University of Central Oklahoma

Jackson College of Graduate Studies

BEYOND DNA: AN EPIGENETIC APPROACH TO IDENTICAL TWIN IDENTIFICATION

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE IN FORENSIC SCIENCE – BIOLOGY/CHEMISTRY

By

COLTAN SCRIVNER
Edmond, Oklahoma
2016

BEYOND DNA: AN EPIGENETIC APPROACH TO IDENTICAL TWIN IDENTIFICATION

By: COLTAN SCRIVNER

A THESIS

APPROVED FOR THE W. ROGER WEBB FORENSIC SCIENCE INSTITUTE

July 2016

By _____

Dr. Dwight E. Adams Committee Chairperson

_____

Dr. Brandt Cassidy Committee Member

_____

Dr. James Creecy Committee Member

Acknowledgements

I would first like to express my gratitude to my advisor and thesis committee chair, Dr. James Creecy, for his guidance, invaluable constructive criticism, and encouragement throughout the entire process of this project. His familiarity with the forensic science community as well as his genetic expertise helped make this project more relevant and robust. I would also like to thank the other two members of my thesis committee, Dr. Dwight Adams and Dr. Brandt Cassidy, for committing their time to be a part of this project. I would also like to thank Keylie Gibson for her assistance with every aspect of the project. Special thanks also go to the twins who participated in the project. Finally, I would like to thank my friends and family for their encouragement and support during my entire Master's degree.

Table of Contents

List of Figures

List of Tables

Abstract

For more than two decades, DNA analysis has helped forensic scientists link suspects to a crime. Often times, DNA evidence is one of the most impactful pieces of evidence available. However, there is still one thing that traditional DNA analysis cannot accomplish – differentiating DNA from identical twins. With identical twins becoming more common than in the past and with numerous examples of cases being dropped because an identical twin was implicated, it would benefit the forensic science community to find a solution to this problem. The goal of this project was to see if the conventional forensic science techniques of cycle sequencing and capillary electrophoresis could be used to distinguish twins via DNA methylation analysis. It was found that the use of cycle sequencing and capillary electrophoresis for the analysis of DNA methylation extracted from human cells was problematic. While small successes were achieved in analyzing the methylation, the results were not consistent. Thus, while cycle sequencing and capillary electrophoresis are convenient and cost efficient for the forensic science community, they may not the best instruments for this problem. The PRKCA locus was shown to be a strong candidate locus that could be targeted by cycle sequencing or high-throughput sequencing technology. Therefore, rather than using an expensive and time-consuming method such as ultra-deep next generation sequencing to differentiate identical twins, the forensic science community should identify several key loci, such as the PRKCA gene analyzed in this study, for DNA methylation analysis.

Chapter 1: Introduction

A Need Within the Field of Forensic DNA Analysis

DNA analysis has emerged as the standard method for determining human identity. Along with fingerprints, DNA evidence serves as a reliable source of individual identification in numerous forensic cases. While the fingerprints of monozygotic, or identical, twins are known to be different, it is relatively easy to prevent leaving a fingerprint at a crime scene by wearing gloves. DNA on the other hand can be left behind in the form of hair, skin cells, and other body tissues by even the most careful of perpetrators. If the short tandem repeat (STR) profile of DNA taken from a crime scene matches the STR profile of a monozygotic twin, the evidence is significantly less impactful because the other identical twin will also display the same STR profile.

The value of devoting resources to resolving the issue of monozygotic (MZ) twin DNA in forensic casework may be called into question by assuming that the number of monozygotic twins is minimal. However, approximately 1 in 167 people born are a monozygotic twin (Bortolus et al., 1999). The Federal Bureau of Investigation (FBI) recently reported that the Combined DNA Index System (CODIS), the national database for STR profiles, maintains over 11 million unique STR profiles after only two decades of operation. With 1 in 167 individuals being born a monozygotic twin, and over 11 million profiles in CODIS, it is likely that more than 30,000 profiles in CODIS belong to a MZ twin. Furthermore, the fact that a defendant is an identical twin is often enough to cast reasonable doubt in court. This would be

particularly true in cases where DNA is one of the only pieces of evidence that is able to link a suspect to a crime.

Discrimination between MZ twins is one of the only areas in which traditional STR analysis cannot be utilized for individualization. To overcome this impediment, a characteristic of the DNA molecule that is unique even to monozygotic twins must be characterized and analyzed. I propose to address this need for a genetic-based method of MZ twin discrimination by analyzing the differences in cytosine methylation at specific genomic loci between monozygotic siblings.

Purpose of This Study

The interrogation of variably methylated loci in DNA shows promise as a discriminatory method for monozygotic twins in forensic casework. Monozygotic twins, colloquially referred to as "identical" twins, arise from the same fertilized egg and later split into two embryos. Consequently, they share the same genome, barring any mutations after the zygote splits. This is in contrast to dizygotic or "fraternal" twins, which arise from separate eggs and share the same DNA similarities as other siblings. Dizygotic twins involved in a forensic case, like other siblings, are distinguishable via traditional STR analysis. However, there is currently no forensic genetic protocol in place to handle a case in which one or both MZ twins are suspects. DNA evidence is less impactful in these cases, as STR analysis cannot discriminate between the DNA from MZ twins. However, if highly variable methylation patterns are determined, a means for distinguishing between twins will be tenable.

Though DNA methylation is relatively new to forensic science, it has been a mainstay in molecular biology and biomedicine for many years. This rapid influx of information is quickly advancing knowledge of DNA methylation. If genetic loci where methylation varies between MZ twins can be determined, then the variation in methylation may be able to be utilized as a discriminatory method. Because STR would allow for inclusion of only the two MZ twins, methylation analysis does not need to be as discriminatory, and would serve as a supplemental, comparative test. Thus, the aim of this thesis is to investigate and introduce an epigenetic method by which MZ twins may be distinguished following STR analysis.

Process and Function of DNA Methylation

The methylation of the DNA base cytosine to 5-methylcytosine (5mC) is one of the principle ways in which eukaryote genomes interacts with the environment and tends to tissue specific needs (Frumkin, Wasserstrom, Budowle, & Davidson, 2011; Langevin et al., 2011; Meaney & Szyf, 2005; Varley et al., 2013). While working on nucleic acid separation by paper chromatography, Hotchkiss (1948) was the first to suggest that 5mC (then called "epi-cytosine") was a pre-existing base in DNA and not an artifact of experimentation. It would be nearly three decades later before Holliday and Pugh (1975) propose that DNA methylation might regulate gene expression in eukaryotes. Since then, studies on DNA methylation, and more broadly the epigenome – the heritable collection of chemical alterations to DNA and histones – have elucidated many aspects of gene expression (Song et al., 2005; Weber et al., 2007; Yang et al., 2014; Zhao et al., 2009).

Conversion of cytosine to 5mC occurs in the catalytic presence of DNA methyltransferase (DNMT). CpG sites (cytosine followed by a guanine) are the primary source of 5mC, with approximately 70% of the 28 million cytosines in CpG sites being methylated (Chen & Riggs, 2011; Smith & Meissner, 2013). Roughly 10-20% of CpGs can be found aggregated in clusters referred to as CpG islands (Smith & Meissner, 2013; van Dongen et al., 2014). CpG islands (CGIs) are stretches of CG rich sequences averaging 1000bp long that are commonly found in the promoter region of genes (often housekeeping and developmentally regulated genes). Unlike the hypermethylated CpG dinucleotide sites scattered throughout the genome, CGIs tend to be hypomethylated (Deaton & Bird, 2011; P. a Jones & Takai, 2001). Saxonov et al. (2005) found that over 70% of promoters in the human genome lie within a CGI. Furthermore, the hypermethylation of CGIs presents a well-documented correlation with gene silencing, though the exact direction of causality is disputed (De Smet, Lurquin, Lethé, Martelange, & Boon, 1999; Gonzalez-Zulueta et al., 1995; P. A. Jones, 2012). The complex and enigmatic relationship between DNA methylation and gene silencing has become more evident, however, the exact mechanisms involved in methylation are still being investigated. With ever increasing research on the epigenome, particularly in the field of cancer biology, our knowledge of DNA methylation and the impact these modifications have on the gene expression is poised for growth.

Maintenance and *de novo* Methylation

There are two systems of DNA methylation: maintenance and *de novo*. Maintaining methylation patterns is the function of DNA methyltransferase 1

(DNMT1). DNMT1 is responsible for the methylation of newly synthesized DNA strands during mitosis, ensuring that previously methylated cytosines are correctly methylated on newly synthesized DNA. While DNMT1 cannot be excluded from functioning in *de novo* methylation, it has a much higher likelihood of participating in maintenance methylation (Bestor, 2000). The current understanding is that DNMT1 is attracted to hemimethylated DNA (where only one strand is methylated) via a protein called Ubiquitin-like PHD and RING finger domain 1 (UHRF1). UHRF1 preferentially binds hemimethylated DNA to its DNA methyl-binding domain (Bostick et al., 2007). Sharif et al. (2007) has shown that mice containing a deletion of both copies of the UHRF1 gene fail to maintain global DNA methylation, providing further evidence for the existence of a link between DMNT1, UHRF1, and hemimethylated DNA.

De novo methylation involves the establishment of new DNA methylation. Though this can occur throughout life, most *de novo* methylation occurs during embryogenesis, where global DNA methylation patterns are established (see Law & Jacobsen, 2010 for review). *De novo* methylation occurs during embryogenesis because primordial germ cells (PGCs), the cells that give rise to the gametes in the developing organism, are largely demethylated during embryogenesis. This demethylation process is known as reprogramming. Reprogramming is a necessary part of development, as it permits cells to become pluripotent, allowing them to develop into any of the cells in the body. However, some research suggests that DNA methylation patterns may somehow escape this erasure. Much of the work has been

done on rats, where the father's diet impacts the epigenetics and gene expression of the offspring (Carone et al., 2010; Ng et al., 2010; Wei et al., 2014).

Transgenerational Epigenetic Inheritance

Another exception to global demethylation during embryogenesis has been noted in transposable elements. Transposable elements (TEs) are sequences of DNA that are capable of moving around in the genome, giving them the nickname "jumping genes." Intracisternal A-Particles (IAPs) – the most recent transposable elements in the mouse genome – are largely unaffected by demethylation (Lane et al., 2003). Additionally, CGIs close to IAPs were less affected by de-methylation, suggesting a spatial mechanism of action (Lanlan Shen & Waterland, 2007). Indeed, these areas near the IAP that are resistant to the global demethylation may serve as a means of epigenetic inheritance in mice (Lane et al., 2003; Lanlan Shen & Waterland, 2007). Because these TEs are active, it is possible that their persisting methylation status is employed to help prevent transposition mutations.

More recently, Dias and Ressler (2014) provide strong evidence of specific exposure in male mice being linked to specific gene expression alterations in the offspring. In the study, male mice were exposed to acetophenone in conjunction with an electric shock, conditioning the mice to fear the smell of acetophenone. Females were then impregnated via in vitro fertilization to ensure that the males had no social contact with the females. The resulting offspring – and their subsequent offspring – had abnormally fearful reactions to the smell of acetophenone. This observation was complemented by the fact that those offspring also showed increased neuroanatomical pathways for the receptor upon which

acetophenone acts. Further analysis on the epigenetic level revealed that the gene that codes for the receptor, Olfr151, was hypomethylated in the offspring. Thus, evidence on the behavioral, anatomical, and epigenetic level all point to a transmission of epigenetic information across generations, an event known as transgenerational epigenetic inheritance.

For the past decade, transgenerational epigenetic inheritance has largely been studied in mice and rats. However, a recent study by Donkin et al. (2015) has demonstrated that obese men have significantly different DNA methylation of sperm than lean men on 9000 different genes, over 250 of which have been associated with appetite control. Though preliminary, this lends strong evidence for the epigenetic modification of germ cells through lifestyle. Taken in light of the mouse studies previously mentioned, there seems to be strong evidence for transgenerational epigenetic inheritance, possibly even in humans.

Demethylation During Life

As well as occurring during embryogenesis, demethylation may also occur throughout life. Several studies have indicated a connection between age and methylation (Fraga et al., 2005; Heyn et al., 2013; Jintaridth & Mutirangura, 2010; Weidner et al., 2014; C. C. Y. Wong et al., 2010). Interestingly, some studies revealed a negative correlation between aging and methylation status of TEs (Bollati et al., 2009; Jintaridth & Mutirangura, 2010). It is suspected that TEs undergo hypomethylation with age due to the downregulation of DNMT1, the maintenance methyltransferase (Casillas, Lopatina, Andrews, & Tollefsbol, 2003). However, it is

not yet understood whether aging impacts methylation or methylation impacts aging.

Hypomethylation over time suggests a mechanism of demethylation during life. Global demethylation may occur through a passive mechanism in which DNMT1 is downregulated, but it is unlikely that this accounts for all instances of DNA methylation during embryogenesis and throughout life. Only in the past few years have experiments been performed that demonstrate a probable mechanism of active DNA demethylation. One possible method of active demethylation involves the deamination of 5mC to thymine via activation-induced deaminase (AID); following deamination, a G/T pairing exists, prompting thymine DNA glycosylase (TDG) to replace thymine with a new unmethylated cytosine (Franchini et al., 2014). In addition to base excision, TDG may also play a regulatory role in deamination. Cortellino et al. (2011) has shown that mice without TDG do not display deamination induced transition mutations (5mC → T). The absence of transition mutations implies that TGD is important not only in excising thymine, but also the deamination process and indeed the very initiation of DNA demethylation.

In addition to deamination, base excision repair may occur via hydroxylation of 5mC to 5-hydroxymethylcytosine (5hmC) by ten-eleven translocation (TET) - enzymes (Guo, Su, Zhong, Ming, & Song, 2011). Furthermore, TET enzymes may continue to oxidize 5hmC to 5-formylcytosine (5fC) and again to 5-carboxylcytosine (5caC) (Li Shen et al., 2013). Figure 1.1 demonstrates the chemical pathway that can occur via TET. Once 5mC has been converted to either 5fC or 5caC, TDG may excise the base and insert a new cytosine in a similar process as occurs during AID assisted

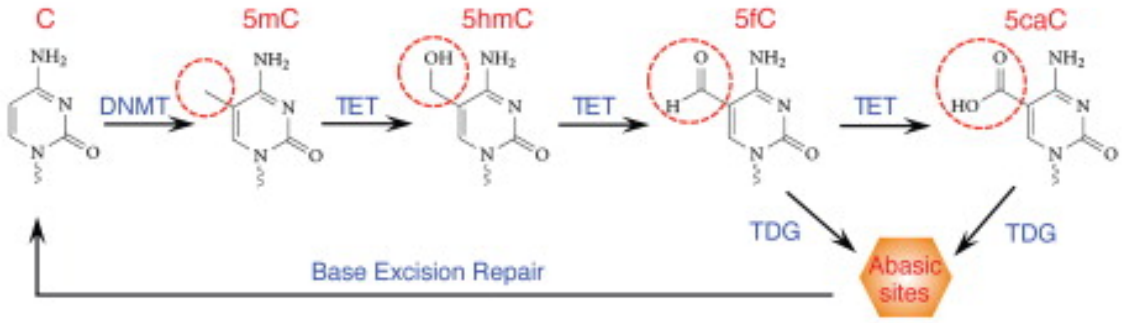


Figure 1.1 – TET Oxidation. Active demethylation may occur through the consecutive oxidation of 5mC to 5fC or 5caC. Once oxidized, base excision can occur through the TDG enzyme. The nucleotide slot will be filled with a new, unmethylated cytosine. Figure borrowed from Li Shen et al. (2013).

