2018

# Evaluating the Recovery of DNA from Adhesive Tape after Exposure to Heat and Humidity: Assessing the Degradation Index and STR Profile

Emily Davis

# Evaluating the recovery of DNA from adhesive tape after exposure to heat and humidity: assessing the degradation index and STR profile

**Emily Davis**

**Thesis submitted**
**To the Eberly College of Arts and Sciences**
**at West Virginia University**

**in partial fulfillment of the requirements for the degree of**

**Master of Science in**
**Forensic and Investigative Science**

**Keith Morris, Ph.D., Chair**
**Tina Moroose, M.S.**
**Clifton Bishop, Ph.D.**

**Department of Forensic and Investigative Science**

**Morgantown, West Virginia**
**2018**

**Keywords: DNA degradation, duct tape, electrical tape, Quantifiler™ Trio Quantification Kit, GlobalFiler™ PCR Amplification Kit**

## Abstract

# Evaluating the recovery of DNA from adhesive tape after exposure to heat and humidity: assessing the degradation index and STR profile

**Emily Davis**

It is not uncommon for forensic DNA samples to be degraded or of low quantity; these types of samples can pose challenges throughout analysis and an informative STR profile is not guaranteed. DNA analysis is also very expensive and timely; therefore, it is necessary to understand where to sample a piece of evidence to obtain the best quality and quantity of DNA from the substrate, especially in the case of trace DNA samples which may be difficult to detect. It is also necessary to know when to proceed with analysis given the quality of the sample. The aim of this research was to evaluate the effect of adhesive tape on the degradation of DNA when exposed to heat and humidity and to assess the usefulness of the Quantifiler™ Trio degradation index to estimate the level of degradation in an STR profile. DNA was deposited on or between layers of duct tape or electrical tape and incubated in a humidity chamber for up to two weeks. The degradation index and profile slope were determined for each sample to assess whether the type of tape or number of layers had a significant effect on the degradation of the DNA sample. Multiple linear regression was also used to assess the relationship between the degradation index, amplification input, and the profile slope. The type of tape, number of layers of tape, and treatment length were found to have a significant effect on the degradation index of DNA samples, however, the treatment length was the only factor that had a significant effect on the slope of the DNA profile. These results indicate that the type of tape and number of layers may significantly affect the degradation of a DNA sample given a long enough treatment period, however the degradation index alone cannot be used to evaluate the level of degradation in a DNA sample or estimate the slope of a DNA profile; factors such as the quantity of DNA amplified and method of degradation affect the profile slope as well.

# Table of Contents

# Introduction

Advancements in forensic DNA technology have broadened the types of samples that can be tested, and when properly collected, stored, and processed DNA evidence can be of high evidentiary value due to its large discriminating power. However, there are still many challenges to address. For example, a forensic sample may be exposed to environmental conditions that compromise the quality of the sample. Additionally, there has been an increased demand for the analysis of trace DNA or touch DNA samples; these types of biological samples contain a limited amount of DNA, typically less than 100 pg, and are difficult to detect[1].

Degraded or low quantity samples may pose challenges for a DNA analyst; it may be difficult to detect the sample or know where to swab the substrate, the extracted sample may need to be further concentrated, sub-optimum amounts of DNA may be available for amplification, artifacts may be present in the STR profile, and analysis may require the sample to be consumed. These challenges may be compounded when the substrate upon which the DNA sample is found requires testing in another section of the laboratory. Knowing where trace DNA samples are more likely to persist can help direct the evidence throughout various laboratory sections and aid the DNA analyst's sampling process to avoid processing multiple samples in an attempt to obtain usable results.

Adhesive tape is an example of a substrate upon which DNA may be found that would require testing in other sections of the laboratory. In addition to carrying DNA, the physical, chemical, and elemental properties of adhesive tape may be analyzed as well as fingerprints or trace evidence that may adhere to the tape. There are many types of adhesive tape with varying degrees of adhesive strength and physical properties which could affect the preservation of DNA as well as the quantity of DNA retained. A better understanding of this relationship may assist in the decision of whether DNA testing is appropriate or where to sample the adhesive tape for DNA in order to preserve appropriate amounts of the tape for testing in other sections.

In addition to knowing how to best sample a piece of evidence to collect DNA, it is also necessary to develop checkpoints throughout the workflow to determine whether DNA analysis should proceed. The quantitation step is a common checkpoint; if there is not enough human DNA present in a sample the analysis may halt at this step. Commercial quantification kits have advanced and assays now contain more than one human target. Multiple human DNA targets have been utilized to calculate a degradation index, which can theoretically be used to indicate the potential success of an STR profile. However, the weight of the degradation index is still somewhat vague.

The aim of this research was to evaluate the effect of adhesive tape on the degradation of DNA when exposed to heat and humidity and to assess the usefulness of the Quantifiler™ Trio degradation index to estimate the level of degradation in an STR profile obtained when amplifying with the GlobalFiler™ PCR Amplification kit. DNA was deposited on or between layers of duct tape or electrical tape and incubated in a humidity chamber for up to two weeks. The degradation index and profile slope were determined for each sample to assess whether the type of tape or the number of layers had a significant effect on the degradation of the DNA sample. Raw data files of degraded DNA samples were also obtained from the PROVEDIt database which were used to develop a linear model to estimate the profile slope of a DNA sample given the degradation index and quantity of DNA available for amplification; this model was applied to the samples degraded in this study to test its usefulness.

# Chapter 1: Background

## 1.1 Deoxyribonucleic acid

Deoxyribonucleic acid (DNA) is the genetic "blueprint" of life. In humans, most DNA is found in the nucleus of cells and contains 22 pairs of autosomal chromosomes and one pair of sex chromosomes. Chromosomes contain genetic information in the form of genes. A given gene has a specific location on a specific chromosome; this is referred to as a locus. There are various forms of a given locus; these are

referred to as alleles. An individual has two alleles for a given locus; one inherited from the father and the other from the mother. If the two alleles for a locus are the same, the locus is homozygous. If the two alleles for a locus are different, the locus is heterozygous.

DNA consists of two anti-parallel helical strands which are held together by purine and pyrimidine bases[2]. Each strand is composed of a backbone of deoxyribose sugars and phosphate groups[2]. The phosphate groups form phosphodiester linkages between the 3'C of one deoxyribose sugar and the 5'C of another deoxyribose sugar[2]. Branching off from the deoxyribose sugar of this backbone are the purine and pyrimidine bases. A purine base on one DNA strand is hydrogen bonded to a pyrimidine base on the other DNA strand[2]. The purine base adenine (A) binds with the pyrimidine base thymine (T) and the purine base guanine (G) binds with the pyrimidine base cytosine (C).

The sequence of bases along a strand of DNA can be thought of as an individual's genetic code, however only about 5% of the human genome codes for proteins[3]. The non-coding region of the genome does not impact the genetic fitness of an individual; for this reason, mutations have occurred throughout these regions with little consequence and have resulted in highly variable regions in the human genome known as polymorphisms, which contribute to the uniqueness of an individual's DNA profile[3].

## 1.2 DNA Analysis

Forensic DNA analysis focuses on short tandem repeats (STRs), a category of polymorphisms that contain repeating units of DNA that range between 3-8 base pairs in length and repeat consecutively between 2-20 times[3]. The number of times an STR is repeated varies among individuals. Currently, STR analysis in the United States looks at 20 Core Loci that consist of units of 4 base pairs that are polymorphic in the number of times they are repeated. The process of generating an STR profile includes extraction, quantitation, amplification, separation, and detection.

The process of extraction is used to lyse cells, remove DNA from the nucleus, and isolate DNA from other cellular components as well as organic and inorganic compounds that may interfere with DNA

analysis methods such as polymerase chain reaction (PCR). During the quantitation step, the quantity of human DNA extracted from a sample is determined. Real-time PCR or quantitative PCR (qPCR) is used to monitor DNA concentration as the template is amplified by measuring the amount of fluorescence given off by the PCR product[4]. This is necessary because commercial STR amplification kits are optimized for a specific range of input DNA and a sample may need to be concentrated or diluted before proceeding to amplification. Too much input DNA may result in incomplete +A nucleotide addition, which occurs when DNA polymerase does not have enough time to incorporate the 3'A nucleotide to the end of the PCR product [5; 6]. Too much input DNA may also result in off-scale data which may cause pull-up and mask alleles of the same base pair size that are labeled with a different fluorescent dye[4; 6] . Too little DNA input may result in unbalanced amplification of alleles due to stochastic fluctuation or allele drop-out[4; 6]. Amplification is performed through PCR, which consists of a series of heating and cooling cycles during which denaturing, primer annealing, and base pair extension occur. This process results in an exponential increase in quantity of the specific DNA fragments to be analyzed.

Capillary electrophoresis is used to separate the amplified DNA fragments as they pass through a narrow, charged capillary containing a liquid polymer matrix, which acts as a sieve[7; 8]. Electrokinetic injection is used to introduce DNA into the capillary, during which the capillary and cathode are placed in a tube containing the DNA sample and a buffer solution[9]. When a voltage is applied to the electrode, the negatively charged DNA fragments are drawn into the capillary[9]. The DNA fragments migrate to the other end of the capillary, which is placed in a separate tube containing a buffer solution and an anode[9]. Fluorescently labelled DNA fragments, obtained through PCR, are separated according to their size (molecular weight) as they migrate through the capillary; smaller DNA fragments migrate more quickly due to less resistance from the polymer[8; 10]. The fluorescently labelled DNA fragments are excited near the end of the capillary; the resulting electropherogram displays the relative fluorescence units (RFU) of the fragments as a function of molecular weight, which is represented by the number of base pairs[10; 11].

## 1.3 DNA Degradation

Exposure to environmental factors such as ultraviolet (UV) light, microbes, humidity, and heat can damage DNA and lead to degradation. There are processes that occur in living organisms to recognize and repair DNA damage, however these processes cease when the cell can no longer maintain homeostasis and begins to breakdown[12]. As cells breakdown, they release nutrients that support bacterial growth; the combination of digestive processes of bacteria and the release of cellular components can alter the pH of a biological stain[12]. An acidic pH catalyzes a reaction mechanism which results in a C-N bond cleavage between a deoxyribose sugar and pyrimidine base; this generates a free pyrimidine base and an apurinic site on the DNA strand[12; 13]. At basic conditions, the apurinic site may undergo a further reaction which cleaves the DNA strand between the 3'C and phosphate group, leading to DNA strand breakage[12; 13].

Deoxyribonucleases (DNase) are endonucleases, or enzymes which catalyze the hydrolysis of DNA[14]. DNases are found throughout all tissue types; they are secreted in the cell where some types are retained while others are secreted extracellularly[14]. There are two classes of DNase; DNase I and DNase II. DNase I enzymes are dependent on $Ca^{2+}$ and $Mg^{2+}$ and function at neutral pH, whereas DNase II can function at acidic pH without the presence of $Ca^{2+}$ or $Mg^{2+}$ [14]. The function of DNase I is negligible in the absence of $Ca^{2+}$ or $Mg^{2+}$ and optimal when both divalent cations are present, however it can function when in the presence of at least one of the two[15]. DNase I also behaves differently when in the presence of $Mg^{2+}$ or $Mn^{2+}$, when in the presence of $Mg^{2+}$, DNase I produces single strand breaks, whereas in the presence of $Mn^{2+}$ DNase I produces single and double strand breaks[16].

When DNA is exposed to UV light, the DNA absorbs photons which can cause the 5-6 double bond of a pyrimidine to open[17]. If this occurs to two adjacent pyrimidine bases, they can form a cyclobutane pyrimidine dimer[17]. A second possible dimer is a pyrimidine 6-4 pyrimidone dimer, which forms when a 5-6 double bond in a pyrimidine opens and reacts with the exocyclic moiety of an adjacent 3' pyrimidine, forming a 6-4 linkage[17]. The formation of these dimers prevents transcription and replication[17].

### 1.3.1 Quantifiler™ Trio DNA Quantification

The electropherogram of a degraded DNA sample has a characteristic ski-slope shape; shorter fragments tend to have higher RFUs compared to longer fragments. The greater the slope of a profile, the more likely peaks are to drop out of the profile or fall below the stochastic threshold. The Quantifiler™ Trio DNA Quantification Kit by Applied Biosystems uses a highly sensitive real-time PCR reaction to determine the concentration of human DNA and level of degradation in a sample[18]. The limit of detection for this kit is less than 1 pg/µL and the kit can quantify three target loci: a small autosomal locus (80 base pairs) which is used to determine the volume of extracted DNA necessary to amplify 1 ng, a large autosomal locus (214 base pairs), and a Y-chromosome locus[18].

Since larger fragments of DNA are more susceptible to degradation than smaller fragments, the ratio of the larger autosomal amplicon signal to the smaller autosomal amplification signal is calculated to determine a degradation index (DI), which is used to indicate the level of DNA degradation in a sample and to help indicate the expected success of STR amplification (Equation 1)[18]. A DI value of less than one indicates there is no DNA degradation, a value between one and ten indicates there is slight to moderate degradation, and a value greater than ten indicates the DNA is significantly degraded[18].

$$DI = \frac{[Small\ Autosomal\ Target]}{[Large\ Autosomal\ Target]} \quad (1)$$

In a stability study on degraded DNA, human male DNA was mechanically sheared with a solicitor and then digested with different amounts of DNase I over a range of incubation times in order to induce various levels of degradation[18]. The samples were amplified with the GlobalFiler™ PCR Amplification kit and the number of alleles present were counted. Results showed that the number of alleles present was influenced by the amount of DNA amplified as well as the DI value. Some samples had comparable allele counts despite a wide range of DI values. A sample with a high DI value and a large concentration of DNA

may provide more DNA for amplification compared to a sample with a low DI value and a small concentration of DNA[18].

### 1.3.2 Degradation Studies

Various studies have been performed in order to better understand how environmental factors affect DNA samples. Al-Kandari, Singh, and Sanger looked at how heat and humidity affect saliva, semen, and blood samples[19]. They exposed samples to various temperatures and degrees of humidity, including 55°C with 41% relative humidity (RH) and 37°C with 55% RH for a total of 28 days, removing samples each day to extract and quantify. They found that saliva samples exposed to 55°C with 41% RH had significantly lower quantities of DNA present after the second day and reached 0 ng/µl by the 17$^{th}$ day. Saliva samples exposed to 37°C with 55% RH degraded more slowly; quantities decreased from day 12 to 28, however these quantities were not significantly different from day one[19].

Barbaro and Cormaci also looked at blood, semen, and saliva samples to assess how DNA from these samples degrade when exposed to heat[20]. They incubated samples in an oven at 50°C, 100°C, 150°C, and 200°C for 20 minutes. They found no significant differences in the quantity of DNA from samples that were untreated and exposed to 50°C or 100°C. However there was a 50% reduction in the quantity of DNA present when comparing samples exposed to 100°C to samples exposed to 150°C and a 75% reduction in the quantity of DNA present when comparing samples exposed to 150°C to samples exposed to 200°C[20].

Thacker *et al.* compared the degradation of blood samples exposed to UV light and heat with humidity[21]. They exposed some blood samples to UV light for various periods of time ranging from 10-120 seconds and other samples to 37°C with 85% RH for up to four weeks. For samples exposed to UV light, there was clear drop-out observed for larger fragments which increased with the length of exposure to UV. There were no full profiles observed for samples exposed to UV light for 120 seconds. Samples incubated at 56°C with 85% RH for four weeks resulted in reduced peak heights, however some samples were able to produce full profiles at the 50 RFU analytical threshold[21].

Bright *et al.* fit the characteristic ski-slope shape of a degraded DNA profile to a linear model as well as an exponential model in order to assess which model provides a better representation of the relationship between peak height and fragment length in degraded samples[22]. If peak height declines constantly with increasing fragment size, a linear model would fit, however, if degradation occurs randomly along the backbone of the DNA strand, you would expect an exponential relationship between peak height and fragment size. The linear model resulted in extreme positive departures from the best-fit line at the high molecular weight end and extreme negative departures from the best-fit line for fragments with mid-ranged lengths. The exponential model reduced the number of departures from the best-fit line[22].

