Identification of Tissue Specific Differential Methylation in Human Body Fluids and its Potential Application in Forensics

Farzeen Kader
209503949

School of Life Sciences, Faculty of Science and Agriculture,
University of KwaZulu-Natal

School of Life Sciences
University of KwaZulu-Natal Westville
Private Bag X 54001
Durban
4000
South Africa

Tel: 031 260 3192 / 2029 **Email:** 209503949@ukzn.ac.za

Master of Science - Genetics
Supervisor: Dr Meenu Ghai
(UKZN - Westville Campus)

IDENTIFICATION OF TISSUE SPECIFIC DIFFERENTIAL METHYLATION IN HUMAN BODY FLUIDS AND ITS POTENTIAL APPLICATION IN FORENSICS

BY

FARZEEN KADER

Submitted in fulfillment of the Academic Requirements for the degree of Master of Science (MSc) in the Discipline of Genetics, School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal (Westville Campus), Durban

As the candidate's supervisor, I have approved this dissertation for submission.

Supervisor: Dr. Meenu Ghai

Signed: _____ **Name:** _____ **Date:** _____

PREFACE

The experimental work described in this dissertation was carried out in the School of Life Sciences, Discipline of Genetics, University of KwaZulu-Natal (Westville Campus), Durban, South Africa from January 2014 to December 2015, under the supervision of Dr. M. Ghai and the co-supervision of Dr. O. T. Zishiri and Prof. A. O. Olaniran.

These studies represent original work performed by the author and have not otherwise been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of work of others, it is duly acknowledged in the text.

Supervisor:	Dr. M. Ghai	Signature: _____
Co-supervisor:	Dr. O. T. Zishiri	Signature: _____
Co-supervisor:	Prof. A. O. Olaniran	Signature: _____

COLLEGE OF AGRICULTURE, ENGINEERING AND SCIENCE

DECLARATION 1 – PLAGIARISM

I, …….. declare that

1. The research reported in this dissertation, except where otherwise indicated, is my original research.

2. This dissertation has not been submitted for any degree or examination at any other university.

3. This dissertation does not contain other person's data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This dissertation does not contain other person's writing, unless specifically acknowledged as being sourced from other researchers. Where other research sources have been quoted, then:
    a) Their words have been rewritten but the general information attributed to them has been referenced;
    b) Where their exact words have been used, then their writing has been placed in italics inside quotation marks and referenced.

5. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed:

…….

COLLEGE OF AGRICULTURE, ENGINEERING AND SCIENCE

DECLARATION 2 – PUBLICATIONS

DETAILS OF CONTRIBUTION TO PUBLICATIONS that form part and/or include research presented in this dissertation (including publications in preparation, submitted *in press* and published).

- Publication 1

Title: DNA Methylation and Application in Forensic Sciences.

Journal: Forensic Science International.

Citation: Kader, F. and M. Ghai. 2015. DNA Methylation and Application in Forensic Sciences. Forensic Science International. 249: 245-265.

- Publication in Preparation

Title: Development of new tDMR-based markers for identification of blood and vaginal fluid for forensic applications.

Authors: Kader, F., M. Ghai, D. Evans and F. Abader.

Signed:

…….

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First and most importantly, I wish to express my profound gratitude to God Almighty for His grace, guidance and granting me the strength to complete this study.

My sincere thanks go to my supervisor, Dr Meenu Ghai. Thank you for your patience, mentorship, assistance and guidance during the course of this study. I appreciate your time to critically review the dissertation and most of all, thank you for constantly motivating me to grow and do better. I am immensely grateful for all that you have done for me.

I wish to extend my gratitude to Dr Oliver Zishiri, for his constant enthusiasm and motivation. Thank you for constantly reminding me that in the end, it will be worth all the hurdles and challenges.

I acknowledge Professor Ademola Olufolahan Olaniran, for his support and advice throughout this study.

I am very thankful not only for Bioinformatic assistance from Dr Dyfed Lloyd Evans, but also for his advice and improvement of the study.

Thank you to Dr Shailesh V. Joshi for statistical analysis and assistance.

I am grateful to the National Research Foundation for financial support.

Thanks to all the volunteers who participated in the present study.

I am immeasurably grateful to my parents Osman and Thohida, and my brother Faizaan. Thank you for your patience and blessings throughout the study. I know I will not succeed without your dua's (prayers).

ABSTRACT

In forensic investigations, human biological traces have great potential to serve as strong evidence in placing a suspect at the scene of a crime. In cases where activities leading to the deposition of an individual's biological materials are disputed, the identification of body fluids and tissues may be crucial. Current catalytic, enzymatic and immunological techniques used to identify body fluids present numerous limitations such as lack of sensitivity and specificity. Hence, forensic scientists are constantly seeking exciting, novel methods to identify and analyse body fluids and aid in the reconstruction of crime scenes. Recently, DNA methylation-based markers have emerged as a reliable tool for identification of body fluids. Genome-wide methylation analysis using high throughput DNA technologies has discovered numerous differentially methylated regions (tDMRs) that differ in levels of methylation in various cell types and tissues. tDMRs may be unmethylated in particular body fluids/tissues yet display varying degrees of methylation in others, hence providing distinguishing characteristics between tissues. tDMRs can be targeted to develop markers for body fluid identification. To date, only a few DNA methylation-based markers have been reported to identify body fluids, most of which have yet to be validated. To enhance the specificity and robustness of DNA methylation-based identification, novel markers are required. Furthermore, as DNA methylation levels have been found to differ between ethnic groups of human populations, it is worthwhile to test previously documented tDMR-based markers on different ethnic groups to determine if there are significant methylation differences.

The present study developed new potential tDMRs-based markers to differentiate between saliva, semen, blood and vaginal fluid; and tested the methylation status of previously documented tDMRs for saliva on the diverse South African population. To identify new tDMRs, 1833 differentially expressed (over-expressed) genes, proposed to be regulated by DNA methylation, were identified in four body fluids; namely saliva, blood, semen and vaginal fluid. CpG dinucleotide methylation information from non-target tissues was mapped to genes and heavily methylated CpG islands (CGIs) were targeted for primer design. Sixty-three CGI sequences were selected and analysed for specificity in the human genome, and a total of four CGIs were targeted to design PCR primers. The primers were tested on saliva, blood, semen and vaginal fluid by methylation-specific restriction enzyme (MSRE)–PCR. The study has identified two potential body fluid-specific tDMRs: a tDMR of the *HPCAL1* gene was identified as a potential blood-specific hypomethylation marker, and a tDMR of the *PTPRS* gene is a potential vaginal fluid-specific hypermethylation marker. To

our knowledge, this is the first study where these genes have been targeted to identify tDMRs and develop markers for body fluid identification.

To determine if DNA methylation levels of previously documented tDMRs differ between four ethnic groups of South Africa, saliva samples were collected from 80 healthy individuals, male and female, belonging to four different ethnic groups of SA; Blacks, Indians, Whites and Coloureds. A multiplex MSRE-PCR assay was used to determine the methylation levels of four tDMRs in the *USP49, DACT1, L81528* and *PFN3* genes. The methylation levels of all selected tDMRs were highest in the Coloured ethnic group, while the lowest methylation levels were evident in the Black ethnic group. Promising results were found for two tDMRs; DACT1 and L81528, as both these markers displayed significant variations between the Coloured and Black ethnic groups. Significant differences in DNA methylation levels could assist forensic analysts in future, not only to accurately identify saliva but also to narrow down the search of sample donors, link them to the crime or exonerate them with confidence.

The analysis of tDMRs represents a novel, efficient and reliable technique to identify biological fluids and tissues and to differentiate between human populations. Future prospects involve validation of new tDMRs based markers on a wider population size and to determine methylation differences in other forensically relevant body fluids among ethnic groups of South Africa.

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

AND

LITERATURE REVIEW

1.1 Introduction

Every cell of a multicellular organism contains essentially the same genome. However, organisms contain a huge variety of differentiated and specialized cells, all of them using only a small proportion of available genes. These patterns of differential gene activity are clonally inherited through cell division. It is crucial to understand why genetically homogenous cells are different from each other and how this is controlled. The precise degree of epigenetic control of gene expression and development has been controversially discussed and researched to a vast extent (Reik, 2007; Storey *et al.,* 2007). Early studies of DNA methylation in genomic DNA successfully demonstrate tissue-specific differences in global methylation levels (Ayoubi and Van de Ven, 1996; Bird, 1986). Determining biological importance of this tissue-specificity is at present an area of active interest, especially since new roles of DNA methylation in growth and development have been discovered. Ever since researchers have discovered that DNA methylation is an epigenetic mark of supreme significance for normal development, the biochemical process is now being researched on a genome-wide as well as a tissue-specific level (Bock *et al.,* 2010; Dyson *et al.,* 2014; Vidaki *et al.,* 2013).

Exploiting DNA methylation for various applications such as medicine and forensic sciences is now well under way. The introduction of genome-wide DNA methylation analysis comparing tumorous and non-malignant tissues resulted in the discovery of many regions that undergo aberrant methylation during carcinogenesis (Tost, 2010; Ma *et al.,* 2013b). Those regions can potentially be used as biomarkers for cancer detection. Once methylation is acquired, it is chemically and biologically stable over time, while expression of mRNA and/or proteins can be modified by non-disease related environmental conditions and varies over the cell cycle (Jain *et al.,* 2013; Simon, 2005; Wild and Flanagan, 2010). DNA methylation profiles are tissue-specific and there are several chromosome segments called tissue-specific differentially methylated regions (tDMRs) that are known to show varying methylation patterns according to tissue or cell type (An *et al.,* 2013; Choi *et al.*, 2014; Lee *et al*., 2012; Rakyan *et al*., 2008). Since methylation patterns of tDMRs differ according to cell and tissue type, which reciprocally influences tissue-specific gene expression, tDMRs serve as dynamic, targetable, candidates for biomarkers in medical and forensic research (Choi *et al.,* 2014; Levenson and Melnikov, 2012; Rakyan *et al.,* 2008; Wan *et al.,* 2015).

In forensic sciences, detection and identification of body fluids present at crime scenes are crucial aspects of forensic investigations. Current molecular technologies allow for methylation profiling of tDMRs to reveal information about exactly what biological tissue or fluid source the sample originated from, as well as provides estimates of age, gender, phenotypic characteristics and ethnic backgrounds of perpetrators or victims (Gunn *et al.,* 2014; Vidaki *et al.,* 2013). Differential DNA methylation profiling may be the absolute best method to use when only a limited quantity of an unknown sample is collected as evidence and little information about the donor is known. Establishing the identity and origin of the body fluid helps to reconstruct the crime scene (An *et al.,* 2013; Choi *et al.,* 2014; Lam *et al.,* 2012; Lee *et al.,* 2012; Melnikov *et al.,* 2005).

Although research has been successful at identifying tDMR-based markers for precise differentiation between body fluids and tissues (An *et al.,* 2013; Frumkin *et al.,* 2011; Lee *et al.,* 2012; Madi *et al.,* 2012; Park *et al.,* 2014a), these studies have encountered various hurdles such as false positives and false negatives, inability to multiplex, incomplete bisulfite conversions, and inconclusive results for certain tDMRs due to over/underestimation of methylation levels. Some of these problems may be alleviated with the identification and addition of more tDMR-based markers and development of new methods for methylation profiling.

Having gathered scientist's attention only in the past few decades, the application of methylation in forensics is still relatively new and hence scientists are continually seeking novel methods to correctly identify biological fluids found at crime scenes. The present study endeavoured to identify new potential gene-specific tDMRs in the human genome for differentiation of four body fluids (saliva, blood, semen and vaginal fluid). The newly discovered tDMRs may serve as novel markers for body fluid and tissue identification and thus aid in the reconstruction of crime scenes.

DNA methylation patterns have also proven to be highly divergent between populations (Bell *et al.,* 2011; Fraser *et al*., 2012; Heyn *et al.,* 2013; Moen *et al.,* 2013; Zhang *et al.,* 2011a). Thus, the second objective was to decipher whether the tissue-specific methylation levels of previously documented tDMRs differ in saliva of four ethnic groups of South Africa; namely Blacks, Whites, Indians and Coloureds. tDMRs for the *USP49, DACT1, PFN3* (Ubiquitin-Specific Peptidase 49, Dapper 1 Isoform II, Profilin III) genes and *L81528*; a semen-specific marker, were used in a multiplex PCR assay for methylation

profiling of saliva. If significant differences are observed, the results could enable scientists to narrow down the search to a particular ethnic group.

Since numerous tDMRs exist in the mammalian genome, the identification of novel body fluid tDMRs, as well as validation of previous tDMRs is expected to spur the improvement of promising body fluid identification methods for medical and forensic applications.

1.2 Literature Review

1.2.1 Epigenetics

The human genome project resulted in vast amounts of genomic raw data. There are two meters of tightly compacted DNA, containing roughly 30 000 genes, coiled into the nucleus of each cell. However, identity and developmental potential of individual cells within organisms is not only defined by its genetic component. It is important to understand changes that occur in the function of genes without a change in the genome sequence. Referred to as epigenetics, the term describes various reversible modifications of the genome. The precise definition of epigenetics has baffled scientists for several years. On top of the genetic code, the epigenetic code comprises an additional layer of information. Whereas the former provides a framework for RNA and structure of protein; the epigenetic code controls packaging of DNA as well as gene regulation (Goldberg *et al.,* 2007; Schilling and Rehli, 2008; Tammen *et al.,* 2013).

The term epigenetics was first introduced by Conrad Waddington in early 1940, who defined epigenetics as ‘‘the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being’’ (Waddington, 1942; Waddington, 1968). Currently, a widely acknowledged definition is the ‘‘study of processes that produce a heritable phenotype that does not strictly depend on the DNA sequence’’ (Lieb *et al.,* 2006). Epigenetic modifications include DNA methylation, histone modifications, chromatin remodelling and non-coding RNAs; all of which play a pertinent role in regulation of gene expression devoid of changes in DNA sequence (Schaukowitch and Kim, 2014; Vidaki *et al.,* 2013). Epigenetic processes include imprinting, reprogramming, gene silencing, X chromosome inactivation and carcinogenesis. In mammals, a vital cell function regulated by epigenetic processes is cell differentiation wherein during embryogenesis; stem cells are completely differentiated (Heyn *et al.,* 2013; Tammen *et al.,* 2013; Vidaki *et al.,* 2013). During the entire lifespan of an organism, epigenetic mechanisms provide a link between environmental factors and phenotypic changes (Figure 1.1).

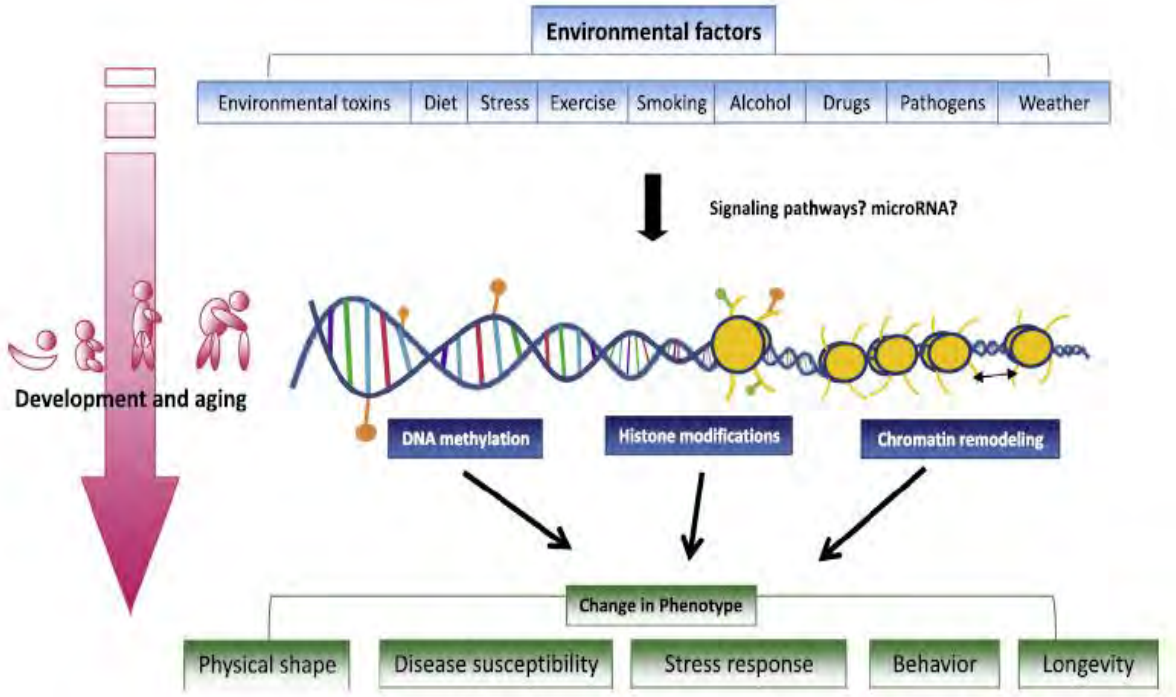

Figure 1.1: The association of epigenetic processes and phenotypic changes throughout an organism's lifespan (Tammen *et al.,* 2013).

The molecular basis of epigenetics is multifaceted and principally involves alterations in the activation of specific genes. Furthermore, chromatin proteins in association with DNA may be silenced or activated, thus ensuring that cells express only necessary genes required for an activity. Epigenetic programming is believed to begin as early as foetal development in the uterus. As DNA is inherited from one generation to the next, so too are epigenetic patterns preserved during cell division, yet modifications have been observed over an individual's lifetime. These changes have been found to occur in response to environmental exposure and various factors such as smoking and diet. Epigenetic research nowadays includes the study of covalent and non-covalent modifications of DNA and histone proteins, and the mechanisms by which such modifications influence overall chromatin structure. Of all the intriguing epigenetic phenomena, investigation in the field of DNA methylation is perhaps progressing at the most astonishing pace, rendering it superbly well explored and thus, best characterized. The present study focuses solely on DNA methylation, in particular tissue-specific differential DNA methylation which provides a stable, heritable and critical component of epigenetic regulation.

1.3 DNA Methylation

DNA methylation is the first epigenetic modification identified on DNA. It is an epigenetic mark of paramount importance for normal development in the human genome. Methyl-cytosine is the product of covalent attachment of a methyl group to a cytosine residue of the DNA sequence (Figure 1.2). The reaction is mediated by methyltransferases (DNMTs), namely; DNMT1, DNMT3A and DNMT3B which introduce onto the C5 position of cytosine residue a methyl group derived from S-Adenosylmethionine (SAM). The enzymes are crucial in mammalian development; partaking in two enzymatic activities: sustaining methylation post-replication and *de novo* methylation to institute new DNA methylation patterns throughout early development (Kader and Ghai, 2015; Lalruatfela, 2013; Zampieri *et al.,* 2015). The loss of DNA methylation leads to apoptosis or growth arrest in normal cells (Lister *et al.,* 2009; Yan *et al.*, 2011). Methyl-cytosine was thought to be as the only chemical modification of the mammalian genomic DNA. However, the existence of hydroxymethyl-cytosine in mammalian cells was proven by Kriaucionis and Heintz (2009) and Tahiliani *et al*. (2009). Hydroxymethyl-cytosine is an oxidation product of methyl-cytosine and the conversion of methyl-cytosine into hydroxymethyl-cytosine could be the first step in a pathway leading towards DNA demethylation. Due to its probable regulatory role in gene transcription, not unlike methyl-cytosine, hydroxymethyl-cytosine has been termed the 'fifth base' (Munzel *et al.,* 2011; Tammen *et al.,* 2013).

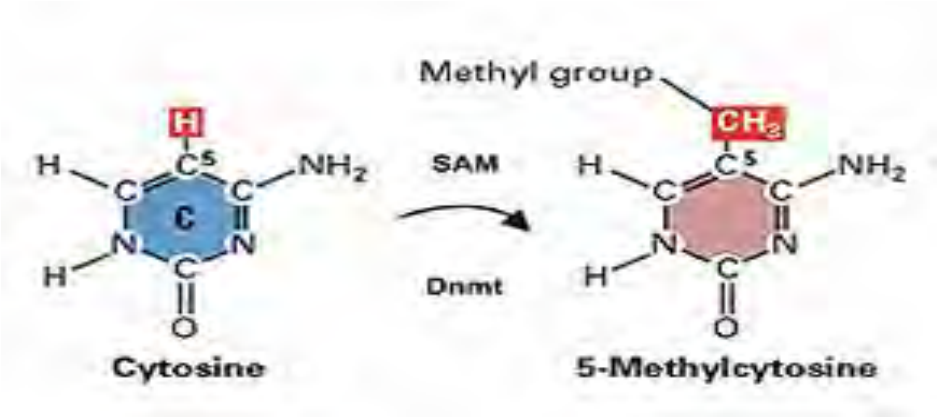

Figure 1.2: Basic overview of DNA methylation. A methyl group (CH_3), donated by S-Adenosylmethionine (SAM) is covalently attached to the C5 position of a cytosine residue.

1.3.1 Localisation of DNA Methylation in the Human Genome

DNA methylation occurs chiefly in CpG dinucleotides; however methylation elsewhere has been documented (Pinney, 2014; Yan *et al*., 2011). Residing within the human genome are approximately 30 million CpG dinucleotides which are unmethylated, hemi-methylated or abundantly methylated; varying according to region on chromosome, alleles, type of cell or phase of development (Rienius *et al*., 2012; Tammen *et al*., 2013).

The arrangement of methylation in human DNA comprises of two categories; a large segment wherein CpGs are intermittent yet highly methylated and a minor portion that is rich in CpGs however, mostly deficient of methylation. This minor portion is referred to as a CpG island; demarcated as a region of unmethylated CpGs that consists of 300-3000 bp and a GC content of less than 55% (Ghosh *et al*., 2010; Illingworth *et al*., 2008; Kader and Ghai, 2015). A minute portion, of below 20% of CpGs within the genome may be located in CGIs. Over 70% of gene promoters are in association with CGIs; the islands are shown to localise with promoters of housekeeping genes, just about 40% of tissue-specific genes and regulatory genes (Deaton and Bird, 2001; Ghosh *et al*., 2010). A conspicuous characteristic of islands are that they are not methylated in germ-cells and most somatic cells; the promoters are generally hypomethylated whilst promoters containing decreased CpG density are highly methylated (Day *et al*., 2013; Illingworth and Bird, 2009; Kader and Ghai, 2015).

In order to decipher the cause of global methylation in the genome, Bell *et al*. (2011) found a link between variations in DNA methylation patterns that surround single nucleotide polymorphisms (SNPs) in *cis.* The study deciphered that the genetic codes of promoters are the crucial contributing factors of methylation of surrounding cytosine bases on the same DNA strand. Lienert *et al.* (2011) reported another important determinant of DNA methylation; mutations occurring at transcription factor-binding sites within the genome that deter the maintenance of DNA methylation in surrounding regions.

Despite numerous studies defining the purpose and function of DNA methylation, there is still an immense lack of understanding of exact characteristics of DNA methylation particularly in individual human tissues. This lack of understanding necessitates detailed analysis of tissue-specific methylation of individual tissues (Igarashi *et al*., 2008). Genome-wide studies have revealed that DNA methylation profiles are tissue-specific and there are several chromosome segments called tissue-specific differentially methylated regions

(tDMRs) that are known to show varying methylation patterns according to tissue or cell type (An *et al.,* 2013; Choi *et al.*, 2014; Lee *et al*., 2012; Rakyan *et al*., 2008).

1.3.2 Tissue-Specific Differential DNA Methylation

Tissue-specific differentially methylated regions, or tDMRs are enriched at the margins of CpG islands and both CpG and G/C content is lower than that of surrounding regions. tDMRs are generally assumed to play the role of developmental switches; providing cells with epigenetic memory by generating cell-type-specific hypo- and hypermethylation patterns. These tDMRs have been implicated in indispensable involvement of mammalian development and tissue differentiation. Differentially methylated regions are believed to function by either preventing or attracting the binding of specific factors in a methyl-dependent manner (Cohen *et al.,* 2011; Igarashi *et al.,* 2008; Illingworth and Bird, 2009; Ohgane *et al*., 2008; Vidaki *et al.,* 2013). Numerous studies have characterised a large number of tDMRs via comparison of DNA methylation profiles amongst various cell lines and tissues (An *et al.,* 2013; Choi *et al.*, 2014; Christensen *et al.,* 2009; Day *et al.,* 2013; Huh *et al.,* 2013; Lee *et al.,* 2012; Rakyan *et al.,* 2008). These tDMRs are categorised based on collective behaviour (hypo- or hypermethylation) of a group of spatially clustered CpGs. Despite the fact that correlations of tDMR methylation levels with transcriptional state have been documented, the precise active regulatory role of tDMRs is not quite clear. However, major links between gene silencing and tDMRs have been established (Cohen *et al.,* 2011; Huh *et al.,* 2013; Ohgane *et al*., 2008).

Research over the past few years has established that individual genetic background and environmental factors are intertwined to lifestyle in determining the overall genetic and hence, health status of individuals. Increasing evidence shows that environmental and lifestyle factors may influence epigenetic mechanisms and these influences are rather apparent in DNA methylation patterns (Alegria-Torres *et al.,* 2011; Hunter, 2005; Tammen *et al.,* 2013). Such influences include, but are certainly not confined to nutrition and diets (King-Batoon *et al.,* 2008; Park *et al.,* 2011), life experiences (McGowan *et al.,* 2009), ageing, stress, exposure to pollutants (Alegria-Torres *et al.,* 2011), alcohol (Hines *et al.,* 2001; Mason and Choi, 2005) as well as economic status and even institutionalised care (Naumova *et al.,* 2012). As methylation patterns are altered and this is portrayed in expression, a few of these are further elaborated.

1.4 Environmental Influences and Differential DNA Methylation

1.4.1 Age

The relationship between environmental signals and epigenetics is not well defined in mammals; however hypo- and hypermethylation have been associated with ageing (Day *et al*., 2013; Fraser *et al*., 2012; Jaenisch and Bird, 2003; Tammen *et al*., 2013). Studies that focus on monozygotic twins have defined links between environment or ageing and long-term epigenetic effects on phenotype. Due to sharing the same genetic basis, monozygotic twins serve as the perfect system to study epigenetics. During early years, twins display similar methylation patterns; however later in life they demonstrate different amounts and patterns of methylation (Fraga *et al*., 2005; Li *et al.,* 2013; Vidaki *et al*., 2013). Studies of the *H19/Igf2* (Insulin-like Growth Factor II) locus by Pirazzini *et al*. (2012) led to the discovery of two regions in which, after 60 years of age, twins displayed considerable increase in intra-couple variation. It was observed that the range of methylation values increased only in the *Igf2* shore region thus emphasizing that the range of variation in methylation depends on the genomic location (Schneider *et al*., 2010). With increase in age, DNA methylation machinery tends to lose its ability to maintain methylation patterns across cellular divisions (Goyal *et al*., 2006). Variation in methylation of imprinted genes, such as the *H19/Igf2* locus has also been previously observed by Woodfine *et al*. (2011).

Furthermore, methylation of CpG islands linked with various genes such as that encoding oestrogen receptor, *MYOD* and *Igf2* was untraceable in young individuals, however, with age became detectable in normal tissue. The correlation between age and DNA methylation spurs questions regarding how epigenetic alterations influence various tissue types over time. It has been proposed that such variation in DNA methylation patterns could have been a stochastic process of random epigenetic drift (Boks *et al*., 2009; Christensen *et al*., 2009; Jaenisch and Bird, 2003; Vidaki *et al*., 2013). Studies by Bjornsson *et al*. (2008), Calvanese *et al*. (2009) and Zhang *et al*. (2011a) have demonstrated that with age, comes a general genomic decrease in DNA methylation. The study by Zhang *et al*. (2011a) found that in a population of individuals between the ages of 45-75, ageing tissues do in fact demonstrate a progressive decrease in methyl-cytosine, however although the age-related reduction in methylation was not too significant; this could possibly have been due to the limited population size and age range. Day and colleagues (2013) provided a few speculations to explain changes in DNA methylation with age. CpG methylation changes

were observed in blood and various other tissues. Such changes were attributed to variation in chromatin structure over time since enrichment of methylation with age has been noted within bivalent or repressive chromatin, and DNTM1 was found to reside in hypermethylated regions of transcribed genes (Day *et al.,* 2013).

1.4.2 Nutrition and Diets

Possibly ensuing an 'epigenetic diet', various dietary bioactive food components have been observed to alter gene expression via changes in DNA methylation (Hardy and Tollefsbol, 2011; Park *et al*., 2011). The availability of the methyl-donor, S-Adenosylmethionine (SAM) is determined by one-carbon metabolism. This is a pathway that involves vitamins B_6 and B_{12}, betaine, folate and choline as well as various amino acids such as glycine, methionine, serine and cysteine. If a component of the pathway is missing, such as deficiency in B vitamins, DNA methylation is altered (Niculescu and Zeisel, 2002). In addition to altering availability of B vitamins, alcohol consumption causes wastage of choline and methionine, reducing the amount of SAM available and thereby altering DNA methylation (Mason and Choi, 2005; Tammen *et al*., 2013). The diet is found to be an imperative determinant in the manifestation of late-onset disease. Vitamins and folates influence activity of enzymes that partake in cellular methylation processes and very much influence the rate of disease symptoms. Genomic instability and hypomethylation is allied with reduced amounts of folates (Jacob, 1999; Jaenisch and Bird, 2003). Fruit and vegetables contain numerous antioxidants and specific enzymes that also participate in methylation processes. Diets rich in fruits and vegetables have been shown to have anticancer properties (Alegria-Torres *et al.,* 2011; Borek, 2004). In mice, increased intake of folic acid increased DNA methylation of an allele in the coat colour agouti locus, resulting in gene silencing and phenotypic modifications (Tost, 2010; Waterland and Jirtle, 2003). Additionally, methyl-deficient diets induce liver cancers associated with hypomethylation and enhanced expression of oncogenes such as *c-fos* and *c-ras* (Friso *et al*., 2002; Jaenisch and Bird, 2003).

Other bioactive ingredients may have the opposite effects on DNA methylation. Epigallocatechin-3-gallate, which is a primary polyphenol found in green tea was found to reduce global DNA methylation in cancer cell lines via competitive inhibition of DNA methyltransferases. This reverses repression of tumour suppressor genes such as *p16*, O-6-methylguanine-DNA methyltransferase and reversion-inducing-cysteine rich protein with kazal motifs (Fang *et al*., 2003; Kato *et al*., 2008). The red carotenoid, lycopene was observed to have demethylating capabilities in a breast cancer cell line (King-Batoon *et al*.,

2008). Selenium, found in some vegetables and grains alter DNA methylation, as a reduction in this mineral leads to decrease in global DNA methylation with reduced expression of *DNMT1* in colon and prostate cancer cell lines, and rat colon and liver tissue (Tammen *et al*., 2013; Xiang *et al*., 2008).

Maternal diets also affect the offspring phenotypes and disease-risks. Intake of B-vitamins is linked to change in the susceptibility of offspring to breast and colon cancers (Sakatani *et al*., 2005; Tammen *et al*., 2013). Protein-restriction in rats has been found to epigenetically program the metabolism of offspring. With mothers that were fed low-protein diets, decreased methylation along with increased expression of peroxisome proliferator-activated receptor α in the liver of pups has been demonstrated (Lillycrop *et al*., 2008). Although a similar trend was observed for the glucocorticoid receptor gene, the effect was lost in pups of mothers fed a low-protein high-folate diet, demonstrating influences of the maternal diet on offspring carbohydrate and fat metabolism. Effects on global DNA methylation in genomes of pigs has been demonstrated; with changes in expression of DNA methyltransferases in liver and skeletal muscles of new-born offspring being observed during high and low protein diets of maternal pigs. Clearly, instead of using the term, 'you are what you eat', the phrase 'you are what your parents ate' seems more suitable (Altmann *et al*., 2012; Jimenez-Chillaron *et al*., 2012; Tammen *et al*., 2013).

1.4.3 Life Experiences

Adult risk factors such as tobacco smoking have been related with DNA methylation patterns in tumour tissues. Links between psycho-social factors such as cortisol output and perceived stress and DNA methylation have been established, as was early life socio-economic status (Christensen *et al*., 2009; Lam *et al*., 2012; Terry *et al*., 2008). Exposure to chemical and environmental pollutants induces changes in DNA methylation without altering the genetic sequence, resulting in epimutation-associated phenotypes. The anti-androgenic fungicide vinclozolin, which is an endocrine disruptor, alters methylation patterns in sperm; effects of which have been shown to persist for at least four generations (Anway *et al*., 2005). During comparative studies of epigenetic patterns of suicide victims without a history of childhood abuse to suicide victims with a history of abuse, increased methylation of the promoter of *Nuclear Receptor Subfamily III* genes was found. This gene encodes neuron-specific glucocorticoid receptor, which when stimulated inhibits the hypothalamic-pituitary-adrenal stress response. These results were also found in a study of rats, where pups raised with less grooming and licking, as well as less arched-back nursing also demonstrated altered

stress response. ‘Cellular memory’ mechanisms cause cells to remember and maintain their chosen fates, even long after the stimulus is gone and hence perturbations at early stages have long-lasting effects (Liu *et al*., 1997; Tammen *et al*., 2013; Tost, 2010).

Understanding DNA methylation marks and their biological regulation is central to understanding and targeting DNA methylation–associated changes. The past few years have seen unprecedented advances made in the development of new technologies to improve the study of DNA methylation. Advances include improvement in high-throughput methods to obtain quantitative data on locus-specific DNA methylation and development of various approaches to study DNA methylation on a genome-wide. No single method of DNA methylation analysis will be appropriate for every application. Nevertheless, by understanding the type of information provided by each method, and the inherent potential for bias and artefact associated with the method, investigators can select the method most appropriate for their specific research needs. Below is a brief review of these technological advances with examples of their adaptation to genome-wide DNA methylation profiling.

1.5 Mapping of Genome-wide DNA Methylation

Just as gene expression microarrays transformed and accelerated the study of transcriptional regulation, rapidly improving technologies are increasingly enabling researchers to assess locus-specific DNA methylation. Since DNA methylation plays an important role in many aspects of biology, including development and disease, and is now explored in forensics, precedence has been given to developing methods to determine the DNA methylation patterns of entire methylomes, individual genes and CpG sites.

Detailed accounts of methods to detect methylation have been described in numerous reviews, however methods can be divided into three main procedures; namely methylation-sensitive restriction enzymes, bisulfite conversion and affinity purification of methylated DNA. Combining these techniques with DNA microarrays and high throughput sequencing has made the mapping of DNA methylation feasible on a genome-wide scale (Ishkanian *et al.,* 2004). For example, Rollins and colleagues (2006) used restriction enzymes and standard cloning and sequencing to analyse over 8 Mb of methylated DNA and nearly 14 Mb of unmethylated human DNA. The Human Epigenome Project has used standard sequencing approaches to sequence a massive amount of bisulfite-converted DNA from human tissues and primary cells, and has identified a substantial number of tissue–specific differentially methylated regions (tDMRs) (Eckhardt *et al*., 2006).

1.5.1 Sodium Bisulfite Treatment of DNA Templates

Standard polymerase chain reaction (PCR) procedures remove methylation markers (Madi *et al*., 2012) and thus cytosine residues may be chemically modified with sodium bisulfite. Methylated cytosines remain unchanged whereas unmethylated cytosines are converted into uracil. The result is different DNA sequences for methylated and unmethylated DNA (Vidaki *et al*., 2013). Treated DNA is thereafter amplified by PCR using specifically designed primers. If DNA is unmethylated, a change in sequence occurs as unmethylated CpG pairs are converted to UpG pairs. Uracil will be converted to thymine during PCR, with corresponding drop in melting temperature and accordingly, specific amplification of methylated or unmethylated DNA may be accomplished by appropriate primer design (Madi *et al*., 2012). A technique known as Methylation-Specific Polymerase Chain Reaction (MS-PCR) may be applied to discriminate methylated from unmethylated cytosine residues (Lee *et al*., 2012). Base changes may also be detected by pyrosequencing which is a sequence-by-synthesis (enzymatic synthesis of DNA that is complementary to the DNA to be sequenced) method that allows monitoring of nucleotide addition and extension of DNA sequence in real-time (Madi *et al*., 2012; Tost and Gut, 2007). Analysis of the PCR product by Sanger sequencing (Eckhardt *et al*., 2006) or mass spectrometry (Ehrich *et al*., 2005; Igarashi *et al.,* 2008), can also be used to quantify the extent of methylation at each cytosine. Conversion of DNA with sodium bisulfite allows for quantitative and qualitative analysis of CpG sites, but one must take into account the incomplete conversion of DNA which may lead to over-estimation of methylation (Frommer *et al*., 1992; Vidaki *et al*., 2013).

1.5.1.1 Infinium HumanMethylation BeadChip Arrays

An outgrowth of their genotyping method, Illumina's HumanMethylation BeadChip arrays is best suited for analysis of bisulfite-treated DNA. Fast becoming a popular technique to analyse DNA methylomes, the technique enables determining the methylation profiles of well over 480 000 CpG sites within the genome (Dedeurwaerder *et al*., 2013; Wu *et al*., 2013). Chemical assays known as Infinium I and Infinium II are employed which are based on quantitative genotyping of C/T polymorphisms. These polymorphisms are generated at CpG sites by sodium bisulfite treatment with subsequent amplification of DNA. Additionally, the bisulfite modification generates a nearly '3-letter genome' which includes the bases adenine, thymine and guanine, with the only cytosines remaining being those that are

methylated (Bibikova *et al*., 2009; Dedeurwaerder *et al*., 2013). Infinium I bears a large resemblance to a single-channel microarray. The method uses two probes, each for a methylated and unmethylated allele. For these alleles, base extension is identical. In contrast, Infinium II has dual-colour readout. This method relies on one probe for both the methylated and unmethylated alleles and base extension is dependent on the methylation pattern of the hybridized DNA. Analysis of differential methylation is by statistical tests such as the Mann-Whitney or *t*-test. Differentially methylated sites are those that generate a *p* value below certain threshold, such as below 0.05 (Dedeurwaerder *et al*., 2013; Geiss *et al.,* 2008; Houseman *et al*., 2012). HumanMethylation BeadChip allows unbiased coverage of CpG island regions and genes reaching up to 96% and 99% respectively, as well as CpG island shores (these are 2 kb regions upstream and downstream of CpG islands) to determine genome-wide methylation profiles (Day *et al*., 2013; Koch and Wagner, 2011; Lokk *et al*., 2014). The chips also target sufficient CpG dinucleotides outside of these regions (Slieker *et al*., 2013). However, analysis and interpretation is more complex than initially considered, as it is necessary to filter probes that contain common single nucleotide polymorphisms (SNPs) as well as those that may be cross-reactive, due to possible generation of artefactual results. Within-array normalization must be conducted which will deal with colour/dye bias adjustment, Infinium I and II bias corrections and background correction (Dedeurwaerder *et al*., 2013). The ratio of the PCR products are ascertained using Illumina's Sentrix Array Matrix bead array platform, which can assay up to 1536 sites in 96 samples in a single experiment. However, the approach provides less coverage than other array-based methods, and necessitates the development and evaluation of a large set of selective primers, thus limiting its utility for *de novo* genome analysis (Zilberman and Henikoff, 2007).

1.5.1.2 MALDI-TOF MS

High throughput identification of methylation sites and semi-quantitative measurement of their single or multiple CpG locations may be accomplished by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry, or more commonly referred to as MALDI-TOF MS. The method employs base-specific cleavage of DNA (Ehrich *et al*., 2005; Stanssens *et al*., 2004). A low molecular weight organic acid matrix with high absorption at wavelength of the excitation laser contains the analytes which are desorbed by a laser pulse. An electric field will extract ionized DNA molecules, which are separated due to charge and mass by time-of-flight to the detector. Especially important for DNA methylation profiling, a quantitative signal is produced, facilitating determination of relative

abundance of products. Mass Spectrometry enables rapid, direct measurements of DNA instead of indirect data commonly obtained with other methods. With automated software, results may be easily stored. At present, the most frequently employed MS-based approaches rely initially on sodium bisulfite treatments followed by PCR amplification with subsequent sequencing and mass spectrometry (Igarashi *et al*., 2008; Meyers, 2012).

1.5.1.3 Pyrosequencing

A real-time sequencing by synthesis method, pyrosequencing is often employed for DNA methylation analysis of bisulfite-treated, amplified DNA (Shen *et al.,* 2007; Zhang *et al.,* 2011a). The method screens the incorporation of nucleotides via the conversion of released pyrophosphate into a relative light signal. The visible light is generated and may be viewed as peaks on a pyrogram. The heights of peaks are indicative of the amount of identical nucleotides that are incorporated into the sequence; a double peak would imply that two nucleotides were added. Degree of CpG methylation is software-calculated from the ratio of thymine to cytosine (Madi *et al.,* 2012; Tost and Gut, 2007). Mainly employed in the analysis of single nucleotide polymorphisms (SNPs), the method is quantitative, reproducible and sequencing may be repeated for the analysis of different CpG sites of a single amplification product. Automation for large-scale screening is also an advantage. However, four enzymes are necessary for accurate detection of nucleic acid sequences, and when this is included in other costs of the reactions such as pre-treatment with sodium bisulfite, primers and use of software, it is very expensive. A high error rate and long fusion primers which may possibly introduce bias greatly contribute to limitations of the method. Moreover, the technique is also time-consuming, providing results in about one day hence reducing feasibility and practicality (Fakruddin *et al.,* 2012; Tost and Gut, 2007).

1.5.2 Methylation-Sensitive Restriction Enzyme PCR

The classic tool of DNA Methylation analysis, Methylation-Sensitive Restriction Enzymes employ the use of restriction enzymes that recognize short pieces of DNA and cleave the DNA at distinct sites within or adjacent to these sequences. Some enzymes are sensitive to methylation and will not cleave the DNA if a cytosine residue is methylated, whilst other enzymes are insensitive to methylation and will specifically digest methylated DNA (Bird, 1986). An example of an enzyme that is insensitive to methylation is McrBC which is an *Escherichia coli* endonuclease that cleaves methylated DNA on one or both strands. Essentially, restriction enzyme-based methods either enrich for methylated DNA or

unmethylated DNA. Comparisons are made in one of the following ways; between a sample treated with an enzyme or a combination of enzymes and an untreated control; between a sample treated with a methylation-sensitive enzyme compared with a control treated with a methylation-insensitive isoschizomer or finally, between two test samples, such as two tissue types both digested with the same enzyme. Enrichment of unmethylated DNA, by digesting methylated DNA or by isolating smaller fragments generated by methylation-inhibited enzymes, is particularly useful for analysis of large, heavily methylated genomes. In the human genome, about 60-90% of CpG sites are methylated (Kader and Ghai, 2015), and hence enriching unmethylated DNA significantly reduces the complexity of the sample. The approach is robust, simple and does not require large quantities of DNA. However, efficient digestion of template DNA is absolutely vital, or else spurious results may be obtained. High quality DNA is necessary for efficient analysis and the main drawback of the method is dependence of the availability of recognition sequences that flank the sequences of interest. Frequently employed enzymes are the isoschizomers *Hpa*II and *Msp*I, both of which recognize the sequence CCGG. Whereas *Msp*I is blocked only by methylation of the outer cytosine, *Hpa*II is blocked by methylation of either cytosine. Since in mammalian genomes, methylation occurs chiefly in CpG sites *Hpa*II is inhibited and *Msp*I is not (Goll and Bestor, 2005). Another useful enzyme employed frequently in genomic studies is McrBC, an *E.coli* endonuclease that cleaves between two methylated cytosines in the context (G/A) metC, (Lippman *et al*., 2006; Rollins *et al*., 2006; Schumacher *et al*., 2006; Sutherland *et al*., 1992). The two sites can be separated by up to 3 kb, but the optimal separation is 55-100 bp (Gowher *et al*., 2000; Zhou *et al*., 2002). For this reason, McrBC is an excellent tool for the removal of densely methylated DNA. Although less of an issue with McrBC, sequence polymorphisms between samples can mimic methylation differences if they affect the enzyme recognition site. Therefore, it is safest to use restriction enzymes to compare samples that have no or little polymorphism, such as different tissues from the same organism. Extensive digestion of genomic DNA by means of the restriction enzyme may be followed by multiplex PCR amplification (Figure 1.3) of user-defined genes via gene-specific primers that flank the recognition site of the enzyme in use, in an amplification reaction termed Methylation-Sensitive Restriction Enzyme Polymerase Chain Reaction (MSRE-PCR) (Choi *et al*., 2014; Melnikov *et al*., 2005).

MSRE combined with PCR may be followed with methylation analysis employing standard capillary electrophoresis platforms (An *et al.,* 2013; Choi *et al.,* 2014; Melnikov *et*

al., 2005). MSRE-PCR is ideal for rapid DNA methylation analysis in a user defined set of genes, which is based on extensive digestion of genomic DNA with a methylation-sensitive restriction enzyme and PCR amplification of surviving fragments. A fragment will be designated 'unmethylated' if no PCR product is observed after digestion. Alternatively, the fragment will be called 'methylated' if it can be amplified after digestion.

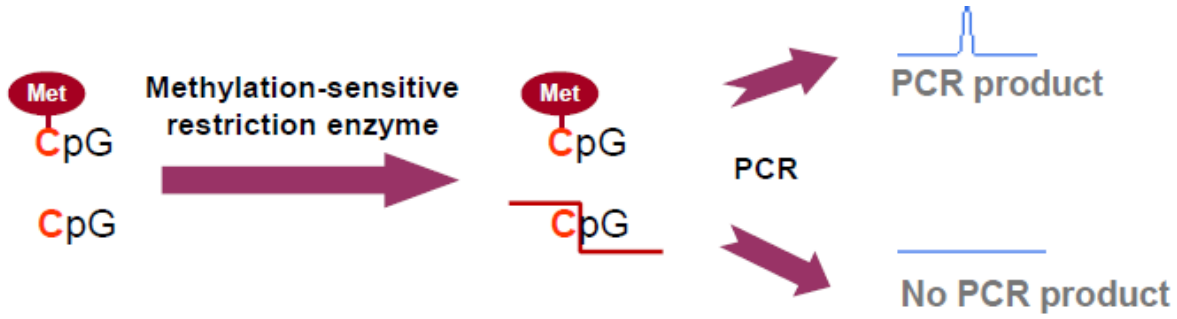

Figure 1.3: Methylation-Sensitive Restriction Enzyme Polymerase Chain Reaction (MSRE-PCR). Genomic DNA is cleaved by the endonuclease, for example *Hha*I which cleaves at recognition sites (GCGC). This is followed by PCR amplification of surviving fragments. If a fragment may be amplified after PCR, it will be termed 'methylated'. If a PCR product is not detected after digestion, the fragment is called 'unmethylated'.

MSRE-PCR allows simple multiplexing and avoids some of the problems inherent in bisulfite conversion, in particular the poorly controlled efficiency of modification, which can be incomplete due to incomplete denaturation or partial renaturation of DNA during treatment (Rein *et al.,* 1997); comprehensive modification of unmethylated cytosines is required for correct readout, which can be influenced by various factors including DNA apurinization during bisulfite treatment (Harrison *et al.,* 1998; Reeben and Pryds, 1994; Stirzaker *et al.,* 1997) and downstream differentiation of the methylated versus unmethylated sequence in many bisulfite-based methods requires two pairs of primers and two PCR reactions for each potentially methylated fragment which reduces the throughput of MSP and similar techniques, making screening of clinical samples more labour-intensive in bisulfite conversions. Finally, the yield of each product depends on the quality of the corresponding primers and can result in biased PCR if the amplification efficiency is different (Melnikov *et al.,* 2005; Stirzaker *et al.,* 1997). MSRE-PCR is a major detection tool that provides many pros. Since analysis may be performed using a universally known method of electrophoresis special training is not necessary. Multiple DNA templates may be analysed in a single assay (An *et al*., 2013; Choi *et al*., 2014). An important feature of the MSRE-PCR assay is its

ability to detect promoter methylation in heterogeneous samples, even when methylated sequences represent a small fraction of the overall specimen (Melnikov *et al.,* 2005).

1.5.2.1 Capillary Electrophoresis

The primary methodology used for separating and detecting short tandem repeat (STR) alleles, capillary electrophoresis is applied to numerous fields of research, especially in forensic DNA typing. To achieve trustworthy STR typing, three conditions must be met. Spatial resolution is needed to separate STR alleles that may differ in size by a single nucleotide; spectral resolution is needed to separate fluorescent dye colours from one another so that PCR products from loci labelled with different dyes can be resolved; and third, DNA sizing precision from run to run must be consistent enough so that samples can be related to allelic ladders that are run for calibration purposes (Butler, 2012). These specific requirements have been met with a variety of CE systems; and numerous studies have validated this with excellent findings. When capillary electrophoresis was applied to analyse 53 promoters of breast cancer cell lines (*MCF-7*, *MDA-MB-231* and *T47D*) that were subjected to MSRE-PCR using *Hin*61 enzyme, the MSRE-PCR followed by the CE system rectified the methylation status of genes analysed by other techniques (Melnikov *et al.,* 2005).

MSRE-PCR and electrophoresis is often used in the analysis of tDMRs to differentiate body fluids that may be located at crime scenes. When using tDMRs as markers, primers are designed to specifically target the methylated region within the tDMR. This is achieved by designing primers that flank the recognition site of the enzyme that will be used (An *et al.,* 2013; Melnikov *et al.,* 2005). Frumkin and colleagues (2011) subjected 50 DNA samples from blood, saliva, semen, and skin epidermis to digestion by *Hha*I, followed by multiplex amplification of specific genomic targets with fluorescent-labelled primers, capillary electrophoresis of amplification products, and automatic signal analysis by dedicated software (Figure 1.4). The investigation profitably yielded the source tissue of the samples. The system was described as fully automatable, provided operator-independent results, and allowed combining tissue identification with profiling in a single procedure which is quite favourable for forensic applications. Detection of semen and DNA profiling were combined into one assay and the ability to detect mixtures of semen and saliva in various ratios was demonstrated. The calculated percentage of semen was comparable to the fraction of semen in the samples. The same enzyme, *Hha*I was employed by An and colleagues (2013) and Choi *et al.* (2014); both studies successfully differentiated between various body fluids by use of MSRE-PCR in combination with capillary electrophoresis.

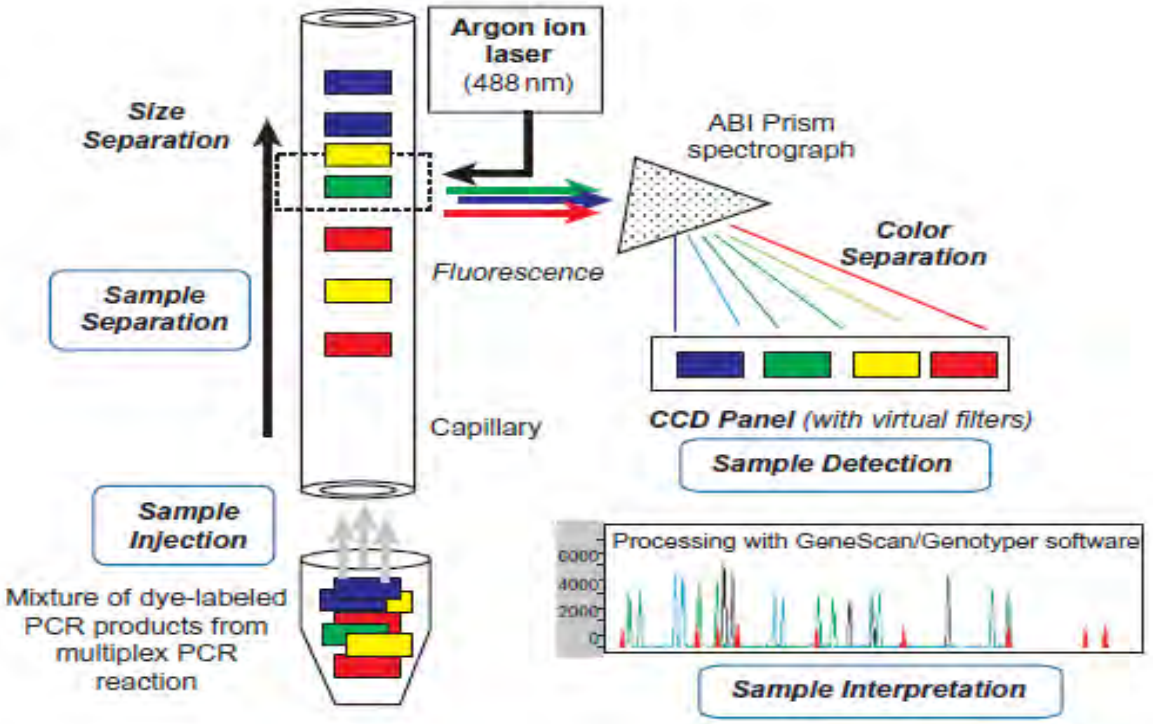

Figure 1.4: Detection of sample is performed automatically by measuring time span from sample injection to sample detection with a laser placed near the capillary end. Laser light is shone on the capillary where a window is burned into the coating of the capillary. DNA fragments are illuminated upon passing the window. Smaller DNA molecules are detected before larger molecules in order of migration speed which correlates with length or number of base pairs. Data from CE separations are plotted as a function of the relative fluorescence intensity observed from fluorescence emission of dyes passing the detector. The fluorescent emission signals from dyes attached to DNA molecules can then be used to detect and quantify the DNA molecules passing the detector (Butler, 2012).

1.5.2.2 Restriction Landmark Genomic Scanning

Based on the notion that within any genome, restriction enzyme sites may signify landmarks, Restriction Landmark Genomic Scanning (RLGS) is a quantitative method (Costello *et al*., 2002; Song *et al*., 2005) that allows for high resolution two-dimensional display of direct radio-labelled genomic DNA digested fragments. It enables recognition of high amounts; possibly over two thousand restriction landmarks in just a single assay. The method may be employed to determine epigenetic alterations in tissues, tumours as well as cancer cell lines (Ando and Hayashizaki, 2006; Rush and Plass, 2002). DNA is restricted with a rare-cutting enzyme, as methylation sensitivity of the endonuclease activity of the particular enzyme serves as a basis for detection of differential methylation patterns.

Although the system is applicable to all organisms and demonstrates high scanning extensibility by use of a combination of enzymes, it is disadvantageous in terms of efficiency; as results may be obtained between five days and two weeks. Small amounts of DNA samples cannot be analysed as the method requires a few good quality micrograms (Ho and Tang, 2007; Smiraglia *et al.,* 2007). Additionally the technique requires use of specific software such as Virtual Image-RLGS (VI-RLGS), expensive high-efficiency scanning capacity instruments and advanced image analysis systems such as a Fuji BAS2500 system (Costello *et al*., 2002; Ho and Tang, 2007; Okuizumi *et al.,* 2010).

1.5.3 Affinity Purification of Methylated DNA

Another method to map DNA methylation uses column- or bead-immobilized recombinant methylated CpG binding domain (MBD) proteins or antibodies, specifically MeCP2 and MBD2 which are able to bind methylated DNA. This is a rather simple approach that exploits the MBD which binds methylated CpGs. A tagged MBD domain expressed in *E. coli* may be affinity purified and the MBD column is subsequently used to purify methylated DNA. This can be followed with detection by PCR or microarray hybridization (Cross *et al.,* 1994; Gebhard *et al.,* 2006).

On the other hand, a commercially available monoclonal antibody that specifically identifies methylated cytosine may be used to immuno-precipitate methylated DNA (Cheong *et al.,* 2006; Weber *et al.,* 2007; Zhang *et al.,* 2006; Zilberman and Henikoff, 2007). Immuno-precipitation of methylated DNA is commonly referred to as Methyl DNA IP or MeDIP. Thereafter DNA undergoes shearing by restriction enzymes or sonication, with subsequent denaturation to obtain fragments. Short fragments are required for reduction of fragment-length effects and bias, as well as to improve efficiency and resolution. Then, immuno-selection and immuno-precipitation can take place using the antibody directed against 5-methyl-cytidine and antibody binding beads. The technique may be combined with PCR (Dalma-Weiszhausz *et al.,* 2006; Vidaki *et al*., 2013). Importantly, affinity-based techniques measure methylation in a given region and thus, a methylated stretch of DNA where methylation target sites are scant might be difficult to differentiate from an unmethylated region. This is particularly an apprehension with mammalian genomes, where CpG density is commonly low and CpG-dense sequences are usually unmethylated (Kreil *et al.,* 2006; Weber *et al.,* 2007). A possible method to overcome the limitation, as suggested by Zilberman and Henikoff (2007) would be to enrich for unmethylated DNA by isolating the

unbound fraction from either affinity method. The ratio of antibody (or MBD domain) to DNA would have to be carefully optimized to ensure removal of all methylated DNA. Instead, another option is that unmethylated DNA, prepared by McrBC (an endonuclease) digestion, could be further enriched by the removal of residual methylated DNA by affinity reagents. This approach would greatly reduce sample complexity.

