

THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOLOGY

FREQUENCY OF ALUYA5 INSERTION IN A REGULATORY REGION OF
MICROSOMAL EPOXIDE HYDROLASE IN HUMAN DISEASED TISSUE

EMILY FAY
Spring 2011

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biology
with honors in Biology

Reviewed and approved* by the following:

Curtis Omiecinski
Professor of Veterinary & Biomedical Sciences
Thesis Supervisor

Stephen Schaeffer
Associate Professor of Biology
Honors Adviser

*Signatures are on file in the Schreyer Honors College

ABSTRACT

Human microsomal epoxide hydrolase (mEH) is an enzyme involved in both detoxification and bioactivation of endogenous and xenobiotic compounds. mEH catalyzes the *trans*-addition of water to reactive epoxide intermediates to form less toxic diols; however, mEH is also responsible for bioactivating metabolic intermediates, resulting in potentially toxic metabolites. mEH is required for the activation of polyaromatic hydrocarbons (PAHs) found in cigarette smoke to genotoxic carcinogens.

Several common genetic polymorphisms exist in the mEH gene (EPHX1) that may cause differences in epoxide hydrolase expression and activity between individuals. Epidemiological studies have shown that certain genetic variants are correlated with altered disease susceptibilities, suggesting they may serve as markers for predicting disease. This study investigates the population frequencies of a double AluYa5 insertion element that occurs in the far upstream promoter region of EPHX1 that impacts the gene's transcriptional status.

Cell or tissue-derived DNA samples from several hundred unrelated individuals were analyzed in these investigations. A focus of the study was to evaluate genotype status of the EPHX1 upstream promoter region using DNA extracts from normal uninvolved tissues that were derived from a human tissue bank, consisting of lung and liver tissues obtained from otherwise diseased individuals. The DNA samples were analyzed for the presence or absence of the double Alu insertion, as well as Alu (-/-), Alu (+/-), and Alu (++) gene frequencies. Using Pearson's chi-square test, we compared the observed and expected genotype frequencies in a previously studied healthy population to

determine whether the insertion was in Hardy-Weinberg equilibrium. We used the allele frequencies of the healthy population to calculate the expected genotype frequencies of the diseased population, and Pearson's chi-square test was performed to identify any significant difference between observed and expected genotype frequencies. Additionally, the allele frequencies of the healthy population were used to calculate the expected genotype frequencies in the lung and liver samples, male and female samples, and samples from donors age 0-49 and age 50 and older, to determine whether any correlations existed between genotype and tissue type, gender, and age of development. Pearson's chi-square test was also used to compare observed and expected genotype frequencies in each subpopulation analyzed. These analyses allowed us to determine whether the double Alu insertion was in Hardy-Weinberg equilibrium in the subpopulations.

Through these investigations, we determined that the double Alu insertion was indeed in Hardy-Weinberg equilibrium in both the healthy and diseased populations, indicating that the insertion was not under selection pressures and therefore did not correlate with protection from disease nor initiation of disease. The double Alu insertion was also in Hardy-Weinberg equilibrium in both lung and liver samples, female and male donors, and donors younger than and older than age 50, and no significant differences in genotype frequency were identified in any subpopulation analyzed. Our findings suggest that the double Alu insertion does not play a major role in altering disease susceptibility in the samples analyzed and does not appear to be a useful biomarker for disease prediction within the spectrum of diseases assessed.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES	iv
ACKNOWLEDGEMENTS	vi
INTRODUCTION	1
Microsomal Epoxide Hydrolase	1
mEH's Role in Detoxification and Bioactivation	1
EPHX1 Variation	3
Experimental Aims and Hypothesis	7
MATERIALS AND METHODS	10
Tissue Sample Preparation	10
Tissue Homogenization	10
RNA and DNA Isolation	11
E1-b PCR Amplification	12
Statistical Analysis	14
RESULTS	15
PCR Optimization	15
Alu Genotype Analysis	17
Comparison to a Healthy Population	18
Comparison of Alu Genotypes by Tissue Type	20
Comparison of Alu Genotypes by Gender	22
Comparison of Alu Genotypes by Age	23
DISCUSSION	25
REFERENCES	30
APPENDIX A	32
Lung Sample Information	32
Liver Sample Information	34
APPENDIX B	37
APPENDIX C	44

LIST OF TABLES

Table 1: Thermal cycler protocol.....	13
Table 2: Alu genotypes in lung and liver samples and overall.....	18
Table 3: Genotype and allele frequencies of the healthy population and a population in Hardy-Weinberg equilibrium.....	19
Table 4: Genotype and allele frequencies of the diseased population and a population in Hardy-Weinberg equilibrium.....	20
Table 5: Observed and expected genotype frequencies in lung samples.....	21
Table 6: Observed and expected genotype frequencies in liver samples.....	21
Table 7: Observed and expected genotype frequencies in female donor samples.....	22
Table 8: Observed and expected genotype frequencies in male donor samples.....	22
Table 9: Observed and expected genotype frequencies in donor samples age 0-49.....	24
Table 10: Observed and expected genotype frequencies in donor samples age 50+.....	24

LIST OF FIGURES

Figure 1: Formation of benzo[a]pyrene diolepoxide.....	3
Figure 2: Alternative exon 1 sequences of human EPHX1.....	4
Figure 3: Tissue-specific expression of E1 and E1-b transcripts.....	5
Figure 4: Structure of the E1-b promoter region.....	7
Figure 5: PCR products of Coriell Panel DNA with temperature gradient.....	16
Figure 6: PCR products of tissue bank samples.....	17
Figure 7: PCR products of lung samples 8749, 8907, 8538, 8466, 8480, 8492, 6270, 6282, 6357, 6399, 6483, 6495, and 6503.....	37
Figure 8: PCR products of lung samples 6765, 6784, 6893, 6964, 8079, 8141, 2509, 2909, 2919, 2949, 2955, 4588, and 6014.....	37
Figure 9: PCR products of lung samples 6040, 6091, 6178, 362, 368, 2091, 2193, 198, 2346, 8758, 8810, 8963, and 14878.....	38
Figure 10: PCR products of liver samples 6412, 6518, 6624, 6668, 8046, 8222, 8762, 5096, 5245, 6013, 6066, 4172, and 2029.....	38
Figure 11: PCR products of liver samples 2225, 2304, 2332, 2433, 2580, 2701, 2999, 3124, 4140, 4423, and 5014.....	39
Figure 12: PCR products of lung samples 8963, 14878, 8556, 8568, 8608, 8631, 8712, 8732, 8187, 8259, 8315, 8331, and 8402.....	39
Figure 13: PCR products of lung samples 8433, 8497, 6379, 6498, 6520, 6653, 6799, 6927, 8010, 2537, 2621, 2924, and 4229.....	40
Figure 14: PCR products of lung and liver samples 4448, 6095, 6204, 2782, 2485, 7695, 8216, 8349, 8716, 8818, 14876, 5067, and 5111.....	40
Figure 15: PCR products of lung samples 5147, 5168, 5211, 5221, 5222, 5233, 5247, 5258, 5334, 6011, 6021, 6056, and 6080.....	41
Figure 16: PCR products of liver samples 6086, 6156, 6247, 6277, 2, 2088, 2127, 2298, 2469, 2511, 2619, 2778, and 2779.....	41
Figure 17: PCR products of liver samples 2786, 2920, 3235, 3243, 3248, 3292, 3664, 3874, 4030, 4568, and 5058.....	42

Figure 18: PCR products of lung and liver samples 8433 (in duplicate), 2909, 4448, 8216, 8349, 8716, 8818, 14876, 5067, 5111, 5147, and 511142
Figure 19: PCR products of liver samples 5211, 5233, 5247, 5258, 5334, 6011, 6056, 6247, 2127, 2464, 2779, and 292043
Figure 20: PCR products of liver samples 2786 and 329243

ACKNOWLEDGEMENTS

I would like to thank my parents for all of their support not only during college, but in every aspect of my life. You have always believed in me, and I truly can never thank you enough for everything that you have given me.

I would also like to thank Dr. Stephen Schaeffer, my honors adviser, for his guidance over the past four years. Since I entered Penn State as a clueless freshman, you steered me in the right direction, kept me grounded, and really made an effort to get to know me. Thank you for all the time you spend in helping students achieve their goals.

Finally, I would like to thank Dr. Curtis Omiecinski and the entire Omiecinski lab for all of their help and support. Dr. Omiecinski, over the past three years, you have constantly provided me with the guidance I needed to pose a research question and to design experiments to answer it. I learned so much about lab techniques and research during my time in your lab, and I am very thankful for the experience. Elizabeth, thank you for all of your assistance with my project. I appreciate all of the time you spent formulating and optimizing the DNA/RNA extraction procedure, as well as your enduring patience when walking me through the procedure and answering my questions. Denise, since I stepped foot into the lab as a sophomore, you have made me feel welcome and involved in the lab. Thank you so much for your mentorship and for dealing with my endless supply of questions about PCR optimization and lab work in general.

Everyone in the lab has been kind and supportive, and it has been a pleasure getting to know each one of you. Thanks for all of your help!

INTRODUCTION

Microsomal Epoxide Hydrolase

Microsomal epoxide hydrolase (mEH) is a hydrolytic enzyme involved in the metabolism of xenobiotic chemicals. Located in the smooth endoplasmic reticulum of cells, mEH functions primarily to produce more stable intermediates from reactive epoxide intermediates, thereby reducing the toxicity of metabolites. While mEH acts on both endogenous and xenobiotic epoxide derivatives, its primary recognized role is to metabolize xenobiotic compounds⁷.

Human mEH is encoded by the gene EPHX1, located on the long arm of chromosome 1. Containing nine exons and eight introns, the gene is approximately ~20kb long and encodes 455 amino acids⁶. All human tissues have been found to express mEH, suggesting that the enzyme plays an essential role in detoxification and protection⁷. Additionally, EPHX1 maintains a 75% sequence similarity between human, rat, and rabbit, indicating a conserved function¹³.

mEH's Role in Detoxification and Bioactivation

Common xenobiotic substrates for mEH include anticonvulsants, such as phenytoin and carbamazepine, as well as environmentally important chemicals that include 1,3-butadiene, benzene, and polycyclic aromatic hydrocarbons⁷. mEH acts by catalyzing the *trans*-addition of water to the epoxide substrates, forming dihydrodiol products. While epoxide hydrolysis does occur spontaneously *in vitro*, mEH increases reaction rate and selectivity by creating an optimum pH for the protonation of the epoxide oxygen and

the nucleophilic addition of a hydroxyl group¹⁶. Depending on the substrate, mEH may play a role in detoxifying or in bioactivating a xenobiotic.

mEH often acts in concert with the cytochrome P450s (CYPs; CYP450), a family of enzymes involved in Phase I conversions that catalyze the oxidation of compounds. CYPs form epoxides from carbon-carbon multiple bonds in aliphatic compounds as well as carbon-carbon bonds in aromatic compounds, forming potentially reactive epoxide intermediates, which in turn may covalently bind with proteins and DNA, resulting in toxicity. mEH also participates as a Phase I enzyme in detoxification reactions, directly hydrolyzing xenobiotic epoxides to create more soluble, and often less reactive metabolites. Thus, in this capacity, mEH plays a protective role against reactive epoxide intermediates.

While mEH plays a major role in detoxification processes, it is also responsible for bioactivation of some substrates. In bioactivation reactions, mEH converts epoxide intermediates into metabolites that will ultimately become toxic. Polyaromatic hydrocarbons (PAHs), aromatic compounds found in cigarette smoke, induce carcinogenesis upon activation by mEH. In the example of the PAH benzo[a]pyrene, mEH initially forms a stable *trans*-diol through hydrolysis of an epoxide residue introduced initially through the activity of the CYPs. However, this *trans*-diol is oxidized again by the CYPs, forming an epoxide substrate that mEH cannot further hydrolyze (Figure 1). The resulting benzo[a]pyrene diolepoxide is genotoxic and carcinogenic, able to covalently bind to and intercalate DNA, resulting in mutations¹⁶.