Vernarecci *et al.* compared the DI obtained from Quantifiler™ Trio to the quality of STR profile obtained when using the GlobalFiler™ PCR Amplification Kit by Applied Biosystems[23]. They assessed 181 forensic casework samples. Average peak heights were calculated for alleles in the 75-95 base pair range as well as for alleles in the 204-224 base pairs range for 95 forensic casework samples. The average peak heights for alleles in the 75-95 base pair range were comparable for nondegraded and degraded samples; this was expected because the concentration of the 80 base pairs amplicon in the Quantifiler™ Trio kit was used to determine the volume of sample necessary to amplify. The ratio between average peaks heights in the 75-95 base pairs range to the 204-224 base pairs range was also calculated and plotted against the degradation index; a linear relationship was observed, however the 75-95 base pairs/204-224 base pairs ratio was generally 1.5 times greater than the degradation index value. This suggests there is more degradation observed in the STR profile compared to the quantification data. Vernarecci *et al*. extended their analysis to the entire STR profile by calculating the log-linear relationship between fragment length and peak height. Their analysis included sub-threshold alleles to increase the quantity of data for severely degraded samples with fewer detected alleles above the analytical threshold. The measure of degradation (p) obtained from this model was then plotted against the degradation index

obtained from quantification. An inverse linear relationship ($R^2$=0.70) was observed between the measure of degradation in the STR profile (p) and the DI obtained from quantification[23].

Kitayama *et al.* artificially degraded DNA samples by digesting them with DNase I for various lengths of time in the presence of either $Mn^{2+}$ or $Mg^{2+}$ as well as treating samples with Methylene blue in the presence of visible light for various lengths of time[16]. A degradation index was determined through real-time PCR quantification of two human specific sequences, one that was 98 base pairs long and the other 207 base pairs long. STR profiles were generated and it was observed that the number of detectable STR loci was inversely related to the degradation index. The number of detectable loci decreased more rapidly when DNA was digested with DNase I in the presence of $Mn^{2+}$ compared to $Mg^{+2}$. The number of detectable loci also decreased more rapidly when DNA was treated with Methylene blue in the presence of visible light compared to when it was digested with DNase I. This indicates that the relationship between the number of detectable loci in an STR profile and the degradation index obtained through quantification varies depending on the method of degradation. They also found that the relationship between the number of loci detected and the degradation index of aged blood stains was more closely correlated with that of DNA artificially degraded by digestion with DNase I in the presence of $Mn^{+2}$ compared to the other methods of artificial degradation[16].

The Project Research Openness for Validation with Empirical Data (PROVEDIt) Database at Rutgers University-Camden is hosted online at the Laboratory for Forensic Technology Development and Integration (LFTDI) website and free to anyone. The database contains .hid and .fsa datafiles for over 25,000 STR profiles which were generated from one to five contributors with various contributor ratios, under various laboratory conditions inducing degradation and inhibition, amplification target masses ranging from 0.007-1 ng, and with different instrumentation and commercial STR kits including a 3500 Genetic Analyzer and the GlobalFiler™ STR multiplex[24]. The known genotypes for all participants were also provided. Laboratory conditions used to induce DNA degradation included exposure to UV light,

enzymes, and sonication. A 7500 Real-Time PCR System and the Quantifiler Trio™ DNA Quantification kit were used to quantify human DNA; the degradation index obtained with Quantifiler™ Trio was reported for each sample.

Alfonse *et al.* utilized the PROVEDIt Database to assess the degradation of an STR profile at a threshold of 1 RFU by modeling the sloped contour of the STR profile by the exponential decay in fluorescence as a function of molecular weight using the following equation: $H_l = Ae^{B\bar{s}_l}$, where $H_l$ is the sum of the peak heights associated with the known genotypes at locus $l$, $\bar{s}_l$ is the average base pair size of the STR alleles at locus $l$, and $A$ and $B$ are the exponential parameters obtained for each sample using least squares regression[24]. In a highly degraded sample, the larger molecular weight STR fragments exhibit lower peak heights compared to the smaller molecular weight STR fragments which causes $B$ to be a large negative value[24]. If there is minimal or no degradation present, the peak heights will be more balanced across all loci in the sample and $B$ will be near zero[24]. Untreated samples were found to have a Quantifiler™ Trio DI near one and $B$ near zero[24]. Alfonse *et al.* plotted the log of the Quantifiler™ Trio degradation index versus the calculated $B$ value for each sample to assess correlation; it was found that correlation was strong for samples that had been exposed to conditions to induce degradation, suggesting the Quantifiler™ Trio degradation index could be used to predict the slope on an STR profile, however, the relationship between the log of the Quantifiler™ Trio degradation index and the calculated $B$ value varied depending on the method used to induce degradation[24].

## 1.4 GlobalFiler™ PCR Amplification

The GlobalFiler™ PCR Amplification Kit by Applied Biosystems is a six-dye kit that amplifies 24 loci, 10 of which are miniSTRs with amplicon lengths of 220 base pairs or less [6; 25]. MiniSTRs can be used to analyze low quantity or low quality DNA. MiniSTR multiplex PCR kits contain primers which have been moved closer to the sequence of interest in order to reduce the resulting amplicon length[26]. Conventional amplicon lengths range from 100-450 base pairs, whereas miniSTR amplicon lengths are below 200 base

pairs[27]. High molecular weight amplicons, which are more likely to drop out of degraded or low quantity DNA samples, can be more reliable typed when amplicon lengths are shorted by using miniSTRs[26-28].

An optimized buffer system and these 10 miniSTRs maximize the sensitivity of the GlobalFiler™ PCR Amplification kit for degraded, inhibited, and low quantity DNA samples[6]. The discrimination power of the ten miniSTR loci included in GlobalFiler™ alone results in a probability of identity (PI) of $9.2x10^{-12}$, based on a US Caucasian database[25]. The PI is $7.12x10^{-26}$ when all 24 GlobalFiler™ loci are taken into account[6; 25].

In a sensitivity study that was performed on the GlobalFiler™ PCR Amplification Kit, the DNA control was serially diluted from 1.0-0.031 ng[6]. The recommended amount of input DNA for amplification is 1.0 ng, however full profiles were consistently observed when at least 0.125 ng of DNA was input for amplification[6]. Partial profiles were occasionally observed when quantities less than 0.125 ng of DNA were amplified[6].

## 1.5 Adhesive Tape

Adhesive tape has been collected as evidence from various crimes involving the immobilization or silencing of victims, kidnappings, homicides, burglary, wrapping drug packets or other objects, the preparation of mechanical or electrical gadgets, postal bombs, and improvised explosive devices[24-29-36]. Forensic analysis of adhesive tapes consists of the examination of physical characteristics, polarized light microscopy, Fourier transform infrared spectroscopy, elemental techniques, and pyrolysis gas chromatography[37]. However, adhesive tape can also serve as a carrier of DNA evidence. The Scientific Working Group for Materials Analysis (SWGMAT) conducted a survey in 2012 in which they received 130 responses from 18 different countries and 105 different laboratory systems[38]. Of the laboratories that responded from the US, about 95% received duct tape samples and about 75% received electrical tape samples. The most common method of separating pieces of tape was to mechanically pull them apart at room temperature. In about 30% of laboratories, adhesive tape samples were sent to the trace evidence

section first, however in about 30% of laboratories it was sent to the DNA section first. In about 20% of laboratories, adhesive tape goes to the trace evidence section second and in about 60% of laboratories it goes to the DNA section second[38].

Zech, Malik, and Thali assessed the applicability of DNA analysis on adhesive tape samples[30]. They reviewed 100 cases from 1999-2010 that involved adhesive tape or similar items and analyzed 152 samples. They were able to obtain 98 DNA profiles. Feasible profiles were defined as having at least six confirmed loci for single source profiles when analyzed in duplicate or eight confirmed loci per contributor for profiles with two contributors. Complete DNA profiles were defined as having confirmed alleles at all analyzed loci. Of the 98 DNA profiles obtained, 12 were complete single source profiles, 19 were complete profiles with two contributors, five were feasible single source profiles, and 62 were feasible profiles with two contributors[30].

### 1.5.1 Electrical Tape

Electrical tape, or vinyl tape, is used in applications that require heat resistance and insulation[37]. There are two layers that comprise electrical tape. The first layer is a plasticized polyvinyl chloride (PVC) film backing and the second layer is an elastomeric adhesive[31]. An elastomer is a material that is easily deformed, but can return to its original form when forces are removed[37]. Electrical tape backing is typically between 4.5-7.5 mils thick[37]. A variety of plasticizers may be added to the backing to impart flexibility in the plastic as well as other inorganic materials[37]. There are also a variety of ways in which the adhesive may be formulated depending on the intended use[37].

In a study by Goodpaster *et al.* 67 rolls of electrical tape from 34 brands and seven manufacturers were analyzed with scanning electron microscopy and energy dispersive spectroscopy[31]. They found the relative amounts of magnesium, aluminum, silicon, sulfur, lead, chlorine, antimony, calcium, titanium, and zinc varied greatly between brands. All 3M tapes examined in this study contained some amount of antimony, which was attributed to flame retardant, and calcium, which was attributed to filler. Premium

grade brands were also found to have higher levels of magnesium, aluminum, antimony, and calcium compared to mid-range or general grades[31].

Before electrical tape can be sold in the US, it is rated by testing laboratories such as the Underwriter's Laboratories (UL)[31]. Properties that are rated include adhesion, temperature resistance, flame resistance, and dielectric properties. The UL Standard 510 is for evaluation of Polyvinyl Chloride, Polyethylene, and Rubber Insulating Tape which are to be used for electrical insulation at not more than 600 V and 80°C in accordance with the National Electrical Code ANSI/NFPA 70[39]. All tapes evaluated by this standard are subjected to a flame test and sunlight resistance test; additional tests are required depending on the specific kind of tape[39].

In accordance with the flame test, insulating tape labeled "flame retardant" may flame no longer than 60 seconds following application of a 15 seconds test flame[39]. In accordance with the sunlight resistance test, insulating tape labeled "sunlight resistant" must retain at least 65% of its original tensile strength and elongation after being exposed to ultraviolet light from a radiation source for 100 hours[39]. One of the additional tests applied to thermoplastic tapes, such as PVC tape, is exposure to heat. In accordance with the exposure to heat test, a PVC insulating tape may not crack when flexed or be adversely affected in any other way after being exposed to temperatures of 113°C for 168 hours in a circulating-air oven[39].

### 1.5.2 Duct Tape

Duct tape, or polycoated cloth tape, consists of three layers: a backing, a fabric reinforcement, and an adhesive[37]. The polymeric backing is a carrier for the adhesive and provides color; the most common color is silver[36]. Inorganic materials may be added to the backing to help improve water repellency or tear strength[37]. The backing is typically between 1.5-4 mils thick[37]. The fabric layer, which is found between the backing and the adhesive, provides strength and bulk to the tape and affects the tearing properties[36]. The fibers are typically cotton, polyester, or a blend of the two[36; 37]. The adhesive consists of an elastomer,

which may be either natural rubber-based or synthetic, to which tackifying resins are added to make the elastomer sticky and inorganic materials are added to increase the bulk or add color[36; 37].

The physical and elemental features of the backing, fabric, and adhesive vary depending on the manufacture, raw materials, commercial end use, and specifications. As of 2005 there were over 150 duct tape references in the US produced by 4-5 manufacturers[36]. In a study conducted by Benson, the physical and elemental components of twelve samples of duct tape were examined[40]. Infrared analysis (IR) showed the adhesives were polypropylene based with carbonate and silicate additives and the backing was mainly polyethylene. The adhesive was analyzed with emission spectroscopy; aluminum, silicon, iron, titanium, and magnesium were found in all samples, while calcium and zinc were found in most, but not all[40].

# Chapter 2: Materials and Methods

## 2.1 Experimental Design

A 2x2x3 factorial design was utilized (Table 1). The factor "weeks" has two levels; one week and two weeks. The factor "tape" has two levels; duct tape and electrical tape. The factor "layers" has three levels; zero layers, one layer, and two layers. Twelve replicates were prepared for each condition, six of which were treated for one week in a humidity chamber set at 40°C and 50% RH and the other six were treated for two weeks in a humidity chamber set at 40°C and 50% RH. Six duct tape blanks and six electrical tape blanks were also prepared; three of each were treated for one week in the humidity chamber set at 40°C and 50% RH and the other three were treated for two weeks in the humidity chamber set at 40°C and 50% RH. Each of the three participants contributed saliva samples for two of the six replicates for each condition.

*Table 1: Factorial experimental design. The factor "tape" has two levels: duct tape and electrical tape. The factor "layers" has 3 levels: zero layers, one layer, and two layers. The factor "weeks" has two levels: one and two.*

| | 0 Layers, Blank | 0 Layers | 1 Layer | 2 Layers |
|---|---|---|---|---|
| **Duct Tape** | 1st Week: 3 | 1st Week: 6 | 1st Week: 6 | 1st Week: 6 |
| | 2nd Week: 3 | 2nd Week: 6 | 2nd Week: 6 | 2nd Week: 6 |
| **Electrical Tape** | 1st Week: 3 | 1st Week: 6 | 1st Week: 6 | 1st Week: 6 |
| | 2nd Week: 3 | 2nd Week: 6 | 2nd Week: 6 | 2nd Week: 6 |

## 2.2 Sample Collection

Multiple buccal swabs were collected from the three participants. Each buccal swab was then placed in a spin basket in a 1.5 mL microcentrifuge tube and centrifuged at 14000 RPM for two minutes to collect the saliva and buccal cells at the bottom of the tube. The samples from a given participant were combined into a single microcentrifuge tube. Saliva samples were stored in the refrigerator until further use.

## 2.3 Sample Preparation

Pieces of black Scotch Super 33+ Vinyl Electrical Tape were draped over a wooden frame. One-inch segments were marked on the backing of the electrical tape. The electrical tape samples were then placed in a Spectroline Select XLE-1000 UV Crosslinker with 254nm wavelength UV tubes for ten minutes, after which the samples were flipped over and treated for another ten minutes. The same process was repeated with pieces of silver Duck Basic Strength Duct Tape.

Samples were prepared in a laminar flow hood which had been sanitized with 10% bleach and UV light for ten minutes. The saliva sample from a participant was removed from the refrigerator and vortexed before dispensing 25 µl onto the center of the adhesive side of electrical tape and duct tape samples. The tube of saliva was inverted several times between dispensing samples. The saliva was allowed to air dry on the tape samples in the laminar flow hood.

The pieces of electrical tape were then cut along the one-inch markings. Samples were prepared as illustrated in Figure 1. Two pieces of electrical tape containing saliva stains were individually stored in

50 mL conical tubes to serve as the zero-layer samples, the other pieces of electrical tape were used to prepare layered samples and blanks, which were also stored individually in 50 mL conical tubes. This process was repeated for the duct tape samples. All samples corresponding to a given participant were prepared concurrently; two electrical tape blanks and two duct tape blanks were also prepared alongside the samples for a given participant. All samples were stored in the refrigerator.

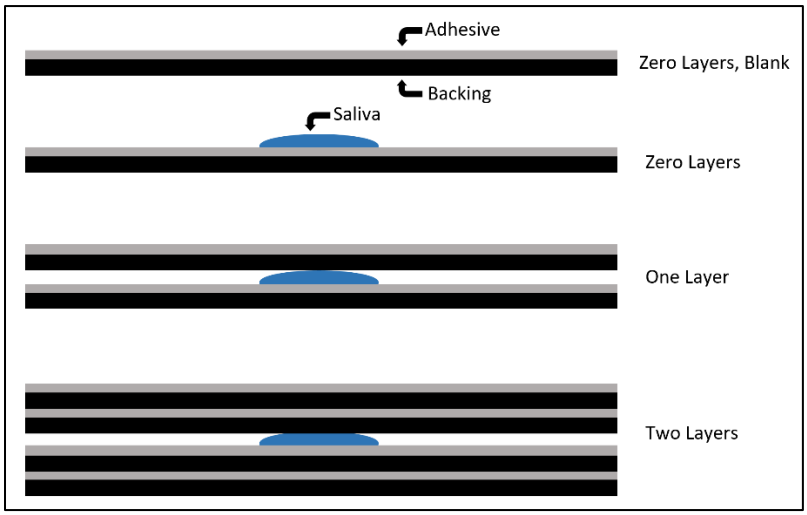


*Figure 1: Tape samples consisted of zero-layer blank samples, zero-layer samples with saliva, one-layer samples with saliva, and two-layer samples with saliva. Layers were prepared by adhering the adhesive side of one piece of tape to the backing of another piece of tape.*

## 2.4 Sample Treatment

A Caron 6105 Fingerprint Chamber set at 40°C and 50% RH was used to incubate the tape samples. All samples were placed in the chamber at the same time. Half of the samples were removed after one week, the rest of the samples were removed after two weeks (Table 1). A DHT22 AM2302 Digital Temperature and Humidity Sensor and a Raspberry Pi 1 Model B was used to log the temperature and humidity of the chamber at ten-minute increments. The average temperature for the samples treated for one week was 57.1°C with a standard deviation of 1.52°C and a range of 36.1-58.6°C. The average %RH for the samples treated for one week was 60.4 with a standard deviation 2.97 and a range of 27.1-64.4. The average temperature for the samples treated for two weeks was 56.8°C with a standard deviation of 1.17°C and a range of 36.1-58.6°C. The average %RH for the samples treated for two weeks 60.6 with a standard

deviation of 2.13 and a range of 27.1-65.0. After samples were removed from the chamber, they were stored in 50 mL conical tubes and placed in the refrigerator.