1.6 Role of Epigenetic Modifications in Gene Expression and Development

In addition to transcription factors, gene expression and hence development are complemented by or are an outcome of epigenetic mechanisms such as post-transcriptional histone alterations, histone variants, several non-coding and small RNAs, ATP-dependant chromatin remodelling and DNA methylation (Kabesch *et al*., 2010; Goldberg *et al.,* 2007; Mazzio and Soliman, 2012; Reik, 2007). Consequently as differentiation and development proceeds, differentiated cells gain epigenetic marks dissimilar to those of pluripotent cells and cells of various lineages accrue altered patterns (Jaenisch and Bird, 2003; Reik, 2007; Remenyi *et al.,* 2004). Epigenetic mechanisms display elasticity in development; at infant stages genes that are only essential for later processes are temporarily repressed by histone modifications. Such repression is reversed when the genes are required. At the time of development, genes necessary for pluripotency are repressed by DNA methylation and histone modifications; these genes may also be repressed in germ cells implying that rapid reversal of epigenetic modifications may be required following fertilization. This enables activation of genes associated with pluripotency in the subsequent generation. Conversely DNA methylation-based silencing of imprinted genes and transposons requires stable maintenance at all gametic, early embryonic and adult stages (Lande-Diner and Cedar, 2005; Matarazzo *et al.,* 2007; Reik, 2007). Surprisingly, without changes in DNA sequence, epigenetic modifications are stably retained after numerous cell divisions and generations (Kader and Ghai, 2015; Reik, 2007).

DNA methylation has an essential regulatory function in mammalian development, which is to establish the correct pattern of gene expression, and that distinct DNA methylation patterns are tightly correlated to specific chromatin structures. Holliday and Pugh (1975) and Riggs (1975) suggested that maintenance of gene expression patterns throughout mitosis was due to DNA methylation. Now, with much substantiation and investigative studies to support this proposition, one simply cannot dispute the role of DNA methylation in gene expression.

1.6.1 Differential DNA Methylation and Gene Expression

Set aside from the greater part of genomic DNA, the methylation profiles of CGIs are a major focus of numerous studies; they represent a portion of the epigenome with incontestable regulatory potential. CGIs may often be associated with gene promoters and the very first exon; however they are also located in gene regions at the 3′ end (Takai and Jones, 2002). CGIs impact gene expression; a common trend observed is increased methylation of promoter CGIs leads to decreased expression or inactivation of downstream genes (Figure 1.5). Research on DNA methylation has exploited the promoter region with studies mainly focussing on this area. The finding that methylation patterns may be tissue-specific encouraged several genome-wide studies; researchers are constantly interested in genome-wide chromosome segments labelled as tissue-specific differentially methylated regions (tDMRs) whose methylation configurations fluctuate according to the cell or tissue type. tDMRs likewise induce variation in gene expression and regulation, hence enriching the wide role of methylation on human genotype and phenotype (Choi *et al*., 2014; Rakyan *et al*., 2008; Song *et al.,* 2005; Wan *et al.,* 2015).

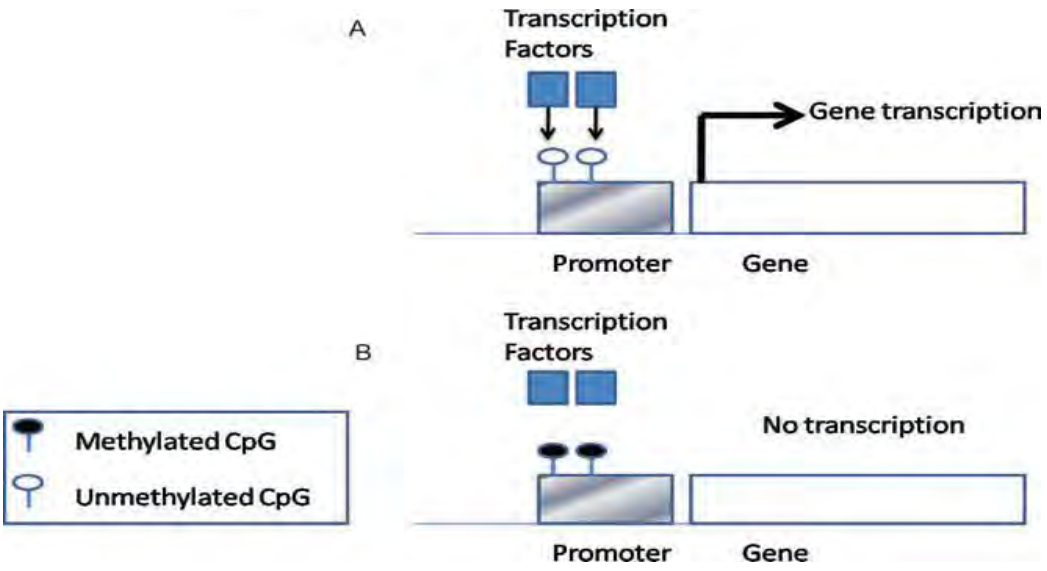

Figure 1.5: Regulation of gene expression by DNA Methylation. (A) The CpG island promoter is hypomethylated and enables binding of transcription factors, which is necessary for transcription initiation. (B) Methylation of the CpG island promoter inhibits binding of transcription factors and results in gene repression (Lim and Maher, 2010).

There are two widely accepted theories to the mechanism of gene silencing. First; in the presence of methyl groups, DNA methylation dependent transcription factors are incapable of gaining access to promoter regions of cognate recognition sites, rendering the gene untranscribed and inactive (Kulis *et al*., 2013; Yu *et al*., 2013). This concept was demonstrated by Macleod and colleagues (1994) who proved that exclusion of the transcription factor *Sp*1 binding sites that flanked a CpG island permitted access of the sites to DNA methyltransferases during development, enabling *de novo* methylation. According to the second theory, proteins consisting of a methyl CpG-binding domain (MBD) bind to methyl groups, with subsequent recruitment of histone deacetylases, chromatin compression and finally, gene inactivation (Dong *et al*., 2000; Kulis *et al*., 2013). Classic examples of genes in which the methylation status of CpG-dense promoters correlate with gene silencing are *MASPIN* and *GATA2* (Mammary Serine Protease Inhibitor and Transcription Factor and Gene encoding GATA Binding Protein, respectively) (Kader and Ghai, 2015).

Mammals are said to use the elevated methylation of CpG-enriched promoters to hinder transcription and ensure that genes on the X chromosome, imprinted genes and parasitic DNA are silenced (Cocozza *et al*., 2011; Jones and Takai, 2001). One must also consider that even though the greater parts of CGIs are unmethylated during development,

only about half the genes are activated in cells or cell types. Other types of silencing mechanisms such as the absence of transcription factors are responsible for inactivating the remaining genes (Kulis *et al*., 2013).

Disease-associated modifications in DNA methylation have been studied extensively and they are starting to play an important role as biomarkers (Hesselink *et al.,* 2014; Heyn and Esteller, 2012; Laird, 2003; Silva *et al.,* 1999). Their usefulness as biomarkers can be attributed to the fact that DNA methylation patterns are stable over time, easy to investigate and cell-type specific (Bloushtain-Qimron *et al.,* 2008; Hesselink *et al.,* 2014). DNA methylation profiles, especially in cancer cells seem to be partially based on their cell of origin. As tumours with different cells of origin may differ in prognostic outcome and therapeutic response, biomarkers distinguishing these groups may have predictive value. In addition, since DNA methylation alterations usually occur early in disease and are stable throughout disease progression, DNA methylation-based markers may be good candidates for early disease detection and monitoring disease development. Additionally, DNA methylation can be used to detect cancer recurrence (Brennan *et al.,* 2012; Hanahan and Weinberg, 2011; Levenson and Melnikov, 2012; Li *et al.,* 2009; Stirzaker *et al.,* 1997).

Whilst results of most DNA methylation-based studies indicate repressed transcription and expression (Rauch *et al.,* 2009; Lam *et al.,* 2012), there are some researchers who have not detected a correlation between DNA methylation and expression (Grunau *et al.,* 2000), and contrastingly a few have even found enhanced expression due to DNA methylation (Archey *et al.,* 1999). Studies that highlight the unpredictable effect of DNA methylation on gene expression are discussed in the sections that follow.

1.6.1.1 Imprinting

Imprinting refers to the rare but extraordinary situation when alleles of the same gene are expressed unequally in a parent-of-origin dependent manner. This means that some genes within the diploid genome are only expressed from alleles inherited from the father, whereas some others are only expressed from alleles derived from the mother. Although a rather infrequent epigenetic phenomenon in humans; genomic imprinting by DNA methylation cannot be overlooked for its role in transport, developmental and regulatory processes in humans (Hamed *et al.,* 2012; Ishida and Moore, 2013; Tycko and Morrison, 2002).

The first mammalian imprinted genes that were identified are the *H19* and *Igf2* (Insulin-like growth factor II) genes. The hypomethylated status of the control region ensures *H19* expression solely in the maternal allele. Similarly, *Igf2* is only expressed in the paternal allele (Bell and Felsenfeld, 2000; Kurukuti *et al*., 2006).

Hark *et al*. (2000) and Kurukuti *et al*. (2006) confirmed that decreased methylation of the control region enables binding of CTCF, the Zinc finger protein. This serves as a boundary, prohibiting interaction of *Igf2* with enhancers located at the 3′ end of *H19*. On the other hand, increased methylation of the control region terminates binding of CTCF, hence allowing *Igf2*-enhancer association and enabling paternal expression of *Ifg2.* These studies, in addition to Bell and Felsenfeld (2000) proved that DNA methylation can govern gene expression by tempering with enhancer contact to the gene promoter.

It has been established that aberrant DNA methylation patterns can lead to a diseased state, especially wherein global hypomethylation concurs with gene-specific elevated methylation, such as cancer (Ehrlich, 2002; Tost, 2010; Yoo and Jones, 2006; Wentzensen *et al.,* 2014). The *H19* and *Igf2* genes are good models of this and have been thoroughly researched in this context on numerous occasions (Goa *et al*., 2002, Guo *et al*., 2014a; Tost, 2010). Elevated methylation levels of the DMR (differentially methylated region) directly upstream of *H19* is linked to anomalous expression of *Igf2* and the *H19* genes. The DMR of *H19* facilitates reciprocal expression at the 11p15.5 locus. High methylation of the DMR leads to slackened state of imprinting of *Igf2* (biallelic as opposed to monoallelic expression), resulting in inexpression of *H19* in illnesses such as breast cancer (Ehrlich, 2002; Goa *et al*., 2002).

CGIs in general may be deprived of methylation, however this is not the case of the female X chromosome and those in the region of imprinted genes; these may contain methylated or unmethylated alleles. Minimal research had focussed upon regions throughout the genome rather than solely imprinted genes to study allelic differences in methylation. This was prior to an analysis of methylation status of 149 CGIs of the human chromosome 21q (Yamada *et al*., 2004). The group employed a method termed *Hpa*II-McrBC PCR, based on complementary sensitivities of *Hpa*II and McrBC endonucleases. The method identified 31 hypermethylated CGIs, five of which were located in the 5′ promoter region of genes. Analysis of expression in these genes demonstrated that three were testis-specific (when compared to peripheral blood leukocytes); namely the *HSF2BP* (Heat-Shock Transcription

Factor II Binding Protein), the *DKFZp434A171*-LIKE gene and *PPP1R2P2* (Protein Phosphatase 1 Regulatory Inhibitor Subunit 2 Pseudogene II). Full methylation and inexpression of *HSF2BP* and *PPP1R2P2* genes were discovered in blood leukocytes. The *H2B*-LIKE (similar to H2B Histone Family Member S) gene demonstrated ubiquitous expression, and this was also the case for *ADAR2* (double-stranded RNA-specific Adenosine Deaminase). The study identified three allele-specific CGIs designated numbers 59, 112 and 130; former two of which were maternally methylated. CGI-59 exhibited a scattered pattern of methylation; one population in leukocytes demonstrated complete unmethylation, the other being maternally methylated. CGI-112 was maternally methylated in leukocytes and five tissues of placental DNA. Investigation of leukocytes revealed that the differential methylation mostly occurred within a tandem repeat sequence; five sequences that were over 80% identical. The final allele-specific CGI-130 provided the extraordinary factor; irrespective of origin, methylation was confined to a C allele. Upon investigation, some C/G heterozygotes had a paternally-methylated C allele, whereas others were maternal. Mono-allelic methylation was also characteristic of those who were homozygous for a G or C allele. This was the initial discovery of allele-specific parental-origin-independent methylation (Yamada *et al*., 2004).

1.6.1.2 Regulation of ncRNAs

Accompanying the many roles of DNA methylation is the regulation of microRNA (miRNA), which are non-coding RNAs (ncRNAs) believed to partake in human growth and development, apoptosis, and connected to disease such as cancer. Widely experimented in the context of cancer, an inverse correlation exists between variation of DNA methylation patterns and miRNA expression, thereby altering expression of the miRNA target genes (Sood *et al.,* 2006; Figure 1.6).

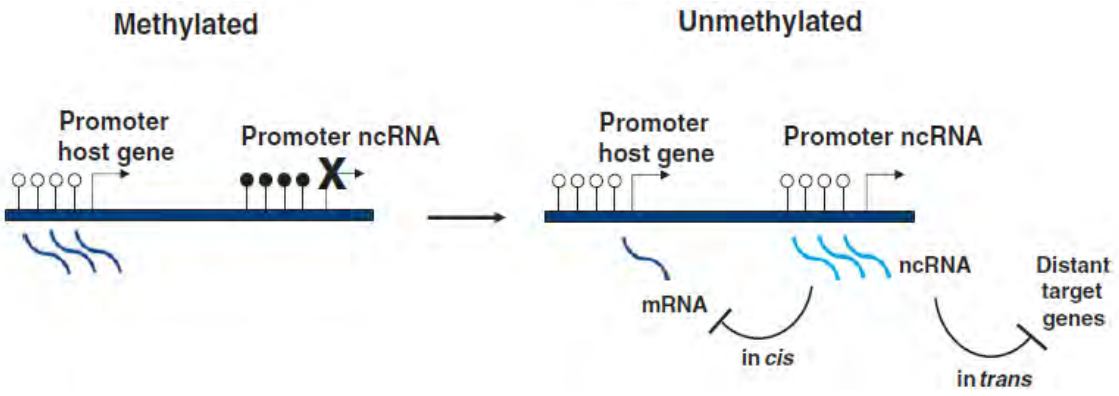

Figure 1.6: Regulation of non-coding RNAs (ncRNAs) by DNA methylation. Reduced methylation of ncRNA-promoters leads to expression of ncRNA and regulation of the ncRNA target genes. Regulation occurs in *cis* or *trans,* affecting the host gene, or distant target genes respectively. Dark lollipops indicate methylation; white indicates unmethylation (Kulis *et al.,* 2013).

The altered expression of miRNA target genes often leads to tumourigenesis (Bierkens *et al.,* 2013; Breuckner *et al*., 2007; He *et al*., 2015). In colon tissue, MiR-124a (human MiR-124-1 stem-loop, microRNA precursor) generally exists in an unmethylated state, however is hypermethylated in colorectal tumours. Inactivation of MiR-124 results in the phosphorylation and expression of the target oncogene, i.e. Cyclin D Kinase 6. Hence *RB1*, the tumour suppressor gene is inactivated (Lujambio *et al*., 2007; Taguchi, 2013). Han *et al*. (2007) demonstrated tight regulation of miRNA by DNA methylation. Contrasting of miRNA expression patterns of a DNMT1 and DNMT3b double knockout cell line to its parental cell line HCT 116 revealed tight controlling of a fraction of the miRNAs by methylation. A futile endeavour to revive activation of miRNA by treatment with 5-aza-2-deoxycytidine (5-aza) or obliteration of *DNMT1* shed light that decreased methylation levels was inadequate for reinstating miRNA expression (Han *et al*., 2007). And yet, only a single microgram dosage of 5-aza was proven sufficient to reduce methylation levels of all promoters analysed by Shen *et al*. (2007).

A medical study of Human Papillomavirus (HPV) by Bierkens *et al.* (2013) found a significant association of DNA methylation and miRNA. Upon analysis of DNA from vaginal secretions, promoter methylation-mediated silencing of MiR-124-2 (MicroRNA-124-2), as well as *CADM1* (Cell Adhesion Molecule I) and *MAL* (T-lymphocyte Maturation

Associated Protein) genes were found to be associated with cervical carcinogenesis. Increased methylation of these genes intensified the severity of cervical disease.

Control of miRNA by methylation may not always be detrimental. The decreased methylation levels of the Let-7a-3 oncogenic miRNA leads to lung adenocarcinoma. Results of the study by Breuckner and colleagues (2007) suggested that aberrant methylation patterns of Let-7a-3 would be a possible mechanism of suppression of the oncogene and reduced expression of the cancer cell line. Similarly, Saito and colleagues (2006) demonstrated that activation of the testis-specific MiR-127 from its promoter in cancer cells is reliant on DNA demethylation. MiR-127 is a tumour suppressor and treatment with 5-aza led to the down regulation of the target proto-oncogene *BCL6.* This particular testis-specific MiR-127 was also analysed, in addition to MiR-142, 338 and 363, by Schilling and Rehli (2008) who detected tissue-specific variation in methylation levels of promoters of various tissues and miRNAs.

1.6.1.3 X Chromosome Inactivation

In mammals, sex is determined by differential inheritance of a pair of dimorphic chromosomes: the gene-rich X chromosome and the gene-poor Y chromosome. To balance the unequal X chromosome dosage between the XX female and XY male, mammals have adopted a unique form of dosage compensation in which one of the two X chromosomes is inactivated in the female.

This highly co-ordinated sequence of events is controlled by X inactivation centre (Xic), except for maintenance. Xic contains *Xist*, a ncRNA gene. It is assumed that a region at the 3′ end of the *Xist* gene is responsible for a counting step in which the number of X chromosomes is measured in relation to haploid autosome sets. Thereafter all but one X chromosome is committed to inactivation and it has been shown that the genes *Xist* and *Tsix* participate in regulating this step. *Tsix* is the second non-translated RNA, transcribed antisense to *Xist*, encoded by the Xic region. Xist RNA accumulates along the X chromosome containing the active *Xist* gene and proceeds to inactivate almost all of the other hundreds of genes on that chromosome (Clerc and Avner, 1998; Lee and Lu, 1999). *Tsix* mRNA has been shown to be a negative regulator of *Xist* and prevents its abundant accumulation on the active X chromosome. Hence, the initiation of silencing relies on *Xist* expression. However, once silencing is established, maintenance of the inactive X is apparently independent of further Xic and *Xist* function. Following this physical deactivation,

large parts of the chromosome are silenced by DNA methylation. For random X inactivation maintenance, DNA methylation is the key stabilizing factor, as deletion of maintenance DNA methyltransferase *DNMT1* results in reactivation of the silenced X in the embryo. In addition, DNA methylation is also required to stably repress the *Xist* gene on the active X chromosome (Gartler and Goldman, 2001; Lee *et al.,* 1999; Wutz and Jaenisch, 2000).

1.6.1.4 Control of Alternate Promoters

Contradictory to the popular notion that an inverse correlation exists between methylated CGIs at promoter regions and gene expression, the concept cannot be applied to CpGs situated within the gene itself. DNA methylation blocks transcription initiation at promoters, but elongation of genes remain unaffected (Bird, 1995; Kulis *et al.,* 2013). Intragenic CpG methylation (which results from elongation-associated nucleosomes such as trimethylated H3K36 (H3K36me3) that recruit DNMTs does not affect the transcription and expression process (Hahn *et al.,* 2011). Thus, gene bodies with high methylation are still transcribed. However, DNA methylation at intragenic CpGs is believed to control use of alternative promoters and create diversity in regulation and expression of main transcripts (Ayoubi and Van de Ven, 1996; Kulis *et al*., 2013; Maunakea *et al.,* 2010).

In a study by Archey *et al*. (1999) to decipher whether CpG methylation at the *TGF-β3* (Transforming Growth Factor-*β3*) gene locus correlates with promoter use, considerable variances in CpG methylation of breast and non-breast cancer cell lines were found to be restricted to just a small group of CpGs flanking the alternate promoter. In breast cancer cell lines, CpGs displayed hypomethylation, whereas non-breast cancer cell lines demonstrated approximately complete methylation. Decreased DNA methylation of the intragenic promoter of the breast cancer cell lines led to transcriptional activation; generating mRNA with enhanced translational potential.

Alternate promoters of 61 genes of liver, stomach, spleen, brain and testis tissues were studied by Cheong and colleagues (2006). Methylation Dependent Restriction Enzyme McrBC-PCR with subsequent bisulfite sequencing showed significant differential methylation patterns amongst promoters that belonged to the same genes (Cheong *et al*., 2006). Additionally, about 62% of the tissue-biased promoters displayed significant variation in methylation profiles among the tissues and tended to exhibit greater variation than non-tissue biased promoters. These tissue-biased promoters displayed greater methylation levels in tissues that they were impartial to, than tissues in which they were favourably expressed in

(Cheong, *et al*., 2006). The pattern of differential methylation levels of a specific promoter in different genes detected in this study may not be applicable to all promoters. Schilling and Rehli (2008) found contrasting results (with a single exception of the *SYBL1* gene); copies of genes such as *VCX* (chromosome X) and *DAZL* (chromosome 3) share practically identical promoter methylation with Y chromosome gene homologues, *VCY* and *DAZ*, respectively.

During a genome-wide analysis employing arrays of BAC clones and the MSREs *Not*I and *Bss*HII, Ching and colleagues (2005) discovered that an intragenic island of the *SHANK3* gene (structural protein in neuronal postsynaptic densities) displays differential methylation patterns and tissue-specific expression. Further investigation of methylation in *SHANK3* and the homologues *SHANK1* and *SHANK2* in brain cortex, lymphocytes, heart and cerebellum revealed numerous differentially methylated CpGs, however only *SHANK3* methylated CpG islands presented promoters displaying high or low methylation with silencing and activation respectively. Majority of the islands were hypermethylated in blood lymphocytes but either reduced or completely nil methylation was observed in brain, heart and cerebellum. By performing MSRE-Analysis and Methylation-Specific PCR (MSP), CpG methylation and expression of *SHANK3* was studied in human neuroblastoma lines; these were found to demonstrate low methylation and were thus expressed. An Epstein Barr Virus-transformed line was abundantly methylated and not expressed (Beri *et al*., 2007). Haploinsufficiency of *SHANK3* is well-documented for its partaking in dendritic spine morphogenesis and is linked to 22q13 deletion syndrome in humans (Ching *et al*., 2005; Roussignol *et al*., 2005).

Rauch and colleagues (2009) compared of expression levels in human male $CD19^+$ B cells in testis, heart and brain and confirmed that promoter methylation contributes to gene silencing. Three genes in the B cells were found to deviate from the hypermethylation/inexpression dogma; despite their highly methylated promoters, *PARP12, MFHAS1* and *MSL2L1* were expressed. Upon 5′ rapid amplification of cDNA end transcripts alternate unmethylated intragenic promoters were detected, not only representative of the role of methylation but also highlighting the origin of cell and tissue-specific alternate transcripts (Kader and Ghai, 2015; Rauch *et al*., 2009).

Shen and colleagues (2007) performed a comprehensive study of promoter-methylation of over 6000 genes, a great portion of which were rich in CGIs, and detected over 200 genes that displayed unusual hypermethylation in normal peripheral blood. This

outlined a clear nonconformity to the philosophy that CGI methylation is restricted to imprinted genes and X inactivation (Shen *et al*., 2007).

1.7 Tissue-Specific Differentially Methylated Regions (tDMRs) and Control of Gene Expression

Tissue-specific gene expression profiling provides vital information about the biology of diverse cell types within an organism and interactions among tissues within multicellular organisms. It is widely known that tissue-specific methylation patterns are a critical aspect of the regulatory mechanisms of tissue-specific gene expression during different phases of development (Doi *et al.,* 2009; Ghosh *et al.,* 2010; Schug *et al.,* 2005; Varley *et al.,* 2013).

tDMRs demonstrate variation in methylation patterns according to tissue/cell/fluid type. Since such regions may be unmethylated in particular tissues, yet display varying degrees of methylation in others, they provide distinguishing characteristics between the tissues (Choi *et al.*, 2014; Schug *et al.,* 2005; Sugimoto *et al.,* 2009). Examples of these tDMRs that provide distinguishing characteristics of tissues have been identified by Eckhardt *et al.* (2006) and Shen *et al.* (2007). tDMRs have been found to play roles in sex-determination (Naito *et al.,* 1993), phenotypic variations (Heyn *et al.,* 2013; Fraser *et al.,* 2012), prediction of age, mortality (Fraga *et al.,* 2005; Marioni *et al.,* 2015) as well as susceptibility to disease and respective treatments (Stirzaker *et al.,* 1997; Tost, 2010). It is assumed that evolutionary dynamics of CpGs where actively functional tDMRs are found in, provide indications for a selective signature (Cohen *et al.,* 2011; Huh *et al.,* 2013). Once established, DNA methyltransferases ensure that methylation patterns of tDMRs remain fixed. This phenomenon was indeed proved by Eckhardt *et al*. (2006) in a study that described methylation profiles at tDMRs to be specific and stable, thus rendering them excellent markers for body tissue identification. Hence it is possible to resolve the identification of extracted DNA samples using epigenetic markers. Studies based on the roles of tissue-specific differential DNA methylation in tissue-specific gene expression are discussed below.

Many researchers have detected an undermethylated trend in male germ-line tissues and fluids. Schilling and Rehli (2008) used a self-developed methyl-CpG immune-precipitation procedure to determine the effect of methylation profiles of CpGs in promoter regions of brain, monocytes and testes upon genes expressed in the respective tissues. The study confirmed tissue-specific expression as in all cases, the methylation levels were less

than 38%. Additionally, even though most of the testes-specific gene promoters displayed elevated methylation levels in somatic tissues, a subgroup of four genes deviated from this category. This subgroup of genes did not display differential methylation; promoters of these genes were hypomethylated in all tissues however, interestingly activated solely in testes. Upon correlation of the data with expression patterns; genes that displayed decreased methylation results of 20% and 19% in brain and monocytes, respectively were expressed specifically in those tissues. Testes-hypomethylated genes, which were the largest group of 38%, were indeed specifically expressed in testes.

A thorough investigation of *FLJ40201, ANKRD30A, FTMT, SOHLH2, C12orf12, INSL6* and *DPPA5* genes in assessment of cell and tissue-specific CGI promoter methylation was conducted by Shen *et al.* (2007). Bisulfite pyrosequencing was used to quantitatively analyse testes, sperm, blood, breast, liver, colon, fibroblast and skeletal muscle of human origin. All promoters demonstrated hypermethylation, with exception of testes and sperm tissues. Sperm DNA presented alleles without methylation, and alleles of testes were either almost completely hypo- or hypermethylated. *FLJ40201, ANKRD30A, SOHLH2, INSL6* and *DPPA5* genes were specifically expressed in testes and the same genes, with exception of *SOHLH2,* were identified in sperm. To confirm that this group of genes fit into an exclusive class of promoter CGI-associated genes that are methylated and therefore silenced in a tissue-specific manner, final analysis revealed promoter hypomethylation of below 21% for both *INSL6* and *SOHLH2* placental tissue. The specific hypermethylation/inexpression and demethylation/derepression of somatic and germ-line promoters respectively was attributed to regulatory sequences (Shen *et al*., 2007). The proposition spurred from previous findings that transcription factors, such as male germ-line specific CTCF-paralogous BORIS (Brother of the Regulator of Imprinted Sites) may initiate demethylation of cancer-testis antigens *MAGE-A1* in somatic cells (Kitamura *et al*., 2007; Schilling and Rehli, 2008; Vatolin *et al*., 2005).

Eckhardt *et al*. (2006) piloted DNA methylation profiling of specifically human chromosomes six, 20 and 22. Forty-three samples consisting of twelve various tissues were selected for the study, which reported the methylation status of nearly two million CpG sites. The pursuit of tDMRs spotted a minimal portion located in the 5′ regions and exons, but a third of all non-coding regions were tDMRs. In the attempt to correlate methylation with expression: a total of 53 genes were arbitrarily selected. The intragenic tDMRs did not display any association with expression and the same could be said for 63% of the 43 genes

that were associated with 5′ untranslated regions (UTRs). The rest within this category demonstrated an indirect correlation. The *OSM* (Oncostatin) gene reflected an inverse relation to expression which was quite interesting as the gene does not contain a CGI in the 5′ region.

Rakyan and colleagues (2008) provided a significant contribution to studies of methylation. The group developed the aptly titled Batman program (Bayesian Tool for Methylation Analysis) which provides estimates of absolute levels of methylation following MeDIP profiling (Down *et al*., 2008) and performed an all-inclusive genome-wide tDMR quest. The methylation profiles that were deciphered by the group, of 13 different somatic tissues including blood, placenta, B cells and lung amongst others; sperm, placenta and the immortalized EBV-transformed cell line GM06990 were included in the initial ENCODE study (ENCODE Project Consortium, 2007; Rakyan *et al*., 2008). In contrast to observations by Eckhardt *et al*. (2006) who studied three chromosomes, and a chromosome-wide search by Weber *et al*. (2007) this study by Rakyan *et al.* (2008) found that most promoters that were not associated with a CGI displayed unmethylated profiles. An overall negative correlation was detected for methylation and expression, for example in tissues wherein expression was observed, the transcription start site was not methylated. Comparing their results to previous studies with contradictory results led to the unreciprocated query as to why some promoters act irrespective of DNA methylation status, and others depend on methylation for regulation. For 16 tissues tested, just above 65% of CGIs associated with promoters were also unmethylated. Although located throughout the genome, the promoter-associated tDMRs were positioned mostly within those with mediocre CpG densities. tDMRs represented 18% of the entire genomic region studied, and conforming to results obtained by many other research studies, a considerable percentage of all tDMRs were sperm-specific. These displayed hypermethylated profiles in all tissues, and low methylation in mature sperm. Intragenic tDMRs displayed a positive methylation/expression profile; the *ICAM3* (Intercellular Adhesion Molecule III) gene presented low methylation at the promoter, but hypermethylation in expressed tissues (Rakyan *et al*., 2008). Additional tDMRs to distinguish semen from other tissues were identified by Igarashi *et al.* (2008), who also detected an age-related linear correlation of DNA methylation in the testes tissues.

Rienius *et al*. (2012) and Lam *et al*. (2012) established that patterns of methylation in blood display greater variations between cell populations rather than between individuals. Rienius and colleagues (2012) explored whole blood samples as well as its components. Variations were found between the low methylated CpGs of the myeloid cell population

(monocytes, neutrophils and eosinophils) and the highly methylated lymphoid population (B, C, NK cells). In whole blood as well as individual populations, distributions of differentially methylated patterns were mostly intragenic. The unmethylated cell populations were subjected to gene ontology enrichment analysis to determine cell-specific functions. Even though enrichment for the eosinophils was more typical of general cell functions instead of specific, the peripheral blood mononuclear cells (PBMCs) demonstrated cell-specific enrichment pathways. This included T cells involved in leukocyte and lymphocyte activation and NK cells in molecular signalling cascades. Functions of genes exhibiting variation in methylation were highlighted more by the cell-specific profiles of methylation of a set of genes according to surface expression. For example, membranous expressions of the *CD14* and *CD3* genes which demonstrated elevated methylation profiles in the cells that they were unexpressed. Strangely, B cells demonstrated overall largest variations in methylation. As the B cells evidently epitomize, methylation patterns found in the study were well correlated to cell specific functions. They are involved in numerous critical pathways and roles such as the humoral immune response, presentation of antigens, and internalization amongst others, which estranges them from T cells. Eighty-five percent of genes that were selected during inferences to inflammatory diseases such as asthma, atopic dermatitis, inflammatory bowel disease, rheumatoid arthritis, and diabetes were differentially methylated. The *TNF* and *LTA* genes (Tumour Necrosis Factor and Lymphotoxin Alpha, respectively), which are both associated with asthma, proved differentially methylated in the promoter regions only. In contrast, *TCF7L2* (Type 2 Diabetes Candidate Gene Transcription Factor 7-Like II), positioned on the 10q25.3 chromosome, exhibited an analogous methylation profile throughout the gene however methylation toward the promoter islands tended to decline. Some CpG sites of $CD14^{+}$ monocytes were specifically unmethylated (Rienius *et al.,* 2012). A thorough expression study consisting of 13 assays by Prokunina-Olsson *et al*. (2009) proved that the ex7–8 isoform of *TCF7L2* is activated specifically in these monocytes with lowest levels of expression in activated T cells and B cells.

A particular example of research that did not find an association between DNA methylation and gene expression was by Grunau and colleagues (2000). The group performed a detailed methylation analysis of three X linked genes, namely *MSSK1* (Muscle-specific Serine Kinase), *CDM*, *SLC6A8* (Creatine Transporter) and the pseudogene *ψSLC6A8* in eight tissues. The *MSSK1* gene presented low methylation patterns in prostate, heart and brain and intermediate methylation in kidney, muscle, pancreas and lung tissues but was only

specifically expressed in muscle and heart. In spite of overall strong methylation in brain and heart, and hypomethylation in the prostate the *CDM* gene was expressed in all tissues. Similar to *MSSK1* and *CDM,* there was no distinct pattern of methylation and gene expression in *SLC6A8.* Hypomethylation was observed in prostate and the additional tissue tested, i.e. testis. A fascinating finding was that even though the methylation profiles of liver and muscle were nearly indistinguishable, the gene is inhibited in the latter and specifically expressed in the former. The autosomal pseudogene *ψSLC6A8* was the only gene that showed an affiliation between methylation and gene expression. In addition to the nine tissues tested, methylation of white cerebral matter from seven more participants were profiled. High intensities of methylation were observed in all, except testis which was completely free of methylation. High methylation rendered the pseudogene silenced in all somatic tissues, and demethylation enabled testis-specific expression. In somatic cells, the pseudogene *ψSLC6A8*, as well as the *PDHA*-2 and *PGK*-2 genes (Pyruvate Dehydrogenase (Lipoamide) Alpha II and Phosphoglycerate Kinase II, respectively) are all silenced due to methylation, yet unmethylated and active in male germ-line cells. A plausible offering was that transcription of testis-specific pseudogenes may be a by-product of the transient demethylation during the course of spermatogenesis (Ariel *et al*., 1991; Grunau *et al*., 2000; Iannello *et al*., 1997). It is conceivable that lack of correspondence with expression would simply be that the studied tDMRs may not have been situated in regions that govern expression. Or rather another rational explanation, as suggested later by Illingworth and Bird (2009) the intragenic methylation status encumbers gene body transcription with subsequent hindering of parent genes.

Recent studies have identified unorthodox non-CpG methylation patterns in the brain (Lister *et al.,* 2013), embryonic stem cells (Lister *et al.,* 2009) and germ cells (Kobayashi *et al.,* 2013). Schultz *et al.* (2015) also detected this occurrence in human post-mortem tissue samples including lung, pancreas, sigmoid colon and liver. A negative relationship between the DMRs and gene expression was observed in *MYH10,* which is linked to blood vessel function. Hypomethylated DMRs in aorta overlapped with aorta-specific enhancers, indicating that decreased methylation levels corresponded with tissue-specific functions. Non-CpG methylation was most evident in purified glia, brain neurons and H1 embryonic stem cells. Non-CpG methylation correlated with positive expression in H1 cells in the study by Lister *et al.* (2009) but both Lister *et al.* (2013) and Schultz *et al.* (2015) found a negative correlation, thus the function of non-CpG methylation is not known. (Schultz *et al.,* 2015).

Such studies of non-CpG methylation and unknown roles in expression and development only highlight the poorly understood role of DNA methylation.

1.8 Tissue-Specific Differentially Methylated Regions (tDMRs) and its Application in Forensic Science

DNA methylation provides a selective signature has been exploited in medical laboratories on a very wide scale to target specific tissues and genes for diagnostic and therapeutic purposes. Now much effort has been placed on researching tDMRs for forensic science applications by various scientists worldwide.

Analysing the methylation status of DNA is favourable as in addition to demonstrating high sensitivity and specificity, simple extraction and purification methods may be attuned with novel nucleic acid technologies. Since the methods target only extracted DNA, there is no additional consumption of physical material. This is essential in criminal cases when sample quantities are limited. Therefore forensic scientists are not faced with decisions of which assays to perform for fear of loss of crucial evidence. A major advantage of DNA methylation-based assays in forensic sciences is efficiency, convenience and the analysis of multiple tissues in a single assay (An *et al.,* 2013; Choi *et al.*, 2014; Frumkin *et al.,* 2011; Gunn *et al.,* 2014; Madi *et al.,* 2012). PCR-based techniques are mainly used to analyse DNA from biological material. DNA databases are then scanned for matches to DNA profiles found at crime scenes. There are countless commercially manufactured DNA systems that have been approved for forensic applications (Gunn *et al.,* 2014; Romeika and Yan, 2013; Vidaki *et al.,* 2013).

tDMRs may be employed on their own, irrespective of whether the genes they represent are expressed or not to differentiate between body fluids, and gene expression studies may also be employed to provide indications of sex, age and ancestry informative markers. Hence, application of tDMRs in forensics is discussed below.

1.8.1 Verification of DNA Samples

It is generally believed that every trace of DNA found at a crime scene is of biological origin, although this is not always the case. This is quite disturbing, especially since DNA evidence holds such heavy weight in the courtroom. Frumkin *et al*. (2010) considered the ease of fabricating DNA evidence by criminals, since only basic equipment, minimal financial expense and no particular expertise would be necessary. There is an alarming

possibility that DNA planted at crime scenes may be overlooked during forensic investigations. By artificially synthesizing blood, saliva and skin, mock forensic samples were generated. A selection of tDMRs, which were known to be methylated or unmethylated at particular regions, were analysed in an authentication assay. Methylation profiles of artificial DNA were generated using bisulfite conversion followed by PCR amplification, and compared to natural human DNA samples. The research revealed that artificially synthesised DNA is consistently unmethylated in all loci whereas biological DNA displays methylation in some loci and unmethylation in others. This study, using differential methylation, provided a mechanism that enables forensic scientists and judicial law to convict the correct criminal, rather than one wrongfully accused. Methylation patterns of tDMRs were used to uphold the credibility of trace DNA located at crime scenes, which is unquestionably of extreme importance in forensic investigations.

1.8.2 Identification of Forensically Relevant Body Fluids/Tissues by Analysis of tDMRs

In forensics-based research, one of the first fruitful attempts at DNA methylation-based tissue identification was by Frumkin *et al*. (2011) leading to successful differentiation between blood, saliva, semen and skin samples. The authors selected 16 loci that displayed considerable differential amplification patterns and subjected a mere 1 ng of DNA from samples to methylation sensitive restriction enzyme PCR (MSRE-PCR) followed by capillary electrophoresis. The tissue identification assay was designed such that loci with low degrees of methylation were amplified with low efficiency and demonstrated weak signals on the resulting electropherogram, and loci with high degrees of methylation conversely displayed strong signals. A methylation ratio for each pair of differentially methylated loci was calculated, allowing for discrimination between the biological fluids. Precise identification of semen was achieved by low methylation levels of the *L91762* gene corresponding with hypermethylation of the *L68346* gene. Methylation levels of *L91762* were much higher in blood, saliva and skin. A low ratio of *L91762/L68346* was again used to differentiate semen samples from other tissues. The methylation ratios of *L76138/ L26688* genes were higher in semen and skin than in blood and saliva. Identification of skin epidermis was confirmed by high ratios of methylation in *L91762/L68346* as well as *L76138/L26688*. The assay positively identified tissues from a single source, as well as those in mixtures as a combined semen assay was also performed. Semen was mixed in varying ratios with urine and saliva, and the absence or presence of semen was detected.

In a study by Madi *et al*. (2012) bisulfite modification and pyrosequencing procedures were employed to identify new tDMRs. Focus was directed at particular CpG sites in tandem as flanking sites may display differing methylation levels. Methylation levels at single CpG sites were compared between samples. Various CpG sites of *ZC3H12D, FGF7, C20orf117* and *BCAS4* genes were analysed to display differential methylation profiles of semen, blood, and saliva and skin tissue. *ZC3H12D* and *FGF7* successfully differentiated semen from other fluids. All five CpG sites tested in *ZC3H12D* displayed high methylation levels ranging from 82-100% in blood, saliva and skin, but hypomethylation, below 12% in sperm. Presumed initially to distinguish skin samples from others; the FGF7 marker enabled identification of semen, with hypermethylation relative to blood, saliva and skin. Blood, in particular white blood cells, was positively identified by the *C20orf117* locus due to hypermethylation. High levels of methylation of $CD4^+$ and $CD8^+$ lymphocytes relative to skin and sperm were demonstrated by Eckhardt *et al*. (2006). The C20orf117 marker showed great prospective for the differentiation of blood from other tissues however, one must consider that accurate differentiation from skin was inconclusive in the experiment as only one skin sample could be amplified due to errors in bisulfite conversions or primer hybridization. Also with this marker the methylation of sperm was particularly low compared to other fluids. Based on the outcome of the report by Eckhardt and colleagues (2006), Madi *et al.* (2012) expected a high level of methylation for semen in the BCAS4 marker. However this result was not reproduced as sperm displayed less than 20% methylation and saliva displayed highest methylation; thus this marker was designated for saliva identification. The same group of researchers conducted a study in 2013, testing the same tDMRs and body fluids. Analogous results were obtained for the markers; and in addition the study determined mean percent methylation of nine year old blood samples, and 20 year old blood and semen samples. Interestingly, methylation patterns were observed to be unwavering over such long periods of time; percent methylation was the same as samples that were recently collected (Antunes *et al*., 2013; Madi *et al.,* 2012).

In a study by Lee *et al*. (2012) DNA methylation profiles of blood, saliva, semen, vaginal fluid and menstrual blood were generated by a selection of five tDMRs. tDMRs for *USP49* and *DACT1* genes (Ubiquitin-Specific Peptidase 49 and Dapper 1 Isoform II, respectively) were selected as semen-specific markers, and tDMRs for *PFN3*, *PRMT2* and *HOXA4* genes (Profilin III, Protein Arginine N-methyltransferase II and Homeobox A4, respectively) were chosen as blood-specific markers as different methylation patterns in

blood, spleen and brain tissues were observed in previous reports by Illingworth and colleagues (2008). Bisulfite conversion of DNA followed by PCR and sequencing demonstrated differentially methylated patterns in all the tested tissues and fluids. *DACT1* and *USP49* were unmethylated in over 90% of clones from semen and hypermethylated in almost all blood, saliva, vaginal fluid and menstrual blood clones. The methylation profiles of tDMRs did not significantly differ in these tissues and fluids, thus they may be used to positively identify semen which will aid greatly in sexual assault cases (Lee *et al.,* 2012). The VASA tDMR may also be used to identify semen, as it is hypomethylated in testis and demonstrates high degrees of methylation in other tissues. Furthermore, it is expressed solely in germ cells (Sugimoto *et al.,* 2009). The HOXA4 tDMR displayed high degrees of methylation in blood and female saliva, but there was no considerable difference in these methylation patterns to allow accurate differentiation. The HOXA4 tDMR was hypomethylated in vaginal fluid and menstrual blood, and the *PFN3* gene displayed methylation of 65% of loci in vaginal fluid whereas more than 80% methylation was observed in other tissues and fluids. PRMT2 was hypermethylated in vaginal fluid and menstrual blood, and demonstrated great differences between semen/vaginal fluid, and semen/menstrual blood. From this data, the authors suggested that low methylation of HOXA4 and high methylation of PRMT2, USP49 and DACT1 may be used to confirm the presence of vaginal fluid and menstrual blood. For testing pooled DNA of semen and vaginal fluid, the combined use of *DACT1, USP49, PRMT2* and *PFN3* genes may be employed. However any issues regarding mixtures of samples, which are often the case in real-life situations, may be alleviated by *USP49* and *DACT1* genes, as methylation patterns are quite distinct in semen compared to the other tissues. The same study also tested aged samples whereby all tissues and fluids were left at ambient temperatures for 30 days. A methylation-specific PCR procedure revealed that all unmethylation and methylation patterns remained consistent over this period (Lee *et al.,* 2012).

An *et al*. (2013) and Choi *et al*. (2014) studied the same tDMRs USP49, DACT1, PFN3 and PRMT2 for body fluid identification. An and colleagues (2013) demonstrated age related changes in methylation (discussed below) and methylation profiles of genes within blood, saliva, semen, vaginal fluid and menstrual blood. In addition to MSRE-PCR, the study also determined methylation profiles using a less common method known as Methylation SNaPshot. The method involves a multiplex individual-base extension, designed using in silico bisulfite converted genomic reference sequences for specific genes, with subsequent

PCR amplification. Using MSRE-PCR, blood and saliva were found to display high methylation yields (indicated as peak height ratios of methylation status of tDMRs to amelogenin) in DACT1, followed by PFN3, USP49 and lowest methylation in PRMT2. Levels of methylation in DACT1 and USP49 were significantly higher than PRMT2 and PFN3 for vaginal fluid and menstrual blood. Analogous results for MSRE-PCR and methylation SNaPshot were observed as blood and saliva displayed over 90% methylation in USP49, DACT1 and PFN3. Similar results to Lee *et al*. (2012) were achieved for vaginal fluid and menstrual blood as high methylation was observed at USP49 and DACT1, and low methylation at PFN3. Yet, in contrast to the same study, semen was found to display complete unmethylation at the tDMR of PFN3 using both methods (Choi *et al.*, 2014; Lee *et al.,* 2012). Choi and colleagues (2014) employed MSRE-PCR and tested the same fluids using tDMRs USP49, DACT1, PFN3 and L81528. Vaginal fluid and menstrual blood, analogous to the two previously described reports by Lee *et al*. (2012) and An *et al*. (2013) displayed low methylation levels at PFN3. The L81528 tDMR was selected as a semen-specific methylation marker; ten out of the 18 non-vasectomised semen samples generated amplicons only at this tDMR. Sensitivity tests were conducted by testing varying quantities of DNA produced by serial dilutions. Methylation patterns could be generated for saliva and semen with just 500 pg or more of DNA and a measly 250 pg of DNA from vaginal fluid was sufficient. Aged samples generated identical results to An *et al*. (2013) as all fluids were analysed except for saliva. Mixtures of saliva and semen, in 1:1 and 1:2 ratios were clearly distinguished by amplification of L81528. Mixtures of semen and vaginal fluid was also distinguishable as profiles were comparable to that of vaginal fluid when using a general DNA extraction method, and a differential extraction method facilitated identification of vaginal fluid from the resultant supernatant, and semen from the pellet. A single post-coital penile sample and three post-coital vaginal samples were tested for an artificial sexual assault case. General DNA extraction methods resulted in two of the vaginal fluid and the penile samples showing a mixed sample profile by generating low peaks at PFN3, and a semen-specific L81528 amplicon. Again, the differential extraction method facilitated a display of a profile consistent with vaginal fluid, for the third post-coital sample (Choi *et al.*, 2014). From the above studies it is clear that additional tDMRs are required for precise discrimination between blood and saliva, and vaginal fluid and menstrual blood.

Wasserstrom *et al*. (2013) introduced a kit known as DNA-source identifier (DSI-semen) for confirmation of semen. The study conducted methylation profiling of five genome

loci using DNA from blood, saliva, semen, vaginal fluid, and menstrual blood and urine samples. The genome loci were located on chromosomes two, four, eleven and 19. All samples that were tested were correctly identified with confidence levels above 0.9999. Real forensic casework samples were also analysed and all were in complete concordance except for one discrepancy; a single sample displayed semen and non-semen components. However, further investigation revealed that the assay was more accurate than even the much-relied microscopic analysis which did not detect semen at all. Mock casework samples were also correctly recognised; vaginal swabs were mixed with varying quantities of semen. In each sample the pellet was classified as semen and supernatant as non-semen. This will enable scientists to apply the assay for analysis of internal vaginal fluid samples obtained from victims of assault. The DSI-semen assay correctly discriminated between semen and male urine, rendering it quite a reliable tool. The assay is user-independent, fully automatable and operating the software would not require any special training. Furthermore, there is now potential for the development of analogous assays for identification other common fluids and tissues found at crime scenes, as the only variable would be primers. The same setup, reagents and computer software would be employed.

Park and colleagues (2014a) performed an excellent pyrosequencing analysis of 2986 differentially methylated CpG sites of blood, saliva, semen and vaginal fluid. Eight markers of 20 body fluids each, namely cg08792630 and cg06379435 for blood, cg20691722 and cg26107890 for saliva, cg17610929 and cg23521140 for semen, and cg14991487 and cg01774894 for vaginal fluid, were examined. Overall, each marker displayed hypermethylation for its respective body fluid with significantly reduced methylation in the others. To confidently apply the markers in forensic casework, the group tested various concentrations of DNA from each fluid, ranging from 5-500 ng. Inaccurate results were obtained only when 5 ng was tested and thus the authors suggested a minimum of 10 ng DNA to be sufficient for precise distinction of fluids. Although the study provided a great contribution to the application of DNA methylation in forensics, the group did experience difficulty in differentiation of methylation levels of saliva and vaginal fluid for a few samples. To alleviate this, a suggestion of more reliable DNA methylation markers was offered (Park *et al.,* 2014a).

1.8.3 Sex Determination in Forensics

Gender identification is vital in forensic sciences not only in sexual assault cases to distinguish between criminals and victims but also for investigating remains of mass disasters and missing persons cases (Butler, 2012). Numerous techniques have been developed over the past few years including visualisation of digested DNA on agarose gels as well as PCR-based methods (Manucci *et al.,* 1994). Currently differential DNA methylation is employed on a regular basis for sex-typing in forensics, and even though DNA methylation profiling is relatively new, it was reported long ago (Naito *et al.,* 1993). The study introduced DNA methylation into forensic sciences by developing a simple procedure for female sex typing based on varying methylation pattern of DXZ4, an X chromosome-specific region. The DXZ4 sequence showed low degrees of methylation on the inactive X, but hypermethylation on the active X chromosome. The protocol was quite sensitive as due to the high copy number of DXZ4 in the genome, only a minute amount of DNA was necessary for accurate sex typing. The researchers also suggested application for detecting sex-reversed patients (Naito *et al.,* 1993; Vidaki *et al.,* 2013).

The amelogenin system, first introduced by Sullivan and colleagues (1993) is a popular sex-differentiating method. It is extremely sensitive and requires only a single set of primers to amplify both X and Y chromosomes, providing equal copy numbers of both sequences. Manucci *et al.* (1994) tested only 1 ng of male DNA in a 100-fold excess of female DNA from blood, bone and muscle samples and successfully differentiated between male and female donors. An *et al.* (2013) and Choi *et al.* (2014) also successfully used amelogenin in combination with other tDMRs for sex-differentiation and body fluid identification.

Sarter *et al.* (2005) reported significant sex differences in four autosomal genes of blood, (*ESR1*, *MTHFR*, *CALCA* and *MGMT*) and suggested that sex is at least as strong a predictor of methylation in certain genes as is age. However, later on, the study by Eckhardt and colleagues (2006) did not find any statistical difference between male and female samples.

The study by Boks *et al.* (2009) revealed independent links of DNA methylation with gender; all X chromosomal probes were extensively more methylated in females than males. According to Shen *et al*. (2008) this pattern of methylation loci on X chromosomes could primarily be due to the X inactivation mechanism in females. Numerous autosomal loci of

CpGs were also found to display differential methylation in females (Boks *et al.,* 2009), and this was also demonstrated previously (El-Maari *et al.,* 2007) signifying that gender-dependent methylation may not be an uncommon occurrence. Overall, the variation in levels of methylation between the two sexes were not too high, but still considerable.

Zhang and colleagues (2011a) examined *LINE1* (Long Interspersed Repeat Sequence) of peripheral blood of males and females. This study reported a 1.8% difference in global methylation between the two sexes, with females being lower. This is a notable pattern; Hsuing *et al*. (2007), and Zhu *et al*. (2012) found a 1.17% and 0.8% methylation difference, respectively, between males and females with females being lower. In addition to the proposed idea that X chromosome inactivation mechanisms diminish the resources necessary for methylation of autosomal loci, it is also believed that low degrees of total methylation in females may be due to consumption of varying amounts of one-carbon nutrients such as methionine and B vitamins, or dietary folate (Shen *et al.,* 2008; Zhang *et al.,* 2011a).

In the research study by Lee *et al*. (2012), a tDMR for the *HOXA4* gene was hypermethylated (above 90%) in male saliva, yet displayed below 53% methylation in female saliva. Additionally, low methylation levels of the HOXA4 were generated upon testing female menstrual blood, with corresponding high methylation of PRMT2, USP49 and DACT1. However, cross-reactivity with other fluids may interfere with results and hence this must be considered during application in forensics. Perhaps other confirmatory methods, such as use of one or more of the mentioned tDMRs or new markers, may be employed in conjunction.

1.8.4 Ancestry Informative Markers of Fluid/Tissue Donor

Direct genotype–phenotype relations are not easily accounted for since majority of variant sites are located in noncoding loci (Kilpinen and Dermitzakis, 2012). Nevertheless, genome-wide association studies have endeavoured to establish genetic associations solely with differences between populations (Bell *et al.,* 2011; Fraser *et al.,* 2012; Lam *et al.,* 2012); diseases (Abdolmaleky *et al.,* 2006; Dunn *et al.,* 2014; Tost, 2010); and the response to external stimuli (Jimenez-Chillaron *et al.,* 2012; Terry *et al.,* 2008). In these contexts, the epigenetic network is expected to add layers of regulation, suggesting interplay between the genotype and epitype in gene regulation and phenotypic variation (ENCODE Project Consortium, 2012). Examples for epigenetic differences between individuals are rare and

mostly, but not exclusively, confined to the level of DNA methylation (Rakyan *et al.,* 2002). With new advances in technology, other than differentiating and finding links between individuals based on gender, age and diseases, great progress has been made in determining ones ancestry informative characteristics using differential methylation.

Research of different racial groups has illustrated distinguishable patterns of DNA methylation, gene expression and the onset of disease (Adkins *et al*., 2012; Kuzawa and Sweet, 2009). Hypermethylation of the tumour-suppressor gene, *PYCARD* (~62% vs. ~22% in benign controls) was detected in non-Black prostate cancer patients, yet this was not observed in Blacks (67% vs. 58% controls) implying that the gene is hypomethylated in non-tumorous tissues of non-Blacks, however upon the onset of the disease increases to the general level that is present in tumorous and non-tumorous tissue of Blacks (Adkins *et al*., 2012; Das *et al*., 2006). Mokarram and colleagues (2009) detected higher levels of methylation in three colorectal cancer candidate genes *ICAM5, GPNMB* and *CHD5* (Intercellular Adhesion Molecule V, Glycoprotein Transmembrane and Chromodomain Helicase DNA Binding Protein V, respectively) of Blacks as opposed to Iranians, accounting for the higher incidence and mortality rates of Blacks.

A study quite relevant to discriminating ancestry informative markers which is highly applicable to forensic sciences, observed differences in Blacks, Whites and Hispanics in New York City birth cohort until mid-life. Blood samples were collected and analysed by [^{3}H]-methyl acceptance assay which involves incubating DNA with [^{3}H] SAM in the presence of *Sss*I prokaryotic methylases. This will methylate unmethylated CpG sites, and is quantified followed by statistical analysis relevant to the study. In the birth record, Blacks were found to display lower levels of methylation than Whites and Hispanics. This association between DNA methylation and the groups did not alter after adjusting life course variables. Other factors such as prenatal, early life, adolescent and adult life variables; namely ethnicity, prenatal smoke exposure, birth length and weight, smoking, amidst others were considered. Exposure to prenatal smoking was associated with considerably elevated levels of DNA methylation however insignificant results were obtained for smoke exposure later in life. Overall, Blacks demonstrated lowest levels of methylation (Terry *et al.,* 2008).

Non-Hispanic Blacks, Non-Hispanic Whites and Hispanics were also studied by Zhang *et al*. (2011a) who examined *LINE1* (Long Interspersed Nuclear Element) of peripheral blood. The correlation of leukocyte DNA methylation to ageing, behavioural and

environmental factors between different ethnic groups was also described in this study. Analogous to Terry *et al.* (2008) Blacks displayed lowest percentage methylation of 73.1%, Hispanics 74% and Whites 75.3%. Females exhibited lower overall methylation compared to males, and obesity similarly demonstrated an inverse relationship with global methylation. There were no observed differences in methylation when adjusting for variables such as smoking, alcohol, body mass index and education (Zhang *et al*., 2011a).

Whilst research studies demonstrating reduced methylation in healthy tissues and genes of Blacks (Adkins *et al.,* 2012; Terry *et al.,* 2008; Zhang *et al.,* 2011a) may provide significant contributions to long-term disease studies and applications such as forensics, this is not a general trend as exemplified by Straughen *et al*. (2015) who examined *Igf1* of the mononuclear fraction of umbilical cord blood to model the role of methylation as a mediator between birth weight and race. This informative study revealed that higher methylation levels in Black neonates accounted for variation of *Igf1* expression thus resulting in a substantial decrease in birth weight when compared to non-Black counterparts. Studies of methylation in Blacks particularly, are unpredictable in different genes and fluids, necessitating further exploration to apply the concept of reduced methylation with confidence.

Elaborate investigations by three groups of researchers have gained insight into differential methylation and expression associated with HapMap lymphoblastoid cell lines (LCLs) with varying results (Bell *et al*., 2011; Fraser *et al*., 2012; Moen *et al*., 2013).