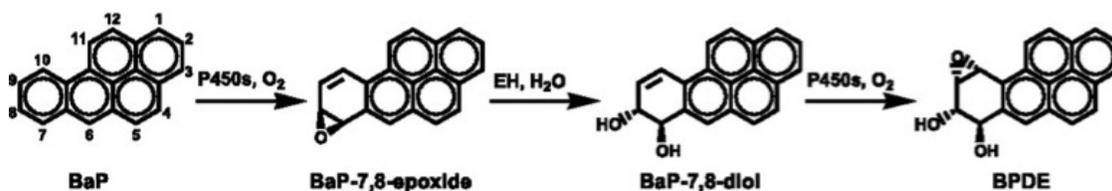


Figure 1: Formation of benzo[a]pyrene diol epoxide. In this bioactivation reaction, mEH hydrolyses an epoxide formed by CYP450 in an oxidation reaction. CYP450 then acts again, producing a reactive diol epoxide that cannot be further hydrolyzed by mEH. Hecht, et al. *Cancer Epidemiol Biomarkers and Prev.* 15 (2006): 1085-1811.

EPHX1 Variation

PAH-induced genotoxicity is believed to be one of the main causes of lung cancer in cigarette smokers; thus, mEH is an important factor in lung cancer development.

Interestingly, studies show that mEH-knockout mice do not develop cancer when exposed to high levels of PAHs, indicating that mEH is essential in converting PAHs into carcinogens⁹. While many genetic and environmental factors are involved in cancer development, variations in EPHX1 expression are implicated in differential susceptibilities to cancer and other diseases. Variations in mEH protein structure, EPHX1 transcriptional regulation, and tissue-specific expression of mEH may all factor into differential disease susceptibility¹².

Interindividual variation in EPHX1 expression arises in several ways. Firstly, two common polymorphisms involving amino acid substitutions exist in humans. At the amino acid 113 position, tyrosine (Y) is frequently substituted with histidine (H), while at the 139 position, histidine is often replaced by arginine (R). These single nucleotide polymorphisms (SNPs) are non-synonymous changes, resulting in altered protein structure⁶. Previous epidemiological studies have associated these polymorphisms with a higher risk of lung cancer development and an increased sensitivity to 1,3-butadiene, an

industrial chemical that causes neurotoxicity¹⁶. Currently, research has not elucidated the exact reason behind these predispositions, as analyses have shown no significant difference in substrate hydrolysis between the enzyme variants; however, allelic variants may still act preferentially on certain substrates, thus producing different types of carcinogens¹⁰.

Alternative splicing also produces variation in human EPHX1 expression. EPHX1 has two alternative exon 1 structures, termed E1 and E1-b, which direct mRNA transcription. E1 lies directly upstream of the EPHX1 coding region, 3.2kb 5' of exon 2. E1-b lies much further upstream, 18.5kb 5' of exon 2 (Figure 2)⁶.

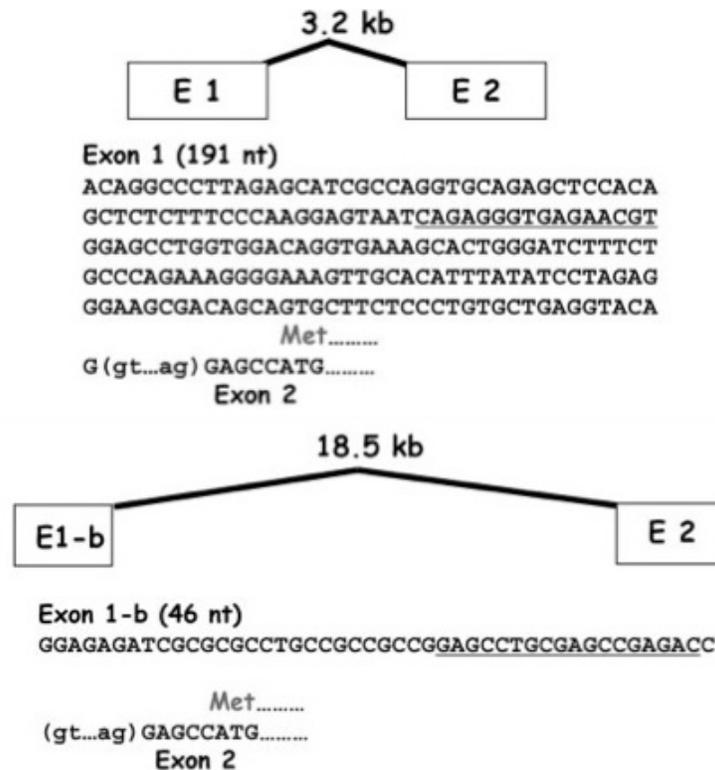


Figure 2: Alternative exon 1 sequences of human EPHX1. Two alternative exon 1 sequences direct mRNA transcription in humans. E1 lies 3.2kb upstream of exon 2, while E1-b lies 18.5kb upstream of exon 2. Liang, et al. “Alternative Promoters Determine Tissue-Specific Expression Profiles of the Human Microsomal Epoxide Hydrolase Gene (EPHX1). *Molecular Pharmacology*. 67 (2005): 220-230.

Real-time PCR assays show that the E1 EPHX1 transcript accounts for only a small portion of EPHX1 expressed in human tissues, found mainly in the liver and in very low levels in the ovary (Figure 3). The E1-b transcript accounts for the majority of EPHX1 expression in humans, detected as the major EPHX1 transcript in every tissue tested¹⁸. The dual-promoter system may provide another line of defense against reactive epoxide metabolites. Because the liver is the major site of xenobiotic detoxification in humans, high levels of liver-selective mEH may protect the organ from increased exposure to reactive metabolites¹³.

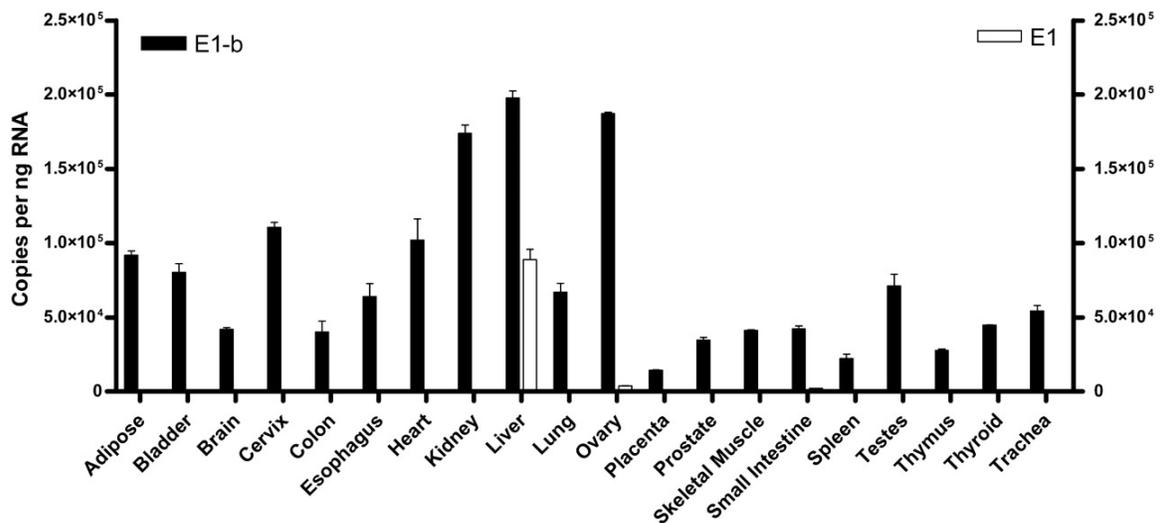


Figure 3: Tissue-specific expression of E1 and E1-b transcripts. E1-b is the predominant transcript in all tissue types analyzed. E1 is found mostly in the liver and in low levels in the ovary.

Yang, et al. “The Expression of Human Microsomal Epoxide Hydrolase Is Predominantly Driven by a Genetically Polymorphic Far Upstream Promoter.” *The Journal of Pharmacology and Experimental Therapeutics*. 330 (2009): 23-30.

In contrast to the highly conserved EPHX1 protein-coding sequence, the 5'-flanking regions vary widely between different species. Because these alternative promoter regions are responsible for directing transcription of EPHX1, sequence differences between species result in altered responses to inducers. For example, mouse

and rat EPHX1 expression can be induced by exposure to xenobiotic chemicals, while human hepatocytes exhibit very little response to inducers¹³.

Human E1 and E1-b promoter region sequences also vary between individuals, as they may contain transposable elements (TEs, Figure 4). TEs are non-coding DNA sequences that integrate into the genome directly as DNA, or after transcription from RNA to DNA². Composing nearly half of the entire human genome, TEs influence gene transcription and can act as alternative promoters, ultimately impacting protein expression¹⁸.

Alu repeats are a group of TEs classified as short interspersed elements (SINES), accounting for approximately 10% of the human genome¹⁸. These TEs are about 300bp long and are specific to primates. Generally, Alu elements that have recently integrated into the genome are homogenous, while older elements have diverged in sequence due to an accumulation of mutations². Within the human genome, Alu elements are most frequently located in transcriptionally active regions, and they may be involved in gene regulation¹⁸. Alu inserts have been shown to disrupt genes, causing inherited diseases; more frequently, however, the inserts only indicate an increased risk of developing a certain disease⁷.

Several variable Alu elements have been identified within a ~3kb region of the E1-b promoter. Two of these inserts are part of the Ya5 Alu family, one of the youngest and most common of the Alu families found in humans³. This double Alu insertion occurs at nucleotide positions -2214 and -1392 (Figure 4). Interestingly, genomic PCR screens of human tissue samples have shown that these AluYa5 insertions always occur together; genotypes never contain a lone Ya5 insertion at either position¹⁸. *In vitro* assays

show that the double Alu insertion downregulates the transcription of EPHX1, thus reducing the quantity of E1-b transcript and mEH protein produced. As mEH functions as a detoxification and bioactivation enzyme, these inserts may play a role in decreasing or increasing susceptibility to certain diseases.

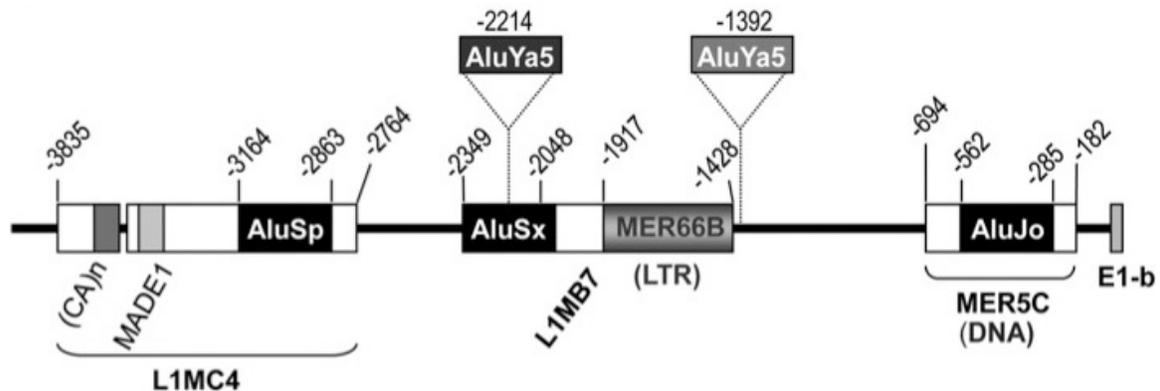


Figure 4: Structure of the E1-b promoter region. The E1-b promoter region contains several Alu inserts within a ~3kb region. The double AluYa5 insertion occurs at nucleotide positions -2214 and -1392. This insertion increases the E1-b transcript size from 1.1kb to 1.7kb, as each insert is approximately 300bp long. Yang, et al. “The Expression of Human Microsomal Epoxide Hydrolase Is Predominantly Driven by a Genetically Polymorphic Far Upstream Promoter.” *The Journal of Pharmacology and Experimental Therapeutics*. 330 (2009): 23-30.

Experimental Aims and Hypothesis

Cigarette smoking is estimated to cause ~90% of all lung cancer cases¹, 3,000 lung cancer deaths per day, and 30% of all cases of cancer mortality⁸. However, less than 20% of smokers will actually develop lung cancer from the habit². PAHs are well-established as potent carcinogens found in cigarette smoke, and PAH bioactivation by mEH plays a fundamental role in lung cancer development. Genetic variation in EPHX1 resulting in functional differences in mEH may play a role in interindividual susceptibility to PAH-induced lung carcinogenesis. Further, understanding how genotypic variations in EPHX1 account for differences in the metabolism of PAHs by

mEH will lead to better treatment and prevention of cancer in individuals. This project focused on determining the presence of the double Alu insertion in the 5' regulatory region of EPHX1 in human tissue samples obtained from diseased individuals. We hypothesized that the frequency of the Alu insertion will be different in the diseased individuals compared with healthy individuals.