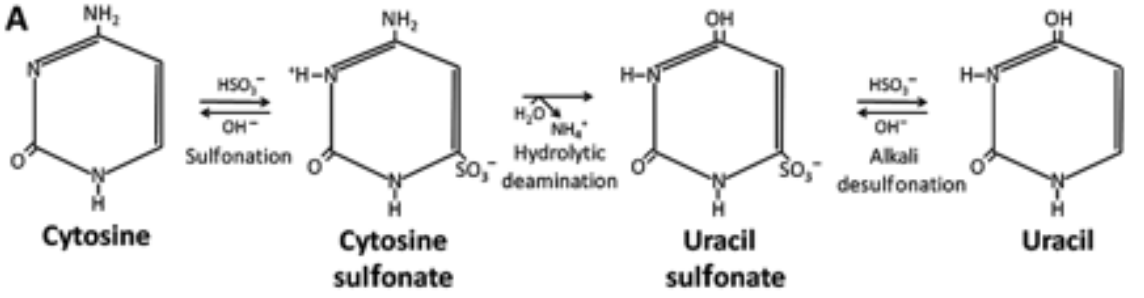
base excision. Though it seems intuitive that DNA demethylation must occur, and potential enzymes involved in demethylation such as AID, TDG and TET have been identified, the details of the molecular mechanisms have yet to be disclosed.

Bisulfite Conversion

Methods for the characterization of DNA methylation vary greatly depending upon the research question. Researchers may want to probe specific CpG islands for transcription studies or even whole genome methylation for twin and age-related studies. With the advent of high-throughput sequencing and the subsequent advances in DNA sequencing technology, whole genome methylation studies – or methylome studies – are now possible. A review of every methylation analysis is beyond the scope of this paper. For reviews on methylation analysis methods, see Fraga and Esteller (2002) and Shen and Waterland (2007). At any rate, the underlying technique for virtually every 5mC analysis method, regardless of the research question, is bisulfite conversion.

Sodium bisulfite sequencing was first described by Frommer et al. (1992) as a positive identification method for 5mC in DNA. The chemical process of bisulfite

conversion is shown in Figure 1.2. Unmethylated cytosines are first sulfonated in the presence of bisulfite (HSO_3^-), followed by the deamination of sulfonated cytosine to sulfonated uracil, and eventual desulfonation of uracil. During sequencing, the newly synthesized uracil is read as thymine, thus effectively distinguishing 5mC from unmethylated cytosine. Bisulfite conversion of cytosine to uracil is the key process underlying most DNA methylation analyses and is considered the "gold standard" (Clark, Statham, Stirzaker, Molloy, & Frommer, 2006).



B

	Original sequence	Sequence after bisulfite treatment
Unmethylated DNA	A-T-**C**-**G**-G-T-C-A-T-**C**-**G**-C-A-T	A-T-**U**-**G**-G-T-U-A-T-**U**-**G**-U-A-T
Methylated DNA	A-T-**C**-**G**-G-T-C-A-T-**C**-**G**-C-A-T	A-T-**C**-**G**-G-T-U-A-T-**C**-**G**-U-A-T

Figure 1.2 – Conversion of cytosine to uracil via sodium bisulfite. Bisulfite conversion is based off chemically deriving uracil from an unmethylated cytosine. Following conversion, 5mCs are positively identified by PCR or direct sequencing, wherein they will remain as "C's." Figure from Kristensen and Hansen (2009).

Chapter 2: Forensic Application of DNA Methylation

In accordance with Federal Rule of Evidence 702, any scientific testimony in a court of law must be based on sufficient facts or data that are derived from reliable principles and methods that are applied to the facts of the case. Consequently, forensic science in the United States must verify scientific procedures before they may be used in court, a task that was historically assumed by the FBI laboratory and now performed by the companies that manufacture DNA analysis reagents. Although epigenetics is an emerging discipline, the application of DNA methylation to forensics was proposed as far back as 1993 for an alternative technique for sex determination (Naito, Dewa, Yamanouchi, Takagi, & Kominami, 1993). Since then, methylation has been studied for a variety of forensic applications, including: phenotype, parental origin of alleles, DNA authentication, cause of death, age, paternity testing, and body fluid identification. While the discrimination of MZ twins by methylation is largely uninvestigated, many of the researched applications listed above have supplied important information for a theoretical basis of MZ twin discrimination. Two of the most studied applications in forensic science involve tissue identification and determination of phenotype/age.

Tissue Identification

The identification of tissue via DNA methylation has been one of the most heavily researched applications of methylation in forensic science (see the following studies: Frumkin et al., 2011; Gomes, Kohlmeier, & Schneider, 2011; Ma, Yi, Huang, Mei, & Yang, 2013; Park et al., 2014). The proper identification of tissue/fluid at a

crime scene is an important part of the investigative process, as it can reveal a wealth of information about the crime. Currently, most fluid identification techniques rely on presumptive and confirmatory tests, wherein a presumptive test guides further confirmatory testing that ultimately identifies the body fluid. A methylation based technique, however, could be universally applied to tissue/fluid samples and identify them in a single step, thus amalgamating the repertoire of tests that are currently used into a single test procedure. This discrimination is possible because DNA methylation, being implicated in gene regulation, is variable among tissues (Kitamura et al., 2007; Song et al., 2005). Different tissues perform different functions, thus they exhibit differential methylation.

Phenotype/Aging

The epigenome can be seen as an interface between the environment and the genome; a sort of link between the age-old debate of nature vs. nurture. This avenue of research offers perhaps the most intriguing potential. Relating genetics to phenotype has been a popular topic of many crime shows and is gaining research support in the scientific community. However, genetic sequences alone cannot account for all of phenotypic diversity. If this were true, a pair of identical twins would be completely indistinguishable.

In regards to genetically identical animals, slight phenotypic differences are likely attributed to epigenetic adaptations to environmental exposure. Studies on diet and smoking are some of the most common studies that back this view, often citing negative effects due to gene silencing (see Oka et al., 2009; Hasegawa et al., 2002; Milagro et al., 2009; Singh, Murphy, & O'Reilly, 2003). However,

environmental exposure does not account for all epigenetic variation. Methylation differences have been shown to exist in twins that are raised together, suggesting that stochastic events may play a larger part in differential methylation than previously thought (D'Aquila, Rose, Bellizzi, & Passarino, 2013; A. H. C. Wong, Gottesman, & Petronis, 2005).

Strong correlations between epigenetic divergence and age in MZ twins have been drawn, so much as to propose a possible mechanism for age determination through a DNA sample (D'Aquila et al., 2013; Koch & Wagner, 2011). Indeed, Bocklandt et al. (2011) constructed a methylation-based model that was able to predict the age of 66 individuals, +/- 5.2 years. This demonstrates the potential power of utilizing DNA methylation in forensic studies. Furthermore, MZ twins offer an excellent opportunity to study the stochastic effects on DNA methylation due to the fact that they are often raised in the same environment and are genetically indistinguishable.

Proposed Use of DNA Methylation for MZ Twin Differentiation

Knowledge gleaned from the previously described forensic applications of DNA methylation provide a strong theoretical basis from which MZ twin discrimination may be derived. Studies in phenotype and aging have shown that methylation is highly specific to individuals. Even when reared in the same environment, MZ twins will show epigenetic differences. Indeed, even at the moment of birth, MZ twins have been shown to exhibit discordance in DNA methylation (Gordon et al., 2012). The converging evidence for epigenetic

differences among MZ twins is foundational for a protocol involving MZ twin discrimination.

Regarding tissue/fluid identification, studies have offered a caveat that must be taken into consideration when methylation profiling MZ twins. Owing to the functional specificity of tissues, gene expression and, consequently, DNA methylation, is tissue specific. This characteristic may raise issues with DNA samples from crime scenes, however this will be discussed towards the end of the review along with other potential issues. Fortunately, the issue is not insurmountable, and may prove to be an inconvenience at worst.

Many studies involving methylation-based tissue identification and methylation-based age/phenotype prediction have been performed (see Frumkin et al., 2011; Gomes et al., 2011; Koch & Wagner, 2011; A. H. C. Wong et al., 2005). Several of these studies involve MZ twins as subjects for greater control over genetic and environmental factors. Previous research on the forensic application of DNA methylation has provided ample grounds for an investigation in the use of DNA methylation in MZ twin discrimination. Moreover, the inability for the DNA analyst to distinguish between MZ twins has produced a need for the implementation of a genetic-based protocol for MZ twin discrimination. However, before any attempt to create a biometric based on methylation, there must first be an evaluation of what a good biometric entails.

Biometrics

The use of human identification methods, or biometrics, has been essential to the success of forensic science. Biometrics use individual characteristics,

characteristics that are unique to an individual, to link a question sample with a reference sample. This often involves the recognition of some physiological pattern, such as fingerprint morphology or STRs (A. Jain, Hong, & Pankanti, 2000). Although individual characteristics cannot prove guilt or innocence, they can provide very strong evidence for a case. A variety of biometric methods are commonly used by forensic science laboratories today, including nuclear and mitochondrial DNA analysis, fingerprinting, palm printing, facial recognition, and iris recognition (Federal Bureau of Investigation, "Fingerprints and other biometric methods").

The history of biometrics began in the late 1800's with a French police official named Alphonse Bertillon. Bertillon devised a system of 11 body measurements that, purportedly, could identify individuals (Cole, 2001). However, the use of the Bertillon method was short-lived. In an unusually coincidental 1903 case, two men who looked very similar, both named Will West, displayed the same Bertillon measurements. The one biometric that would distinguish them, and thus distinguish itself as the gold standard of identification, was fingerprinting (Hoover, 1961). Fingerprinting reigned as the chief forensic identification method for over 80 years. However, after Jeffreys et al. (1985) described a method of identification with tandem-repetitive hypervariable genomic loci, DNA "fingerprints" quickly established prominence alongside traditional fingerprints.

In the United States, DNA profiling involves the analysis of 13 multi-allelic short tandem repeats (STRs), also known as microsatellites (Butler, 2006). In STR analysis, nuclear DNA is extracted and isolated from a biological sample before being amplified via polymerase chain reaction (PCR). Invented in 1983 by Dr. Kary

Mullis, PCR is a seminal technology in molecular biology in which a target sequence of sample DNA is amplified using DNA polymerase, excess nucleotides, buffer, and primers inside of a thermocycler. The idea behind this process is to mimic the natural conditions of the cell under which DNA replicates. Many STR amplification kits use primers that contain fluorescent dyes so that each amplified loci may be detected via color and size during electrophoresis. During STR amplification, the 13 STRs, also called core loci, are preferentially amplified to produce billions of copies.

After amplification, the DNA fragments are electrophoretically separated, generally using capillary electrophoresis, and identified by their dye color and size. Each locus has a variable number of possible alleles (repeat units) that, when taken together, create a statistically unique profile for identification purposes (Office of the Inspector General, 2004). The use of biometrics, chiefly the identification methods of STR and fingerprinting, has revolutionized forensic science by allowing for the identification of individuals under wide variety of circumstances, including criminal, paternal, and humanitarian cases.

Biometric Properties

Jain, Bolle, & Pankanti (1999) previously described four necessary properties that define an effective biometric. 1) The measurement must exist *universally*. It is essential that the biometric be represented in some way in everyone. 2) Each profile needs to be *unique*. The uniqueness of profiles allows for individualization. 3) A degree of *permanence* is essential. Because the metric will identify the individual throughout their lifetime, it must persist throughout the entirety of their life. 4) The

final property that is essential for a biometric is *collectability*. The trait must exist in some quantifiably collectable fashion.

Jain, Bolle, & Pankanti (1999) also suggest three additional characteristics that are desirable in a biometric. 1) The *performance* of the method must be feasible, accurate, and reliable. This characteristic is particularly suitable for biometrics in forensic science, wherein methods must be scientifically validated before they can be used in a court of law. 2) People should generally *accept* the system. A controversial system will prove problematic in legal cases. 3) Lastly, a metric would be more reliable if there exists a low possibility of *circumvention*. That is, it should be difficult to commit a fraudulent act with the system. Any attempt at identifying a novel biometric should comply with the properties previously described.

DNA Methylation as a Biometric for MZ Twins

The identification of a MZ twin via DNA sample will utilize the traditional STR method along with the proposed novel method involving the variably methylated loci. In order to evaluate the effectiveness of the proposed method, the procedure will be assessed according to the necessary and desirable characteristics delineated above. As a necessary component of development and gene expression, DNA methylation is a universal trait, satisfying the first tenet of an effective biometric. The fourth tenet, collectability, is also satisfied by DNA methylation. DNA methylation is acquired in tandem with normal DNA collection, making its collectability equal to that of DNA collection. The third tenet regarding permanence may only be partially satisfied. While DNA methylation will always exist in a genome, the epigenome is a dynamic structure that is subject to change throughout

life. Fortunately, however, studies have shown that DNA methylation patterns exhibit no significant difference over the course of 12 weeks (Fraga et al., 2005). Furthermore, older MZ twins actually exhibit greater discordance than younger pairs, meaning methyl profiles become more distinguishable from one another as MZ twins age (Fraga et al., 2005; Talens et al., 2012). It should be noted that the dynamic nature of the epigenome may not occur uniformly, but may instead occur more often in certain areas of the genome, leading to a change in overall global methylation (Bollati et al., 2009; Ekram, Kang, Kim & Kim, 2012; Ziller et al., 2013).

DNA methylation has been shown to exhibit a higher level of discordance between unrelated individuals and dizygotic (DZ) twins than between MZ twins (Coolen et al., 2011; Kaminsky et al., 2009). This discordance is intuitive given that not all CpG sites would be the same among individuals, whereas monozygotic twins would share all CpG sites. In other words, a CpG site in one individual may not be methylated in another individual because the individual may not contain that CpG site in the genome. This brings us to the second tenet of an effective biometric: uniqueness. The proposed method of discrimination relies on STR analysis in conjunction with 5mC interrogation. STR analysis in and of itself is an identification technique, with discriminatory power well beyond global population. The most obvious pitfall of STR analysis is that regarding two individuals who share all amplified loci in a DNA profile, i.e., MZ twins. STR analysis is able to discriminate down to a single profile that is shared between two individuals. The addition methylation profiling to STR analysis creates a level of uniqueness, even among

individuals with the same DNA. Thus, methylation profiling, used in tandem with STR, provides ample uniqueness for forensic testing, and satisfies the second tenet.

Furthermore, methylation profiling is poised to meet all three of the desirable traits in a biometric, that is: performance, acceptance, and a low level of circumvention. Bisulfite conversion is the foundation for DNA methylation studies. As such, the technique exists in a variety of experiments and is used in many fields. Moreover, bisulfite conversion chemistries, considered the "gold standard" in DNA methylation research, have proven to have high yield products (Holmes et al., 2014). Hence, methylation profiling should easily satisfy the performance and acceptance characteristics. Little research has been performed on the level of circumvention required to manipulate a sample, though it is worth noting that both methylation and demethylation are enzymatic processes, and DNA methylation status is increasingly unlikely to be the victim of a fraudulent act. In fact, the presence of methylation has been proven to be effective in authenticating DNA, as synthetic DNA contains no methylation (Frumkin, Wasserstrom, Davidson, & Grafit, 2010). Given its fulfillment of the seven aforementioned tenets of a good biometric, as well as its use in age prediction and tissue identification, DNA methylation is in a good position to be evaluated as a biometric for MZ twins.

