgtggtgattgaactta
agtccggggccacccctataggggtccgacaatatcccatgagcagagaagctcaagagggtatacgccc
ccaaattaacaaactgctccaacaagggattttggtcccatgcaaatccccttggaacactcctctactt
ccagtaaaaaaaccagggaccagtgactaccgtccagtacaggaccttagagaagtcaacaagagagttc
aggacatacaccccacggtgccaaatccttataacctcctcagcaccttgccacctggtcggacatggta
cacagtcctggatctcaaagacacttttttctgtttgaggttacaccccaacagccagcccttgttcgct
ttcgaatggcgagactccgagagtggacaagccggacagctcacatggacgaggctgcctcagggattca
agaactcgcccactttgttcgatgaagccctacaccgagatcttgctcttttccgagccaataacccaca
ggtgactcttctgcaatatgtagatgacctgctcctagctgcagaaacacacgaggactgtgaaattggg
acctaaaacctcctgggcgagttaggtaacctggggtatcgggcctctgctaaaaaggctcagttatgcc
agatagaagtgacctacctaggatatgtcttgagagatagacaacggtggctcacagaagccagaaaaca
agctgttatgcagatcccgaccccaaccactgctcgccaggtaagagagttcctggggaccgccgggttt
tgcagactctggattcctggatttgccacactggcagctcccttgtatccactaaccaaagagaaagggg
aatttacctggaccagagaacatcagctagcctttgaaactctcaaaaaggcactgctgtaggctccggc
attggccctgccagatttaaacaaacctttcaccctatacattgatgaaagaaatggagtggcaaggggа
gtccttacccaggttttgggaccatggaagcgcccggtagcctacttatcaaagaaactggaccctgtgg
ccagtggatggccctcctgcctgcgagcgatagcagccacggctgtgctagtaagagatgctgacaaact
gactatgggccagaatgttactatagtggccccacactctcttgagagcatcatcaggcaaccactggac
cgctggatgaccaacgcccgaatgacgcactaccagagcctattgctgacagagcgagtaagttttgcac
ccccagccattctcaaccctgcctccttactacctgaggctgacgaggcccctgcacataagtgtgaaga
aatactggcagaagagactggaatccggccagacctcacagaccaaccttcgccaggggcgatgacttgg
ttcacagacggaagcagctttgtggtagaaggtaagcggagggctggggcagcagtagtggatggaaagt
ctgtcatatgggccagcagtctgctggagggtacatcagctcaaaaagcagaactaatcgcattaattca
agccttaaggctggcagaaggaagggctcttaatgtctatactgacagccggtacgcttttgccacggct
catgttcacagagcaatataccgacaccgtggactgctgacgtctgccggcaaagatatcaaaaataaag
aaggaattctcagcttattagaagctgttcatctgccccgtagggtggtaattttccattgcccaggaca
ccagaagggaactgggcccgttgaaaagggaaatcaaatggcagaccaagaagctaaaaaagcagcccaa
gggccaatgactctggtggtgagaacccaacagcccgctgctgaggaaataaataaaagaaccctcacag
aagaagaggggcgagattacttagctaacatacaccatctgactcatttaggaactaaaaaattactaaa
attggttagtaagtccccctattacattcctggattaaaaagaattgtggaagagatagtaaaaaactgc
cgtgcttgtgcacttaccaacgctgggtctagcaggctccaggaaagaaaacgactgcgaggagacaggc
ctggagcctactgggaaactgacttcactgaggtgaaaccggctaggtatggaaataaatatctcctagt
ttttatagacacctttcaggatgggtcaaagcattccccaccaagaaagaaacgactaatgtagtggtc
aagaagatacttgaagaaatccttccccgttttgggatacctaaggtaatggggtcagacaacagacctg
ccttcgtctcccaggtaagtcagggattggccagacaactggggacaaattggaaattacattgtgcata
cagaccccagagttcaggacaggtagaaaggatgaacagaacgctaaaggagactctgactaaaatagcc
ttagaatccggtggaagcgattggacagccattctcccttatgccttgttcagggttcagaatacacctg
gacccccttggcctaactccatttgaattaatgtatggggcgcccccacccatttttatgaccgtaggtga
taagaatcgcctggatgtgtctttctctcctccttctagtcttttggctcgattaaaagctctcgaaata
gtaagaaaagaggtctgggaacagctaaaagaaacctatgttgctggtgacacacaggtgccacatcagt
ttgaagtaggagacgcagtcctggtgaggagacaccgagcaggaaacctagaaccgaggtggaagggacc
ctacttggtgctactgacaacgcccaccgcggtcaaagtggaaggaatccccacttgggtccacgcatcc
cacgtcaagagagcacccccctggagtcagccatgatgagtggactttggagaagactactaatcctttta
agttgcgcctgcttcgtaggagcgatcccaaaagacttcaacccccacagtcctgttcaacaaacgtggg
aggtactcaatgaggagggtagggctgtatggacaatcgccgaggtacaccctctgtggacttggtggcc
tgatcttttccctgacatctgtaagttggctataggagcccctcctggatgggacttggaggggtactct
gacattcagagggcacctttaacacccccctccgtacgtagaaaaacatttgagagacccgtggggtggtt
gctctaaccaaagggatagaagtatgcttcgaacccatcccttctatgtctgccccgggccccaccaaag
tcagtccctcaatccaacgtgtggaggtaaggctgacttcttttgcaagagctggggttgtgagacttca

ggtacagcccgctggaagccctcctcgagctgggactatattagagtaacagccaactattccctagcat
cttatgtacctggaggatttgacctagacgagtgtactgactggtgccatccgctccgtgtcactttcac
tgaaccagggaagagagctctgggatggacaagagggtatacctggggtcttaggatttacaaggaaaga
tatgatgagggattattgttcactatcagattaaaaatagagacccttacaatcctttaggccccccaa
ccaagttcacacccctcacccatacaattactcagcctactccagtgattgcggacccccttaatatggc
cgctatcacccaacctcccactcctcaggtacctctaactattacccccacgattccttcaagacagagg
atgtttaacctagtgagaggagccttttatgcccttaacagaactgatccaagcgctactgaggactgct
ggctatgcctgtcctcgggtccgccttattatgaaggaatcgccttcaatggagatttcaacagaatcag
cagccatacttcctgctcttggggaacaggacaaaaactgaccctgactgaagtatccgtgaggaatcca
ggtctctgtataggcaccccaccttccactcacaaacacctatgcggacaaattcagtccatgtccagaa
cagaagctaattactatcttgtaccttccccggttggatggtgggcttgcaatacaggacttactccctg
tgtatcaactaaggtttttaattcatctcatgatttttgtgtcatgatacagctgttaccccacgtatat
tatcaccctgcatccagtttagaagaaagctatgctggccggaggtcaaaaagagaaccaactactttaa
ccctggctgcattcatgggaataggtatggcagtaggagtggggacgggagtgtcagctttgatagaagg
aagacagggaattcagtctttgagggatgctgtcaatgaagacctagcggcaatagagaagtccattgac
gctttaaaaaaatctttgacctccctgtctgaggtagttttgcagaacaggagaggtcttgatttgttgt
tcctcaaggaaggaggactgtgtgctgcccttaaagaagagtgctgcttctatgcagatcatacaggaat
agttagagactctatgcagaaactgagagaaaaattagagcgaaggaaaccggaacgggatgctcaacgg
gggtggtttgagtcgtggtttgaatcacgaccatcttggataacttctttaatttccgctgtagccggac
caatccttatgatatgcttagctttagttttcagcccttgtataataaatagaggaatggctttcatcca
gagtaaaattgatacagtaaaactcatggttcttcaaaggcaatatcaacctatagttcaggtagatgaa
gagttaggggacaccaatctctaaaattctatgattagaattagtctaaacagaagaagaggggaa

Chapter 5: Future Directions

Endogenous Retroviruses:
Cancer Initiation Markers and Inflammation Mediators?