The overall goal of this study was to quantify the variable presence of the E1-b double Alu insertion in human lung and liver tissue samples from a preexisting tissue bank. Liver tissue was chosen for analysis because it is a major site of epoxide metabolism and detoxification processes, and it has the highest expression of the E1 transcript. In addition to being a major site of epoxide metabolism, lung tissue was chosen because of the correlation between PAH exposure and lung cancer. The vast majority of the tissue samples are from individuals with cancers and other xenobiotic-induced diseases, so we were able to analyze Alu genotype frequencies in a diseased population. Information about the disease states of each donor is listed in Appendix A.

The specific aim of this study was to investigate the inheritance patterns of the double Alu insertion using the Hardy-Weinberg principle. The Hardy-Weinberg principle states that genotype and allele frequencies will remain static between generations, assuming that mutations do not occur, there is no immigration or emigration, there is an infinitely large population size, random mating occurs, and there are no selective pressures for or against a certain genotype⁵. Because these factors do occur in nature, Hardy-Weinberg equilibrium cannot exist in an actual population; however, we can still use the Hardy-Weinberg equilibrium as well as our null hypothesis to determine whether

there are significant differences between expected and actual genotype frequencies in our observed population¹⁷.

Because the double Alu insertion is known to downregulate the transcription of EPHX1, it may have a protective effect against PAH-induced carcinogenesis, as less mEH is available to produce carcinogenic diolepoxides. Conversely, if a decrease in mEH expression results in higher levels of reactive epoxides, the insertion may promote carcinogenesis and disease. Therefore, we used the Hardy-Weinberg equation and chi-square statistics to compare genotype frequencies of the diseased tissue bank samples to frequencies found in a healthy population. We hypothesized that the E1-b double Alu insertion will occur more or less frequently than in a healthy population, and that the insertion will not be in Hardy-Weinberg equilibrium in the diseased population.

MATERIALS AND METHODS

Tissue Sample Preparation

Human lung, liver, and ovary tissue samples, each from a different donor, were obtained from the Penn State Hershey Cancer Institute Tissue Bank. Samples were frozen in liquid nitrogen immediately following surgeries, placed in -80°C storage, then shipped to our laboratory and stored further at -80°C until use. Future directions for this study involve DNA, RNA, and protein-based assays, so methods were taken to ensure that all samples would remain stable. Because the tissues were received frozen, it was necessary to transition the tissues with *RNAlater*®-ICE Frozen Tissue Transition Solution prior to extraction. This transitioning allows the RNA to remain stable during tissue thawing.

To ensure that the *RNAlater*®-ICE solution penetrated the tissue, frozen samples were cut into pieces weighing between 50 and 100mg on dry ice using a scalpel and forceps pre-chilled and pre-treated with RNAase solution (100mM NaOH, 0.1% SDS, 1mM EDTA). Samples were then stored at -20°C until needed for TRIzol® extraction. Remaining untreated tissue was stored at -80°C for future protein-based applications.

Tissue Homogenization

A TRIzol® extraction method was used to allow for the extraction of both DNA and RNA from a single tissue sample. *RNAlater*®-ICE treated tissue samples were blotted with a Kimwipe and transferred to prechilled RNase free 1.5ml tubes. 1ml TRIzol® was added to each tissue sample.

Samples were homogenized using a Bullet Blender (Next Advance Inc., Averill Park, NY). The Bullet Blender uses small beads to homogenize tissue and can homogenize up to 24 samples at a time. For lung samples, 1.4 mm stainless steel beads were added to each tube containing tissue in TRIzol®, while 1mm zirconium oxide beads were added for liver and ovary samples, according to the manufacturers' instructions. For each sample, a volume of beads approximately equal to the mass of the tissue sample was used. The Bullet Blender was placed in a 4°C cold room to ensure that the tissue samples remained cold despite the friction created from the homogenization process. Samples were run at a speed of 8 for 3-5 minutes, depending on the size of the tissue.

RNA and DNA Isolation

Homogenates were transferred to fresh RNase free 1.5ml tubes and were incubated at room temperature for 5 minutes. Chloroform (0.2ml) was added to each sample, and samples were shaken vigorously by hand for 15 seconds. Samples were then centrifuged at 12,000XG for 15 minutes at 4°C to separate the aqueous and organic layers.

Following centrifugation, the RNA-containing aqueous layer was transferred to a fresh RNase free 1.5ml tube and the organic layer was retained for DNA isolation. Isopropanol (0.5ml) was added, and the samples were incubated at room temperature for 10 minutes. Samples were then centrifuged at 12,000XG for 10 minutes at 4°C to pellet the RNA, and the supernatant was discarded. 75% ethanol (1ml) was added to each sample, and samples were stored at -20°C for later analysis. For extraction of DNA, 100% ethanol (0.4ml) was added to the DNA-containing organic layer, and samples were incubated for 20 minutes at room temperature. Samples were centrifuged at 2,000XG for

5 minutes at 4°C to pellet the DNA, and the supernatant was discarded. 0.1M sodium citrate in 10% ethanol (1ml) was added to each DNA pellet, and pellets were incubated on a rocker for 30 minutes at room temperature. Centrifugation and sodium citrate precipitation steps were repeated, and samples were centrifuged at 2,000XG for 5 minutes at 4°C. 1.5ml 75% ethanol was added to each sample, and samples were stored at 4°C.

To prepare the stored DNA samples for PCR, they were centrifuged at 2,000XG for 5 minutes at 4°C. Supernatant was removed, and the pellets were left to air dry for 10 minutes. Pellets were resuspended in 8mM NaOH (100µl) and centrifuged at 12,000XG for 10 minutes at room temperature. Supernatant was transferred to a fresh tube, and 1M HEPES (2.3µl) was added to each sample to adjust the pH. DNA concentrations were determined using a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE).

E1-b PCR Amplification

To analyze the isolated genomic DNA for the presence of Alu inserts, DNA samples underwent PCR to amplify the E1-b proximal promoter region. Our previous studies have shown that the two Alu inserts that occur at separate positions are always linked together; in no instance have we detected only single Alu polymorphic insertions within a given haplotype¹⁸. Thus, only the region from -1305 to -1485, which includes one of the AluYa5 insertions, underwent amplification. This allowed for the most efficient amplification of the Alu insertion at the -1392 nucleotide position. Forward primers (5'-TTGGTCCTCAATCTGGTGTCCAAG-3') and reverse primers (5'-AACTGCAGTCTGGGAGAGTTCTTT-3') were diluted to 10µM using nuclease-free

water. For concentrated DNA samples ($\geq 20\text{ng}/\mu\text{l}$), $5\mu\text{l}$ template was used in the PCR reaction, while for dilute samples ($< 20\text{ng}/\mu\text{l}$), $15\mu\text{l}$ template was used. The total reaction volume was $50\mu\text{l}$ and contained $5\mu\text{l}$ 10x reaction buffer, $2.5\mu\text{l}$ of each $10\mu\text{M}$ primer, $4\mu\text{l}$ 2.5mM DNTPs, and $0.5\mu\text{l}$ Taq. $10\mu\text{l}$ betaine was added to the PCR mix to improve product yield, and reactions were brought up to $50\mu\text{l}$ with $20.5\mu\text{l}$ or $10.5\mu\text{l}$ nuclease-free water for concentrated and dilute DNA samples, respectively.

293A cell lysate was included each time a PCR was run as a control sample homozygous for the double Alu insertion (Alu (+/+)), while HH1484 cell lysate was used as a control sample that lacked the insertion (Alu (-/-)). Additionally, nuclease-free water was used as a negative control. PCR was performed using a BioRad C1000 Thermocycler. The thermocycler protocol conditions are shown in Table 1. The PCR products ($20\mu\text{l}$ of each) were run on a 1.5% agarose gel, alongside a 100bp ladder. Products were visualized with ethidium bromide and UV transillumination.

Table 1: Thermal cycler protocol.

Step	Time (minutes)	Temperature ($^{\circ}\text{C}$)	Cycles
Initial Denaturation	10:00	95	1
Denaturation	0:30	95	37
Annealing	0:30	58	
Elongation	2:00	72	
Final Elongation	5:00	72	1
	∞	4	

PCR reactions ($50\mu\text{l}$) were performed on a thermal cycler using an optimized protocol. The entire cycle took 2:06, and PCR products were kept at 4°C until analysis by gel electrophoresis.

Statistical Analysis

68 lung samples and 69 liver samples were genotyped, for a total of 137 diseased samples. Diseased genotype frequencies were compared with those found in a previously studied healthy population of 449 individuals. Pearson's chi-square test was used to compare genotype frequencies between the healthy and diseased groups, as well as between lung and liver samples, samples from male and female donors, and samples from donors younger than age 50 and donors age 50 and older. For each chi-square test, the null hypothesis was that the population was in Hardy-Weinberg equilibrium, and the alternative hypothesis was that the population was not in Hardy-Weinberg equilibrium. The null hypothesis was rejected if the p-value was less or equal to than 0.05. Example statistical calculations are shown in Appendix C.

RESULTS

PCR Optimization

To ensure that PCR conditions were optimal for genomic DNA amplification, several different PCR mixtures and annealing temperatures were explored. Optimization assays were done using genomic DNA samples from the Coriell panel, a tissue panel analyzed previously in our lab for Alu genotype¹⁸. This allowed us to ensure that Alu (-/-), Alu (+/-), and Alu (+/+) would be properly amplified. A temperature gradient showed that an annealing temperature of 58°C provided the best product yield (Figure 5).

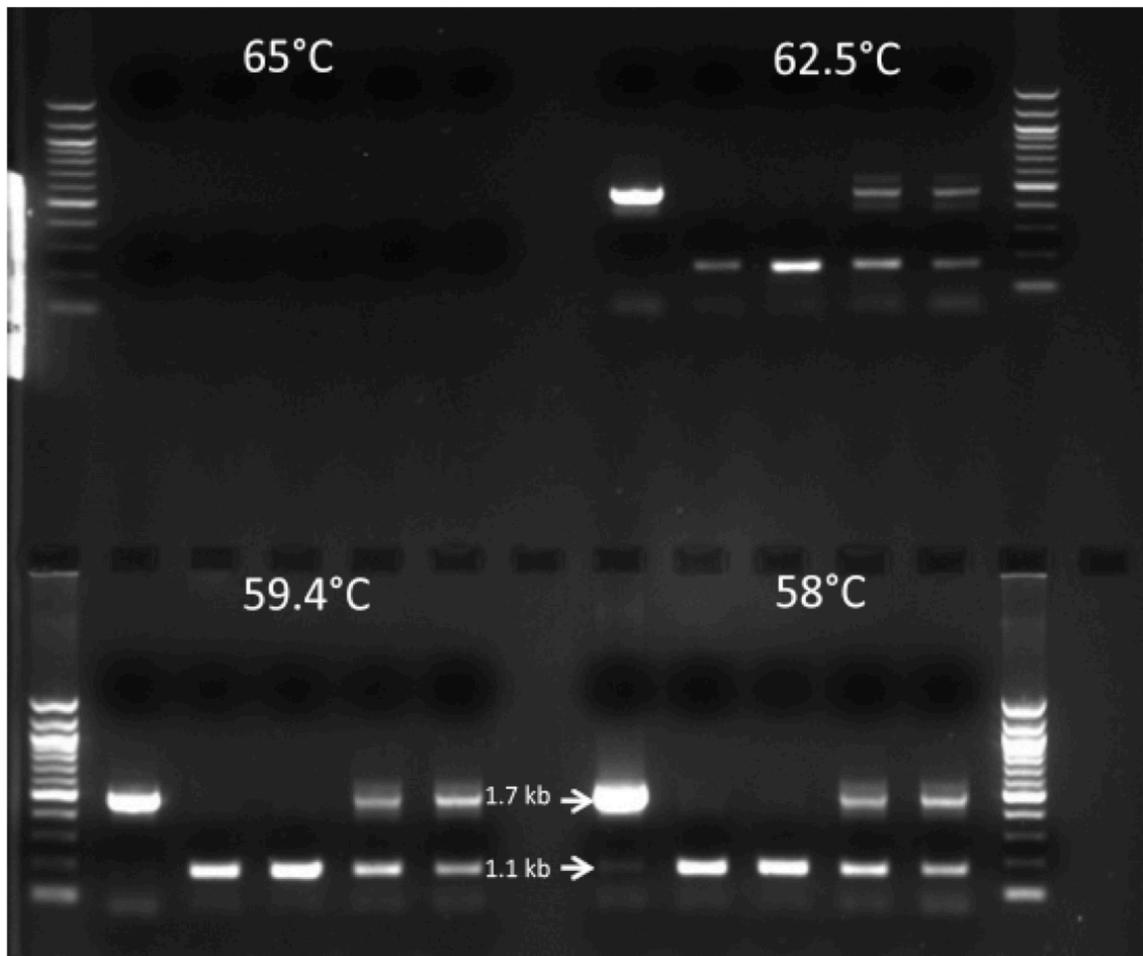


Figure 5: PCR products of Coriell Panel DNA with temperature gradient. 293A cell lysate (Alu (+/+)) and Coriell Panel DNA samples were run at four different temperatures. 58°C was chosen as the optimal annealing temperature, as it provided the highest product yield and closely matched the T_m of the primers (E1b_F $T_m=58.9^\circ\text{C}$, E1b_R $T_m=59.9^\circ\text{C}$). Alu (-/-) insertions appear at 1.1kb, while Alu (+/+) insertions appear at 1.7kb. Heterozygotes have both a 1.1 and 1.7kb PCR product. Some minor bands are seen near the 1.7kb bands in the heterozygotes, but these non-specific products are not part of our analyses.