In a study aiming to relate genetic variation and differential methylation, Bell and colleagues (2011) analysed promoter methylation of 13 236 genes of 77 human LCLs from the HapMap Yoruba Collection. Yorubans are an African ethnic group, chiefly of Nigerian descent. As most studies have observed, a lower level of methylation was detected near the promoter regions of genes as opposed to intergenic regions (Cheong *et al*., 2006; Schilling and Rehli, 2008; Shen *et al*., 2007). A remarkable number of 11 657 genes displayed a negative correlation between DNA methylation and expression; a reduced methylation level near transcription start sites was identified with those genes that were highly expressed. The group confirmed that methylation and histone modifications are not mutually exclusive in regulation of gene expression; a resilient negative correlation was detected between histones that target expressed genes and methylation levels. Lack of methylation positively correlated with histones involved in activating genes, such as H3K9ac, H3K27ac and H3K4me3. These particular histone marks, along with a few others were observed to exhibit lower levels of

differential methylation in HapMap Yoruban LCL samples when compared to those of European origin by Moen *et al*. (2013) who profiled differential methylation and implications for traits between the two populations. Overall, greater differential methylation was observed intergenically rather than at transcription start sites. As exemplified by the *STK39* gene (Serine Threonine Kinase 39), European samples exhibited higher differential methylation in the regions 1 kb upstream of transcription start sites to the first gene body quantile, with fairly depleted regions towards the 3′ UTRs, whereas the opposite was true for the Yoruban population who exhibited elevated differential methylation more towards the 3′ UTRs. A negative correlation between promoter-methylation of cg27270541 and expression levels was eminent for the *PLA2G4C* (Phospholipase A2, group IVC) gene in the European population, contrasting with high expression levels when elevated methylation was detected in the gene body (Moen *et al*., 2013).

A further step was taken to decipher if genetic variants conveyed inter-individual variation in methylation levels and in the resulting expression. The Yoruban LCLs assessed by Bell and colleagues (2011) displayed an association between methylation Quantitative Trait Loci (meQTL) and gene expression. A substantial number of genome-wide meQTLs that were identified demonstrated an indirect relationship between distance and target CpG sites, suggestive of a *cis*-acting pattern. A particular population-specific meQTL noted by Moen *et al*. (2013) was the G allele of rs28544087; associated with higher methylation level of cg09307883 in *ANAPC2* (Anaphase Promoting Complex Subunit II) in the Yoruban, but not European LCLs.

One hundred and eighty SNPs that were meQTLs demonstrated enrichment for expression, for example the single nucleotide polymorphism (SNP) rs8133082, which was a meQTL and expression QTL (eQTL) for *C21orf56.* Located on chromosome 21, the gene is hypothesised to serve as a predictor of inter-individual variations upon exposure to DNA damaging agents and partakes in the reduced sensitivity of LCLs to N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), an alkylating agent (Bell *et al*., 2011; Fry *et al*., 2008). Initially, Bell and colleagues (2011) found ample association between methylation and numerous SNPs that were eQTLs for genes with reduced signals; indicative of elevated methylation mediating lower gene expression. However, upon re-analysis of the eQTL results and considering methylation as the regulator, methylation proved accountable for 10% account for the effects of expression associated with SNPs.

Nevertheless, the investigation by Moen *et al*. (2013) accentuated the concept of methylation playing a major role; *T* allele of meQTL rs10779587 exhibited high methylation levels of cg01313622, concomitant with reduced expression of *FLVCR1* (Feline Leukaemia Virus Subgroup C Cellular Receptor I). *T* allele was less frequent in the Yoruban samples, and thus higher expression levels were obtained. A sizable number of meQTLs were linked to racial differences in neurological, metabolic and autoimmune disorders. For example, three SNPs associated with rheumatoid arthritis, multiple sclerosis and prostate cancer were annotated as a meQTL for a CpG located intergenically within *LOC285830* (Homo sapiens hypothetical LOC285830). All three SNPs demonstrated a higher risk allele frequency in the European LCL samples as opposed to Yoruban.

Additionally, another study based on HapMap LCLs between European and Yoruban populations by Fraser *et al*. (2012), detected an array of differentially methylated CpG sites within, as well as between populations. The group employed HumanMethylation27 BeadChip arrays followed by pyrosequencing to determine methylation of 27 578 CpG sites near transcription start sites of 14 495 genes. The study population comprised of 30 trios (consisting of father, mother and offspring) of the Europeans and 30 trios of families of the Yorubans. Methylation profiles of approximately 30% of CpG sites, which is over a third of the genes examined, differed between the populations. Specifically, 14% of sites differed in methylation by an average of over 5%, and 3.9% of sites differed by over 10%. The differences were small but somewhat significant, and consistent with findings of Bell and colleagues (2011). DNA methylation was liable for little differential gene expression that existed between the groups. Delving more into the reasons for observed variances between the populations, the group deciphered the role of additive genetic variances to methylation of CpG sites, also termed heritability. Measuring the link in methylation levels between the trios of parents and offspring revealed transmissible methylation at about 900 CpG sites in the Yoruban population, and over 700 in the European LCLs were discovered. This was implicative of genetic control of methylated polymorphisms being, albeit less transmissible than expression, quite common. Since failing to establish an analogous pattern of heritability between the populations; divergences impacting methylation were further investigated by comparing the SNPs linked to methylation of CpG sites in the populations, as genetic and environmental interactions may lead to reduced heritability. Eighty-six and 49 mSNPs were identified in the Yoruban and European groups respectively which accounted for 36-92%

observed differential methylation. Also comparable to the results from Bell and colleagues (2011) most of the mSNPs were not in correspondence with eSNPs.

Furthermore, inconsistencies in allele frequencies between the populations may have led to lower heritability. Upon exploring whether population or single SNP genotype were dominant contributors of methylation of each site, about 55% of the differentially methylated sites exhibited a stronger association with a *cis*-acting SNP genotype rather than with population. This was suggestive of genetic variants being responsible for majority of the observed population specificity (Fraser *et al*., 2012).

In a study also concentrated on LCLs of blood, Heyn *et al*. (2013) attempted to identify methylation differences between three human populations, and provided a link between the methylation and the possible natural phenotypic variation that occurs. Research involved LCLs and naive blood of unrelated, healthy African Americans, Caucasian Americans and Han-Chinese Americans. In total, 439 CpG sites were found to display differential methylation between the three populations. The study associated this differential methylation with defining characteristics such as response to various stimuli, disease frequencies, sensory perception and appearances. Analysis of differential DNA methylation was incorporated into genome-wide association studies; and resulted in a direct relationship between differential methylation of CpG sites of the *HLA-DPA1* locus. This locus is strongly correlated with chronic hepatitis B (HBV) infection. HBV infection risk alleles are usually enriched in populations such as Africans or Asians as they have higher frequency of diseases. Ten HBV-infection SNPs displayed high correlation with 17 CpG sites in *HLA-DPA1*. Results revealed that the risk alleles were linked to variation in DNA methylation, showing high incidence in the African and Asian populations. Risk alleles were related to the high degree of methylation of the *HLA-DPA1* promoter and hypomethylation of the gene body and this is associated with low levels of gene expression (Heyn *et al.,* 2013; Jones, 2012). Gene repression of *HLA-DPA1* was associated with DNA methylation in these two populations by identifying the risk alleles that mediate this phenomenon which spurs variation in presentation of cell surface receptors, altered binding of HBV and risk of infection (Heyn *et al.,* 2013). This warrants further research into these particular CpG sites; it may be possible that they appear specifically in other forensically relevant fluids and tissues.

It is clear that by only analysing LCLs, no final conclusion could be drawn on the effect of methylation upon expression between the European and Yoruban populations,

demanding further investigation. Perhaps a panel of methylation-specific markers may be chosen with simultaneous analysis of other body fluids in combination with the LCLs. This may provide more substantial results.

Although using differential methylation to distinguish between certain populations seems quite promising; it requires much consideration. Due to legal and ethical issues of such applications, using these ancestry informative markers requires much consideration. Predominantly, the possibility for misapplication as an ancestry informative marking method remains an extensive concern as discussed by Koops and Schellekens (2008) and M'charek *et al*. (2012). Such reports emphasize that phenotyping should be allowed for externally perceptible traits, such as hair colour and non-sensitive behavioural traits, like left-handedness or a preference for smoking. It should not be allowed for susceptibilities for diseases and for other sensitive information like an inclination toward homosexuality or aggressiveness. The intermediate category of not too sensitive traits may be certified, for example, for early apparent medical disorders, like albinism or teenage-onset alcoholism. Ethnic origin and surname phenotyping are also attuned with fundamental rights, provided measures are taken to eliminate the risk of discrimination. Also, informing the suspect of forensic phenotyping would be necessary, but it should be left to the suspect to request the actual phenotyping results.

1.9 The Current Status of Forensic Fluid and Tissue Identification

Forensic analysis of biological fluids is an indispensable tool in the identification of suspected perpetrators and the exoneration of innocent individuals in criminal investigations. However often in forensic cases, establishing the origin of the body fluid found at a crime scene is required. In these specific situations, the presence of a suspect at a crime scene is not in dispute. Instead, it is the finding of a specific body fluid such as blood, saliva, or semen that can be used to determine that a violent act has taken place. The discovery of such body fluids may significantly affect the outcome of a case (Madi *et al.,* 2012). Often merely a minute quantity of biological fluid is enough to break a case and ability to identify evidence at a crime scene in a non-destructive manner is imperative in order to preserve the sample and DNA evidence for further use.

In particular, body fluids such as blood, saliva, semen and vaginal fluids are often found at crime scenes; however others such as sweat and urine also serve as informative sources of evidence (Nakazono *et al.,* 2008; Quinones and Daniels, 2011; Virkler and

Lednev, 2009; Wickenheiser, 2011). Each of these fluids contains DNA and thus it is imperative that great precautionary measures are implemented when handling such crucial aspects of investigations. The detection and identification of body fluids in a criminal investigation are important in forensic science. Determining whether a body fluid is present and later identifying it allows further testing of protein, RNA and DNA components (Pandeshwar and Das, 2014; Virkler and Lednev, 2009). Since the technique that is used dictates the levels of purity and integrity for further analysis, a final elaboration is provided that will explain the difficulties that one may encounter by employing conventional methods to differentiate between the four most forensically relevant human body fluids.

1.9.1 Conventional Methods Employed to Identify Forensically Relevant Body Fluids

Several presumptive and confirmatory tests have been developed for differentiating between saliva, blood, semen and vaginal fluid (Virkler and Lednev, 2009). Presumptive tests are used as screening tests but tend to have specificity limitations. Confirmatory tests are used for absolute tissue identification and can be useful in reconstructing the events of a crime. Confirmatory tests based on proteins and mRNAs are associated with lack of stability issues (An *et al.,* 2012; Haas *et al.,* 2009; Zubakov *et al.,* 2008). Existing presumptive and confirmatory tests are applicable to only specific biological fluids. The investigator would be required to make a choice as to which test to perform, based on which fluid is most likely to be identified; necessitating a precise method in which the fluid may be identified irrespective of the nature (Virkler and Lednev, 2009).

1.9.1.1 Saliva

A popular presumptive test used for saliva is based on amylase, of which two different forms may be found in the human body. Amylase found in saliva, breast milk and perspiration is coded by the *AMY1* locus on chromosome one, while amylase found in the pancreas, semen, and vaginal secretions is coded by the *AMY2* locus. Although *AMY1* is found more in saliva than any other fluid, it can still only give presumptive information since it is not exclusive to saliva (Greenfield and Sloan, 2003; Quarino *et al.,* 2005). The starch–iodine test may also be employed however; competing proteins such as albumin and gamma-globulin in blood and semen will also break down iodine to form a false positive result (Virkler and Lednev, 2009).

1.9.1.2 Blood

Methods frequently employed for detection of blood primarily rely on the peroxidase-like activity of the haeme unit of haemoglobin in blood. Tests for blood include the Kastle-Meyer phenolphthalein test however potential interferences and only probable identification represent limitations; and the chemi-luminescence of luminol which has once been reported as mutagenic and thus is not recommended due to uncertainty of safety, as well as being limited to use only in dark environments (Bevel and Gardner, 2002; Virkler and Lednev, 2009).

1.9.1.3 Semen

Semen is one of the most prevalent body fluids found during criminal investigations, and in sexual assault cases its conclusive identification is usually required to corroborate the alleged crime (Romero-Montoya *et al.,* 2011). A common technique is microscopic identification of sperm, which may be specific but it impossible to identify sperm cells in cases where the evidence is in low amount or quality, due to degradation. Specificity is also difficult in bacterially-contaminated samples. Microscopy is expensive, requiring skilled personnel, time-consuming and is not amenable to automation. Detecting specific proteins found in semen and sperm may also be attempted using the prostate-specific antigen (PSA) test but high sensitivity comes with a risk of false positives. In addition, protein-based assays might suffer from low sensitivity in relation to DNA-based assays because proteins are generally not as stable as DNA under most environmental conditions. Hence a negative PSA result might be due to degradation of the protein, and not necessarily indicate absence of semen (Virkler and Lednev, 2009; Wasserstrom *et al.,* 2013). Alternative approaches developed include mRNA profiling and Raman spectroscopy, which are applicable to other fluids as well however limitations include relative instability of mRNA and lack of specificity for spectroscopy-based methods (Danaher *et al.,* 2015; Haas *et al.,* 2009; Juusola and Ballantyne, 2007).

1.9.1.4 Vaginal Fluid

Although not as common at crime scenes as blood, semen, and saliva; vaginal fluid evidence can play an important role in sexual assault cases. However, there are not many tests available to test for the presence of this fluid mainly because it is not very well defined. The constituents can change based on the menstrual cycle of the female, and this makes testing for specific components very difficult (Virkler and Lednev, 2009). One test is based on the detection of glycogenated epithelial cells using a Periodic Acid-Schiff (PAS).

However, since glycogenation varies based on the menstrual cycle; this test is not very reliable. Also, some females will likely show no glycogenated cells if they are pre-pubescent or postmenopausal, so this technique can easily have false negative results. False positive results can emerge from the mouth or urethral tract in males. Moreover, the test uses a large amount of sample and will destroy valuable DNA evidence (Greenfield and Sloan, 2003). Since citrate and lactate are found in both semen and vaginal fluids; comparisons of concentrations of these acids have been made in the past to differentiate between the two fluids. However, these carboxylic acids are present, albeit in small amounts, in saliva hence limiting use of this relatively simple method (Martin and Cheshire, 1986).

1.9.2 The Use of Biological Markers

In addition to difficulties encountered with use of traditional methods, recently developed methods also pose various limitations. Whilst great efforts are being made to preserve biological fluids for further analysis, there are also issues associated with the biological markers that are tested. These are further elaborated.

1.9.2.1 Protein-based Biomarkers

Protein-based biomarkers have mainly been employed for the diagnosis and treatments of diseases. The detection of medically relevant proteins has long been used as a biomarker in the field of diagnosis, for example detection of cardiac troponins serve as inference to cardiac disease and carbohydrate deficient transferrine indicates alcoholism (Forch *et al*., 2009). However in forensic contexts, most protein-based techniques are not amenable as they are based on presumptive or confirmatory techniques. Such tests depend on the stability of target molecules for the tissues of interest, lead to consumption of sample, degradation and irreproducibility in laboratories. Protein biomarkers are unstable and have been found to undergo rapid decomposition during downstream processing such as autolysis and putrefaction. This results in concentrations of proteins being below the detection-limits, leading to false results (Mondini *et al.,* 2009; Park *et al.,* 2014a). Protein-based assays rely on intact and functional proteins which will not work on degraded and environmentally-damaged samples. Especially in sexual assault cases, proteins present many drawbacks. To date even with many attempts, proteins that are unique to vaginal epithelial cells and buccal epithelial cells have not been validated. Whilst semen contains polymorphic protein markers, they may not be used in forensic analysis. They are recovered in too small amounts and are generally of low quality, making them unreliable. Discrimination power of protein systems

are low, so individualization is not promising (Budowle and Daal, 2008; Rifai *et al.,* 2006). Cross-reactivity with other tissues and biomolecules also decrease specificity. Furthermore, certain proteins require different methods of detection. This means too many chemicals must be purchased and various techniques have to be employed; hence eliminating the attention of protein markers in forensic laboratories today (Madi *et al*., 2012; Frumkin *et al*., 2011).

1.9.2.2 RNA-based Biomarkers

Gomes *et al*. (2013) developed a pentaplex system to identify saliva and skin by analysis of two saliva genes *STATH* and *HTN3* (Statherin and Histatin III, respectively) and three skin genes *LOR, LCE1C* and *CDS* (Loricrin, Late Cornified Envelope 1C and Diacylglycerol Synthase (Phosphatidate Cytidyltransferase) 1, respectively). The group reported *LCE1C* as most sensitive. Multiplexing in RNA techniques (Figure 1.7) is limited due to dependence on the number of fluorophores (Sijen, 2014). Interpretation of qPCR data requires normalization with reference genes. This presents a limitation since reference genes differ in expression levels between cell types. Also, most house-keeping genes used as controls for normalization and gene expression analysis such as *GADPH* (Glyceraldehyde-3-Phosphate Dehydrogenase) exist in the human genome as processed pseudogenes (Juusola and Ballantyne, 2007).

Several markers for vaginal fluids, such as membrane-associated mucins have been rendered unreliable due to expression in salivary glands (Liu *et al*., 2002; Nussbaumer *et al*., 2006). Nussbaumer and colleagues (2006) found specificity of *KLK* (Kallikrein) for semen however *HBA* (Haemoglobin Alpha) was detected in blood but also showed a weak reaction in saliva; and *MUC4* (Mucin 4) was not specific for vaginal fluid or saliva. Useful mRNA markers have been identified for skin (Gomes *et al*., 2013; Visser *et al*., 2011), saliva, blood, semen (Haas *et al*., 2014; Zubakov *et al*., 2011), sweat and urine (Xu *et al*., 2014). Since a marker employed for saliva is also present in nasal mucosa, only a partially specific marker for nasal secretion and nasal blood was found by Sakurada *et al*. (2012). However both the specificity and sensitivity vary for all of these candidates. Even though approaches of mRNA detection and analysis are constantly being developed, mRNA itself is not stable, being vulnerable to variation of temperature and humidity (An *et al*., 2012). Moreover, tissue-specific expression of mRNA in humans is not absolute and mRNAs exhibit low levels of expression in numerous tissues (Sijen, 2014; Vennemann and Koppelkamm, 2010).

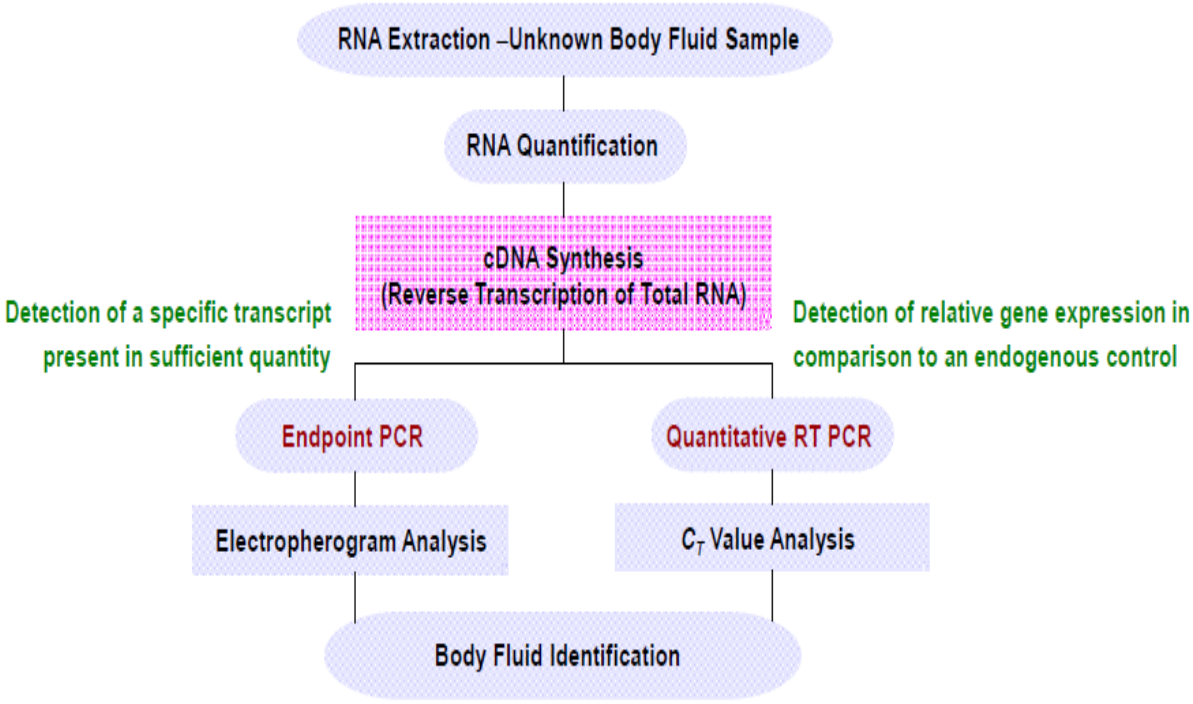

Figure 1.7: Commonly employed techniques for body fluid identification using RNA. Procedures entail end-point PCR which requires capillary electrophoresis for final analysis, or RT-qPCR in which the C*T* value (cycle threshold) is inversely proportional to the amount of RNA in the tested sample.

Hanson and colleagues (2009) introduced miRNA expression analysis into forensics by studying over 400 miRNAs of saliva, blood, semen, vaginal fluid and menstrual blood by RT-qPCR. Nine miRNAs, namely MiR-412, MiR-372, MiR-205, MiR-451, MiR-135b, MiR-10b, MiR-124a, MiR-16 and MiR-658 exhibited differential expression in the fluids. Later, Zubakov *et al*. (2010) performed detailed profiling of 718 miRNAs in these body fluids by microarrays. Fourteen candidates were identified for potential forensic application however when a validation by RT-qPCR was performed, only blood and semen were comparable, but still not the same. Inconclusive results were found for saliva, vaginal fluid and menstrual blood. Whilst studies exemplify use of miRNA expression to differentiate body fluids, it is evident that inconsistent patterns are found when employing different technologies. Additional needs to enhance specificity are necessary in miRNA assays since universal primers greatly reduce specificity (Baker, 2010). Similar to mRNA assays; since RT-qPCR is employed in miRNA detection and quantification, appropriate reference targets are necessary for normalization of miRNA expression. These would include stable and abundant expressed

small ncRNAs such as snRNAs (small nuclear RNAs that mediate pre-mRNA processing) or snoRNAs (small nucleolar RNAs that mediate chemical modification of other RNAs) (Sauer *et al*., 2014; Wang *et al*., 2013). There is no unanimity in appropriate small RNA genes for this purpose. Whilst MiR-16 and snoRNAs SNORD43 and RNU62 are used; MiR-16 is over-expressed in erythrocytes and haemolysis greatly influences levels in blood. RNU62 is unreliable as it synthesised by polymerases different from those that synthesise precursor miRNAs (Baker, 2010; Ma *et al*., 2013b). Not unlike mRNAs, miRNA expression is not fully cell-type specific due to background transcription, hence compromising use in forensic applications (Sijen, 2014).

1.9.2.3 Microbial Biomarkers

Human bodies do not escape the ubiquitous presence of microorganisms. Microbes including viruses, fungi, bacteria, protozoa and parasites aid in food digestion and human health. They are located outside of, and within the body including hair, skin, the vagina, intestines and mouth. Variations of the human microbiome are attributed to illnesses such as obesity, periodontitis, vaginosis and inflammatory bowel disease (Jorth *et al*., 2014). Studies have shown that specific microbes may be found in particular human body sites and fluids. This is now being exploited profusely for indications of body sites, fluids and tissues such as saliva, skin, nasal area, stool and vaginal fluids (Benschop *et al*., 2012; Huse *et al*., 2012). The specificity of microbial biomarkers is found to be best for vaginal fluids and menstrual secretions (Benschop *et al*., 2012; Choi *et al.,* 2003; Fleming and Harbison, 2010; Giampaoli *et al*., 2012). Akutsu and colleagues (2012) reported PCR detection of 16S-rRNA genes of *Lactobacillus jensenii* and *L. crispatus* in not only vaginal fluid but also female urine samples. The presence of *L. gasseri* and *Gardnerella vaginalis* was also found in semen. *L. gasseri* has proven to be unreliable marker, being detected in vaginal fluid (Fleming and Harbison, 2010) as well as semen (Akutsu *et al*., 2012) and furthermore, the use of bacterial RNA is a drawback. Primers may target not only the sequence of interest but also any species-specific region. In RNA-based assays, sequences must first be transcribed (Sijen, 2014). Giampaoli and colleagues (2012) described multiplex RT-PCR detection of microflora DNA obtained from vaginal and faecal samples. Enterococcus spp. was most prevalent in faecal samples. Although strong signals for *L. crispatus* and *L. gasseri* were detected in vaginal fluid, signals for *L. gasseri* were also detected for faecal samples.

Lactobacilli demonstrate potential as a vaginal fluid marker however, Pavlova and Tao (2000) demonstrated that benzo-(a)-pyrene-diol epoxide from smoking disrupts naturally occurring *Lactobacillus* spp. Antibiotics such as erythromycin causes changes in the vaginal microbial communities (Choi *et al*., 2003; Witkin *et al*., 2007). Deliberation exists with the strategy due to intra-individual variability; bacterial fingerprints may differ from hands of a single individual, leading to confusion (Kuczynski *et al*., 2010). Microbes are easily transferred, by touch or contact of another human or surfaces. Vaginal microflora can transfer to the male penis or groin (Benschop *et al*., 2012; Flores *et al*., 2011). Complexity arises in use of microbial markers since the microbiota at a body site does not necessarily relate to microbiota from a crime stain. Compositions may vary due to degradation, contamination and growth of many airborne species upon release from the human body (Sijen, 2014). Also, Dewhirst *et al*. (2011) showed that approximately 16% of oral microbiota is shared between canines and humans; including *Mogibacterium timidum, M. diversum, Treponoma amylovorum, T. vincenti,* three streptococcal species and numerous others. Such findings greatly compromise human specificity.

1.10 Scope of the Present Study

The most prevalent methods to analyse body fluids found at crime scenes are limited by lack of specificity and sensitivity, consumption of valuable biological material and are largely presumptive. However the precise identification of body fluids found at crime scenes provides vital information that can support a link between sample donors and actual criminal acts. Reports of whole genome epigenetic analysis indicate that chromosome segments called tissue-specific differentially methylated regions (tDMRs) show varying DNA methylation profiles according to the type of cell or tissue. Thus, body fluid-specific differential DNA methylation is a promising indicator for precise body fluid identification. Even though forensic tDMR-based body fluid identification shows much potential, not many tDMRs have been confirmed for this purpose. Studies have confirmed tDMRs for semen identification however specific tDMRs for blood, saliva and vaginal fluid have yet to be validated. The development of new tDMRs is the best way forward for accurate identification of biofluids for forensic applications. Accordingly, novel tDMRs were searched for based on the analysis of differential gene expression in human tissues and its correlation with DNA methylation. The tDMRs were targeted to develop primers for use in Methylation-Specific Restriction Enzyme (MSRE-PCR) for efficient identification of saliva, blood, semen and vaginal fluid. The identification of additional body fluid tDMRs is expected to spur the improvement of promising body fluid identification methods for forensic applications.

Additionally, researchers have provided resounding evidence of differential DNA methylation between population groups. Therefore, it is worthwhile to validate previously documented tDMR-based markers on the South African population. For this, methylation profiles of four previously reported tDMRs in saliva obtained from Blacks, Indians, Whites and Coloureds in South Africa were evaluated, to determine if there is differential methylation of targeted tDMRs among ethnic groups. Significant differences could assist forensic analysts in future, to narrow down their search for sample donors.

1.10.1 Hypotheses tested

It is hypothesized that the identification of new tissue-specific differentially methylated regions (tDMRs) in human body fluids would allow the development of markers to differentiate between body fluids.

It is further hypothesized that DNA methylation profiles of saliva, using selected tissue-specific differentially methylated regions (tDMRs) for *USP49, DACT1, L81528* and *PFN3* genes differ among the diverse South African population and thus may be of significance in forensic casework.

1.10.2 Aims

a) To identify novel tissue-specific differentially methylated regions (tDMRs) which could be used as specific markers for body fluid identification and thus aid in the reconstruction of crime scenes.

b) To infer variation between four ethnic groups of South Africa based on differential DNA methylation profiles of four previously reported tDMRs in saliva.

1.10.3 Objectives

The following objectives were established to achieve the stated aims:

a) To mine gene expression and DNA methylation databases for identification of differentially methylated regions in forensically significant body fluids: saliva, blood, semen and vaginal fluid.

b) To design PCR primers targeting tissue-specific differentially methylated regions.

c) To test the selected methylation-targeted primers by MSRE-PCR, on DNA derived from human body fluids (saliva, blood, semen, vaginal fluid).

d) To study the methylation profile of saliva within the diverse South African population by MSRE-PCR.

1.10.4 Experimental Design

In order to achieve the stated objectives, the research was divided into the relevant chapters described below. A precise description of each phase will be described in the chapters which will include all rationale, challenges experienced and deviations from reported methods.

- **Chapter Two**

This chapter focuses on collection of DNA methylation information of CpG islands (CGIs) of tissue-specific differentially expressed genes, followed by the design of PCR primers that target the differentially methylated CGIs. Bioinformatics data acquisition and primer design was followed by testing selected methylation-targeted primers on DNA derived from human body fluids (saliva, blood, semen and vaginal fluid) by Methylation-Sensitive Restriction Enzyme-PCR (MSRE-PCR) to identify specific primers that allow the differentiation of the body fluids.

- **Chapter Three**

This chapter investigated the methylation profile of saliva within various ethnic groups of the South African population. Previously reported tDMR-based markers for identification of saliva, were used to screen the diverse South African population.

- **Chapter Four**

The final chapter, comprising the general discussion and conclusion, places the present research study in perspective, providing an overview of the main objectives and findings reported in each of the chapters of the dissertation. Possible limitations or shortcomings are acknowledged and a scope for future improvement and development is provided.

CHAPTER TWO

IDENTIFICATION OF TISSUE SPECIFIC DIFFERENTIAL DNA METHYLATION IN HUMAN BODY FLUIDS

ABSTRACT

DNA methylation is an epigenetic mark of paramount importance as it plays a vital role in development and differentiation by control of gene expression via alterations in chromatin structure. Studies of whole-genome methylation have reported that chromosome segments known as tissue-specific differentially methylated regions (tDMRs) present varying methylation profiles according to tissue or cell type which also induce tissue-specific variation in gene expression and regulation. Therefore, tDMRs can be targeted to develop markers for body fluid identification, which would be especially useful in forensic casework where the identification of biological materials is in dispute. The present study aimed to develop new DNA methylation-specific markers using a bioinformatics approach based on the analysis of differential gene expression in human tissues and its correlation with DNA methylation. The objective was to target the identified tDMRs to design PCR primers for efficient identification of body fluids.

Tissue-specific gene expression is mostly associated with hypomethylation of gene-based CpG islands (CGIs), and low or no expression level is associated with hypermethylation. Thus, to identify new tDMRs, a total of 1833 genes which were over-expressed in only a single body fluid, were identified from surrogate tissues of four body fluids, namely; saliva, blood, semen and vaginal fluid. CpG methylation information of normal tissues (non-target tissues) was mapped to genes and heavily methylated CGIs were targeted for primer design. Sixty-three CGI sequences were selected and analysed for specificity in the human genome, and a total of four CGIs were targeted to design PCR primers. Methylation profiles of the tDMR-based primers in saliva, blood, semen and vaginal fluid was determined by Methylation-Sensitive Restriction Enzyme PCR (MSRE–PCR). Two potential body fluid-specific tDMRs were identified: a tDMR of the *HPCAL1* gene was identified as a potential blood-specific hypomethylation marker, and a tDMR of the *PTPRS* gene was identified as a potential vaginal fluid-specific hypermethylation marker.

Currently, only a few tDMR-based markers have been reported to identify body fluids, most of which require validation. To enhance the specificity and robustness of DNA methylation-based identification, identification of novel markers is required. The described method of identification and analysis of tDMRs represents an efficient and reliable technique to identify biological fluids and tissues found at crime scenes, and the two novel body fluid specific tDMRs will be of great assistance in accurate identification and differentiation between forensically relevant biological fluids.

2.1 Introduction

DNA methylation is a significant epigenetic alteration that plays a key role in gene expression and regulation (Ghosh *et al.,* 2010; Kurukuti *et al.,* 2006; Varley *et al.,* 2013). Changes in DNA methylation are most frequently associated with cytosines at CpG dinucleotides. However, recent research has described non-CpG methylation in human embryonic stem cells and in differentiated mammalian cell types including human skeletal muscle and brain (Pinney, 2014; Guo *et al.,* 2014b). Although CpG sites are mainly methylated across the mammalian genome (Gutierrez-Arcelus *et al.,* 2013), there are distinct, mostly unmethylated CG-rich regions called CpG islands (CGIs), which have a G+C content greater than 50%. CGI methylation in promoters and first exon regions are strongly associated with reduced gene expression (Bird, 2002; Ginno *et al.,* 2012; Lienert *et al*., 2011). However, positive correlations between intragenic DNA methylation and gene expression have been reported (Irizarry *et al.,* 2009; Lister *et al.,* 2009; VanderKraats *et al.,* 2013). CGI methylation is also important for tissue-specific gene regulation. Certain tissue-specific genes or regions of a gene are methylated in the tissues in which they are not expressed, but unmethylated in tissues where they are expressed (De-Smet *et al.,* 1999; Weber *et al.,* 2007). These tissue-specific differentially methylated regions (tDMRs) are currently seen as potential biomarkers particularly in the field of cancer research for disease detection, progression and therapeutic intervention (Kulis *et al.,* 2012; Mikeska and Craig, 2014; Rawson and Bapat, 2012; Yoo and Jones, 2006).

The amplifiable and stable nature of methylated DNA enables simple transfer of a methylation-based diagnostic test from laboratory settings to diagnostic situations. tDMR-based biomarkers are sensitive, specific, readily detectable, and unaffected by varying methods of collection and analysis (Mikeska and Craig, 2014). Since methylation profiles underpin the regulatory landscape for externally-visible cellular expression programs; scientists have shifted focus from unstable and unreliable protein and RNA biomarkers to differentially methylated regions (Kit *et al.,* 2012; Levenson, 2010). Indeed the risk of heterogeneity, along with the necessity for preservation of tissues, and need for normalization of derived data has limited the use of other biomolecules such as mRNA and miRNA as biomarkers (Tost, 2010).

Scientists are striving to develop methods to identify and characterise CGIs and methylation patterns that associate with genes and phenotypes. For this, computational

programs based solely on DNA methylation have been developed. Recent research has focused on the identification of differentially methylated regions (DMRs) predominantly in the context of diseased and healthy tissues between populations. Early studies for tDMR identification used the Students *t*-test or the Wilcoxin Rank Sum Test (WRST) with normalization between population groups (Dyson *et al.,* 2014; Wang *et al.,* 2012). Expression analysis was performed by the univariate test (Hsiao *et al.,* 2014). Also, the Analysis of Variance model (ANOVA) has been developed for detection of aberrant methylation in diseased samples (Dyson *et al.,* 2014); as well as 'sliding window' technique for DMR analysis (Hackenberg *et al*., 2010; Jones and Takai, 2003; Ponger and Mouchiroud, 2002). Wang and Leung (2004) developed a Java-based application known as CpGIE to identify CGIs made up of ~200 bp, a G+C content of 55%, CpG observed/expected (*o/*e) ratio of 0.65, as specified by Takai and Jones (2002).

Bioinformatics-based methods have also been created to identify DMRs. Slieker and colleagues (2013) identified 3533 tDMRs in peripheral tissues including hair follicles, blood, buccal swabs and saliva; and 5382 tDMRs in internal tissues comprising of liver, pancreas, muscle, fat, spleen and omentum by use of the R package *IlluminaHumanMethylation450k.db.* Upon searching for gene expression data from the TiGER (Tissue-specific Gene Expression and Regulation) database, they discovered that tDMRs mapping to genes with tissue-specific expression were hypomethylated in the tissue in which the gene was over-expressed and hypermethylated in tissues showing no expression, indicative of a negative relationship between DNA methylation and gene expression. When ENCODE data was mapped onto transcription factor binding sites, enrichment of tDMRs was observed. *FOXA2* transcription factor was active in pancreas, liver and hair follicles and was hypomethylated in these tissues. The tDMRs were also enriched in alternative transcription sites. Nearly 50% of both the peripheral tissue and the internal organ datasets mapped to an alternative transcription event (Cheong *et al.,* 2006; Kulis *et al.,* 2013; Slieker *et al.,* 2013).

Software programs mainly employed for analysis and visualization of Illumina array data have also been developed to identify new DMRs (Aryee *et al.,* 2014; Chen *et al.,* 2012; Wang *et al.,* 2012). These include Minfi (Aryee *et al.,* 2014), DMRforPAIRS (Rijlaarsdam *et al.,* 2014), QDMR (Zhang *et al.,* 2011b), BSmooth (Hansen *et al.,* 2012), DMRcate (Peters *et al.,* 2015) and methyAnalysis (Du and Bourgon, 2015). The methods are useful to explore methylation mechanisms further but are restricted to Infinium arrays which are time-

consuming and expensive. Issues exist with normalization, development and understanding of intricate algorithms (Frommer *et al.,* 1992; Levenson, 2010).

Screening of novel tDMR-based biomarkers requires a dependable and accurate method of identification and analysis. Most approaches are based on bisulfite conversion followed by Methylation-Specific PCR, which is sensitive and cost-effective but highly stringent assays are necessary to avoid false positives. Bisulfite treatment provides detailed methylation profiling of abundant and homogenous samples. However, in a medical or forensic situation the DNA is most often heterogeneous and rarely available in high amounts or purity. Epigenotyping arrays such as Infinium enables single-base resolution and high coverage yet comes with the drawbacks of cross-reactivity, expense and vulnerability to genetic variations (Dedeurwaerder *et al*., 2013; Kit *et al.,* 2012). Pyrosequencing, as employed by Madi and colleagues (2012) offers quantitative methylation status of CpGs, but is compromised by mediocre sensitivity, high cost and high error rate (Tost and Gut, 2007). Whole-genome bisulfite sequencing (WGBS) is limited by incomplete bisulfite conversion and DNA degradation. Moreover, hydroxymethyl-cytosine is indistinguishable from methyl-cytosine, leading to false positives. Thus, the technique is neither informative nor reliable (Bock, 2012; Levenson, 2010; Munson *et al.,* 2007). Methods based on pre-treatment with methyl-cytosine specific proteins and antibodies such as MeDIP-seq, MethylCap Sequencing and MBD-isolated Genome Sequencing (MIGS) present limits of low resolution, elevated levels of background signals, and degrees of CpG and DMR methylation levels that bias the techniques (Bock *et al.,* 2010; Down *et al.,* 2008; Harris *et al.,* 2010).

To evade bisulfite conversion and intricate algorithms, researchers have turned to the robust, reliable method of Methylation Sensitive Restriction Enzymes (MSRE) (Choi *et al.,* 2014; Melnikov *et al.,* 2005). MSRE-based treatments are cost-effective, relatively simple and easy to perform. Most analysis methods do not require special instrumentation (Kit *et al.,* 2012; Melnikov *et al.,* 2005). MSRE treatment followed by PCR and electrophoresis presents advantageous features over bisulfite pre-treatment methods; however has not been used to a large extent in gene expression-based research. The method enables numerous regions to be analysed in a single reaction by multiplexing, whilst simultaneously reducing the amount of template DNA required (Choi *et al.,* 2014).

DNA methylation and tDMR-based markers demonstrate paramount potential in forensic casework. Relating methylation profiles of tDMRs and their corresponding

phenotypic effects on various body fluids collected at crime scenes would be extremely useful in discriminating between unknown body fluids. Accurate identification of biological material is crucial, as results could either make or break a case. However, to date, not many tDMRs have been confirmed for this purpose. Whilst tDMR-based markers, targeting CpGs of genes such as *L81528* (Choi *et al.,* 2014), *FGF7* and *ZC3H12D* (Madi *et al.,* 2012) have been developed for semen identification, validated markers do not exist for other forensically significant body fluids such as epithelia, saliva, blood and vaginal fluid. Accurate differentiation between saliva and blood has presented difficulty for numerous researchers (An *et al.,* 2013; Frumkin *et al.,* 2011; Lee *et al.,* 2012; Liu *et al.,* 2010). Frumkin *et al.* (2011) tested methylation ratios of various loci to differentiate body fluids and detected similar methylation patterns for blood and saliva in *L76138/L26688*. Similarly An and colleagues (2013) found hypermethylation for both body fluids using the *PFN3* gene, and Lee *et al.* (2012) experienced difficulty upon employing the HOXA4 tDMR.

Thus far, inconclusive results have been obtained for other body fluids. The BCAS4 marker was found to be hypermethylated in semen by Eckhardt *et al.* (2006) however Madi *et al.* (2012) found saliva to be similarly hypermethylated. Madi and colleagues (2012) also declared C20orf117 inconclusive in differentiating between blood and epithelial skin cells. More recent research by Choi *et al.* (2014) has only been able to classify samples as blood-saliva or semen-vaginal fluid due to similar methylation levels. The group proposed that PFN3 may be able to identify vaginal fluid however the marker has yet to be validated. Additionally, Park and colleagues (2014a) experienced difficulty in finding stable markers to differentiate saliva and vaginal fluid.

Being a relatively novel tool, DNA methylation-based cell-type and body fluid inference still requires much work. DNA collected from crime scenes are typically in low concentrations and amounts and there is no room for error, false positives or wastage of valuable evidential, genetic material (Gomes *et al.,* 2011; Sijen, 2014; Virkler and Lednev, 2009). Body fluid identification by tDMR analysis targets DNA, thus there is no additional consumption of evidence. Multiplex systems for PCR analysis uses standard capillary electrophoresis platforms, which are compatible with STR-typing techniques. The identification and development of new tDMRs will enable accurate identification of biofluids for forensic applications (Choi *et al.,* 2014; Frumkin *et al.,* 2011; Park *et al.,* 2014a).

The present study centres mainly on the identification of tDMRs in four forensically significant body fluids, namely; saliva, blood, semen and vaginal fluid. The objective was to target the identified tDMRs to design PCR primers for efficient identification of body fluids by MSRE-PCR. The approach followed for the identification of differential DNA methylation is based on bioinformatics analysis of differential gene expression in human tissues and its correlation with DNA methylation.

2.2 Materials and Methods

2.2.1 Data Acquisition

Bioinformatics data collection and analysis was performed in correspondence with Dr Dyfed Lloyd Evans. Gene expression data of normal surrogate tissues of four human body fluids, namely saliva (surrogate tissues: salivary glands and tongue), blood, semen (surrogate tissue: prostate) and vaginal fluid (surrogate tissues: uterus and cervix) were mined.

Gene Expression Data were obtained from the following databases:

- Tissue-specific Gene Expression and Regulation (TiGER) expression database (http://bioinfo.wilmer.jhu.edu/tiger/).

TiGER (Tissue-specific Gene Expression and Regulation) is a database for generating comprehensive information about human tissue-specific gene regulation, including both expression and regulatory data. The database contains tissue-specific expression profiles for UniGene genes, combinatorial regulation for interacting transcription factor (TF) pairs, and *cis*-regulatory modules (CRMs) for tissue-specific genes (Liu *et al.,* 2008).

Identification of differentially expressed genes: Any gene with a minimum of 10-fold enrichment in any single surrogate tissue of interest as compared to all other tissues was deemed to be up-regulated in that particular tissue.

- Human BodyMap 2.0 data was downloaded from EnsEMBL (http://www.ensembl.org).

EnsEMBL is a database comprising of numerous species genomes including humans, other vertebrates as well as model organisms. Specifically, RNASeq data from Illumina's Human BodyMap project was obtained from EnsEMBL. The data entails 16 human tissue types including kidney, adrenal, adipose, brain, liver, heart, lymph, breast, skeletal muscle, lung, testes, prostate, colon, ovary, white blood cells and thyroid (Cunningham *et al.,* 2015).

Identification of differentially expressed genes: Differentially expressed genes were determined by using the DEGseq R package: (https://www.bioconductor.org/packages/release/bioc/html/DEGseq.html) (Wang and Wang, 2015).

- GXA (Gene Expression Atlas) (http://www.ebi.ac.uk/gxa/home).

Comprising both microarray and RNASeq data, the GXA provides information on gene expression patterns under various biological conditions. The Expression Atlas provides

information on gene expression patterns under different biological conditions. Gene expression data is re-analysed to detect genes showing interesting baseline and differential expression (Petryszak *et al.,* 2014).

Differentially expressed genes were determined by using DEGseq R package.

- Human Gene Expression (HuGE) dataset (http://zlab.bu.edu/HugeIndex/HugeIndex/HugenIndexCompendiumSupplement.shtml).

The HuGE Index is a comprehensive mRNA expression database that contains expression data of thousands of genes in normal human tissues and organs. The expression data was achieved from use of high-density oligonucleotide array technology (Haverty *et al.,* 2002).

- Serial Analysis of Gene Expression (SAGE) data from NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/).

The GEO is an international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomics data submitted by researchers (Barrett *et al.,* 2013).

From the above databases, gene expression data was integrated and genes specifically up-regulated or over-expressed in surrogate tissues of interest but were not expressed or marginally expressed in other tissues were selected for further study.

2.2.2 Methylation Data Acquisition

Methylation states for selected genes were obtained from the datasets stated below, together with methylation positions, and custom scripts were written to map these methylation states to the CpG islands (CGIs).

- ENCODE data was downloaded from EnsEMBL (http://www.ensembl.org).

The ENCODE (Encyclopaedia of DNA Elements) Consortium is a comprehensive list of functional elements of the human genome. These elements include those that act at RNA and protein levels, as well as regulatory elements involved in cell-control and circumstances wherein a gene is activated (ENCODE Project Consortium, 2012).

- NGSmethDB (http://www.methdb.net/). Human DNA methylation datasets with 3-pass verification were downloaded.

The NGSmethDB databases comprise information of methylated cytosines in human DNA. The database contains many types of data, including that obtained from HPLC and sequencing. Comprehensive background information is obtained from NGSmethDB

including the sample origin, phenotypes and expression of related genes (Grunau *et al.,* 2001). NGSMethDB data was based on an old version of the human genome assembly (hg18). As a result, the entire dataset was remapped onto the latest human genome assembly (GRCh38) coordinate system, enabling the integration of ENCODE and NGSMethDB data.

2.2.3 Data Integration and Exclusion of Differentially Expressed Genes Present in More than One Surrogate Tissue

Gene expression data from various sources mentioned above were integrated to generate a single dataset. Since the data sources use varying transcript and gene identifiers, these were mapped to stable EnsEMBL identifiers using both the HGNC (www.genenames.org) and EnsEMBL APIs (Application Program Interfaces).

Genenames.org is a repository of official human gene nomenclatures, gene families and associated resources including access to genomic, proteomic and phenotypic information. All of the nomenclature in the database has been approved by the HUGO Gene Nomenclature Committee (HGNC) (Gray *et al.,* 2015). Using common, stable identifiers enabled tissue specific expression data to be mapped onto the corresponding human gene set.

Based on EnsEMBL IDs, genes found to be over-expressed in more than one surrogate tissue were excluded from further study (this occurs as most expression data is based on transcript rather than gene sequences).

2.2.4 Elimination of Genes with No Evidence of Protein Synthesis

Even if an mRNA signal does increase, it is still possible for no additional protein to be expressed in a particular tissue (this is known as mRNA sequestration), therefore to confirm that the genes were functional, RNA expression profiles were compared with protein expression data derived from GeneCards. GeneCards is a compendium of genomic, transcriptomic, proteomic, genetic, clinical and functional information of both annotated and predicted human genes. It may be obtained from the Human Gene Expression database (Belinky *et al.,* 2015).

Using an exclusive link for each gene (http://www.genecards.org/cgi-bin/carddisp.pl?<gene>) known as the gene card summary, the remaining genes were checked for their protein expression. All genes that could not be confirmed as protein coding, RNA encoding genes, pseudogenes and non-coding RNAs were removed from the study.

Genes which showed high protein expression in concert with high transcript expression were selected for further evaluation.

2.2.5 Genes and Genome Data

Genome regions corresponding to over-expressed transcripts were obtained from the EnsEMBL online databases using custom scripts developed against EnsEMBL APIs (http://www.ensembl.org).

2.2.6 Mapping CpG Island and Methylation Information

Using codes developed against the EnsEMBL Core API, the genes identified above were used to query the EnsEMBL databases. Co-ordinates of CpG Islands (CGIs) associated with the genes, as well as exons and UTRs were obtained. ENCODE and NGSMethDB data were also mapped to the gene co-ordinates. Genes without association with CGIs were eliminated at this stage. The methylated CpG dinucleotides were mapped onto the genomic co-ordinates of the remaining genes. The methylation signal was averaged across the length of the CGIs on which they were located.

For each CGI, methylation data were integrated and expressed as mean, median and minimum-maximum values of the methylation states across the CGI. Genes in which CGIs presented a mean, median and minimum-maximum methylation percentage above 75% were targeted. These genes had methylation profiles above 75% (in tissues where typical gene expression was normal or low) but were specifically over-expressed in one of the surrogate tissues of the four forensically relevant body fluids.

2.2.7 Final Selection of Candidate Tissue-Specific Differentially Methylated Regions

The targeted CGIs of selected genes which were highly methylated in tissues other than saliva, blood, semen and vaginal fluid were pasted into NEBcutter V2.0, (http://nc2.neb.com/NEBcutter2/) to ensure the CGIs contained the *Hha*I restriction site (GCGC). The CGI sequences which lacked these sites were removed from the list of candidate genes. This step was to ensure that the designed primers targeting the CGIs would flank the *Hha*I recognition site during Methylation-Sensitive Restriction Enzyme Polymerase Chain Reaction (MSRE-PCR).

The sequences of these remaining CGIs that correspond to the genes of interest and had the *Hha*I restriction site were BLASTed against the human genome assembly GRCh38. This was to ensure that unique CGIs were targeted to design the primers. CpG sequences mapping to a single chromosomal location were selected, whilst those mapping to several locations on human chromosomes were discarded.

2.2.8 Primer Design for PCR

To determine the methylation status of the selected CGIs, primers flanking the *Hha*I recognition sites (GCGC) were designed using the Primer3Plus program (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi//). The unmethylated artificial DNA control was obtained by PCR amplification of the 481-bp portion of the pCR®2.1 TOPO® vector (Invitrogen, Carlsbad, CA, USA) (An *et al.,* 2013; Choi *et al*., 2014; Appendix A).

2.2.9 Sample Collection

Body fluid samples (saliva, blood, semen and vaginal fluid) for analysis of designed primers were obtained from the principal investigator and a male member of her family, according to methods specified by the Biomedical Research Ethics Committee of UKZN Westville. Saliva was collected in a 15 ml sterile tube and 200 L aliquots were stored frozen at -20 C until DNA was extracted. Blood was obtained in sterile tubes containing ethylenediamine tetraacetic acid (EDTA, an anticoagulant), and 200 L aliquots were stored frozen at -20 Cuntil DNA was extracted. Semen was obtained in sterile Eppendorf tubes and stored frozen at -20 Cuntil DNA extraction. Vaginal fluid was obtained using sterile cotton swabs and dried at room temperature for approximately two hours, followed by DNA extraction.

2.2.10 DNA Extraction and Quantification

DNA was extracted from each aliquot of blood and saliva using a Quick-g DNA MiniPrep Kit (ZymoResearch); from semen using a QIAamp® DNA Mini Kit (QIAGEN, Germany) and DNA from vaginal fluid was extracted using a QIAamp® DNA Investigator Kit (QIAGEN) according to the manufacturer's instructions. Extracted DNA was quantified using a spectrophotometer (Nanodrop ND-2000, Thermo Fisher Scientific Inc, Waltham, MA, USA).

2.2.11 Restriction Enzyme Treatment

MSRE-PCR was developed to amplify *Hha*I recognition sites (GCGC) of the selected CGIs. Prior to conducting the PCR; approximately 100 ng of saliva, blood, semen and vaginal fluid DNA was restricted separately with *Hha*I in a 10 L reaction containing 1 L of CutSmart Buffer (New England Biolabs, Ipswitch, MA, USA) and *Hha*I (New England Biolabs, Ipswitch, MA, USA) at various concentrations of 10 U, 2 U and 0.2 U. Different incubation and inactivation temperatures and times were attempted for successful restriction (37 Cfor 16 hours; 37 Cfor 1 hour; 37 Cfor 30 minutes; and heat inactivation at 80 Cfor 20 minutes as well as 65 Cfor 20 minutes). The control reaction, to ensure successful restriction, also consisted of restricting 100 ng of the unmethylated artificial DNA template using the same concentrations of *Hha*I as for body fluids. All restrictions were performed in triplicate.

2.2.12 MSRE-PCR

PCR amplification was carried out in a 20 L reaction volume containing 10 L of enzyme-digested DNA, 2.0 U of AmpliTaq Gold® DNA Polymerase (Applied Biosystems), 1 L of Gold ST*R 10× Buffer (Promega, Madison, WI, USA) and 0.3 M each primer (Table 2.8; Table 2.9). PCR cycling was conducted on a BIORAD T100™ Thermal Cycler under the following HotStart conditions: 95˚C for 11 minutes, 28 cycles of 94˚C for 20 seconds, varying annealing temperatures (Table 2.8) for 60 seconds, and 72˚C for 30 seconds and final extension at 72˚C for 60 minutes. The amplification products were analysed by agarose gel electrophoresis using 2% agarose gel, stained with 10 mg/ml ethidium bromide at 80 V for 1 hour.

The stated PCR cycling conditions known as HotStart PCR were used as AmpliTaq Gold® DNA Polymerase (Applied Biosystems) is provided in an inactive state and is only activated upon heat. Whilst typical PCR cycling reactions consist of an initial denaturation step of 94-96˚C for 2-5 minutes, the denaturation time is lengthened in a HotStart PCR.

2.3 Results

2.3.1 Tissue-Specific Gene Expression Data

A total of 1833 genes were found to be over-expressed in the surrogate tissues of saliva, blood, semen and vaginal fluid. The distribution of these genes in the different body fluids are shown in Table 2.1, indicating that majority of these genes were found to be specifically over-expressed in saliva (31.8%) and blood (28.8%). Table 2.2 shows the top 10 over-expressed genes in each of the forensically significant body fluids.

Table 2.1: Over-expressed genes in surrogate tissues of saliva, blood semen and vaginal fluid.

Body Fluid	Number of Over-expressed Genes	Percent Distribution of Over-Expressed Genes
Saliva	583	31.8%
Blood	528	28.8%
Semen	352	19.2%
Vaginal Fluid	370	20.2%
Total	1833	-

Table 2.2: The top 10 over-expressed genes in each of the forensically significant body fluids.

Body Fluid	EnsEMBL ID	Gene Name	Expressed Sequence Tag (EST)-Enrichment
Saliva	ENSG00000225930	DKFZP434L187	92.598
	ENSG00000092758	COL9A3	66.814
	ENSG00000169248	CXCL11	53.569
	ENSG00000017483	SLC38A5	49.807
	ENSG00000198464	ZNF480	45.061
	ENSG00000073282	TP73L	43.041
	ENSG00000169752	NRG4	41.595
	ENSG00000166535	A2ML1	40.992
	ENSG00000169474	SPRR1A	40.992
	ENSG00000241794	SPRR2A	39.284
Blood	ENSG00000133742	CA1	66.412
	ENSG00000262827	FCAR	63.037
	ENSG00000223609	HBD	62.519
	ENSG00000206047	DEFA1	61.841
	ENSG00000172232	AZU1	61.461

	ENSG00000159339	PADI4	61.287
	ENSG00000169385	RNASE2	59.535
	ENSG00000262383	LILRA3	56.909
	ENSG00000124469	CEACAM8	55.282
	ENSG00000101425	BPI	54.929
Semen	ENSG00000100373	UPK3A	25.523
	ENSG00000124233	SEMG1	24.110
	ENSG00000159182	PRAC	23.928
	ENSG00000124157	SEMG2	23.615
	ENSG00000263639	MSMB	23.226
	ENSG00000167751	KLK2	22.775
	ENSG00000229637	PRAC2	22.043
	ENSG00000163810	TGM4	21.665
	ENSG00000167749	KLK4	20.206
	ENSG00000167332	OR51E2	20.109
Vaginal Fluid	ENSG00000205810	KLRC3	44.514
	ENSG00000205809	KLRC2	44.514
	ENSG00000178226	PRSS36	41.334
	ENSG00000177138	FAM9B	33.067
	ENSG00000158483	FLJ10661	30.061
	ENSG00000186523	FAM86B1	30.061
	ENSG00000123388	HOXC11	30.061
	ENSG00000170577	SIX2	29.334
	ENSG00000130720	FIBCD1	28.934
	ENSG00000117148	ACTL8	27.556

2.3.2 Exclusion of Differentially Expressed Genes Present in More than One Surrogate Tissue

Comparisons of RNA and protein expression identified 179 genes that were over-expressed in more than one surrogate tissue. Together with the tissues the genes are over-expressed in, these 179 genes are shown in Table 2.3. The removal of the genes yielded 1654 remaining genes for further study.

Table 2.3: Genes over-expressed in more than a single surrogate tissue and corresponding tissues in which they were over-expressed.