Triggering the immune system to attack regions of the body which are undergoing cellular transformation is the overarching goal of cancer immunotherapy. However, the earliest events that activate an immune response following oncogenic stress remain unknown. This knowledge would provide mechanistic insights into tumorigenesis with the potential to aid in the development of better cancer drugs. To eradicate cancer we ideally wish to target cells at the onset of oncogene activation and simultaneously prevent a cancer promoting milieu in the surrounding tissues. Thus, we wish to identify molecules which serve as early markers of cancer initiation that also promote unfavorable conditions in the surrounding tissues. These key mediators should exhibit two intrinsic features:

1. in response to cellular stress, they should be upregulated at the protein level, and
2. their upregulation should be associated with a distinct immune response *in vivo.*

We have recently identified a molecule that may fit both of these criteria. We developed a genetically engineered mouse strain which harbors an allele that results in the constitutive activation of the universal stress sensor, p53. In this

model, we observe p53 dependent activation of inflammatory signaling pathways. This model led us to seek out the molecular connection between the p53 signaling pathway and inflammation. Interestingly, we discovered a murine endogenous retrovirus (ERV), *Mmergln-int*, whose expression is dependent on the activity of p53. p53 is a tumor suppressor that is upregulated in the early stages of many cancers, and subsequently mutated during malignant transformation. We found that envelope (*env*) of *Mmergln-int* is sufficient to reduce cell viability independent of the other viral components *ex vivo*. *Mmergln-int env* is characterized by a domain that is found in human T-cell leukemia virus type 1, an oncogenic retrovirus that also causes chronic inflammation. We discovered this protein domain family is endogenized in the genomes of multiple vertebrates, including humans, potentially highlighting that these sequences are functional. Interestingly, in human cell lines cellular stress upregulate the expression of distinct ERV envelopes. With this taken together, we propose that *Mmergln-int env* and homologous human ERV (hERV) are the early transducers of cancer initiation events stress which activate an immune response during tumorigenesis.

Hypothesis and Objectives: We hypothesize that ERVs are upregulated during cancer initiation, as a result of cellular senescence observed in premalignant lesions of many types of cancers. Furthermore, we hypothesize that the expression of ERV envelopes induces an immune response to promote the elimination of precancerous cells. Our objectives are to determine the role of p53

in modulating hERV expression human prostate cancer and to delineate the consequences of ERV envelope expression in modulating immune response.

Specific Aim 1: Determine the role of p53 in inducing the expression of human ERV in prostate premalignant lesions and adenocarcinomas

a) Determine the correlation between hERV transcript level and p53 status in prostate adenocarcinoma
b) Determine if hERV transcript level is a marker of prostate premalignant lesions in prostate tissue and serum
c) Develop monoclonal antibodies against hERV envelopes to evaluate protein expression levels in human prostate tissue and serum
d) Investigate the effect of oncogene activation or loss of tumor suppression on hERV expression in human primary prostate fibroblasts

Specific Aim 2: Determine the physiological effects of ERV envelope expression.

a) Evaluate the effect of *Mmergln-int env* expression on tumor volume
b) Determine the ability of *Mmergln-int env* to promote migration of immune cells to tumor *in vivo*
c) Investigate the ability of hERV envelopes to promote migration of immune cells *ex vivo*

Significance: This research aims to elucidate the function poorly understood genomic elements that make up at least 8% of the human genome. It proposes to address undiscovered two aspects of tumorigenesis i.e. what key molecules contribute to cancer initiation and how do they contribute to the elimination of precancerous cells. The implications of these findings may contribute to improved diagnostics and therapeutics for cancer.

SIGNIFICANCE

Understanding how tumors form is essential to eradicate cancer. Therefore the proposed research attempts to address two major unknown questions regarding tumorigenesis, namely 1) what are the earliest molecular markers of cancer initiation and 2) how do these molecules interact with the immune system to promote elimination of precancerous lesions or support the expansion of tumor cells. The answers to these questions could lead to better therapeutic measures to detect, prevent and treat cancer. Our current immunotherapy strategies target antigens found in tumors not premalignant cells. Thus, identification of markers of cancer initiation, like human endogenous retroviral envelopes, could provide new therapeutic targets that allow for earlier medical interventions. The proposed research could lead to novel diagnostic procedures and the development of monoclonal antibodies for adjuvant therapies for prostate cancer.

In addition to studying the molecular underpinnings of tumorigenesis, this proposal aims to investigate the function of a portion of the genome that is poorly

understood. Endogenous retroviruses are a class of repetitive elements that comprise at least 8% of the human genome (99). Despite the fact that these genetic elements make up such a substantial proportion of the human genome, their functions have remained elusive due to technical limitations in accurately detecting their expression. RNA Sequencing (RNA Seq) in combination with bioinformatics tools to analyze repetitive sequences will finally allow us to explore in what context these genomic loci are expressed. Using this approach, the proposed research aims to ascribe a function to endogenous retroviruses with respect to tumor suppression.

APPROACH

Preliminary Studies

The molecular mechanism by which an organism responds to stress is highly conserved. Genes homologous to the universal stress sensor, p53 exists even in the genomes of early animals such as anemones and worms (100). Stressors, such as DNA damage or oncogene expression, lead to the stabilization of the p53 protein. p53 functions as a transcription factor to activate the expression of genes involved in processes which maintain the integrity of the genome like DNA repair, halting the cell cycle or apoptosis (43, 44, 101). One mainstay of cancer research is determining p53 target genes to better understand its function as a tumor suppressor. Although, loss of p53 is one of the most frequent genetic aberrations in human cancer, this is a typically a late stage event occurring often during the process of malignant transformation (101). One

mechanism by which p53 acts as a tumor suppressor is to induce cellular senescence. The expression of oncogenes such as Ras, Raf and Akt or inactivation of tumor supressors such as PTEN or NF1 enhance p53 activity to induce cellular senescence *in vivo* (102, 103). Thus, examining the physiological outcome of constitutively activated p53 may help us to discover the molecular hallmarks of a developing premalignant lesion.

In order to determine the physiological consequences of the constitutive activation of p53, we took advantage of a mouse model containing a duplication (*dp/+*) of an allele that positively regulates p53 expression. Although the mice with the *dp/+* allele die at birth due to extreme senescence and apoptosis (Figure 26A), this phenotype is completely rescued in the absence of p53 (Figure 26B) (56, 104). To better understand the phenotype of excess p53, the transcriptome of *dp/+* mouse embryonic fibroblasts (MEFs) compared to +/+ MEFs. Interestingly, chemokines associated with an NFkB mediated inflammatory response were the most significantly upregulated genes in *dp/+* MEFs (Figure 26C). We wished to validate this phenotype *in vivo*, however the *dp/+* mice exhibit perinatal lethality on postnatal day 1. Therefore we generated *dp/+* mice in which both p53 alleles were inactivated through the introduction of a stop cassette flanked with loxp sites in intron one of the p53 gene (*dp/+, p53 lsl/lsl*). These mice were further crossed to carry the Cre ERT2 recombinase under the Rosa26 promoter. This combination of alleles allowed us to turn on the expression of p53 body wide by the administration of tamoxifen. The reactivation of p53 *p53 lsl/lsl* mice resulted in no significant changes, however the *dp/+, p53*

lsl/lsl mice were moribund within weeks of receiving the tamoxifen injections. Histopathological analysis revealed inflammatory foci in adipose tissue and mesentery in *dp/+, lsl/lsl* mice demonstrating that constitutive activation of p53 results in inflammation *in vivo*.

In the interest of gaining a more complete understanding of how p53 contributes to the activation of an inflammatory response, we coupled RNA Seq with a novel data analysis tool, AITA that detects the expression of regions of the genome that for which no annotated exists (Annotation Independent Transcriptome Assembly). A substantial proportion of mammalian genomes are comprised of poorly or non-annotated repetitive sequences whose functions remain unknown. With repetitive elements make up 30-50% of mammalian genomes (25), and a recent computational analysis demonstrating that 66-69% of the human genome is repetitive in nature (105), it is likely these regions of the genome contain undiscovered functional sequences which contribute to p53 mediated tumor suppression.

Using AITA, we discovered an non-annotated transcript derived from a LTR retrotransposon named *Mmergln-int* whose expression was not detected in the absence of p53 (Figure 27C, D, E). Transposons are the major source of repetitive sequences in mammalian genomes, and it is hypothesized that stress activates their expression to enhance the fitness of an organism (26, 83). Indeed, numerous studies demonstrate the transposons are activated in response to various forms of stress from bacteria to humans (23-25, 82-84) and differential expression of distinct families of transposons has been reported among different

cell and tissue types (25). However the specific functions of these transcripts are yet to be revealed. Therefore, we wished to determine the role of *Mmergln-int* with respect to p53 function. We demonstrate that p53 regulates the expression of *Mmergln-int* by binding a highly conserved p53 response element within the LTR, and that the LTR is capable of driving the expression of a luciferase reporter in the presence of p53 (data not shown). We detect transcriptional upregulation of *Mmergln-int* in response to genotoxic stress in *p53+/+* MEFS and confirm *Mmergln-int* is not expressed in *p53-/-* MEFs (Figure 27D).