After the PCR protocol was optimized for the Coriell DNA samples, the Hershey Tissue Bank samples were tested to ensure that the genomic DNA would still be amplified effectively. Three concentrated ($\geq 20\text{ng}/\mu\text{l}$) lung, liver, and ovary samples were chosen for amplification (Figure 6). PCR reactions included 5 μl of DNA template.

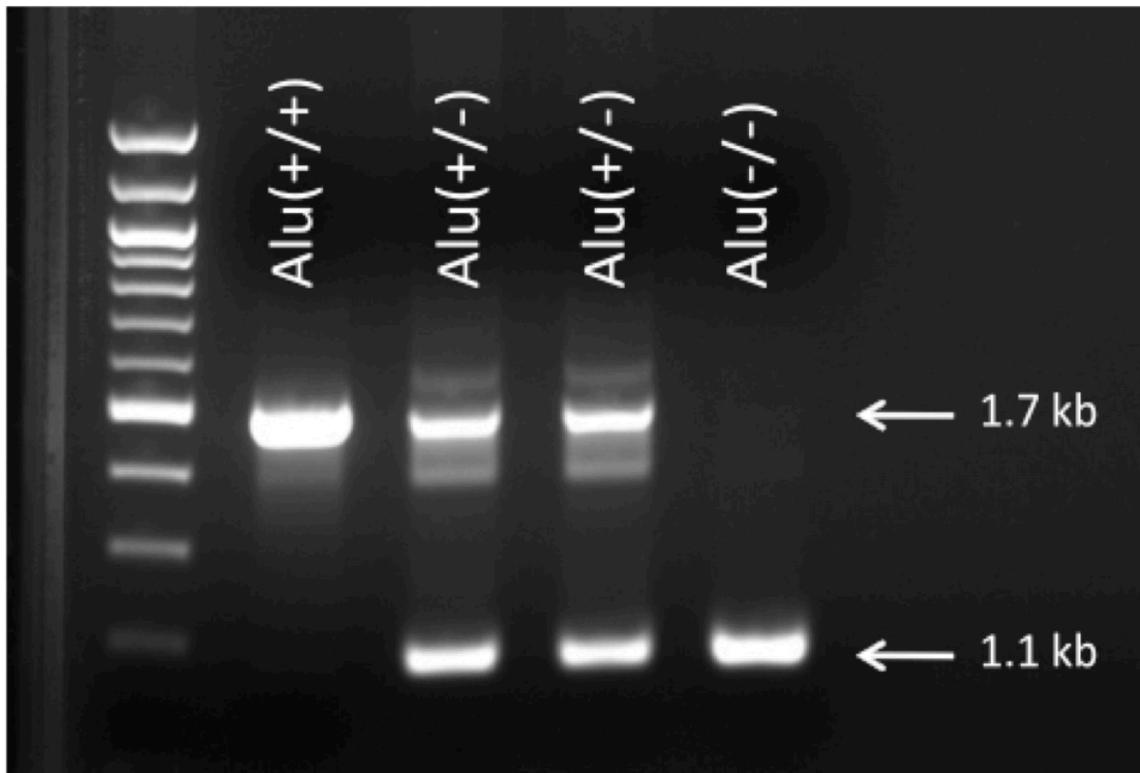


Figure 6: PCR products of tissue bank samples. PCR was performed using the optimized protocol with 5 μ l each of 293A cell lysate and Hershey lung, liver, and ovary* genomic DNA templates. The agarose gel image shows successful amplification of Alu (-/-), Alu (+/-), and Alu (+/+) genotypes. Again, non-specific products are seen near the 1.7kb band, but these products are not part of our analyses.

**Ovary samples were also obtained from the tissue bank to undergo the Alu insertion analysis, but data is not included in the present study.*

Alu Genotype Analysis

Upon successful amplification of the three test Hershey samples, the remaining DNA samples underwent the E1-b PCR protocol, and products were analyzed using gel electrophoresis. Gel images are shown in Appendix B. The number of Alu (-/-), Alu (+/-) and Alu (+/+) genotypes were totaled, and percentages were estimated (Table 2).

Table 2: Alu genotypes in lung and liver samples and overall.

	Genotype Frequencies (n(%))		
	Lung	Liver	Total
Alu (-/-)	37 (55.2)	28 (54.9)	65 (55)
Alu (+/-)	25 (37.3)	22 (43.1)	47 (39.8)
Alu (+/+)	5 (7.5)	1 (2)	6 (5.1)
Total	67	51	118

Alu (-/-), Alu (+/-), and Alu (+/+) genotypes were totaled for each tissue type and overall. Because the total number of tissue samples differed between lung and liver, percentages were calculated to provide a means of comparing the genotype totals.

Comparison to a Healthy Population

Previous studies by our lab examined the double Alu insertion genotype using genomic DNA from a healthy United States population¹⁸. To compare genotype frequencies between the healthy and diseased populations, we first determined whether the Alu insertion was in Hardy-Weinberg equilibrium in the healthy population. Observed genotype and allele frequencies were estimated from this population, and expected genotype frequencies were found using the determined allele frequencies. Pearson's chi-square test was performed to determine whether there was a significant difference between observed and expected genotype frequencies in the healthy population (Table 3). Calculations are shown in Appendix C.

Table 3: Genotype and allele frequencies of the healthy population and a population in Hardy-Weinberg equilibrium.

	Genotype Frequencies (n(%))			Allele Frequencies	
	Alu (-/-)	Alu (+/-)	Alu (+/+)	(-)	(+)
Observed	269 (59.9)	162 (36.1)	18 (4)	0.78	0.22
Expected	273.17 (60.8)	154.10 (34.3)	21.73 (4.8)	0.78	0.22
Chi-square	$\chi^2=1.109$ $p \sim 0.30$			/	

Allele frequencies were calculated from the observed genotype frequencies in the healthy population. Expected genotype frequencies were determined using the calculated allele frequencies. Pearson's chi-square test was performed by comparing observed and expected genotype frequencies, and the p value was located using a statistical table⁵.

Because $p > 0.05$, we could not reject the null hypothesis, and the healthy population appears to be in Hardy-Weinberg equilibrium.

The genotype frequencies for the diseased population were then compared to the expected genotype frequencies found in a population in Hardy-Weinberg equilibrium. Because we determined that the healthy population was in Hardy-Weinberg equilibrium, we used the allele frequencies calculated in the healthy population to determine the expected genotype frequencies of the diseased population. Pearson's chi-square test was performed to determine whether there was a significant difference between observed and expected genotype frequencies in the diseased population (Table 4). Calculations are shown in Appendix C.

Table 4: Genotype and allele frequencies of the diseased population and a population in Hardy-Weinberg equilibrium.

	Genotype Frequencies (n(%))			Allele Frequencies	
	Alu (-/-)	Alu (+/-)	Alu (+/+)	(-)	(+)
Observed	65 (55)	47 (39.8)	6 (5.1)	0.75	0.25
Expected	269 (59.9)	162 (36.1)	18 (4)	0.78	0.22
Chi-Square	$\chi^2=1.70,$ $p\sim 0.20$				

Using the allele frequencies calculated in the healthy population, the expected genotype frequencies were calculated for a population in Hardy-Weinberg equilibrium. Pearson's chi-square test was performed by comparing observed genotype frequencies in the diseased population to expected frequencies in a population in Hardy-Weinberg equilibrium, and the p value was located using a statistical table⁵.

Because $p > 0.05$, the null hypothesis cannot be rejected, and the diseased population appears to be in Hardy-Weinberg equilibrium given the frequencies of the Alu genotypes in the healthy population.

Comparison of Alu Genotypes by Tissue Type

Because E1 and E1-b are expressed in different levels in lung and liver tissues, we wanted to determine whether there was a difference in Alu genotype frequencies between the tissues, as this would indicate that Alu genotype played a role in tissue-specific disease predispositions. Expected genotype frequencies for the lung and liver subpopulations were estimated using allele frequencies from the healthy population, which was found to be in Hardy-Weinberg equilibrium. Pearson's chi-square test was performed to compare the expected and observed genotype frequencies in each subpopulation (Tables 5 and 6).

Table 5: Observed and expected genotype frequencies in lung samples.

	Genotype Frequencies (n(%))			Allele Frequencies	
	Alu (-/-)	Alu (+/-)	Alu (+/+)	(-)	(+)
Observed	37 (55.2)	25 (37.3)	5 (7.5)		
Expected	40.76 (60.8)	23 (37.3)	3.24 (4.8)	0.78	0.22
Chi-Square	$\chi^2=1.479,$ $p=\sim 0.20$				

Using the allele frequencies calculated in the healthy population, the expected genotype frequencies for a population in Hardy-Weinberg equilibrium were estimated for the diseased lung samples. Pearson's chi-square test was performed by comparing observed genotype frequencies in the diseased lung subpopulation to expected frequencies in a population in Hardy-Weinberg equilibrium, and the p value was located using a statistical table⁵.

Table 6: Observed and expected genotype frequencies in liver samples.

	Genotype Frequencies (n(%))			Allele Frequencies	
	Alu (-/-)	Alu (+/-)	Alu (+/+)	(-)	(+)
Observed	28 (54.9)	22 (43.1)	1 (2)		
Expected	31.03 (60.8)	17.5 (34.3)	2.47 (4.8)	0.78	0.22
Chi-Square	$\chi^2=2.328,$ $p=\sim 0.10$				

Using the allele frequencies calculated in the healthy population, the expected genotype frequencies for a population in Hardy-Weinberg equilibrium were estimated for the diseased liver samples. Pearson's chi-square test was performed by comparing observed genotype frequencies in the diseased liver subpopulation to expected frequencies in a population in Hardy-Weinberg equilibrium, and the p value was located using a statistical table⁵.

Because $p > 0.05$ in both the lung and liver samples, the null hypothesis cannot be rejected, and the Alu insertion appears to be in Hardy-Weinberg equilibrium in both tissue types. There is no significant difference in Alu genotype between the two tissue types.

Comparison of Alu Genotypes by Gender

EPHX1 expression may also differ between the genders, as different E1-b transcript levels are expressed in the ovaries and testes (Figure 3). Alu genotype data was categorized by gender to identify any large differences in genotype frequencies. Expected genotype frequencies for female and male subpopulations were estimated using allele frequencies from the healthy population. Pearson's chi-square test was performed to compare the expected and observed genotype frequencies in each gender (Tables 7 and 8).

Table 7: Observed and expected genotype frequencies in female donor samples.

	Genotype Frequencies (n(%))			Allele Frequencies	
	Alu (-/-)	Alu (+/-)	Alu (+/+)	(-)	(+)
Observed	35 (56.5)	26 (41.9)	1 (1.6)		
Expected	37.73 (60.8)	21.28 (34.3)	3 (4.8)	0.78	0.22
Chi-Square	$\chi^2=2.576,$ $p=\sim 0.10$				

Using the allele frequencies calculated in the healthy population, the expected genotype frequencies for a population in Hardy-Weinberg equilibrium were estimated for the female donor samples. Pearson's chi-square test was performed by comparing observed genotype frequencies in the female diseased subpopulation to expected frequencies in a population in Hardy-Weinberg equilibrium, and the p value was located using a statistical table⁵.

Table 8: Observed and expected genotype frequencies in male donor samples.