Chapter 3: DNA Methylation Variability

It is important to understand that methylation profiling alone would not serve as an effective means of individualization. Though this may be theoretically possible, the niche of genetic individualization is filled by STR analysis. However, methylation profiling will prove to be a necessary addition to STR analyses that includes a MZ twin. When considering power of discrimination in the case of methylation profiling, it becomes clear that the power of discrimination needs not to be in the realm of STR analysis. Methylation analysis aims only to discriminate between two individuals who have already been identified to the exclusion of all other individuals. Thus, methylation profiling requires not a level of discrimination on the order of billions or trillions. Indeed, methylation profiling would be a relatively simple comparison technique. Still, highly variable regions should be selected as candidate loci in order to increase the odds of a difference being found. To appropriately select these loci, methylation variability in different genomic regions must be explored.

Variability Between Individuals

Studies on DNA methylation have suggested that methylation status is largely the product of environmental factors and stochastic genetic effects (Fraga et al., 2005; Gordon et al., 2012; Kaminsky et al., 2009). Still, MZ twins have shown higher concordance than DZ twins, suggesting a possible genetic component to methylation (Gervin et al., 2011). A genetic component seems logical, considering not all individuals would have the same CpG sites in their genomes. Thus, a child who

inherited the same genetic sequence containing the same CpG sites as a parent would show more DNA methylation concordance due to having the same locations that can be methylated. Nevertheless, MZ twins have been shown to exhibit discordance in methylation at every age, with many studies showing a positive correlation between age and methylation discordance (Fraga et al., 2005; Gervin et al., 2011; Martino et al., 2013; van Dongen et al., 2014). These studies support the idea that DNA methylation arises from a complex interplay between genetics, the environment, and stochastic events. The element of randomness that is associated with DNA methylation is ideal for distinguishing between two individuals with identical or nearly identical genomes.

Potential Loci for Methylation Profiling

Transposable elements (TEs) are a possible candidate for methylation profiling in twin studies. Studies in mice have detected a high degree of variation among evolutionarily recent TEs in isogenic mice (Ekram, Kang, Kim, & Kim, 2012). Moreover, the methylation status of the TEs has been shown to vary depending on nutrition and environmental factors (Waterland & Jirtle, 2003; Wolff, Kodell, Moore, & Cooney, 1998). This same epigenetic trend is likely to be found in human TEs.

Approximately half of the human genome is comprised of TEs, which also happen to be high in GC content (International Human Genome Sequencing Consortium, 2001). Only a minute percentage (<0.05%) of TEs are active, including *Alu*, L1, SVA, and potentially HERV-K (Mills, Bennett, Iskow, & Devine, 2007). *Alu* repeats, totaling over 1 million copies in the human genome, are the most active TE (Bennett et al., 2008). In line with findings on active TEs in mice, *Alu* repeats have

been shown to be hot spots for methylation variability. In an important study by Fraga et al. (2005), MZ twins were shown to have highly variable methylation among *Alu* repeats. Figure 3.1 demonstrates methylation discordance among twins. Although discordance was much greater in older twins and exhibited a positive correlation with age, no significant difference could be detected between methylation measurements after 12 weeks (Fraga et al., 2005). This stability is important for forensic science cases, where reference samples cannot always be obtained quickly. Furthermore, the variance in methylation between twins was observed in many different tissues, including lymphocyte, buccal, adipose, and skeletal muscle.

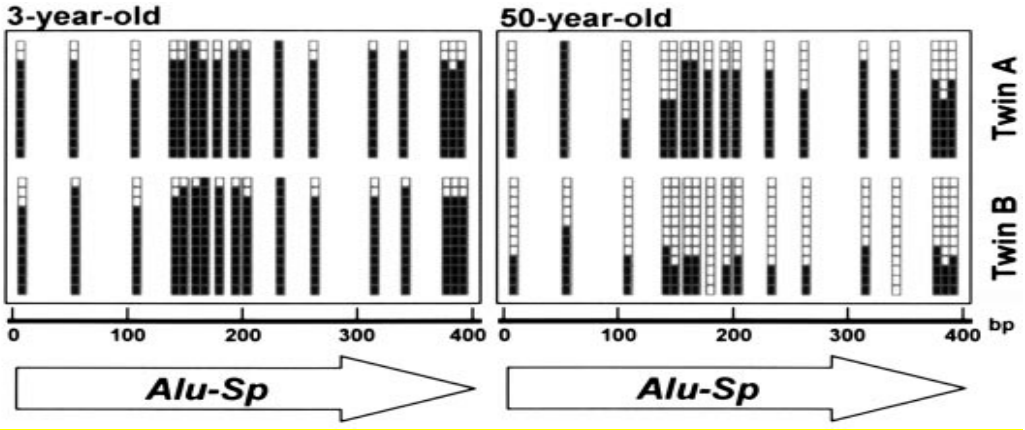


Figure 3.1 – Alu-Sp methylation. Each line contains 12 squares, each representing one of 12 clones of a CpG. Black squares indicate methylated CpGs, while white indicate unmethylated. The discordance with age is obvious, yet even the 3-year-old twins show slight variation within this one region. Figure borrowed from Fraga et al. (2005).

Alu repeats are just one example of a potential candidate for methylation profiling. Other TEs may prove to be equally as capable, though research is lacking on the methylation status of other TE's. *Alu* repeats are particularly attractive candidates because their variation in methylation has already been established. Still, other loci have been investigated for methylation discordance among MZ twins. For

example, a study by Gervin et al. (2011) has shown that the major histocompatibility complex (MHC) is also discordant for methylation among MZ twins. Moreover, the methylation trend among CpG regions of the MHC was an instantiation of the global trend noted by van Dongen et al. (2014), namely higher discordance among MZ twins in non-coding regions.

CpG Islands, Shores, Shelves.

When selecting loci for methylation profiling, global characteristics should be taken into consideration. DNA methylation status is not uniform among CpG regions. While most individual CpG sites are hypermethylated, CGIs, found most often in promoter regions of genes, are typically hypomethylated. The hypomethylation seen in promoter regions is likely due to the role that methylation plays in gene silencing (Gervin et al., 2011; van Dongen et al., 2014). Areas that demonstrate overwhelming hypomethylation or hypermethylation are perhaps not the best loci to interrogate. While CGIs themselves may not be the best loci for methylation profiling, the genetic regions surrounding CGIs might be good candidates.

Genomic regions known as shores and shelves flank CGIs on both sides. Figure 3.2 provides a spatial representation of CGIs, shores, and shelves. These regions, though not functionally defined, lie about 2000 bases from one another. They represent good loci for methylation profiling for two main reasons. Importantly, several studies have shown that variation in methylation tends to be a function of distance from the promoter region (Gordon et al., 2012; Martino et al., 2013; van Dongen et al., 2014). Van Dongen et al. (2014) found that, between MZ twins, non-CGI regions displayed the greatest discordance (Spearman's rho 0.49),

followed by CGI shelves (0.50), shores (0.54), and then CGIs (0.66). In other words, methylation variation among MZ twins increases as you move away from CGIs. These results follow logically from the idea that methylation interferes with transcription. Thus, more functional genomic regions would likely be more epigenetically conserved.

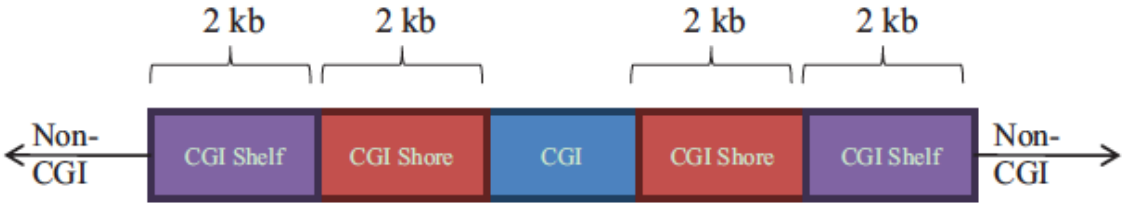


Figure 3.2 – CGI shores and shelves. CGIs are flanked by shores, which are in turn sandwiched between shelves. Each shore and shelf is ~2kb in length. Figure borrowed from van Dongen et al. (2014).

For forensic science purposes, it is most practical to use regions with more possible points of variation. This means two things: 1) the region should contain a relatively high number of possible points of variations and; 2) the region should display a degree of variability between MZ twins. It has already been noted that variation is a function of distance from the promoter region, and thus from the CGI. This implies that a loci located further from the CGI would be most informative for discordance. However, the paucity of CpG sites outside of the CGI means that the selected loci should lie relatively close to the CGI. In an analysis of 59,000 CpG sites, van Dongen et al. (2014) found that CGI shores and shelved displayed the most inter-twin variation, respectively. Although CGI shelves displayed greater variation in the study, they made up only about 10% of CpG sites analyzed. On the other hand, CGI shores demonstrated a good degree of discordance while also comprising about

25% of the CpG sites in the study. Thus, it seems that CGI shores would be the best candidate for methylation profiling given their relatively high number of CpG sites and high degree of discordance among MZ twins.

Finding a Locus

Theoretically, one should be able to scan the genome on a genetic database such as GenBank and find a locus with which to work. Once a gene is found, a CpG island is likely to be found near the promoter region of genes, with shores being approximately 2000 bp upstream. While this "shotgun approach" is plausible, there is no way of knowing whether the specific locus is variably methylated between twins. Fortunately, Lévesque et al. (2014) had previously compiled a list of genes that were hypervariable across individuals (including twins), yet stable for 3-6 months. It can be reasonably inferred that if the loci are hypervariable, then the regions just upstream of them (shores) would also be hypervariable. Thus, a suitable candidate locus for methylation analysis can be drawn from this list. One such example is the Protein Kinase C, Alpha (PRKCA) gene.

PRKCA was searched in GenBank and a location approximately 2000 bp upstream was scanned for CpG sites. A 280 bp candidate amplicon was located at 64297650 – 64297969 on chromosome 7. Both the normal sequence and the bisulfite converted sequence can be seen below with CpG sites highlighted in green and underlined.

Normal Sequence

5' – CTCTGCACCTGACAGTATTGCAGTAATTAGCTTGGGATGCAAAATGATTTCTAAATT

AAGGTCCTGGACTTGTAAAATCAAGAAAGCCTGATAATCAACATCGTCGCTCTGGTGGCC

ATTGAATAACGGGATAAATACCTAAGACAAGTCCTTTAACCATCCCACGTGGTGATAAA

TGCCCGACGTACAAATTAAAAGTCTCTACCACACGCAGGTGAAATAAGGTATAGGTCAA

GAAGAGAAGGTAGGGGAAAGGGTCAATGCTTAATTATTTAAAAAT – 3’

Bisulfite Converted Sequence

5’ – TTTTGTATTTGATAGTATTGTAGTAATTAGTTTGGGATGTAAAATGATTTTTAAAT

TAAGGTTTTGGATTTGTAAAATTAAGAAAGTTTGATAATTAATATCGTCGTTTTGGTGG

TTATTGAATAACGGGATAAATATTTAAGATAAGTTTTTTAATTATTTTACGTGGTGATA

AATGTTCGACGTATAAATTAAAAGTTTTTATTATACGTAGGTGAAATAAGGTATAGGTT

AAGAAGAGAAGGTAGGGGAAAGGGTTAATGTTTAATTATTTAAAAAT – 3’

Primers for the bisulfite converted sequence, including their respective length and approximate melting temperatures according to Bisulfite Primer Seeker (Zymo Research) can be seen in Table 3.1. Because bisulfite converted primer sequences are depleted of Cytosine, longer primers are necessary for higher melting temperatures and greater specificity during PCR.

Direction	Bisulfite Converted Primers (5’ – 3’)	Length	Tm
Forward	TTTTGTATTTGATAGTATTGTAGTAATTAGTTTGGG	36 bp	55.9 °C
Reverse	ATTTTTAAATAATTAAACATTAACCCTTTCCCC	33 bp	55.3 °C

Table 3.1 – Bisulfite converted primers for PRKCA. Primers designed using the Bisulfite Primer Seeker algorithm from Zymo Research.

This sequence is, theoretically, a particularly good amplicon for methylation analysis for a few reasons. First, the amplicon is under 300 bp, which lessens the likelihood that the amplicon will break during degradation, especially during the bisulfite conversion process. Second, the 3’ end of the primers have strong G/C

"clamps" that help keep the primer on the DNA strand. Third, there are seven CpG sites in the amplicon, allowing for modest discriminatory power. Finally, those CpG sites are situated 102 bp (forward) and 70 bp (reverse) from the ends of the sequence. This is important during sequencing analysis, as the first 30-50 bp are often unreadable sequence.

Chapter 4: Cycle Sequencing of Bisulfite Converted DNA

Confirmation of Monozygosity

All experimental methods and protocols were approved by the University of Central Oklahoma's Institutional Review Board prior to subject recruitment and experimentation. Subjects also signed informed consent statements prior to participation. Prior to methylation analysis on twins, monozygosity was confirmed. First, DNA was extracted from a buccal swab using DNeasy Blood and Tissue chemistry (Qiagen 69506). Briefly, 180 μL of Buffer ATL (provided) and 20 μL of Proteinase K were added to the tips of each buccal swab inside of a 2.0 mL tube. Samples were vortexed and heated to 56 °C for 20 minutes. Samples were then vortexed briefly and 200 μL of Buffer AL (provided) was added. Samples were vortexed again before 200 μL of 100% ethanol was added and vortexed. Liquid and swabs were transferred to a DNeasy Mini spin column in a 2.0 mL collection tube (provided).

Samples were centrifuged at 6000 x g for 1 minute. Flow-through and swab were discarded and spin columns were placed in a new 2.0 mL collection tube. Next, 500 μL of Buffer AW1 (provided) was added to each sample. Samples were centrifuged at 6000 x g for 1 minute and flow-through was discarded. Spin columns were added to a new 2.0 mL collection tube and 500 μL of Buffer AW2 (provided) was added. Samples were centrifuged at 20,000 x g for 3 minutes and flow-through was discarded. Spin columns were placed in new 2.0 mL tubes and 200 μL of Buffer AE (provided) was added to the center of each column. Samples were incubated at

room temperature for 1 minute and then centrifuged at 6000 x g for 1 minute to elute.

Extracted DNA was then quantitated using a Nanodrop (Thermo Fisher ND-2000) and amplified using GlobalFiler PCR chemistry (Thermo Fisher 4476135). Briefly, 1 ng of each DNA sample was added to 7.5 μL GlobalFiler Master Mix and 2.5 μL GlobalFiler Primer Set in a PCR tube. Molecular grade water was used to bring each solution to 25 μL. Positive and negative controls were also prepared. Tubes were vortexed, centrifuged, and then placed in a thermocycler with the following conditions: A one minute hold at 95 °C followed by 29 cycles of 94 °C for 10 seconds and 59 °C for 90 seconds. A final extension of 60 °C for 10 minutes was also used.

Ten μL of each sample was mixed with 9.6 μL Hi-Di Formamide (Applied Biosystems 4440753) and 0.4 μL GeneScan 600 LIZ Size Standard v2.0 (Applied Biosystems 4408349) and warmed to 95 °C. Finally, samples were placed on ice for three minutes and ran on a 3500 Genetic Analyzer (Applied Biosystems, 4406017) with POP-7 polymer (Applied Biosystems 4393708) and 36 cm capillary arrays (Applied Biosystems 4404683). Samples were evaluated with the Genemapper ID-X 1.4 (Applied Biosystems). All five sets of twins were confirmed as monozygotic. Allele call sheets for these tests can be seen in Supplemental Tables 1 – 5.

Bisulfite Conversion of Samples

For this study, five pairs of identical twins were recruited to donate buccal swabs and fill out questionnaires. An example of survey questions can be seen in Table 4.1. Buccal swabs were stored at -20 °C until use. Before trying the protocol

on the twin samples, the methylation analysis protocol was optimized on DNA samples from two members of the lab. The purpose of this was twofold. First, this would preserve twin samples for use on an optimized protocol. Second, this would allow the researchers to see if differences existed between unrelated individuals. Furthermore, two different genomic regions with amplicons of different lengths (~900 bp and ~280 bp) were analyzed in this experiment. Primers can be found in Supplemental Table 6.