## 2.5 Sample Processing

A reference profile was obtained for each participant; 50 µL of saliva was removed from the tube containing the samples for a given participant. Extraction was performed with the QIAGEN QIAamp DNA Investigator Kit, following the procedure for isolation of total DNA from small volumes of blood or saliva. Samples were eluted from the QIAamp MiniElute columns with 60 µl of Buffer ATE. Extracted samples were stored in the refrigerator.

Tape samples were swabbed and extracted in three batches; each batch corresponding to one of the participants. Layered samples were pulled apart using forceps. A Puritan Sterile Cotton Tipped Applicator with 100 µl of 100% ethanol was used to swab the samples. For layered samples, the backing and adhesive sides that were in contact with the saliva stain were both swabbed with the same swab. The cotton swab was cut and stored in a 2 mL microcentrifuge tube and placed in the refrigerator. Tape samples were stored in 50 mL conical tubes and placed in the freezer.

Swabs were extracted using the QIAGEN QIAamp DNA Investigator Kit, following the procedure for isolation of total DNA from surface and buccal swabs. As per the manufacturer's suggestion, 1 µg of carrier RNA was added to each sample during the lysis step to enhance binding of DNA to the QIAamp MiniElute column in subsequent steps. Swabs were also centrifuged in a spin basket for two minutes at 14000 RPM to harvest all of the lysate from the swabs before proceeding to the purification step. Samples were eluted from the QIAamp MiniElute columns with 60 µl of Buffer ATE. Extracted samples were stored in the refrigerator.

Quantifiler Trio™ DNA Quantification Kit (Applied Biosystems) was used for human DNA quantification on a 7500 Real-Time PCR System following the manufacturer's suggested protocol. HID Real-Time PCR Analysis Software v1.3 was used for quantification analysis. Samples containing less than

0.125 ng/15µL were concentrated by centrifuging in an Eppendorf Vacufuge until they were dry and then adding 17 µl of nuclease free water; they were then re-quantified.

GlobalFiler™ PCR Amplification Kit (Applied Biosystems) was used for amplification following the manufacturer's suggested protocol. Where possible, 1 ng of DNA was amplified, otherwise the maximum volume of extracted sample (15 µl) was amplified. Samples were amplified using a GeneAmp PCR System 9700 and capillary electrophoresis was performed using a 3500 Series Genetic Analyzer. Capillary electrophoresis was performed using 36 cm long capillaries, POP4 polymer, 60°C oven temperature, a run voltage of 13 kVolts, prerun voltage of 15 kVolts, injection voltage of 1.2 kVolts, run time of 1550 sec, prerun time of 180 sec, injection time of 15 sec, and data delay of 1 sec. GeneMapper® ID-X v1.4 was used for STR analysis.

## 2.6 Profile Slope

The profile slope of each sample was calculated using Equation 2, where $H_l$ is the sum of peak heights at locus $l$ corresponding with the reference profile, $\bar{s}_l$ is the average fragment size (base pair) at locus $l$ corresponding with the reference profile, and $m$ and $b$ are the slope and y-intercept obtained for each sample using linear regression[24]. The STR profile of each sample was analyzed at a threshold of one RFU and all peaks corresponding with the respective reference profile were called. In the case where an allele in the reference profile was missing from the sample profile, the base pair size of the allele in the reference profile was used and a value of zero RFUs was assigned.

$$m = \frac{ln(H_l) - b}{\bar{s}_l} \qquad (2)$$

### 2.6.1 PROVEDIt Database Profile Slopes

Raw .fsa data files were obtained from the PROVEDIt Database[24]. All datafiles downloaded from the PROVEDIt Database were for samples which were quantified with the Quantifiler™ Trio Quantification kit, amplified with the GlobalFiler™ PCR Amplification kit, and analyzed on a 3500 Genetic Analyzer. The raw

.fsa files were downloaded and analyzed with GeneMapper® ID-X v1.4 at a threshold of ten RFUs. All sample profiles were compared to the known genotypes, which were also made available, and all peaks corresponding with the respective reference profile were called.

The profile slope was calculated using Equation 1 for profiles in which all alleles corresponding with the reference profile were called. There were 175 samples treated with UV light and 183 samples treated with DNase I that were analyzed. Of the samples treated with UV light, the quantity of DNA input for amplification ranged from 0.0313-0.504 ng and the DI values ranged from 0.7-58. Of the samples treated with DNase I, the quantity of DNA input for amplification ranged from 0.0156-0.7 ng and the DI values ranged from 0.05-14.

## 2.7 Model for Estimating Profile Slope

Three models for estimating the slope of a profile were calculated based on samples from the PROVEDIt Database; one based on the DNA samples treated with UV to induce degradation, one based on DNA samples treated with DNase I to induce degradation, and one combining DNA samples treated with both methods to induce degradation.

Multiple linear regression was used to calculate the relationship between the profile slope, ln(DI), and the quantity of DNA input for amplification (ng) using Equation 3, where $\beta_0$ is the y-intercept, $\beta_1$-$\beta_3$ are the slopes for their respective variables, $x_D$ is the ln(DI), $x_A$ is the quantity of DNA input for amplification (ng), and $x_{DA}$ is the interaction between the ln(DI) and the quantity of DNA input for amplification (ng).

$$Profile\ Slope = \beta_0 + \beta_1 x_D + \beta_2 x_A + \beta_3 x_{DA} \quad (3)$$

Each of the models were tested with DNA samples that were treated with heat and humidity in this study. In order for a sample to be used to test a model, it had to fall within the DI range and amplification input range (ng) of the samples that were used to create the model. There were 41 samples

used to test the model that was created with the DNase I treated samples, 41 samples used to test the model that was created with the UV treated samples, and 47 samples used to test the model that was created with the combination of DNase I and UV treated samples.

# Chapter 3: Results

The concentration of the small autosomal quantification target in the reference sample of one of the participants was about 74% lower than the concentration of the other two reference samples. The mean concentration for samples corresponding with this participant were consistently lower for samples that were treated for one and two weeks. The concentration of the small autosomal target for each participant are summarized in Table 2. Of the 12 blank tape samples, human DNA was detected in 4 of them; the concentration of DNA in these samples ranged from 0.0001-0.0003 ng/µl. Of the 24 treated samples containing female DNA, male human DNA was detected in 4 of them; the concentration of male DNA ranged from 0.0001-0.0005 ng/µl.

*Table 2: Concentration of the small autosomal quantification target for each participant's reference sample and the mean concentration for the small autosomal quantification target of treated samples corresponding to each participant.*

| Participant ID | Reference Sample Concentration (ng/µl) | Week 1 | | Week 2 | |
|---|---|---|---|---|---|
| | | Small Autosomal Concentration (ng/µl) | SD | Small Autosomal Concentration (ng/µl) | SD |
| 144 | 0.6879 | 0.0129 | 0.0078 | 0.0093 | 0.0020 |
| 331 | 2.6318 | 0.0556 | 0.0290 | 0.0270 | 0.0200 |
| 565 | 2.6271 | 0.0891 | 0.0389 | 0.0418 | 0.0247 |

## 3.1 Degradation Index

The DI value for each sample was calculated with the HID Real-Time PCR Analysis Software v1.3 with a built-in function which utilizes Equation 1. The degradation index could not be calculated for one of the samples because the large autosomal target was not detected. Summary statistics were calculated (Table 3) and boxplots were made (Figure 2) for each treatment group using R version 3.4.2[41].

*Table 3: Summary statistics for the DI of each treatment group.*

| Weeks | Tape | Layers | Sample Size | Mean DI | SD | %RSD |
|---|---|---|---|---|---|---|
| 1 | Electrical | 0 | 6 | 4.537 | 1.190 | 26.24 |
| 1 | Electrical | 1 | 6 | 8.187 | 6.793 | 82.98 |
| 1 | Electrical | 2 | 6 | 5.988 | 2.759 | 46.08 |
| 1 | Duct | 0 | 6 | 4.728 | 1.605 | 33.93 |
| 1 | Duct | 1 | 6 | 6.678 | 2.586 | 38.72 |
| 1 | Duct | 2 | 6 | 6.448 | 2.980 | 46.22 |
| 2 | Electrical | 0 | 6 | 8.643 | 2.557 | 29.58 |
| 2 | Electrical | 1 | 6 | 8.633 | 2.007 | 23.25 |
| 2 | Electrical | 2 | 6 | 7.683 | 3.276 | 42.64 |
| 2 | Duct | 0 | 6 | 8.620 | 1.975 | 22.91 |
| 2 | Duct | 1 | 6 | 10.94 | 4.479 | 40.93 |
| 2 | Duct | 2 | 5 | 17.96 | 6.860 | 38.20 |

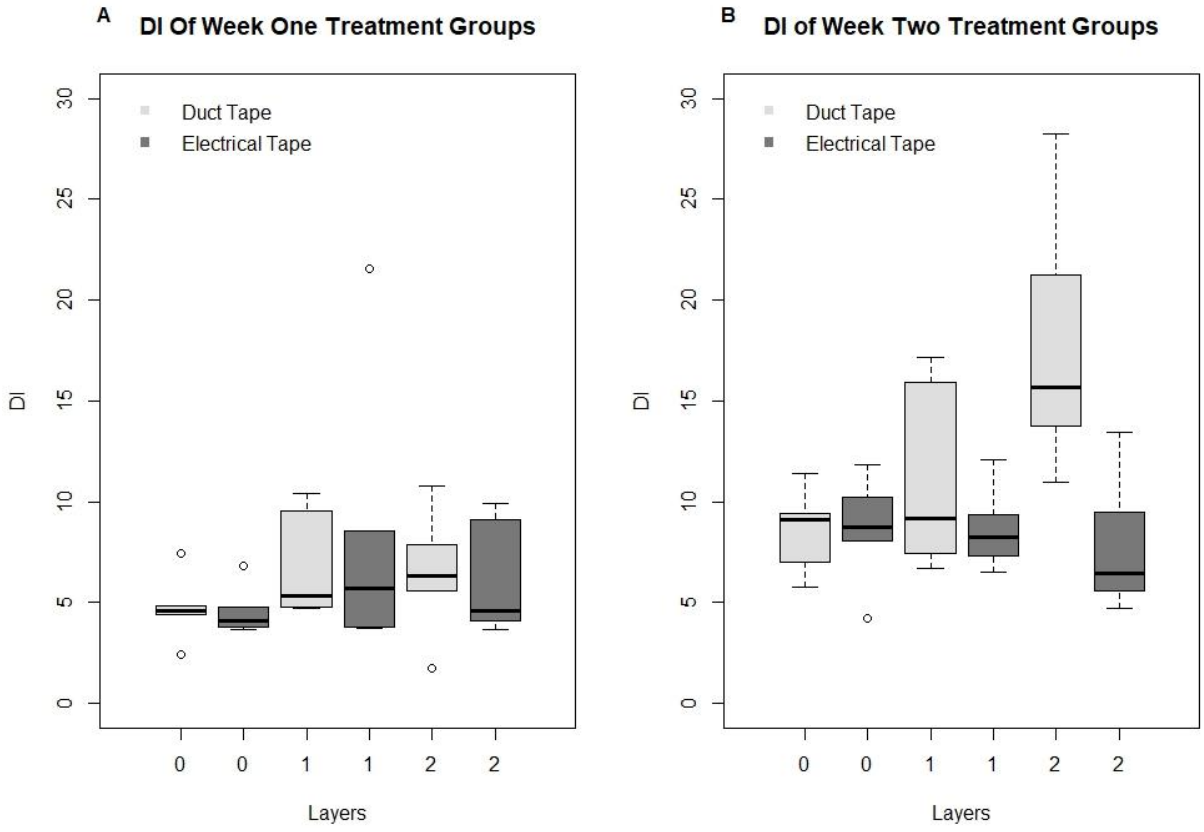


*Figure 2: (A) DI of samples treated for one week. (B) DI of samples treated for two weeks.*

Log likelihood ratio tests were performed on linear mixed-effects models to determine the significance of fixed variables on the DI while taking into account the repeated measures from each participant. The lme4 package in R was used to fit all linear mixed effects models[42]. Fixed variables include the type of tape, the number of layers, the number of weeks, and the interaction between these variables. The random variable accounts for the participant. This is necessary because the quantity of DNA and composition of saliva from each participant will vary and taking multiple measurements from the saliva sample from each participant will result non-independent data. Equation 4 is the full model, where $\beta_0$ is the intercept, $\beta_1$-$\beta_7$ are the slopes for their respective fixed variables, $\gamma_1$ is the slope for the random variable, $x_T$ is the variable for the type of tape, $x_L$ is the variable for the number of layers of tape, $x_W$ is the variable for the number of weeks, $x_{TL}$ is the interaction between the type of tape and number of layers, $x_{TW}$ is the interaction between the type of tape and number of weeks, $x_{LW}$ is the interaction between number of layers and number of weeks, $x_{TLW}$ is the interaction between type of tape, number of layers, and number of weeks, $x_p$ is the random variable for the participant, and $\varepsilon$ is the random error.

To determine the significance of the three-way interaction between tape, layers, and weeks on the DI, a log likelihood ratio test was performed between a reduced model (Equation 5, $H_0$), which dropped the variable $x_{TLW}$, and the full model (Equation 4, $H_1$). The fit of the full model was found to be significantly better than the fit of the reduced model ($\chi^2(2)=7.4596$, $p=0.024$), indicating there is a significant three-way interaction between the fixed variables. The log likelihood ratio test results are summarized in Table 4.

$$DI = \beta_0 + \beta_1 x_T + \beta_2 x_L + \beta_3 x_W + \beta_4 x_{TL} + \beta_5 x_{TW} + \beta_6 x_{LW} + \beta_7 x_{TLW} + \gamma_1 x_p + \varepsilon \quad (4)$$

$$DI = \beta_0 + \beta_1 x_T + \beta_2 x_L + \beta_3 x_W + \beta_4 x_{TL} + \beta_5 x_{TW} + \beta_6 x_{LW} + \gamma_1 x_p + \varepsilon \quad (5)$$

The treatment groups were split up into two main groups, those that were treated for one week and those that were treated for two weeks, in order to assess whether there is a significant interaction

between tape and layers on the DI for the different treatment lengths. A log likelihood ratio test was performed between Equation 6, the full model ($H_1$) and Equation 7 the reduced model ($H_0$) for both treatment lengths. The log likelihood ratio test for samples treated for one week showed there was no significant interaction between tape and layers ($\chi^2(2)=0.784$, p = 0.6757). The log likelihood ratio test for samples treated for two weeks showed there was a significant interaction between tape and layers ($\chi^2(2)=14.385$, p=0.0007522).

$$DI = \beta_0 + \beta_1 x_T + \beta_2 x_L + \beta_3 x_{TL} + \gamma_1 x_p + \varepsilon \quad (6)$$

$$DI = \beta_0 + \beta_1 x_T + \beta_2 x_L + \gamma_1 x_p + \varepsilon \quad (7)$$

*Table 4: Likelihood ratio and Chi Square test for model comparisons to determine variable significance.*

| Model | df | Log Likelihood | Deviance | Chi Squared | df | p-Value |
|---|---|---|---|---|---|---|
| **$H_0$:** Equation 5 | 12 | -183.49 | 366.98 | 7.4596 | 2 | **0.024** |
| **$H_1$:** Equation 4 | 14 | -187.24 | 374.44 | | | |
| **Week 1 $H_0$:** Equation 7 | 6 | -91.977 | 183.95 | 0.784 | 2 | 0.6757 |
| **Week 1 $H_1$:** Equation 6 | 8 | -91.585 | 183.17 | | | |
| **Week 2 $H_0$:** Equation 7 | 6 | -97.529 | 195.06 | 14.385 | 2 | **0.0007522** |
| **Week 2 $H_1$:** Equation 6 | 8 | -90.336 | 180.67 | | | |

There was not a significant interaction between tape and layers for the group treated for one week, so the main effects were assessed for significant differences. There was a significant interaction between tape and layers for the group treated for two weeks, so the simple effects were assessed for significant differences. The emmeans package in R was used to calculate the P-value for the difference between the estimated marginal means for the main effects and the simple effects using the Kenward-Roger method to calculate degrees of freedom and Tukey's test[43]. Table 5 summarizes the P-values obtained for each contrast of the main effects and simple effects.