EnsEMBL ID	Body Fluid/s	EnsEMBL ID	Body Fluid/s
ENSG00000011600	Blood, saliva	ENSG00000168310	Blood, saliva
ENSG00000015133	Blood, vaginal fluid	ENSG00000008441	Semen, lachrymal
ENSG00000057149	Saliva, vaginal fluid	ENSG00000108405	Blood, semen
ENSG00000057704	Blood, saliva	ENSG00000168769	Saliva, semen
ENSG00000079482	Saliva, semen	ENSG00000026508	Vaginal fluid, lachrymal
ENSG00000081277	Saliva, semen	ENSG00000062038	Saliva, vaginal fluid
ENSG00000088002	Saliva, vaginal fluid	ENSG00000063169	Blood, semen
ENSG00000095303	Saliva, semen	ENSG00000079385	Blood, saliva
ENSG00000096696	Saliva, vaginal fluid	ENSG00000090382	Blood, lachrymal
ENSG00000096006	Blood, saliva	ENSG00000099812	Vaginal fluid, semen
ENSG00000100448	Semen, vaginal fluid	ENSG00000114626	Blood, lachrymal
ENSG00000102034	Blood, saliva	ENSG00000116741	Blood, semen
ENSG00000102554	Saliva, vaginal fluid	ENSG00000124102	Saliva, vaginal fluid
ENSG00000117691	Saliva, semen	ENSG00000133195	Blood, vaginal fluid
ENSG00000121316	Blood, saliva	ENSG00000135413	Lachrymal, saliva
ENSG00000121552	Semen, saliva	ENSG00000136522	Saliva, lachrymal
ENSG00000122133	Semen, vaginal fluid	ENSG00000142197	Blood, saliva
ENSG00000128422	Saliva, vaginal fluid	ENSG00000143761	Saliva, lachrymal
ENSG00000128709	Semen, vaginal fluid	ENSG00000145879	Saliva, semen
ENSG00000130720	Semen, vaginal fluid	ENSG00000147873	Saliva, vaginal fluid
ENSG00000132507	Semen, vaginal fluid	ENSG00000159763	Saliva, semen
ENSG00000132746	Semen, vaginal fluid	ENSG00000160862	Semen, lachrymal
ENSG00000134757	Semen, vaginal fluid	ENSG00000162078	Saliva, lachrymal
ENSG00000136689	Saliva, semen	ENSG00000167419	Lachrymal, saliva, blood

ENSG00000137273	Saliva, vaginal fluid	ENSG00000167656	Saliva, vaginal fluid
ENSG00000137440	Saliva, vaginal fluid	ENSG00000170373	Lachrymal, saliva
ENSG00000137699	Saliva, vaginal fluid, lachrymal	ENSG00000185479	Saliva, vaginal fluid
ENSG00000141469	Blood, semen	ENSG00000186832	Saliva, vaginal fluid
ENSG00000143409	Blood, saliva	ENSG00000198496	Blood, saliva
ENSG00000143546	Blood, saliva, vaginal fluid	ENSG00000241794	Saliva, vaginal fluid
ENSG00000143556	Saliva, vaginal fluid	ENSG00000205420	Saliva, vaginal fluid
ENSG00000145819	Blood, saliva	ENSG00000177192	Semen, vaginal fluid
ENSG00000159763	Saliva, semen	ENSG00000181444	Blood, semen
ENSG00000163207	Semen, vaginal fluid	ENSG00000184470	Semen, vaginal fluid
ENSG00000163220	Blood, vaginal fluid	ENSG00000186081	Semen, vaginal fluid
ENSG00000165474	Saliva, vaginal fluid	ENSG00000186723	Semen, saliva
ENSG00000167419	Blood, saliva	ENSG00000230389	Blood, vaginal fluid
ENSG00000167815	Blood, saliva	ENSG00000229035	Saliva, vaginal fluid
ENSG00000167880	Semen, vaginal fluid	ENSG00000235568	Blood, saliva
ENSG00000169469	Saliva, vaginal fluid	ENSG00000012223	Blood, lachrymal
ENSG00000169474	Saliva, semen, vaginal fluid	ENSG00000240508	Blood, saliva
ENSG00000171051	Blood, semen	ENSG00000268569	Blood, saliva
ENSG00000172432	Saliva, vaginal fluid	ENSG00000268682	Semen, vaginal fluid

2.3.3 Elimination of Genes with Unconfirmed Protein Products

Since not all transcribed RNAs are translated into proteins, 190 genes for which protein products could not be confirmed were removed (Table 2.4). Thereafter, 1464 genes remained for further analysis.

Table 2.4: Genes over-expressed in surrogate tissues of saliva, blood, semen and vaginal fluid that could not be confirmed as protein coding.

EnsEMBL ID	Surrogate tissue	EnsEMBL ID	Surrogate tissue
ENSG00000038219	Blood	ENSG00000160877	Lachrymal
ENSG00000048740	Blood	ENSG00000004777	Lachrymal
ENSG00000060138	Blood	ENSG00000105341	Semen
ENSG00000074706	Blood	ENSG00000113719	Semen
ENSG00000100055	Blood	ENSG00000137634	Semen
ENSG00000103381	Blood	ENSG00000139865	Semen
ENSG00000105122	Blood	ENSG00000145331	Semen
ENSG00000108055	Blood	ENSG00000146205	Semen
ENSG00000108669	Blood	ENSG00000148803	Semen
ENSG00000109684	Blood	ENSG00000152931	Semen
ENSG00000122122	Blood	ENSG00000159182	Semen
ENSG00000124193	Blood	ENSG00000160336	Semen
ENSG00000126759	Blood	ENSG00000165501	Semen
ENSG00000129534	Blood	ENSG00000171053	Semen
ENSG00000130038	Blood	ENSG00000171462	Semen
ENSG00000131969	Blood	ENSG00000171517	Semen
ENSG00000134698	Blood	ENSG00000179151	Semen
ENSG00000137674	Blood	ENSG00000179151	Semen
ENSG00000150867	Blood	ENSG00000182165	Semen
ENSG00000156026	Blood	ENSG00000182796	Semen
ENSG00000156265	Blood	ENSG00000185186	Semen
ENSG00000160796	Blood	ENSG00000197588	Semen
ENSG00000163009	Blood	ENSG00000204174	Semen
ENSG00000163116	Blood	ENSG00000205133	Semen
ENSG00000165097	Blood	ENSG00000207720	Semen
ENSG00000166501	Blood	ENSG00000214049	Semen
ENSG00000168071	Blood	ENSG00000222036	Semen

ENSG00000169877	Blood	ENSG00000225937	Semen
ENSG00000170006	Blood	ENSG00000226516	Semen
ENSG00000170456	Blood	ENSG00000227418	Semen
ENSG00000172197	Blood	ENSG00000229637	Semen
ENSG00000175106	Blood	ENSG00000236699	Semen
ENSG00000177663	Blood	ENSG00000243766	Semen
ENSG00000178449	Blood	ENSG00000251321	Semen
ENSG00000180071	Blood	ENSG00000255794	Semen
ENSG00000184730	Blood	ENSG00000257227	Semen
ENSG00000183688	Blood	ENSG00000266109	Semen
ENSG00000183597	Blood	ENSG00000266722	Semen
ENSG00000185811	Blood	ENSG00000268369	Semen
ENSG00000187240	Blood	ENSG00000269936	Semen
ENSG00000187605	Blood	ENSG00000271942	Semen
ENSG00000187866	Blood	ENSG00000167925	Vaginal fluid
ENSG00000197405	Blood	ENSG00000168476	Vaginal fluid
ENSG00000197561	Blood	ENSG00000170523	Vaginal fluid
ENSG00000254470	Blood	ENSG00000171711	Vaginal fluid
ENSG00000008382	Saliva	ENSG00000172867	Vaginal fluid
ENSG00000008869	Saliva	ENSG00000173581	Vaginal fluid
ENSG00000073282	Saliva	ENSG00000174206	Vaginal fluid
ENSG00000103544	Saliva	ENSG00000174564	Vaginal fluid
ENSG00000104413	Saliva	ENSG00000175575	Vaginal fluid
ENSG00000104979	Saliva	ENSG00000175938	Vaginal fluid
ENSG00000109775	Saliva	ENSG00000179941	Vaginal fluid
ENSG00000129460	Saliva	ENSG00000183242	Vaginal fluid
ENSG00000133401	Saliva	ENSG00000185761	Vaginal fluid
ENSG00000134256	Saliva	ENSG00000186130	Vaginal fluid
ENSG00000135372	Saliva	ENSG00000186205	Vaginal fluid
ENSG00000136231	Saliva	ENSG00000196422	Vaginal fluid

ENSG00000136688	Saliva	ENSG00000196943	Vaginal fluid
ENSG00000141499	Saliva	ENSG00000197079	Vaginal fluid
ENSG00000143624	Saliva	ENSG00000197565	Vaginal fluid
ENSG00000144468	Saliva	ENSG00000205426	Vaginal fluid
ENSG00000144535	Saliva	ENSG00000214374	Vaginal fluid
ENSG00000145860	Saliva	ENSG00000223414	Vaginal fluid
ENSG00000145934	Saliva	ENSG00000143374	Vaginal fluid
ENSG00000146410	Saliva	ENSG00000149311	Vaginal fluid
ENSG00000154102	Saliva	ENSG00000158483	Vaginal fluid
ENSG00000156931	Saliva	ENSG00000166938	Vaginal fluid
ENSG00000164221	Saliva	ENSG00000135443	Vaginal fluid
ENSG00000164941	Saliva	ENSG00000136630	Vaginal fluid
ENSG00000165102	Saliva	ENSG00000137267	Vaginal fluid
ENSG00000169599	Saliva	ENSG00000137309	Vaginal fluid
ENSG00000170465	Saliva	ENSG00000108759	Vaginal fluid
ENSG00000170802	Saliva	ENSG00000109805	Vaginal fluid
ENSG00000174579	Saliva	ENSG00000110104	Vaginal fluid
ENSG00000175643	Saliva	ENSG00000113119	Vaginal fluid
ENSG00000179104	Saliva	ENSG00000119943	Vaginal fluid
ENSG00000182318	Saliva	ENSG00000125967	Vaginal fluid
ENSG00000182463	Saliva	ENSG00000129474	Vaginal fluid
ENSG00000184076	Saliva	ENSG00000130193	Vaginal fluid
ENSG00000254126	Saliva	ENSG00000130707	Vaginal fluid
ENSG00000266074	Saliva	ENSG00000131351	Vaginal fluid
ENSG00000266714	Saliva	ENSG00000131738	Vaginal fluid
ENSG00000127325	Saliva	ENSG00000100300	Vaginal fluid
ENSG00000110075	Saliva	ENSG00000100350	Vaginal fluid
ENSG00000110328	Saliva	ENSG00000042286	Vaginal fluid
ENSG00000116785	Saliva	ENSG00000064545	Vaginal fluid
ENSG00000204361	Saliva	ENSG00000070731	Vaginal fluid

ENSG00000218089	Saliva	ENSG00000080986	Vaginal fluid
ENSG00000223380	Saliva	ENSG00000084693	Vaginal fluid
ENSG00000198315	Saliva	ENSG00000088986	Vaginal fluid
ENSG00000184330	Saliva	ENSG00000089195	Vaginal fluid
ENSG00000196652	Saliva	ENSG00000094796	Vaginal fluid
ENSG00000254127	Lachrymal	ENSG00000005302	Vaginal fluid
ENSG00000083168	Lachrymal	ENSG00000011143	Vaginal fluid
ENSG00000167671	Lachrymal	ENSG00000170540	Saliva

2.3.4 Methylation Data

CpG Island (CGI) methylation information was obtained for the remaining 1464 genes in tissues which displayed low-to-normal gene expression. A total of 952 genes outlined in Table 2.5, have CGIs positioned within them and were therefore included in the study. The genes which did not have any CGIs positioned within them and were excluded from further analysis are outlined in Table 1, Appendix A.

Table 2.5: List of genes harbouring CGIs.

ENSG00000155545	ENSG00000145860	ENSG00000170802	ENSG00000196353
ENSG00000178226	ENSG00000176597	ENSG00000162040	ENSG00000132205
ENSG00000069011	ENSG00000136490	ENSG00000050628	ENSG00000100266
ENSG00000215440	ENSG00000109255	ENSG00000161798	ENSG00000082175
ENSG00000185947	ENSG00000100359	ENSG00000167130	ENSG00000140564
ENSG00000111328	ENSG00000107077	ENSG00000125965	ENSG00000181577
ENSG00000161992	ENSG00000136161	ENSG00000198464	ENSG00000115963
ENSG00000164221	ENSG00000242110	ENSG00000198863	ENSG00000107554
ENSG00000088205	ENSG00000089639	ENSG00000092758	ENSG00000184825
ENSG00000173581	ENSG00000046604	ENSG00000145592	ENSG00000125347
ENSG00000135636	ENSG00000174851	ENSG00000134516	ENSG00000152359
ENSG00000172244	ENSG00000183242	ENSG00000103174	ENSG00000204842
ENSG00000164164	ENSG00000101333	ENSG00000179111	ENSG00000173535
ENSG00000065268	ENSG00000148110	ENSG00000198001	ENSG00000143147
ENSG00000167004	ENSG00000100461	ENSG00000163660	ENSG00000103126
ENSG00000158019	ENSG00000170961	ENSG00000123609	ENSG00000005073
ENSG00000158715	ENSG00000130707	ENSG00000134574	ENSG00000144724
ENSG00000119559	ENSG00000141298	ENSG00000107738	ENSG00000165502
ENSG00000185024	ENSG00000196943	ENSG00000185215	ENSG00000128641
ENSG00000106780	ENSG00000170581	ENSG00000130119	ENSG00000077782
ENSG00000140332	ENSG00000149554	ENSG00000174564	ENSG00000112234
ENSG00000166451	ENSG00000174292	ENSG00000112062	ENSG00000119760
ENSG00000241399	ENSG00000157540	ENSG00000145934	ENSG00000130300

ENSG00000083307	ENSG00000125257	ENSG00000168875	ENSG00000149177
ENSG00000068650	ENSG00000114209	ENSG00000152822	ENSG00000122335
ENSG00000097007	ENSG00000104267	ENSG00000111186	ENSG00000168461
ENSG00000134201	ENSG00000155657	ENSG00000188372	ENSG00000188130
ENSG00000048140	ENSG00000205133	ENSG00000186205	ENSG00000075290
ENSG00000062282	ENSG00000185088	ENSG00000161055	ENSG00000005381
ENSG00000171055	ENSG00000114779	ENSG00000106714	ENSG00000121671
ENSG00000101624	ENSG00000151617	ENSG00000184381	ENSG00000144668
ENSG00000075218	ENSG00000170485	ENSG00000135698	ENSG00000165097
ENSG00000180739	ENSG00000109743	ENSG00000160145	ENSG00000103245
ENSG00000184056	ENSG00000164284	ENSG00000125967	ENSG00000166938
ENSG00000158417	ENSG00000156482	ENSG00000188419	ENSG00000060138
ENSG00000132604	ENSG00000065621	ENSG00000021762	ENSG00000166006
ENSG0000007278	ENSG00000203747	ENSG00000149489	ENSG00000166509
ENSG00000120899	ENSG00000158470	ENSG00000137709	ENSG00000124251
ENSG00000136937	ENSG00000182858	ENSG00000266714	ENSG00000137841
ENSG00000162676	ENSG00000167034	ENSG00000038219	ENSG00000080815
ENSG00000096070	ENSG00000182796	ENSG00000074370	ENSG00000113013
ENSG00000177200	ENSG00000196652	ENSG00000148300	ENSG00000070371
ENSG00000172197	ENSG00000167202	ENSG00000198496	ENSG00000169174
ENSG00000151726	ENSG00000008283	ENSG00000108669	ENSG00000169599
ENSG00000107159	ENSG00000165527	ENSG00000117691	ENSG00000144339
ENSG00000104856	ENSG00000153179	ENSG00000064115	ENSG00000179583
ENSG00000082293	ENSG00000182318	ENSG00000150867	ENSG00000133067
ENSG00000152229	ENSG00000140993	ENSG00000105427	ENSG00000167705
ENSG00000180340	ENSG00000156427	ENSG00000151835	ENSG00000198604
ENSG00000162493	ENSG00000058866	ENSG00000151702	ENSG00000047648
ENSG00000168092	ENSG00000068831	ENSG00000158483	ENSG00000088247
ENSG00000072195	ENSG00000169213	ENSG00000068878	ENSG00000038002
ENSG00000113296	ENSG00000198729	ENSG00000253293	ENSG00000113119
ENSG00000169020	ENSG00000184708	ENSG00000169621	ENSG00000182400
ENSG00000104960	ENSG00000134215	ENSG00000077063	ENSG00000182162
ENSG00000125845	ENSG00000130720	ENSG00000175264	ENSG00000175938
ENSG00000184307	ENSG00000184956	ENSG00000114480	ENSG00000059378
ENSG00000122585	ENSG00000198382	ENSG00000122592	ENSG00000076928
ENSG00000173545	ENSG00000099953	ENSG00000141574	ENSG00000070031
ENSG00000184012	ENSG00000128652	ENSG00000009413	ENSG00000121316
ENSG00000187605	ENSG00000140279	ENSG00000198211	ENSG00000196357
ENSG00000104047	ENSG00000156983	ENSG00000117525	ENSG00000197622
ENSG00000130803	ENSG00000197561	ENSG00000146205	ENSG00000142347
ENSG00000142765	ENSG00000136378	ENSG00000108932	ENSG00000135655
ENSG00000168071	ENSG00000262209	ENSG00000121742	ENSG00000103257
ENSG00000196428	ENSG00000140577	ENSG00000153113	ENSG00000138780
ENSG00000008869	ENSG00000167815	ENSG00000244274	ENSG00000171517
ENSG00000171462	ENSG00000157613	ENSG00000112699	ENSG00000077420
ENSG00000076662	ENSG00000186523	ENSG00000145569	ENSG00000115163
ENSG00000133704	ENSG00000204160	ENSG00000152256	ENSG00000084693
ENSG00000148803	ENSG00000000938	ENSG00000110934	ENSG00000117280
ENSG00000165501	ENSG00000126822	ENSG00000123179	ENSG00000198315
ENSG00000001167	ENSG00000036549	ENSG00000133028	ENSG00000106003
ENSG00000167658	ENSG00000173480	ENSG00000103222	ENSG00000198951
ENSG00000118197	ENSG00000035664	ENSG00000197965	ENSG00000140945
ENSG00000171840	ENSG00000043355	ENSG00000171174	ENSG00000104064
ENSG00000225921	ENSG00000185761	ENSG00000145331	ENSG00000186635
ENSG00000183166	ENSG00000178038	ENSG00000169981	ENSG00000131591
ENSG00000109184	ENSG00000230989	ENSG00000145916	ENSG00000103381
ENSG00000170540	ENSG00000153094	ENSG00000149488	ENSG00000110075
ENSG00000121691	ENSG00000136231	ENSG00000176788	ENSG00000183943
ENSG00000105538	ENSG00000130741	ENSG00000101773	ENSG00000165879
ENSG00000032219	ENSG00000134283	ENSG00000119986	ENSG00000182230

ENSG00000131351	ENSG00000128268	ENSG00000105835	ENSG00000178449
ENSG00000169851	ENSG00000179361	ENSG00000124243	ENSG00000124614
ENSG00000137193	ENSG00000169083	ENSG00000184922	ENSG00000115380
ENSG00000103343	ENSG00000174669	ENSG00000005302	ENSG00000118849
ENSG00000101596	ENSG00000170577	ENSG00000148655	ENSG00000109881
ENSG00000010017	ENSG00000267645	ENSG00000196628	ENSG00000187583
ENSG00000154553	ENSG00000101336	ENSG00000066427	ENSG00000110400
ENSG00000131018	ENSG00000019186	ENSG00000171135	ENSG00000040608
ENSG00000132773	ENSG00000006611	ENSG00000139289	ENSG00000085719
ENSG00000214160	ENSG00000167580	ENSG00000175575	ENSG00000182566
ENSG00000006016	ENSG00000049319	ENSG00000142197	ENSG00000128040
ENSG00000137699	ENSG00000136819	ENSG00000180818	ENSG00000145113
ENSG00000139190	ENSG00000185721	ENSG00000124164	ENSG00000197261
ENSG00000183570	ENSG00000176386	ENSG00000137273	ENSG00000135372
ENSG00000198554	ENSG00000108443	ENSG00000158615	ENSG00000057704
ENSG00000101654	ENSG00000127989	ENSG00000146733	ENSG00000150760
ENSG00000186340	ENSG00000099812	ENSG00000112541	ENSG00000172006
ENSG00000106031	ENSG00000015285	ENSG00000136286	ENSG00000187079
ENSG00000115756	ENSG00000225930	ENSG00000136877	ENSG00000166923
ENSG00000170425	ENSG00000089195	ENSG00000104976	ENSG00000182580
ENSG00000179104	ENSG00000117151	ENSG00000124193	ENSG00000108055
ENSG00000005961	ENSG00000167693	ENSG00000096696	ENSG00000183091
ENSG00000167513	ENSG00000135604	ENSG00000115750	ENSG00000112146
ENSG00000177105	ENSG00000168769	ENSG00000058056	ENSG00000064270
ENSG00000166974	ENSG00000173511	ENSG00000184076	ENSG00000149499
ENSG00000123146	ENSG00000173442	ENSG00000223414	ENSG00000005020
ENSG00000167799	ENSG00000115361	ENSG00000130635	ENSG00000137414
ENSG00000134901	ENSG00000174125	ENSG00000128585	ENSG00000026103
ENSG00000187735	ENSG00000100311	ENSG00000185567	ENSG00000104447
ENSG00000182481	ENSG00000150281	ENSG00000128713	ENSG00000155849
ENSG00000102554	ENSG00000197361	ENSG00000136718	ENSG00000042286
ENSG00000169896	ENSG00000116478	ENSG00000152078	ENSG00000174579
ENSG00000053747	ENSG00000110324	ENSG00000249859	ENSG00000134046
ENSG00000066336	ENSG00000168476	ENSG00000162129	ENSG00000145819
ENSG00000106351	ENSG00000185345	ENSG00000204271	ENSG00000163067
ENSG00000118898	ENSG00000110375	ENSG00000146457	ENSG00000064012
ENSG00000113719	ENSG00000174206	ENSG00000134698	ENSG00000125657
ENSG00000102362	ENSG00000156931	ENSG00000105122	ENSG00000173546
ENSG00000187840	ENSG00000214021	ENSG00000144909	ENSG00000124466
ENSG00000170178	ENSG00000124440	ENSG00000141655	ENSG00000183688
ENSG00000090339	ENSG00000198858	ENSG00000198088	ENSG00000174562
ENSG00000115138	ENSG00000180071	ENSG00000164920	ENSG00000196767
ENSG00000162526	ENSG00000141380	ENSG00000136536	ENSG00000028137
ENSG00000124766	ENSG00000183762	ENSG00000117143	ENSG00000140836
ENSG00000030304	ENSG00000187266	ENSG00000172432	ENSG00000221866
ENSG00000140519	ENSG00000095261	ENSG00000128709	ENSG00000171097
ENSG00000160191	ENSG00000120457	ENSG00000144674	ENSG00000254087
ENSG00000021355	ENSG00000140396	ENSG00000165533	ENSG00000130038
ENSG00000138162	ENSG00000137198	ENSG00000211456	ENSG00000167700
ENSG00000115306	ENSG00000165102	ENSG00000122644	ENSG00000134419
ENSG00000156017	ENSG00000166133	ENSG00000166710	ENSG00000171700
ENSG00000168209	ENSG00000146676	ENSG00000156273	ENSG00000130592
ENSG00000124831	ENSG00000138180	ENSG00000048740	ENSG00000158406
ENSG00000100599	ENSG00000111913	ENSG00000198736	ENSG00000182287
ENSG00000196358	ENSG00000174775	ENSG00000121749	ENSG00000137857
ENSG00000160703	ENSG00000181222	ENSG00000133401	ENSG00000168310
ENSG00000151657	ENSG00000142156	ENSG00000171307	ENSG00000074657
ENSG00000074319	ENSG00000163357	ENSG00000197548	ENSG00000143815
ENSG00000173530	ENSG00000169752	ENSG00000183597	ENSG00000113448
ENSG00000166501	ENSG00000158711	ENSG00000119121	ENSG00000144029

ENSG00000049759	ENSG00000104413	ENSG00000129933	ENSG00000139874
ENSG00000130844	ENSG00000100350	ENSG00000159202	ENSG00000173171
ENSG00000204219	ENSG00000172530	ENSG00000197045	ENSG00000148842
ENSG00000131389	ENSG00000119231	ENSG00000156113	ENSG00000101868
ENSG00000143761	ENSG00000117148	ENSG00000130775	ENSG00000160796
ENSG00000163795	ENSG00000078061	ENSG00000141404	ENSG00000169855
ENSG00000185009	ENSG00000109775	ENSG00000198900	ENSG00000149212
ENSG00000107625	ENSG00000166477	ENSG00000183878	ENSG00000116815
ENSG00000119714	ENSG00000160712	ENSG00000100285	ENSG00000104979
ENSG00000124383	ENSG00000135838	ENSG00000071054	ENSG00000035928
ENSG00000081277	ENSG00000141480	ENSG00000101096	ENSG00000137809
ENSG00000118855	ENSG00000123329	ENSG00000181274	ENSG00000008516
ENSG00000105426	ENSG00000100416	ENSG00000079482	ENSG00000196422
ENSG00000157570	ENSG00000106078	ENSG00000167996	ENSG00000198000
ENSG00000159842	ENSG00000101425	ENSG00000169180	ENSG00000187664
ENSG00000140368	ENSG00000107551	ENSG00000092036	ENSG00000168386
ENSG00000111790	ENSG00000184937	ENSG00000159648	ENSG00000119684
ENSG00000104154	ENSG00000134202	ENSG00000136867	ENSG00000132906
ENSG00000006194	ENSG00000164647	ENSG00000159461	ENSG00000243766
ENSG00000132570	ENSG00000132507	ENSG00000063169	ENSG00000108306
ENSG00000130755	ENSG00000100985	ENSG00000217555	ENSG00000186642
ENSG00000180447	ENSG00000015475	ENSG00000182568	ENSG00000154229
ENSG00000182022	ENSG00000101236	ENSG00000178935	ENSG00000165689
ENSG00000145088	ENSG00000117411	ENSG00000103855	ENSG00000122420
ENSG00000143537	ENSG00000135069	ENSG00000184470	ENSG00000254806
ENSG00000151651	ENSG00000112208	ENSG00000011198	ENSG00000184281
ENSG00000116747	ENSG00000160753	ENSG00000163069	ENSG00000105656
ENSG00000131759	ENSG00000132334	ENSG00000146410	ENSG00000114115
ENSG00000187778	ENSG00000077943	ENSG00000089154	ENSG00000069275
ENSG00000198755	ENSG00000147889	ENSG00000138346	ENSG00000075643
ENSG00000139880	ENSG00000105204	ENSG00000103202	ENSG00000182257
ENSG00000108264	ENSG00000179029	ENSG00000163637	ENSG00000113231
ENSG00000203685	ENSG00000136167	ENSG00000167419	ENSG00000172572
ENSG00000147065	ENSG00000150893	ENSG00000008130	ENSG00000177663
ENSG00000175643	ENSG00000089248	ENSG00000155506	ENSG00000066468
ENSG00000128714	ENSG00000123388	ENSG00000076242	ENSG00000141002
ENSG00000158042	ENSG00000130733	ENSG00000113648	ENSG00000127191
ENSG00000107187	ENSG00000204138	ENSG00000206560	ENSG00000017797
ENSG00000115761	ENSG00000064687	ENSG00000077157	ENSG00000099341
ENSG00000198108	ENSG00000162714	ENSG00000165816	ENSG00000164032
ENSG00000117115	ENSG00000151503	ENSG00000101255	ENSG00000131969
ENSG00000169567	ENSG00000050767	ENSG00000185811	ENSG00000012061
ENSG00000091536	ENSG00000176390	ENSG00000110328	ENSG00000165752
ENSG00000117713	ENSG00000083454	ENSG00000177192	ENSG00000131873
ENSG00000140406	ENSG00000168077	ENSG00000100983	ENSG00000173068
ENSG00000135363	ENSG00000181444	ENSG00000213390	ENSG00000175054
ENSG00000124225	ENSG00000159733	ENSG00000173214	ENSG00000179151
ENSG00000091831	ENSG00000141524	ENSG00000107262	ENSG00000172789
ENSG00000168010	ENSG00000008382	ENSG00000085982	ENSG00000128283
ENSG00000139865	ENSG00000100802	ENSG00000138735	ENSG00000167880
ENSG00000167173	ENSG00000109572	ENSG00000123364	ENSG00000259112
ENSG00000154025	ENSG00000172987	ENSG00000105483	ENSG00000120158
ENSG00000103544	ENSG00000173575	ENSG00000170456	ENSG00000159231
ENSG00000129657	ENSG00000114626	ENSG00000075073	ENSG00000166592
ENSG00000144554	ENSG00000090470	ENSG00000088543	ENSG00000007541
ENSG00000120616	ENSG00000125457	ENSG00000171444	ENSG00000029534
ENSG00000112996	ENSG00000160255	ENSG00000147324	ENSG00000133627
ENSG00000117614	ENSG00000196655	ENSG00000100504	ENSG00000072401
ENSG00000104043	ENSG00000180370	ENSG00000112964	ENSG00000171862
ENSG00000130590	ENSG00000165312	ENSG00000168067	ENSG00000127616

ENSG00000137642	ENSG00000178403	ENSG00000102034	ENSG00000108591
ENSG00000145808	ENSG00000157554	ENSG00000111432	ENSG00000088387
ENSG00000180758	ENSG00000163798	ENSG00000119699	ENSG00000123700
ENSG00000088386	ENSG00000167114	ENSG00000100393	ENSG00000175106
ENSG00000109180	ENSG00000150051	ENSG00000115524	ENSG00000198918
ENSG00000182463	ENSG00000161513	ENSG00000125691	ENSG00000133961
ENSG00000159788	ENSG00000107968	ENSG00000175356	ENSG00000100314
ENSG00000187866	ENSG00000149269	ENSG00000143578	ENSG00000177548
ENSG00000198924	ENSG00000105676	ENSG00000102879	ENSG00000138376
ENSG00000172845	ENSG00000005513	ENSG00000172936	ENSG00000169509
ENSG00000187741	ENSG00000163497	ENSG00000144535	ENSG00000170370
ENSG00000171729	ENSG00000162302	ENSG00000108474	ENSG00000062038
ENSG00000171763	ENSG00000116741	ENSG00000168765	ENSG00000171777
ENSG00000221829	ENSG00000007038	ENSG00000049239	ENSG00000144063
ENSG00000197299	ENSG00000107018	ENSG00000083223	ENSG00000129534
ENSG00000129451	ENSG00000144468	ENSG00000132849	ENSG00000165474
ENSG00000090889	ENSG00000133805	ENSG00000187210	ENSG00000143801
ENSG00000164941	ENSG00000102539	ENSG00000109805	ENSG00000088826
ENSG00000159184	ENSG00000087460	ENSG00000170571	ENSG00000163975
ENSG00000133392	ENSG00000076944	ENSG00000034713	ENSG00000173110
ENSG00000171608	ENSG00000158717	ENSG00000169826	ENSG00000151715
ENSG00000133302	ENSG00000106591	ENSG00000145824	ENSG00000136630
ENSG00000136040	ENSG00000106038	ENSG00000139132	ENSG00000150361
ENSG00000015133	ENSG00000254470	ENSG00000136522	ENSG00000156026
ENSG00000089159	ENSG00000115271	ENSG00000114738	ENSG00000100030
ENSG00000135312	ENSG00000124171	ENSG00000068024	ENSG00000154102
ENSG00000143669	ENSG00000102854	ENSG00000143768	ENSG00000140044
ENSG00000138696	ENSG00000134996	ENSG00000170265	ENSG00000135441
ENSG00000148572	ENSG00000143476	ENSG00000168556	ENSG00000182307
ENSG00000129474	ENSG00000114026	ENSG00000170615	ENSG00000179008
ENSG00000104738	ENSG00000113387	ENSG00000104093	ENSG00000165912
ENSG00000164078	ENSG00000129514	ENSG00000088992	ENSG00000139842
ENSG00000176170	ENSG00000102921	ENSG00000128965	ENSG00000182132
ENSG00000101412	ENSG00000132436	ENSG00000198768	ENSG00000130193
ENSG00000110090	ENSG00000181192		

2.3.5 Mapping CpG Island and Methylation Information

After finding the 952 genes in which CGIs were located, the mean, median and min-max methylation levels for each of the CGIs were calculated (based on a merged ENCODE and NGSMethDB dataset). CGIs of 50 genes which displayed a mean, median and min-max methylation level above 75% (heavily methylated and low-to-normal expression in non-target tissues) but over-expression in saliva, blood, semen and vaginal fluid were selected (methylation levels are listed in Table 2, Appendix A). The genes where CGIs displayed a methylation level below 75% are presented in Table 3, Appendix A. Since multiple CGIs within certain genes displayed high methylation, all of these CGIs were included. Table 2.6 outlines the genes with CGIs displaying above 75% methylation and the body fluids in which those genes are over-expressed.

Table 2.6: Genes displaying methylation levels above 75% (hypermethylation) in tissues with low-to-normal expression, the body fluid it is over-expressed in and the number of CGIs selected from the gene.

EnsEMBL ID	Gene Abbreviation	Name of Gene	Body Fluid	Number of CGIs selected
ENSG00000135636	DYSF	Dysferlin	Blood	1
ENSG00000172244	C5orf34	Chromosome 5 open reading frame 34	Vaginal fluid	1
ENSG00000065268	WDR18	WD repeat domain 18	Vaginal fluid	1
ENSG00000082293	COL19A1	Collagen, type XIX, alpha 1	Blood	1
ENSG00000160145	KALRN	Kalirin, RhoGEF kinase	Semen	1
ENSG00000107077	KDM4C	Lysine (K)-specific demethylase 4C	Blood	3
ENSG00000155657	TTN	Titin	Saliva	1
ENSG00000174564	IL20RB	Interleukin 20 receptor beta	Vaginal fluid	1
ENSG00000145934	TENM2	Teneurin transmembrane protein 2	Saliva	2
ENSG00000196353	CPNE4	Copine IV	Semen	1
ENSG00000100266	PACSIN2	Protein kinase C and casein kinase substrate in neurons 2	Saliva	2
ENSG00000128641	MYO1B	Myosin IB	Saliva	1
ENSG00000130300	PLVAP	Plasmalemma vesicle associated protein	Saliva	1
ENSG00000149596	JPH2	Junctophilin 2	Saliva	1
ENSG00000006459	KDM7A	Lysine (K)-specific demethylase 7A	Semen	1
ENSG00000188419	CHM	Choroideremia (Rab escort protein 1)	Semen	1
ENSG00000187605	TET3	Tet methylcytosine dioxygenase 3	Blood	1
ENSG00000130803	ZNF317	Zinc finger protein 317	Saliva	1
ENSG00000146205	ANO7	Anoctamin 7	Blood	1
ENSG00000088247	KHSRP	KH-type splicing regulatory protein	Semen	1
ENSG00000167658	EEF2	Eukaryotic translation elongation factor 2	Saliva	1
ENSG00000136231	IGF2BP3	Insulin-like growth factor 2 mRNA binding protein 3	Saliva	1
ENSG00000183570	PCBP3	Poly(rC) binding protein 3	Blood	4
ENSG00000115756	HPCAL1	Hippocalcin-Like 1	Vaginal fluid	1
ENSG00000064012	CASP8	Caspase 8, apoptosis-related cysteine peptidase	Blood	1

ENSG00000030304	MUSK	Muscle, skeletal, receptor tyrosine kinase	Saliva	1
ENSG00000115306	SPTBN1	Spectrin, beta, non-erythrocytic 1	Saliva	1
ENSG00000196358	NTNG2	Netrin G2	Blood	1
ENSG00000165102	HGSNAT	Heparan-alpha-glucosaminide N-acetyltransferase	Saliva	1
ENSG00000197548	ATG7	Autophagy related 7	Blood	1
ENSG00000119121	TRPM6	Transient receptor potential cation channel, subfamily M, member 6	Blood	1
ENSG00000113448	PDE4D	Phosphodiesterase 4D, cAMP-specific	Semen	1
ENSG00000101868	POLA1	Polymerase (DNA directed), alpha 1, catalytic subunit	Semen	1
ENSG00000105426	PTPRS	Protein tyrosine phosphatase, receptor type, S	Saliva	2
ENSG00000124225	PMEPA1	Prostate transmembrane protein, androgen induced 1	Semen	1
ENSG00000091831	ESR1	Estrogen receptor 1	Vaginal fluid	1
ENSG00000015475	BID BH3	BH3 interacting domain death agonist	Blood	1
ENSG00000101096	NFATC1	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2	Semen	1
ENSG00000079482	OPHN1	Oligophrenin 1	Saliva	2
ENSG00000063169	GLTSCR1	Glioma tumor suppressor candidate region gene 1	Semen	3
ENSG00000184470	TXNRD2	Thioredoxin reductase 2	Vaginal Fluid	1
ENSG00000050767	COL23A1	Collagen, type XXIII, alpha 1	Blood	1
ENSG00000159733	ZFYVE28	Zinc finger, FYVE domain containing 28	Saliva	1
ENSG00000109572	CLCN3	Chloride channel, voltage-sensitive 3	Saliva	1
ENSG00000127191	TRAF2	TNF receptor-associated factor 2	Vaginal Fluid	1
ENSG00000206560	ANKRD28	Ankyrin repeat domain 28	Blood	1
ENSG00000090889	KIF4A	Kinesin family member 4A	Vaginal fluid	2
ENSG00000083223	ZCCHC6	Zinc finger, CCHC domain containing 6	Blood	1
ENSG00000068024	HDAC4	Histone deacetylase 4	Blood	2
ENSG00000170265	ZNF282	Zinc finger protein 282	Vaginal fluid	1

2.3.6 Final Selection of CGIs for Primer Design

The heavily methylated CGIs of the selected genes were BLASTed against the human genome assembly to ensure that the CGIs mapped to a single location in the human genome (Table 4 and Figure 2-11, Appendix A) and were also checked for the presence of *Hha*I restriction site. Finally, out of 50 genes, 10 candidate genes met the above criteria and were selected for primer design (Table 2.7).

Table 2.7: Candidate genes selected for primer design and chromosomal locations of targeted CGIs.

EnsEMBL ID	Chromosome Location	Abbreviation	Name of Gene
ENSG00000135636	Chromosome 2	DYSF	Dysferlin
ENSG00000115756	Chromosome 2	HPCAL1	Hippocalcin-Like 1
ENSG00000196353	Chromosome 3	CPNE4	Copine IV
ENSG00000091831	Chromosome 6	ESR1	Estrogen Receptor 1
ENSG00000170265	Chromosome 7	ZNF282	Zinc Finger Protein 2
ENSG00000136231	Chromosome 7	IGF2BP3	Insulin-like Growth Factor 2 mRNA Binding Protein 3
ENSG00000030304	Chromosome 9	MUSK	Muscle, Skeletal, Receptor Tyrosine Kinase
ENSG00000167658	Chromosome 19	EEF2	Eukaryotic Translation Elongation Factor 2
ENSG00000105426	Chromosome 19	PTPRS	Protein Tyrosine Phosphatase, Receptor Type S
ENSG00000124225	Chromosome 20	PMEPA1	Prostate Transmembrane Protein, Androgen Induced 1

The sequences of all the candidate tDMR-based primers along with their specific PCR melting temperatures, annealing temperatures and amplicons are presented in Table 2.8.

Table 2.8: Sequence of PCR primers targeting selected CGIs.

Gene abbreviation/ Marker	Forward primer sequences (5'→3')	Reverse primer sequences (5'→3')	Forward primer Melting Temperature	Reverse primer Melting Temperature		Amplicon Size (bp)
DYSF	GCATAAATCACCCTGGTTGG	GGAAAATGGAGAGGAGTCCC	55	56	50	599
HPCAL1	GCCTTCTTGGTGGTCCATAA	TGTCTTGCAGGTGTAGTCGC	56	59	51	178
CPNE4	CGCCTGACTGGTTTTCGTAT	GAACCCGTATGGGAGGTGA	56	58	51	554
ESR1	ACGGGTGACTTCTGCATTTC	GACAGCTTTGAAGAGGGCAG	56	57	51	568
ZNF282	CTATCTCCCCAGGTGACAGC	CGAGAAGAGCTTCAACTGCC	58	58	53	504
IGF2BP3	CATCACCAAGTCCCTGTACG	GTTCTTTCGCCTTTCAGCAC	56	55	50	584
MUSK	GAACAGCTCCGGTCTACAGC	TCTGTAGGCTCCACCTCTGG	59	60	54	528
EEF2	TCAGCACACTGGCATAGAGG	CACCTCGCCTTTATCGATGT	58	56	51	565
PTPRS	CATAACCCACAAACCGCTCT	GATGCTGTAACGTGTGGTGG	56	57	51	517
PMEPA1	CTGATGCCCGAGTTACTGCT	CGTGTGCAGAGAGCAGAGAG	58	59	53	580

Upon initial analysis of the status of methylation of the targeted CGIs in saliva, blood, semen and vaginal fluid by MSRE-PCR, only four candidate genes (*HPCAL1, ZNF282, PTPRS* and *PMEPA1*) showed differential methylation profiles and thus were selected for further analysis.

Table 2.9: PCR primers and concentrations used for amplification of potential tDMRs.

Marker	Forward primer sequences (5'→3')	Reverse primer sequences (5'→3')	Conc µM	Amplicon size (bp)
HPCAL1	GCCTTCTTGGTGGTCCATAA	GCGACTACACCTGCAAGACA	0.3	178
ZNF282	CTATCTCCCCAGGTGACAGC	GGCAGTTGAAGCTCTTCTCG	0.3	503
PTPRS	CATAACCCACAAACCGCTCT	CCACCACACGTTACAGCATC	0.3	516
PMEPA1	CTGATGCCCGAGTTACTGCT	CTCTCTGCTCTCTGCACACG	0.3	579

2.3.7 Optimisation of the *Hha*I Restriction Reaction

Optimisation of the restriction reaction was performed under different conditions. High concentrations of *Hha*I enzyme such as 10 U and 2 U hindered complete digestion, evident by the presence of a band for the unmethylated artificial DNA template (481-bp portion of the pCR®2.1 TOPO® vector) even after restriction (results not shown). Increasing the incubation time from 37 C for 30 minutes to 37 C for 16 hours and 37 C for 1 hour did not increase the efficiency of restriction (results not shown). Heat inactivation at 80 C for 20 minutes showed similar results as heat inactivation at 65 C for 20 minutes. However, efficient digestion was obtained by using 0.2 U of enzyme, 37 C for 30 minutes for incubation and heat inactivation at 65 C for 20 minutes.

2.3.8 Body Fluid Identification using MSRE-PCR

2.3.8.1 HPCAL1 Gene-based Primer Set

The expected product size of the amplicon was 178 bp. Amplification was observed in saliva (Lane 2), semen (Lane 4) and vaginal fluid (Lane 5) but not in blood (Lane 3) (Figure 2.1). This implies that the target region is hypermethylated in saliva, semen and vaginal fluid and was not cleaved by *Hha*I; but it is hypomethylated in blood. The absence of a band at 481 bp in Lane 10 (restricted unmethylated artificial DNA template, size 481 bp) proves successful digestion. Thus, the primer set could be used as a potential tDMR to differentiate blood from other fluids.

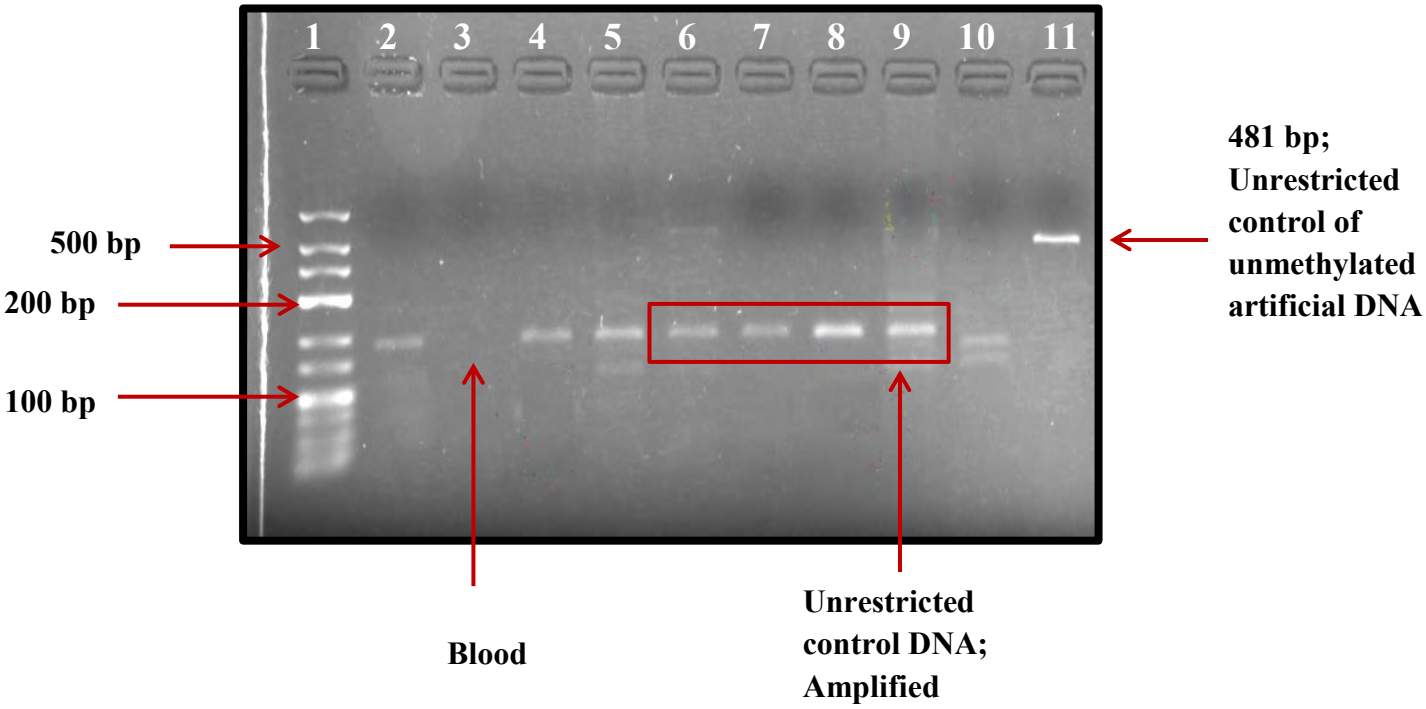

Figure 2.1: MSRE-PCR based methylation profiling of body fluids using the HPCAL1-based primer set. Lane 1 - GeneRuler™ Low Range DNA Ladder. Lane 2 – Saliva DNA (100 ng/µl); Lane 3; Blood DNA (100 ng/µl); Lane 4 – Semen DNA (100 ng/µl); Lane 5 – Vaginal fluid DNA (100 ng/µl). Lanes 6-9 – Unrestricted control of saliva, blood, semen and vaginal fluid respectively. Lane 10 – Restricted unmethylated artificial DNA template (481-bp portion of the pCR®2.1 TOPO® vector); Lane 11 – Unrestricted control of the unmethylated artificial DNA template.

2.7.8.2 PTPRS Gene-based Primer Set

The expected product size of the primers was 516 bp. Amplification was not observed in saliva (Lane 2), blood (Lane 3) and semen (Lane 4) but amplification was observed in vaginal fluid (Lane 5) (Figure 2.2). Clear distinct bands were obtained. This implies that the region is potentially hypermethylated in vaginal fluid and thus was not cleaved by *Hha*I; but it is potentially hypomethylated in saliva, blood and semen. Thus, the primer set could be used as a potential tDMR to differentiate vaginal fluid from other fluids.

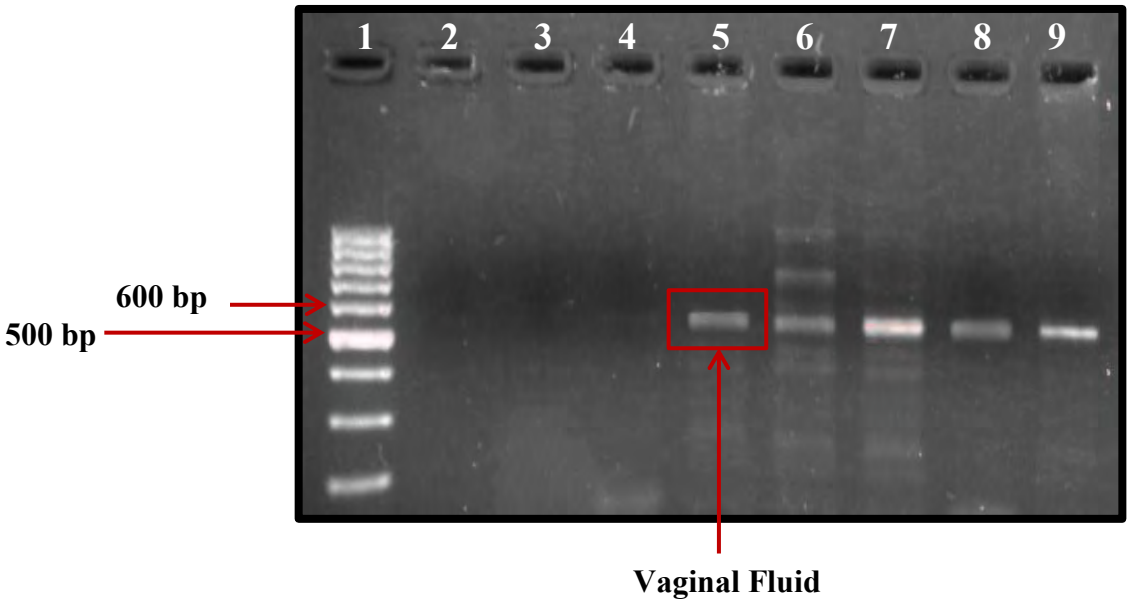

Figure 2.2: MSRE-PCR based methylation profiling of body fluids using the PTPRS-based primer set. Lane 1 - GeneRuler™ 100 bp DNA Ladder. Lane 2 – Saliva DNA (100 ng/µl); Lane 3; Blood DNA (100 ng/µl); Lane 4 – Semen DNA (100 ng/µl); Lane 5 – Vaginal fluid DNA (100 ng/µl). Lanes 6-9 – Unrestricted controls of saliva, blood, semen and vaginal fluid respectively.

2.3.8.3 ZNF282 Gene-based Primer Set

The expected product size of the primers was 503 bp. Clear bands were not observed at 503 bp. However, amplification was observed in saliva (Lane 2) and vaginal fluid (Lane 5) but not in blood (Lane 3) and semen (Lane 4) (Figure 2.3). Non-specific amplification was observed. Upon repetition of the reaction, inconsistent results were observed.

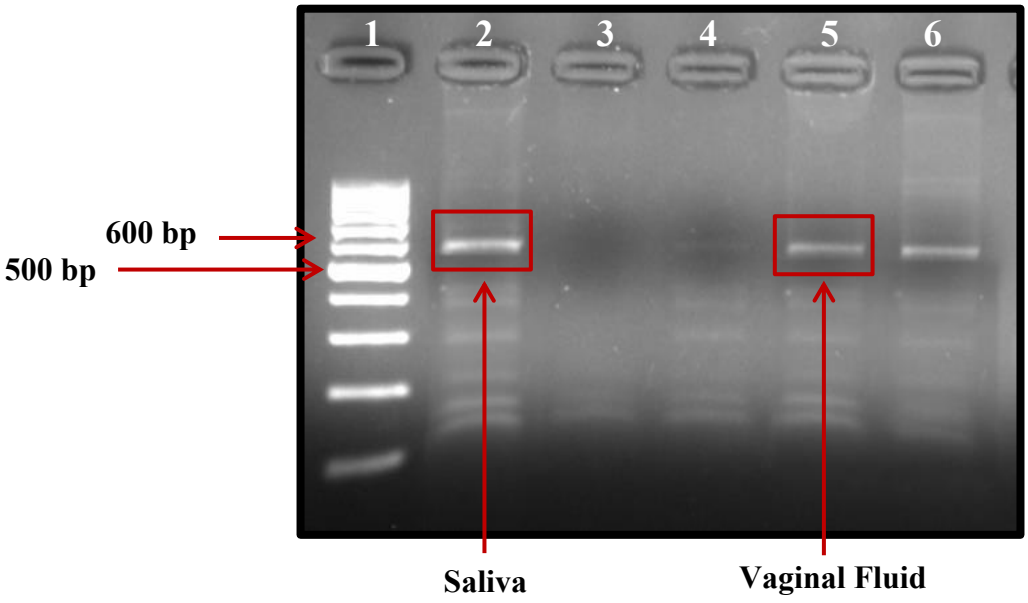

Figure 2.3: MSRE-PCR based methylation profiling of body fluids using the ZNF282-based primer set. Lane 1 - GeneRuler™ 100 bp DNA Ladder. Lane 2 - Saliva DNA (100 ng/µl); Lane 3; Blood DNA (100 ng/µl); Lane 4 – Semen DNA (100 ng/µl); Lane 5 – Vaginal fluid DNA (100 ng/µl). Lane 6 – Unrestricted control of saliva.

2.3.8.4 PMEPA1 Gene-based Primer Set

The expected product size was 579 bp. Amplification was not observed in saliva (Lane 2) and semen (Lane 4) but amplification was observed in blood (Lane 3) and vaginal fluid (Lane 5) (Figure 2.4). A single distinct band was not observed for this primer set at 579 bp. Faint bands were obtained for vaginal fluid. The results indicate potential hypermethylation in blood and vaginal fluid and potential hypomethylation in saliva and semen. However, reproducible results were not obtained for this primer set.

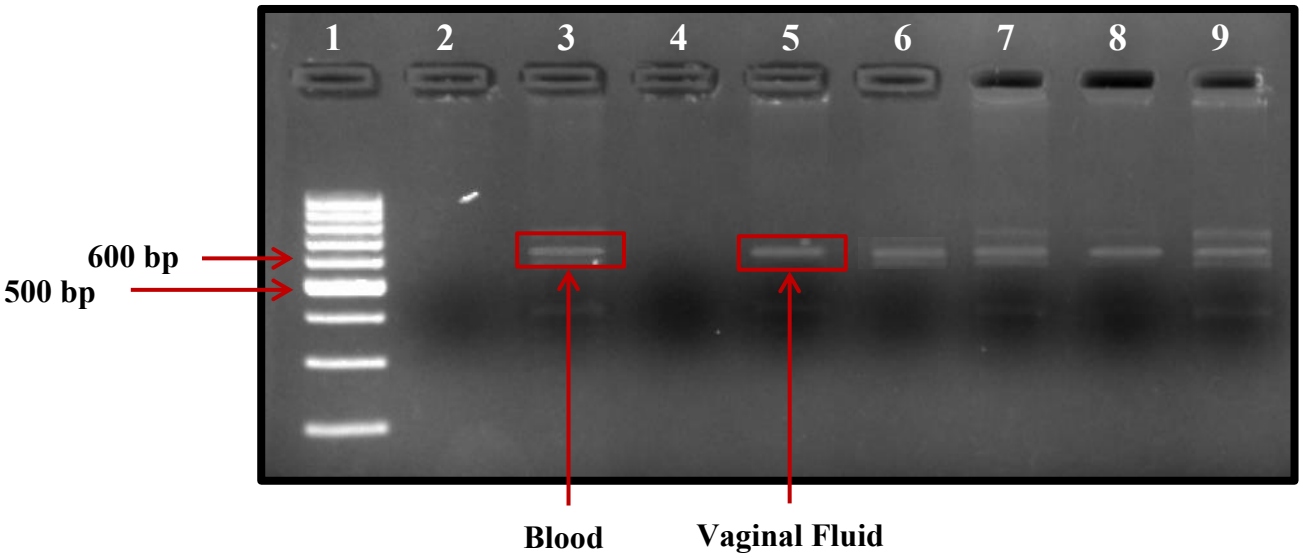

Figure 2.4: MSRE-PCR based methylation profiling of body fluids using the PMEPA1-based primer set. Lane 1 - GeneRuler™ 100 bp DNA Ladder. Lane 2 – Saliva DNA (100 ng/µl); Lane 3; Blood DNA (100 ng/µl); Lane 4 – Semen DNA (100 ng/µl); Lane 5 – Vaginal fluid DNA (100 ng/µl). Lanes 6-9 – Unrestricted controls of saliva, blood, semen and vaginal fluid respectively.

2.4 Discussion

The aim of the present study was to identify tissue-specific differentially methylated regions (tDMRs) in four forensically significant body fluids: namely saliva, blood, semen and vaginal fluid. tDMRs exhibit different DNA methylation profiles according to tissue or cell type (An *et al.,* 2013; Choi *et al.,* 2014; Lee *et al.,* 2012; Rakyan *et al.,* 2008). The basis behind identification of potential tDMRs was to design PCR primers targeting the differentially methylated regions, to differentiate between the four fluids by Methylation-Sensitive Restriction Enzyme PCR (MSRE-PCR). Body fluid identification is achieved by analysing the DNA methylation status of tDMRs.

Research studies have provided resounding evidence for a correlation between DNA methylation and gene expression (Kulis *et al.,* 2013; Schilling and Rehli, 2008). Differential expression of genes in different tissues may be regulated by DNA methylation. tDMRs have been associated with tissue-specific expression of genes and also show hypomethylation specifically in the tissue expressing those genes (De-Smet *et al.,* 1999; Slieker *et al.,* 2013; Weber *et al.,* 2007).