After showing that p53 regulates the expression of *Mmergln-int*, we investigated how its expression influences cell viability. Overexpression of *Mmergln-int* in leads to decreases in cell viability in *p53+/+* MEFS, and the transient overexpression of the *env* is sufficient to induce this change in *p53+/+* MEFs (Figure 28). A further examination of *Mmergln-int's env* protein reveals that it is classified as member of the HTLV-1-like heptad repeat 1-heptad repeat 2 protein domain family. This highly conserved domain is present in modern infectious viruses and is also found in endogenized in the genomes of many vertebrate species, including humans. We observe an ~2 fold upregulation of *envR*, *env1*, and *envW* transcripts in human MCF7 cells after UV exposure, a known agent to induce p53. Although very few antibodies exist to detect expression of human endogenous retroviruses, we were able to confirm that protein level of *envW* increases at 8 and 24 hours after UV exposure relative to non-treated MCF7cells (Figure 28).

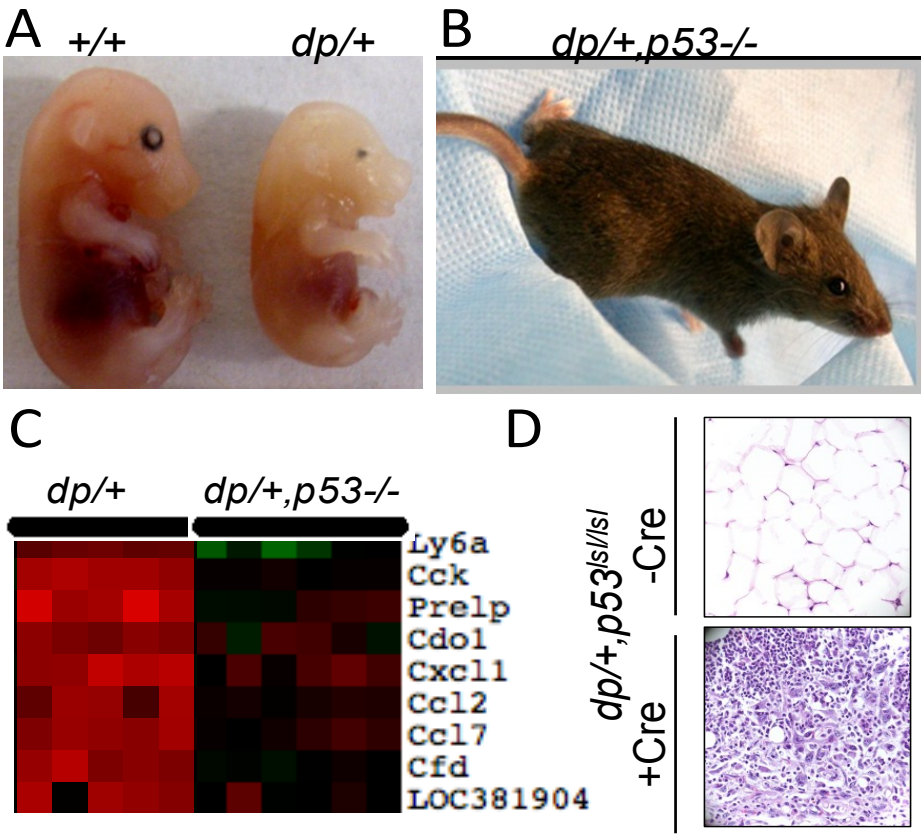


Figure 26. Constitutive activation of p53 induces an inflammatory response

(A) dp/+ mice exhibit high levels of p53 and p21 which results in perinatal lethality

(B) Lethality is rescued in dp/+, p53-/- mice

(C) Microarray analysis demonstrates inflammatory chemokines are transcriptionally upregulated in *dp/+* MEFs

(D) Inducing the expression of p53 in *dp/+, p53 lsl/lsl* 8 week old mice results in inflammation of adipose tissue and mesentery which is especially severe adjacent to the pancreas that is not observed in *lsl/lsl* mice. Inflammatory foci are and abnormal fibroblastic like cells are present in *dp/+ lsl/lsl* mice.

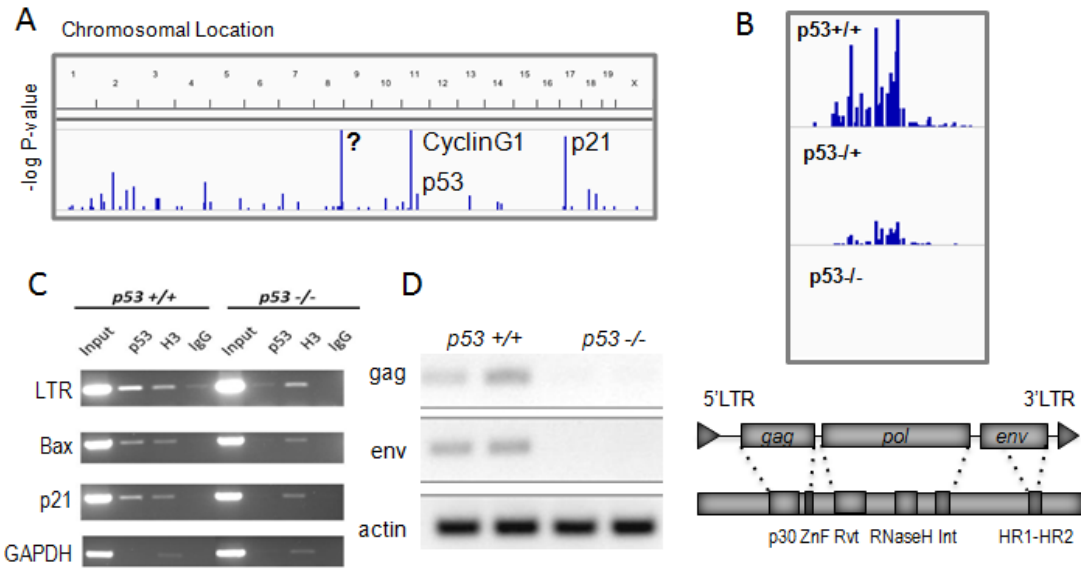


Figure 27. Annotation Independent Transcriptome Assembly identifies Mmergln-int in dp/+ MEFs

(A) AITA identifies a differentially expressed transcript in *dp/+* MEFS which is non-annotated

(B) The transcript encodes for Mmergln-int, an ERV which demonstrates p53 dependent expression

(C) p53 binds the LTR of Mmergln-int similar to other p53 target genes Bax and p21