*Table 5: P-Values calculated from the difference in the estimated marginal means of the DI for contrasts between main effects for samples treated for one week and simple effects contrasts for samples treated for two weeks.*

| Contrast | Estimated Difference | SE | df | t-Value | p-Value |
|---|---|---|---|---|---|
| Week 1: One Layer - Two Layers | 1.2142 | 1.2719 | 30 | 0.955 | 0.6107 |
| Week 1: One Layer- Zero Layers | 2.8000 | 1.2719 | 30 | 2.201 | 0.0872 |
| Week 1: Two Layers - Zero Layers | 1.5858 | 1.2719 | 30 | 1.247 | 0.4357 |
| Week 1: Electrical Tape - Duct Tape | 0.2856 | 1.0385 | 30 | 0.275 | 0.7852 |
| Week 2: Two Layers Duct Tape - One Layer Duct Tape | 6.9594 | 1.9800 | 27.06 | 3.515 | **0.0176** |
| Week 2: One Layer Duct Tape - Zero Layers Duct Tape | 2.3233 | 1.8806 | 27.00 | 1.235 | 0.8158 |
| Week 2: One Layer Duct Tape - One Layer Electrical Tape | 2.3100 | 1.8806 | 27.00 | 1.228 | 0.8193 |
| Week 2: Two Layers Duct Tape - Zero Layers Duct Tape | 9.2827 | 1.9800 | 27.06 | 4.688 | **0.0009** |
| Week 2: Two Layers Duct Tape - Two Layers Electrical Tape | 10.2194 | 1.9800 | 27.06 | 5.161 | **0.0003** |
| Week 2: Zero Layers Electrical Tape - Zero Layers Duct Tape | 0.0233 | 1.8806 | 27.00 | 0.012 | 1.0000 |
| Week 2: One Layer Electrical Tape - Two Layers Electrical Tape | 0.9500 | 1.8806 | 27.00 | 0.505 | 0.9955 |
| Week 2: Zero Layers Electrical Tape - One Layer Electrical tape | 0.0100 | 1.8806 | 27.00 | 0.005 | 1.000 |
| Week 2: Zero Layers Electrical Tape - Two Layers Electrical Tape | 0.9600 | 1.8806 | 27.00 | 0.510 | 0.9953 |

Of the samples treated for one week, there were no significant difference in DI due to the main effects (Table 5). Of the samples treated for two weeks, there was found to be a significant difference in DI when comparing samples with one layer of duct tape to samples with two layers of duct tape ($t(27.06)=3.515$, $p=0.0176$), a significant difference in DI when comparing samples with two layers of duct tape to samples with zero layers of duct tape ($t(27.06)=4.688$, $p=0.0009$), and a significant difference in DI when comparing samples with two layers of duct tape to samples with two layers of electrical tape ($t(27.06)= 5.161$, $p=0.0003$). The mean DI of the simple effects for samples treated for two weeks are summarized in Table 6.

*Table 6: Mean DI of simple effects for samples treated for two weeks.*

| | Zero Layers | One Layer | Two Layers |
|---|---|---|---|
| **Duct Tape** | **8.620** | **10.94** | **17.96** |
| **Electrical Tape** | 8.643 | 8.633 | **7.683** |

The mean DI of samples treated for two weeks with two layers of duct tape was 17.96, which is significantly greater than the DI of samples treated for two weeks with one layer of duct tape, which had a mean DI of 10.94 (t(27.06)=3.515, p=0.0176). The mean DI of samples treated for two weeks with two layers of duct tape was also significantly greater than the mean DI of samples treated for two weeks with zero layers of duct tape, which was 8.620 (t(27.06)=4.688, p=0.0009). The mean DI of samples treated for two weeks with two layers of duct tape was also significantly greater than the mean DI of samples treated for two weeks with two layers of electrical tape, which had a mean DI of 7.683 (t(27.06)= 5.161, p=0.0003).

## 3.2 Profile Slope

Summary statistics for profile slopes were calculated (Table 7) and boxplots were made (Figure 3) for each treatment group using R version 3.4.2[41]. The slope was not calculated for one of the samples due to contamination by the allelic ladder. There were extraneous peaks in some of the electropherograms, which were attributed to shared resources and facilities of the laboratory.

*Table 7: Summary statistics for the profile slope of each treatment group*

| Weeks | Tape | Layers | Sample Size | Mean Slope | SD | %RSD |
|---|---|---|---|---|---|---|
| 1 | Electrical | 0 | 6 | -0.0097 | 0.0009 | 9.212 |
| 1 | Electrical | 1 | 5 | -0.0101 | 0.0039 | 38.65 |
| 1 | Electrical | 2 | 6 | -0.0104 | 0.0023 | 22.30 |
| 1 | Duct | 0 | 6 | -0.0086 | 0.0012 | 14.25 |
| 1 | Duct | 1 | 6 | -0.0101 | 0.0015 | 14.42 |
| 1 | Duct | 2 | 6 | -0.0087 | 0.0013 | 14.56 |
| 2 | Electrical | 0 | 6 | -0.0129 | 0.0008 | 6.273 |
| 2 | Electrical | 1 | 6 | -0.0115 | 0.0024 | 20.64 |
| 2 | Electrical | 2 | 6 | -0.0106 | 0.0016 | 14.73 |
| 2 | Duct | 0 | 6 | -0.0122 | 0.0017 | 14.22 |
| 2 | Duct | 1 | 6 | -0.0118 | 0.0018 | 15.13 |
| 2 | Duct | 2 | 6 | -0.0120 | 0.0023 | 19.18 |

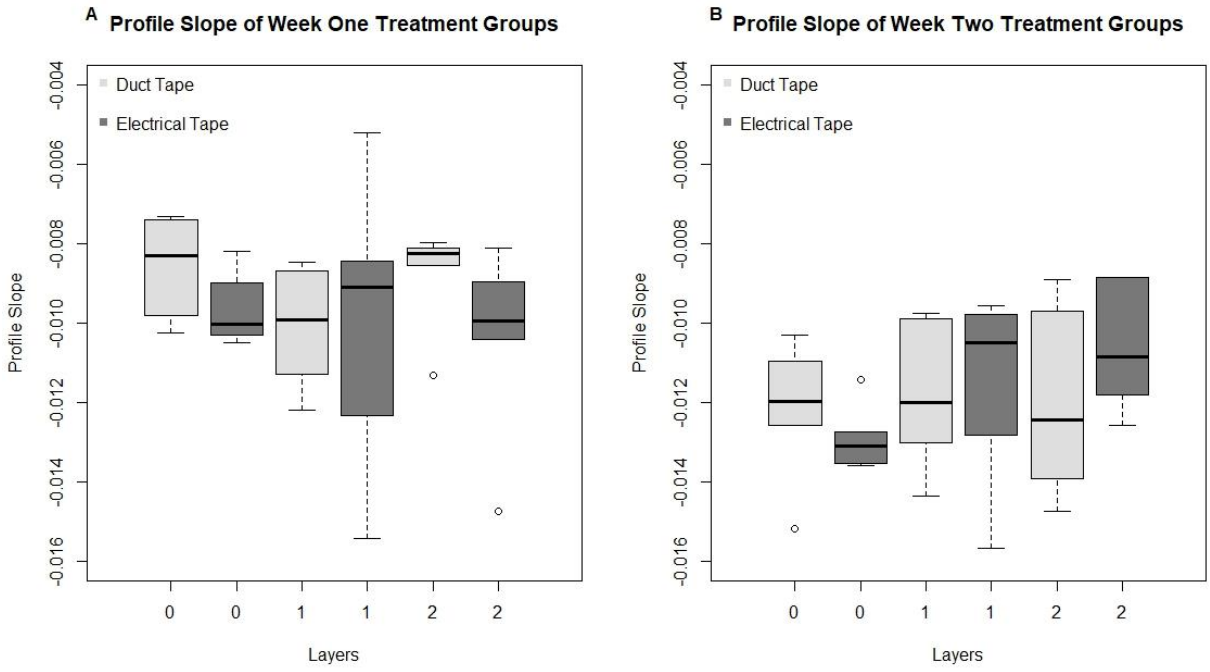


*Figure 3: (A) Profiles slopes for samples treated for one week. (B) Profile slopes for samples treated for two weeks.*

Log likelihood ratio tests were performed on linear mixed-effects models to determine the significance of fixed variables on the profile slope while taking into account the repeated measures from each participant. To determine the significance of the three-way interaction between tape, layers, and weeks on the profile slope, a log likelihood ratio test was performed between a reduced model (Equation 8, $H_0$), which dropped the variable $x_{TLW}$, and the full model (Equation 9, $H_1$). The fit of the full model was not found to be significantly better than the fit of the reduced model ($\chi^2(2)=2.2875$, $p=0.3186$), indicating there is not a significant three-way interaction between the fixed variables. The log likelihood ratio test results are summarized in Table 8.

To determine the significant of the two-way interactions on the profile slope, log likelihood ratio tests were performed between Equation 8 and three reduced equations in which the interaction between layers and weeks was removed (Equation 10), the interaction between tape and weeks was removed (Equation 11), and the interaction between tape and layers was removed (Equation 12). None of the two-

way interactions were found to be significant ($\chi^2(2)=3.949$, p=0.1388; $\chi^2(2)=2.2803$, p=0.131; $\chi^2(2)=1.2269$, p=0.5415 respectively). The log likelihood ratio tests are summarized in Table 8.

$$Profile\ Slope = \beta_0 + \beta_1 x_T + \beta_2 x_L + \beta_3 x_W + \beta_4 x_{TL} + \beta_5 x_{TW} + \beta_6 x_{LW} + \gamma_1 x_p + \varepsilon \quad (8)$$

$$Profile\ Slope = \beta_0 + \beta_1 x_T + \beta_2 x_L + \beta_3 x_W + \beta_4 x_{TL} + \beta_5 x_{TW} + \beta_6 x_{LW} + \beta_7 x_{TLW} + \gamma_1 x_p + \varepsilon \quad (9)$$

$$Profile\ Slope = \beta_0 + \beta_1 x_T + \beta_2 x_L + \beta_3 x_W + \beta_4 x_{TL} + \beta_5 x_{TW} + \gamma_1 x_p + \varepsilon \quad (10)$$

$$Profile\ Slope = \beta_0 + \beta_1 x_T + \beta_2 x_L + \beta_3 x_W + \beta_4 x_{TL} + \beta_5 x_{LW} + \gamma_1 x_p + \varepsilon \quad (11)$$

$$Profile\ Slope = \beta_0 + \beta_1 x_T + \beta_2 x_L + \beta_3 x_W + \beta_4 x_{TW} + \beta_5 x_{LW} + \gamma_1 x_p + \varepsilon \quad (12)$$

*Table 8: Likelihood ratio and Chi Square test for model comparisons to determine variable significance*

| Model | df | Log Likelihood | Deviance | Chi Squared | df | p-Value |
|---|---|---|---|---|---|---|
| **$H_0$:** Equation 8 | 12 | 349.04 | -698.07 | 2.2875 | 2 | 0.3186 |
| **$H_1$:** Equation 9 | 14 | 350.18 | -700.36 | | | |
| **$H_0$:** Equation 10 | 10 | 347.06 | -694.12 | 3.949 | 2 | 0.1388 |
| **$H_1$:** Equation 8 | 12 | 349.04 | -698.07 | | | |
| **$H_0$:** Equation 11 | 11 | 347.90 | -695.79 | 2.2803 | 1 | 0.131 |
| **$H_1$:** Equation 8 | 12 | 349.04 | -698.07 | | | |
| **$H_0$:** Equation 12 | 10 | 348.42 | -696.85 | 1.2269 | 2 | 0.5415 |
| **$H_1$:** Equation 8 | 12 | 349.04 | -698.07 | | | |

There were no significant two-way interactions, so the main effects were assessed for significant differences. The emmeans package in R was used to calculate the P-value for the difference between the estimated marginal means for the main effects using the Kenward-Roger method to calculate degrees of freedom and Tukey's test[43]. Table 9 summarizes the P-values obtained for each contrast of the main effects. There was found to be a significant difference in profile slope between samples treated for one week and samples treated for two weeks ($t(64.02)=5.022$, $p<0.0001$). There were no significant differences found in profile slope between the number of layers or types of tape (Table 9). The mean profile slopes for samples treated for one week and samples treated for two weeks are summarized in Table 10. The mean profile slope of samples treated for two weeks was -0.0118, which is significantly

greater than the mean profile slope of samples treated for one week, which was -0.0096 (t(64.02)=5.022, p<0.0001).

*Table 9: P-Values calculated from the difference in the estimated marginal means of the profile slope for contrasts between main effects.*

| Contrast | Estimated Difference | SE | df | t-Value | p-Value |
|---|---|---|---|---|---|
| Two Layers - One Layer | 0.0005 | 0.0006 | 64.03 | 0.844 | 0.6775 |
| Zero Layers - One Layer | 0.0001 | 0.0006 | 64.03 | 0.140 | 0.9892 |
| Two Layers-Zero Layers | 0.0004 | 0.0006 | 64.00 | 0.711 | 0.7577 |
| Duct Tape-Electrical Tape | 0.0003 | 0.0005 | 64.02 | 0.687 | 0.4944 |
| One Week-Two Weeks | 0.0023 | 0.0005 | 64.02 | 5.022 | **<0.0001** |

*Table 10: Estimated marginal mean profile slopes for samples treated for one week and two weeks.*

| | | Zero Layers | One Layer | Two Layers | Estimated Marginal Mean |
|---|---|---|---|---|---|
| **One Week** | **Duct Tape** | -0.0086 | -0.0101 | -0.0087 | **-0.0096** |
| | **Electrical Tape** | -0.0097 | -0.0101 | -0.0104 | |
| **Two Weeks** | **Duct Tape** | -0.0122 | -0.0118 | -0.0120 | **-0.0118** |
| | **Electrical Tape** | -0.0129 | -0.0115 | -0.0106 | |

## 3.3 Model for Estimating Profile Slope

The multiple linear regression results for the DNase I treated samples are summarized in Table 11; these values were used to construct the DNase I model (Equation 13). The adjusted R-squared value for the DNase I model (Equation 13) is 0.711, with F(179)=150.3 and p<0.0000. The root mean square error (RMSE) of the estimated profile slope values and the actual values for the heat and humidity treated samples was calculated to be 0.0023. The DI and profile slope of the DNase I treated samples, the heat and humidity treated samples, and the estimated values for the heat and humidity samples are plotted in Figure 4A.