Based on this knowledge, the approach for the identification of tDMRs was entirely based on information acquired from bioinformatics databases on DNA methylation and gene expression in the human genome. Other methods employed to identify tDMRs include Methylation Specific Karyotyping (MSDK) (Hu *et al.,* 2006; Bloushtain-Qimron *et al.,* 2008), MethylCap Sequencing and MBD-isolated Genome Sequencing (MIGS) (Harris *et al.,* 2010), Whole Genome Bisulfite Sequencing (WGBS), pyrosequencing (Edwards *et al.,* 2010; Madi *et al.,* 2012; Park *et al.,* 2014a), base resolution methylomes genome sequencing (MethylC-Seq) (Schultz *et al.,* 2015) and epigenome-typing arrays such as Infinium HumanMethylation BeadChip (Dedeurwaerder *et al*., 2013). Major advantages of bioinformatics approaches compared to others are that it is relatively simple and low-cost since databases are freely available online, and very importantly it provides high resolution screening for tDMRs throughout the genome.

The initial steps of the study involved bioinformatics data mining to obtain a list of candidate genes that were over-expressed in surrogate tissues representing four body fluids of interest; namely saliva, blood, semen and vaginal fluid. Other than blood, gene expression data does not exist for the body fluids of interest. Therefore surrogate tissues that represented primary cell or fluid sources for the fluids of interest were selected. Salivary glands and

tongue were chosen for saliva; prostate gland for semen and vaginal and cervical tissues for vaginal fluid.

As genes showing under-expression or low expression could be due to factors other than DNA methylation, only over-expressed genes were specifically targeted. It is also not possible, in many databases, to differentiate between genes that were studied but demonstrated no expression and genes that were not analysed. It was assumed that the over-expression was more likely to be associated with low intragenic methylation levels compared to hypermethylation in other tissues.

Once genes that were over-expressed in the specific body fluids were identified (1833), genes that were found to be up-regulated in more than one of the body fluids were removed. This was to obtain genes that were specifically over-expressed in only one of the body fluids. Genes for which substantial data was unavailable for protein expression or association was not found between transcript expression and translation were eliminated. This was done because the aim was to target functional genes and their expression, and proof of a protein product reinforces that gene expression data is correct. After the exclusion of these genes, 1464 genes remained for further analysis.

Corresponding CGI Island (CGI) methylation information was obtained for these remaining 1464 genes. Obtaining methylation states in normal tissues enabled identification of the genes containing highly methylated CGIs. It is likely that this methylation is associated with low expression in normal tissues. Genes displaying CGIs with a mean and median methylation of above 75% were categorized as hypermethylated. Not only were promoter-associated CpGs analysed; but also those located across the entire gene. Hypermethylated CGIs which were unique to specific regions in the human genome and had a *Hha*I restriction site were selected for primer design.

As a result, a total of ten genes harboured CGIs which could be differentially methylated between the four forensically relevant tissues of interest and were selected for primer design. Methylation-Sensitive Restriction Enzyme PCR (MSRE-PCR) was used for the analysis of DNA methylation status of the targeted CGIs (potential tDMRs) in saliva, blood, semen and vaginal fluid.

Significant tissue-specific differential DNA methylation patterns between saliva, blood, semen and vaginal fluid were apparent upon methylation analysis using four gene-

based primer sets; *HPCAL1* (Hippocalcin-Like 1), *ZNF282* (Zinc Finger 282), *PTPRS* (Protein Tyrosine Phosphatase) and *PMEPA1* (Prostate Transmembrane Protein). In this study, two potential tDMRs showing differential DNA methylation between the four body fluids of interest were identified. These were blood-specific and vaginal fluid-specific markers.

The HPCAL1 gene-based CGI was initially presumed to show hypomethylation in vaginal fluid, as it was over-expressed in vaginal fluid. However from the results of MSRE-PCR, the marker displayed blood-specific hypomethylation, and hypermethylation, indicated by bright bands on the agarose gel, was evident in other fluids. The *HPCAL1* gene is found on chromosome two and encodes a protein which is a member of a neuron-specific calcium-binding protein family that is typically found in the brain and retina. It is assumed to be of significance for neuronal signalling in the central nervous system and also plays a role in calcium-dependent regulation of rhodopsin phosphorylation (Burgoyne, 2007). Blood-specific hypomethylation of HPCAL1 instead of vaginal fluid could be because of mix up of cervical tissue with uterine tissue of the wall of the cervix during gene expression studies. Uterine tissue is very rich in blood vessels so there is a possibility that over-expression of HPCAL1 was actually associated with blood. Over-expression yet hypermethylation in vaginal fluid could also be because of the presence of alternate promoters. DNA methylation at intragenic CpGs is believed to control use of alternative promoters and create diversity in regulation and expression of main transcripts (Ayoubi and Van de Ven, 1996; Kulis *et al*., 2013). When used in conjunction with other markers, the HPCAL1 may aid in forensic identification of blood. Frumkin *et al.* (2011) detected inconclusive methylation ratios between L76138/L26688 loci for blood and saliva. Similarly Lee *et al.* (2012) and An *et al.* (2013) detected similar methylation between blood and saliva for HOXA4 and DACT1 markers respectively. Ma *et al.* (2013a) showed promising results for blood identification. Six blood-specific differentially methylated fragments, termed BL 1-6 were identified. Two were potential blood-specific hypomethylation markers and four were potential hypermethylation markers. The *ELOVL2* marker was also studied to be a potential marker for blood by Zbiec-Piekarska *et al.* (2015).

The PTPRS gene-based CGI was presumed to show hypomethylation in saliva. However it displayed hypomethylation not only in saliva, but also in blood and semen. Successful differentiation of vaginal fluid from other tested fluids was achieved using the marker as it was found to be hypermethylated. The marker may be useful in future forensic

assays for identifying vaginal fluid since currently only a single tDMR; PFN3 exists for this purpose. Moreover, the PFN3 marker requires further validation (Choi *et al.,* 2014). Also, current methods using miRNA markers such as those tested by Zubakov *et al.* (2010) are incapable of differentiating vaginal fluid and saliva. The *PTPRS* gene is located on chromosome 19. PTPs are well-known as signalling molecules that partake in cellular processes such as growth, proliferation and mitosis. It is of great significance in oncogenic transformation; illnesses associated with *PTPRS* include ureteroceles and pineal gland cancer (Belinky *et al.,* 2015).

The PMEPA1 gene-based CGI showed hypomethylation in semen and saliva. Vaginal fluid and blood were hypermethylated. Further investigation of this marker is necessary for conclusive results since the results were not reproducible. Perhaps once the HPCAL1 and PTPRS markers are validated they may be used in a multiplex reaction with PMEPA1 for accurate differentiation between blood and vaginal fluid. The *PMEPA1* gene is situated on chromosome 20 and encodes a protein that harbours a Smad interacting motif. The gene is androgen and transforming growth factor *β*-inducible. Via interaction with Smad proteins, PMEPA1 inhibits the androgen receptor and transforming growth factor *β* signalling pathways. Importantly, it negatively regulates prostate cancer cell growth (Li *et al.,* 2008).

To our knowledge, this unique panel of genes targeted to design primers have not been previously used for identification of tDMRs or differentiation of forensically significant body fluids. Recent studies on the use of tDMRs for identification of body fluids have shown variable methylation levels in saliva, blood, semen and vaginal fluids (Igarashi *et al.,* 2008; Madi *et al.,* 2012; Schilling and Rehli, 2008; Smith *et al.,* 2014). The unique feature of the present study is the identification of potential methylation sites that demonstrate presence or absence of methylation, as opposed to varying degrees of methylation levels. This approach simplifies identification of body fluids considerably, since quantification of methylated CpGs is rendered unnecessary. Assessment of DNA methylation status using MSRE-PCR is less time-consuming and labour-intensive than bisulfite conversion (Melnikov *et al.,* 2005; Rein *et al.,* 1997). Samples of low quality and concentrations may be tested and because analysis is performed by the universal method of electrophoresis, the requirement for special training is circumvented. Furthermore, it is highly advantageous due to compatibility with current STR typing (An *et al.,* 2013).

The results of this study will complement previous forensics-based research in the development of novel tDMRs for body fluid identification. Since tDMR-based markers such as L81528, USP49, ZC3H2D and FGF7 have been developed and validated for semen (An *et al.,* 2013; Choi *et al.,* 2014; Madi *et al.,* 2012), there is much potential for development of tDMR markers for other fluids. As such, two potential hypermethylated tDMRs each for blood, saliva, semen and vaginal fluid were identified by Park *et al.* (2014a), six potential blood-specific markers were identified by Ma *et al.* (2013a) and now the present study has identified potential blood-specific and vaginal fluid-specific tDMRs. These results are promising and pave the way for application of differential DNA methylation for body fluid identification in forensic sciences.

Regarding the observed lack of sync between the expression dataset and the MSRE–PCR results, it can be said that the surrogate tissues used for each body fluid may be associated with a small amount of cell debris from surrounding body fluid that can induce cross-talk. So even though expression levels are validated and can be confirmed, and their association with hypermethylated CpGs and tDMRs can also be confirmed, the link between surrogate tissue and targeted bodily fluid may not be as direct as anticipated. This is clear from cervical tissue, where the rich blood supply to this tissue led to the discovery of blood-associated markers rather than the expected vaginal fluid markers. Despite these drawbacks, the method presented here still offers the most rapid approach to discovering new tissue-specific differential methylation markers.

It is evident that the analysis of DNA methylation and tDMR-based biomarkers in forensic sciences would be favourable. There is considerable potential for identification of body fluids and tissues by employing DNA methylation assays, but there are a few factors to consider first. When selecting a particular CpG site to analyse, it is vital to examine several samples to certify that the site does not demonstrate low levels of inter-individual variation. Also, it must be established that the site is not influenced by external stimuli or is age-dependent. Future research will involve quantification of methylation and validation of the potential tDMRs identified in this study by different profiling methods such as Methylation-Specific PCR (Beri *et al.,* 2007; Lee *et al.,* 2012) or the development of multiplex assays such as methylation SNaPshot (An *et al.,* 2013). Also, the identified potential tDMRs should be tested on a larger sample size comprising individuals of varying age groups. Identification of more tDMRs will contribute to make methylation-based detection of forensic body fluid more robust.

CHAPTER THREE

METHYLATION PROFILING OF SALIVA FROM THE DIVERSE SOUTH AFRICAN POPULATION USING tDMR-BASED MARKERS

ABSTRACT

Forensic DNA analysis is an indispensable tool which aids in the identification of suspected perpetrators and exoneration of innocent persons in criminal investigations, identification of the dead or missing, and identification of human remains in mass disasters both by natural causes and by criminal intent. While current methods for human identification include the use of short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs), certain disadvantages such as allele drop-ins and drop-outs pose a hindrance and too high numbers of markers (50-200) are required for powerful discrimination. Moreover, STR and SNP markers only provide inference of persons who are already known to the investigating authorities. DNA methylation-based markers show potential as an alternative tool for forensic human identification. Studies of genome-wide DNA methylation have discovered numerous differentially methylated regions (tDMRs) that differ in levels of methylation in various cell types and tissues. Furthermore while studies of genetic variation between human populations have shown that methylation levels of tDMRs differ between human populations; to our knowledge differentiation of ethnic groups by differential DNA methylation for forensic application has not been studied in South Africa. Therefore based on analysis of a forensically significant fluid, saliva, the present study aimed to determine if there is differential methylation of previously reported tDMR-based markers between ethnic groups of South Africa.

For DNA methylation-based screening of the diverse SA population, saliva was collected from 80 healthy individuals, male and female, belonging to the Black, Indian, White and Coloured ethnic groups. A multiplex MSRE-PCR assay was used to determine the methylation levels of four tDMRs of the *USP49, DACT1, L81528* and *PFN3* genes. The Coloured group displayed the highest methylation levels for all selected tDMRs and lowest methylation levels were evident for the Black ethnic group. Two tDMRs; DACT1 and L81528 displayed significant variations between the four ethnic groups; with $p = 0.02$ and $p = 0.03$, respectively. The significant differences in DNA methylation levels could potentially assist forensic analysts in future, not only to accurately identify saliva but also to narrow down the search of sample donors. This assists in identification of human remains in mass disasters, as well as linking donors to crimes or exonerating them with confidence. However, further validation of differential methylation among ethnic groups is required on a larger population size.

3.1 Introduction

Forensic analysis and identification of body fluid samples found at crime scenes is essential for crime scene reconstruction. The accurate identification of cell types and fluids are especially necessary when activities leading to deposition of the person's biological material are disputed. Biological evidence for assessment at crime scenes are mainly in the form of saliva traces, blood stains, semen, vaginal fluid, organ tissue such as brain or kidney as well as skin shreds (Sijen, 2014). However, conventional presumptive and confirmative tests for identifying these fluids, including identification of phosphatase in semen, haeme in blood and amylase in saliva lack sensitivity and specificity, result in false positives and lead to degradation of what little sample is collected (Haas *et al.,* 2009; Quarino *et al.,* 2005; Virkler and Lednev, 2009). It is necessary to explore novel strategies that aid in identification of body fluids. Core aspects in forensics are the necessity to develop markers not only for identification of body fluids, but also to gain inference to the origin and donor of the fluid. This is especially important to reveal identities of perpetrators or victims, which can direct conviction or exoneration, respectively.

Determining clues of the donor, in particular their racial/ethnic group is vastly researched by detection of genetic variants within and among human populations. Some researchers state that genetic variations account for a modest amount of differences between populations while others showed that individuals of different populations are more genetically similar than individuals from a single population (Rosenberg *et al.,* 2002; Tang *et al.,* 2005; Witherspoon *et al.,* 2007). Nevertheless, the prospect of genetic variation between populations is gaining momentum as it is advantageous to know if individuals within a population are more or less susceptible to diseases and who will most likely benefit from therapeutic interventions (Ku *et al.,* 2010; Lohmueller *et al.,* 2003; Risch *et al.,* 2002; Wilson *et al.,* 2001). It has recently become evident that mRNA and miRNA expression profiles, short tandem repeats (STRs), single nucleotide polymorphisms (SNPs) and DNA methylation may be reliably used to categorize humans into population groups (Bryc *et al.,* 2010; D'Amato *et al.,* 2008; Heyn *et al.,* 2013; Rawlings-Goss *et al.,* 2014; Rosenberg *et al.,* 2002).

MiRNA-based biomarkers have been developed for several diseases including diabetes, asthma, breast, prostate and lung cancers as well as leukaemia (Berezikov, *et al.,* 2005; Breuckner *et al.,* 2007; He *et al.,* 2015; Rawlings-Goss *et al.,* 2014; Taguchi, 2013). Studies have provided evidence of differing miRNA profiles between African Americans and

European Americans in lung cancers and in early stages of breast cancer (Heegaard *et al.,* 2012; Zhao *et al.,* 2010). Rawlings-Goss and colleagues (2014) found common SNPs in miRNA sequences between Hadza and Sandawe populations, and between Hadza and Kalahari San individuals of Africa. A large number of population differentiated miRNA (pd-miRNA) alleles between Africans and Asians were found, while only a few between Asians and Europeans were identified. A lower frequency of the hsa-mir-196a-2 *T* allele at SNP rs11614913 in Africans was found compared to non-Africans (Rawlings-Goss *et al.,* 2014). This particular variant is linked to a higher risk of oesophageal cancers in European males but lower risk for breast and gastric cancers in Chinese populations (Ryan *et al.,* 2010). Higher mortality rates have been reported for renal cell carcinoma and prostate and testicular-germ cancers in Africans compared to Europeans or Asians (Chornokur *et al.,* 2012; Sun *et al.,* 2011). Studies of Alzheimer's found that high expression of MiRNA-34a and MiRNA-125b is significant in brain synaptogenesis and amyloidogenesis (Cui *et al.,* 2010; Zhao *et al.,* 2013). This could explain onset and severity of Alzheimer's amongst African American, Caucasian American and Hispanics. Interestingly, African Americans and Hispanics have a higher incidence and severity of Alzheimer's when compared to Caucasians (Lukiw, 2013; Reitz *et al.,* 2013).

STRs which are regions of DNA that contain a variable number of tandemly repeated short sequence motifs (Kayser and Kniff, 2011) were examined by Rosenberg *et al*. (2002) to verify a relationship between patterns of genetic variation and different populations. The group was able to cluster 1056 individuals into Europeans and Asians west of the Himalayas, sub-Saharan Africans, inhabitants of New Guinea and Melanesia, Native Americans and East Asians. In South African research, D'Amato *et al.* (2008) genotyped nine Y-STR loci and identified high levels of genetic diversity in a Mixed Ancestry group (Coloureds) and minimal variation in a Xhosa population. The same group of researchers (D'Amato *et al.,* 2009) found high discriminatory capacity of DYS447, DYS437, DYS448, and 18 other Y-STR markers for forensic human identification in South Africa (D'amato *et al.,* 2009). By genotyping STR markers from the AmpFlSTR Identifiler kit against from eight Mexican Amer-Indian populations, Rangel-Villalobos and colleagues (2013) found that D18S51, D19S433, FGA, and D21S11 were the most informative loci to decipher genetic variation between various populations of the North Mexican Region and West Regions.

Variation of diets and climates as well as exposure to disease has resulted in high levels of diversity in genotypic and phenotypic characteristics of populations in Africa (Bryc *et al.,* 2010; Campbell and Tishkoff, 2008; Tishkoff *et al.,* 2010). In evaluating this diversity Tishkoff and colleagues (2010) genotyped SNPs and found high levels of genetic diversity within Africa and African American populations. Middle Easterns and Africans shared a high number of alleles and private alleles were found in between African Khoesan (SAK) San and Xun/Khwe groups of South Africa. Possibly due to genetic drift, Oceanic and Native American populations displayed highest proportions of genetic variation amongst their populations. Within Africa, two SAK populations clustered together and were most distant from other populations suggesting that these populations had the highest number of diverged genetic lineages (Tishkoff *et al.,* 2010). Bryc *et al.* (2010) also genotyped SNPs and found significant differences within West African populations. Africans were clearly distinguished from Europeans, and African Americans demonstrated highly variable levels of West African and European ancestry. A Fulani population was genetically distinct from Yoruba, Igbo and Brong groups, however, the Yoruba and Igbo groups were indistinguishable even when comparing over 300 000 markers.

While studies have shown that SNPs and miRNAs play significant roles in variation of gene expression between HapMap Lymphoblastoid cell lines (LCLs) obtained from Yorubans and Caucasians (Altshuler *et al.,* 2010; Huang *et al.,* 2011; Stranger *et al.,* 2007), recent research has also found significant differential DNA methylation between populations using LCLs (Bell *et al.,* 2011; Fraser *et al.,* 2012; Moen *et al.,* 2013). Fraser *et al.* (2012) and Moen *et al.* (2013) both evaluated European and Yoruban populations. Over 4300 differentially methylated CpG sites within and between populations were found by Fraser and colleagues (2012), and over 36 500 were identified by Moen *et al.* (2013). The *FLJ32569* and *STK39* (Serine Threonine Kinase 39) genes were differentially methylated between the populations whereas there was little variation in *AP4S1* (Adaptor-Related Protein Complex 4, Sigma 1 Subunit). A higher degree of methylation-SNP (mSNP) associations were found for the Yorubans than Europeans. Several mSNPs were associated with cardiovascular diseases, metabolic, autoimmune and neurological disorders. Five mSNPs linked with cholesterol, cardiovascular and coronary artery disease were found in *APOA5* (Apolipoprotein A-V) and a lower risk allele frequency was observed in Europeans, however mSNPs associated with rheumatoid arthritis, multiple sclerosis and prostate cancer demonstrated a higher risk allele frequency in the European LCLs as opposed to Yorubans (Cullen, 2000; Moen *et al.,* 2013).

The *HLA-DPA1* locus which is correlated with chronic hepatitis B (HBV) infection is abundant in Africans or Asians as they have higher frequency of diseases. Heyn and colleagues (2013) and Jones (2012) found that hypermethylation of the *HLA-DPA1* promoter conjoined with intragenic hypomethylation is associated with decreased gene expression. This alters cell surface receptor presentation for HBV-binding and thus, higher risk of infection in these populations.

In prostate cancer research, Enokida and colleagues (2005) demonstrated that methylation levels of *GSTP1* (π-class Glutathione S-transferase) were significantly higher in African Americans when compared to Caucasians and Asians. Kwabi-Addo *et al.* (2010) examined the methylation status of prostate cancer-related genes; *GSTP1, AR, RARβ2, SPARC, TIMP3,* and *NKX2-5* in African American and Caucasian American populations. The study also found higher methylation levels of the genes in African Americans compared to Caucasians, which may potentially contribute to the racial differences that are observed in prostate cancer pathogenesis.

Terry and colleagues (2008) and Zhang *et al.* (2011a) measured DNA methylation of peripheral blood samples between Blacks, Whites and Hispanics and found that Whites displayed highest levels of methylation, followed by Hispanics, while Blacks displayed lowest methylation levels.

At present, body fluid analysis is frequently employed in identification of individuals following vehicle accidents, victims and criminals of armed conflicts, situations of criminal acts and disasters (Goodwin *et al.,* 2009). For this, body fluid-specific mRNA and miRNA markers were identified (Juusola and Ballantyne, 2007; Nussbaumer *et al.,* 2006; Park *et al.,* 2014b), but RNA is highly unstable and susceptible to degradation, fluctuating pH and UV light (Liu *et al.,* 2002; Park *et al.,* 2014a). After many years of forensic STR- and SNP-typing, analysts found that PCR amplicon sizes of STRs included in kits are too large which lead to genotyping failures and degraded samples (common in forensic laboratories) cannot be genotyped. Stutter artefacts due to strand slippage in PCR compromise analysis and not every STR locus varies between individuals (Butler, 2012; Gill, 2002; Kayser and Kniff, 2011; Parson and Steinlechner, 2001). Whilst SNPs have been considered as potential replacements of STRs due to ability to analyse degraded samples and generation of smaller PCR amplicons (Dixon *et al.,* 2005; Sanchez *et al.,* 2006), allele drop-ins and drop-outs pose a hindrance and too many SNP markers are required for powerful discrimination (Amorim

and Periera, 2005; Gill *et al.,* 2004). Very importantly, STRs and SNPs cannot be used to identify those individuals that are not known to investigating authorities (Butler *et al.,* 2007; Kayser and Kniff, 2011).

In contrast, profiles of tissue-specific differentially methylated regions (tDMRs) show potential for the identification of unknown individuals since phenotypic differences between twins (Fraga *et al.,* 2005), clues to donors ethnicity and diets have been estimated by tDMR profiling (Terry *et al.,* 2008; Lam *et al.,* 2012). Analysts are able to confidently estimate age of fluids, time of fluid deposition as well as differentiation of sex (An *et al.,* 2013; Bockland *et al.,* 2011). tDMRs of *EDARADD* (EDAR Associated Death Domain) and *ELOVL2* (Fatty Acid Elongase II) genes from blood and teeth were validated to accurately estimate age of donors (Bekeart *et al.,* 2015; Garagnani *et al.,* 2012).

Analysis of tDMRs is cost-effective, reproducible, robust, sensitive and specific, and only requires low amounts of template DNA (Choi *et al.,* 2014; Frumkin *et al.,* 2011; Sijen, 2014). tDMR-based markers are not influenced by fluctuating temperatures, varying methods of collection and handling, transport and storage. Multiplexing enables rapid analysis of large amounts of samples within a short time period (An *et al.,* 2013; Choi *et al.,* 2014; Frumkin *et al.,* 2011; Park *et al.,* 2014a). Since it is evident that epigenetic variation does account for differences in human populations, the aim of the present study was to determine if the diverse South African population shows differences in DNA methylation. MSRE-PCR and capillary electrophoresis was used due to rapidity, ease of use and can easily be implemented in real casework due to compatibility with STR-typing methods (An *et al.,* 2013; Choi *et al.,* 2014; Frumkin *et al.,* 2011). Unlike bisulfite treatment, MSRE-PCR enables analysis of heterogenous mixtures as well as multiplexing; thereby reducing the number of amplifications and overall costs. In addition to sequence specificity, fluorescent primers enable closed-tube detection of numerous targets within a single reaction which eliminates chances of cross-contamination. The only factor be cautious of is efficient digestion of the template DNA, therefore the use of an unmethylated control is recommended (Frumkin *et al.,* 2011). Since PCR is followed by the universally well-known method of capillary electrophoresis, training of personnel is circumvented which saves time and reduces human error.

For the present study, the methylation profile of saliva obtained from four ethnic groups in South Africa was evaluated, to determine if there is differential methylation of targeted tDMRs among ethnic groups. Significant differences in DNA methylation among four ethnic groups could assist forensic analysts in future, to narrow down their search for sample donors.

3.2 Materials and Methods

3.2.1 Sample collection

Approximately 5 mL of saliva was collected in sterile 15 mL tubes from 80 healthy volunteers (with no chronic disorders). Participants included males and females, from four ethnicities (Blacks, Indians, Whites and Coloureds). The saliva samples were stored frozen at -20 C until DNA extraction. To ensure anonymity, each donor was designated a number code. The collection tube was numbered and labelled according to gender and ethnicity of donor. The ages of participants used for the study is presented in Table 3.1. The donors provided signed informed consent after the goals of the study were described (Appendix B). Additionally, each participant was requested to provide written disclosure of their ethnicity/race, age and gender in the form of a questionnaire (Appendix B). The study was conducted according to the methods specified by the Biomedical Research Ethics Committee of the University of Kwa-Zulu Natal (Westville Campus).

3.2.2 DNA Extraction and Quantification

DNA was extracted from 200 L aliquots of saliva using a Quick-g DNA MiniPrep Kit (ZymoResearch) according to the manufacturer's instructions. Extracted DNA was quantified using a spectrophotometer (Nanodrop ND-2000, Thermo Fisher Scientific Inc, Waltham, MA, USA). The concentration of saliva DNA of each participant is listed in Table 1, Appendix B. Following quantification, each saliva sample was stored frozen at -20˚C until further use.

Table 3.1: Demographics for study participants.

Ethnic group	Gender	Number of Participants (n)	Age Range (years)
Black	Male	11	Below 30 n=6
			30-50 n=5
	Female	10	Below 30 n=7
			30-50 n=3
Indian	Male	9	30-50 n=8
			Above 50 n=1
	Female	10	Below 30 n=10
White	Male	12	Below 30 n=5
			30-50 n=6
			Above 50 n=1
	Female	8	Below 30 n=5
			30-50 n=3
Coloured	Male	10	30-50 n=10
	Female	10	Below 30 n=4
			30-50 n=5
			Above 50 n=1

Black participants can belong to any one, or be a mixture of the four major ethnic groups living in South Africa including the Nguni, Sotho, Shangaan-Tsonga and Venda groups. The Black participants were part of the Nguni languages (Zulu, Xhosa, Swati and Ndebele). All Indian participants had Indian grandparents and parents. Whites refer to individuals having only European ancestry. Coloureds of South Africa are said to be the outcome of admixture which is mating between individuals from reproductively isolated ancestral populations, therefore having mixed ancestry.

3.2.3 Selection of tDMR Markers and Primer Design for PCR

For DNA methylation-based differentiation of four ethnic groups, four tDMRs for *DACT1, USP49, L81528* and *PFN3* genes were selected as markers. tDMRs for *DACT1, USP49* and *PFN3* genes have been reported to display varying amplification in saliva (An *et al.,* 2013; Choi *et al.,* 2014; Lee *et al.,* 2012) and L81528 shows no amplification in saliva as it was reported to be a semen-specific hypermethylation marker (Choi *et al.,* 2014; Frumkin *et al.,* 2011). To ensure that the markers would be applicable in Methylation-Sensitive Restriction Enzyme PCR (MSRE-PCR), primers developed by Choi *et al*. (2014) that flank the *Hha*I recognition sites (GCGC) of the four tDMRs were used (Table 3.2). Amplicon sizes were smaller than 150 bp and the forward or reverse primer were labelled with FAM

(Fluorescein), a fluorescent dye (Choi *et al.,* 2014). Amelogenin was used as a sex-typing marker. Primers for the amplification of an amplification control, was designed and synthesised. The artificial DNA templates of the amplification control for PCR success was obtained by PCR amplification of the 481 bp portion of the pCR®2.1 TOPO® vector (Invitrogen, Carlsbad, CA, USA) (An *et al.,* 2013; Choi *et al*., 2014). The USP49, DACT1, L81528, PFN3 and Amelogenin markers were labelled with FAM, and the amplification control was labelled with NED.

Table 3.2: Multiplex PCR primers for amplification of tDMRs. Adapted from Choi *et al.* (2014) with minor modifications.

Marker	Forward primer sequences (5'→3')	Reverse primer sequences (5'→3')	Conc µM	Amplicon size (bp)
USP49	GTAGCAGGTGTTGCCCAGGTT	FAM-CCCTCCCTACCTCACGCAGA	1.0	107
DACT1	FAM-CACTCCTCCCCTGCTGTCTA	GATAAACTGGCTTGACCA	0.70	118
L81528	FAM-CTTCTGGGGCGACTACCTG	AGTCAGCCTCATCCACACTGA	0.40	128
PFN3	CCTGGCAGCCTCTAGACTCA	FAM-GGGCCAAATAAACTGTGACC	0.20	137
Amelogenin	CCCCTTTGAAGTGGTACCAGAG	FAM-GCATGCCTAATATTTTCAGGGAATAA	0.25	81; 84
Amplification Control	CTGTTCTTCTAGTGTAGCCGTAGTT	NED-CAACCCGGTAAGACACGACT	0.15	131

3.2.4 *Hha*I Enzyme Restriction Reaction

Multiplex PCR was developed to amplify *Hha*I recognition sites (GCGC) of the tDMRs of *USP49, L81528, DACT1* and *PFN3* genes. Amelogenin is devoid of a restriction site. Approximately 100 ng of DNA from each participant was digested with *Hha*I in a 10 L reaction containing 1 L of CutSmart Buffer (New England Biolabs, Ipswitch, MA, USA) and 0.2 U of the *Hha*I restriction enzyme (New England Biolabs, Ipswitch, MA, USA). To ensure complete digestion, the unmethylated artificial DNA template (481 bp portion of the pCR®2.1 TOPO® vector) was restricted. Additionally, each sample was run with an unrestricted control reaction. Herein the enzyme was not added therefore the reaction simply contained 100 ng of DNA and 1 L of CutSmart Buffer (New England Biolabs, Ipswitch, MA, USA). All digestions were incubated at 37 C for 30 minutes with subsequent heat inactivation at 65 C for 20 minutes on a BIORAD T100™ Thermal Cycler (An *et al*., 2013; Choi *et al*., 2014).

3.2.5 Multiplex MSRE-PCR

Multiplex PCR was carried out in a 96 well plate, in 20 L reaction volumes that contained 10 L of enzyme-digested DNA and undigested DNA controls, 2.0 U of AmpliTaq Gold DNA Polymerase, 1 L of Gold ST*R 10× Buffer (Promega, Madison, WI, USA) and the stated primer concentrations (Table 3.2). PCR was conducted under the following HotStart conditions: 95˚C for 11 minutes, 28 cycles of 94˚C for 20 seconds, 59˚C for 60 seconds, and 72˚C for 30 seconds and final extension at 60˚C for 60 minutes on a BIORAD T100™ Thermal Cycler (Choi *et al.,* 2014).

To prepare samples for analysis, a mix containing 1 L of amplification products, 20 L of Hi-Di formamide and 0.2 L of GeneScan™-500 LIZ® internal lane size standard was denatured at 95˚C for 5 minutes followed immediately by 3 minutes on ice (An *et al.,* 2013; Choi *et al*., 2014). These preparation steps were essential as the amplification products were run on an ABI 310 genetic analyser (Applied Biosystems) (Figure 3.1) according to manufacturer's instructions.

The electropherograms were analysed using GeneMapper ID Software 5 (Applied Bioystems). The threshold for a positive peak was set to 100 relative fluorescent units (rfu).

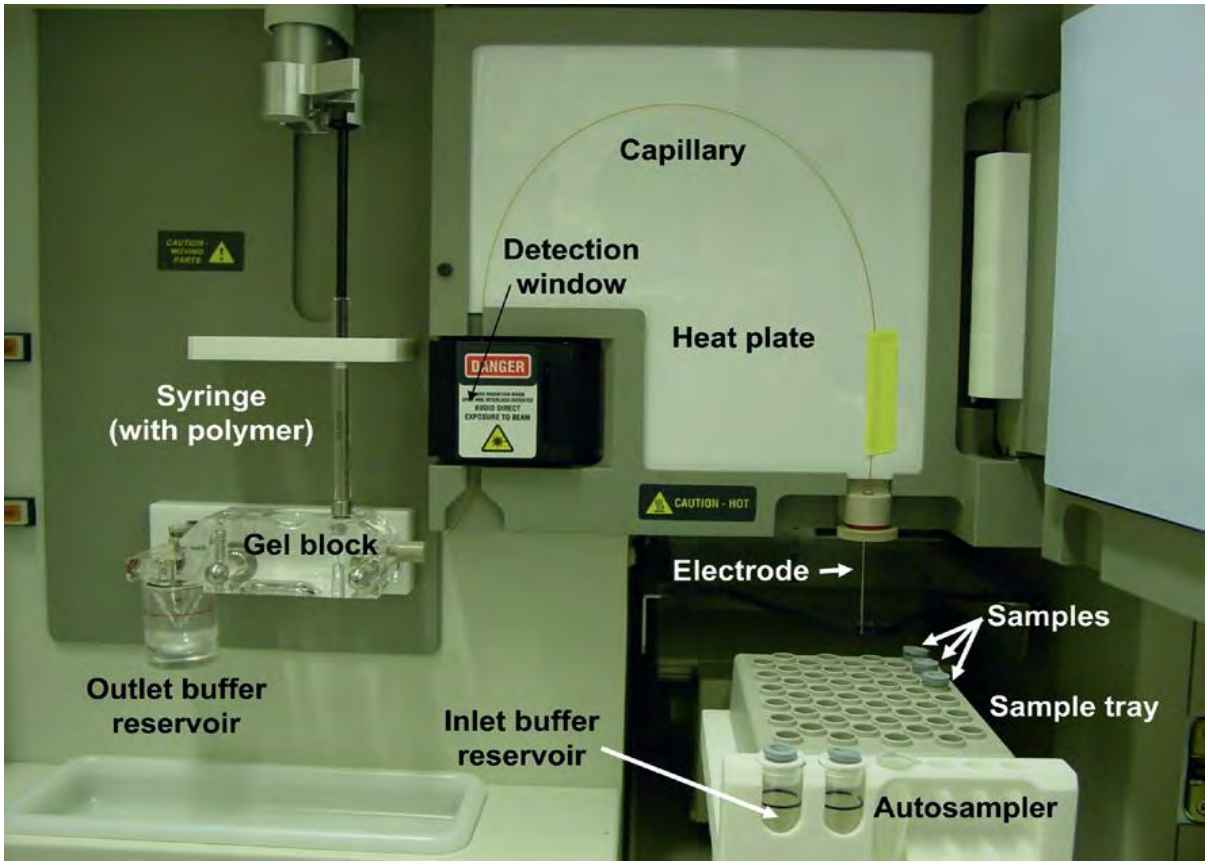

Figure 3.1: An internal view of the ABI 310 Genetic Analyser. A single capillary is located between the gel block and the inlet electrode. The capillary is filled with polymer solution through the gel block. A heat plate is used to raise the temperature of the capillary to a specified temperature. Samples are placed in an autosampler tray that moves up and down to insert the sample onto the capillary and electrode for the injection process (Butler, 2012; Moretti *et al.,* 2001).

3.2.6 Statistical Analyses

To determine whether the tDMRs showed a significant difference in DNA methylation profiles of saliva between the four ethnic groups, ANOVA (Analysis of Variance) was carried out using SAS Software Version 9.2 (SAS Institute Inc., Cary, NC, USA). Comparison of methylation status of each tDMR, for each ethnic group, was performed by calculating the peak height ratios/tDMRs. Differences were statistically significant when *p* values were less than 0.05. To test for normality of data, Shapiro-Wilk, Kolmorogorov-Smirnov Cramer-von Mises and Anderson-Darling tests were carried out. All tests confirmed that the data followed a normal distribution. The decision to utilize analysis of variance (ANOVA) to analyse the data was therefore justified. The methylation status is shown as peak height ratios of each of four tDMRs to Amelogenin for each sample in MSRE-PCR. For men, the peak height of Amelogenin was calculated by the sum of two peak heights (An *et al.,* 2013).

3.3 Results

3.3.1 Individual DNA Methylation Profiling using MSRE-PCR

MSRE-PCR was conducted on saliva DNA from 80 individuals. From a total of 80, amplification was observed only for 42 samples; comprising of 9 Blacks, 10 Indians, 15 Whites and 8 Coloureds. The non-amplified samples which did not show any amplification in the restricted as well as unrestricted controls were eliminated from further analysis.

Figure 3.2 represents an example of the multiplex MSRE-PCR results of *Hha*I-treated saliva DNA. The figure depicts a restricted profile (A) followed by a control unrestricted profile (B). The sequence of Amelogenin, the sex-typing marker, is devoid of the *Hha*I site and will thus generate peaks following MSRE-PCR for every sample. The electropherograms for 42 samples are presented in Appendix B.

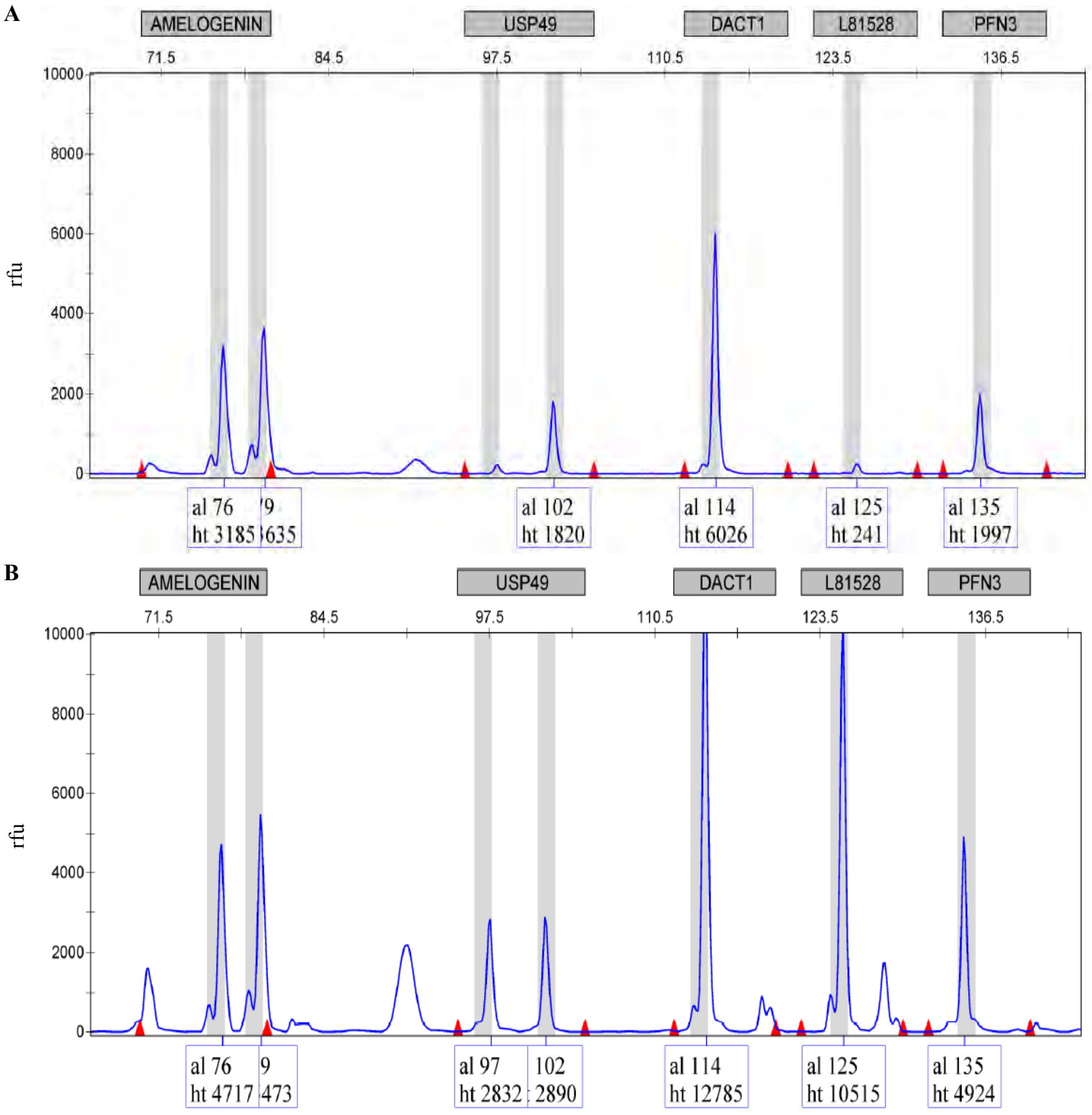

Figure 3.2: Methylation profile of saliva obtained from a Black Male by multiplex MSRE-PCR. Samples were run on an Applied Biosystems 310 Genetic Analyser and analysed by GeneMapper ID Software 5. A – Restricted DNA. B – Unrestricted DNA (Control). The grey bars represent bins assigned to the peaks. The allele/loci calls and peak heights are shown at the bottom of each plot.

3.3.2 Calculation and Comparison of Methylation Levels of Selected tDMRs between Four Ethnic Groups

Methylation status of each tDMR was calculated as a peak height ratio of each sample. The peak height ratios for each individual in each ethnic group are listed in Table 3.3 – Table 3.6 Peak height ratios revealed a distinct variation of methylation levels for all four tDMRs between the four ethnic groups. The DACT1 tDMR displayed the highest methylation status between all four tDMRs. The methylation status of the tDMRs in saliva of the Coloured population was highest between the four ethnic groups (Table 3.6), whereas the lowest methylation status of the tDMRs was found in the Black population group (Table 3.3). A graphical representation of the mean peak height ratio for each tDMR in each ethnic group is shown in Figure 3.3.

Table 3.3: The peak height ratio (%) for each marker depicting the methylation status in every individual in the Black ethnic group.

Sample Number	Amelogenin	PHR USP49 (%)	PHR DACT1 (%)	PHR L81528 (%)	PHR PFN3 (%)
3	1706	0	20.57	0	0
4	6822	26.68	88.33	3.35	29.27
5	3906	0	11.11	0	0
11	7039	20.88	130.39	25.87	39.8
12	6029	3.67	21.06	0	0
14	3325	14.41	151.16	7.64	61.62
15	6652	6.98	69.14	8.75	18.08
17	5474	3.54	27.3	0	0
18	1105	0	0	0	0
Average/Mean	4673.11	8.46	57.67	5.07	16.53

Table 3.4: The peak height ratio (%) for each marker depicting the methylation status in every individual in the Indian ethnic group.

Sample Number	Amelogenin	PHR USP49 (%)	PHR DACT1 (%)	PHR L81528 (%)	PHR PFN3 (%)
24	2279	0	0	0	0
29	4497	12.63	77.21	2.58	18.61
30	13154	20.83	82.81	24.22	18.83
31	4727	2.34	13.81	0	0
31	8887	11.63	65.01	1.64	3.5
33	743	0	76.04	0	21.93
34	9819	13.14	78.81	1.92	13.71
37	8281	26.84	105.34	51.72	36.01
38	11674	27.44	112.88	40.11	35.99
40	742	0	54.58	0	19.14
Average/Mean	6480.3	11.49	66.65	12.22	16.77

Table 3.5: The peak height ratio (%) for each marker depicting the methylation status in every individual in the White ethnic group.

Sample Number	Amelogenin	PHR USP49 (%)	PHR DACT1 (%)	PHR L81528 (%)	PHR PFN3 (%)
41	12437	28.25	84.5	9.01	24.41
42	432	0	56.25	0	36.11
43	2594	17.66	111.6	17.81	45.41
44	5036	24.5	90.48	24.46	28.69
45	5428	5.6	43.28	0	3.78
46	10050	19.4	89.29	22.03	29.52
47	8687	15.72	100.71	15.16	34.76
48	9084	7.38	37.12	3.76	7.55
49	7493	20.47	110.09	59.76	36.02
53	5223	31.05	117.9	62.44	37.81
56	3879	13.22	90.92	8.89	19.93
57	6796	0	3.39	0	0
58	8413	31.58	140.68	13.65	49.64
59	2677	9.82	127.53	28.05	29.59
60	8240	20.13	113.86	7.56	32.68
Average/Mean	6431.27	16.32	87.84	18.17	27.73

Table 3.6: The peak height ratio (%) for each marker depicting the methylation status in every individual in the Coloured ethnic group.

Sample Number	Amelogenin	PHR USP49 (%)	PHR DACT1 (%)	PHR L81528 (%)	PHR PFN3 (%)
61	3635	29.38	134.97	42.28	62.17
64	10280	25.27	120.56	43.02	44
71	7375	18.87	92.5	5.44	14.54
72	5869	29.9	127.3	55.14	44.73
74	4678	7.2	54.44	4.15	6.54
75	6096	35.16	136.63	46.7	47.55
76	6101	18.95	122.41	39.91	45.44
77	12635	17.15	86.88	29.29	26.78
Average/Mean	7083.63	22.74	109.46	33.24	36.47

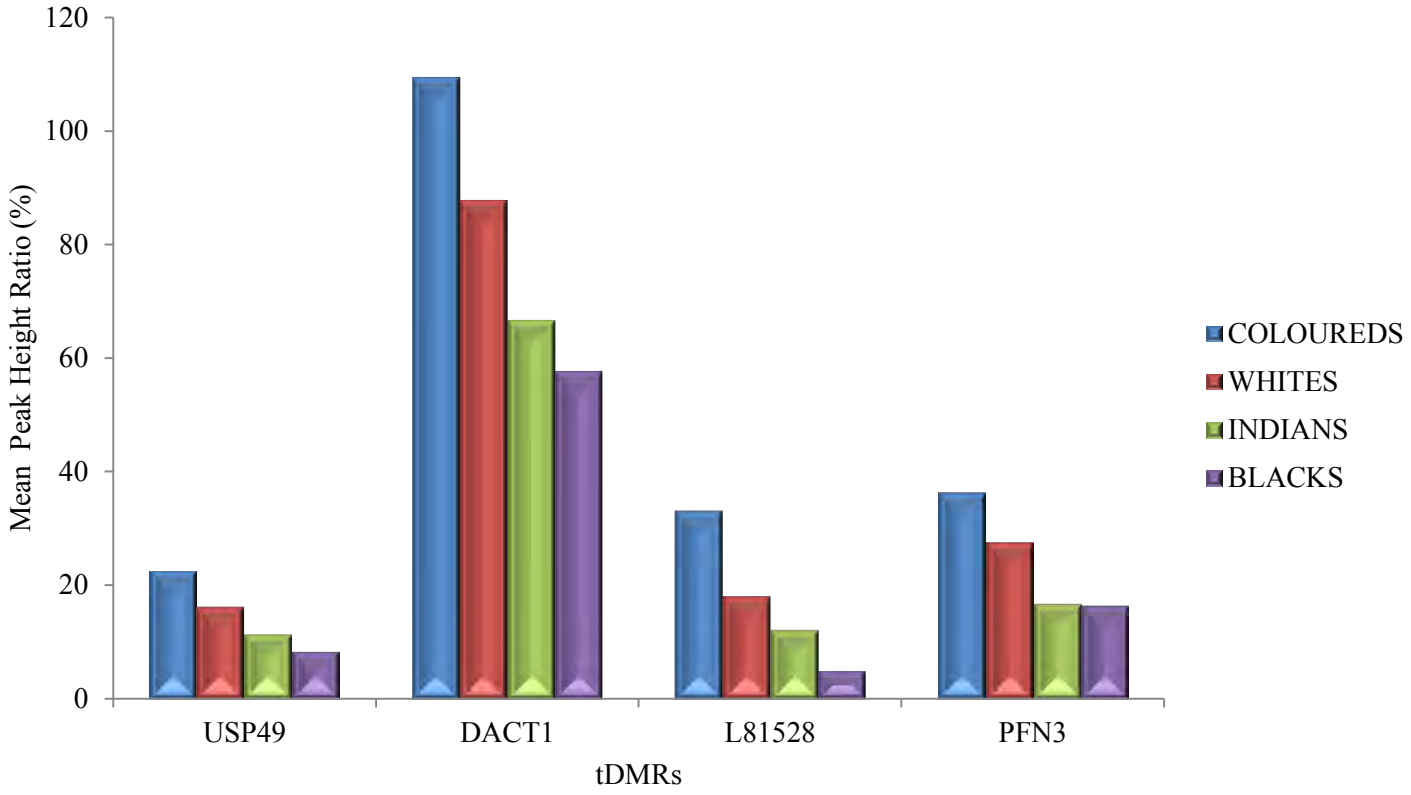

Figure 3.3: Methylation status of the USP49, DACT1, L81528 and PFN3 tDMRs in saliva DNA of all four ethnic groups. Methylation status is indicated as (mean) peak height ratios of each of four tDMRs to Amelogenin for each sample in MSRE-PCR. For men, peak height of Amelogenin was calculated by the sum of two peak heights. Blacks, n=9; Indians, n=10; Whites, n=15; Coloureds, n=8. All four tDMRs are display highest methylation in the Coloured population and lowest methylation in the Black population.

3.3.3 Statistical Analyses

ANOVA was performed to determine if there was any significant difference in methylation status of the four tDMRs in saliva of four ethnic groups. Methylation profiles of each tDMR was analysed for each race. Effect of age and gender on methylation status was also analysed. All ANOVA results followed a normal distribution. Normality plots are shown in Appendix B.

3.3.3.1 USP49 tDMR-based Marker

The statistical analysis of the methylation status of the USP49 tDMR demonstrated that the marker did show slight variation between the four ethnic groups (p = 0.05) (Table 3.7). The Coloured group displayed the highest methylation level and the Black groups displayed minimum methylation levels (Figure 3.4).

Table 3.7: Analysis of variance to test the effect of race and age on methylation status of the USP49 tDMR.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Race	3	869.71	289.90	2.82	0.05
Age	2	388.04	194.02	1.89	0.17
Race*Age	5	355.93	71.18	0.69	0.63

The age of individuals did not have any significant effect on the methylation status of USP49, however an increased methylation level was observed in subjects between 30-50 years of age when compared to subjects below 30 years of age (Figure 3.5). Interaction of race and age did not have any significant effect (data not shown). There was no significant difference observed in methylation status between genders (data not shown).

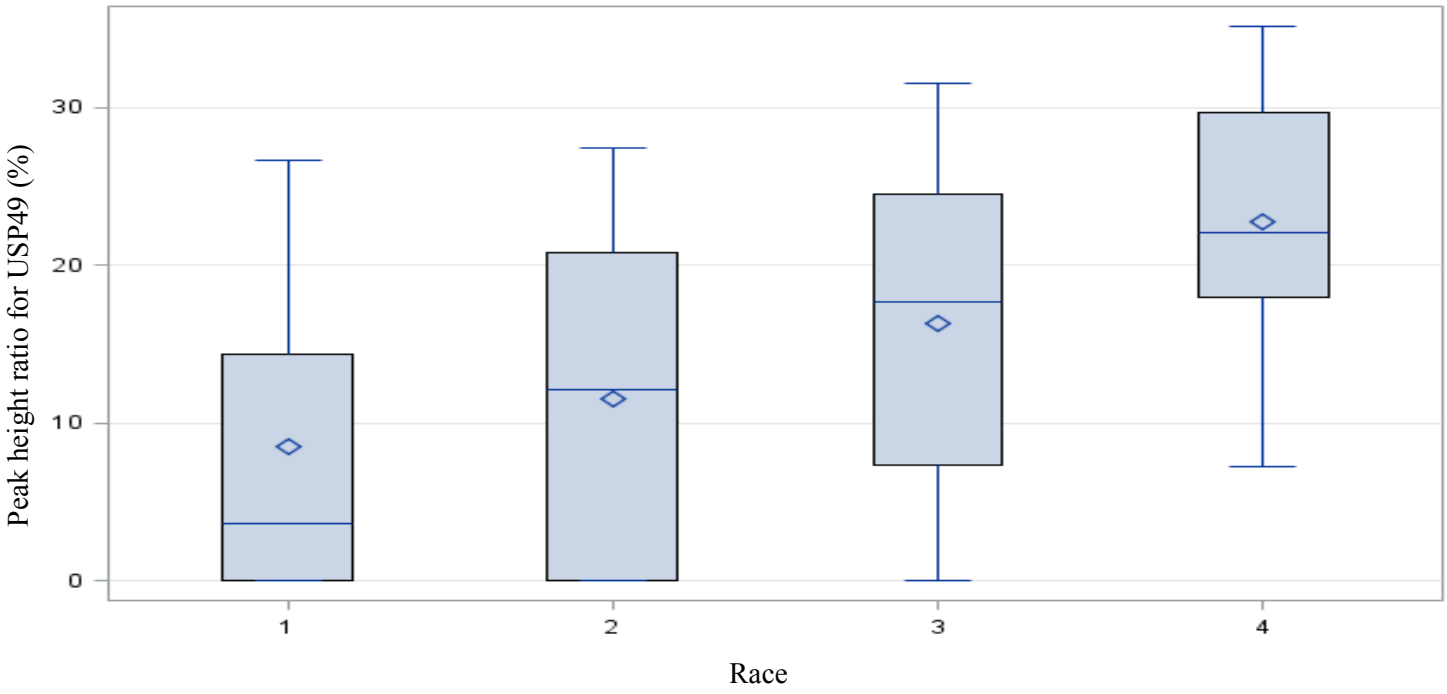

Figure 3.4: Box-Whisper plot depicting the effect of ethnicity on methylation status of the USP49 tDMR across four ethnic groups. Races 1-4 represent Blacks, Indians, Whites and Coloureds respectively.

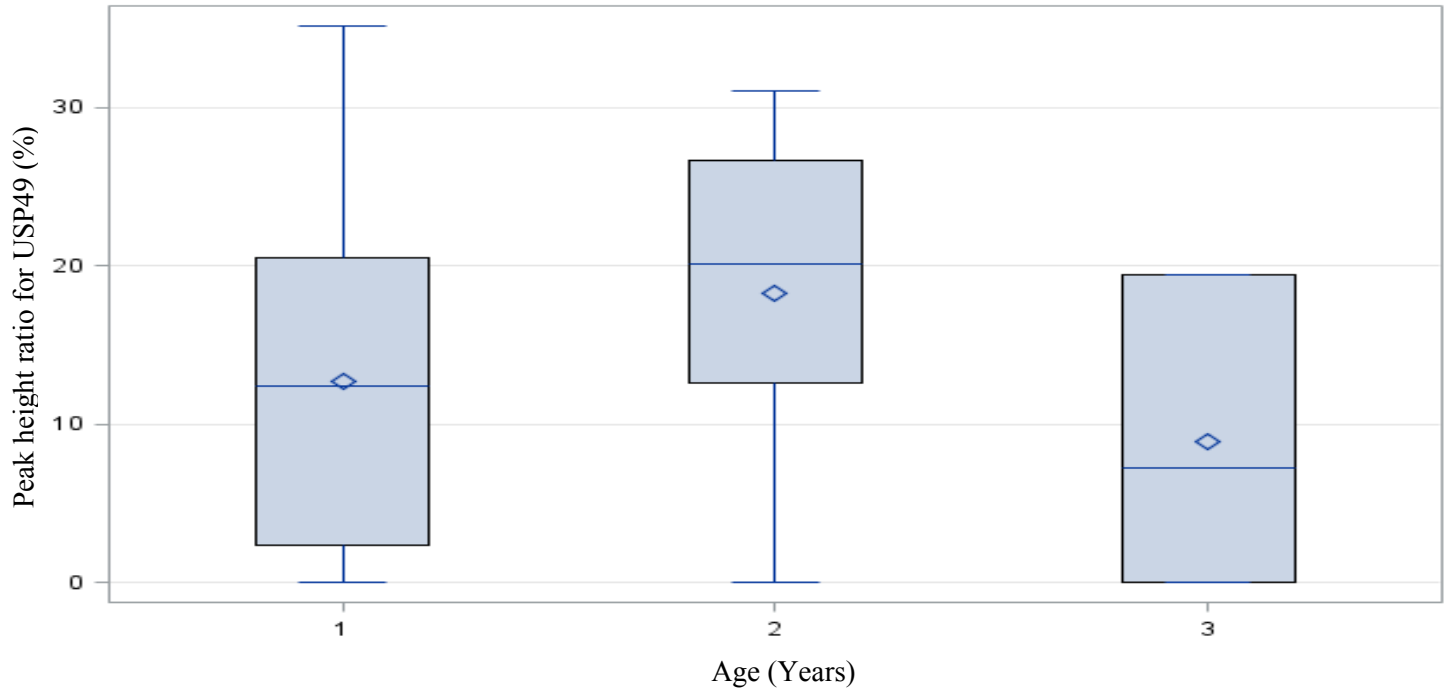

Figure 3.5: Distribution of methylation levels for the USP49 tDMR across three age groups. 1 - <30 years; 2 – 30-50 years; 3 - >50 years.

3.3.3.2 DACT1 tDMR-based Marker

The statistical analysis of the methylation status of the DACT1 tDMR demonstrated that the marker showed significant variation between the four ethnic groups ($p = 0.02$) (Table 3.8). The Coloured group displayed the highest methylation level and the Black group displayed minimum methylation levels (Figure 3.6).