(D) RT PCR confirms expression of Mmergln-in *p53+/+* but *p53-/-* MEFs

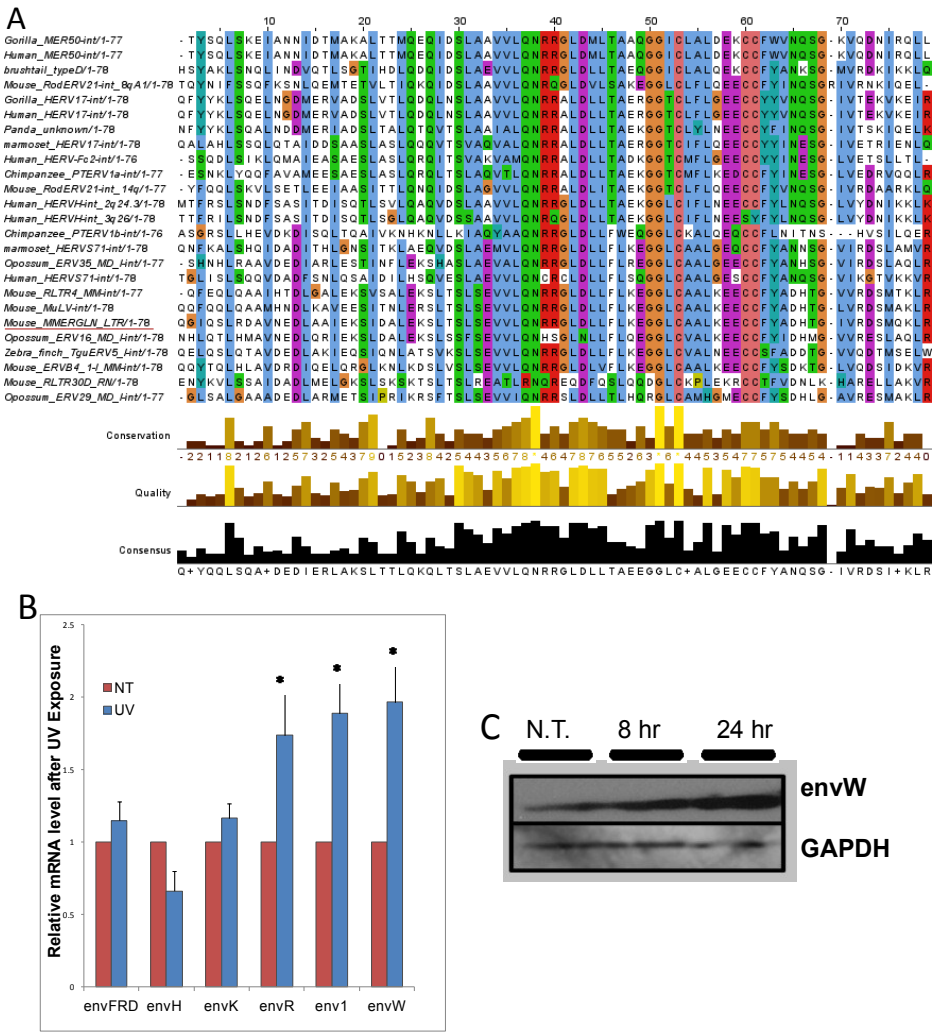


Figure 28. Envelopes homologous to HTLV-1 are endogenized in the genomes of multiple vertebrate species and some are stress responsive in humans

(A) The HTLV 1-like HR1-HR2 domain was identified in human, gorilla, chimpanzee, bushtail, mouse panda, marmoset, opossum, and zebra finch. Amino acids are highlighted with different colors to depict conservation among organisms. The degree of conservation is illustrated by calculating the conservation, quality and consensus of the alignments (black and yellow bar graph)
(B) *envR*, *env1* and *envW* are transcriptionally upregulated in response to UV stress in human Mcf7 cells and
(C) envW is upregulated at the protein level at 12 and 24 hr after UV exposure.

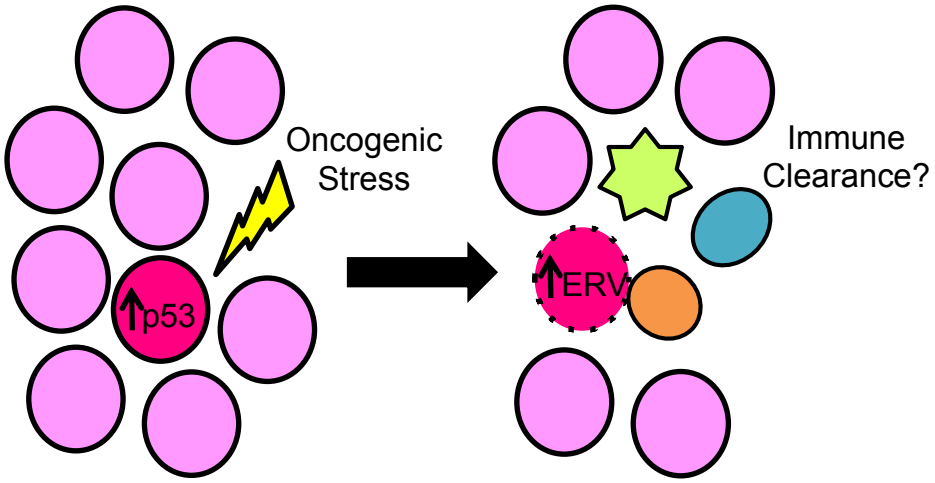


Figure 29. p53 activates the expression of ERV envelopes to promote the clearance of premalignant cells

Our model proposes that in premalignant lesions endogenous retrovirus are upregulated in response to enhanced p53 activity. The expression of transmembrane ERV envelopes activates an immune response which may promote the elimination of precancerous cells.

Specific Aim 1: Determine the role of p53 in inducing the expression of hERV in prostate premalignant lesions and adenocarcinomas

We will investigate the expression of human ERVs in prostate tissue, because of the high rate of prostate cancer in the male population and the severe limitations in current prostate cancer screening procedures. According to the American Cancer Society, approximately 1 in 6 men will be diagnosed with prostate cancer in their lifetime. The Center for Disease Control and Prevention declared prostate cancer to be the most common cancer among men of all races, and the second leading cause of cancer related deaths in men. Furthermore, since many men have no symptoms or very mild symptoms, such as difficult or frequent urination, screening procedures are imperative in preventing the development of life threatening forms of prostate cancer. Recently, the US Preventative Services Task Force (USPSTF) recommended that men of all ages should no longer be screened for prostate cancer, because the potential benefits of the current screening procedures do not outweigh the harm they cause (106). Prostate cancer screening based on Prostate Specific Antigen (PSA) levels prevents 0-1 prostate cancer deaths per year. In fact, the USPSTF found current diagnostic procedures to be harmful, since over 10% of men screened receive a false positive resulting in a biopsy with potential side effects including severe infection, bleeding, pain and anxiety. Furthermore, in the event that prostate cancer is accurately diagnosed at an early stage, no histological or molecular marker can predict is a premalignant lesion will progress to adenocarcinoma (106-109). Considering the side effects of prostate cancer treatment include

range from sexual dysfunction, bladder and bowel incontinence to death, it is essential to better define individuals who will develop life-threatening cancer (106). These findings highlight an extreme need for more basic science research to identify markers of cancer initiation and progression.

We wish to study expression levels of hERVs in prostate tissue during cancer initiation and progression. We demonstrated that distinct human ERV envelopes are upregulated in human breast adenocarcinoma cell lines after induction of p53 (Figure 3B,C) and preliminary data show that in *envFc1* and *envFr*d are upregulated in human LNCaP prostate adenocarcinoma cell lines in response to genotoxic stress. This finding is consistent with a report indicating that over 30% of human p53 binding sites reside in endogenous retroviral LTRs (110, 111) Therefore, we hypothesize that senescent cells, which are known to express high levels of p53, can also be distinguished by increased levels of ERV transcripts and proteins. Furthermore, we hypothesize that premalignant lesions, which are enriched with senescent cells due to the activation of oncogenes or loss of tumor suppressors (112, 113), express higher levels of ERVs than benign tissue. We speculate that the activation of human ERV expression may serve as a general marker of cancer initiation with diagnostic and therapeutic implications. We chose to focus first on prostate cancer not only because of dire need to develop better diagnostic procedures, but also due to distinct molecular hallmarks of the disease. Previous studies detect transcription of hERV-K (114) in prostate cancer tissue and antigen in serum of prostate cancer patients (115). Furthermore, one of the few known markers premalignant lesions of the prostate

are elevated p53 levels (106, 116). Though few genes which promote prostate cancer initiation have been discovered, aberrant expression of Myc and the PI3K/PTEN/AKT signaling pathway contribute to enhanced p53 expression (117-122). Therefore, we propose:

a) Determine the correlation between hERV transcript level and p53 status in prostate adenocarcinoma

In order to establish if a correlation exists between human ERV expression and p53 expression, we will analyze publically available human prostate adenocarcinoma transcriptome data from The Cancer Genome Atlas. Using our novel AITA approach, we will determine if expression levels of hERV correlate with enhanced p53 activity. (see letter of collaboration, Sarver)

b) Determine if hERV transcript level is a marker of prostate premalignant lesions in prostate tissue and serum

Transcriptome analysis will be performed on de-identified, staged human prostate tissue and serum from urologist Dr. Badrinath Konety at the Masonic Cancer Center (see letter of collaboration, Konety). Biostatistical analysis projects that less than 10 samples per stage will be needed in order to obtain statically significant results. RNA and protein will be harvested from the tissue samples to compare the expression of level of hERV of staged, high-grade prostatic intraepithelial neoplasia (HGPIN, grade PIN2 or PIN3) and acinar-type adenocarcinoma respect to benign tissue. Protein will be harvested in RIPA buffer supplemented with protease inhibitors and will be stored for later analysis. RNA Seq will be used to analyze the transcriptome of tissue samples. Briefly,

RNA will be isolated from tissue with Qiagen's RNeasy Minikit and from serum with miRNeasy Serum/Plasma kit. The RNA integrity is to be verified by quantification with a Ribo Green Assay and by running the samples on an Agilent Nano chip. Samples with a RIN number of 8 or above will be selected for RNA Seq library preparation. For library preparation, the mRNA will be purified and fragmented. Next, cDNA will be synthesized, ends repair, and the 3' ends will be adenylated for ligation to adaptors. The library will be quantified and validated using an Agilent High Sensitivity Chip, Pico Green assay and KAPA qPCR. Differential expression of p53 target genes and hERVs will be determined by bioinformatics analysis. We will use quantitative reverse transcription polymerase chain reaction (qRT-PCR) to validate the expression levels of hERVs (90) and selected p53 target genes. We will determine if it is possible to detect expression of hERVs in serum by qRT-PCR. For hERV sequence analysis, the consensus sequences will be extracted from Repbase (87) and mapped to hg19 with the UCSC Genome Browser (123).