	Genotype Frequencies (n(%))			Allele Frequencies	
	Alu (-/-)	Alu (+/-)	Alu (+/+)	(-)	(+)
Observed	30 (53.6)	21 (37.5)	5 (8.9)		
Expected	34.07 (60.8)	19.22 (34.3)	2.71 (4.8)	0.78	0.22
Chi-Square	$\chi^2=2.586,$ $p=\sim 0.10$				

Using the allele frequencies calculated in the healthy population, the expected genotype frequencies for a population in Hardy-Weinberg equilibrium were estimated for the male donor samples. Pearson's chi-square test was performed by comparing observed genotype frequencies in the male diseased subpopulation to expected frequencies in a population in Hardy-Weinberg equilibrium, and the p value was located using a statistical table⁵.

Because $p > 0.05$ in both the female and male donor samples, the null hypothesis cannot be rejected, and the Alu insertion appears to be in Hardy-Weinberg equilibrium in both genders. There is no significant difference in Alu genotype between the two genders.

Comparison of Alu Genotypes by Age

The Alu double insertion may alter susceptibility to age-related diseases, such as cancer. We wanted to determine whether there was an association between diseased tissue samples from individuals under age 50 and a certain genotype, as this would indicate that specific genotypes increase susceptibility to developing cancer. Alu genotype data was categorized by age, and subpopulations were created for donors ages 0-49 and donors age 50+. Expected genotype frequencies for the young and old subpopulations were estimated using allele frequencies from the healthy population. Pearson's chi-square test was performed to compare the expected and observed genotype frequencies in each age group (Tables 9 and 10).

Table 9: Observed and expected genotype frequencies in donor samples age 0-49.

	Genotype Frequencies (n(%))			Allele Frequencies	
	Alu (-/-)	Alu (+/-)	Alu (+/+)	(-)	(+)
Observed	12 (63.2)	6 (31.6)	1 (5.3)		
Expected	11.56 (60.8)	6.52 (34.3)	0.92 (4.8)	0.78	0.22
Chi-Square	$\chi^2 = .0652,$ $p = \sim 0.80$				

Using the allele frequencies calculated in the healthy population, the expected genotype frequencies for a population in Hardy-Weinberg equilibrium were estimated for the age 0-49 donor samples. Pearson's chi-square test was performed by comparing observed genotype frequencies in the age 0-49 diseased subpopulation to expected frequencies in a population in Hardy-Weinberg equilibrium, and the p value was located using a statistical table⁵.

Table 10: Expected and observed genotype frequencies in donor samples age 50+.

	Genotype Frequencies (n(%))			Allele Frequencies	
	Alu (-/-)	Alu (+/-)	Alu (+/+)	(-)	(+)
Observed	53 (53.5)	41 (41.4)	5 (5)		
Expected	60.23 (60.8)	33.98 (34.3)	4.79 (4.8)	0.78	0.22
Chi-Square	$\chi^2 = 1.022,$ $p = \sim 0.30$				

Using the allele frequencies calculated in the healthy population, the expected genotype frequencies for a population in Hardy-Weinberg equilibrium were estimated for the age 50+ donor samples. Pearson's chi-square test was performed by comparing observed genotype frequencies in the age 50+ diseased subpopulation to expected frequencies in a population in Hardy-Weinberg equilibrium, and the p value was located using a statistical table⁵.

Because $p > 0.05$ in both the female and male donor samples, the null hypothesis cannot be rejected, and the Alu insertion appears to be in Hardy-Weinberg equilibrium in both the young and old age group. There is no significant difference in Alu genotype between the two age groups.

DISCUSSION

The EPHX1 sequence varies widely between individuals, which may affect the enzyme's structure, function, and transcriptional regulation. Because mEH is an important enzyme involved in both detoxification and bioactivation reactions, these polymorphisms may alter susceptibility to xenobiotic-induced diseases, such as cancer. The double AluYa5 insertion within the E1-b promoter region is known to downregulate transcription of E1-b, and it may serve as a marker for certain disease associations¹⁸.

We determined the double Alu insertion genotypes of lung and liver samples from a diseased population to identify any associations between genotype and diseased state. We estimated genotype frequencies in the diseased population, and we estimated genotype and allele frequencies in a previously analyzed healthy population. Using a chi-square test to compare observed and expected genotype frequencies, we determined that the double Alu insertion was in Hardy-Weinberg equilibrium in the healthy population. After using the allele frequencies from the healthy population to determine the expected genotype frequencies for the diseased population, our chi-square test determined that the insertion was also in Hardy-Weinberg equilibrium in the diseased population.

These findings imply that while the double Alu insertion may affect EPHX1 transcription and ultimately influence mEH protein levels, the Alu insertion is not more or less prevalent in diseased tissue. Because the double insertion was in Hardy-Weinberg equilibrium in both a healthy and a diseased population, the insertion does not appear to be under positive or negative selection, and it does not seem to impact reproductive success. Also, as there is no association between genotype and disease development, we cannot use the double Alu insertion as a marker or predictor of disease.

Still, we cannot conclude that the double Alu insertion does not have a role in carcinogenesis. The origins, roles, and consequences of alternative promoters and TEs are poorly understood, but recent research has shown that these non-coding sequences generate diversity among protein expression and regulation¹². The presence of one Alu insertion may also promote the integration of others; the clustering of Alu inserts within the E1-b promoter region supports this hypothesis⁷ (Figure 4). The double Alu insertion may function as a binding site for transcription factors and hormone receptors¹⁸. Consequently, its function may be impacted by other polymorphisms or by environmental factors.

A haplotype block analysis of the double Alu insertion, the Tyr113/His SNP, and the Arg139/His SNP shows that none of the three polymorphisms lies within the same haplotype block⁷. This indicates that the polymorphisms are unlinked, and each polymorphism has an independent effect on phenotype. However, the polymorphisms may still have a synergistic action, and an analysis that considers both SNPs and TEs may provide more information about genotypes and disease associations.

While this study focused determining the presence of the double AluYa5 insertion within the E1-b far upstream promoter of diseased tissue, other Alu insertions and TEs within EPHX1 may also influence mEH protein expression and thus impact disease susceptibility. There are several other Alu insertions in the upstream promoter, and four of the gene's eight introns contain clusters of Alu elements⁷. To elucidate the impact of the double Alu insertion on phenotype, future studies will investigate the roles of the other Alu inserts within EPHX1. With more information about how different Alu

insertions affect gene transcription and expression and how the insertions function together, we can better understand the implications of transposons in disease.

In addition to further analyzing genetic regulatory elements, the functional impact of polymorphisms in the protein coding sequence of mEH should be further characterized. Our previous studies suggested that the coding region polymorphisms (H113Y and H139R) do not impact mEH enzymatic activity. However, these studies only tested one substrate. Additional PAH substrates should be tested, as mEH bioactivation of PAHs is essential in initiating carcinogenesis¹⁸. By identifying PAH metabolites produced by different mEH isoforms, we can determine which genotypes are most efficient in detoxification and bioactivation, as well as which produce the most carcinogenic metabolites.

After comparing expected and observed genotype frequencies in lung and liver samples, we found that the double Alu insertion was in Hardy-Weinberg equilibrium in both subpopulations. Thus, we can make no correlations between the double Alu insertion and tissue-specific expression of EPHX1. Previous studies have shown that the insertion downregulates the transcriptional level of the associated E1-b EPHX1 transcript, suggesting that individuals with the Alu (+/+) genotype may possess lower EPHX1 activities. However, this downregulation may have different consequences depending on the tissue type and the substrates found within the tissue. Further studies will determine mRNA transcript levels in both tissue types to investigate whether the Alu (+/+) genotype has a greater effect on EPHX1 expression in either tissue. Additionally, mEH protein level and metabolic activity will be assessed in these tissues.

Alternative promoters and TEs may allow for tissue-specific regulation of transcription and protein expression, as determined by the needs of the tissue. Although we did not identify any association between an Alu genotype and disease development in lung or liver tissue, the insertion may still have consequences that vary between tissues.

We also found that the double Alu insertion was in Hardy-Weinberg equilibrium in both male and female subpopulations. A difference in genotype frequencies between the males and females would have suggested that the insertion is sex-linked, so our findings support previous studies that show no association between sex and Alu genotype. Unfortunately, time constraints prevented us from analyzing diseased ovary samples for Alu genotype. Ovary expresses a high level of E1-b transcript, while testes express E1-b at a much lower level. Interestingly, ovary is one of the few tissue types with detectable levels of E1 transcript (Figure 3).

While little is known about mEH's role in the ovary, it is thought to aid in estrogen production and may act on estrogen epoxide substrates¹⁸. Further studies are needed to determine mEH's exact function in epoxide metabolism in the ovary, as well as the impact of EPHX1 polymorphisms in the ovary. While we identified no difference in genotype frequencies among males and females from the diseased population, the double Alu insertion may play tissue-specific roles and may have different consequences in each gender.

The double Alu insertion was found to be in Hardy-Weinberg equilibrium in both the young and old subpopulations. The majority of cancer cases occur in people over 50, due to both environmental factors and the aging process⁹. We expected that if the double Alu insertion increased or decreased cancer risk, the younger population would have had

a higher frequency of Alu (+/+) or Alu (-/-) individuals, respectively. However, these results show no apparent association between age and genotype, as both populations have similar genotype frequencies. This supports our finding that the double Alu insertion occurs at the same rate in both healthy and diseased populations, as the insertion is not shown to correlate with cancer development.

Because PAH-induced carcinogenesis requires exposure to environmental carcinogens, data about the donors' lifestyles might provide additional insight regarding age-related disease and Alu genotype. Increased exposure to PAHs results in an increased risk of cancer development. However, the Alu insertion may alter sensitivity to PAHs, potentially increasing or decreasing the amount of exposure needed to cause disease. With more information about the donors' lifetime PAH exposure, we could speculate whether Alu genotype has an impact on sensitivity to PAHs.

In summary, we determined that within both a healthy and a diseased population, the allele for the double Alu insertion is in Hardy-Weinberg equilibrium. There is no significant difference between allele and genotype frequencies in the diseased and healthy populations we analyzed. Additionally, we found that the double Alu insertion is in Hardy-Weinberg equilibrium within subpopulations categorized by tissue type, gender, and age. Therefore, the insertion does not appear to greatly impact disease susceptibility, and it does not serve as a predictive marker for disease. Further studies on the impact of the double Alu insertion on E1-b transcript levels and resulting protein level may provide clinically useful information about genetic variation in mEH and its role in development of disease.

REFERENCES

1. Amos, C.I., et al. "Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1." *Nature Genetics*. 40 (2008): 616-622.
2. Deininger, P.L., Batzer, M.A. "Alu Repeats and Human Disease." *Molecular Genetics and Metabolism*. 67 (1999): 183-193.
3. Caceres, J.F., Kornblihtt, A.R. "Alternative splicing: multiple control mechanisms and involvement in human disease." *TRENDS in Genetics*. 18 (2002): 186-193.
4. Davuluri, R.V., Yutaka, S., Sugano, S., Christoph, P., Huang, T.H. "The functional consequences of alternative promoter use in mammalian genomes." *TRENDS in Genetics*. 24 (2008): 167-176.
5. Emigh, T.H. "A Comparison of Tests for Hardy-Weinberg Equilibrium." *Biometrics*. 36 (1980): 627-642.
6. Fretland, A.J., Omiecinski, C.J. "Epoxide hydrolases: biochemistry and molecular biology." *Chemico-Biological Interactions*. 129 (2000): 41-59.
7. Hassett, C., Robinson, K.B., Beck, N.B., Omiecinski, C.J. "The Human Microsomal Epoxide Hydrolase Gene (EPHX1): Complete Nucleotide Sequence and Structural Characterization." *Genomics*. 23 (1994): 433-444.
8. Hecht, S.S. "Tobacco carcinogenesis: mechanisms and biomarkers." Tobacco: Science, Policy, and Public Health. 2nd ed. New York: Oxford University Press, 2010. 127-154.
9. Hobson, Katherine. "Cancer and Age: Why We May Face a Tradeoff Between Cancer Risk and Aging." U.S. News: Health. 20 February 2009.
<<http://health.usnews.com/health-news>>
10. Hosagrahara, V.P., Rettie, A.E., Hassett, C., Omiecinski, C.J. "Functional analysis of human microsomal epoxide hydrolase genetic variants." *Chemico-Biological Interactions*. 150 (2004): 149-159.
11. Laurenzana, E.M., Hassett, C., Omiecinski, C.J. "Post-transcriptional regulation of human microsomal epoxide hydrolase." *Pharmacogenetics*. 8 (1998): 157-167.
12. Landry, J.R., Mager, D.L., Wilhelm, B.T. "Complex controls: the role of alternative promoters in mammalian genomes." *TRENDS in Genetics*. 19 (2003): 640-648.