Table 3.8: Analysis of variance to test the effect of race and age on methylation status of the DACT1 tDMR.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Race	3	17733.31	5911.10	3.89	0.02
Age	2	5952.50	2976.23	1.96	0.13
Race*Age	5	8411.68	1682.33	1.11	0.38

Distribution of methylation levels for the DACT1 tDMR across all four ethnic groups and ages are represented in Figures 3.6 and Figure 3.7 respectively. Even though a significant variation in methylation levels was observed across all races, the largest difference of methylation was observed between Coloureds and Blacks (Figure 3.6), and a consistent level of methylation of the DACT1 tDMR in subjects <30 years and between 30-50 years of age was observed (Figure 3.7). Interaction of race and age, as well as gender, did not have an effect (data not shown).

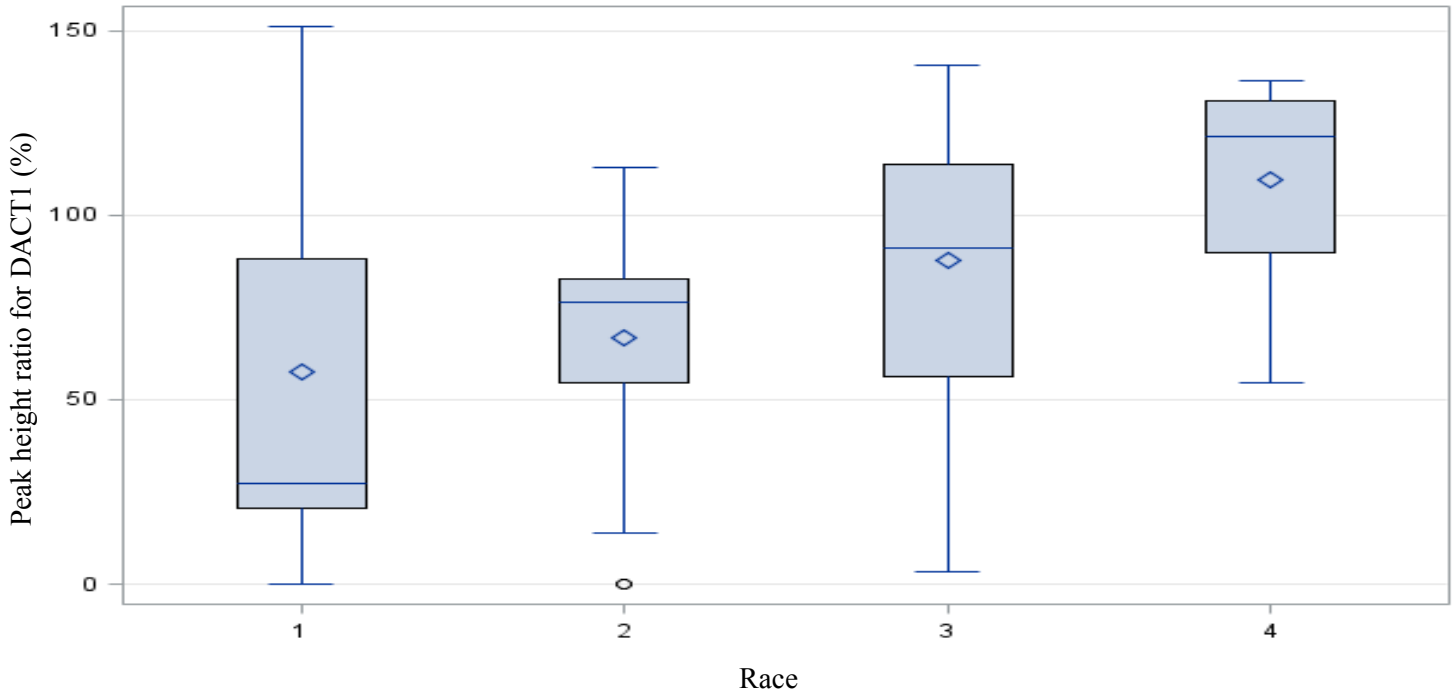

Figure 3.6: Box-Whisper plot depicting the effect of ethnicity on methylation status of the DACT1 tDMR across four ethnic groups. Races 1-4 represent Blacks, Indians, Whites and Coloureds respectively. Significant variations were observed for all four races, but the largest differences between the Black and Coloured populations were observed for the marker.

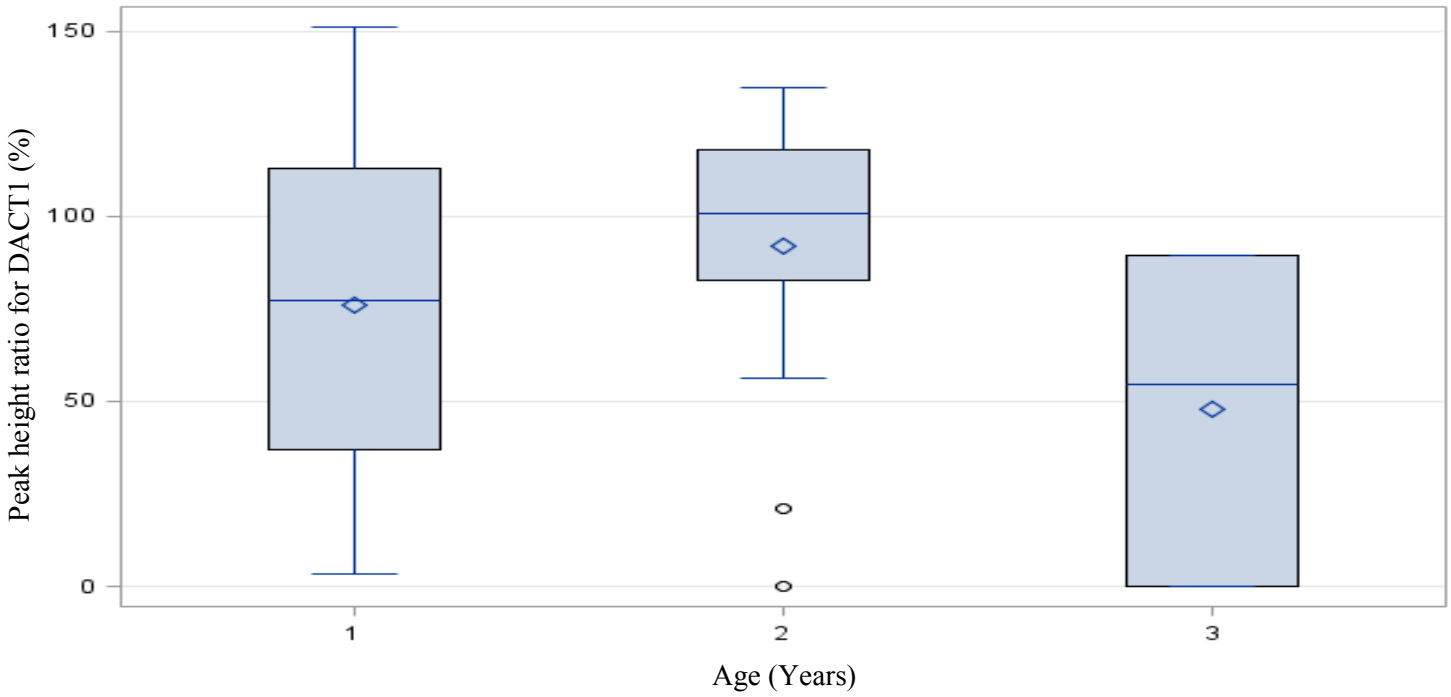

Figure 3.7: Distribution of methylation levels for the DACT1 tDMR across three age groups. 1 - <30 years; 2 – 30-50 years; 3 - >50 years.

3.3.3.3 L81528 tDMR-based Marker

The statistical analysis of the methylation status of the L81528 tDMR demonstrated that the marker showed significant variation between the four ethnic groups ($p = 0.03$) (Table 3.9). Similar to USP49 and DACT1, the Coloured group displayed the highest methylation level and the Black group displayed minimum methylation levels (Figure 3.8).

Table 3.9: Analysis of variance to test the effect of race and age on methylation status of the L81528 tDMR.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Race	3	2976.08	992.03	3.27	0.03
Age	2	1181.18	590.59	1.94	0.16
Race*Age	5	1262.19	252.44	0.83	0.54

Distribution of methylation levels for the L81528 tDMR across all four ethnic groups and ages are represented in Figure 3.8 and Figure 3.9, respectively. Methylation levels for this marker across all age groups did not significantly differ (Figure 3.9). Similarly, the interaction of race and age, as well as gender, did not have an effect (data not shown).

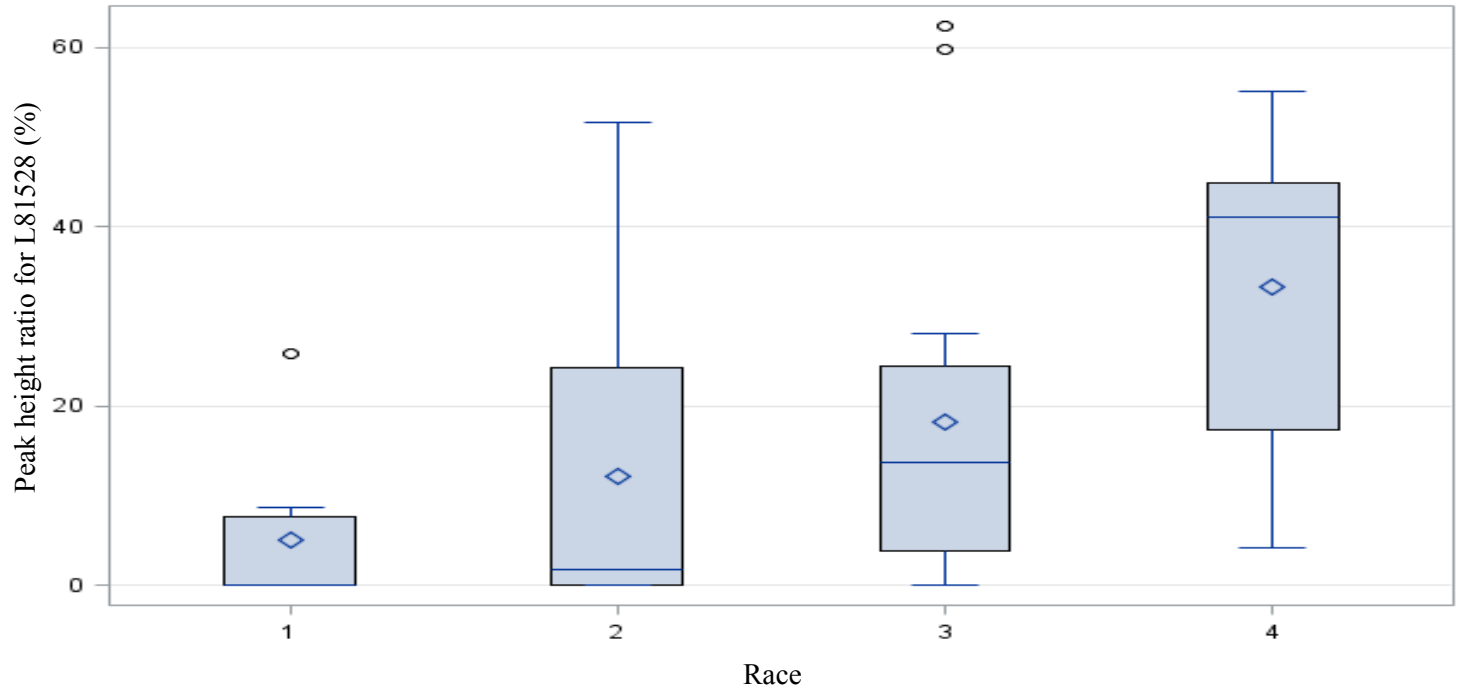

Figure 3.8: Box-Whisper plot depicting the effect of ethnicity on methylation status of the L81528 tDMR across four ethnic groups. Races 1-4 represent Blacks, Indians, Whites and Coloureds respectively. Significant variations between all four ethnic groups were observed for the marker, but the largest differences between the Black and Coloured populations were observed for the marker.

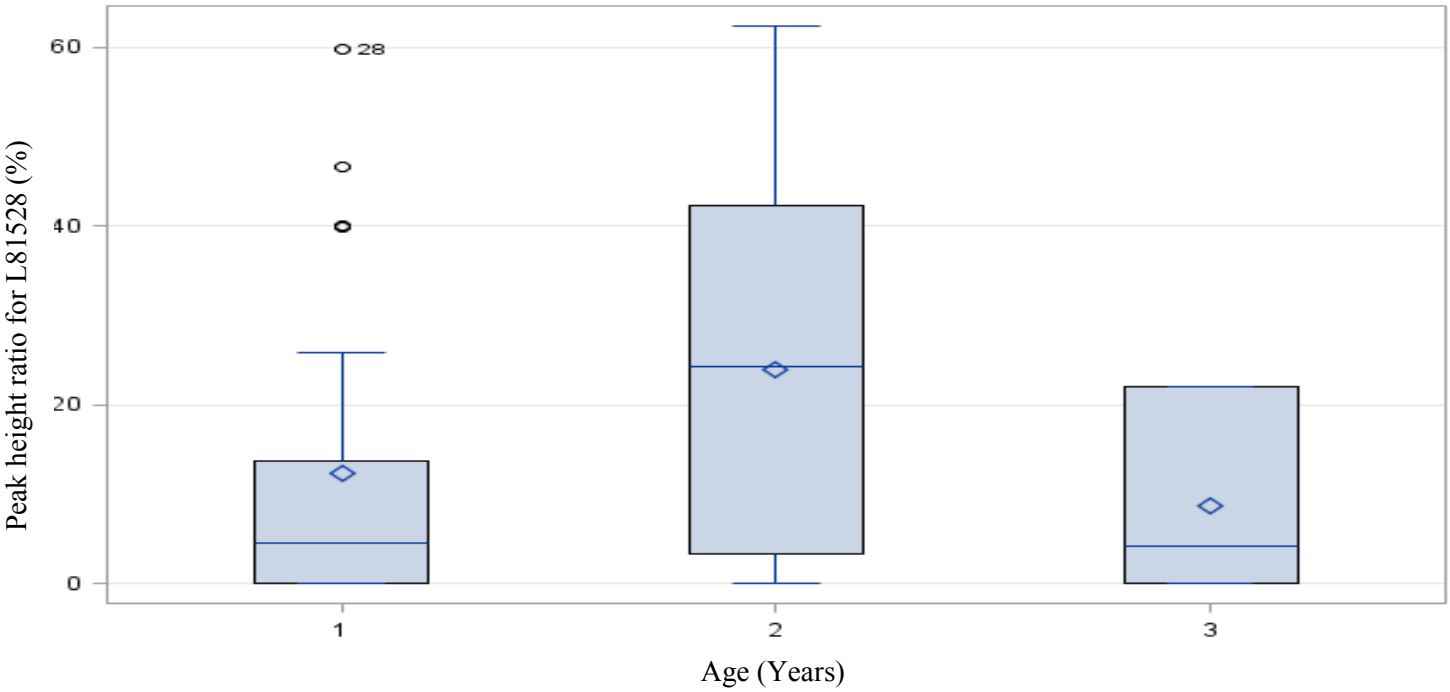

Figure 3.9: Distribution of methylation levels in the L81528 tDMR across three age groups. 1 - <30 years; 2 – 30-50 years; 3 - >50 years.

3.3.3.4 PFN3 tDMR-based Marker

The statistical analysis of the methylation status of the PFN3 tDMR demonstrated that the marker did show slight variation between the four ethnic groups (p = 0.05) (Table 3.10). The Coloured group displayed the highest methylation level and the Black groups displayed minimum methylation levels (Figure 3.10).

Table 3.10: Analysis of variance to test the effect of race and age on methylation status of the PFN3 tDMR.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Race	3	2266.35	755.45	2.78	0.05
Age	2	1046.00	523.00	1.92	0.16
Race*Age	5	1398.62	279.72	1.03	0.42

Distribution of methylation levels for the PFN3 tDMR across all four ethnic groups and ages are represented in Figure 3.10 and Figure 3.11, respectively. A consistent level of methylation of the PFN3 tDMR in subjects <30 years and between 30-50 years of age was observed (Figure 3.11). Interaction of race and age, as well as gender, did not have an effect (data not shown).

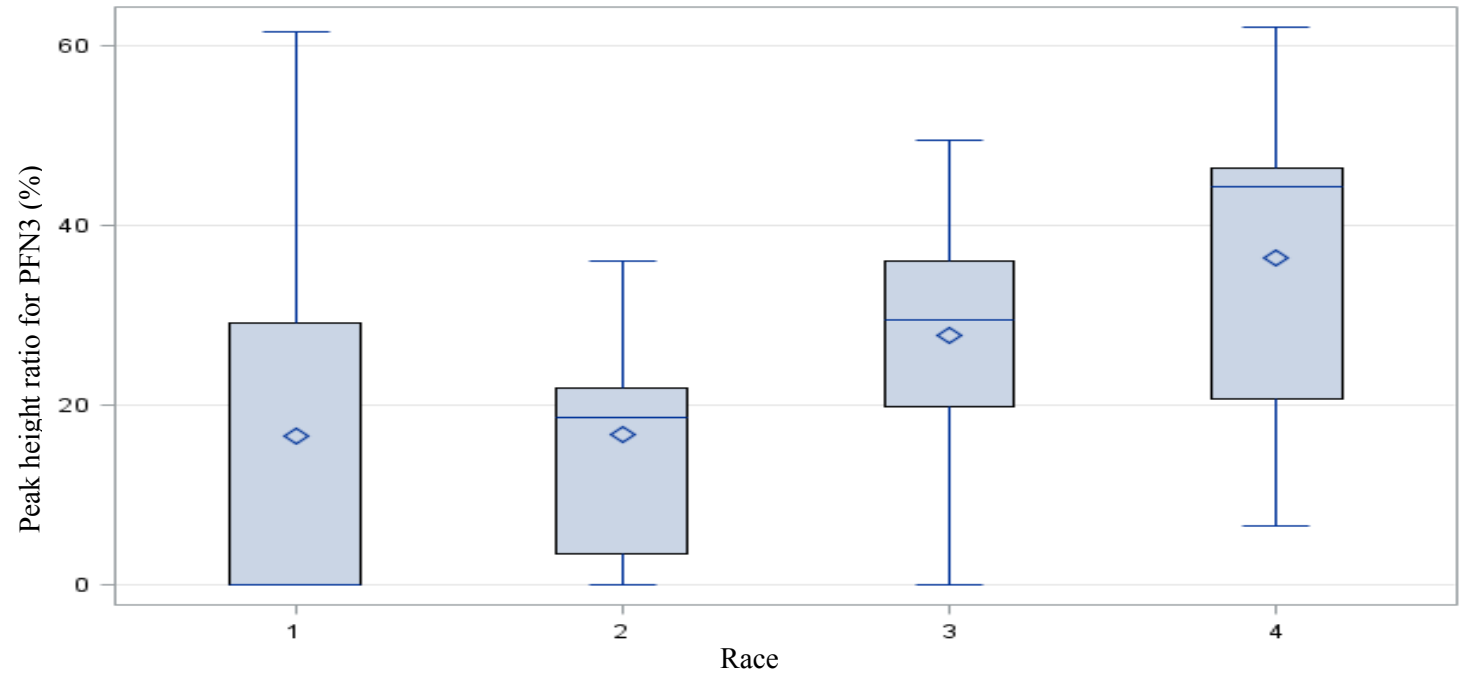

Figure 3.10: Box-Whisper plot depicting the effect of ethnicity on methylation status of the PFN3 tDMR across four ethnic groups. Races 1-4 represent Blacks, Indians, Whites and Coloureds respectively.

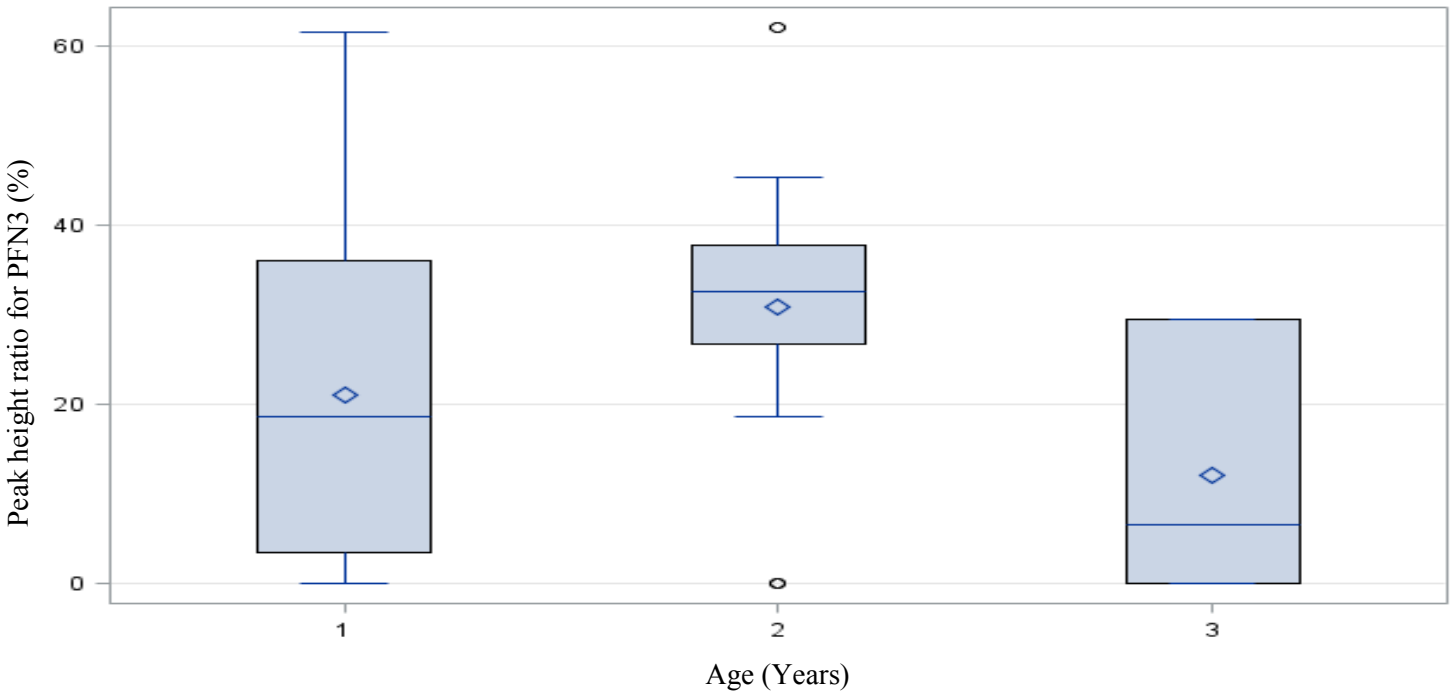

Figure 3.11: Distribution of methylation levels in the PFN3 tDMR across three age groups. 1 - <30 years; 2 – 30-50 years; 3 - >50 years.

3.4 Discussion

Predominantly for medical research purposes, studies have proven that variations in human populations exist at the genetic and epigenetic level (Lukiw, 2013; Reitz *et al.,* 2013; Ryan *et al.,* 2010; Tang *et al.,* 2005). South Africa is a multi-ethnic country comprising of various races, religions and traditions. While variation in outward appearances is apparent, little is understood of the extent of these variances at the epigenetic level. Thus, the present study was undertaken to explore this epigenetic variation in the context of tissue-specific DNA methylation. The aim was to determine if the methylation profile of tissue-specific differentially methylated regions (tDMRs) in saliva differs between four ethnic groups in South Africa, namely; Blacks, Whites, Indians and Coloureds. The analysis of these tDMR-based biomarkers may be of relevance in surveying phenotypic and ethnic traits for forensic applications.

The approach to screen the diverse South African population first began by collecting a forensically significant body fluid from study subjects. The body fluid of choice, saliva; was selected as advantages lie in ease of access and handling, non-invasive collection techniques, as well as no cost involvement (Pandeshwar and Das, 2014; Rajshekar *et al.,* 2013). Saliva was collected from 80 healthy individuals (with no history of any chronic disorder). As volunteers were from different age groups, the effect of age on methylation status of the selected tDMRs was also investigated. To facilitate the differentiation of the four ethnic groups using the saliva DNA, a multiplex PCR assay was used to detect the methylation status of tDMRs. Four tDMRS from the *USP49, DACT1, L81528* and *PFN3* genes were selected. These markers have been previously reported to differ between forensically significant fluids including saliva, blood, semen and vaginal fluid (An *et al.,* 2013; Choi *et al.,* 2014; Lee *et al.,* 2012). Two tDMRs for the *DACT1* and *USP49* genes display semen-specific hypomethylation (An *et al.,* 2013; Choi *et al.,* 2014; Lee *et al.,* 2012) therefore they should ideally be hypermethylated in saliva. Another semen-specific hypermethylation marker; L81528 was included (Choi *et al.,* 2014; Frumkin *et al.,* 2011). This tDMR should ideally display hypomethylation in saliva. The PFN3 tDMR was chosen as a vaginal fluid and menstrual blood-specific hypomethylation marker (An *et al.,* 2013; Lee *et al.,* 2012). This marker would similarly be hypermethylated in saliva. The multiplex assay used was based on Methylation-Sensitive Restriction Enzyme PCR (MSRE-PCR). The methylation-sensitive enzyme used for restriction was *Hha*I.

Methylation profiling of saliva indicated a distinct variation in the methylation status of the four tDMRs; especially between the Coloured and Black ethnic groups. All four tDMR markers displayed highest methylation levels in the Coloured ethnic group. The methylation level/status of all tDMRs for all study subjects was graphically represented as peak height ratios. A trend in methylation levels is apparent; the tDMRs of the Coloured population displayed highest methylation levels, followed by Whites, Indians and Blacks. The Coloured population of South Africa is said to be the outcome of admixture, which is mating between individuals from reproductively isolated ancestral populations; in this case the admixture of various ethnic groups namely Khoi-San, Xhosa, and descendants from European, South East Asian, Indian and Middle East immigrants (D'Amato *et al.,* 2008). Notably, the Black population shows the lowest methylation levels for all four tDMRs. Similar results were observed by Terry *et al.* (2008) and Zhang *et al.* (2011a), who compared DNA methylation levels of blood DNA obtained from Hispanics, Whites and Blacks and found lowest levels of methylation in the Black populations. In this regard, there have been reports of similar methylation patterns between saliva and blood (Frumkin *et al.,* 2011; Liu *et al.,* 2010; Philibert *et al.,* 2008; Thompson *et al.,* 2013).

In the present study, the DACT1 tDMR marker displayed the highest peak height ratio, followed by PFN3, L81528 and USP49 which displayed the lowest average/mean peak height ratio (between the four markers) of 22.74% in the Coloured population. Other studies have also reported high methylation of the DACT1 tDMR in saliva. The DACT1 tDMR was also found to display highest methylation followed by USP49 and PFN3 in saliva obtained from 34 Korean males and females in the study by Choi and colleagues (2014). An and colleagues (2013) found highest peak height ratio for the DACT1 tDMR followed by PFN3, USP49 and PRMT2 tDMRs; and Lee *et al.* (2012) found approximately 97% methylation for the DACT1 tDMR.

ANOVA analysis showed that methylation levels for the USP49 and PFN3 tDMRs did not differ significantly across all races ($p = 0.05$). However, the distribution plot showed a marked difference between the Coloureds and Black groups. There were no significant variations between the Indian and White populations for these markers therefore they would not be good candidates to differentiate between these two populations. Choi and colleagues (2014) have reported the PFN3 marker to generate significant methylation yields in saliva

DNA in their study, wherein 34 study participants were involved. Similarly Lee *et al.* (2012) found methylation levels above 80% for the PFN3 marker in saliva.

The DACT1 ($p = 0.02$) and L81528 ($p = 0.03$) markers displayed significant variation in methylation levels across all four ethnic groups. Since the L81528 tDMR is a semen-specific hypermethylation marker (Choi *et al.,* 2014) either complete unmethylation or hypomethylation should be observed in saliva. Consistently low, yet varying methylation levels was observed across all races and ages. A significant variation in methylation levels were observed across all four races ($p = 0.03$), and analogous to USP49 and DACT1, largest differences in methylation levels between the Coloured and Black populations were observed.

There was no significant effect of age and gender on the methylation profile of saliva for the ethnic groups. A larger sample size would shed more light on the effect of age on methylation levels of tDMRs. An *et al.* (2013) used Methylation Snapshot and found that, in saliva, there was a significant difference in methylation levels between young (below 30 years) and elderly males (above 50 years) for the DACT1 tDMR ($p = 0.008$).

Promising results were found in this study for two markers, DACT1 and L81528 as both these markers displayed significant variations between the Coloured and Black ethnic groups. The analysis of DNA extracted from saliva has been proven to be informative in medical as well as forensic research (Pandeshwar and Das, 2014; Malamud and Rodriguez-Chavez, 2011; Tabak, 2001). Once deemed as merely a digestive juice, researching salivary biomarkers is widely acclaimed due to non-invasive methods of collection and hence compliance of study subjects, as well as providing good quantities of sound quality DNA for large amounts of applications. Minimum volumes of 0.1 ml have been sufficient for excellent quality DNA despite being subjected to processing and storage during shipping, and when stored without refrigeration for up to eight months (Abraham *et al.,* 2012; Nunes *et al.,* 2012; Rogers *et al.,* 2007). However, detecting methylation status of the tDMRs in other body fluids, such as blood, semen, vaginal fluid and menstrual blood obtained from the four ethnic groups will further introduce more accuracy and precision in the analysis.

The optimisation of multiplex PCR was initially a challenging task. Numerous factors were taken into consideration, such as primer concentration and thermocycling conditions. However, the method facilitated rapid amplification of the four tDMRs of interest as well as the sex-typing marker Amelogenin in a single reaction. Multiplexing in forensic DNA investigations would be advantageous since only a limited amount of template DNA would

be needed for a detailed analysis of samples. Overall costs are reduced, time is saved and the added use of fluorescent primers supported detection of numerous targets within a single reaction which eliminates chances of cross-contamination. The multiplex MSRE-PCR used in the present study for differentiation of ethnic groups based on methylation levels of tDMRs was designed to use a standard capillary electrophoresis platform; which is compatible with STR-typing methods. The method ensures forensic applicability, high sensitivity, specificity and reliability. Developments in novel technologies for analysis of DNA methylation enable easy compatibility with current STR-typing methods which makes the method more attractive. Analysis of DNA methylation status in tDMRs, especially DACT1 and L81528, may be potentially implemented in differentiation of ethnic groups in real forensic casework. However, further quantification of the methylation status of the tDMRs and testing in a larger sample size is necessary to substantiate these findings.

CHAPTER FOUR

GENERAL DISCUSSION

AND

CONCLUSION

4.1 Research in Perspective

Human biological traces, such as saliva, blood, semen, vaginal fluid and menstrual blood have the potential to serve as strong evidence to place a suspect or victim at a crime scene. For forensic investigations, activities leading to the deposition of an individual's biological material are increasingly disputed. Therein, the identification of the biological material could be essential in order to determine that a violent act has indeed taken place, as well as to determine the circumstances leading to deposition of the biological material, which may have resulted in injury or death.

Traditional presumptive and confirmative methods developed for differentiation of body fluids, such as identification of phosphatase in semen, haeme in blood and amylase in saliva lack sensitivity, are applicable to a few types of body fluids and result in consumption of evidential material (Haas *et al.,* 2009; Quarino *et al.,* 2005; Virkler and Lednev, 2009). Following the introduction of DNA/RNA co-extraction techniques in forensics, RNA analysis became a routine application (Alvarez *et al.,* 2004; Haas *et al.,* 2014). Numerous body fluid-specific mRNA and miRNA markers have been identified (Juusola and Ballantyne, 2007; Nussbaumer *et al.,* 2006; Park *et al.,* 2014b), but RNA itself is highly unstable and susceptible to degradation by ribonucleases, fluctuating pH, moisture and UV light (Liu *et al.,* 2002; Park *et al.,* 2014a). Whereas mRNA and miRNA markers are sensitive to a certain extent, there are issues in applying these markers for routine applications. These include normalization in analysis, susceptibility to environmental degradation and expression not being fully cell-type specific (Ma *et al.,* 2013b; Sijen, 2014; Vennemann and Koppelkamm, 2010). DNA is highly regarded as a biological source for forensic identification profiling due to high stability and insusceptibility to external factors such as fluctuations in pH, temperature and humidity (An *et al.,* 2012; Frumkin *et al.,* 2011). Currently, identification of body fluids based on differential DNA methylation is an area of much interest because tissue-specific differentially methylated regions (tDMRs) exhibit different DNA methylation profiles according to tissue or cell type (An *et al.,* 2013; Choi *et al.,* 2014; Lee *et al.,* 2012; Rakyan *et al.,* 2008).

There is a dire need for the identification of new tDMRs and development of novel markers to increase specificity and sensitivity of DNA methylation-based profiling. Thus, the main aim of the present study was to identify new tDMRs for forensically significant body fluids (saliva, blood, semen and vaginal fluid). The search for novel tDMRs was based on

information acquired from bioinformatics databases on DNA methylation and gene expression in the human genome. Over-expressed or up-regulated genes in body fluids of interest were identified. Numerous studies have demonstrated that gene expression is regulated by DNA methylation (Choi *et al.,* 2014; Kulis *et al.,* 2013; Rakyan *et al.,* 2008). Hence, it was proposed that genes over-expressed in a single body fluid are likely to be hypomethylated in that tissue, and hypermethylated in normal tissues. As gene expression data was only available for blood, gene expression data from surrogate tissues for saliva (surrogate tissues: salivary glands and tongue), semen (surrogate tissue: prostate) and vaginal fluid (surrogate tissues: uterus and cervix) was obtained. Differential DNA methylation markers are persistent, resulting in cellular differentiation with specific expression of those specific genes (Eckhardt *et al.,* 2006; Jones and Takai, 2001). The present study sought to exploit this characteristic; using unique methylation levels of differentiated cells to identify them. The underlying principle behind identifying potential tDMRs in these body fluids was to target the differentially methylated regions to design PCR primers to differentiate between the four fluids.

Numerous researchers have reported experimental methods for identification of tDMRs including Whole Genome Bisulfite Sequencing (WGBS), pyrosequencing (Madi *et al.,* 2012; Park *et al.,* 2014a), base resolution methylomes genome sequencing (MethylC-Seq) (Schultz *et al.,* 2015) and epigenome-typing arrays such as Infinium HumanMethylation BeadChip (Dedeurwaerder *et al*., 2013). However, the present study describes a completely bioinformatics-based search for novel tDMRs which is relatively simple and low-cost since databases are freely available online, and very importantly it provides high resolution screening for tDMRs throughout the genome. After intricate screening of genes and their associated potential tDMRs in surrogate tissues, four genes and their heavily methylated CGIs were selected for primer design. These genes were *HPCAL1* (Hippocalcin-Like 1), *ZNF282* (Zinc Finger 282), *PTPRS* (Protein Tyrosine Phosphatase) and *PMEPA1* (Prostate Transmembrane Protein).

For studying DNA methylation profiles of identified potential tDMRs in four body fluids of interest, Methylation-Sensitive Restriction Enzyme PCR (MSRE-PCR) using a methylation-sensitive enzyme, *Hha*I was employed. MSRE treatment followed by PCR and capillary electrophoresis presents advantageous features over bisulfite pre-treatment methods; however it has not been used to a large extent in gene expression-based research. Most research focuses on bisulfite conversion since it has been deemed the standard method of

methylation analysis (Madi *et al.,* 2012; Park *et al.,* 2014a), however, these studies have not taken into account the high error rates, unreliability, low sensitivity and specificity. To surpass these limitations, MSRE-PCR was selected.

The HPCAL1 gene-based primer enabled successful differentiation of blood from saliva, semen and vaginal fluid. HPCAL1 displayed hypomethylation in blood relative to other fluids. The *HPCAL1* gene encodes a protein which is a member of a neuron-specific calcium-binding protein family that is typically found in the brain and retina. Not only is the gene important for neuronal signalling in the central nervous system and calcium-dependent regulation of rhodopsin phosphorylation (Burgoyne, 2007), but also recently Wang and colleagues (2014) found that *HPCAL1* is involved in Alzheimer's Disease. Currently, only a few promising markers for identification of blood have been reported. Frumkin *et al.* (2011), An *et al.* (2013) and Lee *et al.* (2012) have reported difficulty in differentiation between blood and saliva. Additionally, the ELOVL2 marker was also studied to be a potential marker for blood by Zbiec-Piekarska *et al.* (2015). The HPCAL1 gene-based marker will aid in providing accurate differentiation of blood in forensic casework.

The PTPRS gene-based CGI was specifically hypermethylated in vaginal fluid and hypomethylated in saliva, blood and semen. PTPs are well-known as signalling molecules that partake in cellular processes such as growth, proliferation and mitosis. They are important in control of cellular phosphotyrosine levels and have also been implicated in neural development, and importantly, axon growth and guidance (Aricescu *et al.,* 2002). It is of great significance in oncogenic transformation; illnesses associated with PTPRS include ureteroceles and pineal gland cancer (Belinky *et al.,* 2015), and it has been proposed that SNPs in the *PTPRS* gene is associated with ulcerative colitis (Muise *et al.,* 2007). Particularly, vaginal fluid-specific markers are of utmost importance in one of the most common genital cancers in females, which is endometrial cancer (Doufekas *et al.,* 2014; Fiegl *et al.,* 2004; Jones *et al.,* 2013). The PTPRS gene-based primer shows great potential in forensic applications for specific identification of vaginal fluid but also in medical research, seeing as it is involved in many diseases. The other two gene-based markers, ZNF282 and PMEPA1, were not specific for single fluids.

The second aim of the study was to decipher whether the tissue-specific methylation status of previously documented tDMRs (An *et al.,* 2013; Choi *et al.,* 2014; Lee *et al.,* 2012)

differ in saliva of four ethnic groups of South Africa; which could enable scientists to narrow down the forensic search to a particular ethnic group.

Mainly for medical research purposes, numerous studies have investigated genetic variation amongst populations by employing protein, mRNA, miRNA, STR and SNP markers (Bryc *et al.,* 2010; Chornokur *et al.,* 2012; D'Amato *et al.,* 2009; Lukiw, 2013; Tishkoff *et al.,* 2010). However in forensics, human identification based on population variation has disregarded protein and RNA-based markers due to instability, insensitivity and susceptibility to environmental insults. To identify human remains from vehicle accidents, victims and criminals of armed conflicts, situations of criminal acts and disasters (Goodwin *et al.,* 2009), forensic analysts have mainly used STR and SNP markers. However, it has come to light that sizes of PCR amplicons of forensic STRs included in kits are still too large, allele drop-ins and drop-outs pose a limitation, markers are required in high numbers for accurate identification and STRs and SNPs cannot identify those individuals that are not captured on a national database. Recent studies have targeted variation in DNA methylation to differentiate between populations (Fraser *et al.,* 2012; Heyn *et al.,* 2013; Moen *et al.,* 2013). DNA methylation-based markers are stable, reproducible, sensitive and specific (Choi *et al.,* 2014; Frumkin *et al.,* 2011; Sijen, 2014).

For investigation of DNA methylation status of four previously reported tDMRs; USP49, DACT1, L81528 and PFN3 between Blacks, Indians, Whites and Coloureds, saliva was collected from 80 healthy volunteers. These tDMRs have been previously reported to differ between forensically significant fluids including saliva, blood, semen and vaginal fluid (An *et al.,* 2013; Choi *et al.,* 2014; Lee *et al.,* 2012). MSRE-PCR was employed to detect the methylation status of the tDMRs, using the methylation-sensitive enzyme, *Hha*I. Even though DNA was isolated from all saliva samples under similar conditions, the quality of DNA ranged from very high to very poor. From a total of 80 samples, 42 amplified and the final analysis was based on them.

Two tDMRs; namely DACT1 ($p = 0.02$) and L81528 ($p = 0.03$) displayed significant differences in methylation levels between the Coloured and Black populations. Amplification yield (based on peak height ratios) for all the four tDMRs was highest in the Coloured population, followed by Whites, Indians and Blacks. This implies that the methylation level of the four tDMRs in saliva from the Coloured ethnic group was higher when compared to the other ethnic groups. Blacks showed the lowest methylation levels for all four tDMRs.

Variation in DNA methylation levels between ethnic groups has been studied by many groups (Fraser *et al.,* 2012; Heyn *et al.,* 2013; Winnefeld *et al.,* 2012; Zhang *et al.,* 2011a). Terry *et al.* (2008) and Zhang *et al.* (2011a) reported DNA methylation levels in peripheral blood from Blacks, Whites and Hispanics. Both groups of researchers, similar to this study, identified the lowest methylation levels in the Black population. Fraser *et al*. (2012) compared differences of methylation patterns between European and Yorubans using lymphoblastoid cell lines (LCLs). In their investigation of 14 495 genes, over 4300 genes were found to display differential DNA methylation between the populations. In analysis of skin samples from Africans, Whites and Asians, Winnefeld and colleagues (2012) found that the intragenic CpG island of the *CPXM2* (Carboxypeptidase X Member II) was hypomethylated in Asians, but hypermethylated in Whites and Africans; and the *PM20D1* peptidase gene was hypermethylated in Caucasians and hypomethylated in the other two ethnic groups. In South African research, two recent medical studies have found differential DNA methylation in Blacks (Matatiele *et al.,* 2015) and Coloureds (Masemola *et al.,* 2015). However, while these studies have researched differential DNA methylation in South Africa, they have not studied it across various ethnic groups. Thus, to our knowledge, differentiation of ethnic groups by differential DNA methylation for forensic application has not been studied in South Africa prior to the present study.

While none of the markers displayed a significant variation across all three age groups (below 30, between 30 and 50, and above 50 years), a more consistent high level of methylation was observed in ages between 30-50 years when compared to those below 30 years. At this stage it is inconclusive if age really affected the methylation levels of the tDMRs. Furthermore, it is acknowledged that the population size of the study was rather small.

An *et al*. (2013) studied USP49, DACT1, PFN3 and PRMT2 tDMRs for body fluid identification. Saliva was found to display highest peak height ratio in DACT1, followed by PFN3, USP49 and lowest methylation in PRMT2. Similarly, Choi and colleagues (2014) and Lee and colleagues (2012) found high methylation for the DACT1 tDMR in saliva. In the present study using saliva, highest methylation levels were also observed in the DACT1 tDMR (phr = 109.46%). The semen-specific tDMR L81528 was found to be hypomethylated in saliva in the study by Choi *et al.* (2014). Consistently, the present study also showed a low methylation range of 5.07% in Blacks and 33.24% in Coloureds.

Previous studies have reported differential DNA methylation of the four tested tDMRs between forensically significant body fluids (An *et al.,* 2013; Choi *et al.,* 2014; Lee *et al.,* 2012) and methylation variation between ethnic groups using body fluids (Fraser *et al.,* 2012; Heyn *et al.,* 2013; Winnefeld *et al.,* 2012; Zhang *et al.,* 2011a). In this study, both aspects were integrated; studying differential DNA methylation of four tDMRs in saliva, between four ethnic groups.

While the study motivates for the concept of differentiation between ethnic groups based on analysis of tDMRs, much debate exists around the forensic technique of categorizing humans by race or ethnicity, irrespective of the biomarker in use. The notion that humans can indeed be classified into various races has been enshrined by some researchers, especially those in medical research seeking information about disease susceptibility and therapeutic intervention (Fuhrman *et al.,* 2000). However, the concept of human classification is also dismissed by others due to legal and ethical issues. Predominantly, using genetic variation for a description or inference of ethnicity of the perpetrator of a crime may lead to stigmatization of certain groups within society, especially concerning skin colour and diseases (Bamshad *et al.,* 2004; Koops and Schellekens, 2008). The use of biological material from a crime scene sample to identify physical traits causes controversy because of fears that the technology may be abused. For example, privacy issues arise when sensitive matters such as inclination towards homosexuality and susceptibility to disease are made public (Koops and Schellekens, 2008; M'charek *et al.,* 2012). Nevertheless, provided that the DNA analysis is restricted to gaining intelligence related to the case in question and not otherwise, inference of ethnic background adds a new dimension in forensic casework and will enable forensic scientists to narrow down the circle of possible suspects. Currently the technique is not yet fully developed, however the present study shows promising results for use of tDMRs to differentiate between the four ethnic groups of South Africa.

The multiplex MSRE-PCR method used in this study boasts several salient features for forensic casework. Analysis of DNA methylation by restriction enzymes is much simpler and less time-consuming than bisulfite conversions and is compatible with STR-typing methods. Furthermore, bisulfite treated DNA must be analysed within a certain time period due to instability of single strands (An *et al.,* 2012). The only factor to be wary of is efficient restriction of the DNA template. Hence, unmethylated DNA was used as a control.

4.2 Conclusion and Future Work

The present study has demonstrated that the analysis of tDMRs holds much potential in forensics and once implemented, will likely be an indispensable tool to identify the nature of the biological fluid in question. To our knowledge, the panel of genes (*HPCAL1, PTPRS, ZNF282* and *PMEPA1*) have not been previously targeted for identification of tDMRs and differentiation of forensically significant body fluids. The HPCAL1 tDMR was found to be a potential blood-specific hypomethylation marker and the PTPRS tDMR was found to be a potential vaginal fluid-specific hypermethylation marker. Future work would involve quantification of methylation and validation of the potential tDMRs identified in this study by different profiling methods such as Methylation-Specific PCR (Beri *et al.,* 2007; Lee *et al.,* 2012) or the development of multiplex assays such as methylation SNaPshot (An *et al.,* 2013). Testing the identified markers on a larger and geographically diverse sample size as well as on a broad age range is also recommended. Also, the identification of more tDMRs will contribute towards making methylation-based detection of forensic body fluid more robust.

Additionally, two tDMRs; namely DACT1 and L81528 showed potential to differentiate between Coloured and Black ethnic groups of South Africa. As this study was based solely on DNA from saliva, future work will involve analysis on other body fluids to determine if the same methylation status/level is observed. While studies have reported genetic variation between ethnic groups in South Africa (D'Amato *et al.,* 2008; D'Amato *et al.,* 2009), these were based on STR markers. Thus, to our knowledge differential DNA methylation between ethnic groups has not been previously studied. Future work in this regard would also involve testing the potential population-specific tDMRs on a larger sample size. The results are encouraging and pave the way for application of differential DNA methylation for body fluid identification, as well as to infer ethnicity of perpetrators in forensic applications.

REFERENCES

Abdolmaleky HM, Cheng K, Faraone SV, Wilcox M, Glatt SJ, Gao F, et al. 2006. Hypomethylation of MB-COMT promoter is a major risk factor for schizophrenia and bipolar disorder. Human Molecular Genetics. 15: 3132–3145.

Abraham JE, Maranian MJ, Spiteri I, Russell R, Ingle S, Luccarini C, et al. 2012. Saliva samples are a viable alternative to blood samples as a source of DNA for high throughput genotyping. BioMedCentral Medical Genomics. 5: 19.

Adkins RM, Krushkal J, Tylavsky FA, Thomas F. 2012. Racial Differences in Gene-Specific DNA Methylation Levels are Present at Birth. Birth defects research (Part A). Clinical and Molecular Teratology. 91: 728–736.

Akutsu T, Motani H, Watanabe K, Iwase H, Sakurada K. 2012. Detection of bacterial 16S ribosomal RNA genes for forensic identification of vaginal fluid. Legal Medicine. 14: 160–162.

Alegria-Torres TAA, Baccarelli A, Bolati V. 2011. Epigenetics and Lifestyle. Epigenomics. 3: 267–277.

Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, et al. 2010. International HapMap 3 Consortium-Integrating common and rare genetic variation in diverse human populations. Nature. 467: 52-58.

Alvarez M, Juusola J, Ballantyne J. 2004. An mRNA and DNA co-isolation method for forensic casework samples. Analytical Biochemistry. 335: 289–298.

Amorim A, Pereira L. 2005. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. Forensic Science International. 150: 17–21.

An JH, Choi A, Shin KJ, Yang WI, Lee HY. 2013. DNA methylation-specific multiplex assays for body fluid identification. International Journal of Legal Medicine. 127: 35–43.

An JH, Shin KJ, Yang WI, Lee HY. 2012. Body fluid identification in forensics. Biochemistry and Molecular Biology Reports. 45: 545-553.

Ando Y, Hayashizaki Y. 2006. Restriction Landmark Genomic Scanning. Nature Protocols. 1: 2774–2783.

Antunes JP, Madi T, Balamurugan K, Bombardi R, Duncan G, McCord B. 2013. DNA methylation markers as a powerful technique to discriminate body fluids present in crime scenes. Available at: http://au.promega.com/~/media/files/resources/conference%20proceedings/ishi%2024/oral%20presentations/antunes-manuscript.pdf. [Accessed 25 May 2014].

Anway MD, Cupp AS, Umzusu M, Skinner MK. 2005. Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science. 308: 1466–1469.

Applied Biosystems. 2003. GeneMapper ID Software Version 3.1 Human Identification Analysis User Guide. Foster City, California. Available at https://tools.thermofisher.com/content/sfs/brochures/cms_041475.pdf. [Accessed 05 August 2015].

Archey WB, Sweet MP, Alig GC, Bradley AA. 1999. Methylation of CpGs as a Determinant of Transcriptional Activation at Alternative Promoters for Transforming Growth Factor-*β3*. Cancer Research. 59: 2292–2296.

Aricescu AR, McKinnell IW, Halfter W, Stoker AW. 2002. Heparan Sulfate Proteoglycans are Ligands for Receptor Protein Tyrosine Phosphatase σ. Molecular and Cellular Biology. 22: 1881–1892.

Ariel M, McCarrey J, Cedar H. 1991. Methylation patterns of testis specific genes. Proceedings of the National Academy of Sciences USA. 88: 2317–2321.

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. 2014. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays. Bioinformatics. 30: 1363–1369.

Ayoubi TAY, Van de Ven WJM. 1996. Regulation of gene expression by alternative promoters. The Journal of the Federation of the American Societies for Experimental Biology. 10**:** 453-460.

Baker M. 2010. MicroRNA profiling: Separating signal from noise. Nature Methods. 7: 687–692.

Bamshad M, Wooding S, Salisbury BA, Stephens JC. 2004. Deconstructing the relationship between genetics and race. Nature Reviews. 5: 598–609.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. 2013. NCBI GEO: Archive for functional genomics data sets – update. Nucleic Acids Research. 41 (Database Issue): D991–D995.

Bekeart B, Kamalandua A, Zapico SC, Van de Voorde W, Decorte R. 2015. Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. Epigenetics. 9: 1559–2294.

Belinky F, Nativ N, Stelzer G, Zimmerman S, Iny-Stein T, Safran M, et al. 2015. PathCards: multi-source consolidation of human biological pathways. Database 2015: ID bav006.

Bell AC, Felsenfeld G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. Letters to Nature. 405: 482–485.

Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biology. 12: 1–13.

Benschop CC, Quaak FC, Boon ME, Sijen T, Kuiper I. 2012. Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context. International Journal of Legal Medicine. 126: 303–310.

Berezikov E, Guryev V, Van de Belt J, Wienholds E, Plasterk RH, Cuppen E. 2005. Phylogenetic shadowing and computational identification of human microRNA genes. Cell. 120: 21–24.

Beri S, Tonna N, Menozzi G, Bonaglia MC, Sala C, Giorda R. 2007. DNA methylation regulates tissue-specific expression of Shank3. Journal of Neurochemistry. 101: 1380–1391.

Bevel T, Gardner RM. 2002. Bloodstain Pattern Analysis: With an Introduction to Crime Scene Reconstruction (2nd Edn). CRC Press, Boca Raton.

Bibikova M, Jennie L, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, et al. 2009. Genome-wide DNA Methylation Profiling using Infinium Assay. Epigenomics. 1: 177–200.

Bierkins M, Hesselink TA, Meijer CJ, Heideman DAM, Wisman GB, Van de Zee AG, et al. 2013. CADM1 and MAL promoter methylation level in hr-HPV-positive cervical scrapes increase proportional to degree and duration of underlying cervical disease. International Journal of Cancer. 133: 1293-1299.

Bird AP. 1986. CpG-rich islands and the function of DNA methylation. Nature. 321: 209–213.

Bird AP. 1995. Gene number, noise reduction and biological complexity. Trends in Genetics. 11: 94–100.

Bird AP. 2002. DNA methylation patterns and epigenetic memory. Genes and Development. 16: 6-21.

Bjornsson HT, Sigurdsson MI, Fallin MD, Irrizary RA, Aspelund T, Yu W, et al. 2008. Intra-individual change over time in DNA methylation with familial clustering. Journal of the American Medical Association. 299: 2877–2883.

Bloushtain-Qimron N, Yaoa J, Eric LS, Michail S, Campbella LL, Manie SA, et al. 2008. Cell type-specific DNA methylation patterns in the human breast. Proceedings of the National Academy of Sciences USA. 37: 14076–14081.

Bock C. 2012. Analysing and interpreting DNA methylation data. Nature Reviews: Genetics. 13: 705–719.

Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, et al. 2010. Quantitative comparison of genome-wide DNA methylation mapping technologies. Nature Biotechnology. 28: 1106-1114.

Bockland S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S, et al. 2011. Epigenetic Predictor of Age. Public Library of Sciences One. 6**:** 1–6.

Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE, et al. 2009. The Relationship of DNA Methylation with Age, Gender and Genotype in Twins and Healthy Controls. Public Library of Sciences One. 4: 1-8.

Borek C. 2004. Dietary antioxidants and human cancer. Integrative Cancer Therapies. 3: 333–341.

Brennan K, Garcia-Closas M, Orr N, Fletcher O, Jones M, Ashworth A, et al. 2012. Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. Cancer Research. 72: 2304-2313.

Breuckner B, Stresemann C, Kuner R, Mund C, Musch T, Meister M, et al. 2007. The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. Cancer Research. 67: 1419–1423.

Bryc K, Autona A, Nelson MR, Oksenberg JR, Hauserc SL, Williams S, et al. 2010. Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proceedings of the National Academy of Sciences. 107: 786–791.

Budowle B, Daal A. 2008. Forensically Relevant SNPs. BioTechniques. 44: 603-610.

Burgoyne RD. 2007. Neuronal Calcium Sensor Proteins: Generating Diversity in Neuronal Ca2+ Signalling. Nature Reviews: Neuroscience. 8: 182–193.

Butler JM. 2012. Advanced Topics in Forensic DNA Typing: Methodology. Elsevier Academic Press.

Butler JM, Michael D, Coble D, Vallone MP. 2007. STRs vs. SNPs: thoughts on the future of forensic DNA testing. Forensic Science, Medicine and Pathology. 3: 200–205.

Calvanese V, Lara E, Kahn A, Fraga MF. 2009. The role of epigenetics in ageing and age-related diseases. Ageing Research Reviews. 8: 268–276.

Campbell MC, Tishkoff SA. 2008. African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins and Complex Disease Mapping. Annual Reviews of Genomics and Human Genetics. 9: 403–433.

Chen Z, Liu Q, Nadarajah S. 2012. A new statistical approach to detecting differentially methylated loci for case control Illumina array methylation data. Bioinformatics. 28: 1109-1113.

Cheong J, Yamada Y, Yamashita R, Irie T, Kanai A, Wakaguri H, et al. 2006. Diverse DNA methylation statuses at alternative promoters of human genes in various tissues. DNA Research. 13: 155–167.

Ching TT, Maunakea AK, Jun P, Hong C, Zardo G, Pinkel D, et al. 2005. Epigenome analyses using BAC microarrays identify evolutionary conservation of tissue-specific methylation of Shank3. Nature Genetics. 37: 645–651.

Choi A, Shin KJ, Yang WI, Lee HY. 2014. Body fluid identification by integrated analysis of DNA methylation and body fluid-specific microbial DNA. International Journal of Legal Medicine. 128: 33–41.

Choi SY, Chang CE, Kim SC, So JS. 2003. Antimicrobial susceptibility and strain prevalence of Korean vaginal Lactobacillus spp. Anaerobe. 9: 277–280.

Chornokur G, Amankwah EK, Schildkraut JM Phelan CM. 2012. Global ovarian cancer health disparities. Gynaecologic Oncology. 129: 258–264.

Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, et al. 2009. Ageing and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent on CpG island Context. Public Library of Sciences: Genetics. 5: 1–13.

Clerc P, Avner P. 1998. Role of the region 3' to Xist exon 6 in the counting process of X-chromosome inactivation. Nature Genetics. 19: 249–253.

Cocozza S, Akhtar MM, Miele G, Monticelli A. 2011. CpG Islands undermethylation in human genomic regions under selective pressure. Public Library of Sciences One. 6: e23156.

Cohen NM, Kenigsberg E, Tanay A. 2011. Primate CpG Islands are maintained by Heterogeneous Evolutionary Regimes Involving Minimal Selection. Cell. 145: 773–786.

Costello JF, Smiraglia DJ, Plass C. 2002. Restriction Landmark Genomic Scanning. Methods. 27: 144–149.

Cross SH, Charlton JA, Nan X, Bird AP. 1994. Purification of CpG islands using a methylated DNA binding column. Nature Genetics. 6: 236-244.

Cui JG, Li YY, Zhao Y, Bhattacharjee S, Lukiw WJ. 2010. Differential regulation of interleukin-1 receptor-associated kinase-1 (IRAK-1) and IRAK-2 by microRNA-146a and NF-kappa B in stressed human astroglial cells and in Alzheimer disease. Journal of Biological Chemistry. 285**:** 38951–38960.