Age	<18, 18-24, 35-34, 35-44, 45-54, >55				
Sex	Male			Female	
On a scale of 1 to 5, how similar (behaviorally) are you and your twin? 1 being very dissimilar and 5 being very similar.	More dissimilar			More similar	
	1	**2**	**3**	**4**	**5**
For how many total years did you live in the same house as your twin?	<5	5-10	11-15	16-20	>20
Over the past year, how many cigarettes did you smoke per week, on average?	0	Less than 2	3-5	More than 6	
Over the past year, how many alcoholic drinks (8 oz beer, 6 oz wine, 1 oz liquor) did you have per week, on average?	0	Less than 2	3-5	More than 6	
Over the past year, how many days did you exercise (at least 30 minutes) per week, on average?	0	Less than 2	3-5	More than 6	

Table 4.1 – Twin questionnaire. These questions were asked to get an idea of how divergent intra-twin pairs were in lifestyle and to see if greater divergence correlated with greater differences in methylation. Age was also queried to see if older twins were more epigenetically different than younger twins.

Chemistries such as EZ DNA Methylation-Direct (Zymo Research D5020) have been optimized to produce single base methylation resolution from a sample containing as little as 50 pg of DNA. In this experiment, the EZ DNA Methylation-Direct chemistry was used for sample extraction and bisulfite conversion according

to the provided protocol. However, one change was made to the sample in order to accommodate the sample type (buccal swab) for the protocol. Two versions of this change were used throughout the experiment. For the initial change, the tip of the buccal swab was placed in a spin basket inside of a 2 ml collection tube. The basket was filled with DNAse free water to the top of the swab tip (about 450 μL) and spun at low speed in a vacuum centrifuge for 10 minutes. This allowed the cells present in the swab to collect in the eluate. The cell-containing eluate was vortexed and kept at 4 °C until use. The second change will be discussed later alongside its relevant section.

Following collection of cells in the eluate, the EZ DNA Methylation-Direct cell-containing liquid extraction protocol was followed. Briefly, 9 μL of cell-containing liquid was added to 1 μL of proteinase K and 10 μL of the provided digestion buffer in a PCR tube. The mixture was incubated at 50 °C for 20 minutes. Next, 130 μL of previously prepared CT conversion reagent was added to the mixture and vortexed. The new mixture was separated into two PCR tubes, each containing 75 μL. These were then placed in a thermocycler at 98 °C for 8 minutes followed by a 3.5 hour incubation at 64 °C. During this time, bisulfite conversion of the unmethylated cytosines takes place, as shown previously in Figure 1.2.

Bisulfite conversion is a harsh chemical process that can degrade double stranded DNA, particularly through depurination (Ehrich, Zoll, Sur, & van den Boom, 2007). Furthermore, DNA is single stranded and prone to further degradation and non-specific binding following bisulfite conversion. For this reason, amplicons longer than a few hundred base pairs are not often successfully amplified. Results

from this experiment demonstrate this with an amplicon of ~ 900 bp. The original area of interest in this experiment was an ~ 900 bp amplicon in a CGI shore near dopamine receptor D4 (DRD4) on chromosome 11. This location was found by searching GenBank at random for a CGI then moving ~ 2000 bp upstream to the CGI shore.

DRD4 Primer Optimization

As with any new amplification protocol, the primers needed to be optimized. To do this, a Failsafe PCR PreMix Selection Kit (Epicentre FS99060) was used to determine the optimal buffer concentration for PCR. Twelve 25 μL PCR reactions were created using with the following reagents and amounts for each tube: 3.8 μL bisulfite converted DNA, 1.2 μL forward primer, 1.2 μL reverse primer, 0.2 μL AmpliTaq Gold DNA Polymerase (Applied Biosystems N8080248), 3.5 μL of buffer (one of A through L), and 15.1 μL molecular grade water. The reaction had an initial step of 94 °C for 5 minutes and was ran for 30 cycles at the following temperatures and times: 94 °C for 30 seconds, 54 °C for 30 seconds, and 72 °C for 30 seconds. A final step at 72 °C for 7 minutes was followed by a 4 °C hold. A 2% agarose gel was ran with the products of each tube following PCR.

Figure 4.1 displays the agarose gel results of the first amplification event on the DRD4 region using the Failsafe premixes. Twelve different premixes (A-L) with varying salt concentrations and pH were used. The DNA is from one of six buccal swabs taken from the experimenter and prepared as described above. While amplification of this region did seem to take place, it was consistently incomplete.

Bands appeared at about the 230-280 bp range each time with the smear going all the way up to about the 900 bp mark. This falls in line with the 300 bp amplicon limit suggested by Zymo Research (Bisulfite Beginner Guide). As expected, the DNA was very degraded, as shown by the band smears in Figure 4.1. These smears represent partial amplification of some templates during PCR and are fairly common in bisulfite treated DNA due to genomic degradation during the conversion process.

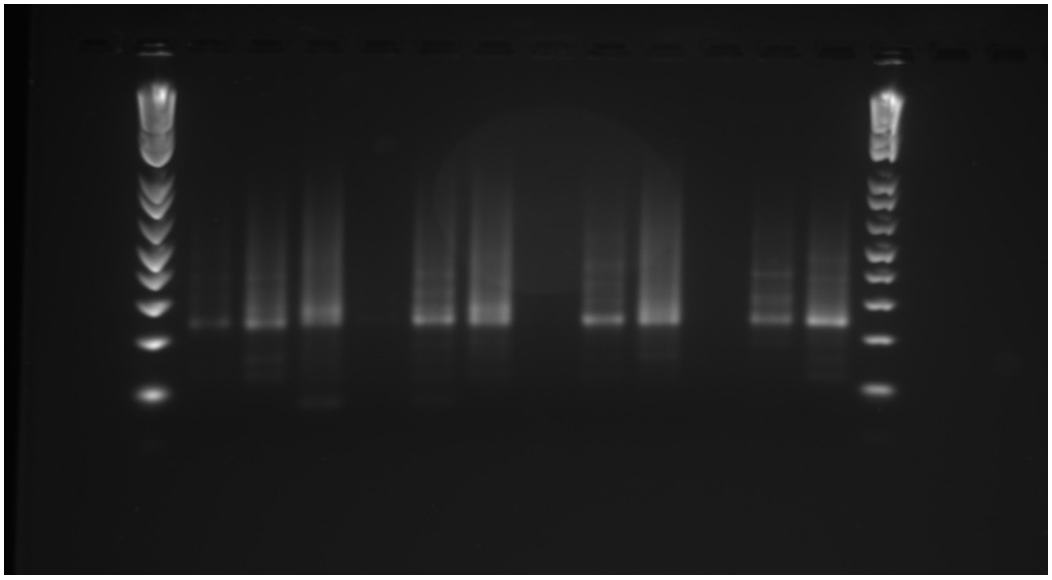


Figure 4.1 – Partially amplified DRD4 region. Buffers are A-L from left to right. The first three ladder bands represent 100, 200, and 300 bp length, respectively. Bands appeared at ~ 240 bp under all conditions except buffers D, G, and H.

Following visualization of the DNA on the gel, the same extraction and conversion protocol was used on Twins 4A, 4B, 5A, and 5B. Instead of using all 12 premix buffers, only the three brightest bands in the previous gel – F, I, and L, - were used for each sample. In an attempt to eliminate the smearing on the gels, touchdown PCR was implemented. The annealing gradient was a 10 °C gradient between 66 °C and 56 °C with a 0.5 °C step down per cycle. PCR conditions were as follows: An initial step of 98 °C for 10 minutes followed by 20 cycles with a 30

second denaturing step at 95 °C, an annealing temperature between 66 °C and 56 °C for 3 minutes, and 72 °C for 2 minutes. A second set of 20 cycles were ran with the following conditions: 95 °C for 30 seconds, 59.4 °C for 3 minutes, and 72 °C for 2 minutes. There was a final step at 72 °C for 7 minutes and a hold at 4 °C.

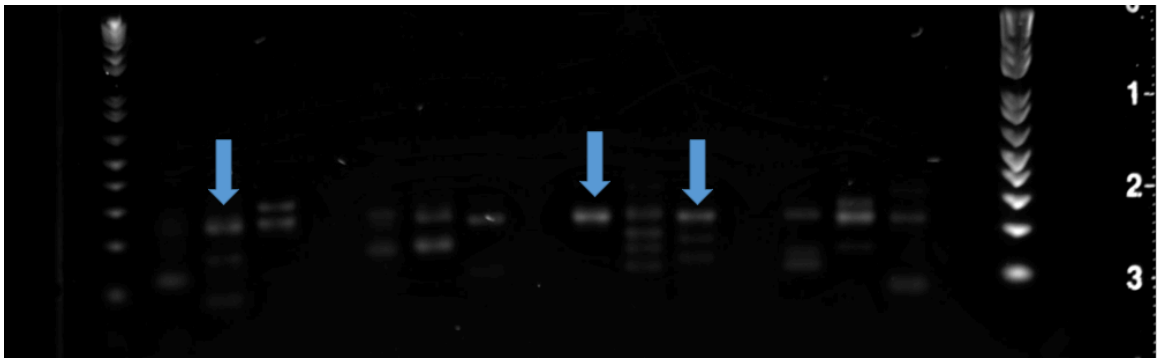


Figure 4.2 – Twins 4-5, DRD4 gel. A 2% agarose gel with PCR products from Twins 4A, 4B, 5A, and 5B from left to right, respectively. DNA from each sample was ran with buffers F, I, and L from left to right, respectively. Some of the bands near the 250 bp mark are indicated by arrows.

PCR products were ran on a 2% agarose gel, as shown in Figure 4.2. While the smearing seen in Figure 4.1 is not present, several bands can be seen for each sample. However, the brightest bands can be seen around the 250 bp mark, consistent with the bands in Figure 4.1. This is an interesting observation given the length of the amplicon of interest (~ 900 bp). While the whole amplicon was never observed, the brightest bands occurred consistently around 250 bp, even among different individuals.

Sequencing and Analysis of DRD4

Following visual confirmation of amplification, the PCR products that amplified were purified via QIAquick PCR Purification chemistry (Qiagen 28104) and sequenced using Big Dye v3.1 Cycle Sequencing chemistry (Applied Biosystems 4337455). Briefly, 4 μL of Ready Reaction Premix was added to 2 μL Big Dye

Sequencing Buffer, 1 μL forward primer, 3 μL molecular grade water, and 10 μL DNA. A second reaction was ran for each sample with reverse primer substituted for forward primer. A positive control with 2 μL of 0.1 ng/μL genomic DNA and a negative control were also ran. A standard cycle sequencing protocol was used, i.e., 96 °C for 1 minute followed by 25 cycles of 96 °C for 10 seconds, 50 °C for 5 seconds, and 60 °C for 4 minutes. Sequencing clean-up was performed using Performa DTR Gel Filtration Cartidges (EdgeBio 42453). Ten μL of Hi-Di Formamide was warmed to 37 °C and added to 10 μL of each sample DNA prior to injection. A 3500 Genetic Analyzer was used to electrophoretically separate DNA. Upon analysis, only the positive control DNA presented readable sequencing results.

The previously described extraction, conversion, and PCR protocol was performed using varying DNA concentrations in PCR (1.5, 2, and 2.5 μL) on Twins 4A and 4B. Figure 4.3 displays the 2% agarose gel results. Bands appeared at all three concentrations for both twins around the 250 bp mark, just as before. Though the results were less sporadic, the bands were faint, suggesting a low level of amplification. Bright bands can also be seen in the wells, suggesting that genomic DNA is present. Upon sequencing and analysis, no samples other than the positive control contained a readable sequence.

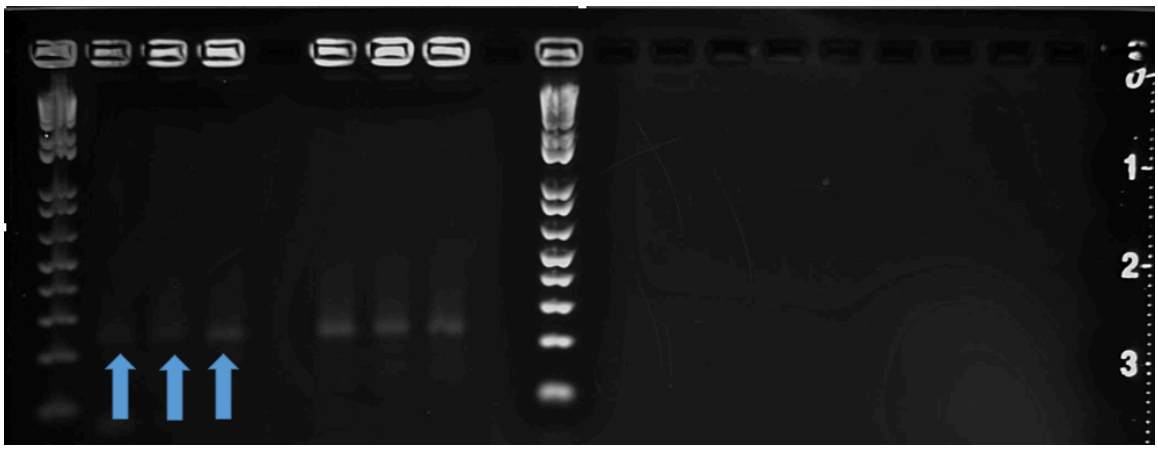


Figure 4.3 – Twins 4A and 4B, DRD4 gel. A 2% agarose gel with Twins 4A (lanes 2-4) and 4B (lanes 6-8). DNA concentrations of 1.5 μL, 2.0 μL, and 2.5 μL (left to right for each twin) were used in PCR. DNA from Twin 4B had slightly better amplification than DNA Twin 4A, which is indicated by the arrows.

PRKCA Primer Optimization

Since no sequencing results could be obtained from the DRD4 region, a new region of interest was chosen. Protein Kinase C, Alpha, or PRKCA, was chosen due to it being noted as a hypervariable yet stable epigenetic location (Lévesque et al., 2014). A 280 bp amplicon was chosen using Zymo Research's primer design algorithm. The amplicon is theoretically a good location to analyze due to its strong primers, small size, and relatively high concentration of CpG sites in the center of the amplicon. Primer sequence, length, and Tm can be seen in Table 3.1. The same PCR protocol described above was used on the PRKCA region, including the touchdown PCR step. Buccal swabs from two lab members (K1 and C1) were used as samples. Figure 4.4 displays the gel results for the first PRKCA amplification.

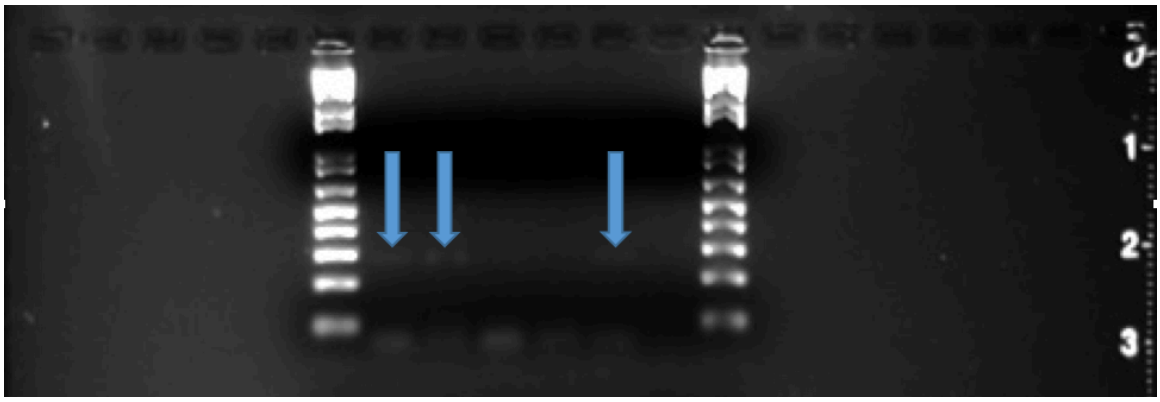


Figure 4.4 - First PRKCA amplification gel. Only very faint bands can be seen just below the 300 bp mark, as indicated by the arrows. The left 3 lanes are amplified, converted DNA from subject K1 and the right three are from subject C1. Buffers used were F, I, and L and each subject from left to right, respectively.