*Table 11: Multiple linear regression of DNase degraded samples.*

| | Estimate | Std. Error | t-Value | p-Value |
|---|---|---|---|---|
| **Intercept** | -0.0045 | 0.0002 | -21.158 | <0.00000 |
| **ln(DI)** | -0.0030 | 0.0003 | -9.481 | <0.00000 |
| **Amp Input** | 0.0018 | 0.0008 | 2.119 | 0.03549 |
| **ln(DI)*Amp Input** | -0.0036 | 0.0011 | -3.192 | 0.00167 |

$$Profile\ Slope = -0.0045 - 0.0030x_D + 0.0018x_A - 0.0036x_{DA} \quad (13)$$

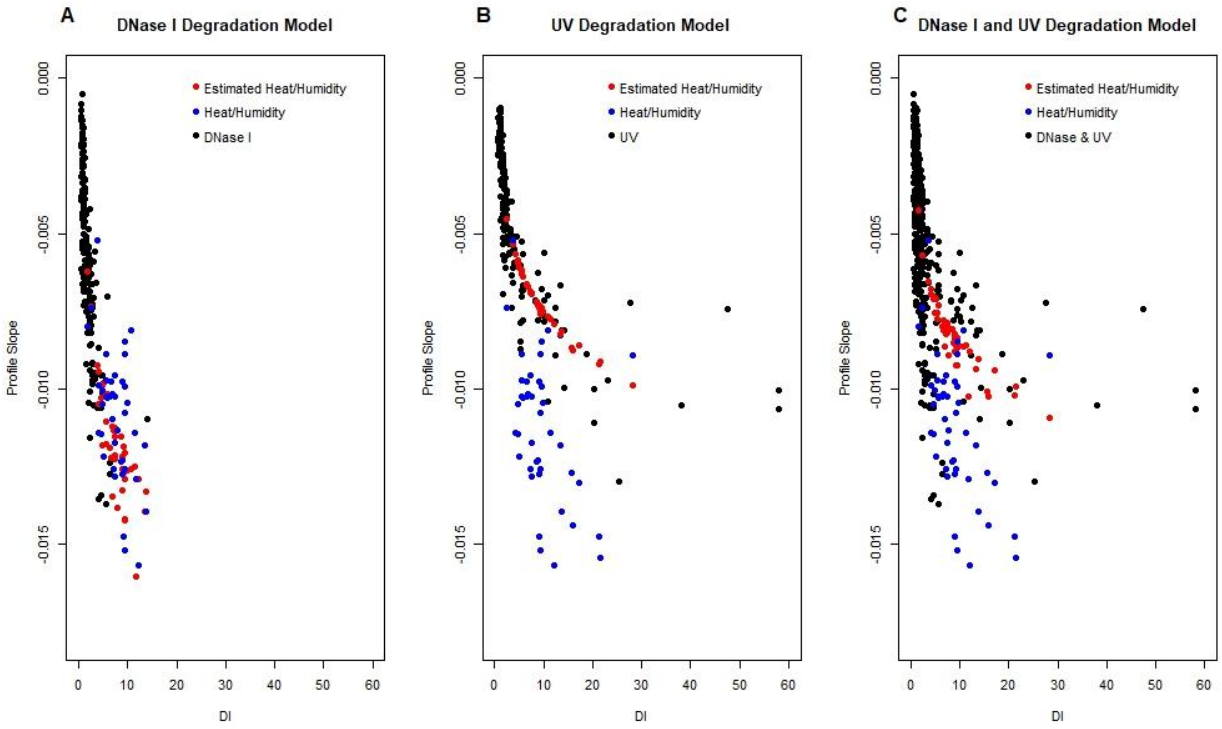


*Figure 4: (A) Profile slope versus DI for DNase I treated samples (n=183, black), heat/humidity treated samples (n=41, blue), and the estimated profile slope of the heat/humidity treated samples based on the DNase I degradation model (red). (B) Profile slope versus DI for UV treated samples (n=175, black), heat/humidity treated samples (n=41, blue), and the estimated profile slope of the heat/humidity treated samples using the UV degradation model (red). (C) Profile slope versus DI for the combined DNase I and UV treated samples (n=358, black), heat/humidity treated samples (n=47, blue), and the estimated profile slope based of the heat/humidity treated samples based on the combined DNase I and UV degradation model (red).*

The multiple linear regression results for the UV treated samples are summarized in Table 12; these values were used to construct the UV model (Equation 14). The adjusted R-squared value of the UV model (Equation 14) is 0.8326 with F(171)= 289.5 and p<0.0000. The RMSE of the estimated profile slope values and the actual values for the heat and humidity treated samples was calculated to be 0.0045. The

DI and profile slope of the UV treated samples, the heat and humidity treated samples, and the estimated values for the heat and humidity samples are plotted in Figure 4B.

*Table 12 Multiple linear regression of UV degraded samples.*

| | Estimated | Std. Error | t-Value | p-Value |
|---|---|---|---|---|
| **Intercept** | -0.0030 | 0.0002 | -12.317 | <0.000000 |
| **ln(DI)** | -0.0019 | 0.0002 | -10.908 | <0.000000 |
| **Amp Input** | 0.0026 | 0.0007 | 3.925 | 0.000126 |
| **ln(DI)*Amp Input** | -0.0014 | 0.0005 | -2.646 | 0.008914 |

$$Profile\ Slope = -0.0030 - 0.0019x_D + 0.0026x_A - 0.0014x_{DA} \quad (14)$$

The multiple linear regression results for the combined DNase I and UV treated samples are summarized in Table 13; these values were used to construct the DNase and UV model (Equation 15). The adjusted R-squared value of the DNase I and UV model (Equation 15) is 0.5142 with F(354)= 126.9 and p<0.0000. The RMSE of the estimated profile slope values and the actual values for the heat and humidity treated samples was calculated to be 0.0035. The DI and profile slope of the DNase I and UV treated samples, the heat and humidity treated samples, and the estimated values for the heat and humidity samples are plotted in Figure 4C.

*Table 13: Multiple linear regression of UV and DNase degraded samples.*

| | Estimated | Std. Error | t-Value | p-Value |
|---|---|---|---|---|
| **Intercept** | -0.0044 | 0.0002 | -19.502 | <0.00000 |
| **ln(DI)** | -0.0016 | 0.0002 | -7.131 | <0.00000 |
| **Amp Input** | 0.0033 | 0.0007 | 4.533 | <0.00000 |
| **lnDI)*Amp Input** | -0.0027 | 0.0007 | -3.548 | 0.00044 |

$$Profile\ Slope = -0.0044 - 0.0016\ x_D + 0.0033x_A - 0.0027\ x_{DA} \quad (15)$$

# Chapter 4: Discussion

When assessing the level of degradation using the degradation index, a significant three-way interaction was found between the type of tape, the number of layers of tape, and the treatment length. The effect of duct tape on the degradation index depended on the number of layers of tape and the treatment length; there was a greater degradation index when more layers of duct tape were used and the samples were treated for two weeks. The effect of the number of layers on the degradation index depended on the type of tape and the treatment length; there was a greater degradation index for duct tape samples compared to electrical tape when the samples were treated for two weeks and two layers of tape were used.

When assessing the level of degradation using the calculated profile slope, there were no significant interactions present between the type of tape, number of layers of tape, or treatment length although the profile slope became less negative with increasing layers of electrical tape when treated for two weeks. The profile slope was only significantly affected by the treatment length; samples that were treated for two weeks had a more negative slope compared to samples that were treated for one week.

Based on the degradation index results, when a sample of duct tape is received, DNA should be swabbed for on an outer layer rather than pulling all of the layers apart and swabbing from within them to obtain a less degraded sample. Furthermore, DNA obtained from electrical tape may be less degraded compared to DNA obtained from duct tape. However, these results were not confirmed when assessing the level of degradation based on the profile slope. Based on the results of the profile slope, neither the type of tape nor the number of layers had a significant effect on the degradation of the DNA sample. The one factor that influenced both the degradation index and the profile slope was the treatment length. Further research is needed to determine whether the effect of the type of tape or number of layers on profile slope is significant at a longer treatment period. Depositing a controlled quantity of DNA on each tape sample or increasing the number of replicates may yield more consistent results among the samples

treated with a specific set of conditions, which would aid in the determination of whether an interaction is present.

Differences in the degradation of DNA obtained from duct tape versus electrical tape could be due to differences in the physical properties of the tape. Electrical tape has a PVC backing and is generally thicker than duct tape, which is a polycoated cloth tape[31; 37]. While the duct tape backing may have water repellent characteristics, this wouldn't necessarily prevent the fabric layer from soaking in moisture from the edges of the tape, whereas electrical tape is made from PVC which wouldn't soak up moisture. The elemental characteristics of the tape may also influence DNA degradation. Both Ca and Mg are commonly found in duct tape and electrical tape[31; 40]. The activity of DNase I is dependent on $Ca^{2+}$ and $Mg^{2+}$ [14; 15]. Further investigation into the presence of $Ca^{2+}$ and $Mg^{2+}$ or the difference in their proportions may indicate whether one type of tape is more favorable for the activity of DNase I compared to the other.

Of the metrics utilized to assess DNA degradation, the profile slope is of greater interest. The degradation index is an indicator of what the level of degradation may be in the profile, whereas the profile slope is a measure of the degradation observed in the profile. The profile slope has been shown to become more negative as the degradation index increases[23; 24]. Therefore, the degradation index could theoretically be used to predict the slope of a DNA profile, however, when comparing the degradation index values and profile slope in this study, the results were not consistent.

It has been observed that STR profiles may have comparable allele counts despite differences in their degradation index values due to differences in the amount of DNA which is amplified[18]. This may be partially due to stochastic sampling of the Quantifiler Trio™ targets; the concentration of one of the targets is more likely to be misrepresented in a sample with a small concentration of human DNA, thereby misrepresenting the level of degradation in the sample. Consequently, two samples with relatively the same degradation index, but very different concentrations, may have very different profile slopes. Therefore, the amount of DNA available for PCR amplification and the degradation index should be taken

into consideration when evaluating the level of degradation in a DNA sample and estimating the slope of an STR profile.

The development of a model which takes into account the degradation index as well as the amount of DNA amplified may provide an estimation of the expected slope for an STR profile. Three models were created using multiple linear regression in an attempt to better understand this relationship. The model based on DNase I treated samples estimated the profile slopes of the samples collected in this study with a RMSE of 0.0023, whereas the model based on UV treated samples estimated the profile slopes with a RMSE of 0.0045. This indicates that DNA samples degraded with heat and humidity more closely resemble samples degraded with DNase compared to UV.

It has already been suggested that the relationship between the degradation index and slope of a profile depends on the method by which the DNA was degraded[16; 24]. However, it is unlikely that the source of degradation would be known in a forensic casework sample, so a third model was created which combined the DNase I and UV treated samples; this model estimated the profile slopes with a RMSE of 0.0035, however the $R^2$ value was only 0.5142. This suggests that a model should be developed using samples which have been treated with a combination of environmental factors; for example, mock casework samples which have been left outside and exposed to various temperatures, levels of humidity, and UV light. A model such as this might be more applicable to DNA samples regardless of the method by which they were degraded.

Another factor to consider is the relationship between the size of the Quantifiler™ Trio targets and the GlobalFiler™ STR loci. The small and large autosomal targets in the Quantifiler™ Trio kit are 80 base pairs and 214 base pairs long, respectively[18]. The length of these targets corresponds with the length of the miniSTR loci in the GlobalFiler™ PCR Amplification kit. There are 10 miniSTR loci out of the 24 STR loci in the GlobalFiler™ kit which fall entirely below 220 base pairs and there are none that fall entirely below 80 base pairs[6; 44]. DNA degradation is more likely to affect larger DNA fragments however the

degradation index only considers DNA fragments with lengths comparable to miniSTR loci. The relationship between the degradation index and the slope of the miniSTR loci should be investigated to determine whether the degradation index is a better indicator of the degradation in this region of the profile compared to the entire profile.

# Chapter 5: Conclusion

The type of tape, number of layers of tape, and treatment length were found to have a significant effect on the degradation index of DNA samples, however, the treatment length was the only factor that had a significant effect on the slope of the DNA profile. Further investigation may determine whether a longer treatment period results in a significant interaction between the type of tape and number of layers on the slope of the DNA profile. These results suggest that the degradation index alone cannot be used to evaluate the level of degradation in a DNA sample or estimate the slope of an STR profile. A multiple linear regression model, which takes into account the degradation index as well as the amount of DNA amplified, may provide a better representation of the profile slope, however degraded DNA samples which are more representative of casework samples are necessary in order to construct the model. This knowledge may assist a forensic laboratory to determine whether DNA analysis is appropriate for a specific piece of evidence given the type of tape or number of layers present as well as when to proceed with DNA amplification given the degradation index and amount of DNA present. Considerations for future work include evaluating the relationship between the degradation index and the miniSTR loci utilized in the GlobalFiler™ PCR Amplification kit.

# References


1. van Oorschot RA, Ballantyne KN, Mitchell RJ. 2010. Forensic trace DNA: A review. Investigative Genetics. 1(1):1-17.
2. Watson JD, Crick FHC. 2003. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. 1953. Annals of Internal Medicine. 138(7):581.
3. Daniel R, Walsh SJ. 2006. The continuing evolution of forensic DNA profiling - from STRs to SNPs. Australian Journal of Forensic Sciences. 38(2):59.
4. Bowyer VL. 2007. Real-time PCR. Forensic science, medicine, and pathology. 3(1):61.
5. National Institute of J. 2014. Forensic DNA: Human DNA quantitation.
6. Applied Biosystems. 2015. Globalfiler™ PCR Amplification kit user guide. Thermo Fisher Scientific.
7. Friedman AL. 1999. Forensic DNA profiling in the 21st century. International Journal of Offender Therapy and Comparative Criminology. 43(2):168-179.
8. Buel E, Schwartz MB, LaFountain MJ. 1998. Capillary electrophoresis STR analysis: Comparison to gel-based systems. Journal of forensic sciences. 43(1):164-170.
9. Butler J. 2011. Capillary electrophoresis: Principles and instrumentation. Advanced topics in forensic DNA typing: Methodology. Academic Press. p. 141.
10. Jobling MA, Gill P. 2004. Encoded evidence: DNA in forensic analysis. Nature Reviews Genetics. 5(10):739-751.
11. Butler J. 2009. DNA separation and detection. Fundamentals of forensic DNA typing. Academic Press.
12. Hall A, Sims L, Foster A, Ballantyne J. 2016. DNA damage and repair in forensic science. In: Katz E, Halamek J, editors. Forensic science : chemistry, physics, biology and engineering. 1 ed. Weinheim, Germany: John Wiley & Sons, Incorporated. p. 193-214.
13. Gates KS. 2009. An overview of chemical processes that damage cellular DNA: Spontaneous hydrolysis, alkylation, and reactions with radicals. Chemical research in toxicology. 22(11):1747-1760.
14. Kolarevic A, Yancheva D, Kocic G, Smelcerovic A. 2014. Deoxyribonuclease inhibitors. European Journal of Medicinal Chemistry. 88:101-111.
15. Guéroult M, Picot D, Abi-Ghanem J, Hartmann B, Baaden M. 2010. How cations can assist DNase I in DNA binding and hydrolysis. PLOS Computational Biology. 6(11):e1001000.
16. Kitayama T, Fujii K, Nakahara H, Mizuno N, Kasai K, Yonezawa N, Sekiguchi K. 2013. Estimation of the detection rate in STR analysis by determining the DNA degradation ratio using quantitative PCR. Legal Medicine. 15(1):1-6.
17. McGregor WG. 1999. DNA repair, DNA replication, and UV mutagenesis. The Journal of Investigative Dermatology Symposium Proceedings. 4(1):1-5.
18. Applied Biosystems. 2015. Quantifiler™ HP and Trio DNA Quantification kits user guide. Thermo Fisher Scientific
19. Al-Kandari N, Singh J, Sanger VC. 2016. Time-dependent effects of temperature on quality of DNA in samples of human saliva, blood and semen in kuwait. International Journal of Pharmaceutical Sciences and Research. 7(7):2852-2873.
20. Barbaro A, Cormaci P. 2008. Study about the effect of high temperatures on strs typing. Forensic Science International: Genetics Supplement Series. 1(1):92-94.
21. Thacker CR, Oguzturun C, Ball KM, Syndercombe Court D. 2006. An investigation into methods to produce artificially degraded DNA. International Congress Series. 1288:592-594.
22. Bright J-A, Taylor D, Curran JM, Buckleton JS. 2013. Degradation of forensic DNA profiles. Australian Journal of Forensic Sciences. 45(4):445-449.