13. Liang, S., Hassett, C., Omiecinski, C.J. "Alternative Promoters Determine Tissue-Specific Expression Profiles of the Human Microsomal Epoxide Hydrolase Gene (EPHX1). *Molecular Pharmacology*. 67 (2005): 220-230.
14. Omiecinski, C.J., Aicher, L., Swenson, L. "Developmental Expression of Human Microsomal Epoxide Hydrolase." *The Journal of Pharmacology and Experimental Therapeutics*. 269 (1994): 417-423.
15. Omiecinski, C.J., Hassett, C., Hosagrahara, Vinayak. "Epoxide hydrolase-polymorphism and role in toxicology." *Toxicology Letters*. 112-113 (2000) 365-370.
16. Sterner, Olav. Chemistry, Health, and Environment. 2nd ed. Great Britain: Wiley-Blackwell, 2010. 110-123.
17. Wang, B., Zhou, X., Dang, A., Liu, G., He, R. "Alu Repeat Polymorphism in the Gene Coding for Tissue-Type Plasminogen Activator and the Risk of Hypertension in a Chinese Han Population." *Hypertension Research*. 25 (2002): 949-953.
18. Yang, X., Liang, S., Weyant, D.M., Lazarus, P., Gallagher, C.J., Omiecinski, C.J. "The Expression of Human Microsomal Epoxide Hydrolase Is Predominantly Driven by a Genetically Polymorphic Far Upstream Promoter." *The Journal of Pharmacology and Experimental Therapeutics*. 330 (2009): 23-30.

APPENDIX A

Lung Sample Information

Donor Number	Age	Sex	DX	Genotype
8749	57	F	Adenocarcinoma	Alu (+/-)
8758	66	M	Mucinous adenocarcinoma	Alu (-/-)
8810	81	M	Squamous Cell Carcinoma	Alu (-/-)
8907	69	M	Squamous Cell Carcinoma	Alu (-/-)
8963	74	M	Adenocarcinoma	Alu (-/-)
14878	49	F	Carcinoid Tumor	Alu (-/-)
8538	55	F	Hyperplasia	Alu (+/-)
8556	73	F	Adenocarcinoma	Alu (+/-)
8568	46	M	Interstitial Fibrosis	Alu (+/+)
8608	56	M	Metastatic carcinoma origin renal	Alu (+/-)
8631	69	M	Peripheral T cell lymphoma	Alu (+/-)
8712	49	F	Large cell neuroendocrine carcinoma	Alu (-/-)
8732	75	F	Adenocarcinoma	Alu (-/-)
8187	60	F	Interstitial lung disease	Alu (-/-)
8259	60	F	Bronchioalveolar carcinoma	Alu (+/-)
8315	55	M	No viable tumor seen	Alu (+/+)
8331	77	M	Adenosquamous carcinoma	Alu (-/-)
8402	79	F	Clear Cell Carcinoma origin renal	Alu (-/-)
8433	74	M	Metastatic High Grade pleomorphic sarcoma	Alu (+/-)
8466	60	F	Squamous Cell Carcinoma	Alu (-/-)
8480	61	F	Neuroendocrine carcinoma	Alu (+/-)
8492	47	M	Small Cell Carcinoma	Alu (-/-)
8497	61	M	Adenocarcinoma	Alu (-/-)
6270	69	F	Fibrosis	Alu (+/-)
6282	61	M	Metastatic Adenocarcinoma origin colon	Alu (+/+)
6357	65	M	Fibrosis	Alu (-/-)
6379	77	F	Adenocarcinoma	Alu (-/-)
6399	66	M	Squamous Cell Carcinoma	Alu (+/-)
6483	66	M	Adenocarcinoma	Alu (+/+)
6495	72	M	Squamous Cell Carcinoma	Alu (-/-)
6498	80	M	Adenocarcinoma	Alu (-/-)
6503	50	F	Carcinoid Tumor	Alu (+/+)
6520	58	F	Bronchioalveolar carcinoma	Alu (+/-)
6653	66	M	Squamous Cell Carcinoma	Alu (-/-)
6765	53	M	Adenocarcinoma	Alu (+/-)
6784	53	F	Adenocarcinoma	Alu (-/-)
6799	50	M	Squamous Cell Carcinoma	Alu (-/-)

6893	73	F	Adenocarcinoma	Alu (-/-)
6927	79	M	Squamous Cell Carcinoma	Alu (-/-)
6964	74	M	Squamous Cell Carcinoma	Alu (+/-)
8010	64	M	Emphysema	Alu (-/-)
8079	23	M	Metastatic Osteosarcoma	Alu (+/-)
8141	75	M	Large cell neuroendocrine carcinoma	Alu (-/-)
2509	78	F	Caseating granuloma	Alu (-/-)
2531	69	F	Squamous Cell Carcinoma	Alu (+/-)
2621	69	F	Adenocarcinoma	Alu (+/-)
2909	79	M	Squamous Cell Carcinoma	Alu (+/-)
2919	67	F	Squamous Cell Carcinoma	Alu (+/-)
2924	60	F	Adenocarcinoma	Alu (+/-)
2949	59	F	Adenocarcinoma	Alu (-/-)
2955	46	F	Fungal bronchopneumonia	Alu (-/-)
4229	75	M	Malignant lymphoma	Alu (-/-)
4448	71	M	No Path Dx	Degraded
4588	66	F	Adenocarcinoma	Alu (-/-)
6014	62	M	Adenocarcinoma	Alu (+/-)
6040	80	M	Adenocarcinoma	Alu (+/-)
6091	36	F	Squamous Cell Carcinoma	Alu (-/-)
6095	68	F	Adenocarcinoma	Alu (+/-)
6178	47	F	Inflammation	Alu (-/-)
6204	46	F	Adenocarcinoma	Alu (+/-)
362	68	M	Adenocarcinoma	Alu (-/-)
368	74	F	Adenocarcinoma	Alu (-/-)
2091	44	F	Adenocarcinoma	Alu (-/-)
198	65	M	No Path Dx	Alu (-/-)
2346	49	F	Adenocarcinoma	Alu (+/-)
2783	55	M	Squamous Cell Carcinoma	Alu (-/-)
2485	62	F	Squamous Cell Carcinoma	Alu (-/-)

Liver Sample Information

Donor Number	Age	Sex	DX	Genotype
6412	70	F	Metastatic adenocarcinoma origin colon	Alu (+/-)
6518	66	M	Intrahepatic Cholangiocarcinoma	Alu (-/-)
6624	68	F	Metastatic adenocarcinoma origin colon	Alu (-/-)
6668	73	F	Negative for tumor	Alu (-/-)
7695	67	M	Chronic Inflammation	Alu (+/+)
8046	56	F	Hepatocellular adenoma	Alu (-/-)
8216	62	F	Biliary cystadenoma	Alu (+/-)
8222	83	M	Benign fibrovascular nodule	Alu (-/-)
8349	75	F	Solitary Cyst	Alu (-/-)
8716	62	F	Focal nodular hyperplasia	Alu (-/-)
8762	47	F	Metastatic adenocarcinoma	Alu (+/-)
8818	73	F	Macrovesicular steatosis	Degraded
14876	44	F	Cystadenoma	Degraded
5067	59	M	Hepatic Carcinoma	Degraded
5096	66	M	No Path Dx	Alu (-/-)
5111	46	F	Metastatic RCC origin kidney	Degraded
5147	71	F	Metastatic cancer origin colon	Degraded
5168	61	F	Focal nodular hyperplasia	Alu (-/-)
5211	73	M	Metastatic adenocarcinoma origin colon	Degraded
5221	67	M	Metastatic adenocarcinoma origin colon	Alu (-/-)
5222	67	M	Adenocarcinoma	Alu (+/-)
5233	55	F	Endometriosis w/ hyperplasia	Degraded
5245	67	F	Metastatic adenocarcinoma origin colon	Alu (+/-)
5247	74	F	Metastatic adenocarcinoma origin colon	Alu (-/-)
5258	76	M	Metastatic adenocarcinoma origin colon	Degraded
5334	51	M	Metastatic cancer origin colon	Alu (-/-)
6011	67	M	No residual malignancy	Degraded
6013	40	F	Metastatic carcinoid tumor origin lung	Alu (-/-)
6021	76	M	Metastatic adenocarcinoma origin colon	Degraded
6056	60	F	Focal nodular hyperplasia	Degraded
6066	63	M	Metastatic adenocarcinoma origin colon	Alu (+/-)
6080	62	M	Metastatic adenocarcinoma origin colon	Alu (+/-)
6086	79	F	Metastatic adenocarcinoma origin colon	Alu (-/-)

6156	36	F	Cavernous hemangioma	Alu (-/-)
6247	59	F	Benign cyst	Degraded
6277	50	F	No residual tumor	Alu (-/-)
4172	35	F	No Path Dx Autopsy	Alu (-/-)
2	50	F	Hyperplasia	Alu (+/-)
2029	71	M	Metastatic adenocarcinoma origin colon	Alu (+/-)
2088	51	F	Metastatic Carcinoid Tumor origin colon	Alu (+/-)
2127	58	M	Hemangioma	Degraded
2225	62	M	Metastatic mucinous adenocarcinoma origin colon	Alu (+/-)
2298	72	M	Metastatic adenocarcinoma origin colon	Alu (-/-)
2304	71	M	Metastatic adenocarcinoma origin colon	Alu (-/-)
2332	26	F	Sarcoma	Alu (-/-)
2433	65	M	Metastatic adenocarcinoma origin colon	Alu (+/-)
2464	64	F	Metastatic carcinoma origin pancreas	Degraded
2511	53	M	Metastatic adenocarcinoma origin colon	Alu (-/-)
2580	54	F	Metastatic adenocarcinoma origin colon	Alu (+/-)
2619	86	M	Metastatic adenocarcinoma origin colon	Alu (-/-)
2701	66	F	Metastatic adenocarcinoma origin colon	Alu (+/-)
2778	53	F	Cholangiocarcinoma,G2	Alu (-/-)
2779	68	F	Metastatic adenocarcinoma origin pancreas	Degraded
2786	44	F	Metastatic adenocarcinoma origin colon	Degraded
2920	57	M	Metastatic adenocarcinoma origin colon	Degraded
2999	45	F	Focal nodular hyperplasia	Alu (+/-)
3124	73	M	Metastatic adenocarcinoma origin small intestine	Alu (-/-)
3235	78	F	Cholangiocarcinoma	Alu (-/-)
3243	83	F	Hepatocellular Carcinoma	Alu (+/-)
3248	59	M	Metastatic adenocarcinoma origin colon	Alu (+/-)
3292	71	F	Metastatic adenocarcinoma origin colon	Degraded
3664	39	F	Metastatic adenocarcinoma origin tongue	Alu (+/-)
3874	50	M	Metastatic adenocarcinoma origin colon	Alu (+/-)
4030	75	M	Hepatocellular Carcinoma	Alu (+/-)
4140	71	M	Metastatic Adenocarcinoma	Alu (+/-)

			origin unknown	
4423	14	F	Infantile Hemangioendothelioma	Alu (-/-)
4568	72	F	Metastatic Carcinoma origin colon	Alu (-/-)
5014	57	F	Metastatic Carcinoma origin colon	Alu (+/-)
5058	63	M	Hepatocellular Carcinoma G1 to G2	Alu (-/-)

APPENDIX B

Each gel includes a 100bp DNA ladder (lane 1), 293A cell lysate (lane 2), HH1484 cell lysate (lane 3), and nuclease-free water (lane 4). Alu (-/-) appears as a single band at 1.1kb, Alu (+/-) appears as bands at both 1.1kb and at 1.7kb, and Alu (+/+) appears as a single band at 1.7kb.

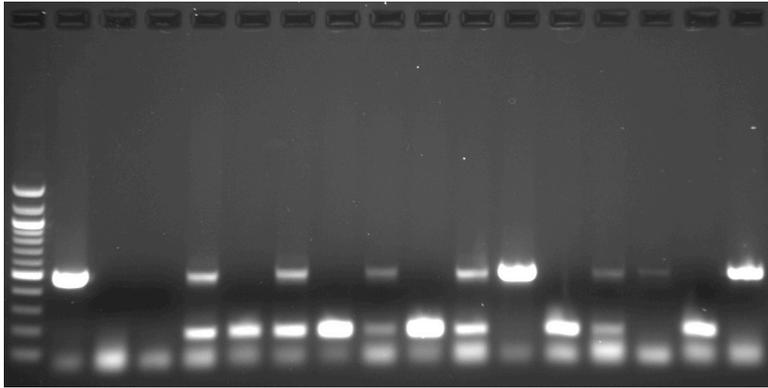


Figure 7: PCR products of lung samples 8749, 8907, 8538, 8466, 8480, 8492, 6270, 6282, 6357, 6399, 6483, 6495, and 6503.