The DNA from Figure 4.4 was not sequenced due to the appearance of low amplification. Instead, several temperature variants of touchdown PCR were ran with the PRKCA primers using 0.1 ng/μL genomic DNA. A high temperature touchdown protocol ranging from 68 °C to 58 °C was tried as well as low temperature touchdown protocol ranging from 58 °C to 48 °C. However, none of the protocols yielded DNA when ran on a 2% agarose gel. To ensure that the DNA was amplifiable, non-converted control DNA with non-converted primers was ran alongside converted DNA with converted primers. Figure 4.5 shows the resulting gel. Only the non-converted amplicon in the presence of the non-converted primers was amplified.

Figure 4.5 – Buffers F, I, & L on control DNA, PRKCA gel. A 2% agarose gel with buffers F, I, and L ran in groups of three with control DNA. The 4th group of three is non-converted control with converted primers. The last group of three is non-converted DNA with non-converted primers.

Figure 4.6 – Control DNA on buffers A-L, PRKCA gel. A 2% agarose gel with control DNA ran using all buffers (A-L) and four replicates with buffer F. No DNA was amplified. Primer bands can be seen at the bottom of the gel. [font is different]

At this point, the amplification of non-converted DNA but lack of amplification in converted DNA suggested that either the converted PCR reaction was not optimized or the DNA was being lost during conversion. In an attempt to diagnose the issue, control DNA was tested on all premix buffers using the 60 °C to 50°C touchdown PCR protocol mentioned previously. Additionally, four replicates of control DNA in buffer F were ran. Figure 4.6 shows the gel containing the products

from this experiment. Since none of the DNA could be seen on the gel, some of the fundamentals of the experiment had to be rethought.

ZymoTaq PreMix on Control DNA

The next step was to try the experiment using buffer and polymerase designed specifically for PCR of bisulfite converted DNA. Eight samples, each containing 2 ng genomic DNA were bisulfite converted. The new PCR reaction for each tube was set up as follows: 4 μL DNA, 1.25 μL forward primer, 1.25 μL reverse primer, 12.5 μL ZymoTaq PreMix (Zymo Research E2003), and 6 μL molecular grade water. The ZymoTaq PreMix contained DNA polymerase and optimized buffer for PCR of bisulfite converted DNA.

In order to find the optimal temperature, a 10 °C temperature gradient was used during PCR. Ten total samples were ran at eight different temperatures: 60 °C, 59.4 °C, 58.3 °C, 56.3 °C, 53.9 °C, 52 °C, 50.7 °C, and 50 °C. Eight of the samples were ran at different temperatures, with duplicates at 56.3 °C and 53.9°C. PCR conditions were as follows: A 10-minute denaturation step at 95 °C followed by 40 cycles of 95 °C for 30 seconds, 50 – 60 °C for 40 seconds, and 72 °C for 1 minute. A final extension at 72 °C for 7 minutes was followed by a 4 °C hold. Figure 4.7 displays the PCR products on a 2% agarose gel. The brightest bands, representing the best amplification, can be seen just below the 300 bp mark in lanes 3, 5, 7, and 8. The DNA in these lanes had annealing temperatures at 60 °C, 58.3 °C, and 56.3 °C, respectively. Interestingly, the DNA in lane 4 with an annealing temperature of 59.4 °C was less bright than those around it.

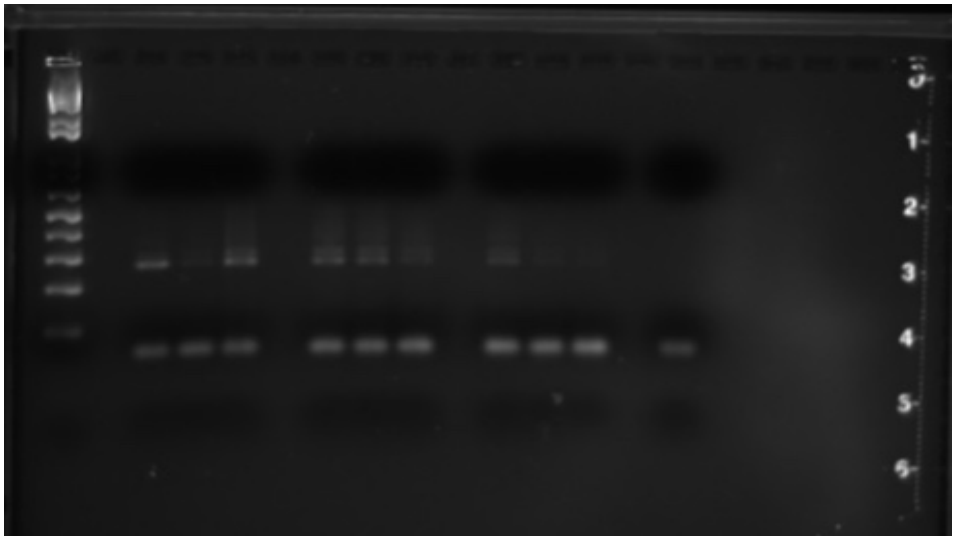


Figure 4.7 – PRKCA PCR gradient gel. A PCR gradient was utilized on 0.1 ng/μL control DNA in order to optimize the annealing temperature for the reaction. Annealing temperatures were: Lane 3: 60 °C; Lane 4: 59.4 °C; Lane 5: 58.3 °C; Lanes 7 and 8: 56.3 °C; Lanes 9 and 11: 53.9 °C; Lane 12: 52 °C; Lane 13: 50.7 °C; and Lane 14: 50 °C. Bands appear just below the 300 bp mark at most temperatures.

Following PCR clean-up, the products were sequenced using the cycle sequencing and sequencing clean-up protocols described previously. Sequenced amplicons were separated using a 3500 Genetic Analyzer and evaluated with Sequencher v4.2. Table 4.2 displays the results of the analysis. Electropherograms from this experiment were fairly clean, however a lot of Cytosine noise was observed in the forward sequences and Guanine noise observed in the reverse sequences. Figure 4.8 displays an example of this. It is possible that this noise is due to either incomplete bisulfite conversion or due to unincorporated Cytosines in the forward sequence and unincorporated Guanines in the reverse sequence since these nucleotides are largely absent in the bisulfite converted sequence.

Samples in Table 4.2 are labeled A-F according to the lanes in Figure 4.7. Lane 3 is Sample A, 4 is Sample B, 5 is Sample C, 7 is Sample D, 9 is Sample E1, 11 is Sample E2, and 12 is Sample F. Five total samples amplified and sequenced well

enough to be evaluated. Among these five, all CGs except F7 and R7 could be easily determined. Forward sequences for Samples A and E1 did not produce enough data to be analyzed.

CG Site	Sample A Status	Sample C Status	Sample D Status	Sample E1 Status	Sample E2 Status	Sample F Status
F1	N/A	U	U	N/A	U	U
F2	N/A	Mixed (U)	U	N/A	U	U
F3	N/A	U	U	N/A	U	U
F4	N/A	U	U	N/A	M (U)	U
F5	N/A	U	U	N/A	U	U
F6	N/A	U	U	N/A	U	U
F7	N/A	X	X	N/A	X	X
R1	U	U	U	U	Mixed (U)	U
R2	U	U	U	U	U	U
R3	U	U	U	U	U	U
R4	U	Mixed (U)	M (X)	M (X)	U	U
R5	U	U	U	U	U	U
R6	U	U	U	U	U	U
R7	X	X	X	X	M (Mixed)	X

Table 4.2 – Control DNA sequence data. CpG sites denoted as "M" were methylated, and CpG sites denoted as "U" were unmethylated. CpG sites signified by "X" could not be determined because of noise. N/A denotes a sample that did not produce good enough data on the 3500 Genetic Analyzer to be analyzed. Mixed denotes a mixture of methylated and unmethylaed CpG sites. Finally, parenthesis indicate the base call made by the researcher if it was different from the call made by Sequencher. F1 represents the first CpG site from the forward primer. R1 represents the first CG site from the reverse primer. F1 and R7 were the same CpG site, F2 and R6 were the same CpG site, and so on.

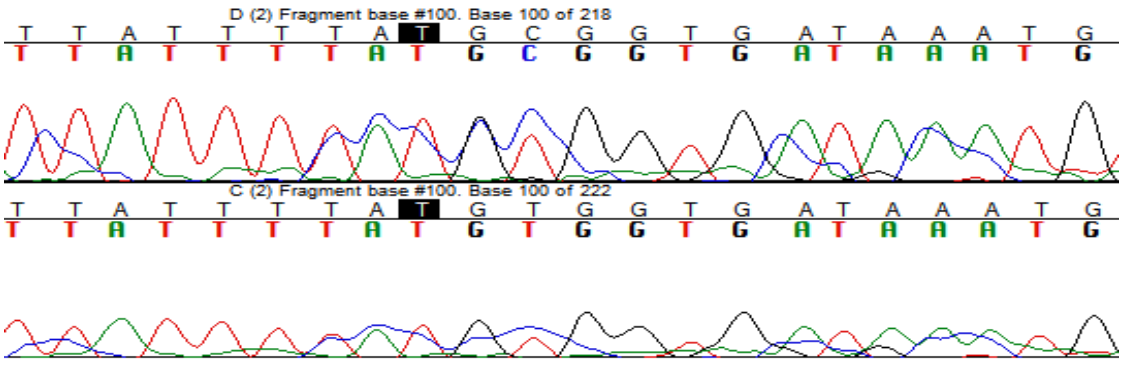


Figure 4.8 – Cytosine noise. Excessive Cytosine noise was observed in the forward sequences. Some peaks are misshapen and easily identified as noise while others have strong peaks and may indicate mixed methylation or incomplete bisulfite conversion.

ZymoTaq PreMix on Extracted DNA

Following successful amplification and sequencing of purified genomic DNA, the next step was to see if the protocol would work on DNA extracted from buccal swabs. It was at this point that the second change was made to the extraction protocol. In order to prepare a sample for extraction, the tip of the buccal swab was placed in a spin basket inside of a 2 ml collection tube. The basket was filled with DNAse free water to the top of the swab tip (about 450 μL) and incubated at room temperature for two minutes. The tube was then centrifuged at full speed (17,900 RCF) for 30 minutes to allow a cell pellet to form on the bottom of the tube. Following centrifugation, most of the liquid supernatant was siphoned off, leaving a small volume (~50 μL) behind for mixing. The sample was then vortexed until the cell pellet was completely dissolved in the remaining liquid. Once a cell-containing liquid was obtained from the cell pellet, the extraction and conversion protocol was followed as before.

The idea behind this change was to increase the concentration of cells in the cell-containing liquid, thus providing more template DNA for PCR. Because bisulfite conversion tends to degrade DNA, a higher DNA input is optimal. By forming the cell pellet and siphoning off some of the eluate, the concentration of cells per μL would be increased over previous samples. Ten total reactions were performed, with four samples from C1 and six samples from K1. Because 60 °C had the brightest band and was the hottest temperature in the previous experiment, an annealing temperature at 60 °C for 40 seconds was used in place of the gradient, while all other steps were kept the same (Table 4.3). PCR products were ran on a 2% agarose gel (Fig. 4.9).

PCR Step	Temperature	Time
Initial Denaturation	95 °C	10 Minutes
Denaturation Step	95 °C	30 Seconds (x40)
Annealing Step	60 °C	40 Seconds (x40)
Extension Step	72 °C	1 Minute (x40)
Final Extension	72 °C	7 Minutes
Hold	4 °C	Until Retrieved

Table 4.3 – PCR conditions for PRKCA. These PCR conditions were found to be most optimal for the PRKCA region.

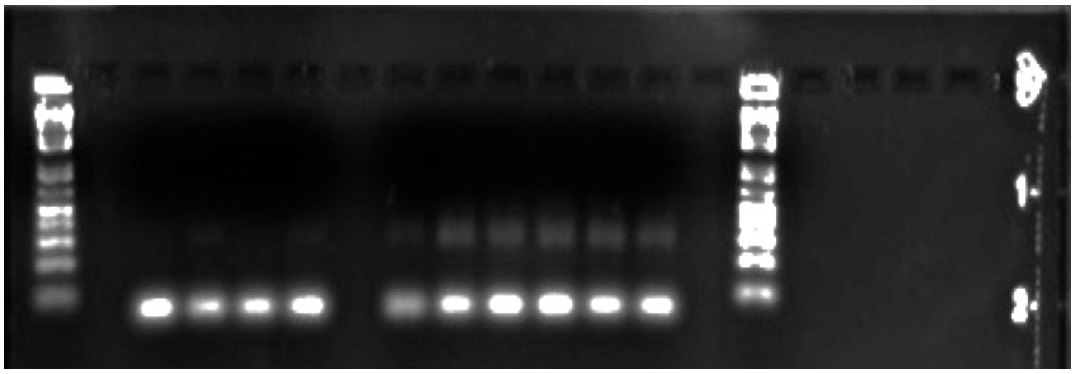


Figure 4.9 – PRKCA on extracted DNA. Amplified DNA from C1 can be seen in lanes 3 – 6 and amplified DNA from K1 can be seen in lanes 8 – 13 around the 300 bp mark. Bright primer bands can also be seen at the bottom, suggesting primer dimers or incomplete use of primers.

As with previous experiments, K1 had greater amplification than C1. However, bright primer bands can also be seen at the bottom of the gel, suggesting a low level of amplification and/or primer-dimers. Clean-up was performed on the amplified DNA and cycle sequencing was performed as in previous experiments. Sequenced DNA was separated on a 3500 Genetic Analyzer and analyzed using Sequencher 4.2 software. Table 4.4 displays the results for K1 and Figure 4.10 displays an example of the electropherograms. C1 did not amplify or sequence well enough to analyze.

CpG Site	Sample 1 Status	Sample 2 Status	Sample 3 Status	Sample 4 Status
F1	Mixed	Mixed	Mixed	Mixed
F2	Mixed	Mixed	Mixed	Mixed
F3	Mixed (M)	Mixed	Mixed	Mixed
F4	X	X	X	X
F5	X	X	X	X
F6	X	X	X	X
F7	X	X	X	X
R1	X	X	X	X
R2	Mixed	Mixed (U)	U	Mixed
R3	U	U	U	U
R4	X	Mixed (U)	X	Mixed (X)
R5	M	Mixed	Mixed	M
R6	X	X	X	X
R7	X	X	X	X

Table 4.4 – K1 sequence data from 25 μL run. These results are from K1's sequence data using the 25 μL PCR reactions with the ZymoTaq PreMix. Results were mostly inconsistent, both between samples and within samples from the forward and reverse directions. Some data was consistent, such as R3, which was clearly unmethylated in all samples.

Another experiment was performed with DNA from C1 and K1. This time, four swabs were taken from each subject. Swabs were spun to create a cell pellet and bisulfite conversion was performed. One additional change was made to the PCR reaction. Instead of 25 μL PCR reactions, 50 μL reactions were used, as recommended by the manufacturer. The volume of reagents used in the previous experiment were doubled, keeping the ratio the same and bringing the total reaction volume to 50 μL. Due to issues with the gel imaging software, an image of the gel could not be captured.