c) Develop monoclonal antibodies against hERV envelopes to evaluate protein expression levels in human prostate tissue and serum

To further investigate hERV expression levels in human prostate precancerous lesions and adenocarcinomas, we will produce antibodies against transcriptionally upregulated hERVs for further characterize expression levels as previously described (124). To generate the immunogen, the ERV env will be cloned into pcDNA3.3 vector and the recombinant protein will be harvested from CHO cells. Female white leghorn chickens will be immunized with 200 μg of the

recombinant protein bimonthly, and bled one a week post injection to evaluate antibody titer. To clone the variable gene rearrangements of the heavy (V_H) and light (V_λ) chains of immunoglobulin, RNA will be isolated from the spleen of the chickens and reverse transcribed with PCR primers V_H and V_λ primers. The size of the PCR products will be confirmed on an agarose gel, then isolated from the gel for overlap extension PCR to construct the single chain variable fragments (scFv). The overlap extension scFv PCRs will be purified from an agarose gel, restriction digested and ligated into the pComb3HSS vector. The library will be transformed into *E. coli* and supplemented with the VCSM13 helper phage. To establish scFv that bind the antigen with high affinity, the library will be panned several times. The phages which bind the immobilized antigen are eluted then re-amplified up to 6 rounds, before analysis by western blot and flow cytometry. The most specific scFvs will expressed in *E. coli* and purified by metal chelate chromatography. The protein levels of p53, p16, p21 and the human ERVs will be evaluated in prostate premalignant lesions and prostate adenocarcinoma tissues and serum by western blot.

d) Investigate the effect of oncogene activation or loss of tumor suppression on hERV expression in human primary prostate fibroblasts

To better understand the molecular events which occur during prostate cancer initiation, we will overexpress oncogene $Hras^{V12}$ and knockout tumor suppressor PTEN in primary human prostate epithelial cells. The $Hras^{V12}$ expression construct or GFP control will be transduced in prostate epithelial cells. Transcription activator-like effector nucleases (TALENs TAL2344, TAL2345) will

be used to delete the PTEN locus. (125) We will assay for markers of senescence p16 and p21 by western blot and senescence-associated β-galactosidase expression. The expression of human ERV envelope transcripts will be determined by qRT-PCR (126).

Outcomes, potential problems alternative approach

We expect to observe upregulation of hERV envelopes such as *envFc1* and *envFRD* at the transcript and protein level in premalignant lesions of the prostate but not in benign tissue samples. We expect this will correlate with the induction of the p53 target genes, and p53, p16 and p21 protein levels. Moreover, we predict that in prostate adenocarcinoma tissue we will see a decrease in hERV expression levels the associated with p53 loss of function. These findings would suggest that if an individual demonstrates a decrease in hERV expression level their cancer maybe progressing to a more advanced stage. Based on previous studies, we expect to detect hERV protein in the serum of patients (115, 127). Future studies could focus on the validation of hERVs as early cancer biomarkers. Additionally, we could determine if the monoclonal antibodies against hERVs can be used to target xenografted tumors derived from human prostate adenocarcinoma cells *in vivo*.

In the event we are not able to observe activation of hERV expression in RNA Seq data from prostate tissue, we will choose an alternative cancers to generate a basis for further molecular analysis such as lung, liver or colon adenocarcinoma, breast ductal carcinoma in situ (DCIS) or pancreatic

intraepithelial lesions (PANIN 1,2,3). Premalignant lesions of these tissues are also characterized by senescence *in vivo* (102, 116). If there is a delay in obtaining an adequate number of samples or we need an alternative tissue type, we will obtain tissues from Bionet, the tissue procurement facility at the University of Minnesota (see letter of collaboration, Bouley). With over 20,000 samples, this facility serves as a centralized resource for de-identified, standardized, staged benign and malignant tissues from human patients for researchers across the Unites States. Finally, if it is not possible to detect human ERV expression in serum samples, but an increase in expression is seen in prostate premalignant lesions, we will use SELDI-TOF Mass Spectrometry and use prostate secretion for molecular analysis, because it is a less invasive source of prostate cells than needle biopsy (B. Konety, personal communication).

Specific Aim 2: Determine the physiological effects of ERV envelope expression

We identified stress responsive ERV envelopes in human and mouse that share a high degree of homology among to the human pathogenic virus, HTLV-1. Interestingly, previously studies observed that while the *gag* and *pro-pol* genes of ERVs accumulate mutations that render them inactive, there is selective pressure to maintain the *env* suggesting it is functional (91, 128). Although the precise mechanism by which these glycoproteins decrease cellular viability is yet to be uncovered, the upregulation of ERV envelopes is associated with the activation of p53 mediated cell death pathways. Furthermore, multiple studies

demonstrate that the expression of ERV correlates with disease states associated with inflammation. For example, aberrant expression of *envW* evokes an endoplasmic reticulum stress response as well as inflammation in the brain in MS patients (93). Furthermore, *envW* has even been employed as an oncolytic agent which when overexpressed shrinks tumors *in vivo* by induction of apoptosis (94). Thus, the correlation between the expression of endogenous retroviral envelopes in response to cellular stress and the induction of cell death pathways suggests they may have tumor suppressive properties yet possess the ability to induce an immune response (Fig. 29). In order to understand the physiologically effects of ERV envelope expression, we aim to:

a) Evaluate the effect of *Mmergln-int env* expression on tumor volume

Our previous studies demonstrate that overexpression of *Mmergln-int* results in a decrease in cell viability. We hypothesize the ERV envelopes may contribute to p53 mediated tumor suppression and possess the ability to inhibit the growth of tumors *in vivo*. To test this hypothesis, we will generate cell lines in which the expression of *Mmergln-int env* or a GFP control can be induced. The envelope of *Mmergln-int* or GFP will be regulated under in a doxycycline inducible promoter with and contain IRES-luciferase reporter for *in vivo* imaging. We will stably integrate the constructs in Lewis lung carcinoma cells (LLC1), a cell line isolated from lung tumor of C57BL mouse which is tumorigenic in mice (129, 130). We will transplant 10^5 cells of with the doxycycline inducible *Mmergln-int env* or GFP control LLC1 cells into the right suprascapular region 10 mice per

cell line. At 48 hours post injection, doxycycline will be administered (625mg/kg feed or 10 mg/ml gavage daily) (131). The tumor volume and metastasis will be determined by external caliper measurement and *in vivo* imaging of live mice as described by Jenkins *et al* (132).

b) Determine the ability of Mmergln-int *env* to promote migration of immune cells to tumor *in vivo*

We hypothesize that expression of ERV envelopes activates an immune response which contributes to the clearance of cells with enhanced p53 activity. We predict that the induction of *Mmergln-int env* expression will evoke an immune response *in vivo*. To determine this, we will assess the infiltration cells of T cells and polarized M1 and M2 macrophages of the 3 weeks post doxycycline induction, the mice will be euthanized, the tumor tissue will be harvested to determine the presence of immune cell infiltration by flow cytometry. Tumor tissues will be dissociated with collagenase, and stained with T cells by flow cytometry using anti–CD4-PE, anti–CD8-PerCP-Cy5.5, and anti–TCR-β and macrophages by anti-CD86 and anti-MCR1.

c) Investigate the ability of hERV envelopes to promote migration of immune cells *in vitro*

The role of the immune system in cancer is seemly contradictory; whereas the innate and adaptive immune response contributes to recognizing and destroying cancer cells during the process of cancer immunosurveillance, the

inflammatory tumor microenvironment promotes tumorigenesis (133, 134). In order to better understand the general role of hERV envelopes, we wish to study their ability to promote migration of immune cells. Using the Conserved Domain Database (88), we determined that hERVs of the ENV1-like and Rb-like domain families contain CKS17 peptides in their envelopes. CKS17 domains suppress lymphocyte proliferation, induce changes in cytokine production, and suppress pathways involved in T cell activation (135). Thus, it is possible that some hERV envelopes promote tumor clearance while other contribute to cancer progression. To determine the unique contributions of hERV envelopes we will test the ability of hERV promote T cell chemotaxis by *in vitro* cell migration assays (136). Briefly, we will harvest T cells from 6 mn old mice thymi and plate them in transwell chambers with 5μm chemotaxis inserts. The lower chamber will contain either medium harvested from cells transiently overexpressing a hERV from the HTLV-1-like, ENV1-like and Rb-like domain families. Medium with 1 μg/ml recombinant mouse CCL21 will used as a positive control. We will quantify the number of T cells by flow cytometry using anti–CD4-PE, anti–CD8-PerCP-Cy5.5, and anti–TCR-β. To obtain the percentage migration, the number of CD4-positive/CD8-negative/TCR-βhi cells will be counted and divided by input number.