Cullen P. 2000. Evidence that triglycerides are an independent coronary heart disease risk factor. American Journal of Cardiology. 86: 943–949.

Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. 2015. Nucleic Acids Research. 43 (Database Issue): D662-D669.

D'Amato ME, Ehrenreich L, Benjeddou M, Davison S, Leat N. 2008. Ancestry and genetic relationships between groups within the Cape Town metropolitan population inferred using Y-STRs genotyping. Forensic Science International: Genetics Supplement Series. 1: 318–319.

D'Amato ME, Benjeddou M, Davison S. 2009. Evaluation of 21 Y-STRs for population and forensic studies. Forensic Science International: Genetics Supplement Series. 2: 446–447.

Dalma-Weiszhausz DD, Warrington J, Tanimoto EY, Miyada CG. 2006. The Affymetrix GeneChip platform: an overview. Methods in Enzymology. 410: 3-28.

Danaher P, White RL, Hanson E, Ballantyne J. 2015. Facile semi-automated forensic body fluid identification by multiplex solution hybridization of NanoString® barcode probes to specific mRNA targets. Forensic Science International: Genetics. 14 : 8–30.

Das PM, Ramachandran K, Vanwert J, Ferdinand L, Gopisetty G, Reis IM, et al. 2006. Methylation mediated silencing of *TMS1/ASC* gene in prostate cancer. Molecular Cancer. 5: 28.

Day K, Waite LL, Mercer AT, West A, Bamman MM, Brooks JD, et al. 2013. Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. Genome Biology. 14: 1–19.

Deaton AM, Bird AP. 2001. CpG Islands and the Regulation of Transcription. Genes and Development. 25: 1010–1022.

Dedeurwaerder S, Defrance M, Bizet M, Calonne E, Bontempe G, Fuks F. 2013. A Comprehensive Overview of Infinium HumanMethylation450 Data Processing. Briefings in Bioinformatics. Page 1-13. Available at http://bib.oxfordjournals.org/. [Accessed 24 September 2014].

De-Smet C, Lurquin C, Lethe B, Martelange V, Boon T. 1999. DNA methylation is the primary silencing mechanism for a set of germ-line and tumour-specific genes with a CpG-rich promoter. Molecular Cell Biology. 19: 7327-7335.

Dewhirst FE, Klein EA, Thompson EC, Blanton JM, Chen T, Milella L, et al. 2012. The canine oral microbiome. Public Library of Sciences One. 7: e36067.

Dixon LA, Murray CM, Archer EJ, Dobbins AE, Koumi P, Gill P. 2005. Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. Forensic Science International. 154: 62-77.

Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, et al. 2009. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nature Genetics. 12: 1350–1353.

Dong Z, Wang X, Evers BM. 2000. Site-specific DNA methylation contributes to neurotensin/neuromedin N expression in colon cancers. American Journal of Physiology-Gastrointestinal and Liver Physiology. 279: G1139–G1147.

Doufekas K, Hadwin R, Kandimalla R, Jones A, Mould T, Crowe S, et al. 2014. *GALR1* Methylation in Vaginal Swabs is Highly Accurate in Identifying Women with Endometrial Cancer. International Journal of Gynaecological Cancer. 23: 1050–1055.

Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, et al. 2008. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nature Biotechnology. 26: 779–785.

Du P, Bourgon R. 2015. methyAnalysis: an R package for DNA methylation data analysis and visualization. R package Version 1.12. Available at https://www.bioconductor.org/packages/3.3/bioc/vignettes/methyAnalysis/inst/doc/methyAnalysis.pdf [Accessed 20 July 2015].

Dunn J, Qiu H, Kim S, Jjingo D, Hoffman R, Kim CW, et al. 2014. Flow-dependent epigenetic DNA methylation regulates endothelial gene expression and atherosclerosis. Journal of Clinical Investigation. 124: 3187–3199.

Dyson MT, Roqueiro D, Monsivais D, Mutlu-Ercan C, Pavone ME, Brooks DC, et al. 2014. Genome-Wide DNA Methylation Analysis Predicts an Epigenetic Switch for GATA Factor Expression in Endometriosis. Public Library of Sciences: Genetics. 10: e1004158.

Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. Nature Genetics. 38: 1378–1385.

Edwards JR, O'Donnell AH, Rollins RA, Peckham HE, Lee C, Milekic MH, et al. 2010. Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. Genome Research. 20: 972–980.

Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liliglou T, Xinarianos G, et al. 2005. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proceedings of the National Academy of Sciences. 44: 15785–15790.

Ehrlich M. 2002. DNA methylation in cancer: Too much, but also too little. Oncogene. 21: 5400–5413.

El-Maari O, Becker T, Junen J, Manzoor SS, Diaz-Lacava A, Schwaab R, et al. 2007. Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. Human Genetics. 122: 505–514.

ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 447: 799-816.

ENCODE Project Consortium. 2012. An integrated encyclopaedia of DNA elements in the human genome. Nature. 489: 57-74.

Enokida H, Shiina H, Urakami S, Igawa M, Ogishima T, Pookot D, et al. 2005. Ethnic group-related differences in CpG hypermethylation of the *GSTP1* gene promoter among African-American, Caucasian and Asian patients with prostate cancer. International Journal of Cancer. 116: 174–181.

Fakruddin MD, Chowdhury A, Nur-Hussain MD, Shahnewaj Bin Mannan K, Mazumdar RM. 2012. Pyrosequencing – Principles and Applications. International Journal of Life Sciences and Pharma Research. 2: 65–76.

Fang MZ, Wang Y, Ai N, Hou Z, Sun Y, Lu H, et al. 2003. Tea polyphenol (-)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines. Cancer Research. 63: 7563–7570.

Fiegl H, Gattringer C, Widschwendter A, Schneitter A, Ramoni A, Sarlay D, et al. 2004. Methylated DNA Collected by Tampons – A New Tool to Detect Endometrial Cancer. Cancer Epidemiology, Biomarkers and Prevention. 13: 882–888.

Fleming RI, Harbison S. 2010. The use of bacteria for the identification of vaginal secretions. Forensic Science International: Genetics. 5: 311–315.

Flores GE, Bates ST, Knights D, Lauber CL, Stombaugh J, Knight R, et al. 2011. Microbial biogeography of public restroom surfaces. Public Library of Sciences One. 6: e28132.

Forch R, Schonherr H, Jenkins ATA. 2009. Surface Design: Applications in Bioscience and Nanotechnology. John Wiley and Sons.

Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. 2005. Epigenetic differences arise during the lifetime of monozygotic twins. Proceedings of the National Academy of Sciences. 102: 10604–10609.

Fraser HB, Lam LL, Nuemann SM, Kobor MS. 2012. Population-specificity of human DNA methylation. Genome Biology. 13: 1–13.

Friso S, Choi SW, Girelli D, Mason JB, Dolnikowski GG, Bagley PJ, et al. 2002. A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. Proceedings of the National Academy of Sciences USA. 99. 5606-5611.

Frommer M, McDonald LE, Miller DS, Collis CM, Watt F, Grigg GW, et al. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the Natural Academy of Sciences USA. 89: 1827-1831.

Frumkin D, Wasserstrom A, Davidson A, Grafit A. 2010. Authentication of forensic DNA samples. Forensic Science International: Genetics. 4: 95–103.

Frumkin D, Wasserstrom A, Budowle B, Davidson A. 2011. DNA methylation-based forensic tissue identification. Forensic Science International: Genetics. 5: 517–524.

Fry RC, Svensson JP, Valiathan C, Wang E, Hogan BJ, Bhattacharya S, et al. 2008. Genomic predictors of inter-individual differences in response to DNA damaging agents. Genes and Development. 22: 2621–2626.

Fuhrman S, Cunningham MJ, Wen X, Zweiger G, Seilhamer JJ, Somogyi R. 2000. The application of Shannon entropy in the identification of putative drug targets. Biosystems. 55: 5–14.

Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, et al. 2012. Methylation of *ELOVL2* gene as a new epigenetic marker of age. Ageing Cell. 11: 1132-1134.

Gartler SM, Goldman MA. 2001. X Chromosome Inactivation. Encyclopaedia of Life Sciences. Available at http://web.udl.es/usuaris/e4650869/docencia/segoncicle/genclin98/recursos_classe_(pdf)/revisionsPDF/XChromoInac.pdf. [Accessed 09 November 2015].

Gebhard C, Schwarzfischer L, Pham T, Schilling E, Klug M, Andreesen R, et al. 2006. Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukaemia. Cancer Research. 66: 6118–6128.

Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature Biotechnology. 26: 317–325.

Ghosh S, Yates AJ, Fruhwald MC, Miecznikowski JC, Plass C, Smiraglia DJ. 2010. Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. Epigenetics. 5: 527–538.

Giampaoli S, Berti A, Valeriani F, Gianfranceschi G, Piccolella A, Buggiotti L, et al. 2012. Molecular identification of vaginal fluid by microbial signature. Forensic Science International: Genetics. 6: 559–564.

Gill P. 2002. Role of Short Tandem Repeat DNA in Forensic Casework in the UK-Past, Present and Future Perspectives. BioTechniques. 32: 366–385.

Gill P, Werritt DJ, Budowle B, Guerrieri R. 2004. An assessment of whether SNPs will replace STRs in national DNA databases-Joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). Science and Justice. 44: 51-53.

Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F. 2012. R-Loop formation is a distinctive characteristic of unmethylated human CpG Island promoters. Molecular Cell. 45: 814-825.

Goa ZH, Suppola S, Liu J, Heikkila P, Janne J, Voutilainen R. 2002. Association of H19 promoter methylation with the expression of *H19* and *IGF-II* genes in adrenocortical tumours. Journal of Endocrinology and Metabolism. 87: 1170–1176.

Goldberg AD, Allis CD, Bernstein E. 2007. Epigenetics: A Landscape Takes Shape. Cell. 128: 635–638.

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. Annual Review of Biochemistry. 74: 481-514.

Gomes I, Strohbucker B, Rothschild MA, Schneider PM. 2013. Evaluation of mRNA specific markers using a pentaplex system for the identification of skin and saliva from contact trace evidence. Forensic Science International: Genetics Supplement Series. 4: e180–e181.

Gomes I, Kohlmeier P, Schneider PM. 2011. Genetic markers for body fluid and tissue identification in forensics. Forensic Science International: Genetics Supplement Series. 3: e469–e470.

Goodwin W, Butler JM, Carracedo A, Orego C, Wysocki M, Simmons T, et al. 2009. Missing People: DNA Analysis and Identification of Human Remains: A guide to best practice in armed conflicts and other situations of armed violence (2nd Edn). International Committee of the Red Cross (ICRC). Available at https://www.icrc.org/eng/assets/files/other/icrc_002_4010.pdf. [Accessed 10 October 2015].

Gowher H, Leismann O, Jeltsch A. 2000. DNA of Drosophila melanogaster contains 5-methyl-cytosine. European Molecular Biology Organization Journal. 19: 6918-6923.

Goyal R, Reinhardt R, Jeltsch A. 2006. Accuracy of DNA Methylation Patterns Preservation by the DNMT1 Methyltransferase. Nucleic Acids Research. 34: 1182-1188.

Gray KA, Yates B, Seal RL, Wright MW, Bruford EA. 2015. Genenames.org: the HGNC resources in 2015. Nucleic Acids Research. 43 (Database Issue): D1079–D1085.

Greenfield A, Sloan MA. 2003. Identification of biological fluids and stains. S. H. James and J. J. Nordby (Eds.). Forensic Science: An Introduction to Scientific and Investigative Techniques. CRC Press, Boca Raton.

Grunau C, Renault E, Rosenthal A, Roizes G. 2001. MethDB - A public database for DNA methylation data. Nucleic Acids Research. 29: 270-274.

Grunau C, Hindermann W, Rosenthal A. 2000. Large scale methylation analysis of human genomic DNA reveals tissue-specific differences between the methylation profiles of genes and pseudogenes. Human Molecular Genetics. 9: 2651–2663.

Gunn P, Walsh S, Roux C. 2014. The nucleic acid revolution continues – Will forensic biology become forensic molecular biology? Frontiers in Genetics. 5: 1-4.

Guo G, Kang Q, Chen Q, Chen Z, Wang J, Tan L, et al. 2014a. High expression of long non-coding RNA H19 is required for efficient tumourigenesis induced by *Bcr-Abl* oncogene. Federation of European Biochemical Societies Letters. 588: 1780-1786.

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, et al. 2014b. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nature Neuroscience. 17: 215–222.

Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, et al. 2013. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. eLIFE. 2: e00523.

Haas C, Klesser B, Maake C, Bar W, Kratzer A. 2009. mRNA profiling for body fluid identification by reverse transcription endpoint PCR and real-time PCR. Forensic Science International: Genetics. 3: 80–88.

Haas C, Hanson E, Anjos MJ, Ballantyne KN, Banemann R, Bhoelai B, et al. 2014. RNA/DNA co-analysis from human menstrual blood and vaginal secretion stains: results of a fourth and fifth collaborative EDNAP exercise. Forensic Science International: Genetics. 8: 203–212.

Hackenberg M, Barturen G, Carpena P, Luque-Escamilla PL, Previti C, Oliver JL. 2010. Prediction of CpG-island function: CpG clustering vs. sliding-window methods. BioMedCentral Genomics. 11: 327.

Hahn XA, Wu X, Li AX, Hahn T, Pfeifer GP. 2011. Relationship between Gene Body DNA Methylation and Intragenic H3K9me3 and H3K36me3 Chromatin Marks. Public Library of Sciences One. 6: e18844.

Hamed M, Ismael S, Paulsen M, Helms V. 2012. Cellular Functions of Genetically Imprinted Genes in Human and Mouse as Annotated in the Gene Ontology. Public Library of Sciences One. 7: e50285.

Han L, Witmer PDW, Casey E, Valle D, Sukumar S. 2007. DNA methylation regulates microRNA expression. Cancer Biology and Therapy. 6: 1290-1294.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. Cell. 144: 646-674.

Hansen KD, Langmead B, Irizarry RA. 2012. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biology. 13: R83.

Hanson EK, Lubenow H, Ballantyne J. 2009. Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. Analytical Biochemistry. 387: 303–314.

Hardy TM, Tollefsbol TO. 2011. Epigenetic diet: impact on the epigenome and cancer. Epigenomics. 4: 503–518.

Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman MS. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. Nature. 405: 486-489.

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, et al. 2010. Comparison of sequencing based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nature Biotechnology. 28: 1097-1105.

Harrison J, Stirzaker C, Clark SJ. 1998. Cytosines adjacent to methylated CpG sites can be partially resistant to conversion in genomic bisulfite sequencing leading to methylation artefacts. Analytical Biochemistry. 264: 129–132.

Haverty PM, Weng Z, Best N, Auerbach K, Hsiao L, Jensen R, et al. 2002. Huge Index: A Database of High Density Oligonucleotide Array Data from Normal Human Tissues. Nucleic Acids Research. 1: 214-217.

He X, Kuang S, Liao J, Xu C, Chang Y, Wu Y, et al. 2015. The regulation of microRNA expression by DNA methylation in hepatocellular carcinoma. Molecular Biosystems. 11**:** 532–539.

Heegaard NH, Schetter AJ, Welsh JA, Yoneda M, Bowman ED, Harris CC. 2012. Circulating micro-RNA expression profiles in early stage non-small cell lung cancer. International Journal of Cancer. 130: 1378–1386.

Hesselink TA, Heideman DAM, Steenbergen RDM, Gok M, Van Kemenade FL, Wilting SM, et al. 2014. Methylation marker analysis of self-sampled cervico-vaginal lavage specimens to triage high-risk HPV-positive women for colposcopy. International Journal of Cancer. 135: 880–886.

Heyn H, Esteller M. 2012. DNA methylation profiling in the clinic: applications and challenges. Nature Reviews. 13: 679–692.

Heyn H, Moran S, Herraez IH, Sayols S, Gomez A, Sandovel J, et al. 2013. DNA Methylation contributes to natural human variation. Genome Research. 23: 1363–1372.

Hines LM, Stampfer MJ, Ma J, Gaziano M, Ridker PM, Hankinson SE, et al. 2001. Genetic Variation in Alcohol Dehydrogenase and the Beneficial Effect of Moderate Alcohol Consumption on Myocardial Infarction. New England Journal of Medicine. 344: 549–555.

Ho S, Tang W. 2007. Techniques used in studies of epigenome dysregulation due to aberrant DNA methylation: an emphasis on foetal-based adult diseases. Reproductive Toxicology. 23: 267–282.

Holliday R, Pugh JE. 1975. DNA Modification Mechanisms and Gene Activity during Development. Science. 187: 226–232.

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. BioMedCentral Bioinformatics. 13: 86.

Hsiao CH, Hsieh AR, Lian I, Lin YC, Wang HM, Fann CSJ. 2014. A Novel Method for Identification and Quantification of Consistently Differentially Methylated Regions. Public Library of Sciences. 9: e97513.

Hsiung DT, Marsit CJ, Houseman EA, Eddy K, Furniss CS, McClean MD, et al. 2007. Global DNA Methylation in whole blood as a biomarker in head and neck squamous cell carcinoma. Cancer Epidemiology, Biomarkers and Prevention. 16: 108–114.

Hu M, Yao J, Polyak K. 2006. Methylation-specific digital karyotyping. Nature Protocols. 1: 1621–1636.

Huang RS, Gamazon ER, Ziliak D, Wen Y, Im HK, Zhang W, et al. 2011. Population differences in microRNA expression and biological implications. RNA Biology. 8: 692–701.

Huh I, Zeng J, Park T, Yi SV. 2013. DNA Methylation and Transcriptional Noise. Epigenetics and Chromatin. 6: 1–10.

Hunter DJ. 2005. Gene-environment interactions in human disease. Nature Reviews: Genetics. 6: 287–298.

Huse SM, Ye Y, Zhou Y, Fodor AA. 2012. A core human microbiome as viewed through 16S rRNA sequence clusters. Public Library of Sciences One. 7: e34242.

Iannello RC, Young J, Sumarsono S, Tymms MJ, Dahl HH, Gould J, et al. 1997. Regulation of *Pdha-2* expression is mediated by proximal promoter sequences and CpG methylation. Molecular Cell Biology. 17: 612–619.

Igarashi J, Muroi S, Kawashima H, Wang X, Shinojima Y, Kitamura E, et al. 2008. Quantitative analysis of human tissue-specific differences in methylation. Biochemical and Biophysical Research Communications. 376: 658–664.

Illingworth RS, Bird A. 2009. CpG islands – A rough guide. Federation of European Biochemical Societies Letters. 583: 1713–1720.

Illingworth RS, Kerr A, DeSousa D, Jorgensen H, Ellis P, Stalker J, et al. 2008. A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci. Public Library of Sciences: Biology. 6: e22.

Irizarry RA, Ladd-Costa C, Wen B, Wu Z, Montano C, Onyango P, et al. 2009. Genome-wide methylation analysis of human colon cancer reveals similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature Genetics. 41: 178-186.

Ishida M, Moore GE. 2013. The role of imprinted genes in humans. Molecular Aspects of Medicine. 34: 826–840.

Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, et al. 2004. A tiling resolution DNA microarray with complete coverage of the human genome. Nature Genetics. 36: 299–303.

Jacob S. 1999. The Role of Micronutrients in DNA Synthesis and Maintenance. Advances in Experimental Medicine and Biology. 472: 101–113.

Jaenisch R, Bird A. 2003. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. Nature Genetics. 33: 245–254.

Jain S, Wojdacs TK, Su YH. 2013. Challenges for the application of DNA methylation biomarkers in molecular diagnostic testing for cancer. Expert Review of Molecular Diagnostics. 13: 283-294.

Jimenez-Chillaron JC, Diaz R, Martinez D, Pentinat T, Ramon-Krauel M, Ribo S, et al. 2012. The role of nutrition on epigenetic modifications and their implications on health. Biochimie. 94: 2242–2263.

Jones A, Teschendorff AE, Quanxi L, Hayward JD, Kannan A, Mould T, et al. 2013. Role of DNA Methylation and Epigenetic Silencing of *HAND2* in Endometrial Cancer Development. Public Library of Sciences: Medicine. 10: e1001551.

Jones PA. 2012. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nature Reviews: Genetics. 41: 484–492.

Jones PA, Takai D. 2001. The role of DNA methylation in mammalian epigenetics. Science. 293: 1067-1070.

Jones PA, Takai D. 2003. The CpG island searcher: a new WWW resource. In silico Biology. 3: 235–240.

Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. 2014. Metatranscriptomics of the human oral microbiome during health and disease. MBio. 5: e01012–e01014.

Juusola J, Ballantyne J. 2007. mRNA Profiling for Body Fluid Identification by Multiplex Quantitative RT-PCR. Journal of Forensic Sciences. 52: 1252–1262.

Kabesch M, Michel S, Tost J. 2010. Epigenetic mechanisms and the relationship to childhood asthma. European Respiratory Journal. 36: 950–961.

Kader F, Ghai M. 2015. DNA methylation and application in Forensic Science. Forensic Science International. 249: 255-265.

Kato K, Long N, Makita H, Tolda M, Yamashita T, Htayakema D, et al. 2008. Effects of green tea polyphenol on methylation status of *RECK* gene and cancer cell invasion in oral squamous cell carcinoma cells. British Journal of Cancer. 99: 647–654.

Kayser M, Kniff P. 2011. Improving human forensics through advances in genetics, genomics and molecular biology. Nature Reviews: Genetics. 12: 179–132.

Kilpinen H, Dermitzakis ET. 2012. Genetic and epigenetic contribution to complex traits. Human Molecular Genetics. 21: R24–R28.

King-Batoon A, Leszczynska JM, Klein CB. 2008. Modulation of gene methylation by genistein or lycopene in breast cancer cells. Environmental and Molecular Mutagenesis. 49: 36–45.

Kit AH, Nielsen HM, Tost J. 2012. DNA methylation based biomarkers: Practical considerations and applications. Biochimie. 94: 2314–2337.

Kitamura E, Igarashi J, Morohashi A, Hida N, Oinuma T, Nemoto N, et al. 2007. Analysis of tissue-specific differentially methylated regions (TDMs) in humans. Genomics. 89: 326–337.

Kobayashi H, Sakurai T, Miura F, Imai M, Mochiduki K, Yanagisawa E, et al. 2013. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. Genome Research. 23: 616–627.

Koch CM, Wagner W. 2011. Epigenetic-Ageing-signature to determine age in different tissues. Ageing. 3: 1–10.

Koops BJ, Schellekens M. 2008. Forensic DNA phenotyping: regulatory issues. The Columbia Science and Technology Law Review. 9: 158–202.

Kreil DP, Russell RR, Russell S. 2006. Microarray oligonucleotide probes. Methods in Enzymology. 410: 73-98.

Kriaucionis S, Heintz N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 324: 929-930.

Ku CS, Loy EY, Salim A, Pawitan Y, Chia KS. 2010. The discovery of human genetic variations and their use as disease markers: past, present and future. Journal of Human Genetics. 55: 403-415.

Kuczynski J, Costello EK, Nemergut DR, Zaneveld J, Lauber CL, Knights D, et al. 2010. Direct sequencing of the human microbiome readily reveals community differences. Genome Biology. 10: 210.

Kulis M, Queiros AC, Beekman R, Martin-Subero JI. 2013. Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. Biochimica et Biophysica Acta. 1829: 1161–1174.

Kulis M, Heath S, Bibikova M, Queiros AC, Navarro A, Clot G, et al. 2012. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukaemia. Nature Genetics. 44: 1236-1242.

Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, et al. 2006. CTCF binding at the *H19* imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *Igf2*. Proceedings of the National Academy of Sciences USA. 103: 10684–10689.

Kuzawa CW, Sweet E. 2009. Epigenetics and the embodiment of race: Developmental origins of US racial disparities in cardiovascular health. American Journal of Human Biology. 21: 2–15.

Kwabi-Addo B, Wang S, Chung W, Jelinek J, Patierno SR, Wang B, et al. 2010. Identification of Differentially Methylated Genes in Normal Prostate Tissues from African American and Caucasian Men. Clinical Cancer Research. 16: 3539–3547.

Laird PW. 2003. The power and the promise of DNA methylation markers. Nature Reviews: Cancer. 3: 253–266.

Lalruatfela B. 2013. On DNA methylation: An introductory review. Science Vision. 13: 1-7.

Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, et al. 2012. Factors underlying variable DNA methylation in a human community cohort. Proceedings of the National Academy of Sciences. 109: 17253–17260.

Lande-Diner L, Cedar H. 2005. Silence of the genes-Mechanisms of long-term repression. Nature Reviews: Genetics. 6: 648–654.

Lee HW, Park MJ, Choi A, An JH, Yang WI, Shin KJ. 2012. Potential forensic application of DNA methylation profiling to body fluid identification. International Journal of Legal Medicine. 126: 55–62.

Lee JT, Lu N. 1999. Targeted mutagenesis of Tsix leads to non-random X inactivation. Cell. 99: 47–57.

Lee JT, Davidow LS, Warshawsky D. 1999. Tsix, a gene antisense to Xist at the X-inactivation centre. Nature Genetics. 21: 400–404.

Levenson VV. 2010. DNA methylation as a universal biomarker. Expert Review of Molecular Diagnostics. 10: 481–488.

Levenson VV, Melnikov AA. 2012. DNA Methylation as Clinically Useful Biomarkers—Light at the End of the Tunnel. Pharmaceuticals. 5: 94–113.

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: 5275-5280.

Li H, Xu LL, Masuda LK, Raymundo E, McLeod D, Dobi A, et al. 2008. A feedback loop between the androgen receptor and a NEDD4-binding protein, PMEPA1, in prostate cancer cells. Journal of Biological Chemistry. 43: 28988-289895.

Li M, Chen WD, Papadopoulos N, Goodman SN, Bjerregaard NC, Laurberg S, et al. 2009. Sensitive digital quantification of DNA methylation in clinical samples. Nature Biotechnology. 27: 858–863.

Lieb LD, Beck S, Bulyk ML, Farnham P, Hattori N, Henikoff S, et al. 2006. Applying whole-genome studies of epigenetic regulation to study human disease. Cytogenic and Genome Research. 114: 1–15.

Lienert F, Wirbelaur C, Som I, Dean A, Mohn F, Schubeler D. 2011. Identification of genetic elements that autonomously determine DNA methylation states. Nature Genetics. 43: 1091-1097.

Lillycrop KA, Phillips ES, Torrens C, Hanson MA, Jackson AA, Burdge GC. 2008. Feeding pregnant rats a protein-restricted diet persistently alters the methylation of specific cytosines in the hepatic PPAR alpha promoter of the offspring. British Journal of Nutrition. 100: 278-282.

Lim DHK, Maher ER 2010. DNA Methylation: A form of epigenetic control of gene expression. The Obstetrician and Gynaecologist. 12: 37–42.

Lippman Z, Gendrel A, Black M, Vaughn MW, Dedhia N, McCombie WR, et al. 2006. Role of transposable elements in heterochromatin and epigenetic control. Nature. 430: 471–476.

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, et al. 2013. Global epigenomic reconfiguration during mammalian brain development. Science. 341: 1237905.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 462: 315–322.

Liu B, Lague JR, Nunes DP, Toselli P, Oppenheim FG, Soares RV, et al. 2002. Expression of membrane-associated mucins MUC1 and MUC4 in major human salivary glands. Journal of Histochemistry and Cytochemistry. 50: 811-820.

Liu D, Diorio J, Tannenbaum B, Caldji C, Francis D, Freedman A, et al. 1997. Maternal care, hippocampal glucocorticoid receptors, and hypothalamic–pituitary–adrenal responses to stress. Science. 277: 1659–1662.

Liu J, Morgan M, Hutchison K, Calhoun VD. 2010. A Study of the Influence of Sex on Genome Wide Methylation. Public Library of Sciences One. 5: e10028.

Liu X, Yu X, Zack DJ, Zhu H, Qian J. 2008. TiGER: A database for tissue-specific gene expression and regulation. BioMedCentral Bioinformatics. 9: 271.

Lohmueller KE, Pearce CL, Pike M, Lande ES, Nirschhorn JN. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nature Genetics. 33: 177–182.

Lokk K, Modhukur V, Rajashekar B, Martens K, Magi R, Kolde R, et al. 2014. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biology. 15: 1–14.

Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, Setien F, et al. 2007. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Research. 67: 1424–1429.

Lukiw WJ. 2013. Variability in micro RNA (miRNA) abundance, speciation and complexity amongst different human populations and potential relevance to Alzheimer's disease (AD). Frontiers in Cellular Neuroscience. 7: 133.

M'charek A, Toom V, Prainsack B. 2012. Bracketting off population does not advance ethical reflection on EVCs: a reply to Kayser and Schneider. Forensic Science International: Genetics. 6: e16–e17.

Ma LL, Yi SH, Xin-Huang D, Mei K, Yang RZ. 2013a. Screening and identification of tissue-specific methylation for body fluid identification. Forensic Science International: Genetics Supplement Series. 4: e37–e38.

Ma Y, Wang H, Jin H. 2013b. Methylated DNA and microRNA in Body Fluids as Biomarkers for Cancer Detection. International Journal of Molecular Sciences. 14: 10307–10331.

Macleod D, Charlton J, Mullins J, Bird AP. 1994. Spl sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. Genes and Development. 8: 2282-2292.

Madi T, Balamurugan K, Bombardi R, Duncan G, McCord B. 2012. The determination of tissue-specific DNA methylation patterns in forensic biofluids using bisulfite modification and pyrosequencing. Electrophoresis. 33: 1736–1745.

Malamud D, Rodriguez-Chavez IR. 2011. Saliva as a Diagnostic Fluid. Dental Clinics of North America. 55: 159–178.

Manucci, A, Sullivan KM, Ivanov PL, Gill P. 1994. Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous Amelogenin gene. International Journal of Legal Medicine. 106: 190–193.

Marioni RE, Shah S, McRae AF, Chen BH, Colicin E, Harris SE, et al. 2015. DNA methylation age of blood predicts all-cause mortality in later life. Genome Biology. 16: 25.

Martin PD, Cheshire SK. 1986. A comparative study of the citrate and lactate concentrations in stains from semen, vaginal secretion and mixtures of the two using isotachophoresis. Methodological and Biochemical Aspects. 1: 299-303.

Masemola ML, Van der Merwe L, Lombard Z, Viljoen D, Ramsay M. 2015. Reduced DNA methylation at the PEG3 DMR and KvDMR1 loci in children exposed to alcohol in utero: a South African Foetal Alcohol Syndrome cohort study. Frontiers in Genetics. 6: 85.

Mason JB, Choi SW. 2005. Effects of Alcohol on Folate Metabolism: Implications for Carcinogenesis. Alcohol. 35: 235–241.

Matarazzo MR, De-Bonis ML, Strazzullo M, Cerase A, Ferraro M, Vastarelli P, et al. 2007. Multiple binding of methyl-CpG and polycomb proteins in long-term gene silencing events. Journal of Cell Physiology. 210: 711-719.

Matatiele P, Tikly M, Tarr G, Gulumian M. 2015. DNA methylation similarities in genes of black South Africans with systemic lupus erythematosus and systemic sclerosis. Journal of Biomedical Science. 22: 34.

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 466: 253-257.

Mazzio EA, Soliman KFA. 2012. Basic concepts of epigenetics: Impact of environmental signals on gene expression. Epigenetics. 7: 119–130.

McGowan PO, Sasaki A, D'Alessio AC, Dymov S, Labonte B, Szyf M, et al. 2009. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nature Neuroscience. 12: 342–348.

Melnikov AA, Gartenhaus RB, Levenson AS, Motchoulskaia NA, Levenson VV. 2005. MSRE-PCR for analysis of gene-specific DNA methylation. Nucleic Acids Research. 33: 1–7.

Meyers RA. 2012. Epigenetic Regulation and Epigenomics: Advances in Molecular Biology and Medicine. John Wiley and Sons.

Mikeska T, Craig JM. 2014. DNA Methylation Biomarkers: Cancer and Beyond. Genes. 5: 821–864.

Moen EL, Zhang X, Mu W, Delaney SM, Wing C, McQuade J, et al. 2013. Genome-Wide Variation of Cytosine Modifications between European and African Populations and the Implications for Complex Traits. Genetics. 194: 987–996.

Mokarram P, Kumar K, Brim H, Naghibalhossaini F, Saberi-Firoozi M, Nouraie M, et al. 2009. Distinct high-profile methylated genes in colorectal cancer. Public Library of Sciences One. 4: e7012.

Mondini L, Noorani A, Pagnotta MA. 2009. Assessing Plant Genetic Diversity by Molecular Tools. Diversity. 1: 19-35.

Montgomery A, Feinberg P, Daniele-Fallin M, Multhaup ML, Jaffe AE, Joehanes R, et al. 2015. DNA methylation age of blood predicts all-cause mortality in later life. Genome Biology. 16: 25.

Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B. 2001. Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. Journal of Forensic Sciences. 46: 647–660.

Muise AM, Walters T, Wine E, Griffiths AM, Turner D, Duerr RH, et al. 2007. Protein-Tyrosine Phosphatase Sigma is Associated with Ulcerative Colitis. Current Biology. 17: 1212–1218.

Munson K, Clark J, Lamparska-Kupsik K, Smith SS. 2007. Recovery of bisulfite-converted genomic sequences in the methylation-sensitive QPCR. Nucleic Acids Research. 35: 2893–2903.

Munzel M, Globisch D, Carell T. 2011. 5-Hydroxymethylcytosine, the Sixth Base of the Genome. Angewandte Chemie International. 29: 6460-6468.

Naito E, Dewa K, Yamanouchi H, Takagi S, Kominami R. 1993. Sex determination using the hypomethylation of a human macro-satellite DXZ4 in female cells. Nucleic Acids Research. 21: 2533–2534.

Nakazono T, Kashimura S, Hayashiba Y, Hara K, Matsusu A, Augustin C. 2008. Dual Examinations for Urine as Being of Human Origin and for DNA-Typing from Small Stains of Human Origin. Journal of Forensic Sciences. 53: 359–363.

Naumova MY, Lee M, Koposov R, Szyf M, Dozier M, Grigorenko EL. 2012. Differential patterns of whole-genome DNA methylation in institutionalized children and children raised by their biological parents. Development and Psychopathology. 24: 143–155.

Niculescu MD, Zeisel S. 2002. Diet, methyl donors and DNA methylation: interactions between dietary folate, methionine and choline. Journal of Nutrition. 132: 2333S-2335S.

Nunes AP, Oliveira IO, Santos BR, Millech C, Silva LP, Gonzalez DA, et al. 2012. Quality of DNA extracted from saliva samples collected with the Oragene™ DNA self-collection kit. BMC Medical Research Methodology. 12: 65.

Nussbaumer C, Gharehbaghi-Schnell E, Korschineck I. 2006. Messenger RNA profiling: a novel method for body fluid identification by real-time PCR. Forensic Science International. 157: 181-186.

Ohgane J, Yagi S, Shiota K. 2008. Epigenetics: The DNA Methylation Profile of Tissue-Dependent and Differentially Methylated Regions in Cells. Placenta (Supplement A, Trophoblast Research). 29: S29-S35.

Okuizumi H, Noguchi T, Saguchi T, Eri T, Shinsuke-Yamanaka N, Kombate K, et al. 2010. Application of the Restriction Landmark Genome Scanning (RLGS) Method for Analysis of Genetic Diversity between Asian and African Sorghum. Electronic Journal of Plant Breeding. 1: 1144–1147.

Pandeshwar P, Das R. 2014. Role of oral fluids in DNA investigations. Journal of Forensic and Legal Medicine. 22: 45–50.

Park JL, Kwon OH, Kim JH, Yoo HS, Lee HC, Woo KM, et al. 2014a. Identification of body fluid-specific DNA methylation markers for use in forensic science. Forensic Science International: Genetics. 13: 147–153.

Park JL, Park SM, Kwon OH, Lee HC, Kim J, Seok HH, et al. 2014b. Microarray screening and qRTPCR evaluation of microRNA markers for forensic body fluid identification. Electrophoresis. 35: 21 – 22.

Park L, Friso S, Choi S. 2011. Nutritional Influences on Epigenetics and Age-Related Disease. The Proceedings of the Nutritional Society. 1: 75–83.

Parson W, Steinlechner M. 2001. Efficient DNA database laboratory strategy for high through-put STR typing of reference samples. Forensic Science International. 122: 1-6.

Pavlova SI, Tao L. 2000. Induction of vaginal *Lactobacillus* phages by the cigarette smoke chemical Benzo-[a]-pyrene-diol epoxide. Mutation Research. 466: 57–62.

Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, Lord RV, et al. 2015. De novo identification of differentially methylated regions in the human genome. Epigenetics and Chromatin. 8: 6.

Petryszak R, Burdett T, Fiorelli B, Fonseca NA, Gonzalez-Porta M, Hastings E, et al. 2014. Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. Nucleic Acids Research. 42 (Database Issue): D926–D932.

Philibert RA, Zadorozhnyaya O, Beach SR, Brody GH. 2008. Comparison of the genotyping results using DNA obtained from blood and saliva. Psychiatric Genetics. 18: 275–281.

Pinney SE. 2014. Mammalian Non-CpG Methylation: Stem Cells and Beyond. Biology (Basel). 3: 739-751.

Pirazzini C, Giulana C, Bacalini MG, Boattini A, Capri M, Fontanesi E, et al. 2012. Space/Population and Time/Age in DNA methylation variability in humans: a study on IGF2/H19 locus in different Italian populations and in mono- and di-zygotic twins of different age. Ageing. 4: 509–520.

Ponger L, Mouchiroud D. 2002. CpGProD: identifying CpG islands associated with transcription start sites in large genomic mammalian sequences. Bioinformatics. 4: 631-633.

Prokunina-Olsson L, Welch C, Hansson O, Adhikari N, Scott LJ, Usher N, et al. 2009. Tissue-specific alternative splicing of TCF7L2. Human Molecular Genetics. 18: 3795–3804.

Quarino L, Dang Q, Hartmann J, Moynihan N. 2005. An ELISA method for the identification of salivary amylase. Journal of Forensic Sciences. 50: 873-876.

Quinones I, Daniels B. 2011. Cell free DNA as a component of forensic evidence recovered from touched surfaces. Forensic Science International: Genetics. 6: 26–30.

Rajshekar M, Tennant M, Thejaswini BDS. 2013. Salivary biomarkers and their applicability in forensic identification. Sri Lanka Journal of Forensic Medicine, Science and Law. 4: 10-15.

Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. 2002. Metastable epialleles in mammals. Trends in Genetics. 18: 348–351.

Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, et al. 2008. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). Genome Research. 18: 1518–1529.

Rangel-Villalobos H, Martinez-Sevilla VM, Salazar-Flores J, Martinez-Cortez G, Munoz-Valle JF, Galaviz-Hernandez C, et al. 2013. Forensic parameters for 15 STRs in eight Amerindian populations from the north and west of Mexico. Forensic Science International: Genetics. 7: e62–e65.

Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP. 2009. A human B cell methylome at 100 bp resolution. Proceedings of the National Academy of Sciences USA. 106: 671–678.

Rawlings-Goss RA, Campbell MC, Tishkoff SA. 2014. Global population-specific variation in miRNA associated with cancer risk and clinical biomarkers. BioMedCentral Medical Genomics. 7: 53.

Rawson JB, Bapat B. 2012. Epigenetic biomarkers in colorectal cancer diagnostics. Expert Review of Molecular Diagnostics. 12: 499-509.

Reeben M, Prydz H. 1994. An improved method for detection of 5-methylcytosine by PCR-based genomic sequencing. BioTechniques. 16: 416–417.

Reik W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. Nature. 447: 435–432.

Rein T, Zorbas H, DePamphilis ML. 1997. Active mammalian replication origins are associated with a high-density cluster of mCpG dinucleotides. Molecular Cell Biology. 17: 416–426.

Reitz C, Jun GA, Naj R, Rajbhandary BN, Vardarajan LS, Wang LS, et al. 2013. Variants in the ATP-binding cassette transporter (ABCA7), Apolipoprotein E4, and the risk of late-onset Alzheimer's disease in African Americans. Journal of the American Medical Association. 309: 1483–1492.

Remenyi A, Scholer HR, Wiliams M. 2004. Combinatorial control of gene expression. Nature, Structural and Molecular Biology. 11: 812-815.

Rienius LE, Acevedo N, Joerink M, Pershagen G, Dahlen S, Grecol D, et al. 2012. Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. Public Library of Sciences One. 7: e41361.

Rifai N, Gillette MA, Carr SA. 2006. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nature Biotechnology. 24: 971–983.

Riggs AD. 1975. X Inactivation, Differentiation and DNA Methylation. Cytogenetics and Cell Genetics. 14: 9-25.

Rijlaarsdam MA, Van der Zwan YG, Dorssers LCJ, Looijenga LHJ. 2014. DMRforPairs: Identifying Differentially Methylated Regions between unique samples using array based methylation profiles. BioMed Central Bioinformatics. 15: 141.

Risch N, Burchard E, Ziv E, Tang H. 2002. Categorization of humans in Biomedical Research: Genes, Race and Disease. Genome Biology. 3: 7.

Rogers NL, Cole SA, Lan HC, Crossa A, Demerath EW. 2007. New saliva DNA collection method compared to buccal cell collection techniques for epidemiological studies. American Journal of Human Biology. 19: 319–326.

Rollins RA, Haghighi F, Edwards JR, Das R, Zhang MQ, Ju J, et al. 2006. Large-scale structure of genomic methylation patterns. Genome Research. 16: 157-163.

Romeika JM, Yan F. 2013. Recent Advances in Forensic DNA Analysis. Journal of Forensic Research. 12: 1–13.

Romero-Montoya L, Martinez-Rodriguez H, Perez MA, Arguello-Garcia R. 2011. Relationship of spermatoscopy, prostatic acid phosphatase activity and prostate-specific antigen (p30) assays with further DNA typing in forensic samples from rape cases. Forensic Science International. 206: 111-118.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. 2002. Genetic Structure of Human Populations. Science. 298: 2831–2835.

Roussignol G, Ango F, Romorini S, Tu JC, Sala C, Worley PF, et al. 2005. Shank expression is sufficient to induce functional dendritic spine synapses in aspiny neurons. Journal of Neuroscience. 25: 3560–3570.

Rush LJ, Plass C. 2002. Restriction landmark genomic scanning for DNA methylation in cancer: past, present, and future applications. Analytical Biochemistry. 307: 191–201.

Ryan BM, Robles AI, Harris CC. 2010. Genetic variation in microRNA networks: the implications for cancer research. Nature Reviews: Cancer. 10: 389–402.

Saito Y, Liang G, Egger G, Friedman JM, Chuang JC, Coetzee GA, et al. 2006. Specific activation of microRNA-127 with downregulation of the proto-oncogene *BCL6* by chromatin-modifying drugs in human cancer cells. Cancer Cell. 9: 435–443.

Sakatani T, Kaneda A, Iacobuzio-Donahue CA, Carter MG, de Boom Witzel S, Okano H, et al. 2005. Loss of imprinting of *Igf2* alters intestinal maturation and tumorigenesis in mice. Science. 307: 1976-1978.

Sakurada K, Akutsu T, Watanabe K, Yoshino M. 2012. Identification of nasal blood by real-time RT-PCR. Legal Medicine (Tokyo). 14: 201-204.

Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, et al. 2006. A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis. 9: 1713-1724.

Sarter B, Long TI, Tsong WH, Koh WP, Yu MC, Laird PW. 2005. Sex differential in methylation patterns of selected genes in Singapore Chinese. Human Genetics. 117: 402-403.

Sauer E, Madea B, Courts C. 2014. An evidence based strategy for normalization of quantitative PCR data from miRNA expression analysis in forensically relevant body fluids. Forensic Science International: Genetics. 11: 174–181.

Schaukowitch K, Kim KT. 2014. Emerging Epigenetic Mechanisms of Long Non-Coding RNAs. Neuroscience. 264: 25–38.

Schilling E, Rehli M. 2008. Global, comparative analysis of tissue-specific promoter CpG methylation. Genomics. 90: 314–323.

Schneider E, Pliushch G, El Hajj N, Galetzka D, Puhl A, Schorsch M, et al. 2010. Spatial, temporal and interindividual epigenetic variation of functionally important DNA methylation patterns. Nucleic Acids Research. 38: 3880-3890.

Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert Jr CJ. 2005. Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biology. 6: R33.

Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, et al. 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature. 523: 212–216.

Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, Yau P, et al. 2006. Microarray-based DNA methylation profiling: technology and applications. Nucleic Acids Research. 34: 528-542.

Shen L, Kondo Y, Guo Y, Zhang L, Zhang L, Ahmed A, et al. 2007. Genome-Wide Profiling of DNA Methylation Reveals a Class of Normally Methylated CpG Island Promoters. Public Library of Sciences: Genetics. 3: 2023–2036.

Shen Y, Mutsino Y, Fouse SD, Rao N, Root S, Xu R, et al. 2008. X-inactivation in female human embryonic stem cells is in a non-random pattern and prone to epigenetic alterations. Proceedings of the National Academy of Sciences. 105: 4709–4717.

Sijen T. 2014. Molecular approaches for forensic cell type identification: On mRNA, miRNA, DNA methylation and microbial markers. Forensic Science International: Genetics. 18: 21–32.

Silva JM, Dominguez G, Garcia JM, Gonzalez R, Villanueva MJ, Navarro F, et al. 1999. Presence of tumour DNA in plasma of breast cancer patients: Clinicopathological correlations. Cancer Research. 59: 3251–3256.

Simon R. 2005. Roadmap for developing and validating therapeutically relevant genomic classifiers. Journal of Clinical Oncology. 23: 7332-7341.

Slieker RC, Bos SD, Goeman JJ, Bovee JVMG, Talens RP, Breggen RVD, et al. 2013. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. Epigenetics and Chromatin. 6: 1-12.

Smiraglia DJ, Kazhiyur-Mannar R, Oakes CC, Wu YZ, Liang P, Ansari T, et al. 2007. Restriction landmark genomic scanning (RLGS) spot identification by second generation virtual RLGS in multiple genomes with multiple enzyme combinations. BioMed Central Genomics. 8: 446.

Smith AK, Kilaru V, Klengel T, Mercer KB, Bradley B, Conneely KN, et al. 2014. DNA Extracted From Saliva for Methylation Studies of Psychiatric Traits: Evidence Tissue Specificity and Relatedness to Brain. American Journal of Medical Genetics (Part B). 9999B: 1–9.

Song F, Smith JK, Kimura MT, Morrow AD, Matsuyama T, Nagase H, et al. 2005. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proceedings of the National Academy of Sciences USA. **102:** 3336–3341.

Sood P, Krek A, Zavolan M, Macino G, Rajewsky N. 2006. Cell-type-specific signatures of microRNAs on target mRNA expression. Proceedings of the National Academy of Sciences USA. 103: 2746–2751.

Stanssens P, Zabeau M, Meersseman G, Remes G, Gansemans Y, Storm N, et al. 2004. High-Throughput MALDI-TOF Discovery of Genomic Sequence Polymorphisms. Genome Research. 14: 126–133.

Stirzaker C, Millar DS, Paul CL, Warnecke PM, Harrison J, Vincent PC, et al. 1997. Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumours. Cancer Research. 57: 2229–2237.

Storey JD, Madeoy J, Strout JL, Wurfel, M, Ronald J, Akey JM. 2007. Gene-Expression Variation Within and Among Human Populations. The American Journal of Human Genetics. 80: 502-509.

Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science. 9: 848–853.

Straughen JK, Sipahi L, Uddin M, Misra DP, Misra VK. 2015. Racial differences in *Igf1* methylation and birth weight. Clinical Epigenetics. 7: 47.

Sugimoto K, Koh E, Sin HS, Maeda Y, Narimoto K, Izumi K, et al. 2009. Tissue-specific differentially methylated regions of the human *VASA* gene are potentially associated with maturation arrest phenotype in the testis. Journal of Human Genetics. 54: 450–456.

Sullivan KM, Manucci A, Kimpton CP, Gill P. 1993. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. BioTechniques. 4: 636–638.

Sun M, Abdollah F, Liberman D, Abdo A, Thuret R, Tian Z, et al. 2011. Racial disparities and socioeconomic status in men diagnosed with testicular germ cell tumours: a survival analysis. Cancer. 117: 4277–4285.

Sutherland E, Coe L, Raleigh EA. 1992. McrBC: a multisubunit GTP dependent restriction endonuclease. Journal of Molecular Biology. 225: 327-348.

Tabak LA. 2001. A Revolution in Biomedical Assessment: The Development of Salivary Diagnostics. Journal of Dental Education. 65: 1335-1339.

Taguchi Y. 2013. MicroRNA-mediated regulation of target genes in several brain regions is correlated to both microRNA-targeting-specific promoter methylation and differential microRNA expression. BioData Mining. 6: 11.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 24: 930-935.

Takai D, Jones PA. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proceedings of the National Academy of Sciences USA. 99: 3740–3745.

Tammen SA, Friso S, Choi SW. 2013. Epigenetics: The link between nature and nurture. Molecular Aspects of Medicine. 34: 753–764.

Tang H, Peng J, Wang P, Risch NJ. 2005. Estimation of Individual Admixture: Analytical and Study Design Considerations. Genetic Epidemiology 28: 289–301.

Terry MB, Ferris JS, Pilsner R, Flom JD, Tehranifar P, Sentellar RM, et al. 2008. Genomic DNA Methylation among Women in a Multiethnic New York City Birth Cohort. Cancer Epidemiology, Biomarkers and Prevention. 17: 2306–2310.

Thompson TM, Sharfi D, Lee M, Yrigolen CM, Naumova OY, Grigorenko EL. 2013. Comparison of Whole-Genome DNA Methylation Patterns in Whole Blood, Saliva, and Lymphoblastoid Cell Lines. Behaviour Genetics. 43: 168–176.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, et al. 2010. The Genetic Structure and History of Africans and African Americans. Science. 324: 1035–1044.

Tost J. 2010. DNA Methylation: An Introduction to the Biology and the Disease-Associated Changes of a Promising Biomarker. Molecular Biotechnology. 44: 71–81.

Tost J, Gut IG. 2007. Analysis of gene-specific DNA methylation patterns by pyrosequencing technology. Methods in Molecular Biology. 373: 89-102.

Tycko B, Morrison IM. 2002. Physiological Functions of Imprinted Genes. Journal of Cellular Physiology. 192: 245–258.

VanderKraats ND, Hiken JF, Decker KF, Edwards JR. 2013. Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. Nucleic Acids Research. 14: 6816–6827.

Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, et al. 2013. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Research. 23: 555–567.

Vatolin S, Abdullaev Z, Pack SD, Flanagan PT, Custer M, Loukinov DI, et al. 2005. Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of *MAGE-A1* and reactivation of other cancer-testis genes. Cancer Research. 65: 7751–7762.

Vennemann M, Koppelkamm A. 2010. mRNA Profiling in Forensic Genetics: Possibilities and limitations. Forensic Science International. 203: 71-75.

Vidaki A, Daniel B, Court DS. 2013. Forensic DNA methylation profiling – Potential opportunities and challenges. Forensic Science International: Genetics. 7: 499–507.

Virkler K, Lednev IK. 2009. Analysis of body fluids for forensic purposes: From laboratory testing to non-destructive rapid confirmatory identification at a crime scene. Forensic Science International. 188: 1–17.

Visser M, Zubakov D, Ballantyne KN, Kayser M. 2011. mRNA-based skin identification for forensic applications. International Journal of Legal Medicine. 125: 253–263.

Waddington CH. 1942. The epigenotype. Endeavour. 1: 18–20.

Waddington CH. 1968. Towards a Theoretical Biology; The Basic Ideas of Biology Edinburgh, Scotland: Edinburgh University Press.

Wan J, Oliver VF, Wang G, Zhu H, Zack DJ, Merbs SL, et al. 2015. Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. BioMedCentral Genomics. 16: 49.

Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, et al. 2012. IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinformatics. 28: 729-730.

Wang L, Wang X. 2015. DEGseq – Differentially Expressed Gene Identification for RNA-seq data. R Package Version 1.24.0. Available at http://www.bioconductor.org/packages/3.3/bioc/manuals/DEGseq/man/DEGseq.pdf. [Accessed 24 September 2015].

Wang W, Zhong Q, Teng L, Bhatnagar N, Sharma B, Zhang X, et al. 2014. Mutations that disrupt PHOXB interaction with the neuronal calcium sensor HPCAL1 impede cellular differentiation in neuroblastoma. Oncogene. 33: 3316–3324.

Wang Y, Leung FCC. 2004. An evaluation of new criteria for CpG islands in the human genome as gene markers. Bioinformatics. 20: 1170–1177.

Wang Z, Zhang J, Lou H, Ye Y, Yan J, Hou Y. 2013. Screening and confirmation of miRNA markers for forensic body fluid identification. Forensic Science International: Genetics. 7: 116–123.

Wasserstrom A, Frumkin D, Davidson A, Schpitzen M, Herman Y, Gafny R. 2013. Demonstration of DSI-semen – A novel DNA methylation-based forensic semen identification assay. Forensic Science International: Genetics. 7: 136–142.

Waterland RA, Jirtle RL. 2003. Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. Molecular and Cellular Biology. 23: 5293–5300.

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, et al. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genetics. 39: 457–466.

Wentzensen N, Bakkum-Gamez JN, Killian JK, Sampson J, Guido R, Glass A, et al. 2014. Discovery and validation of methylation markers for endometrial cancer. International Journal of Cancer. 135: 1860–1868.

Wickenheiser RA. 2011. Trace DNA: A review, Discussion of Theory, and Application of the Transfer of Trace Quantities of DNA through Skin Contact. Journal of Forensic Sciences. 47: 442–450.

Wild L, Flanagan JM. 2010. Genome-wide hypomethylation in cancer may be a passive consequence of transformation. Biochimica et Biophysica Acta: Reviews on Cancer. 1806: 50–57.

Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MF, et al. 2001. Population genetic structure of variable drug response. Nature Genetics. 29: 265-269.

Winnefeld M, Brueckner B, Gronniger E, Stab F, Wenck H, Lyko F. 2012. Stable Ethnic Variations in DNA Methylation Patterns of Human Skin. Journal of Investigative Dermatology. 132: 466–468.

Witherspoon DJ, Wooding S, Rogers AR, Marchani EE, Watkins WS, Batzer MA, et al. 2007. Genetic Similarities Within and Between Human Populations. Genetic Society of America. 176: 351 – 359.

Witkin SS, Linhares IM, Giraldo P. 2007. Bacterial flora of the female genital tract: Function and Immune Regulation. Best Practice and Research Clinical Obstetrics and Gynaecology. 21: 347-354.

Woodfine K, Huddleston JE, Murrell A. 2011. Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. Epigenetics and Chromatin. 4: 1.

Wu G, Gu J, Zhang MK. 2013. FastDMA: An Infinium HumanMethylation 450 BeadChip Analyzer. Public Library of Sciences One. 8: e74275.

Wutz A, Jaenisch RA. 2000. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Molecular Cell. 5: 695–705.

Xiang N, Zhao R, Song G, Zhong W. 2008. Selenite reactivates silenced genes by modifying DNA methylation and histones in prostate cancer cells. Carcinogenesis. 29: 2175–2181.

Xu Y, Xie J, Cao Y, Zhou H, Ping Y, Chen L, et al. 2014. Development of highly sensitive and specific mRNA multiplex system (XCYR1) for forensic human body fluids and tissues identification. Public Library of Sciences One. 9: e100123.

Yamada Y, Watanabe H, Miura F, Soejima H, Uchiyama M, Iwasaka T, et al. 2004. A Comprehensive Analysis of Allelic Methylation Status of CpG Islands on Human Chromosome 21q. Genome Research. 14: 247–266.

Yan J, Zierath JR, Barres R. 2011. Evidence for non-CpG methylation in mammals. Experimental Cell Research. 317: 2555-2561.

Yoo CB, JonesPA. 2006. Epigenetic therapy of cancer: Past, present and future. Nature Reviews: Drug Discovery. 5: 37–50.

Yu D, Ware C, Waterland RA, Zhang J, Chen M, Kunde-Ramamoorthy MGG, et al. 2013. Developmentally Programmed 3’ CpG Island Methylation Confers Tissue- and Cell-Type-Specific Transcriptional Activation. Molecular and Cell Biology. 33: 1845–1858.

Zampieri M, Ciccarone F, Calabrese R, Franceschi C, Burkle A, Caiafa P. 2015. Reconfiguration of DNA methylation in ageing. Mechanisms of Ageing and Development. In Press.

Zbiec-Piekarska R, Spolnicka M, Kupiec T, Makowska Z, Spas A, Parys-Proszek A, et al. 2015. Examination of DNA Methylation Status of the *ELOVL2* marker May Be Useful for Human Age Prediction in Forensic Science. Forensic Science International: Genetics. 14: 161–167.

Zhang FF, Carderelli R, Carroll J, Fulda KG, Kaur M, Gonzalez K, et al. 2011a. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. Epigenetics. 6: 623–629.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, et al. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell. 126: 1189–1201.

Zhang Y, Liu H, Liv J, Xiao X, Zhu J, Liu X, Su J, et al. 2011b. QDMR: a quantitative method for identification of differentially methylated regions by entropy. Nucleic Acids Research. 39: e58.