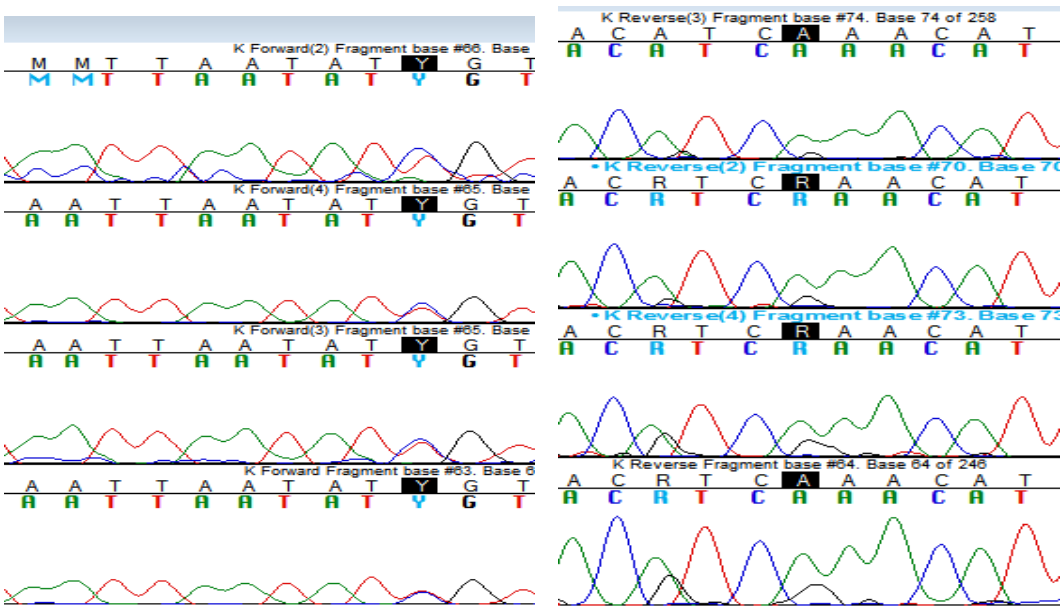


Figure 4.10 – Examples of mixed and unmethylated sites. The image on the left is of F2 for samples 1-4. It's clear that each sample has a mix of methylated/unmethylated C's at this site. The image on the right is of R4. Though a small Guanine peak can be seen, these four samples seem to be largely unmethylated at this CpG site.

PCR products were sequenced and separated on a 3500 Genetic Analyzer. Results from Sequencher evaluation can be seen in Table 4.5. Only K1 results are available, as C1 did not amplify and sequence well enough to analyze. Also, a different sequencing clean-up protocol was used for sample 1 forward and sample 3 forward to test its efficacy. Instead of using sequencing clean-up, the PCR clean-up was used. As seen in Table 4.5, using PCR clean-up after sequencing resulted in no data being produced from the genetic analyzer.

Sequencing results from the extracted DNA in this experiment and the previous one provided some insight into the nature of methylation analysis via cycle sequencing and capillary electrophoresis that will be discussed in the concluding chapter. While C1 samples did not amplify well enough to sequence and evaluate, K1 samples had low, but successful, amplification. Given K1's consistently greater

amplification than C1, it is possible that K1 simply sheds more cells during a buccal swab than C1. These are biological differences that will be difficult or impossible to control. Still, base calls were more consistent with the 50 μL PCR reactions. In some instances, a base call may need to be overridden by the researcher, as shown in Figure 4.11. In forensic contexts, parameters would need to be established concerning when a base can be called as mixed and when it can be called as a single base.

CpG Site	Sample 1 Status	Sample 2 Status	Sample 3 Status	Sample 4 Status
F1	N/A	Mixed	N/A	Mixed
F2	N/A	X	N/A	X
F3	N/A	Mixed (U)	N/A	U
F4	N/A	U	N/A	U
F5	N/A	U	N/A	U
F6	N/A	U	N/A	U
F7	N/A	U	N/A	U
R1	X	X	X	X
R2	Mixed (U)	U	U	U
R3	U	U	U	U
R4	X	X	X	X
R5	U (Mixed)	U	Mixed	U (Mixed)
R6	M	Mixed	Mixed	M
R7	M	Mixed	Mixed	Mixed(U)

Table 4.5 – K1 sequence data from 50 μL run. These results are from K1's sequence data using the 50 μL PCR reactions with the ZymoTaq PreMix. Results from this experiment were more consistent than with the 25 μL PCR reactions. The lack of results for sample 1 forward and sample 3 forward is likely do to the altered sequencing clean-up protocol that was used on these two samples.

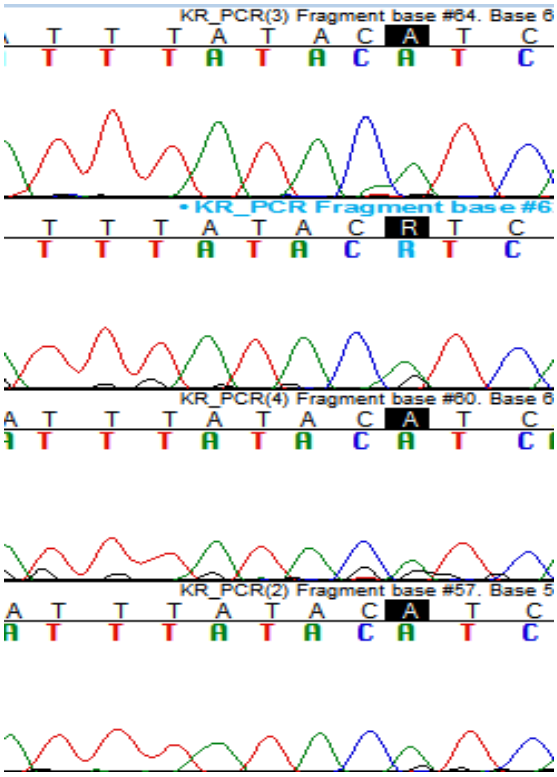


Figure 4.11 – Override of Sequencher call. This sequence data is from R2 using the 50 μL PCR protocol. In this instance, 3 out of 4 base calls were consistent. Sequencher called the nucleotide in the second electropherogram a mixed base. However, the researcher may override this call and change it to an Adenine based on peak height and morphology.

Because amplification and sequencing consistently worked better on samples from K1, the next experiment involved samples only from K1. The point of this was to maximize the efficiency of K1 in order to get an idea of inter-person and inter-sample consistency. For this, 5 extracts were used from the cell-containing liquid from the previous experiment. Because extracts from this cell-containing liquid worked well in the previous experiment, a second analysis using cells from the same liquid should also be successful. A new buccal swab from K1 was also spun into a cell pellet and 5 extracts were taken from it. All 10 samples and a positive control were bisulfite converted and 50 μL PCR reactions were prepared as before. Samples were ran on a gradient PCR with the same conditions as in the first ZymoTaq experiment with the control DNA. The purpose of this was to be sure that 60 °C was

the best primer annealing temperature for amplification and downstream applications.

PCR products were ran on a 2% agarose gel. Because of unresolvable issues with the gel imaging software, a gel image could not be captured. However, the samples from K1 did not produce bands, while the converted positive control did. This suggested that the conversion process and PCR worked properly, but K1 did not amplify from either extract.

The lack of amplification in extracted DNA was surprising given the previous successes. In an attempt to diagnose the problem, a follow-up experiment was performed. Three buccal swabs were taken from lab member J1 and two buccal swabs were taken from K1. These swabs were spun down into cell pellets and extracted. The 5 extracted samples and a control DNA sample were bisulfite converted and 50 μL PCR reactions were prepared. PCR was performed according to the protocol in Table 4.3. Products were ran on a 2% agarose gel (Fig. 4.12).

Just as before, none of the extracted DNA amplified. This time, however, not even the bisulfite converted control DNA amplified. There are a few possible reasons for this. First, it is possible that the DNA, including the control DNA, is poor quality. It could also be the case that an experimenter error was made during the conversion or amplification process. Third, it might be the case that the thermocycler failed. Finally, one or more of the reagents might not have been performing properly.

Figure 4.12 – K1, J1, and control DNA gel. From left to right the samples are K1, K1, control, space, three J1s. None of these bisulfite converted samples amplified.

In an attempt to diagnose the problem, an experiment with bisulfite converted and non-converted control DNA was performed. Four control DNA samples were bisulfite converted and ran on a thermocycler according to the conditions in Table 4.3. One sample of non-converted DNA was also ran with non-converted primers. The five samples were ran on a 2% agarose gel (Fig. 4.13). None of the bisulfite converted DNA was amplified, but the regular DNA did amplify.

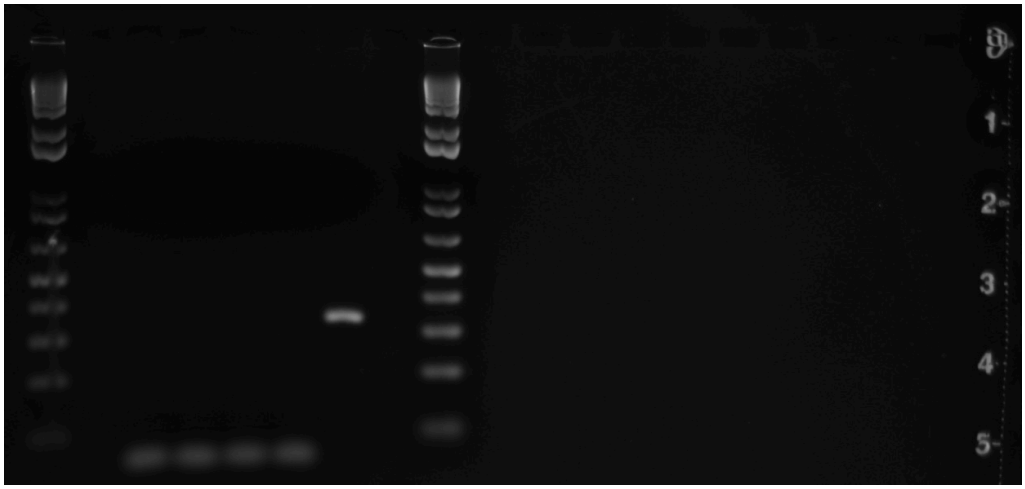


Figure 4.13 – Converted and non-converted control DNA. As in the previous gel, no bands could be seen in lanes containing converted DNA. The band at ~ 350 bp is the non-converted control amplified with non-converted primers. Due to slight primer differences, the non-converted is about 50 bp larger than the converted amplicon.

The final experiment for the project was to sequence the five sets of twins. Samples were spun into a cell pellet as before and bisulfite converted. PCR was performed using 50 μL reactions and the thermocycler conditions in Table 4.3. Products were ran on a 2% agarose gel, but no bands appeared, so sequencing was not performed on the samples.

Chapter 5: Discussion and Future Considerations

Discussion of Results

The results of this experiment established two key findings. First, using the traditional instrumentation of the forensic science laboratory may not be the most efficient or reliable means by which DNA methylation can be analyzed. While some DNA methylation could be visualized, the results using cycle sequencing and capillary electrophoresis were too inconsistent for forensic casework. This suggests that, if the forensic science community wishes to remedy one of the remaining challenges in DNA analysis, it would be beneficial to look to faster and stronger technology that is being used successfully in other fields, namely, targeted high-throughput sequencing.

While it will be worthwhile to investigate the use of targeted high-throughput sequencing as an alternative tool to cycle sequencing and capillary electrophoresis, this project did provide some evidence that those tools may be used. The experiment in which 50 μL PCR reaction was performed on K1 DNA (Table 5.1) provided some support that the method could be consistent and reliable. Four of the seven Cpg sites analyzed were consistent among every sample that produced data and one CpG site was consistent in five out of six samples. Because extracted DNA could not be successfully analyzed again, the data could not be replicated. More research into this method may improve its efficiency and demonstrate reliability.

	CpG 1	CpG 2	CpG 3	CpG4	CpG 5	CpG 6	CpG 7
Sample 2.1 F	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Sample 2.2 F	Mixed	X	Mixed (U)	U	U	U	U
Sample 2.3 F	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Sample 2.4 F	Mixed	X	U	U	U	U	U
Sample 2.1 R	M	M	U (Mixed)	X	U	Mixed (U)	X
Sample 2.2 R	Mixed	Mixed	U	X	U	U	X
Sample 2.3 R	Mixed	Mixed	Mixed	X	U	U	X
Sample 2.4 R	Mixed (U)	M	U (Mixed)	X	U	U	X
Results	5 Mixed / 1 M	2 Mixed / 2 M	3 Mixed / 2 U	2 U	6 U	6 U	2 U

Table 5.1 – 50 μL PCR rxns for K1. CpG sites are numbered 1 through 7 from the forward direction.

Another key finding was the establishment of a potential locus and primers for this locus. It was shown that the bisulfite converted primers for this region could be used to successfully amplify the target region in bisulfite converted DNA. While this amplification was not as reliable as it needs to be for forensic casework, it is likely that more sensitive instrumentation and better technology could remedy the inconsistency seen in this project. Still, the region of interest was shown to be a strong candidate locus for DNA methylation analysis for a few reasons. First, the amplicon is under the 300 bp, which increases the odds of obtaining a full-length amplicon. Second, the 3' end of the primers have strong G/C "clamps" that help keep

the primer on the DNA strand. Third, there are seven CpG sites in the amplicon, allowing for modest discriminatory power. Moreover, those CpG sites are situated 102 bp (forward) and 70 bp (reverse) from the ends of the sequence. This is important during sequencing analysis, as the first 30-50 bp of a sequence are often more difficult to analyze. Finally, CpG sites near this locus have been shown to be hypervariable between individuals, yet stable over 3 – 6 months (Lévesque et al., 2014). Future research that investigates other loci should look for sequence characteristics similar to those found in the PRKCA locus.

Why the SNP Method is not Practical

Forensic cases involving MZ twins prove more difficult to solve due to the inability of STR analysis to discriminate between nearly identical genomes. The interrogation of DNA evidence, one of the greatest milestones in forensic science, is ineffective when a MZ twin is implicated. Only recently has a genetic technique emerged in which there is potential to distinguish between twins. The technique, referred to as ultra-deep next generation sequencing, is described by Weber-Lehmann et al. (2014). The technique relies on sequencing the genomes of both twins nearly 100 times via high-throughput sequencing. This is performed in order to obtain an accurate picture of single nucleotide mutations (SNPs) that can be used to distinguish the twins. In one experiment, Weber-Lehmann et al. (2014) found five SNPs that differed between one pair twins, successfully differentiating between them.

However, there are a few problems with the SNP-based approach. To be different between all tissues in MZ twins, SNPs must occur within a relatively small window of development, specifically, after the twins split and before the separation of the three germ layers. This leaves only a few SNPs between the twins. With so few SNPs, each twin's genome must be sequenced upwards of 100 times for proper resolution. This "ultra deep" sequencing is expensive, costing over $100,000 and requiring over two months to complete. Thus, even if the method passes its Daubert trial, it will be too expensive for most crime labs to outsource.

Weber-Lehmann et al. (2014) mention MZ twin discordance in methylation, but dismiss the use of DNA methylation to distinguish between MZ twins, attributing their disregard to the medically-oriented nature of the research. Interestingly, the authors don't seem to mention the fact that a large portion of SNP research is also biomedically oriented. Moreover, Weber-Lehmann et al. (2014) only cite three sources for epigenetic differentiation of twins, which vastly underestimates the number of sources which have investigated this area of research and found positive results, even outside of the medical literature. Indeed, a search of the forensic science literature will reveal examples of reviews involving DNA methylation in forensic science (Gršković, Zrnec, Vicković, Popović, & Mršić, 2013; Kader & Ghai, 2015; Vidaki, Daniel, & Court, 2013), not to mention research articles like those discussed in Chapter 2.