Outcomes, potential problems alternative approach

We hypothesize that the upregulation of ERV envelopes induce an immune response. We this would be apparent by the infiltration of macrophages,

lymphocytes, Natural Killer cells, neutrophils or dendritic cells in tumors upon induction of *Mmergln-int env* whereas fewer or a different class of immune cells would be detected in the tumors of the GFP expressing control cohorts. We would also expect a reduction in tumor burden and metastasis in mice with allografts expressing *Mmergln-int* compared to the GFP control. It is possible that we observe no significant difference between the experimental and control mice, which could be possibly attributed to the high tumorigenicity of the LLC1 cell line. This could mask significant reduction in tumor size, or the LLC1 transplantation may alone promote an inflammatory response despite the fact that syngenic cells are engrafted. In the event we do not detect differences in cancer progression or immune response, we will selectively induce the expression of *Mmergln-int* or GFP in the livers of FAH-/- mice. First, will retrofit the constructs outlined in part b into a Sleeping Beauty transposon gene delivery vector and add the fumaryl acetoacetate hydrolase (FAH) gene. The constructs will be delivered to hepatocytes by hydrodynamic tail vein injection and we would select for the integration of DNA by removing the NTBC, which supplements the FAH deficiency, from the drinking water. After induction with dox, we would harvest the liver and assess infiltration of immune cells. We suspect the ERV envelopes posses distinct features which contribute the paradoxical. With over 1,500 perfect p53 binding sites in the promoters of endogenous retroviruses in humans (110), it is intriguing to speculate about how the contribute to the stress response in the state of human disease.

Bibliography

1. Crick FH. The genetic code—yesterday, today, and tomorrow. Cold spring harbor symposia on quantitative biology; Cold Spring Harbor Laboratory Press; 1966.

2. Collins F, Lander E, Rogers J, Waterston R, Conso I. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931-45.

3. Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R. Molecular biology of the gene. 7th ed. Pearson; 2014.

4. Sparmann A, van Lohuizen M. Polycomb silencers control cell fate, development and cancer. Nature Reviews Cancer. 2006;6(11):846-56.

5. Branco MR, Pombo A. Chromosome organization: New facts, new models. Trends Cell Biol. 2007;17(3):127-34.

6. Wang GG, Allis CD, Chi P. Chromatin remodeling and cancer, part I: Covalent histone modifications. Trends Mol Med. 2007;13(9):363-72.

7. Wang GG, Allis CD, Chi P. Chromatin remodeling and cancer, part II: ATP-dependent chromatin remodeling. Trends Mol Med. 2007;13(9):373-80.

8. Berger SL. The complex language of chromatin regulation during transcription. Nature. 2007;447(7143):407-12.

9. Shilatifard A. Chromatin modifications by methylation and ubiquitination: Implications in the regulation of gene expression. Annu Rev Biochem. 2006;75:243-69.

10. Schones DE, Zhao K. Genome-wide approaches to studying chromatin modifications. Nature Reviews Genetics. 2008;9(3):179-91.

11. Ringrose L, Paro R. Polycomb/trithorax response elements and epigenetic memory of cell identity. Development. 2007;134(2):223-32.

12. Spector DL. The dynamics of chromosome organization and gene regulation. Annu Rev Biochem. 2003;72(1):573-608.

13. Kumaran RI, Thakar R, Spector DL. Chromatin dynamics and gene positioning. Cell. 2008;132(6):929-34.

14. Heslop-Harrison J, Bennett M. Nuclear architecture in plants. Trends in Genetics. 1990;6:401-5.

15. Cremer T, Cremer C, Baumann H, Luedtke E, Sperling K, Teuber V, et al. Rabl's model of the interphase chromosome arrangement tested in chinise hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. Hum Genet. 1982;60(1):46-56.

16. Marshall WF, Dernburg AF, Harmon B, Agard DA, Sedat JW. Specific interactions of chromatin with the nuclear envelope: Positional determination within the nucleus in drosophila melanogaster. Mol Biol Cell. 1996;7(5):825.

17. Cook PR. A model for all genomes: The role of transcription factories. J Mol Biol. 2010;395(1):1-10.

18. Fraser P, Bickmore W. Nuclear organization of the genome and the potential for gene regulation. Nature. 2007;447(7143):413-7.

19. Misteli T. Beyond the sequence: Cellular organization of genome function. Cell. 2007;128(4):787-800.

20. Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. 2005;24(4):800-12.

21. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. Cytogenetic and genome research. 2005;110(1-4):462-7.

22. Craig JM, Bickmore WA. Genes and genomes: Chromosome bands–flavours to savour. Bioessays. 1993;15(5):349-54.

23. Beauregard A, Curcio MJ, Belfort M. The take and give between retrotransposable elements and their hosts. Annu Rev Genet. 2008;42:587.

24. Bannert N, Kurth R. Retroelements and the human genome: New perspectives on an old relation. Proc Natl Acad Sci U S A. 2004;101(Suppl 2):14572-9.

25. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet. 2009;41(5):563-71.

26. Oliver KR, Greene WK. Transposable elements: Powerful facilitators of evolution. Bioessays. 2009;31(7):703-14.

27. Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, et al. Retrohoming of a bacterial group II intron: Mobility via complete reverse splicing, independent of homologous DNA recombination. Cell. 1998;94(4):451-62.

28. Kazazian HH. Mobile elements: Drivers of genome evolution. Science. 2004;303(5664):1626-32.

29. Servomaa K, Rytömaa T. UV light and ionizing radiations cause programmed death of rat chloroleukaemia cells by inducing retropositions of a mobile DNA element (L1Rn). Int J Radiat Biol. 1990;57(2):331-43.

30. Stribinskis V, Ramos KS. Activation of human long interspersed nuclear element 1 retrotransposition by benzo (a) pyrene, an ubiquitous environmental carcinogen. Cancer Res. 2006;66(5):2616-20.

31. Kurth R, Bannert N. Beneficial and detrimental effects of human endogenous retroviruses. International journal of cancer. 2010;126(2):306-14.

32. Pennisi E. Shining a light on the genome's' dark matter'. Science. 2010;330(6011):1614-.

33. Belyi VA, Ak P, Markert E, Wang H, Hu W, Puzio-Kuter A, et al. The origins and evolution of the p53 family of genes. Cold Spring Harbor perspectives in biology. 2010;2(6).

34. Lin J, Chen J, Elenbaas B, Levine AJ. Several hydrophobic amino acids in the p53 amino-terminal domain are required for transcriptional activation, binding to mdm-2 and the adenovirus 5 E1B 55-kD protein. Genes Dev. 1994;8(10):1235-46.

35. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science. 1996;274(5289):948-53.

36. Walker KK, Levine AJ. Identification of a novel p53 functional domain that is necessary for efficient growth suppression. Proceedings of the National Academy of Sciences. 1996;93(26):15335-40.

37. El-Deiry WS, Kern SE, Pietenpol JA, Kinzler KW, Vogelstein B. Definition of a consensus binding site for p53. Nat Genet. 1992;1(1):45-9.

38. Kitayner M, Rozenberg H, Kessler N, Rabinovich D, Shaulov L, Haran TE, et al. Structural basis of DNA recognition by p53 tetramers. Mol Cell. 2006;22(6):741-53.

39. Jayaraman L, Prives C. Activation of p53 sequence-specific DNA bindingby short single strands of DNA requires the p53 C-terminus. Cell. 1995;81(7):1021-9.

40. Bourdon J, Fernandes K, Murray-Zmijewski F, Liu G, Diot A, Xirodimas DP, et al. p53 isoforms can regulate p53 transcriptional activity. Genes Dev. 2005;19(18):2122-37.

41. Meek DW, Anderson CW. Posttranslational modification of p53: Cooperative integrators of function. Cold Spring Harbor perspectives in biology. 2009;1(6).