23. Vernarecci S, Ottaviani E, Agostino A, Mei E, Calandro L, Montagna P. 2015. Quantifiler® Trio kit and forensic samples management: A matter of degradation. Forensic Science International: Genetics. 16:77-85.
24. Alfonse LE, Garrett AD, Lun DS, Duffy KR, Grgicak CM. 2018. A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt. Forensic Science International: Genetics. 32:62-70.
25. Oldroyd N. 2012. Webinar: Introducing GlobalFiler™-- the world's most powerful STR kit. Thermo Fisher Scientific.
26. Wiegand P, Kleiber M. 2001. Less is more – length reduction of STR amplicons using redesigned primers. International journal of legal medicine. 114(4):285-287.
27. Dixon LA, Dobbins AE, Pulker HK, Butler JM, Vallone PM, Coble MD, Parson W, Berger B, Grubwieser P, Mogensen HS et al. 2006. Analysis of artificially degraded DNA using STRs and SNPs—results of a collaborative European (ednap) exercise. Forensic Science International. 164(1):33-44.
28. Grubwieser P, Mühlmann R, Berger B, Niederstätter H, Pavlic M, Parson W. 2006. A new "ministr-multiplex" displaying reduced amplicon lengths for the analysis of degraded DNA. International Journal of Legal Medicine. 120(2):115-120.
29. Phetpeng S, Kitpipit T, Asavutmangkul V, Duangshatome W, Pongsuwan W, Thanakiatkrai P. 2013. Touch DNA collection from improvised explosive devices: A comprehensive study of swabs and moistening agents. Forensic Science International: Genetics Supplement Series. 4(1):e29-e30.
30. Zech WD, Malik N, Thali M. 2012. Applicability of DNA analysis on adhesive tape in forensic casework. Journal of forensic sciences. 57(4):1036-1041.
31. Goodpaster JV, Sturdevant AB, Andrews KL, Brun-Conti L. 2007. Identification and comparison of electrical tapes using instrumental and statistical techniques: I. Microscopic surface texture and elemental composition. Journal of forensic sciences. 52(3):610-629.
32. Berti A, Barni F, Virgili A, Colozza C, Maiorino F, Tocca M. 2011. The recovery of DNA profiles from saliva and touch evidences after postal bomb explosion. Progress in Forensic Genetics 14Proceedings of the 24th International ISFG Congress. 3(1):e471-e472.
33. Esslinger KJ, Siegel JA, Spillane H, Stallworth S. 2004. Using STR analysis to detect human DNA from exploded pipe bomb devices. Journal of forensic sciences. 49(3):JFS2003127-2003124.
34. Hoffmann SG, Stallworth SE, Foran DR. 2012. Investigative studies into the recovery of DNA from improvised explosive device containers. Journal of forensic sciences. 57(3):602-609.
35. Phetpeng S, Kitpipit T, Thanakiatkrai P. 2015. Systematic study for DNA recovery and profiling from common IED substrates: From laboratory to casework. Forensic Science International: Genetics. 17:53-60.
36. Mehltretter AH, Bradley MJ. 2012. Forensic analysis and discrimination of duct tapes. Journal of American Society of Trace Evidence Examiners. 3(1):2-20.
37. SWGMAT. 2011. Guideline for forensic examination of pressure sensitive tapes. Journal of the American Society of Trace Examiners. 2(1):88-97.
38. SWGMAT. 2012. A 2012 survey regarding the status of forensic tape analysis.
39. Underwriters Laboratories I. 2005. Ul 510 polyvinyl chloride, polythylene, and rubber insulating tape.
40. Forensic examination of duct tape. Proceedings of the International Symposium on the Analysis and Identification of Polymers; 1984; FBI Academy, Quantico, VA.
41. R Core Team. 2017. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
42. Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software. 67(1):1-48.
43. Lenth R. 2018. Emmeans: Estimated marginal means, aka least-squares means. p. R package version 1.1.3.

44. Globalfiler PCR amplification kit. [accessed 2018 March 15]. https://www.thermofisher.com/order/catalog/product/4476135.

# Appendix A

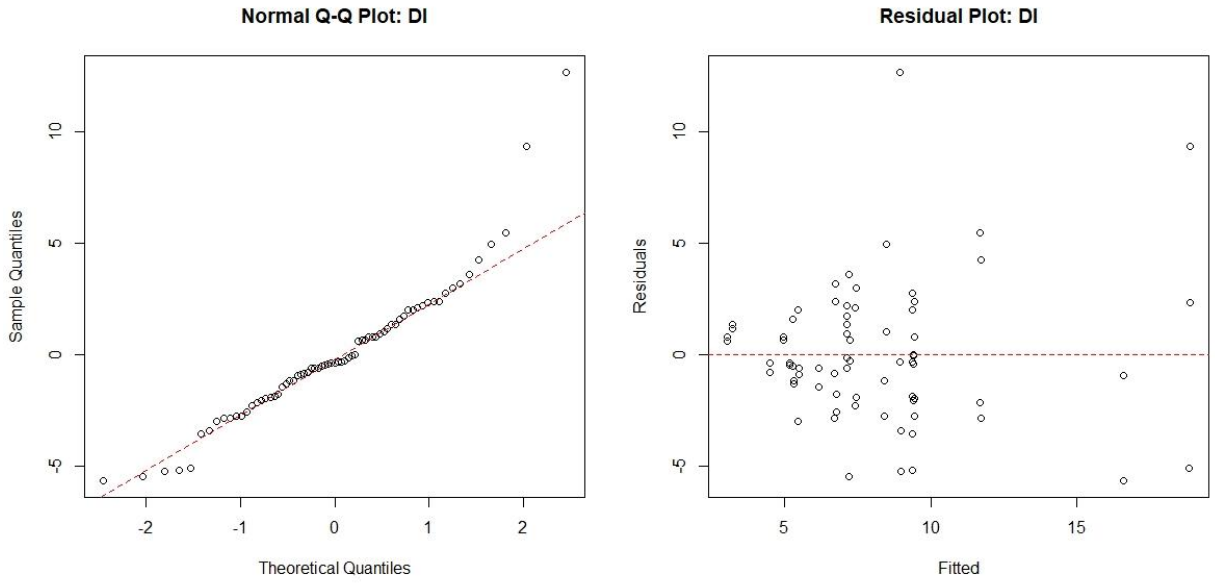


*Figure 5: Normal Q-Q plot and residual plot of the DI values for the samples treated with heat and humidity in this study.*

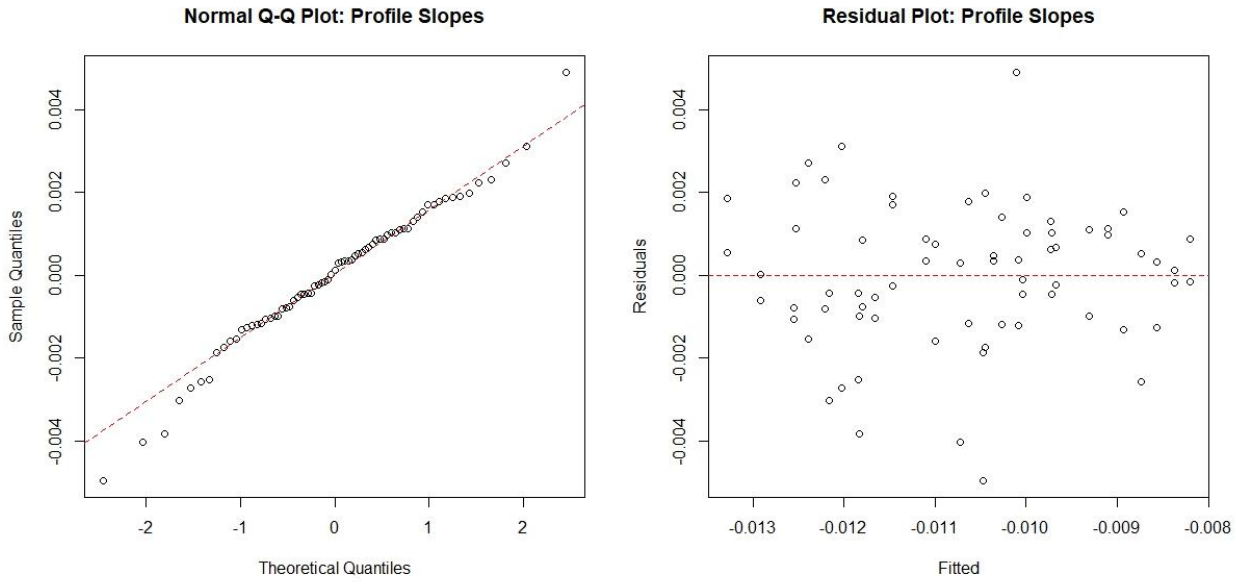


*Figure 6: Normal Q-Q plot and residual plot of the profile slopes for the samples treated with heat and humidity in this study.*

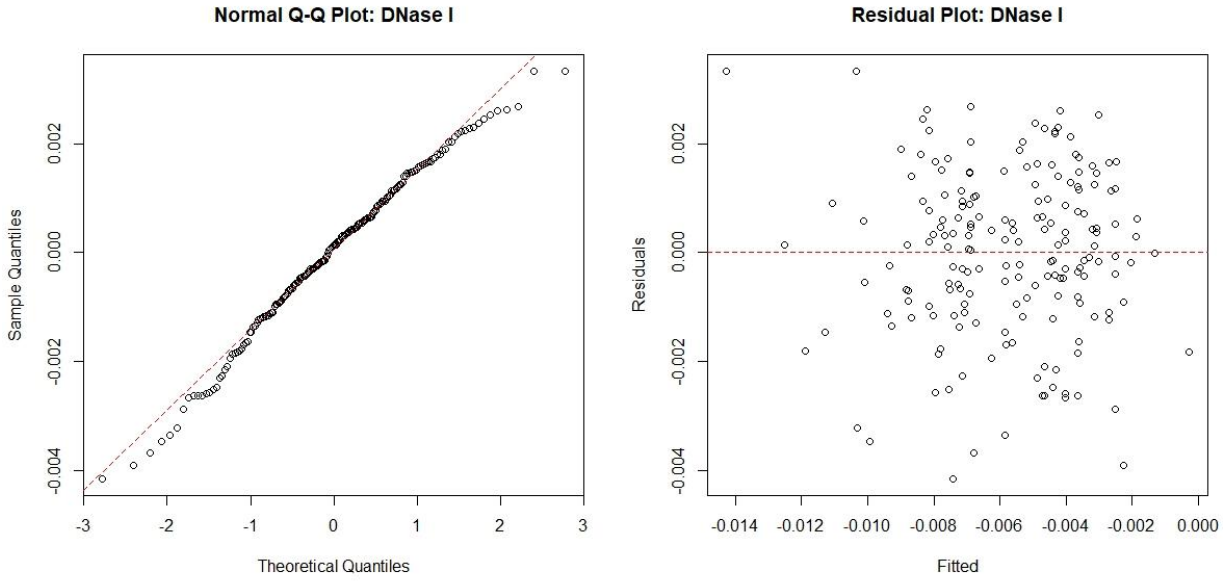


*Figure 7: Normal Q-Q plot and residual plot of the profile slopes for the DNase I treated samples obtained from the PROVEDIt Database that were used to created the DNase I model.*

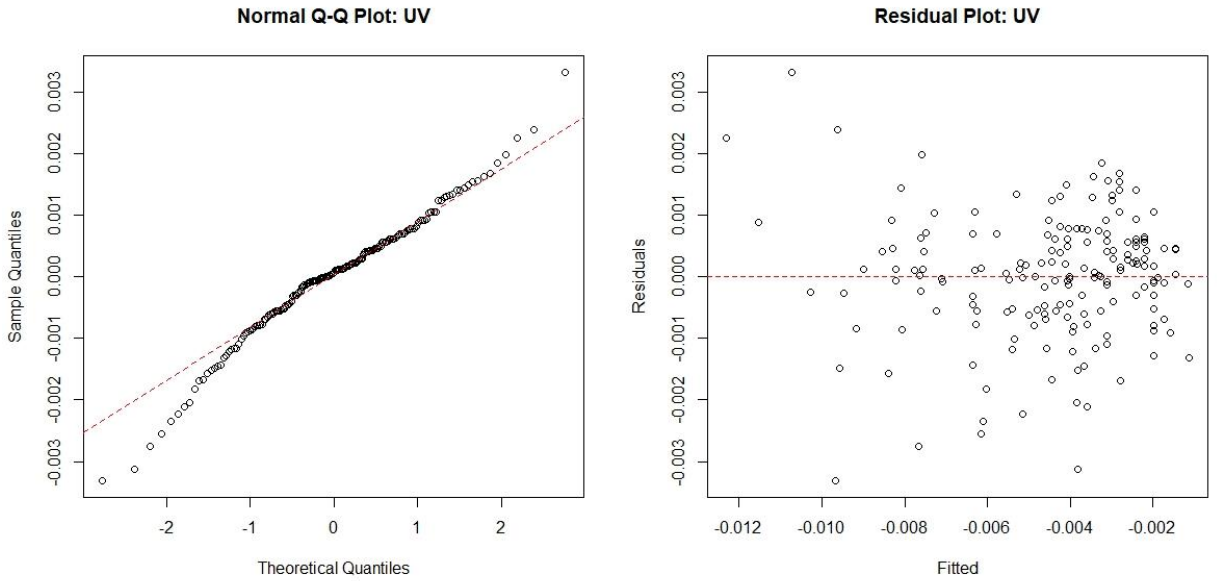


*Figure 8: Normal Q-Q plot and residual plot of the profile slopes for the UV treated samples obtained from the PROVEDIt Database that were used to created the UV model.*

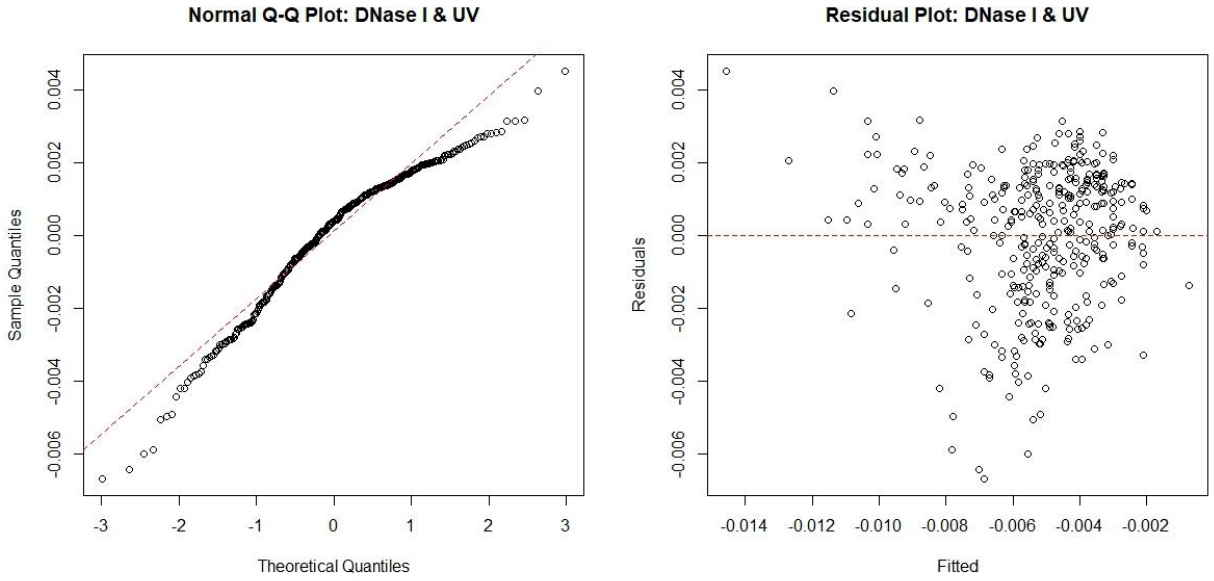


*Figure 9: Normal Q-Q plot and residual plot of the profile slopes for the DNase I and UV treated samples obtained from the PROVEDIt Database that were used to created the combined DNase I and UV model.*