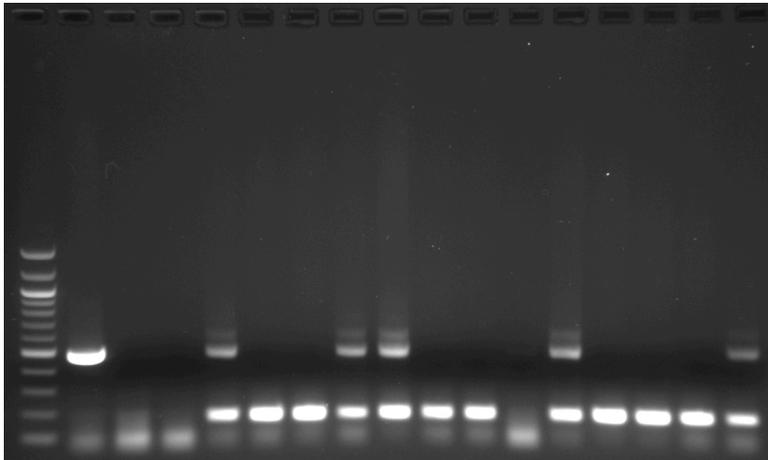


Figure 8: PCR products of lung samples 6765, 6784, 6893, 6964, 8079, 8141, 2509, 2909, 2919, 2949, 2955, 4588, and 6014.

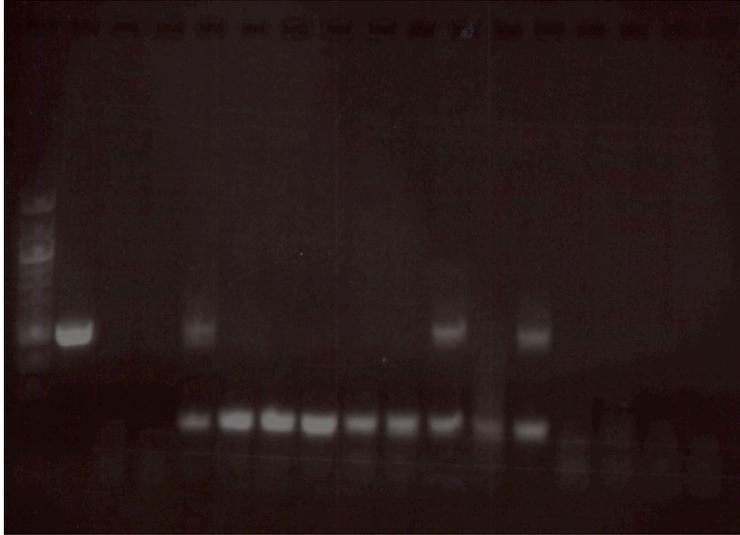


Figure 9: PCR products of lung samples 6040, 6091, 6178, 362, 368, 2091, 2193, 198, 2346, 8758, 8810, 8963, and 14878.

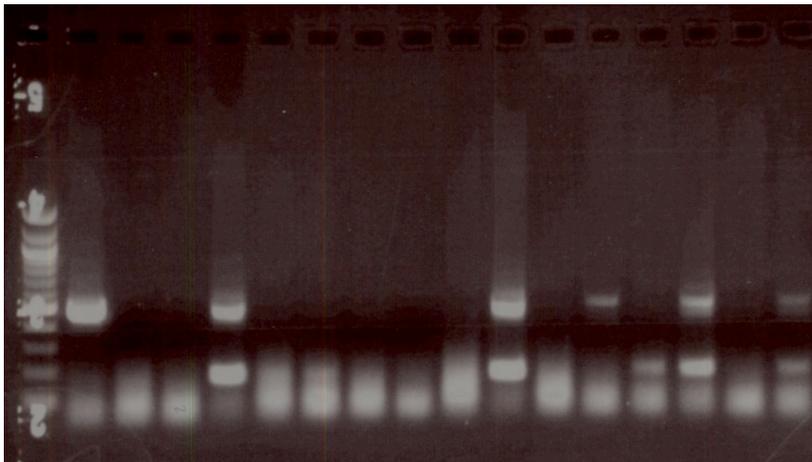


Figure 10: PCR products of liver samples 6412, 6518, 6624, 6668, 8046, 8222, 8762, 5096, 5245, 6013, 6066, 4172, and 2029.

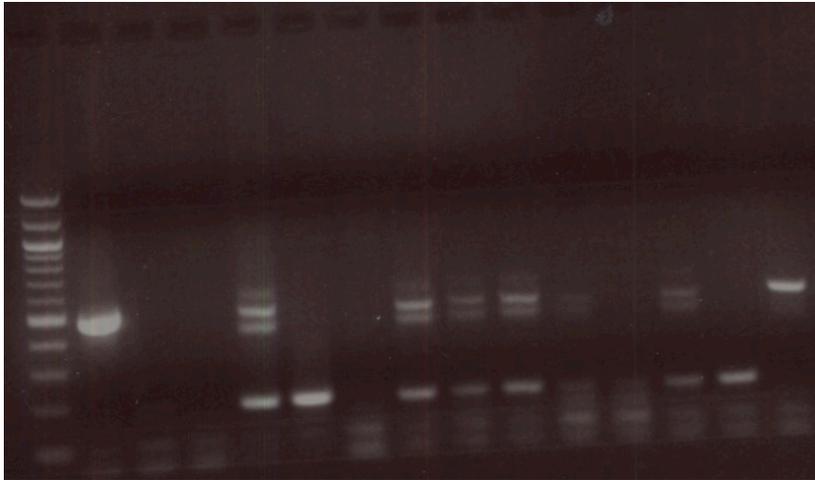


Figure 11: PCR products of liver samples 2225, 2304, 2332, 2433, 2580, 2701, 2999, 3124, 4140, 4423, and 5014.

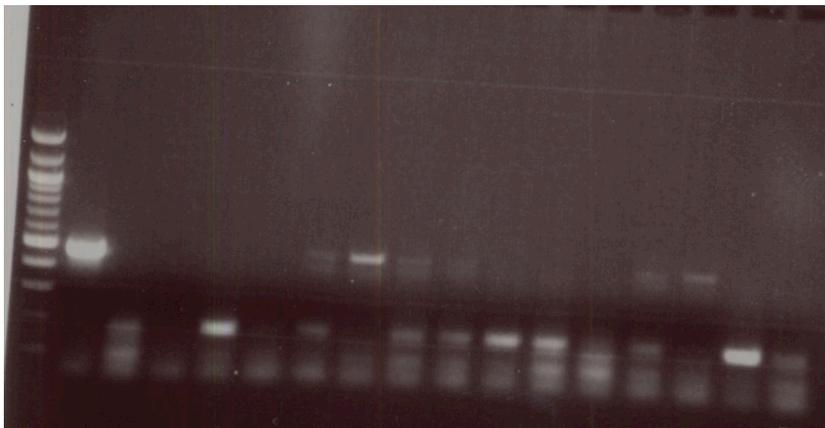


Figure 12: PCR products of lung samples 8963, 14878, 8556, 8568, 8608, 8631, 8712, 8732, 8187, 8259, 8315, 8331, and 8402.

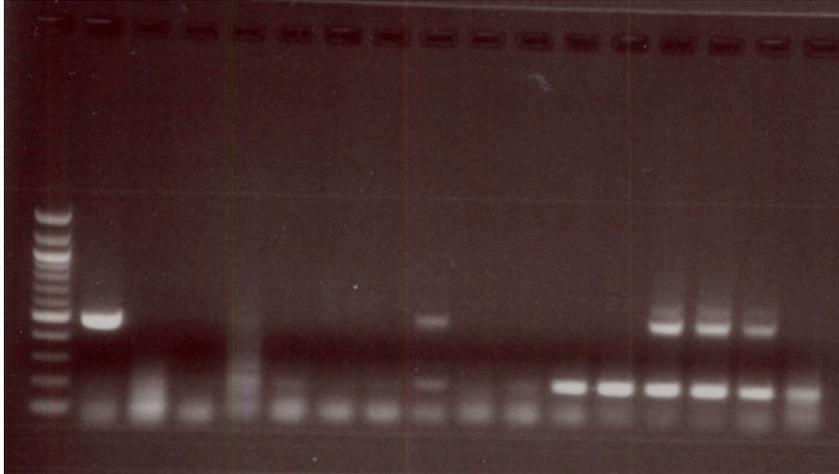


Figure 13: PCR products of lung samples 8433, 8497, 6379, 6498, 6520, 6653, 6799, 6927, 8010, 2537, 2621, 2924, and 4229.

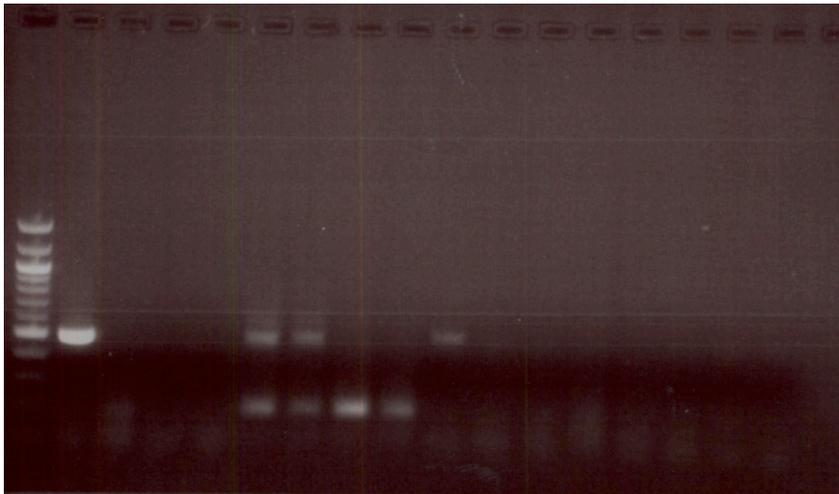


Figure 14: PCR products of lung and liver samples 4448, 6095, 6204, 2782, 2485, 7695, 8216, 8349, 8716, 8818, 14876, 5067, and 5111.

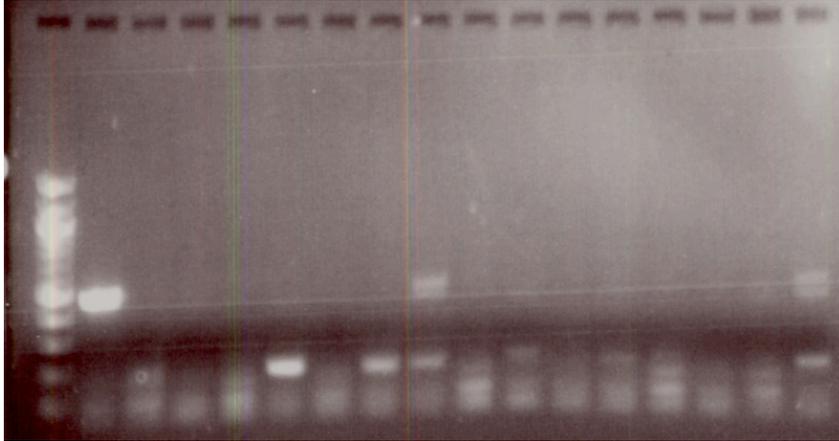


Figure 15: PCR products of lung samples 5147, 5168, 5211, 5221, 5222, 5233, 5247, 5258, 5334, 6011, 6021, 6056, and 6080.

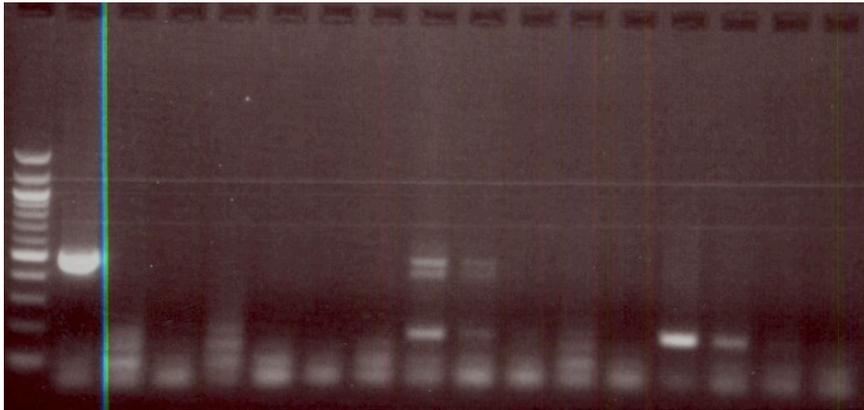


Figure 16: PCR products of liver samples 6086, 6156, 6247, 6277, 2, 2088, 2127, 2298, 2469, 2511, 2619, 2778, and 2779.