42. Laptenko O, Prives C. Transcriptional regulation by p53: One protein, many possibilities. Cell Death & Differentiation. 2006;13(6):951-61.

43. Beckerman R, Prives C. Transcriptional regulation by p53. Cold Spring Harbor perspectives in biology. 2010;2(8).

44. Vousden KH, Prives C. Blinded by the light: The growing complexity of p53. Cell. 2009;137(3):413-31.

45. Ventura A, Kirsch DG, McLaughlin ME, Tuveson DA, Grimm J, Lintault L, et al. Restoration of p53 function leads to tumour regression in vivo. Nature. 2007;445(7128):661-5.

46. Lane DP, Cheok CF, Lain S. p53-based cancer therapy. Cold Spring Harbor perspectives in biology. 2010;2(9).

47. Hemann MT, Fridman JS, Zilfou JT, Hernando E, Paddison PJ, Cordon-Cardo C, et al. An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes in vivo. Nat Genet. 2003;33(3):396-400.

48. Robles AI, Harris CC. Clinical outcomes and correlates of TP53 mutations and cancer. Cold Spring Harbor perspectives in biology. 2010;2(3).

49. Zambetti GP. The p53 mutation “gradient effect” and its clinical implications. J Cell Physiol. 2007;213(2):370-3.

50. Donehower LA. Using mice to examine p53 functions in cancer, aging, and longevity. Cold Spring Harbor perspectives in biology. 2009;1(6).

51. Rudolph KL, Chang S, Lee H, Blasco M, Gottlieb GJ, Greider C, et al. Longevity, stress response, and cancer in aging telomerase-deficient mice. Cell. 1999;96(5):701-12.

52. Vogel H, Lim D, Karsenty G, Finegold M, Hasty P. Deletion of Ku86 causes early onset of senescence in mice. Proceedings of the National Academy of Sciences. 1999;96(19):10770-5.

53. Tyner SD, Venkatachalam S, Choi J, Jones S, Ghebranious N, Igelmann H, et al. p53 mutant mice that display early ageing-associated phenotypes. Nature. 2002;415(6867):45-53.

54. Maier B, Gluba W, Bernier B, Turner T, Mohammad K, Guise T, et al. Modulation of mammalian life span by the short isoform of p53. Genes Dev. 2004;18(3):306-19.

55. García-Cao I, García-Cao M, Martín-Caballero J, Criado LM, Klatt P, Flores JM, et al. 'Super p53'mice exhibit enhanced DNA damage response, are tumor resistant and age normally. EMBO J. 2002;21(22):6225-35.

56. Bagchi A, Papazoglu C, Wu Y, Capurso D, Brodt M, Francis D, et al. < I> CHD5 is a tumor suppressor at human< i> 1p36. Cell. 2007;128(3):459-75.

57. Zigouris A, Alexiou GA, Batistatou A, Voulgaris S, Kyritsis AP. The role of matrix metalloproteinase 9 in intervertebral disc degeneration. Journal of Clinical Neuroscience. 2011;18(10):1424-5.

58. Mazin P, Xiong J, Liu X, Yan Z, Zhang X, Li M, et al. Widespread splicing changes in human brain development and aging. Molecular systems biology. 2013;9(1).

59. Wood SH, Craig T, Li Y, Merry B, de Magalhães JP. Whole transcriptome sequencing of the aging rat brain reveals dynamic RNA changes in the dark matter of the genome. Age. 2012:1-14.

60. Khoury MP, Bourdon J. p53 isoforms an intracellular microprocessor? Genes & cancer. 2011;2(4):453-65.

61. Vo N, Seo H, Robinson A, Sowa G, Bentley D, Taylor L, et al. Accelerated aging of intervertebral discs in a mouse model of progeria. Journal of Orthopaedic Research. 2010;28(12):1600-7.

62. Nasto LA, Wang D, Robinson AR, Clauson CL, Ngo K, Dong Q, et al. Genotoxic stress accelerates age-associated degenerative changes in intervertebral discs. Mech Ageing Dev. 2012.

63. Feng Z, Zhang C, Kang H, Sun Y, Wang H, Naqvi A, et al. Regulation of female reproduction by p53 and its family members. The FASEB Journal. 2011;25(7):2245-55.

64. Hu W. The role of p53 gene family in reproduction. Cold Spring Harbor perspectives in biology. 2009;1(6).

65. Prokocimer M, Barkan R, Gruenbaum Y. Hutchinson–Gilford progeria syndrome through the lens of transcription. Aging cell. 2013.

66. Scaffidi P, Misteli T. Reversal of the cellular phenotype in the premature aging disease hutchinson-gilford progeria syndrome. Nat Med. 2005;11(4):440-5.

67. Goldman RD, Shumaker DK, Erdos MR, Eriksson M, Goldman AE, Gordon LB, et al. Accumulation of mutant lamin A causes progressive changes in nuclear architecture in Hutchinson–Gilford progeria syndrome. Proc Natl Acad Sci U S A. 2004;101(24):8963-8.

68. Columbaro M, Capanni C, Mattioli E, Novelli G, Parnaik V, Squarzoni S, et al. Rescue of heterochromatin organization in hutchinson-gilford progeria by drug treatment. Cellular and molecular life sciences. 2005;62(22):2669-78.

69. Ishimi Y, Kojima M, Takeuchi F, Miyamoto T, Yamada M, Hanaoka F. Changes in chromatin structure during aging of human skin fibroblasts. Exp Cell Res. 1987;169(2):458-67.

70. O'Sullivan RJ, Kubicek S, Schreiber SL, Karlseder J. Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. Nature structural & molecular biology. 2010;17(10):1218-25.

71. O'Sullivan RJ, Karlseder J. The great unravelling: Chromatin as a modulator of the aging process. Trends Biochem Sci. 2012.

72. Allison SJ, Milner J. Remodelling chromatin on a global scale: A novel protective function of p53. Carcinogenesis. 2004;25(9):1551-7.

73. Rubbi CP, Milner J. p53 is a chromatin accessibility factor for nucleotide excision repair of DNA damage. EMBO J. 2003;22(4):975-86.

74. Oberdoerffer P, Michan S, McVay M, Mostoslavsky R, Vann J, Park S, et al. SIRT1 redistribution on chromatin promotes genomic stability but alters gene expression during aging. Cell. 2008;135(5):907-18.

75. Mammoto A, Mammoto T, Ingber DE. Mechanosensitive mechanisms in transcriptional regulation. J Cell Sci. 2012;125(13):3061-73.

76. Méjat A. LINC complexes in health and disease. Nucleus. 2010;1(1):40-52.

77. Wang N, Tytell JD, Ingber DE. Mechanotransduction at a distance: Mechanically coupling the extracellular matrix with the nucleus. Nature reviews Molecular cell biology. 2009;10(1):75-82.

78. Dreuillet C, Tillit J, Kress M, Ernoult- Lange M. In vivo and in vitro interaction between human transcription factor MOK2 and nuclear lamin A/C. Nucleic Acids Res. 2002;30(21):4634-42.

79. Haraguchi T, Holaska JM, Yamane M, Koujin T, Hashiguchi N, Mori C, et al. Emerin binding to btf, a death- promoting transcriptional repressor, is disrupted by a missense mutation that causes Emery–Dreifuss muscular dystrophy. European Journal of Biochemistry. 2004;271(5):1035-45.

80. Liu B, Zhou Z, Lam Y. Quantitative nucleolar proteomics reveals nuclear re-organization during stress-induced senescence in mouse fibroblast. BMC cell biology. 2011;12(1):33.

81. Faulkner GJ. Retrotransposons: Mobile and mutagenic from conception to death. FEBS Lett. 2011;585(11):1589-94.

82. de Koning AJ, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. PLoS genetics. 2011;7(12):e1002384.

83. Rebollo R, Horard B, Hubert B, Vieira C. Jumping genes and epigenetics: Towards new species. Gene. 2010;454(1):1-7.

84. Wilkins AS. The enemy within: An epigenetic role of retrotransposons in cancer initiation. Bioessays. 2010;32(10):856-65.

85. Rebollo R, Romanish MT, Mager DL. Transposable elements: An abundant and natural source of regulatory sequences for host genes. Annu Rev Genet. 2012;46:21-42.

86. Oliver KR, Greene WK. Transposable elements: Powerful facilitators of evolution. Bioessays. 2009;31(7):703-14.

87. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. Cytogenetic and genome research. 2005;110(1-4):462-7.

88. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, et al. CDD: Conserved domains and protein three-dimensional structure. Nucleic Acids Res. 2013;41(D1):D348-52.