# Appendix B

## Part 1: Calculations for heat and humidity treated samples

```
##########          DI Tables        ##########

Quant_Data<-read.csv("C:/Users/Emily/OneDrive/Documents/grad
research/Data/Quant/Total Quant Summary_With Conc Values.csv")
Quant_Data<-subset(Quant_Data,DI>=0)
OneWeek<-subset(Quant_Data, Weeks=="one")
Electrical1<-subset(OneWeek,Tape=="Electrical")
Electrical10<-subset(Electrical1,Layers=="zero")
Electrical11<-subset(Electrical1,Layers=="one")
Electrical12<-subset(Electrical1,Layers=="two")
Duct1<-subset(OneWeek,Tape=="Duct")
Duct10<-subset(Duct1,Layers=="zero")
Duct11<-subset(Duct1,Layers=="one")
Duct12<-subset(Duct1,Layers=="two")

TwoWeeks<-subset(Quant_Data, Weeks=="two")
Electrical2<-subset(TwoWeeks,Tape=="Electrical")
Electrical20<-subset(Electrical2,Layers=="zero")
Electrical21<-subset(Electrical2,Layers=="one")
Electrical22<-subset(Electrical2,Layers=="two")
Duct2<-subset(TwoWeeks,Tape=="Duct")
Duct20<-subset(Duct2,Layers=="zero")
Duct21<-subset(Duct2,Layers=="one")
Duct22<-subset(Duct2,Layers=="two")

OneFourFour_1W<-subset(OneWeek,Participant=="OneFourFour")
ThreeThreeOne_1W<-subset(OneWeek, Participant=="ThreeThreeOne")
FiveSixFive_1W<-subset(OneWeek, Participant=="FiveSixFive")

OneFourFour_2W<-subset(TwoWeeks,Participant=="OneFourFour")
ThreeThreeOne_2W<-subset(TwoWeeks, Participant=="ThreeThreeOne")
FiveSixFive_2W<-subset(TwoWeeks, Participant=="FiveSixFive")

mean(OneFourFour_1W$Small.Target); sd(OneFourFour_1W$Small.Target)
mean(ThreeThreeOne_1W$Small.Target); sd(ThreeThreeOne_1W$Small.Target)
mean(FiveSixFive_1W$Small.Target); sd(FiveSixFive_1W$Small.Target)

mean(OneFourFour_1W$Small.Target); sd(OneFourFour_1W$Small.Target)
mean(ThreeThreeOne_1W$Small.Target); sd(ThreeThreeOne_1W$Small.Target)
mean(FiveSixFive_1W$Small.Target); sd(FiveSixFive_1W$Small.Target)

mean(OneFourFour_2W$Small.Target); sd(OneFourFour_2W$Small.Target)
mean(ThreeThreeOne_2W$Small.Target); sd(ThreeThreeOne_2W$Small.Target)
mean(FiveSixFive_2W$Small.Target); sd(FiveSixFive_2W$Small.Target)

summary_E10<-
c(length(Electrical10$DI),mean(Electrical10$DI),sd(Electrical10$DI),
100*(sd(Electrical10$DI)/mean(Electrical10$DI)))
summary_E11<-
c(length(Electrical11$DI),mean(Electrical11$DI),sd(Electrical11$DI),
100*(sd(Electrical11$DI)/mean(Electrical11$DI)))
summary_E12<-
c(length(Electrical12$DI),mean(Electrical12$DI),sd(Electrical12$DI),
100*(sd(Electrical12$DI)/mean(Electrical12$DI)))
summary_D10<-c(length(Duct10$DI),mean(Duct10$DI),sd(Duct10$DI),
100*(sd(Duct10$DI)/mean(Duct10$DI)))
summary_D11<-c(length(Duct11$DI),mean(Duct11$DI),sd(Duct11$DI),
100*(sd(Duct11$DI)/mean(Duct11$DI)))
summary_D12<-c(length(Duct12$DI),mean(Duct12$DI),sd(Duct12$DI),
100*(sd(Duct12$DI)/mean(Duct12$DI)))
```

```
summary_E20<-
c(length(Electrical20$DI),mean(Electrical20$DI),sd(Electrical20$DI),
100*(sd(Electrical20$DI)/mean(Electrical20$DI)))
summary_E21<-
c(length(Electrical21$DI),mean(Electrical21$DI),sd(Electrical21$DI),
100*(sd(Electrical21$DI)/mean(Electrical21$DI)))
summary_E22<-
c(length(Electrical22$DI),mean(Electrical22$DI),sd(Electrical22$DI),
100*(sd(Electrical22$DI)/mean(Electrical22$DI)))
summary_D20<-c(length(Duct20$DI),mean(Duct20$DI),sd(Duct20$DI),
100*(sd(Duct20$DI)/mean(Duct20$DI)))
summary_D21<-c(length(Duct21$DI),mean(Duct21$DI),sd(Duct21$DI),
100*(sd(Duct21$DI)/mean(Duct21$DI)))
summary_D22<-c(length(Duct22$DI),mean(Duct22$DI),sd(Duct22$DI),
100*(sd(Duct22$DI)/mean(Duct22$DI)))

DuctMeansW1<-c(mean(Duct10$DI),mean(Duct11$DI),mean(Duct12$DI))
ElectricalMeansW1<-
c(mean(Electrical10$DI),mean(Electrical11$DI),mean(Electrical12$DI))

DuctMeansW2<-c(mean(Duct20$DI),mean(Duct21$DI),mean(Duct22$DI))
ElectricalMeansW2<-
c(mean(Electrical20$DI),mean(Electrical21$DI),mean(Electrical22$DI))

#Table for Week 1 Samples DI
table1<-matrix(c(summary_E10, summary_E11, summary_E12,
summary_D10,summary_D11,summary_D12), byrow=TRUE, ncol=4)
rownames(table1)<-c("Electrical, 0 Layers","Electrical, 1 Layer",
"Electrical, 2 Layers", "Duct, 0 Layers", "Duct, 1 Layer", "Duct, 2 Layers")
colnames(table1)<-c("Sample Size", "Mean","SD","RSD"); table1

#Table for week 2 Samples DI
table2<-matrix(c(summary_E20, summary_E21, summary_E22,
summary_D20,summary_D21,summary_D22), byrow=TRUE, ncol=4)
rownames(table2)<-c("Electrical, 0 Layers","Electrical, 1 Layer",
"Electrical, 2 Layers", "Duct, 0 Layers", "Duct, 1 Layer", "Duct, 2 Layers")
colnames(table2)<-c("Sample Size", "Mean","SD","RSD"); table2

##########          DI Box Plots        ##########

par(mfrow = c(1, 2))
boxplot(Duct10$DI,Electrical10$DI,Duct11$DI,Electrical11$DI,Duct12$DI,Electri
cal12$DI,
        col=c("gray87","gray47","gray87","gray47","gray87","gray47"),
        names=c("0","0","1","1","2","2"),
        xlab="Layers",ylab="DI",main="DI Of Week One Treatment Groups",
        xlim = c(0,7),ylim=c(0,30))
legend(-2.5,31, legend=c("Duct Tape","Electrical
Tape"),pch=c(15,15),col=c("gray87","gray47"),
       box.lty=0,bg="NA",y.intersp=1,x.intersp=0.2)
mtext(c("A"),side=3,line=2,at=-0.2,font=2)

boxplot(Duct20$DI,Electrical20$DI,Duct21$DI,Electrical21$DI,Duct22$DI,Electri
cal22$DI,
        col=c("gray87","gray47","gray87","gray47","gray87","gray47"),
        names=c("0","0","1","1","2","2"),
        xlab="Layers",ylab="DI",main="DI of Week Two Treatment Groups",xlim =
c(0,7),ylim=c(0,30))
legend(-2.5,31, legend=c("Duct Tape","Electrical
Tape"),pch=c(15,15),col=c("gray87","gray47"),
       box.lty=0,bg="NA",y.intersp=1,x.intersp=0.2)
mtext(c("B"),side=3,line=2,at=-0.2,font=2)
```

```
##########          DI Analysis        ##########

library(lme4)

# test to see if week 1 and week 2 should be combined or assessed separately
full_1<-lmer(DI~Tape*Layers*Weeks+(1|Participant),data=Quant_Data)
red_1<-
lmer(DI~Tape+Layers+Weeks+Tape:Layers+Tape:Weeks+Layers:Weeks+(1|Participant)
,data=Quant_Data)

par(mfrow = c(1, 2))
qqnorm(resid(full_1), main="Normal Q-Q Plot: DI")
qqline(resid(full_1),col=2,lty=2)
plot(fitted(full_1),
resid(full_1),xlab="Fitted",ylab="Residuals",main="Residual Plot: DI")
abline(0,0, col=2,lty=2)

anova(full_1,red_1)
# p<0.05 -> reduced model is not as good, need to keep Weeks
# Assess 1 week and 2 weeks separately rather than combining all samples

# Test to see if there is a significant interaction between tape and layers--
Week One
full_week1I<-lmer(DI~Tape+Layers+Tape:Layers+(1|Participant),data=OneWeek)
red_week1TL<-lmer(DI~Tape+Layers+(1|Participant),data=OneWeek)

anova(full_week1I,red_week1TL)
# p>0.05 -> there is no significant difference between the models
# there is not a significant interaction between layers and tape at week 1

# No interaction -> look at main effects
full_week1<-lmer(DI~Tape+Layers+(1|Participant),data=OneWeek)
library(emmeans)
emmeans(full_week1, list(pairwise~Layers),adjust="tukey",weight="cells")
# There are no significant difference between layers of tape (averaged acros
tape)
emmeans(full_week1, list(pairwise~Tape),adjust="tukey",weight="cells")
# Theere are no significant difference between types of tape (averaged across
layers)

# Test to see if there is a significant interaction between tape and layers--
Week Two
full_week2I<-lmer(DI~Tape+Layers+Tape:Layers+(1|Participant),data=TwoWeeks)
red_week2TL<-lmer(DI~Tape+Layers+(1|Participant),data=TwoWeeks)

anova(full_week2I,red_week2TL)
# p<0.05 -> there is a significant difference between the models
# There is a significant interaction between layers and tape at week 2

# Interaction -> look at simple effects
emmeans(full_week2I,list(pairwise~Layers:Tape),adjust="tukey")
# p<0.05 for one,Duct-two,Duct
# p<0.05 for two,Duct-zero,Duct
# p<0.05 for two,Duct-two,Electrical

# Calculate Means for week 2
emmeans(red_week2TL,list(pairwise~Layers),adjust="tukey",weight="cells")
emmeans(red_week2TL, list(pairwise~Tape),adjust="tukey",weight="cells")

##########          SLOPE CALCULATION          ##########

Slope_Values<-matrix(byrow=TRUE,ncol=2,nrow=0);colnames(Slope_Values)<-
c("Sample","Slope")
```

```
Slope<-function(x,SaveLocation) {

  for(i in 1:length(x)){

    profile<-read.csv(x[i],header=TRUE)
    sample<-matrix(profile[1,1],byrow=FALSE,ncol=1)
    marker<-as.character(profile[,2])

    tble<-
matrix(c(profile[,5],profile[,6],profile[,7],profile[,8]),byrow=FALSE,ncol=4)
    colnames(tble)<-c("Height.1","Height.2","Size.1","Size.2")
    row.names(tble)<-marker
    tble[is.na(tble)]<-(-1)
    marker<-row.names(tble)

    adj_height<-vector("numeric")
    for(i in 1:length(tble[,1])){
      h<-ifelse(tble[i,2]>=0,(tble[i,1]+tble[i,2]),
                ifelse(tble[i,1]>=0,(tble[i,1]),NA))
      adj_height[i]<-h
      i=i+1
    }

    adj_size<-vector("numeric")
    for(i in 1:length(tble[,3])){
      s<-ifelse(tble[i,4]>=0,((tble[i,3]+tble[i,4])/2),
                ifelse(tble[i,3]>=0,(tble[i,3]),NA))
      adj_size[i]<-s
      i=i+1
    }

    adj_data<-data.frame(adj_height,adj_size)
    colnames(adj_data)<-c("Adjusted_Height","Adjusted_Size")
    row.names(adj_data)<-marker
    adj_data[adj_data==0]<-NA
    adj_data<-subset(adj_data,Adjusted_Height!="NA")

    model<-lm(log(Adjusted_Height)~Adjusted_Size,data=adj_data)
    slope<-model$coefficients[2]

    results<-c(sample,slope)
    Slope_Values<-rbind(Slope_Values,results)
  }
  #print(Slope_Values)
  write.csv(Slope_Values, file=SaveLocation)
}

Files<-list.files(path="C:/Users/Emily/OneDrive/Documents/grad
research/Data/Profiles/Weeks 1 and 2 and
Conc_Corrected",full.names=TRUE,recursive=FALSE)
Slopes<-("C:/Users/Emily/Desktop/Grad_Research/R
Data/Conc_Slopes_NewCalc_Corrected.csv")
Slope(Files,Slopes)

Slopes<-read.csv("C:/Users/Emily/Desktop/Grad_Research/R
Data/Conc_Slopes_NewCalc_Corrected.csv")
Quant_Data<-read.csv("C:/Users/Emily/OneDrive/Documents/grad
research/Data/Quant/Total Quant Summary_With Conc Values.csv")
Quant_Data$Sample.Name<-(as.character(Quant_Data$Sample.Name))
class(Quant_Data$Tape)

TotalData<-cbind(Quant_Data,Slopes$Slope)
TotalData<-TotalData[,(2:11)]
```

```
colnames(TotalData)<-
c("Large_Target","Small_Target","Y_Target","DI","Layers","Tape","Weeks","Amp_
Target","Participant","Slope")
row.names(TotalData)<-Slopes$Sample
TotalData<-TotalData[-42,]
plot(Slope~DI,data=TotalData)
plot(Slope~log(DI),data=TotalData)

Names<-Slopes$Sample
Names<-Names[-42]

SlopeNGDI<-cbind(TotalData$Slope,TotalData$Amp_Target,TotalData$DI)
row.names(SlopeNGDI)<-Names
SlopesNGDI<-write.csv(SlopeNGDI, "C:/Users/Emily/Desktop/Grad_Research/R
Data/SlopesNGDI.csv")

##########          Slope Tables        ##########

OneWeek_S<-subset(TotalData, Weeks=="one")
Electrical1_S<-subset(OneWeek_S,Tape=="Electrical")
Electrical10_S<-subset(Electrical1_S,Layers=="zero")
Electrical11_S<-subset(Electrical1_S,Layers=="one")
Electrical12_S<-subset(Electrical1_S,Layers=="two")
Duct1_S<-subset(OneWeek_S,Tape=="Duct")
Duct10_S<-subset(Duct1_S,Layers=="zero")
Duct11_S<-subset(Duct1_S,Layers=="one")
Duct12_S<-subset(Duct1_S,Layers=="two")

TwoWeeks_S<-subset(TotalData, Weeks=="two")
Electrical2_S<-subset(TwoWeeks_S,Tape=="Electrical")
Electrical20_S<-subset(Electrical2_S,Layers=="zero")
Electrical21_S<-subset(Electrical2_S,Layers=="one")
Electrical22_S<-subset(Electrical2_S,Layers=="two")
Duct2_S<-subset(TwoWeeks_S,Tape=="Duct")
Duct20_S<-subset(Duct2_S,Layers=="zero")
Duct21_S<-subset(Duct2_S,Layers=="one")
Duct22_S<-subset(Duct2_S,Layers=="two")

summary_E10S<-
c(length(Electrical10_S$Slope),mean(Electrical10_S$Slope),sd(Electrical10_S$S
lope), 100*(sd(Electrical10_S$Slope)/mean(Electrical10_S$Slope)))
summary_E11S<-
c(length(Electrical11_S$Slope),mean(Electrical11_S$Slope),sd(Electrical11_S$S
lope), 100*(sd(Electrical11_S$Slope)/mean(Electrical11_S$Slope)))
summary_E12S<-
c(length(Electrical12_S$Slope),mean(Electrical12_S$Slope),sd(Electrical12_S$S
lope), 100*(sd(Electrical12_S$Slope)/mean(Electrical12_S$Slope)))
summary_D10S<-
c(length(Duct10_S$Slope),mean(Duct10_S$Slope),sd(Duct10_S$Slope),
100*(sd(Duct10_S$Slope)/mean(Duct10_S$Slope)))
summary_D11S<-
c(length(Duct11_S$Slope),mean(Duct11_S$Slope),sd(Duct11_S$Slope),
100*(sd(Duct11_S$Slope)/mean(Duct11_S$Slope)))
summary_D12S<-
c(length(Duct12_S$Slope),mean(Duct12_S$Slope),sd(Duct12_S$Slope),
100*(sd(Duct12_S$Slope)/mean(Duct12_S$Slope)))

summary_E20S<-
c(length(Electrical20_S$Slope),mean(Electrical20_S$Slope),sd(Electrical20_S$S
lope), 100*(sd(Electrical20_S$Slope)/mean(Electrical20_S$Slope)))
summary_E21S<-
c(length(Electrical21_S$Slope),mean(Electrical21_S$Slope),sd(Electrical21_S$S
lope), 100*(sd(Electrical21_S$Slope)/mean(Electrical21_S$Slope)))
```

```
summary_E22S<-
c(length(Electrical22_S$Slope),mean(Electrical22_S$Slope),sd(Electrical22_S$S
lope), 100*(sd(Electrical22_S$Slope)/mean(Electrical22_S$Slope)))
summary_D20S<-
c(length(Duct20_S$Slope),mean(Duct20_S$Slope),sd(Duct20_S$Slope),
100*(sd(Duct20_S$Slope)/mean(Duct20_S$Slope)))
summary_D21S<-
c(length(Duct21_S$Slope),mean(Duct21_S$Slope),sd(Duct21_S$Slope),
100*(sd(Duct21_S$Slope)/mean(Duct21_S$Slope)))
summary_D22S<-
c(length(Duct22_S$Slope),mean(Duct22_S$Slope),sd(Duct22_S$Slope),
100*(sd(Duct22_S$Slope)/mean(Duct22_S$Slope)))

DuctMeansW1_S<-
c(mean(Duct10_S$Slope),mean(Duct11_S$Slope),mean(Duct12_S$Slope))
ElectricalMeansW1_S<-
c(mean(Electrical10_S$Slope),mean(Electrical11_S$Slope),mean(Electrical12_S$S
lope))

DuctMeansW2_S<-
c(mean(Duct20_S$Slope),mean(Duct21_S$Slope),mean(Duct22_S$Slope))
ElectricalMeansW2_S<-
c(mean(Electrical20_S$Slope),mean(Electrical21_S$Slope),mean(Electrical22_S$S
lope))

#Table for Week 1 Samples Slopes
table1S<-matrix(c(summary_E10S, summary_E11S, summary_E12S,
summary_D10S,summary_D11S,summary_D12S), byrow=TRUE, ncol=4)
rownames(table1S)<-c("Electrical, 0 Layers","Electrical, 1 Layer",
"Electrical, 2 Layers", "Duct, 0 Layers", "Duct, 1 Layer", "Duct, 2 Layers")
colnames(table1S)<-c("Sample Size", "Mean","SD","RSD"); table1S

#Table for week 2 Samples Slope
table2S<-matrix(c(summary_E20S, summary_E21S, summary_E22S,
summary_D20S,summary_D21S,summary_D22S), byrow=TRUE, ncol=4)
rownames(table2S)<-c("Electrical, 0 Layers","Electrical, 1 Layer",
"Electrical, 2 Layers", "Duct, 0 Layers", "Duct, 1 Layer", "Duct, 2 Layers")
colnames(table2S)<-c("Sample Size", "Mean","SD","RSD"); table2S

##########          Slope Box Plots        ##########

par(mfrow = c(1, 2))
boxplot(Duct10_S$Slope,Electrical10_S$Slope,Duct11_S$Slope,Electrical11_S$Slo
pe,Duct12_S$Slope,Electrical12_S$Slope,
        col=c("gray87","gray47","gray87","gray47","gray87","gray47"),
        names=c("0","0","1","1","2","2"),
        xlab="Layers",ylab="Profile Slope",main="Profile Slope of Week One
Treatment Groups",
        xlim = c(0,7),ylim=c(-0.016,-0.004))
legend(-1,-0.003, legend=c("Duct Tape","Electrical
Tape"),pch=c(15,15),col=c("gray87","gray47"),box.lty=0,bg="NA",y.intersp=1,x.
intersp=0.2)
mtext(c("A"),side=3,line=2,at=-0.2,font=2)

boxplot(Duct20_S$Slope,Electrical20_S$Slope,Duct21_S$Slope,Electrical21_S$Slo
pe,Duct22_S$Slope,Electrical22_S$Slope,
        col=c("gray87","gray47","gray87","gray47","gray87","gray47"),
        names=c("0","0","1","1","2","2"),
        ylab="Profile Slope",main="Profile Slope of Week Two Treatment
Groups",
        xlab="Layers",xlim = c(0,7),ylim=c(-0.016,-0.004))
legend(-1,-0.003, legend=c("Duct Tape","Electrical
Tape"),pch=c(15,15),col=c("gray87","gray47"),box.lty=0,bg="NA",y.intersp=1,x.
intersp=0.2)
```

```
mtext(c("B"),side=3,line=2,at=-0.2,font=2)

##########          Slope Analysis      ##########

library(lme4)
# test to see if week 1 and week 2 should be combined or assessed separately
full_1_S<-lmer(Slope~Tape*Layers*Weeks+(1|Participant),data=TotalData)
red_1_S<-
lmer(Slope~Tape+Layers+Weeks+Tape:Layers+Tape:Weeks+Layers:Weeks+(1|Participa
nt),data=TotalData)

par(mfrow=c(1,2))
qqnorm(resid(full_1_S),main="Normal Q-Q Plot: Profile Slopes")
qqline(resid(full_1_S),col=2,lty=2)
plot(fitted(full_1_S),
resid(full_1_S),xlab="Fitted",ylab="Residuals",main="Residual Plot: Profile
Slopes")
abline(0,0, col=2,lty=2)

anova(full_1_S,red_1_S)
# p>0.05 -> full model is not significantly better than the reduced model
# Do not assess the samples separately

# Removed Layers:Weeks
red_2_S<-
lmer(Slope~Tape+Layers+Weeks+Tape:Layers+Tape:Weeks+(1|Participant),data=Tota
lData)
# Removed Tape:Weeks
red_3_S<-
lmer(Slope~Tape+Layers+Weeks+Tape:Layers+Layers:Weeks+(1|Participant),data=To
talData)
# Removed Tape:Layers
red_4_S<-
lmer(Slope~Tape+Layers+Weeks+Tape:Weeks+Layers:Weeks+(1|Participant),data=Tot
alData)

# Test if Layers:Weeks is significant
anova(red_1_S,red_2_S)
# p > 0.05
# Layers:Weeks is not significant

# Test if Tape:Weeks is significant
anova(red_1_S,red_3_S)
# p > 0.05
# Tape:Weeks is not significant

# Test if Tape:Layers is significant
anova(red_1_S,red_4_S)
# p > 0.05
# Tape:Layers is not significant

red_5_S<-lmer(Slope~Tape+Layers+Weeks+(1|Participant),data=TotalData)

# No two-way interactions -> look at main effects
red_5_S
library(emmeans)
emmeans(red_5_S, list(pairwise~Layers),adjust="tukey",weight="cells")

emmeans(red_5_S, list(pairwise~Tape),adjust="tukey",weight="cells")

emmeans(red_5_S, list(pairwise~Weeks),adjust="tukey",weight="cells")
```