At any rate, the claim that data produced in medical research cannot be applied to forensic science is unfounded. Had Leone Lattes followed this advice, he would have never developed the forensic application of ABO blood groups, which

were originally distinguished by Dr. Karl Landsteiner – a physician. Moreover, restriction fragment length polymorphism (RFLP) research was originally used in medical genetics to discover mutant genes. Yet, RFLP analysis was the basis of Alec Jeffrey's genetic fingerprinting, which kickstarted the use of DNA in forensic science. Ultimately, forensic science is the application of good science to the law. The application should be irrespective of the discipline from which it is derived, so long as the discipline is an established science.

Challenges Associated with Methylation Profiling

Methylation profiling will not be without its challenges. While it is the most studied of the epigenetic phenomena, there is still much that is unknown about DNA methylation. As mentioned earlier, there are differences among tissues due to tissue-specific functioning. This poses a potential issue with forensic sampling. DNA left behind at a crime scene may be from blood, skin, hair, saliva, semen, or other fluids and tissues. If methylation profiling is only applicable on a tissue-by-tissue basis, its power is limited. However, while a single test that encompasses all tissues would be ideal, tissue-specificity may not pose a problem to methylation profiling. If the tissue-origin of the crime scene sample can be determined, a matching sample type can be obtained from the suspect for comparison.

The use of methylation profiling would only apply post-STR, and in cases where the DNA is from one or both MZ twins. This means that the power of discrimination need not be the same degree as with STR profiles. The test would simply be a comparative test between the forensic sample, a known sample from twin A, and a known sample from twin B. With a majority of methylation being

stochastically and environmentally derived across tissue-types, it may not matter which tissue is being tested as long as the tissue type from the known matches the crime scene tissue type. The results rely only on comparison to an unknown, not a database.

In a review of a new SNP based method for distinguishing twins by Weber-Lehmann et al., (2014), Budowle (2014) suggests that statistical analysis does not appear to be necessary. His rationale, which it seems would also apply to methylation profiling, is that STR has already excluded all potential sources, except for the twins involved. So long as methylation profiles can be shown to be relatively stable, or if a time frame can be established for when methylation profiling can be performed post-perpetration, then the same statistical rationale should apply.

Rates of methylation and demethylation, as well as the exact molecular mechanisms involved, are not yet fully elucidated. It is still unknown exactly what role methylation plays in gene silencing and transcription as well as the details surrounding passive and active demethylation both during and after embryonic development. It is important that more research be done on DNA methylation, particularly as it pertains to the forensic questions it is used to answer. For MZ twin discrimination, the most important research on DNA methylation would investigate epigenetic stability over time and error rates relating to all aspects of the analysis. That being said, many applications and methods in every field from medicine to forensic science are successfully implemented without complete knowledge of all aspects of the phenomenon.

The dichotomous nature of methylation profiling diminishes the need for a fully disclosed or comprehensive understanding of the mechanisms at play, though a solid scientific foundation is necessary and, indeed, available. One of the reasons that DNA analysis is seen as the "gold standard" in forensic science is its basis in underlying scientific theory. Methylation analysis would share much of this basis, as epigenetic research began outside of the forensic science community, as opposed to being developed solely for it.

Considerations for Developmental Validation

Through the Scientific Working Group on DNA Analysis Methods (SWGDAM), the forensic science community has laid out developmental validation guidelines for DNA analysis methods. Because of the close theoretical and methodological relationship between DNA analysis and analysis of epigenetic phenomenon such as DNA methylation, any new DNA methylation protocol should strive to abide by the developmental validation guidelines set forth by SWGDAM. SWGDAM defines developmental validation as "the acquisition of test data and determination of conditions and limitations of a new or novel DNA methodology for use on forensic, database, known or casework reference samples" (SWGDAM, 2012). In developing a new protocol for validation, the following studies should be performed:

1. Characterization of genetic markers, to include inheritance, mapping, detection, and polymorphism;
2. Species specificity;

3. Sensitivity, to include upper and lower limits of input DNA for reliable results;
4. Stability of samples;
5. Precision and accuracy of the technique;
6. Repeatability;
7. Reproducibility;
8. Case-type samples
9. Population studies;
10. Mixture studies;
11. Publication of primer sequences;
12. PCR reaction conditions;
13. Assessment of differential amplification;
14. Effects of multiplexing;
15. Assessment of controls;
16. Criteria for detection; and
17. Establishment of appropriate measurement standards.

Following these criteria for the development of a DNA methylation analysis protocol will help ensure that the protocol is a reliable method that may be used in forensic science casework. While much work is left to be done for a proper DNA methylation analysis protocol, this project has helped establish a foundation upon which many of the aforementioned required studies may be performed. In particular, the establishment of the PRKCA locus for DNA methylation analysis and the appropriate PCR conditions will allow other required studies on this locus to be

performed. In addition to the PRKCA locus, other loci should be investigated. When looking for other DNA methylation loci, future researchers should look to Lévesque et al. (2014) for a list of genetic loci that are epigenetically stable over the course of 3 – 6 months, yet hypervariable between individuals.

Examples of Crimes Involving Identical Twins

Across all types of crime, both in the United States and abroad, instances can be found where a suspect cannot be prosecuted because they have an identical twin. Aside from having the same DNA profiles, MZ twins often cannot be told apart by witnesses or victims. For example, a 1999 rape case has remained unsolved due to the fact that the suspect and his twin both had records of sexual assault (Gee, 2014). While the MZ twin DNA problem lies mostly with violent crime, where DNA and witness testimony are often the biggest pieces of evidence linking someone to a crime, problems have arisen in non-violent crimes. In Malaysia in 2009, a man was arrested for possessing over 150 lbs of marijuana and almost 4 lbs of raw opium. Just after the arrested, the man's twin brother pulled into the driveway and was also taken into custody. During holding, the twins got mixed up, and the case was dropped because the guilty twin could not be identified in court (Gee, 2014). Other obscure examples of twins obfuscating a case can also be found. In what sounds like a movie plot, three suspects robbed a high-end jewelry store in Germany, getting away with over $8 million in stolen goods. DNA evidence from the crime scene and a glove led investigators – once again – to twin brothers. Due to reasonable doubt, both twins were set free (Gee, 2014).

Unless a fingerprint is left behind, it can be difficult or even impossible to distinguish MZ twins. Unfortunately, fingerprints are relatively transient, and leaving them behind can be easily avoided by wearing gloves. Without the help of DNA evidence, law enforcement officials are left using more unconventional methods to differentiate the twins. In a serial rape case in France in 2012, DNA evidence led to identical twins. Both brothers denied the charges, and could not be visually distinguished by the victim. In this case, the prosecution is relying on a verbal stutter, which the suspect and only one of the twins displays (Gee, 2014). While this case will go to court, it is possible that reasonable doubt could be cast due to a lack of evidence other than the stutter pointing to only one of the twins and not the other. As with many other crimes where MZ twins are involved, the suspect could be set free due to a relative inability to distinguish him from his twin.

Why Methylation Profiling Should Be Implemented

Forensic analysis of methylation variation among MZ twins should be studied and implemented in casework for two main reasons: 1) A practical test is needed to distinguish MZ twin DNA in forensic cases. Approximately 1 in every 333 births is a MZ twin birth, meaning that about 1 in 167 individuals is a MZ twin (Bortolus et al., 1999). As such, it will not be uncommon for a MZ twin to be implicated in a forensic case, as Weber-Lehmann (2014) duly noted. In the event that a MZ is the suspect of a crime, methylation profiling will serve as an important supplementary technique to STR analysis. 2) While creative and revolutionary, the recent method described by Weber-Lehmann et al. (2014) is not practical for forensic casework. Ultra-deep next generation sequencing requires several weeks to complete and analyze data.

Perhaps one of the biggest barriers to ultra-deep next generation sequencing is the cost. The method, particularly the bioinformatic aspect, is expensive. According to Anderson (2014), the Boston police department and the Boston District Attorney split the $120,000 cost of the DNA test developed by Weber-Lehmann et al. (2014) to help solve a rape case.

Unfortunately, this price tag is far too high for practical use in forensics. However, the willingness to pay over $100,000 demonstrates the community's need for a molecular test to distinguish MZ twins. Methylation profiling would be a relatively affordable and efficient alternative to SNP testing. The cost of the methylation profiling technique would be far less, even if targeted high-throughput sequencing was used. Furthermore, the analysis would take significantly less time to complete, which is imperative in many forensic cases. For example, when the judge denied a delay in the trail in order for Eurofins Scientific to complete the 10 week SNP test, the District Attorney had to drop the initial indictment until test results were obtained (Anderson, 2014). This points again to the importance of both a quick and cost effective method for genetically discriminating between MZ twins.

The evidence for MZ twin discordance is abundant in the literature. Though a large portion of the research is biomedically-focused, the data, as well as that produced by the molecular biology community, are applicable to forensic science. Many forensic scientists and researchers are embracing this explosion of epigenetic knowledge and applying it to forensic questions from tissue identification to age estimation and even MZ twin discrimination. Li and colleagues (2013) performed one of the first genome-wide analyses on MZ twin discrimination for forensics

purposes. Using a BeadChip with 27,578 CpG sites across 14,473 genes, Li and colleagues (2013) discovered 92 highly variable loci. With these 92 loci, Li and colleagues were able to successfully discriminate between 22 sets of MZ twins.

Some researchers are thinking further outside the box when approaching MZ twin differentiation. One group of researchers have had some success using high-resolution meltcurve analysis to distinguish between MZ twins. Taking advantage of a slight melting difference between Cytosine and 5mC, Stewart et al. (2015) were able to distinguish between five sets of MZ twins. While this method does have limitations, such as requiring relatively high DNA input and lower power of discrimination, it would be worthwhile to see if the method could be validated using a larger number of twins. At any rate, experiments like those from Li et al. (2013) and Stewart et al. (2015) are pioneering the field of forensic epigenetics, and have paved the way for further research into methylation-based identical twin discrimination.

References

Anderson, T. (2014, September 10). Rape suspect indicted with cutting-edge DNA testing. *The Boston Globe.* Retrieved from https://www.bostonglobe.com/metro/2014/09/10/prosecutors-indict-man-using-dna-testing-that-distinguishes-him-from-his-twin brother/gKfLy3XjmxvB2OysYg0WlM/story.html

Bennett, E. A., Keller, H., Mills, R. E., Schmidt, S., Moran, J. V., Weichenrieder, O., & Devine, S. E. (2008). Active Alu retrotransposons in the human genome. *Genome Research*, *18*(12), 1875–1883. http://doi.org/10.1101/gr.081737.108

Bestor, T. H. (2000). The DNA methyltransferases of mammals. *Human Molecular Genetics*, *9*(16), 2395–2402. http://doi.org/10.1093/hmg/9.16.2395

Bollati, V., Schwartz, J., Wright, R., Litonjua, A., Tarantini, L., Suh, H., … Baccarelli, A. (2009). Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mechanisms of Ageing and Development*, *130*(4), 234–239. http://doi.org/10.1016/j.mad.2008.12.003

Bortolus, R., Parazzini, F., Chatenoud, L., Benzi, G., Bianchi, M. M., & Marini, A. (1999). The epidemiology of multiple births. *Human Reproduction Update*, *5*(2), 179–187. http://doi.org/10.1093/humupd/5.2.179

Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science*, *317*(5845), 1760–4. http://doi.org/10.1126/science.1147939

Butler, J. M. (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. *Journal of Forensic Sciences*.

http://doi.org/10.1111/j.1556-4029.2006.00046.x

Carone, B. R., Fauquier, L., Habib, N., Shea, J. M., Hart, C. E., Li, R., … Rando, O. J. (2010). Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell*, *143*(7), 1084–1096. http://doi.org/10.1016/j.cell.2010.12.008

Chen, Z., & Riggs, A. D. (2011). DNA methylation and demethylation in mammals. *The Journal of Biological Chemistry*, *286*(21), 18347–53. http://doi.org/10.1074/jbc.R110.205286

Cole, S. A. (2001). *Suspect Identities. A History of Fingerprinting and Criminal Identification*. *History*. http://doi.org/10.1086/345029

D'Aquila, P., Rose, G., Bellizzi, D., & Passarino, G. (2013). Epigenetics and aging. *Maturitas*. http://doi.org/10.1016/j.maturitas.2012.11.005

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, *25*(10), 1010–1022. http://doi.org/10.1101/gad.2037511

Ehrich, M., Zoll, S., Sur, S., & van den Boom, D. (2007). A new method for accurate assessment of DNA quality after bisulfite treatment. *Nucleic Acids Research*, *35*(5). http://doi.org/10.1093/nar/gkl1134

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., … Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(30), 10604–10609. http://doi.org/10.1073/pnas.0500398102

Franchini, D. M., Chan, C. F., Morgan, H., Incorvaia, E., Rangam, G., Dean, W., …

Petersen-Mahrt, S. K. (2014). Processive DNA demethylation via dna deaminase-induced lesion resolution. *PLoS ONE*, *9*(7). http://doi.org/10.1371/journal.pone.0097754

Frumkin, D., Wasserstrom, A., Budowle, B., & Davidson, A. (2011). DNA methylation-based forensic tissue identification. *Forensic Science International: Genetics*, *5*(5), 517–524. http://doi.org/10.1016/j.fsigen.2010.12.001

Frumkin, D., Wasserstrom, A., Davidson, A., & Grafit, A. (2010). Authentication of forensic DNA samples. *Forensic Science International: Genetics*, *4*(2), 95–103. http://doi.org/10.1016/j.fsigen.2009.06.009

Gervin, K., Hammerø, M., Akselsen, H. E., Moe, R., Nygard, H., Brandt, I., … Lyle, R. (2011). Extensive variation and low heritability of DNA methylation identified in a twin study. *Genome Research*, *21*(11), 1813–1821. http://doi.org/10.1101/gr.119685.110

Gomes, I., Kohlmeier, F., & Schneider, P. M. (2011). Genetic markers for body fluid and tissue identification in forensics. *Forensic Science International: Genetics Supplement Series*, *3*(1). http://doi.org/10.1016/j.fsigss.2011.09.096

Gonzalez-Zulueta, M., Bender, C. M., Yang, A. S., Nguyen, T., Beart, R. W., Van Tornout, J. M., & Jones, P. A. (1995). Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. *Cancer Research*, *55*(20), 4531–4535.

Gordon, L., Joo, J. E., Powell, J. E., Ollikainen, M., Novakovic, B., Li, X., … Saffery, R. (2012). Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors,

subject to tissue-specific influence. *Genome Research*, *22*(8), 1395–1406. http://doi.org/10.1101/gr.136598.111

Gršković, B., Zrnec, D., Vicković, S., Popović, M., & Mršić, G. (2013). DNA methylation: The future of crime scene investigation? *Molecular Biology Reports*. http://doi.org/10.1007/s11033-013-2525-3

Guo, J. U., Su, Y., Zhong, C., Ming, G. L., & Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, *145*(3), 423–434. http://doi.org/10.1016/j.cell.2011.03.022

Hasegawa, M., Nelson, H. H., Peters, E., Ringstrom, E., Posner, M., & Kelsey, K. T. (2002). Patterns of gene promoter methylation in squamous cell cancer of the head and neck. *Oncogene*, *21*(27), 4231–6. http://doi.org/10.1038/sj.onc.1205528

Heyn, H., Moran, S., Hernando-Herraez, I., Sayols, S., Gomez, A., Sandoval, J., … Esteller, M. (2013). DNA methylation contributes to natural human variation. *Genome Research*, *23*(9), 1363–1372. http://doi.org/10.1101/gr.154187.112

Holmes, E. E., Jung, M., Meller, S., Leisse, A., Sailer, V., Zech, J., … Dietrich, D. (2014). Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PLoS ONE*, *9*(4). http://doi.org/10.1371/journal.pone.0093933

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. http://doi.org/10.1038/35057062

Jain, A., Hong, L., & Pankanti, S. (2000). Biometric identification. *Communications of*

the ACM, *43*(2), 90–98. http://doi.org/10.1145/328236.328110

Jain, A. K., Bolle, R., & Pankanti, S. (1999). *Biometrics: personal identification in networked society*. *Biometrics* (Vol. 1).