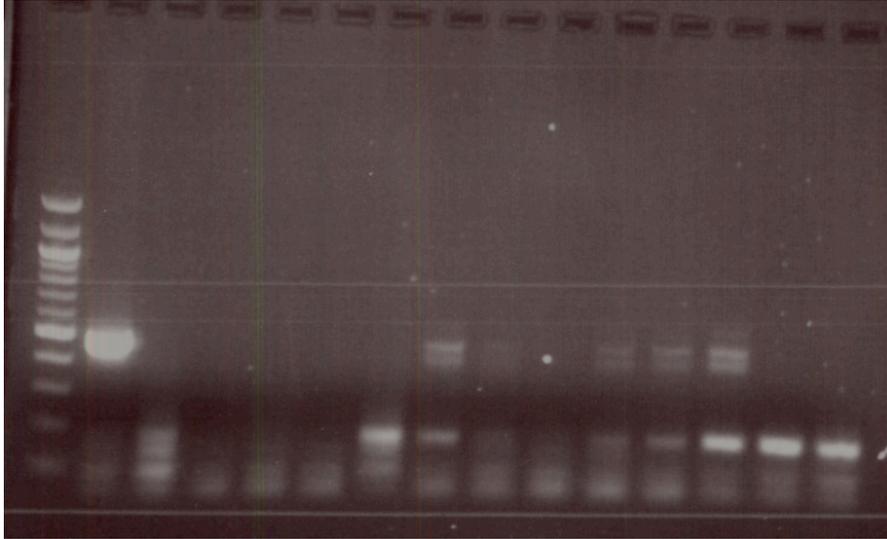


Figure 17: PCR products of liver samples 2786, 2920, 3235, 3243, 3248, 3292, 3664, 3874, 4030, 4568, and 5058.

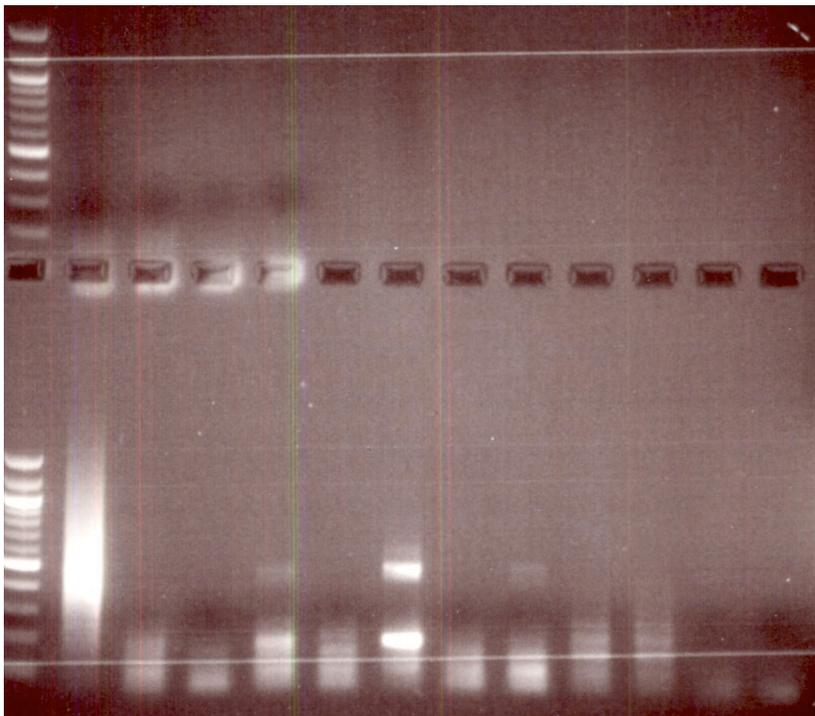


Figure 18: PCR products of lung and liver samples 8433 (in duplicate), 2909, 4448, 8216, 8349, 8716, 8818, 14876, 5067, 5111, 5147, and 5111.

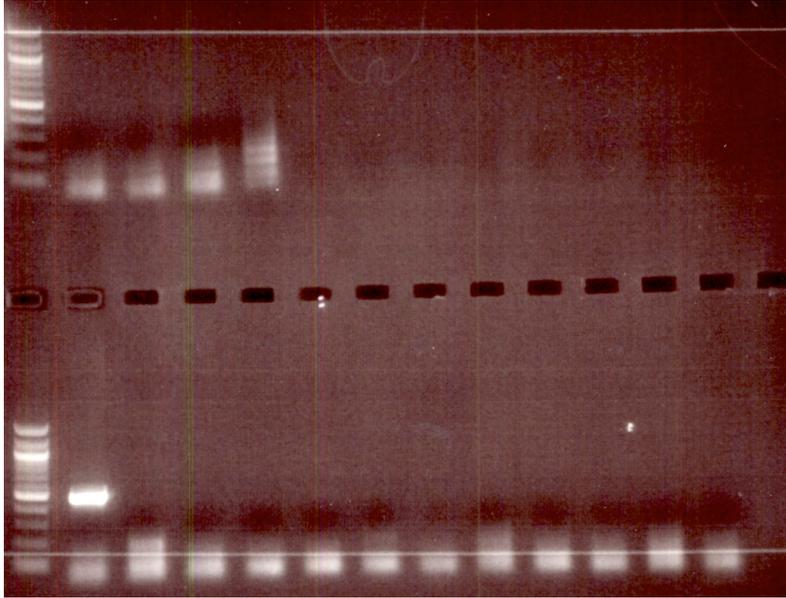


Figure 19: PCR products of liver samples 5211, 5233, 5247, 5258, 5334, 6011, 6056, 6247, 2127, 2464, 2779, and 2920.

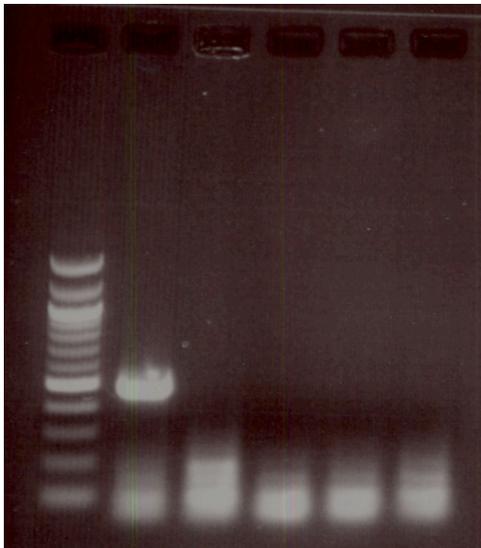


Figure 20: PCR products of liver samples 2786 and 3292.

APPENDIX C

The Hardy-Weinberg principle states that when a population is in equilibrium:

$$p^2 + 2pq + q^2 = 1$$

Where p^2 is the genotype frequency of Alu (-/-), $2pq$ is the genotype frequency of Alu (+/-), and q^2 is the genotype frequency of Alu (+/+).

The principle also states that for a population in equilibrium:

$$p + q = 1$$

Where p is the allele frequency of Alu (-), and q is the allele frequency of Alu (+)⁵.

Allele frequencies were calculated for the healthy population using the following rearrangement of the Hardy-Weinberg equation, where *obs*= observed number of individuals of each genotype:

$$p = \frac{2 \times \text{obs}(-/-) + \text{obs}(+/-)}{2(\text{obs}(-/-) + \text{obs}(+/-) + \text{obs}(+/+))}$$

$$p = \frac{2 \times 269 + 162}{2(269 + 162 + 18)}$$

$$p = \frac{700}{898}$$

$$p = 0.78$$

After solving for p , q was calculated using the following equation:

$$p + q = 1$$

$$q = 1 - p$$

$$q = 1 - 0.78$$

$$q = 0.22$$

We used these allele frequencies to calculate the expected genotype frequencies for a population in Hardy-Weinberg equilibrium, where n =sample size (449 individuals):

$$Exp(-/-) = p^2 n = (0.78)^2 (449) = 273.17$$

$$Exp(+/-) = 2 p q n = 2(0.78)(0.22)(449) = 154.10$$

$$Exp(+/+) = q^2 n = (0.22)^2 (449) = 21.73$$

To determine whether the expected genotype frequencies differed significantly from the observed genotype frequencies, Pearson's chi-square test was performed. For Hardy-Weinberg proportions, degrees of freedom is calculated as:

$$df = \# \text{ genotypes} - \# \text{ alleles}$$

$$df = 3 - 2$$

$$df = 1$$

The chi-square test was calculated using observed and expected frequencies for Alu (-/-), Alu (+/-), and Alu (+/+) frequencies. H_0 : The healthy population is in Hardy-Weinberg equilibrium. H_a : The healthy population is not in Hardy-Weinberg equilibrium. O = observed number of individuals of each genotype, and E = expected number of individuals of each genotype:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = \frac{(269 - 273.17)^2}{273.17} + \frac{(162 - 154.10)^2}{154.10} + \frac{(18 - 21.73)^2}{21.73}$$

$$\chi^2 = 0.0637 + 0.4050 + 0.6403$$

$$\chi^2 = 1.109$$

The p-value for $\chi^2=1.109$ and $df=1$ is $\sim 0.3^5$. Because $p > 0.05$, the null hypothesis is not rejected, and the Alu insertion is considered to be in Hardy-Weinberg equilibrium within the healthy population.

After determining that the Alu insertion was in Hardy-Weinberg equilibrium in the healthy population, we wanted to determine whether the insertion was in equilibrium in the diseased population. We calculated expected genotype frequencies in the diseased population using the allele frequencies calculated in the healthy population, as the healthy population was found to be in Hardy-Weinberg equilibrium. n =sample size (118 individuals):

$$Exp(-/-) = p^2n = (0.78)^2(118) = 71.79$$

$$Exp(+/-) = 2pqn = 2(0.78)(0.22)(118) = 40.50$$

$$Exp(+/+) = q^2n = (0.22)^2(118) = 5.71$$

We performed another chi-square test using expected and observed genotype frequencies within the diseased population. H_0 : The diseased population is in Hardy-Weinberg equilibrium. H_a : The diseased population is not in Hardy-Weinberg equilibrium. O = observed number of individuals of each genotype, and E = expected number of individuals of each genotype:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = \frac{(65 - 71.79)^2}{71.79} + \frac{(47 - 40.50)^2}{40.50} + \frac{(6 - 5.71)^2}{5.71}$$

$$\chi^2 = 0.642 + 1.043 + 0.0147$$

$$\chi^2 = 1.70$$

The p-value for $\chi^2=1.70$ and $df=1$ is $\sim 0.2^5$. Because $p > 0.05$, the null hypothesis is not rejected, and the Alu insertion is considered to be in Hardy-Weinberg equilibrium within the diseased population.

Academic Vita:

Emily Fay
D10 Armenara Plaza
131 Sowers Street
State College, PA 16801
efay456@gmail.com

Education:

B.S. in Biology, Vertebrate Physiology Option
The Pennsylvania State University, *Spring 2011*
Honors in Biology
Frequency of AluYa5 Insertion in a Regulatory Region of
Microsomal Epoxide Hydrolase in Human Diseased Tissue
Thesis Supervisor: Curtis Omiecinski

Related Experience:

Penn State Department of Chemistry Employee
Chem 202, 212 Exam Grader. Chem 110, 210 Proctor.
Spring 2010-Spring 2011
Internship with Vitae Pharmaceuticals
Supervisor: Yuri Bukhtiyarov
Completed *in vitro* primary and secondary screening assays of
novel BACE-1 inhibitors to treat Alzheimer's disease.
Summer 2008 and 2009

Awards:

Alpha Epsilon Delta Premedical Honors Society Inductee
Hammond Scholarship
Dean's List. *Fall 2007-Fall 2011*

Activities:

Student Red Cross Club Volunteer. *Fall 2010-Spring 2011*
OPPerations Committee Member, *Fall 2009-Spring 2011*
Mount Nittany Medical Center Volunteer, *Fall 2009-Fall 2010*
Knitivism Member, *Spring 2009-Fall 2010*