## Part 2: Calculations for linear model to estimate profile slope

```
##########   CALCULATE SLOPE FOR FILE OF PROFILES--MODEL   ##########

Slope_Values<-matrix(byrow=TRUE,ncol=3,nrow=0);colnames(Slope_Values)<-
c("Sample","Slope","AmpInput")

Slope<-function(x,L){
  for(i in 1:length(x)){
    profile<-read.csv(x[i],header=TRUE)
    SampleName<-matrix(profile[1,1],byrow=FALSE,ncol=1)
    sample<-data.frame(profile[1,1])

    marker<-as.character(profile[,2])

    tble<-
matrix(c(profile[,5],profile[,6],profile[,7],profile[,8]),byrow=FALSE,ncol=4)
    colnames(tble)<-c("Height.1","Height.2","Size.1","Size.2")
    row.names(tble)<-marker
    tble[is.na(tble)]<-(-1)
    marker<-row.names(tble)

    adj_height<-vector("numeric")
    for(i in 1:length(tble[,1])){
      h<-ifelse(tble[i,2]>=0,(tble[i,1]+tble[i,2]),
                ifelse(tble[i,1]>=0,(tble[i,1]),NA))
      adj_height[i]<-h
      i=i+1
    }

    adj_size<-vector("numeric")
    for(i in 1:length(tble[,3])){
      s<-ifelse(tble[i,4]>=0,((tble[i,3]+tble[i,4])/2),
                ifelse(tble[i,3]>=0,(tble[i,3]),NA))
      adj_size[i]<-s
      i=i+1
    }

    adj_data<-data.frame(adj_height,adj_size)
    colnames(adj_data)<-c("Adjusted_Height","Adjusted_Size")
    row.names(adj_data)<-marker
    adj_data<-subset(adj_data,Adjusted_Height!="NA")

    model<-lm(log(Adjusted_Height)~Adjusted_Size, data=adj_data)
    slope<-model$coefficients[2]

    library(tidyr)
    target1<-separate(data=sample,col="profile.1..1.",into=c("a"),sep="GF")
    target<-separate(data=target1,col="a",into=c("a","b","c","d","e"),sep="-
")
    target[is.na(target)]<-(-1)
    ng<-if(target$e==(-1)) print(target$d) else print(target$e)
    ng<-as.numeric(as.character(ng))

    results<-c(SampleName,slope,ng)
    Slope_Values<-rbind(Slope_Values,results)
  }
  print(Slope_Values)
  write.csv(Slope_Values, file=L)
}

SlopeNGDI<-read.csv("C:/Users/Emily/Desktop/Grad_Research/R
Data/SlopesNGDI.csv")
RNames<-SlopeNGDI$X
```

```
SlopeNGDI<-SlopeNGDI[,2:4]
colnames(SlopeNGDI)<-c("Slope","AmpInput","DI")
row.names(SlopeNGDI)<-RNames
SlopeNGDI<-subset(SlopeNGDI,DI!="NA")

SlopeNGDI2<-subset(SlopeNGDI,DI<=14)
SlopeNGDI2<-subset(SlopeNGDI2,AmpInput<=0.7)
SlopeNGDI3<-subset(SlopeNGDI,AmpInput<=0.504)
SlopeNGDI4<-subset(SlopeNGDI,AmpInput<=0.7)

length(SlopeNGDI$DI) #70
length(SlopeNGDI2$DI) #40 ->41
length(SlopeNGDI3$DI) #40 ->41
length(SlopeNGDI4$DI) #46 ->47

min(SlopeNGDI$DI) #1.73
max(SlopeNGDI$DI) #28.22

min(SlopeNGDI$AmpInput) #0.099
max(SlopeNGDI$AmpInput) #1

min(SlopeNGDI$Slope) # -0.01566286
max(SlopeNGDI$Slope) # -0.005207175

mean(SlopeNGDI$Slope) # -0.01072611

min(SlopeNGDI2$Slope) # -0.01566286
max(SlopeNGDI2$Slope) # -0.01566286

######## DNase Model ########

Data_Files<-list.files(path="C:/Users/Emily/Desktop/PROVEDit
Database/PROVEDit Reports/DNase_All
Peaks/All",full.names=TRUE,recursive=FALSE)
Slopes_File_Location<-"C:/Users/Emily/Desktop/Grad_Research/R
Data/SlopesNew_DNase All Peaks.csv";Slopes_File_Location
Slope(Data_Files,Slopes_File_Location)

slope_dat<-read.csv(Slopes_File_Location)
Samples<-data.frame(slope_dat$Sample)

library(tidyr)
DI<-separate(data=Samples, col="slope_dat.Sample", into=c("Sample","DI"),
sep="-Q")
DI<-data.frame(DI$DI)
row.names(DI)<-slope_dat$Sample
colnames(DI)<-c("DI")
DI<-DI$DI<-as.numeric(as.character(DI$DI))
dat<-cbind(slope_dat,DI)

min(dat$DI) # 0.5
max(dat$DI) # 14
min(dat$AmpInput) #0.0156
max(dat$AmpInput) #0.7

modelA<-lm(Slope~log(DI)+AmpInput+log(DI):AmpInput, data=dat)

par(mfrow=c(1,2))
qqnorm(modelA$residuals,main="Normal Q-Q Plot: DNase I")
qqline(modelA$residuals, col=2, lty=2)
plot(modelA$residuals~modelA$fitted,xlab="Fitted",ylab="Residuals",main="Resi
dual Plot: DNase I")
abline(0,0,col=2,lty=2)
```

```
summary(modelA)

y<-predict(modelA,SlopeNGDI2)

Original_New_DNase<-data.frame(SlopeNGDI2,y)
write.csv(Original_New_DNase,"C:/Users/Emily/Desktop/Grad_Research/R
Data/Original_New_DNase.csv")

RMSE<-sqrt(mean((y-SlopeNGDI2$Slope)^2))
RMSE

##########          UV Model          ##########

Data_Files2<-list.files(path="C:/Users/Emily/Desktop/PROVEDit
Database/PROVEDit Reports/15sec_UV_all peaks/All",
full.names=TRUE,recursive=FALSE)
Slopes_File_Location2<-"C:/Users/Emily/Desktop/Grad_Research/R
Data/SlopesNew_UV All Peaks.csv"
Slope(Data_Files2,Slopes_File_Location2)

slope_dat2<-read.csv(Slopes_File_Location2)
Samples2<-data.frame(slope_dat2$Sample);Samples2

library(tidyr)
DI2<-separate(data=Samples2, col="slope_dat2.Sample", into=c("Sample","DI"),
sep="-Q")
DI2<-data.frame(DI2$DI)
row.names(DI2)<-slope_dat2$Sample
colnames(DI2)<-c("DI")
DI2<-DI2$DI<-as.numeric(as.character(DI2$DI))
dat2<-cbind(slope_dat2,DI2);colnames(dat2)<-
c("X","Sample","Slope","AmpInput","DI")

min(dat2$DI) # 0.7
max(dat2$DI) # 58
min(dat2$AmpInput) # 0.0313
max(dat2$AmpInput) # 0.504

modelA2<-lm(Slope~log(DI)+AmpInput+log(DI):AmpInput, data=dat2)

par(mfrow = c(1, 2))
qqnorm(modelA2$residuals,main="Normal Q-Q Plot: UV")
qqline(modelA2$residuals, col=2, lty=2)
plot(modelA2$residuals~modelA2$fitted,xlab="Fitted",ylab="Residuals",main="Re
sidual Plot: UV")
abline(0,0,col=2,lty=2)

summary(modelA2)

y2<-predict(modelA2,SlopeNGDI3)

RMSE2<-sqrt(mean((y2-SlopeNGDI3$Slope)^2))
RMSE2 # Root Mean Square Error

##########          DNase & UV Model          ##########

Data_Files3<-list.files(path="C:/Users/Emily/Desktop/PROVEDit
Database/PROVEDit Reports/UV_DNase_All
Peaks",full.names=TRUE,recursive=FALSE)
Slopes_File_Location3<-"C:/Users/Emily/Desktop/Grad_Research/R
Data/SlopesNew_DNase UV All Peaks.csv"
Slope(Data_Files3,Slopes_File_Location3)

slope_dat3<-read.csv(Slopes_File_Location3)
```

```
Samples3<-data.frame(slope_dat3$Sample);Samples3

library(tidyr)
DI3<-separate(data=Samples3, col="slope_dat3.Sample", into=c("Sample","DI"),
sep="-Q")
DI3<-data.frame(DI3$DI)
row.names(DI3)<-slope_dat3$Sample
colnames(DI3)<-c("DI")
DI3<-DI3$DI<-as.numeric(as.character(DI3$DI))
dat3<-cbind(slope_dat3,DI3);colnames(dat3)<-
c("X","Sample","Slope","AmpInput","DI")

min(dat3$DI) # 0.5
max(dat3$DI) # 58
min(dat3$AmpInput) # 0.0156
max(dat3$AmpInput) # 0.7

modelA3<-lm(Slope~log(DI)+AmpInput+log(DI):AmpInput, data=dat3)

par(mfrow = c(1, 2))
qqnorm(modelA3$residuals,main="Normal Q-Q Plot: DNase I & UV")
qqline(modelA3$residuals, col=2, lty=2)
plot(modelA3$residuals~modelA3$fitted,xlab="Fitted",ylab="Residuals",main="Re
sidual Plot: DNase I & UV")
abline(0,0,col=2,lty=2)

summary(modelA3)

y3<-predict(modelA3,SlopeNGDI4)

RMSE3<-sqrt(mean((y3-SlopeNGDI4$Slope)^2))
RMSE3

##########          MODEL PLOTS          ##########

par(mfrow = c(1, 3))

plot(dat$DI,dat$Slope,col="black",xlab="DI",ylab="Profile Slope",main="DNase
I Degradation Model",
     xlim = c(0,60),ylim=c(-0.018,0),pch=16,mtext(c("A"),side=3,line=2,at=-
2,font=2))
points(SlopeNGDI2$DI,y,col="red",pch=16)
points(SlopeNGDI2$DI,SlopeNGDI2$Slope, col="blue",pch=16)
legend(15,0.0005, legend=c("Estimated Heat/Humidity","Heat/Humidity","DNase
I"),col=c("red","blue","black"),pch=c(16,16,16,16),
       bg="NA",box.lty=0,y.intersp=1,x.intersp=0.2)

plot(dat2$DI,dat2$Slope,col="black",xlab="DI",ylab="Profile Slope",main="UV
Degradation Model",
     xlim = c(0,60),ylim=c(-0.018,0),pch=16,mtext(c("B"),side=3,line=2,at=-
2,font=2))
points(SlopeNGDI3$DI,y2,col="red",pch=16)
points(SlopeNGDI3$DI,SlopeNGDI3$Slope, col="blue",pch=16)
legend(15,0.0005, legend=c("Estimated
Heat/Humidity","Heat/Humidity","UV"),col=c("red","blue","black"),pch=c(16,16,
16,16),
       bg="NA",box.lty=0,y.intersp=1,x.intersp=0.2)

plot(dat3$DI,dat3$Slope,col="black",xlab="DI",ylab="Profile
Slope",main="DNase I and UV Degradation Model",
     xlim = c(0,60),ylim=c(-0.018,0),pch=16,mtext(c("C"),side=3,line=2,at=-
2,font=2))
points(SlopeNGDI4$DI,y3,col="red",pch=16)
points(SlopeNGDI4$DI,SlopeNGDI4$Slope, col="blue",pch=16)
```

```
legend(15,0.0005, legend=c("Estimated Heat/Humidity","Heat/Humidity","DNase &
UV"),col=c("red","blue","black"),pch=c(16,16,16,16),
       bg="NA",box.lty=0,y.intersp=1,x.intersp=0.2)
```