Jintaridth, P., & Mutirangura, a. (2010). Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. *Physiological Genomics*, *41*(February 2010), 194–200. http://doi.org/10.1152/physiolgenomics.00146.2009

Jones, P. a, & Takai, D. (2001). The role of DNA methylation in mammalian epigenetics. *Science (New York, N.Y.)*, *293*(5532), 1068–1070. http://doi.org/10.1126/science.1063852

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews. Genetics*, *13*(7), 484–92. http://doi.org/10.1038/nrg3230

Kader, F., & Ghai, M. (2015). DNA methylation and application in forensic sciences. *Forensic Science International*. http://doi.org/10.1016/j.forsciint.2015.01.037

Kaminsky, Z. a, Tang, T., Wang, S.-C., Ptak, C., Oh, G. H. T., Wong, A. H. C., … Petronis, A. (2009). DNA methylation profiles in monozygotic and dizygotic twins. *Nature Genetics*, *41*(2), 240–245. http://doi.org/10.1038/ng.286

Kitamura, E., Igarashi, J., Morohashi, A., Hida, N., Oinuma, T., Nemoto, N., … Nagase, H. (2007). Analysis of tissue-specific differentially methylated regions (TDMs) in humans. *Genomics*, *89*(3), 326–337. http://doi.org/10.1016/j.ygeno.2006.11.006

Koch, C. M., & Wagner, W. (2011). Epigenetic-aging-signature to determine age in

different tissues. *Aging*, *3*(10), 1018–1027. http://doi.org/100395 [pii]

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., & Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis*, *35*(2), 88–93. http://doi.org/10.1002/gene.10168

Langevin, S. M., Houseman, E. A., Christensen, B. C., Wiencke, J. K., Nelson, H. H., Karagas, M. R., … Kelsey, K. T. (2011). The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood. *Epigenetics*, *6*(7), 908–919. http://doi.org/10.4161/epi.6.7.16431

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews. Genetics*, *11*(3), 204–20. http://doi.org/10.1038/nrg2719

Lévesque, M. L., Casey, K. F., Szyf, M., Ismaylova, E., Ly, V., Verner, M. P., … Booij, L. (2014). Genome-wide DNA methylation variability in adolescent monozygotic twins followed since birth. *Epigenetics*, *9*(10), 1410–1421. http://doi.org/10.4161/15592294.2014.970060

Li, C., Zhao, S., Zhang, N., Zhang, S., & Hou, Y. (2013). Differences of DNA methylation profiles between monozygotic twins' blood samples. *Molecular Biology Reports*, *40*(9), 5275–5280. http://doi.org/10.1007/s11033-013-2627-y

Ma, L. L., Yi, S. H., Huang, D. X., Mei, K., & Yang, R. Z. (2013). Screening and identification of tissue-specific methylation for body fluid identification. *Forensic Science International: Genetics Supplement Series*, *4*(1). http://doi.org/10.1016/j.fsigss.2013.10.018

Martino, D., Loke, Y. J., Gordon, L., Ollikainen, M., Cruickshank, M. N., Saffery, R., & Craig, J. M. (2013). Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biology*, *14*(5), R42. http://doi.org/10.1186/gb-2013-14-5-r42

Meaney, M. J., & Szyf, M. (2005). Environmental programming of stress responses through DNA methylation: Life at the interface between a dynamic environment and a fixed genome. *Dialogues in Clinical Neuroscience*, *7*(2), 103–123.

Milagro, F. I., Campión, J., García-Díaz, D. F., Goyenechea, E., Paternain, L., & Martínez, J. A. (2009). High fat diet-induced obesity modifies the methylation pattern of leptin promoter in rats. *Journal of Physiology and Biochemistry*, *65*(1), 1–10. http://doi.org/10.1007/BF03165964

Mills, R. E., Bennett, E. A., Iskow, R. C., & Devine, S. E. (2007). Which transposable elements are active in the human genome? *Trends in Genetics : TIG*, *23*(4), 183–91. http://doi.org/10.1016/j.tig.2007.02.006

Ng, S.-F., Lin, R. C. Y., Laybutt, D. R., Barres, R., Owens, J. a, & Morris, M. J. (2010). Chronic high-fat diet in fathers programs β-cell dysfunction in female rat offspring. *Nature*, *467*(7318), 963–966. http://doi.org/10.1038/nature09491

Park, J. L., Kwon, O. H., Kim, J. H., Yoo, H. S., Lee, H. C., Woo, K. M., … Kim, Y. S. (2014). Identification of body fluid-specific DNA methylation markers for use in forensic science. *Forensic Science International: Genetics*, *13*, 147–153. http://doi.org/10.1016/j.fsigen.2014.07.011

Shen, L., & Waterland, R. A. (2007). Methods of DNA methylation analysis. *Curr Opin Clin Nutr Metab Care*, *10*(5), 576–81. http://doi.org/10.1097/MCO.0b013e3282bf6f43

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A. C., Fung, H. L., … Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*, *153*(3), 692–706. http://doi.org/10.1016/j.cell.2013.04.002

Singh, S. M., Murphy, B., & O'Reilly, R. L. (2003). Involvement of gene-diet/drug interaction in DNA methylation and its contribution to complex diseases: From cancer to schizophrenia. *Clinical Genetics*. http://doi.org/10.1046/j.1399-0004.2003.00190.x

Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nature Reviews. Genetics*, *14*(3), 204–20. http://doi.org/10.1038/nrg3354

Song, F., Smith, J. F., Kimura, M. T., Morrow, A. D., Matsuyama, T., Nagase, H., & Held, W. a. (2005). Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(9), 3336–3341. http://doi.org/10.1073/pnas.0408436102

Stewart, L., Evans, N., Bexon, K. J., Van Der Meer, D. J., & Williams, G. A. (2015). Differentiating between monozygotic twins through DNA methylation-specific high-resolution melt curve analysis. *Analytical Biochemistry*, *476*, 36–39. http://doi.org/10.1016/j.ab.2015.02.001

Talens, R. P., Christensen, K., Putter, H., Willemsen, G., Christiansen, L., Kremer, D., … Heijmans, B. T. (2012). Epigenetic variation during the adult lifespan: Cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell*, *11*(4), 694–703. http://doi.org/10.1111/j.1474-9726.2012.00835.x

van Dongen, J., Ehli, E. A., Slieker, R. C., Bartels, M., Weber, Z. M., Davies, G. E., … Boomsma, D. I. (2014). Epigenetic variation in monozygotic twins: A genome-wide analysis of DNA methylation in buccal cells. *Genes*, *5*(2), 347–365. http://doi.org/10.3390/genes5020347

Varley, K. E., Gertz, J., Bowling, K. M., Parker, S. L., Reddy, T. E., Pauli-Behn, F., … Myers, R. M. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Research*, *23*(3), 555–567. http://doi.org/10.1101/gr.147942.112

Vidaki, A., Daniel, B., & Court, D. S. (2013). Forensic DNA methylation profiling—Potential opportunities and challenges. *Forensic Science International: Genetics*, *7*(5), 499–507. http://doi.org/10.1016/j.fsigen.2013.05.004

Waterland, R. A., & Jirtle, R. L. (2003). Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Molecular and Cellular Biology*, *23*(15), 5293–300. http://doi.org/10.1128/MCB.23.15.5293

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D. C., Wiehler, J., & Rolf, B. (2014). Finding the needle in the haystack: Differentiating “identical” twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Science International: Genetics*, *9*(1), 42–46. http://doi.org/10.1016/j.fsigen.2013.10.015

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., & Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*, *39*(4), 457–466. http://doi.org/10.1038/ng1990

Wei, Y.-P. P. Y., Yang, C.-R. R., Zhao, Z.-A. a, Hou, Y., Schatten, H., & Sun, Q.-Y. Y. (2014). Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals. *Proc Natl Acad Sci U S A*, *111*(5), 1873–1878. http://doi.org/10.1073/pnas.1321195111

Weidner, C. I., Lin, Q., Koch, C. M., Eisele, L., Beier, F., Ziegler, P., … Wagner, W. (2014). Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology*, *15*(2), R24. http://doi.org/10.1186/gb-2014-15-2-r24

Wolff, G. L., Kodell, R. L., Moore, S. R., & Cooney, C. A. (1998). Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J*, *12*(11), 949–957. Retrieved from http://www.fasebj.org/content/12/11/949.short

Wong, A. H. C., Gottesman, I. I., & Petronis, A. (2005). Phenotypic differences in genetically identical organisms: The epigenetic perspective. *Human Molecular Genetics*. http://doi.org/10.1093/hmg/ddi116

Wong, C. C. Y., Caspi, A., Williams, B., Craig, I. W., Houts, R., Ambler, A., … Mill, J. (2010). A longitudinal study of epigenetic variation in twins. *Epigenetics*, *5*(6), 516–526. http://doi.org/10.4161/epi.5.6.12226

Yang, X., Han, H., DeCarvalho, D., Lay, F., Jones, P., & Liang, G. (2014). Gene body

methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell*, *26*(4), 577–590. http://doi.org/10.1016/j.ccr.2014.07.028

Zhao, Q., Rank, G., Tan, Y. T., Li, H., Moritz, R. L., Simpson, R. J., … Jane, S. M. (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nature Structural & Molecular Biology*, *16*(3), 304–311. http://doi.org/10.1038/nsmb.1568

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T.-Y., Kohlbacher, O., … Meissner, A. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature*, *500*(7463), 477–81. http://doi.org/10.1038/nature12433

Supplemental Tables

Marker	Twin 1A	Twin 1B
D3S1358	15, 16	15, 16
vWA	14, 16	14, 16
D16S539	9, 11	9, 11
CSF1PO	10	10
TPOX	8,9	8, 9
Yindel	2	2
AMEL	X, Y	X, Y
D8S1179	10, 15	10, 15
D21S11	30, 31.2	30, 31.2
D18S51	13, 15	13, 15
DYS391	11	11
D2S441	12, 14	12, 14
D19S433	13, 14	13, 14
TH01	7	7
FGA	24, 25	24, 25
D22S1045	11, 16	11, 16
D5S818	11	11
D13S317	8, 11	8, 11
D7S820	9	9
SE33	22.2, 30.2	22.2, 30.2
D10S1248	14, 16	14, 16
D1S1656	13, 14	13, 14
D12S391	18	18
D2S1338	19, 20	19, 20

Supplemental Table 1 – Allele Call Sheets for Twin 1A and Twin 1B

Marker	Twin 2A	Twin 2B
D3S1358	15, 18	15, 18
vWA	16	16
D16S539	11, 12	11, 12
CSF1PO	12	12
TPOX	8, 11	8, 11
Yindel	2	2
AMEL	X, Y	X, Y
D8S1179	12, 14	12, 14
D21S11	28, 29	28, 29
D18S51	16, 18	16, 18
DYS391	10	10
D2S441	11	11
D19S433	14	14
TH01	6, 9.3	6, 9.3
FGA	21	21
D22S1045	15, 16	15, 16
D5S818	9, 14	9, 14
D13S317	11, 12	11, 12
D7S820	7, 10	7, 10
SE33	15, 29.2	15, 29.2
D10S1248	14	14
D1S1656	15, 19.3	15, 19.3
D12S391	20, 21	20, 21
D2S1338	17,25	17, 25

Supplemental Table 2 – Allele Call Sheets for Twin 2A and Twin 2B

Marker	Twin 3A	Twin 3B
D3S1358	15, 17	15, 17
vWA	17, 18	17, 18
D16S539	8, 12	8, 12
CSF1PO	12, 14	12, 14
TPOX	8, 12	8, 12
Yindel	2	2
AMEL	X, Y	X, Y
D8S1179	11, 13	11, 13
D21S11	29, 31.2	29, 31.2
D18S51	13, 15	13, 15
DYS391	10	10
D2S441	11	11
D19S433	14, 15.2	14, 15.2
TH01	9.3	9.3
FGA	21, 22	21,22
D22S1045	15, 16	15, 16
D5S818	11, 12	11, 12
D13S317	9, 11	9, 11
D7S820	12	12
SE33	12, 25.2	12, 25.2
D10S1248	15	15
D1S1656	12, 15.3	12, 15.3
D12S391	18, 19.3	18, 19.3
D2S1338	17	17

Supplemental Table 3 – Allele Call Sheets for Twin 3A and Twin 3B

Marker	Twin 4A	Twin 4B
D3S1358	15, 17	15, 17
vWA	14, 17	14, 17
D16S539	12, 13	12, 13
CSF1PO	12	12
TPOX	8	8
Yindel		
AMEL	X	X
D8S1179	10, 14	10, 14
D21S11	29, 30	29, 30
D18S51	13, 14	13, 14
DYS391		
D2S441	10, 11	10, 11
D19S433	14, 15	14, 15
TH01	8	8
FGA	21, 23	21, 23
D22S1045	15	15
D5S818	10, 11	10, 11
D13S317	12, 13	12, 13
D7S820	11	11
SE33	16, 26.2	16, 26.2
D10S1248	12, 14	12, 14
D1S1656	13, 16	13, 16
D12S391	17, 18	17, 18
D2S1338	16, 17	16, 17

Supplemental Table 4 – Allele Call Sheets for Twin 4A and Twin 4B

Marker	Twin 5A	Twin 5B
D3S1358	16, 17	16, 17
vWA	18	18
D16S539	9, 13	9, 13
CSF1PO	12	12
TPOX	9, 11	9, 11
Yindel	2	2
AMEL	X, Y	X, Y
D8S1179	12, 15	12, 15
D21S11	32, 34.2	32, 34.2
D18S51	12, 13	12, 13
DYS391	9	9
D2S441	14	14
D19S433	14	14
TH01	7, 9.3	7, 9.3
FGA	22, 23.2	22, 23.2
D22S1045	15	15
D5S818	11, 12	11, 12
D13S317	9, 13	9, 13
D7S820	9, 10	9, 10
SE33	24.2, 33.2	24.2, 33.2
D10S1248	14	14
D1S1656	14, 15	14, 15
D12S391	20, 21	20, 21
D2S1338	17, 23	17, 23

Supplemental Table 5 – Allele Call Sheets for Twin 5A and Twin 5B

Location	Primer Dir. / Status	5’ – 3’ Sequence
DRD4	**For / Converted**	GTTTGGTTAATTATTTGTATTTTTAGTAGAGATGGGG
DRD4	**Rev / Converted**	CACTCTTATCACCCAAACTAAAATACAACAAC
PRKCA	**For / Converted**	TTTTGTATTTGATAGTATTGTAGTAATTAGTTTGGG
PRKCA	**Rev / Converted**	ATTTTTAAATAATTAAACATTAACCCTTTCCCC
PRKCA	**For / Non-Converted**	AGCTTGGGATGCAAAATGAT
PRKCA	**Rev / Non-Converted**	GTTTCACCTGGCCAAAATGT

Supplemental Table 6 – Primers