89. Kurth R, Bannert N. Beneficial and detrimental effects of human endogenous retroviruses. International journal of cancer. 2010;126(2):306-14.

90. De Parseval N, Lazar V, Casella JF, Benit L, Heidmann T. Survey of human genes of retroviral origin: Identification and transcriptome of the genes with coding capacity for complete envelope proteins. J Virol. 2003;77(19):10414-22.

91. Aagaard L, Villesen P, Kjeldbjerg AL, Pedersen FS. The∼ 30-million-year-old ERVPb1 envelope gene is evolutionarily conserved among hominoids and old world monkeys. Genomics. 2005;86(6):685-91.

92. Chen H, Cheong ML. Syncytins: Molecular aspects. Cell Fusions. 2011:117-37.

93. Antony JM, Ellestad KK, Hammond R, Imaizumi K, Mallet F, Warren KG, et al. The human endogenous retrovirus envelope glycoprotein, syncytin-1, regulates neuroinflammation and its receptor expression in multiple sclerosis: A

role for endoplasmic reticulum chaperones in astrocytes. The Journal of Immunology. 2007;179(2):1210-24.

94. Lin E, Salon C, Brambilla E, Lavillette D, Szecsi J, Cosset F, et al. Fusogenic membrane glycoproteins induce syncytia formation and death in vitro and in vivo: A potential therapy agent for lung cancer. Cancer Gene Ther. 2009;17(4):256-65.

95. Gao P, Zheng J. Oncogenic virus-mediated cell fusion: New insights into initiation and progression of oncogenic viruses-related cancers. Cancer Lett. 2011;303(1):1-8.

96. Duelli D, Lazebnik Y. Cell-to-cell fusion as a link between viruses and cancer. Nature Reviews Cancer. 2007;7(12):968-76.

97. Lu X, Kang Y. Cell fusion as a hidden force in tumor progression. Cancer Res. 2009;69(22):8536.

98. Lu X, Kang Y. Cell fusion hypothesis of the cancer stem cell. Cell Fusion in Health and Disease. 2011:129-40.

99. Kurth R, Bannert N. Beneficial and detrimental effects of human endogenous retroviruses. International journal of cancer. 2010;126(2):306-14.

100. Belyi VA, Ak P, Markert E, Wang H, Hu W, Puzio-Kuter A, et al. The origins and evolution of the p53 family of genes. Cold Spring Harbor perspectives in biology. 2010;2(6).

101. Rivlin N, Brosh R, Oren M, Rotter V. Mutations in the p53 tumor suppressor gene important milestones at the various steps of tumorigenesis. Genes & cancer. 2011;2(4):466-74.

102. Courtois-Cox S, Jones S, Cichowski K. Many roads lead to oncogene-induced senescence. Oncogene. 2008;27(20):2801-9.

103. Braig M, Schmitt CA. Oncogene-induced senescence: Putting the brakes on tumor development. Cancer Res. 2006;66(6):2881-4.

104. Bagchi A, Mills AA. The quest for the 1p36 tumor suppressor. Cancer Res. 2008;68(8):2551-6.

105. de Koning AJ, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. PLoS genetics. 2011;7(12):e1002384.

106. Moyer VA. Screening for prostate cancer: US preventive services task force recommendation statement. Ann Intern Med. 2012;157(2):120-34.

107. Dickinson SI. Premalignant and malignant prostate lesions: Pathologic review. Cancer Control. 2010;17:214-22.

108. Ayala AG, Ro JY. Prostatic intraepithelial neoplasia: Recent advances. Arch Pathol Lab Med. 2007;131(8):1257-66.

109. McNeal JE. Origin and development of carcinoma in the prostate. Cancer. 1969;23(1):24-34.

110. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, et al. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proceedings of the National Academy of Sciences. 2007;104(47):18613-8.

111. Wei C, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, et al. A global map of p53 transcription-factor binding sites in the human genome. Cell. 2006;124(1):207-19.

112. Collado M, Serrano M. The power and the promise of oncogene-induced senescence markers. Nature Reviews Cancer. 2006;6(6):472-6.

113. Sharpless NE, DePinho RA. Cancer: Crime and punishment. Nature. 2005;436(7051):636-7.

114. Goering W, Ribarska T, Schulz WA. Selective changes of retroelement expression in human prostate cancer. Carcinogenesis. 2011;32(10):1484-92.

115. Ishida T, Obata Y, Ohara N, Matsushita H, Sato S, Uenaka A, et al. Identification of the HERV-K gag antigen in prostate cancer by SEREX using autologous patient serum and its immunogenicity. Cancer Immunity: a Journal of the Academy of Cancer Immunology. 2008;8.

116. Grizzle WE, Srivastava S, Manne U. The biology of incipient, pre-invasive or intraepithelial neoplasia. Cancer Biomarkers. 2011;9(1):21-39.

117. Chen Z, Carracedo A, Lin H, Koutcher JA, Behrendt N, Egia A, et al. Differential p53-independent outcomes of p19Arf loss in oncogenesis. Science signaling. 2009;2(84):ra44.

118. Chan C, Gao Y, Moten A, Lin H. Novel ARF/p53-independent senescence pathways in cancer repression. Journal of molecular medicine. 2011;89(9):857-67.

119. Freeman DJ, Li AG, Wei G, Li H, Kertesz N, Lesche R, et al. PTEN tumor suppressor regulates p53 protein levels and activity through phosphatase-dependent and-independent mechanisms. Cancer cell. 2003;3(2):117-30.

120. Zhang X, Liu F, Wang W. Two-phase dynamics of p53 in the DNA damage response. Proceedings of the National Academy of Sciences. 2011;108(22):8990-5.

121. Lei Q, Jiao J, Xin L, Chang C, Wang S, Gao J, et al. NKX3. 1 stabilizes p53, inhibits AKT activation, and blocks prostate cancer initiation caused by PTEN loss. Cancer cell. 2006;9(5):367-78.

122. Schrecengost R, Knudsen KE. Molecular pathogenesis and progression of prostate cancer. Seminars in oncology; Elsevier; 2013.

123. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12(6):996-1006.

124. Barbas CF, Burton DR, Scott JK, Silverman GJ. Phage displays: A laboratory manual. . 2001.

125. Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol. 2012;30(5):460-5.

126. de Parseval N, Lazar V, Casella J, Benit L, Heidmann T. Survey of human genes of retroviral origin: Identification and transcriptome of the genes with coding capacity for complete envelope proteins. J Virol. 2003;77(19):10414-22.

127. Karlsson H, Schröder J, Bachmann S, Bottmer C, Yolken R. HERV-W-related RNA detected in plasma from individuals with recent-onset schizophrenia or schizoaffective disorder. Mol Psychiatry. 2003;9(1):12-3.

128. Cohen M, Powers M, O'Connell C, Kato N. The nucleotide sequence of the< i> env gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. Virology. 1985;147(2):449-58.

129. Sharma S, Stolina M, Lin Y, Gardner B, Miller PW, Kronenberg M, et al. T cell-derived IL-10 promotes lung cancer growth by suppressing both T cell and APC function. The Journal of Immunology. 1999;163(9):5020-8.

130. Bertram JS, Janik P. Establishment of a cloned line of lewis lung carcinoma cells adapted to cell culture. Cancer Lett. 1980;11(1):63-73.

131. Cawthorne C, Swindell R, Stratford IJ, Dive C, Welman A. Comparison of doxycycline delivery methods for tet-inducible gene expression in a subcutaneous xenograft model. Journal of biomolecular techniques: JBT. 2007;18(2):120.

132. Jenkins DE, Oei Y, Hornig YS, Yu S, Dusich J, Purchio T, et al. Bioluminescent imaging (BLI) to improve and refine traditional murine models of tumor growth and metastasis. Clin Exp Metastasis. 2003;20(8):733-44.

133. Ben-Baruch A. Inflammation-associated immune suppression in cancer: The roles played by cytokines, chemokines and additional mediators. Seminars in cancer biology; Elsevier; 2006.

134. Ha T. The role of regulatory T cells in cancer. Immune network. 2009;9(6):209-35.

135. Haraguchi S, Good RA, Cianciolo GJ, Engelman R, Day N. Immunosuppressive retroviral peptides: Immunopathological implications for immunosuppressive influences of retroviral infections. J Leukoc Biol. 1997;61(6):654-66.

136. Bunnell TM, Burbach BJ, Shimizu Y, Ervasti JM. B-actin specifically controls cell growth, migration, and the G-actin pool. Mol Biol Cell. 2011;22(21):4047-58.