Zhao Y, Bhattacharjee S, Jones BM, Dua P, Alexandrov PN, Hill JM, et al. 2013. Regulation of TREM2 expression by an NF-eB-sensitive miRNA-34a. Neuroreport. 24: 318–323.

Zhao H, Shen J, Medico L, Wang D, Ambrosone CB, Liu S. 2010. A pilot study of circulating miRNAs as potential biomarkers of early stage breast cancer. Public Library of Sciences One. 5: e13735.

Zhu ZZ, Hou L, Bollati V, Tarantini L, Marinelli B, Cantone L, et al. 2012. Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. International Journal of Epidemiology. 41: 126–139.

Zilberman D, Henikoff S. 2007. Genome-wide analysis of DNA methylation patterns. Development. 134: 3959–3965.

Zubakov D, Boersma AW, Choi Y, Van Kuijk PF, Wiemer EA, Kayser M. 2010. MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. International Journal of Legal Medicine. 124: 217-226.

Zubakov D, Hanekamp E, Kokshoorn M, Van Ijcken W, Kayser M. 2008. Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples. International Journal of Legal Medicine. 122: 135–142.

Zubakov D, Kokshoorn M, Kloosterman A, Kayser M. 2011. New markers for old stains: Stable mRNA markers for blood and saliva identification from up to 16-year-old stains. International Journal of Legal Medicine. 123: 71–74.

APPENDIX A

CGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTT
AGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGA
CCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCT
TCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGT
GGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGA
TACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCT
ACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGG
GTTG

Figure 1: Sequence of the artificial DNA template obtained from PCR amplification of pCR®2.1 TOPO® vector. Primers for template amplification are in blue. MSRE-PCR primers to confirm complete digestion with *Hha*I enzyme are in yellow. Enzyme recognition site in the MSRE-PCR amplicon is in green.

Table 1: Genes in which CGIs were not positioned.

ENSG00000000971	ENSG00000085465	ENSG00000109205
ENSG00000001561	ENSG00000085514	ENSG00000109208
ENSG00000004939	ENSG00000086205	ENSG00000109272
ENSG00000005844	ENSG00000086570	ENSG00000109471
ENSG00000006534	ENSG00000086991	ENSG00000109684
ENSG00000007952	ENSG00000088002	ENSG00000109846
ENSG00000011143	ENSG00000089356	ENSG00000110077
ENSG00000011465	ENSG00000090020	ENSG00000110104
ENSG00000011600	ENSG00000090382	ENSG00000110195
ENSG00000012223	ENSG00000091138	ENSG00000110203
ENSG00000012822	ENSG00000092295	ENSG00000110484
ENSG00000014257	ENSG00000094796	ENSG00000110852
ENSG00000017427	ENSG00000095303	ENSG00000110876
ENSG00000017483	ENSG00000096006	ENSG00000111331
ENSG00000019582	ENSG00000099721	ENSG00000111348
ENSG00000021826	ENSG00000100055	ENSG00000111358
ENSG00000023445	ENSG00000100196	ENSG00000111361
ENSG00000025434	ENSG00000100342	ENSG00000111404
ENSG00000026652	ENSG00000100365	ENSG00000111704
ENSG00000039139	ENSG00000100368	ENSG00000111729
ENSG00000042832	ENSG00000100373	ENSG00000111796
ENSG00000043462	ENSG00000100448	ENSG00000111846
ENSG00000057149	ENSG00000100450	ENSG00000112077
ENSG00000057657	ENSG00000100453	ENSG00000112079
ENSG00000058085	ENSG00000101109	ENSG00000112303
ENSG00000058668	ENSG00000101213	ENSG00000113522
ENSG00000059804	ENSG00000101266	ENSG00000114638
ENSG00000060140	ENSG00000101307	ENSG00000115165
ENSG00000064201	ENSG00000101441	ENSG00000115523
ENSG00000064545	ENSG00000101916	ENSG00000115607
ENSG00000065615	ENSG00000102384	ENSG00000115919
ENSG00000065618	ENSG00000102837	ENSG00000115956
ENSG00000070526	ENSG00000102962	ENSG00000116701
ENSG00000070915	ENSG00000103569	ENSG00000116785
ENSG00000072042	ENSG00000103671	ENSG00000117595
ENSG00000073282	ENSG00000104921	ENSG00000117600
ENSG00000073803	ENSG00000105141	ENSG00000118217
ENSG00000074706	ENSG00000105341	ENSG00000118520
ENSG00000075884	ENSG00000105374	ENSG00000118640
ENSG00000077984	ENSG00000105501	ENSG00000119535
ENSG00000078098	ENSG00000105552	ENSG00000119922
ENSG00000079257	ENSG00000105668	ENSG00000119943
ENSG00000079263	ENSG00000105967	ENSG00000120915
ENSG00000079385	ENSG00000106331	ENSG00000121552
ENSG00000080986	ENSG00000106809	ENSG00000121807
ENSG00000081237	ENSG00000106819	ENSG00000121858
ENSG00000082074	ENSG00000107165	ENSG00000122122
ENSG00000083782	ENSG00000108405	ENSG00000133048
ENSG00000084110	ENSG00000108759	ENSG00000133063
ENSG00000085265	ENSG00000109063	ENSG00000133110
ENSG00000122133	ENSG00000128335	ENSG00000133195
ENSG00000122180	ENSG00000128340	ENSG00000133574

ENSG00000122224
ENSG00000123338
ENSG00000123405
ENSG00000123843
ENSG00000124102
ENSG00000124107
ENSG00000124157
ENSG00000124233
ENSG00000124469
ENSG00000124664
ENSG00000124701
ENSG00000124721
ENSG00000124731
ENSG00000124935
ENSG00000125780
ENSG00000125910
ENSG00000126233
ENSG00000126262
ENSG00000126353
ENSG00000126549
ENSG00000126550
ENSG00000126759
ENSG00000127074
ENSG00000127084
ENSG00000127325
ENSG00000127507
ENSG00000127951
ENSG00000128322
ENSG00000170006
ENSG00000170180
ENSG00000170255
ENSG00000170367
ENSG00000170369
ENSG00000170373
ENSG00000170465
ENSG00000170477
ENSG00000170523
ENSG00000171049
ENSG00000171051
ENSG00000171053
ENSG00000171195
ENSG00000171199
ENSG00000171201
ENSG00000171209
ENSG00000178776
ENSG00000179593
ENSG00000179639
ENSG00000179869
ENSG00000143185
ENSG00000143217
ENSG00000143226
ENSG00000143297
ENSG00000143314
ENSG00000143341
ENSG00000143369
ENSG00000128383
ENSG00000128422
ENSG00000128510
ENSG00000128513
ENSG00000129460
ENSG00000130176
ENSG00000131050
ENSG00000131355
ENSG00000131686
ENSG00000131738
ENSG00000131746
ENSG00000132109
ENSG00000132297
ENSG00000132517
ENSG00000132704
ENSG00000132746
ENSG00000132965
ENSG00000140067
ENSG00000140379
ENSG00000140506
ENSG00000140749
ENSG00000140798
ENSG00000141378
ENSG00000141469
ENSG00000141499
ENSG00000141756
ENSG00000141968
ENSG00000142512
ENSG00000142515
ENSG00000142619
ENSG00000142676
ENSG00000143105
ENSG00000143110
ENSG00000143119
ENSG00000143178
ENSG00000171346
ENSG00000171401
ENSG00000171711
ENSG00000172232
ENSG00000172243
ENSG00000172331
ENSG00000172349
ENSG00000172543
ENSG00000176024
ENSG00000176714
ENSG00000176920
ENSG00000177138
ENSG00000177954
ENSG00000153029
ENSG00000153233
ENSG00000154451
ENSG00000154589
ENSG00000155561
ENSG00000156234
ENSG00000156265
ENSG00000133710
ENSG00000133742
ENSG00000134200
ENSG00000134256
ENSG00000134489
ENSG00000134588
ENSG00000134757
ENSG00000134760
ENSG00000134765
ENSG00000135114
ENSG00000135218
ENSG00000135374
ENSG00000135413
ENSG00000135443
ENSG00000135749
ENSG00000135898
ENSG00000135899
ENSG00000135914
ENSG00000136235
ENSG00000136250
ENSG00000136688
ENSG00000136689
ENSG00000136929
ENSG00000137440
ENSG00000137462
ENSG00000137634
ENSG00000137648
ENSG00000137673
ENSG00000137674
ENSG00000137752
ENSG00000137975
ENSG00000137976
ENSG00000138271
ENSG00000138435
ENSG00000138615
ENSG00000138964
ENSG00000172867
ENSG00000173431
ENSG00000173612
ENSG00000174130
ENSG00000174226
ENSG00000174444
ENSG00000174460
ENSG00000174502
ENSG00000175065
ENSG00000175393
ENSG00000175793
ENSG00000176009
ENSG00000163406
ENSG00000163464
ENSG00000163563
ENSG00000163736
ENSG00000163737
ENSG00000163810
ENSG00000163823

ENSG00000143374
ENSG00000143409
ENSG00000143486
ENSG00000143536
ENSG00000143546
ENSG00000143556
ENSG00000143624
ENSG00000143851
ENSG00000144061
ENSG00000144481
ENSG00000145649
ENSG00000145879
ENSG00000146094
ENSG00000146143
ENSG00000146411
ENSG00000146859
ENSG00000147162
ENSG00000147168
ENSG00000147606
ENSG00000147689
ENSG00000147873
ENSG00000148215
ENSG00000148346
ENSG00000148604
ENSG00000149021
ENSG00000149516
ENSG00000149781
ENSG00000149968
ENSG00000150540
ENSG00000152266
ENSG00000152931
ENSG00000179934
ENSG00000179941
ENSG00000180353
ENSG00000180785
ENSG00000180871
ENSG00000181631
ENSG00000182165
ENSG00000182795
ENSG00000182885
ENSG00000183036
ENSG00000183486
ENSG00000183542
ENSG00000184148
ENSG00000184330
ENSG00000184357
ENSG00000184564
ENSG00000184635
ENSG00000184730
ENSG00000184752
ENSG00000185186
ENSG00000197641
ENSG00000198077
ENSG00000198223
ENSG00000198523
ENSG00000157017
ENSG00000158373
ENSG00000158578
ENSG00000158786
ENSG00000158825
ENSG00000158869
ENSG00000159167
ENSG00000159173
ENSG00000159182
ENSG00000159337
ENSG00000159339
ENSG00000159763
ENSG00000160336
ENSG00000160349
ENSG00000160593
ENSG00000160862
ENSG00000160883
ENSG00000162078
ENSG00000162511
ENSG00000162614
ENSG00000162645
ENSG00000162747
ENSG00000162892
ENSG00000162896
ENSG00000163009
ENSG00000163017
ENSG00000163116
ENSG00000163131
ENSG00000163207
ENSG00000163209
ENSG00000163219
ENSG00000163220
ENSG00000163221
ENSG00000185479
ENSG00000185862
ENSG00000186081
ENSG00000186130
ENSG00000186226
ENSG00000186517
ENSG00000186526
ENSG00000186529
ENSG00000186723
ENSG00000186803
ENSG00000186832
ENSG00000186847
ENSG00000186912
ENSG00000186943
ENSG00000186971
ENSG00000187010
ENSG00000187145
ENSG00000189068
ENSG00000189299
ENSG00000189377
ENSG00000196154
ENSG00000196329
ENSG00000163914
ENSG00000163993
ENSG00000164047
ENSG00000164077
ENSG00000164509
ENSG00000164609
ENSG00000164691
ENSG00000165071
ENSG00000165125
ENSG00000165168
ENSG00000165272
ENSG00000165349
ENSG00000165390
ENSG00000165794
ENSG00000166523
ENSG00000166535
ENSG00000166670
ENSG00000167207
ENSG00000167210
ENSG00000167261
ENSG00000167332
ENSG00000167346
ENSG00000167656
ENSG00000167749
ENSG00000167751
ENSG00000167759
ENSG00000167768
ENSG00000167825
ENSG00000167851
ENSG00000167925
ENSG00000168515
ENSG00000168907
ENSG00000169035
ENSG00000169248
ENSG00000169385
ENSG00000169397
ENSG00000169413
ENSG00000169442
ENSG00000169469
ENSG00000169474
ENSG00000169605
ENSG00000169877
ENSG00000169908
ENSG00000187871
ENSG00000187908
ENSG00000188015
ENSG00000188153
ENSG00000188293
ENSG00000188404
ENSG00000188710
ENSG00000211967
ENSG00000213398
ENSG00000213417
ENSG00000213762
ENSG00000213809

ENSG00000198692	ENSG00000196748	ENSG00000214049
ENSG00000198734	ENSG00000196754	ENSG00000214374
ENSG00000198771	ENSG00000196805	ENSG00000215203
ENSG00000203813	ENSG00000196878	ENSG00000218089
ENSG00000204174	ENSG00000196954	ENSG00000221852
ENSG00000204361	ENSG00000197061	ENSG00000222036
ENSG00000204472	ENSG00000197079	ENSG00000223380
ENSG00000204482	ENSG00000197405	ENSG00000223609
ENSG00000204539	ENSG00000197487	ENSG00000223865
ENSG00000204936	ENSG00000197565	ENSG00000225937
ENSG00000205021	ENSG00000197588	ENSG00000226516
ENSG00000205420	ENSG00000197591	ENSG00000226777
ENSG00000205426	ENSG00000197617	ENSG00000227418
ENSG00000205649	ENSG00000189068	ENSG00000244734
ENSG00000205810	ENSG00000229035	ENSG00000248098
ENSG00000205981	ENSG00000229637	ENSG00000250361
ENSG00000206047	ENSG00000230389	ENSG00000251321
ENSG00000206073	ENSG00000231887	ENSG00000251655
ENSG00000206075	ENSG00000232258	ENSG00000254126
ENSG00000206301	ENSG00000235568	ENSG00000254127
ENSG00000207720	ENSG00000235942	ENSG00000255398
ENSG00000211592	ENSG00000236699	ENSG00000255794
ENSG00000211689	ENSG00000237330	ENSG00000256812
ENSG00000211895	ENSG00000237599	ENSG00000257227
ENSG00000260199	ENSG00000239590	ENSG00000257335
ENSG00000262128	ENSG00000239839	ENSG00000258227
ENSG00000262320	ENSG00000241058	ENSG00000268943
ENSG00000262383	ENSG00000241794	ENSG00000268954
ENSG00000262406	ENSG00000242180	ENSG00000269936
ENSG00000262462	ENSG00000262827	ENSG00000271481
ENSG00000262754	ENSG00000262936	ENSG00000271503
ENSG00000262800	ENSG00000263639	ENSG00000271942
ENSG00000260199	ENSG00000264162	ENSG00000272611
ENSG00000266722	ENSG00000266109	ENSG00000272743
ENSG00000268682	ENSG00000268090	ENSG00000272939
ENSG00000268369	ENSG00000000460	

Table 2: The mean, median and min-max methylation levels of CGIs associated with 50 selected genes, with greater than 75% methylation.

EnsEMBL ID	Gene	Mean|Median|Min-Max Methylation
ENSG00000135636	*DYSF*	11.2750532534482|3.61764705882353|1.36842105263158-63.5333333333333
		13.1832706177719|3.15789473684211|0.736842105263158-95.4285714
		28571454.0085511074985|44.8571428571429|8.4-96.4285714285714
		86.2662823868664|91.1016949152542|74.2272727272727-93.4698795180723
ENSG00000172244	*C5orf34*	84.0185164329823|85|62.5-96
ENSG00000065268	*WDR18*	86.7600793650794|91.31|76.9285714285714-92.0416666666667
		19.5931142410016|16 |0.112676056338028-42.6666666666667
ENSG00000082293	*COL19A1*	86.1407193375884|84.7058823529412|84.5227272727273-89.1935483870968
		5.07732374702386|1.6530612244898|0.518518518518518-96.5918367346939
ENSG00000160145	*KALRN*	85.7425535680253|93.9047619047619|68.7547169811321-94.5681818181818
		7.22430604197057|2|0.636363636363636-76.2727272727273
ENSG00000107077	*KDM4C*	91.7308360992571|91.8333333333333|89.2105263157895-94.1486486486486
		96.8451937074094|97|91-100
		2.53033039684938|1.3030303030303|0.36101083032491-29.5
		93.8157722940332|93.1739130434783|90.2463768115942-98.027027027027
ENSG00000155657	*TTN*	82.245447788926|79.4927536231884|77.5-89.7435897435897
ENSG00000174564	*IL20RB*	92.1183932346723|96.7272727272727|82.5348837209302-97.0930232558139
		81.4848484848485|83|61.4545454545455-100
ENSG00000145934	*TENM2*	95.0315734989648|95.9642857142857|89.5-99.6304347826087
		83.8511977925498|88.8235294117647|73.0285714285714-89.7014925373134
ENSG00000196353	*CPNE4*	85.3280980371009|84.328125|75.1315789473684-96.5245901639344
		71.0537988581467|74.5833333333333|57.7954545454545-80.7826086956522
ENSG00000100266	*PACSIN2*	2.8875871343697|1.33567774936061|0.333333333333333-42.8
		5.08376673818981|1.57627118644068|0.26530612244898-77.4090909090909
		97.1091051805338|96.5|96.1538461538462-98.6734693877551
		89.4076115650534|93.33 |80.6486486486486-94.2441860465116
ENSG00000128641	*MYO1B*	5.98434426784891|1.6|0.59375-82.925
		85.8585651936716|93.7872340425532|69.7884615384615-94
ENSG00000130300	*PLVAP*	87.4240990990991|91.6125|77.3625-93.2972972972973
		9.0953819784508|3.13194444444444|1.23188405797101-96.3787878787879
ENSG00000149596	*JPH2*	83.0004094819862|79.3349753694581|76.2918454935622-93.3744075829384
		37.9815157901293|27.1111111111111|6.6-92.8620689655172
ENSG00000006459	*KDM7A*	87.85744596346|91.7761194029851 |78.9285714285714-92.8676470588235
		3.45142039708166|1.55414012738854|0.531531531531532-47.8
		12.8796572088108|7|1.25-62.5
ENSG00000188419	*CHM*	89.9805617147081|91.7|86.3636363636364-91.8780487804878
ENSG00000187605	*TET3*	94.6275452393032|94.925|93.5740740740741-95.3835616438356
		6.98777996251653|1.58928571428571|0.761904761904762-92.2307692307692
ENSG00000130803	*ZNF317*	97.0165996057797|97.6666666666667|84.1082802547771-100
		22.1580086580087|10|0.974025974025974-55.5
ENSG00000146205	*ANO7*	7.565024458572|2.54054054054054|1.16666666666667-61.2903225806452
		82.680303030303|83.7833333333333|74.6212121212121-89.6363636363636
ENSG00000088247	*KHSRP*	2.50888888888889|1.41666666666667|0.166666666666667-39.75
		89.9094017473675|90.8518518518518|75.8857142857143-97.9
ENSG00000167658	*EEF2*	94.1518947161268|94.4444444444444|85.4285714285714-99
		2.17457682596726|1.08333333333333|0.444444444444444-16.25

ENSG00000136231	*IGF2BP3*	96.1947419668939|98|92.4303797468354-98.1538461538462
		5.56103087595058|2.28|0.634146341463415-83.925
ENSG00000183570	*PCBP3*	81.0320132779149|86|11-100
		85.8027065527065|86.4358974358974|74.0972222222222-96.875
		85.5989289946222|85.3|81.6163522012579-89.8804347826087
		88.5947978107232|88.650406504065|82.7450980392157-94.3888888888889
		20.4766002493438|10.4461538461538|4.1-79.1666666666667
		82.2842273521431|89.1836734693878|65.0952380952381-92.5737704918033
ENSG00000115756	*HPCAL1*	4.96868581389298|1.99038461538462|1.25438596491228-51.3939393939394
		94.2708596345888|94.7857142857143|83.3571428571429-98.8571428571429
ENSG00000064012	*CASP8*	94.6370347998255|97.2790697674419|88.9047619047619-97.7272727272727
ENSG00000030304	*MUSK*	85.0794205794206|84.1538461538462|82.8571428571429-88.2272727272727
ENSG00000115306	*SPTBN1*	82.6227450980392|82.5882352941177|75.68-89.6
		4.97683349565548|1.44|0.220125786163522-100
		3.7819546472277|1.651515151515|0.516483516483517-43.5
ENSG00000196358	*NTNG2*	4.62538587305964|2.35443037974684|0.704918032786885-53.7272727272727
		89.3429103500532|93.4285714285714|24-100
		8.45191976953081|3.33695652173913|1.31132075471698-45.7582417582418
		85.4282495905878|88.5686274509804|74.1594202898551-93.5567010309278
ENSG00000165102	*HGSNAT*	2.57680019706837|1.53791574279379|0.0608695652173913-43
		89.7561680469289|95.0869565217391|78.0357142857143-96.1458333333333
ENSG00000197548	*ATG7*	95.8862716532579|96.75|82.75-100
ENSG00000119121	*TRPM6*	84.3155187524457|86.4215686274 51 |75.1063829787234-91.4186046511628
		3.83575549085048|1.3125|0.52-50
ENSG00000113448	*PDE4D*	87.1639869281046|88.5166666666667 |83.24-89.7352941176471
		17.7656531784963|3.90634920634921|1.12244897959184-97.0714285714286
		3.50603286883218|1.69047619047619|0.478260869565217-27.5
ENSG00000101868	*POLA1*	9.1268851358876|3.18181818181818|0.266666666666667-61.6324786324786
		95.7310938845822|95.3488372093023 |91.8444444444444-100
ENSG00000105426	*PTPRS*	95.3194876473019|95.9090909090909|88.2857142857143-98.6428571428571
		66.3326482732733|75.1126126126126|21-100
		93.6196047269877|94.725|81.2430555555556-97.6388888888889
		5.19480788751105|2.28947368421053|0.760869565217391-30.25
		5.14010592434024|1.6734693877551|0.516666666666667-76.1428571428571
ENSG00000124225	*PMEPA1*	81.4122257053292|91.2424242424242 |61.1666666666667-91.8275862068966
		66.5824894360971|70.5454545454545|1.11111111111111-97.3333333333333
		5.51692921853267|1.75384615384615|0.775700934579439-59.25
ENSG00000091831	*ESR1*	85.9675862593099|89.3488372093023 |76.4705882352941-92.0833333333333
		11.9466195289074|2.79591836734694|0.566666666666667-92.4137931034483
ENSG00000015475	*BID BH3*	3.57634733358342|1.84090909090909|0.230769230769231-50
		83.7469158127386|85.8095238095238|78.1645569620253-87.2666666666667
ENSG00000101096	*NFATC1*	10.0024597251936|3.75|0.9375-54.8701298701299
		95.0845607407532|95.0217391304348 |86.1176470588235-98.1666666666667
		4.84010094054851|1.90625|0.5-47.1875
		42.0693740555536|37.8181818181818|0.636363636363636-97.2
ENSG00000079482	*OPHN1*	22.2248068272873|22.9230769230769|0.714285714285714-72.4666666666667
		76.6743445916378|83.1224489795918 |59.7894736842105-87.1111111111111
		92.6984126984127|90|88.0952380952381-100
		89.6658507893139|90.6885245901639|84.7777777777778-93.53125
ENSG00000063169	*GLTSCR1*	88.4584657578446|89.5|76.6666666666667-95.5
		93.4555915992439|93.2068965517241|87.3-97.625
		93.781545600047|94.4|86.2-98.4

		3.64126293531793|1.88961038961039|0.71900826446281-41.75
ENSG00000184470	*TXNRD2*	91.4721568627451|92.4705882352941 |89.24-92.7058823529412
ENSG00000050767	*COL23A1*	85.9724206349206|88.5333333333333|78.5714285714286-90.8125
		87.625577836082|92.5|38.3333333333333-100
		9.15378658084283|3.76388888888889|1.05555555555556-98.2142857142857
		27.8839109466504|12.3269230769231|1.75510204081633-97.4
		91.6685518143184|91.8695652173913|86.7789473684211-96.3571428571429
ENSG00000159733	*ZFYVE28*	83.8349462365591|88.8548387096774 |71.7666666666667-90.8833333333333
		5.0399131624867|2.1875|1.11242603550296-48.4590163934426
		35.7332857307299|21.3333333333333|3.76190476190476-97.4444444444444
		83.4936779765021|88.5679012345679|71.3837209302326-90.5294117647059
ENSG00000109572	*CLCN3*	9.99881796690308|12 |1.32978723404255-16.6666666666667
		2.42065810560668|1.35294117647059|0.470588235294118-41.5
		85.7906258850184|87.4333333333333|80.3030303030303-89.6355140186916
ENSG00000127191	*TRAF2*	4.23557411210648|1.44761904761905|0.693877551020408-77.6666666666667
		91.7861729704254|94.2317073170732|86.0434782608696-95.0833333333333
ENSG00000206560	*ANKRD28*	3.05839206517669|1.54838709677419 |0.652173913043478-48.6666666666667
		93.3603603603604|94.5945945945946 |88.7297297297297-96.7567567567568
ENSG00000090889	*KIF4A*	82.1080986878281|87.9072164948454|69.4383561643836-88.9787234042553
		91.5691884280594|93.4166666666667|84.9516129032258-96.3392857142857
ENSG00000083223	*ZCCHC6*	92.9225957049487|91.7058823529412|90.5142857142857-96.5476190476191
		3.65700724741376|1.54545454545455|0.540540540540541-63.8333333333333
ENSG00000068024	*HDAC4*	93.0201739103857|93.7142857142857|81.725-100
		70.9198343079922|73.3333333333333|1-100
		3.62191859201892|1.80645161290323|0.720338983050847-39.65
ENSG00000170265	*ZNF282*	95.1818913608771|95.5238095238095|88.358024691358-98.0714285714286
		3.08869851242618|1.44444444444444|0.166666666666667-39.6666666666667

Table 3: List of genes that displayed methylation levels below 75%.

ENSG00000155545	ENSG00000184056	ENSG00000170961	ENSG00000182318
ENSG00000178226	ENSG00000158417	ENSG00000130707	ENSG00000170802
ENSG00000069011	ENSG00000132604	ENSG00000141298	ENSG00000162040
ENSG00000215440	ENSG00000072786	ENSG00000196943	ENSG00000050628
ENSG00000185947	ENSG00000120899	ENSG00000170581	ENSG00000161798
ENSG00000111328	ENSG00000136937	ENSG00000149554	ENSG00000167130
ENSG00000161992	ENSG00000162676	ENSG00000174292	ENSG00000125965
ENSG00000164221	ENSG00000096070	ENSG00000157540	ENSG00000117480
ENSG00000088205	ENSG00000177200	ENSG00000125257	ENSG00000135842
ENSG00000173581	ENSG00000172197	ENSG00000114209	ENSG00000185100
ENSG00000135636	ENSG00000151726	ENSG00000104267	ENSG00000198464
ENSG00000172244	ENSG00000107159	ENSG00000155657	ENSG00000198863
ENSG00000164164	ENSG00000104856	ENSG00000205133	ENSG00000092758
ENSG00000167004	ENSG00000082293	ENSG00000185088	ENSG00000145592
ENSG00000158019	ENSG00000184381	ENSG00000114779	ENSG00000134516
ENSG00000158715	ENSG00000135698	ENSG00000151617	ENSG00000103174
ENSG00000119559	ENSG00000145860	ENSG00000170485	ENSG00000179111
ENSG00000185024	ENSG00000176597	ENSG00000109743	ENSG00000198001
ENSG00000106780	ENSG00000136490	ENSG00000164284	ENSG00000163660
ENSG00000140332	ENSG00000109255	ENSG00000065621	ENSG00000123609
ENSG00000166451	ENSG00000100359	ENSG00000203747	ENSG00000134574
ENSG00000241399	ENSG00000136161	ENSG00000158470	ENSG00000107738
ENSG00000083307	ENSG00000242110	ENSG00000182858	ENSG00000185215
ENSG00000134201	ENSG00000089639	ENSG00000167034	ENSG00000130119
ENSG00000048140	ENSG00000046604	ENSG00000182796	ENSG00000174564
ENSG00000062282	ENSG00000174851	ENSG00000196652	ENSG00000112062
ENSG00000171055	ENSG00000183242	ENSG00000167202	ENSG00000168875
ENSG00000101624	ENSG00000101333	ENSG00000008283	ENSG00000152822
ENSG00000075218	ENSG00000148110	ENSG00000165527	ENSG00000111186
ENSG00000180739	ENSG00000100461	ENSG00000153179	ENSG00000188372
ENSG00000186205	ENSG00000137841	ENSG00000266714	ENSG00000151702
ENSG00000161055	ENSG00000080815	ENSG00000038219	ENSG00000158483
ENSG00000148773	ENSG00000113013	ENSG00000070371	ENSG00000068878
ENSG00000082175	ENSG00000140993	ENSG00000169174	ENSG00000253293
ENSG00000140564	ENSG00000156427	ENSG00000169599	ENSG00000169621
ENSG00000181577	ENSG00000058866	ENSG00000144339	ENSG00000077063
ENSG00000115963	ENSG00000091127	ENSG00000179583	ENSG00000175264
ENSG00000184825	ENSG00000166900	ENSG00000133067	ENSG00000114480

ENSG00000125347	ENSG00000068831	ENSG00000106714	ENSG00000122592
ENSG00000152359	ENSG00000169213	ENSG00000196914	ENSG00000141574
ENSG00000204842	ENSG00000198729	ENSG00000073756	ENSG00000009413
ENSG00000173535	ENSG00000184708	ENSG00000152229	ENSG00000198211
ENSG00000143147	ENSG00000134215	ENSG00000180340	ENSG00000117525
ENSG00000005073	ENSG00000130720	ENSG00000162493	ENSG00000146205
ENSG00000165502	ENSG00000113916	ENSG00000168092	ENSG00000108932
ENSG00000077782	ENSG00000099953	ENSG00000072195	ENSG00000153113
ENSG00000112234	ENSG00000128652	ENSG00000113296	ENSG00000121742
ENSG00000119760	ENSG00000140279	ENSG00000169020	ENSG00000244274
ENSG00000149177	ENSG00000156983	ENSG00000104960	ENSG00000112699
ENSG00000122335	ENSG00000104356	ENSG00000125845	ENSG00000145569
ENSG00000168461	ENSG00000197561	ENSG00000184307	ENSG00000152256
ENSG00000188130	ENSG00000136378	ENSG00000122585	ENSG00000110934
ENSG00000075290	ENSG00000262209	ENSG00000173545	ENSG00000133028
ENSG00000005381	ENSG00000148300	ENSG00000184012	ENSG00000140577
ENSG00000121671	ENSG00000108669	ENSG00000104047	ENSG00000167815
ENSG00000144668	ENSG00000117691	ENSG00000142765	ENSG00000157613
ENSG00000165097	ENSG00000064115	ENSG00000168071	ENSG00000076662
ENSG00000166938	ENSG00000150867	ENSG00000196428	ENSG00000133704
ENSG00000060138	ENSG00000125967	ENSG00000008869	ENSG00000148803
ENSG00000166006	ENSG00000021762	ENSG00000171462	ENSG00000165501
ENSG00000166509	ENSG00000149489	ENSG00000105427	ENSG00000001167
ENSG00000124251	ENSG00000137709	ENSG00000151835	ENSG00000167705
ENSG00000198604	ENSG00000171840	ENSG00000153094	ENSG00000182230
ENSG00000047648	ENSG00000225921	ENSG00000230989	ENSG00000178449
ENSG00000038002	ENSG00000183166	ENSG00000178038	ENSG00000124614
ENSG00000113119	ENSG00000109184	ENSG00000049319	ENSG00000115380
ENSG00000182400	ENSG00000170540	ENSG00000185721	ENSG00000118849
ENSG00000141506	ENSG00000121691	ENSG00000136819	ENSG00000109881
ENSG00000182162	ENSG00000105538	ENSG00000176386	ENSG00000187583
ENSG00000175938	ENSG00000043355	ENSG00000103222	ENSG00000110400
ENSG00000059378	ENSG00000035664	ENSG00000197965	ENSG00000040608
ENSG00000076928	ENSG00000006016	ENSG00000171174	ENSG00000085719
ENSG00000070031	ENSG00000214160	ENSG00000145331	ENSG00000182566
ENSG00000121316	ENSG00000132773	ENSG00000169981	ENSG00000145113
ENSG00000196357	ENSG00000131018	ENSG00000137699	ENSG00000197261
ENSG00000197622	ENSG00000154553	ENSG00000139190	ENSG00000135372
ENSG00000142347	ENSG00000010017	ENSG00000145916	ENSG00000139289
ENSG00000117280	ENSG00000101596	ENSG00000149488	ENSG00000180818

ENSG00000198315	ENSG00000103343	ENSG00000101773	ENSG00000128040
ENSG00000106003	ENSG00000137193	ENSG00000119986	ENSG00000137273
ENSG00000186523	ENSG00000169851	ENSG00000105835	ENSG00000198554
ENSG00000204160	ENSG00000131351	ENSG00000124243	ENSG00000101654
ENSG00000000938	ENSG00000032219	ENSG00000184922	ENSG00000106031
ENSG00000084693	ENSG00000185761	ENSG00000005302	ENSG00000170425
ENSG00000115163	ENSG00000101336	ENSG00000148655	ENSG00000179104
ENSG00000077420	ENSG00000019186	ENSG00000066427	ENSG00000005961
ENSG00000171517	ENSG00000006611	ENSG00000171135	ENSG00000167513
ENSG00000138780	ENSG00000167580	ENSG00000198951	ENSG00000177105
ENSG00000103257	ENSG00000170577	ENSG00000104064	ENSG00000166974
ENSG00000135655	ENSG00000174669	ENSG00000186635	ENSG00000123146
ENSG00000126822	ENSG00000169083	ENSG00000103381	ENSG00000167799
ENSG00000036549	ENSG00000179361	ENSG00000110075	ENSG00000134901
ENSG00000173480	ENSG00000267645	ENSG00000183943	ENSG00000187735
ENSG00000118197	ENSG00000130741	ENSG00000165879	ENSG00000182481
ENSG00000102554	ENSG00000124193	ENSG00000026103	ENSG00000173530
ENSG00000169896	ENSG00000096696	ENSG00000104447	ENSG00000143061
ENSG00000053747	ENSG00000115750	ENSG00000042286	ENSG00000130844
ENSG00000066336	ENSG00000184076	ENSG00000174579	ENSG00000112164
ENSG00000106351	ENSG00000223414	ENSG00000134046	ENSG00000165025
ENSG00000108443	ENSG00000130635	ENSG00000145819	ENSG00000204219
ENSG00000127989	ENSG00000128585	ENSG00000163067	ENSG00000105329
ENSG00000015285	ENSG00000185567	ENSG00000125657	ENSG00000131389
ENSG00000225930	ENSG00000128713	ENSG00000173546	ENSG00000124440
ENSG00000089195	ENSG00000136718	ENSG00000124466	ENSG00000198858
ENSG00000117151	ENSG00000152078	ENSG00000174562	ENSG00000180071
ENSG00000167693	ENSG00000249859	ENSG00000196767	ENSG00000141380
ENSG00000135604	ENSG00000134698	ENSG00000140836	ENSG00000183762
ENSG00000168769	ENSG00000105122	ENSG00000221866	ENSG00000188672
ENSG00000173511	ENSG00000144909	ENSG00000171097	ENSG00000187266
ENSG00000173442	ENSG00000141655	ENSG00000254087	ENSG00000095261
ENSG00000115361	ENSG00000198088	ENSG00000113719	ENSG00000120457
ENSG00000100311	ENSG00000164920	ENSG00000187840	ENSG00000136536
ENSG00000150281	ENSG00000057704	ENSG00000170178	ENSG00000117143
ENSG00000197361	ENSG00000172006	ENSG00000090339	ENSG00000172432
ENSG00000116478	ENSG00000187079	ENSG00000115138	ENSG00000128709
ENSG00000110324	ENSG00000166923	ENSG00000162526	ENSG00000144674
ENSG00000168476	ENSG00000182580	ENSG00000124766	ENSG00000129355
ENSG00000110375	ENSG00000108055	ENSG00000140519	ENSG00000165533

ENSG00000174206	ENSG00000146457	ENSG00000021355	ENSG00000211456
ENSG00000214021	ENSG00000204271	ENSG00000138162	ENSG00000122644
ENSG00000158615	ENSG00000162129	ENSG00000156017	ENSG00000166710
ENSG00000146733	ENSG00000112146	ENSG00000168209	ENSG00000156273
ENSG00000112541	ENSG00000064270	ENSG00000100599	ENSG00000140396
ENSG00000136286	ENSG00000149499	ENSG00000160703	ENSG00000137198
ENSG00000136877	ENSG00000005020	ENSG00000151657	ENSG00000166133
ENSG00000104976	ENSG00000137414	ENSG00000074319	ENSG00000146676
ENSG00000138180	ENSG00000156113	ENSG00000115761	ENSG00000112208
ENSG00000111913	ENSG00000148842	ENSG00000198108	ENSG00000160753
ENSG00000130038	ENSG00000143761	ENSG00000117115	ENSG00000132334
ENSG00000167700	ENSG00000163795	ENSG00000169567	ENSG00000077943
ENSG00000134419	ENSG00000185009	ENSG00000117713	ENSG00000147889
ENSG00000171700	ENSG00000107625	ENSG00000140406	ENSG00000105204
ENSG00000130592	ENSG00000124383	ENSG00000135363	ENSG00000179029
ENSG00000048740	ENSG00000081277	ENSG00000168010	ENSG00000136167
ENSG00000198736	ENSG00000118855	ENSG00000139865	ENSG00000150893
ENSG00000121749	ENSG00000157570	ENSG00000167173	ENSG00000089248
ENSG00000133401	ENSG00000140368	ENSG00000154025	ENSG00000123388
ENSG00000171307	ENSG00000111790	ENSG00000103544	ENSG00000130733
ENSG00000181222	ENSG00000104154	ENSG00000129657	ENSG00000130775
ENSG00000163357	ENSG00000006194	ENSG00000117148	ENSG00000141404
ENSG00000169752	ENSG00000132570	ENSG00000078061	ENSG00000198900
ENSG00000158711	ENSG00000130755	ENSG00000109775	ENSG00000183878
ENSG00000104413	ENSG00000180447	ENSG00000166477	ENSG00000100285
ENSG00000100350	ENSG00000182022	ENSG00000160712	ENSG00000071054
ENSG00000119231	ENSG00000145088	ENSG00000135838	ENSG00000181274
ENSG00000158406	ENSG00000143537	ENSG00000141480	ENSG00000167996
ENSG00000182287	ENSG00000151651	ENSG00000123329	ENSG00000169180
ENSG00000137857	ENSG00000116747	ENSG00000100416	ENSG00000092036
ENSG00000168310	ENSG00000131759	ENSG00000101425	ENSG00000159648
ENSG00000074657	ENSG00000187778	ENSG00000107551	ENSG00000136867
ENSG00000183597	ENSG00000198755	ENSG00000184937	ENSG00000159461
ENSG00000129933	ENSG00000108264	ENSG00000134202	ENSG00000182568
ENSG00000159202	ENSG00000203685	ENSG00000164647	ENSG00000178935
ENSG00000143815	ENSG00000147065	ENSG00000132507	ENSG00000103855
ENSG00000144029	ENSG00000175643	ENSG00000100985	ENSG00000011198
ENSG00000139874	ENSG00000128714	ENSG00000101236	ENSG00000204138
ENSG00000173171	ENSG00000158042	ENSG00000117411	ENSG00000151503
ENSG00000197045	ENSG00000107187	ENSG00000135069	ENSG00000176390

ENSG00000168077	ENSG00000122420	ENSG00000166592	ENSG00000165312
ENSG00000181444	ENSG00000254806	ENSG00000177192	ENSG00000178403
ENSG00000160796	ENSG00000184281	ENSG00000100983	ENSG00000157554
ENSG00000149212	ENSG00000105656	ENSG00000213390	ENSG00000163798
ENSG00000116815	ENSG00000114115	ENSG00000173214	ENSG00000167114
ENSG00000104979	ENSG00000069275	ENSG00000107262	ENSG00000150051
ENSG00000035928	ENSG00000075643	ENSG00000085982	ENSG00000161513
ENSG00000137809	ENSG00000182257	ENSG00000138735	ENSG00000107968
ENSG00000008516	ENSG00000113231	ENSG00000123364	ENSG00000149269
ENSG00000198000	ENSG00000066468	ENSG00000105483	ENSG00000171729
ENSG00000196422	ENSG00000141002	ENSG00000170456	ENSG00000171763
ENSG00000168386	ENSG00000127191	ENSG00000075073	ENSG00000221829
ENSG00000119684	ENSG00000099341	ENSG00000144554	ENSG00000197299
ENSG00000132906	ENSG00000164032	ENSG00000120616	ENSG00000129451
ENSG00000163069	ENSG00000131969	ENSG00000112996	ENSG00000164941
ENSG00000146410	ENSG00000113648	ENSG00000117614	ENSG00000104765
ENSG00000089154	ENSG00000077157	ENSG00000130590	ENSG00000159184
ENSG00000138346	ENSG00000165816	ENSG00000137642	ENSG00000133392
ENSG00000103202	ENSG00000101255	ENSG00000145808	ENSG00000100612
ENSG00000163637	ENSG00000185811	ENSG00000180758	ENSG00000105676
ENSG00000167419	ENSG00000110328	ENSG00000088386	ENSG00000005513
ENSG00000173575	ENSG00000165752	ENSG00000109180	ENSG00000163497
ENSG00000114626	ENSG00000076242	ENSG00000182463	ENSG00000116741
ENSG00000141524	ENSG00000155506	ENSG00000187866	ENSG00000005339
ENSG00000008382	ENSG00000012061	ENSG00000198924	ENSG00000007038
ENSG00000100802	ENSG00000131873	ENSG00000172845	ENSG00000107018
ENSG00000172987	ENSG00000173068	ENSG00000187741	ENSG00000067113
ENSG00000243766	ENSG00000175054	ENSG00000090470	ENSG00000144468
ENSG00000108306	ENSG00000179151	ENSG00000125457	ENSG00000133805
ENSG00000186642	ENSG00000172789	ENSG00000160255	ENSG00000102539
ENSG00000154229	ENSG00000128283	ENSG00000196655	ENSG00000087460
ENSG00000165689	ENSG00000259112	ENSG00000163421	ENSG00000132676
ENSG00000171608	ENSG00000102879	ENSG00000156026	ENSG00000151715
ENSG00000133302	ENSG00000172936	ENSG00000115271	ENSG00000165912
ENSG00000136040	ENSG00000168765	ENSG00000124171	ENSG00000139842
ENSG00000015133	ENSG00000049239	ENSG00000145824	ENSG00000182132
ENSG00000089159	ENSG00000132849	ENSG00000139132	ENSG00000130193
ENSG00000135312	ENSG00000187210	ENSG00000136522	ENSG00000181192
ENSG00000143669	ENSG00000109805	ENSG00000114738	ENSG00000101412
ENSG00000076944	ENSG00000170571	ENSG00000143768	ENSG00000110090

ENSG00000158717	ENSG00000007541	ENSG00000179008	ENSG00000164078
ENSG00000106591	ENSG00000029534	ENSG00000182307	ENSG00000176170
ENSG00000106038	ENSG00000133627	ENSG00000135441	ENSG00000129514
ENSG00000254470	ENSG00000072401	ENSG00000140044	ENSG00000102921
ENSG00000088543	ENSG00000171862	ENSG00000113387	ENSG00000168556
ENSG00000171444	ENSG00000127616	ENSG00000114026	ENSG00000170615
ENSG00000147324	ENSG00000108591	ENSG00000143476	ENSG00000104093
ENSG00000100504	ENSG00000088387	ENSG00000134996	ENSG00000088992
ENSG00000112964	ENSG00000123700	ENSG00000169509	ENSG00000128965
ENSG00000168067	ENSG00000175106	ENSG00000170370	ENSG00000198768
ENSG00000102034	ENSG00000198918	ENSG00000062038	ENSG00000132436
ENSG00000111432	ENSG00000133961	ENSG00000171777	ENSG00000138696
ENSG00000119699	ENSG00000100314	ENSG00000144063	ENSG00000148572
ENSG00000100393	ENSG00000177548	ENSG00000129534	ENSG00000111845
ENSG00000115524	ENSG00000138376	ENSG00000165474	ENSG00000129474
ENSG00000125691	ENSG00000034713	ENSG00000143801	ENSG00000104738
ENSG00000175356	ENSG00000169826	ENSG00000088826	ENSG00000150361
ENSG00000143578	ENSG00000136630	ENSG00000163975	ENSG00000173110
ENSG00000159231			

Table 4: Sequences of CGIs targeted for primer design. The primer sequences are highlighted in yellow.

Gene	Amplicon Sequence
DYSF	GCATAAATCACCCTGGTTGGCTCGGCTCTCCCTATCCCTCTCCAGACCCGCTTGCCCCCGCTCTCAGGCCCACTCAGGCCT
	GGTTCCCAGGCAGGCCCCCGCCCCGCTACCCACCGAGCATGCCCCTCCACCACCGAGGAGCGGCCTGGGGCAGCCCATCC
	GCGGCACCCTCGGGGCAGTAAATATTATATTACTACAGGGTGTGCAGTACGTGACCTGAGACAGGCCCCGGGGCTTAAGTT
	TCCCGAGATGGAGTTTGCGTACTGCTCCCCCCATCCTAAAAATTCAATTAAAAAAAATAACAGGAAAAATGTAGTGCGCCG
	AGGGCCTGCCAGAGCGGTGTATCTGGCGGCGGCGGTGCCGCTGCGCTCCGGCGGTAATTGGTTCTGCATAGCTCGGGGAA
	ATTGGCAGAAAAATAAATCAGCCGGGAGCCGGGAGTCAGACGCTGCAGGGAGGGGAGGTGGGGGGTGCTGCAAGTGGTG
	GGAGCAGACCTGGTGGAGCGGGTTTGACTAATGTCCTGTTTGGGAGTCAAGGGGGTGAGGTGGGGGCAGGGCAGCCCCCAT
	ACCCTCAGTGGAGAATGGAAAATGGAGAGGAGTCCC
HPCAL1	GCCTTCTTGGTGGTCCATAAGATAGCGGCATGCCTTCCGATGGGGGACCCGGGAGTTGCTGAGTCGCAGCGCTTGCA
	CAGCGATGGGCTTATTTGGTCTCTCCGGCACATGGCTCAGCCCTGCTCCGTGGCCGTGGGTGGCGTCCCCGGCTGA
	CCCCCTGTCTTGCAGGTGTAGTCGC
CPNE4	CGCCTGACTGGTTTTCGTATTTTTTTGGTGGAGATGGGGTTTCGCTGTGTTGGCCGGGCTGGTCTCCAGCTCCTAACCGCG
	AGTGATCCGCCAGCCTCGGCCTCCCGAGGTGCCGGGATTGCAGACGGAGTCTCGTTCACTCAGTGGTCAATGGTGCCCAG
	GCTGGAGTGCAGTGGTGTGATCTCGGCTCGCTACAACCTCCACCTCCCAGCCACCTGCCTTGGCCTCCCAAAGTGCCAAGA
	TTGCAGCCTCTGCCCAGCCGCCACCCCGTCTGGGAAGTGAGGAGCGTCTCTGCCTGGCCGCCCATCGTCTGGGATGTGAGG
	AGCCCCTCTGCCTGGCTGCCCAGTCTGGAAAGTGAGGAGCGTCTCTGCCCGGCCGCCATCCCATCTAGGAAGTGAGGAGTG
	CCTCTTCCCGGCCGCCATCACATCTAGGAAGTGAGGAGCATCTCTGCCCGGCCGCCCATCGTCTGAGATGTGGGGAGCGCC
	TCTGCCCCGCCGCCCCGTCTGGGATGTGAGGAGCGCCTCTGCCCGGCCGCGAACCCGTATGGGAGGTGA

ESR1 ACGGGTGACTTCTGCATTTCCATCTGACATACCGGGTTCATCTCACTAGGGAGTGCCAGACATTGGGCGGAGGTCAGTGGGTGCGC
GCACCGTGCGCGAGCCGAAGCAGGGCGAGACATTGCCTCACTTGGGAAGCGCAAGGGGTCAGCGAGTTCCCTTTCTGAGTCAAA
GAAAGGGGTGACGGATGGCACCTGGAAAATCGGGTCACTCCCACCCGAATACTGCGCTTTTCCGACGGGCTTAAAAAACGGAGC
ACCACGAGATTATATCCCGCACCTGGCTCGGAGGATCCTACGCCCACGCAGTCTCGCTGATTGCTAGCACAGCAGTCTGAGATCA
AACTGCAAGGCGGCAGCGAGGCTGGGGGAGGGGCGCCCGCCATTGCCCAGGCTTGCATAGGTAAACAAAGCAGCCAGGAAGCTC
GAACTGGGTGGAGCCCACCACAGCTCAAGGAGGCCTGCCTGCCTCTGTAGGCTTCATCTCTGGGGGCAGGGCACAGACAAACAA
AAAGACAGCAGTAACCTCTGCAGACTTAAATGTCCCTGTCTGACAGCTTTGAAGAGGGCAG

ZNF282 CTATCTCCCCAGGTGACAGCCTGCTGATGGTGAAGAACCCACCCCCGGCCCCGCCACAGCCCCAGCCCCAGCCCCAGCCACCGCAGCC
GCAGCTGCAGTCGCAGCCCCAGCCCCAGAGCCTGCCCCCCATCGCGGTGGCCGAGAACCCGGGCGGCCCCCCGAGCCGAGGGCTGCT
GGACGACGGTTTCCAGGTGCTGCCCGGGGAGCGTGGCTCCGGCGAGGCGCCGCCGGGTGGGGACCGCAGCACCGGGGGCGGCGGGGG
CGATGGGGGCGGTGGGGGCGGCGGCGCGGAGGCGGGGACGGGGGCAGGCGGCGGCTGTGGCAGCTGCTGCCCTGGCGGGCTGCGGCG
GAGCCTCCTCCTGCACGGCGCCCGCAGCAAGCCCTACTCGTGCCCCGAGTGCGGCAAGAGCTTCGGCGTGCGCAAGAGCCTCATCATC
CACCACCGCAGCCACACCAAGGAGCGGCCCTACGAGTGCGCTGAGTGCGAGAAGAGCTTCAACTGCC

MUSK GAACAGCTCCGGTCTACAGCTCCCAGCGTGAGCGACGCAGAAGACGGGTGATTTCTGCATTTCCATCTGAGGTACCGGGTTCATCTCAC
TAGGGAGTGCCAGACAGTGGGCGCAGGCCAGTGTGTGTGCGCACCGTGCGCGAGCCGAAGCAGGGCGAGGCATTGCCTCACCTGGGA
AGCGCAAGGGGTCAGGGAGTTCCCTTTCTGAGTCAAAGAAAGGGGTGACGGACGCACCTGGAAAATCGGGTCACTCCCACCCGAATA
TTGCGCTTTTCAGACCGGCTTAAGAAACGGCGCACCACGAGACTATATCCCACACCTGGCTCGGAGGGTCCTACGCCCACGGAATCTCG
CTGATTGCTAGCACAGCAGTCTGAGACCAAACTGCAAGGCGGCAACGGGGCTGGGGGAGGGGCGCCCGCCATTGCCCAGGCTTGCTT
AGGTAAACAAAGCAGCCAGGAAGCTCGAACTGGGTGGAGCCCACCACAGCTCAAGGAGGCCTGCCTGCCTCTGTAGGCTCCACCTCTGG

EEF2 TCAGCACACTGGCATAGAGGCAGCGCCGTGCTGTGGGGATGATCTGGCCCCCTCCGCGGTGGATGGCGTCGGCGTGCAGGGTGACGTC
GTGGACGTCGAAGCGCACACCCCGCATGTTCTCCTCACACAGTGCGCCCTGGGGGGAGGGGGAGAGCCACCGTCAAGGGCCGGACACAC
CTCGGCTGCTTGCCCTCCACCTGCCAAGTCCTGCAGGTCTCCACCAGGGGGACCTGGGGCCTTGCCCGCCTTGGCCCCATTAGGGTCTCT
GTCTCGGGAGGCAGGACCATGAGGTCCCTCTAGAGCCTGGAAACGGGTGTGGTCTGCACATGCTGAGCCGTGCCTCACCTCCTTGGTGG
CCCACTGGAAGCCGGCCACCACACTGTCCTTGATCTCGTTGAGGTACTGCACACCCTTGGTGATGTCGGTGAGGATGTTGGGGCCGGTG
CCGTCGGGCCCAAAGCACCAGATCTTGCGGGCCTCAGCCACGTCCCACTCGTACTTCTCGGCCAGGTAGCGCGCCCGCTGCTTGAGCTC
CTGACGGGCGGACACCTCGCCTTTATCGATGT

IGF2BP3 CATCACCAAGTCCCTGTACGTCAGTCAGCATCACCACATCCGCAGGTCCCCGTTCATCGGCTTCACCGCATCCGCAGGTCCCGTCCATCA
GTCGGCATCACCGCATCGCCAGGTCCCCCTCCATCAGTCGGCATCACCGCATCCGCAGGTCCCCCTCCGTCAGCCGGCATCACCGCATCC
GCAGGTCCCCCTCCGTCAGTCGGCATCACCGCATCCGCAGGTCCCCCTCCGTCAGTCGGCATCACTGCATCCGCAGGTCCCGTCCGTCAGT
CGGCATCACCGCATCCGCAGGTCCCTGTCTGTCAGCCGGCGTCACTGAATCAGCAGTTGGTTACAGCCTCTAGCTCCTCTGGACACTCTGA
ACAAGATGTATAACTCCTCTCTCCTCCCCACATCCCAGGCGCCAGCCATAGCCAAGGAGCTCCCATGGCTCAGAAATCTAAGTCTCAGGTG
CTCAGGGCTCTTCCCTCCAAGTCTCTAGGATCCAAAACCTCAACTTCCTCTTTGGCCTTCCAACCCTCAGGCCGTGGCTGCTCCTTAC
AGTCACTACTCTGGATTCTCTTTGTTCTTTCGCCTTTCAGCAC

PTPRS CATAACCCACAAACCGCTCTGCGGCTTCCAGGGAGCATGCAAAGAACCAAGCCGCACGGCCGGGGCCCCGAGTCCTGCCGACCCCAC
GCCTCACCTCCCTGCTGCGTCTTGACCTGGATGGGGTCCGAGAGGGGCCCGTCGCCGACGGAGGTGAAGGCGAGCACCCGCACGGTG
TAGGTCTCGTCCTCCAGCAGGCTGCCCACGGTGGTCAGCAGGCTGTCGTCCACGTTGTGCTTCTGCCAGTTGCCCACGGGGTGCTCCG
GTTCCATGGTGTAGTAGACGCGGTAGCCGCGGATCAGGCCGTTGGGCTCCACCGGCTCCTCCCACTGCACAATCATGGTGGTCGCGCT
GAGCATCCGGGCTTGCACGTTCCGCGGCGCGCTGGCCGGGGCCTGCTCGCCTGTGCGGGTGACCACGGACTCGCTGGGGGGCCCCTG
GCCGATGGAGTTGACGGCCGACACCCAGATCTCGTACTCCGAGTTGGGGCTCAGGCCGCCGATGCTGTAACGTGTGGTGG

PMEPA1 CTGATGCCCGAGTTACTGCTGGGGGGGCAGGGGCCGCCCAGCCTGGCACTATCCATCAGGTCACTGTCGAAGATGGTTCTGTTTGGG
GGTGCGCGCACCGACTCCCGGTTCAGTTCCAGCTGCTGCTCGGGGTCCCGAAGCTGGAGGGTGCAGGGGCCCTGGTAGGGTGGGGG
CTCCTCCCCGTCTGACAGCGAGATGGTGGGTGGCAGGTCGATCTCGTGCTGCAGGTACGGATAGGTGGGCTGGAAGCGGTGGAAGC
GCTCCCGCTGGGCGAAGGGCGGCACGGCCAGGCGGTCGGTGGGCCGAGGCGGGGCGTAGACCTGCGGCTGGAGGAAGCAGAGCGG
GAGTGAGGGAGGGCGGCTGTCTCAGGTGGGTTGTCCAGAAGCGATCCTGAGACTGGAGTTCTGCTGCATGGGACTTGGCTCTCTGG
GGAAAGGGAGAGGGCAGCAGGGCTGGCGGGGGCGGGAGCCCAGAGCCTGATCCCGCGGGGGCCCTGGCCTGGGGGGCAGCACTG
CAGAGCTGCACCGCCCTCCTCCAAACAGTGGCAGCGTCCGTGCTCCCGTGTGCAGAGAGCAGAGAG

BLAST results for CGI sequences mapping uniquely to a single gene.

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
2:71560383-71560921 [Sequence]	DYSF	Forward	1	539	539 [Sequence]	1068	0.0	100.00 [Alignment]
2:66313933-66313958 [Sequence]		Reverse	468	493	26 [Sequence]	44.1	1.2	96.15 [Alignment]
2:90009223-90009248 [Sequence]		Reverse	220	245	26 [Sequence]	44.1	1.2	96.15 [Alignment]
10:132569812-132569833 [Sequence]	INPP5A	Reverse	218	239	22 [Sequence]	44.1	1.2	100.00 [Alignment]
10:132570107-132570128 [Sequence]	INPP5A	Reverse	218	239	22 [Sequence]	44.1	1.2	100.00 [Alignment]
7:5816172-5816192 [Sequence]		Reverse	233	253	21 [Sequence]	42.1	4.7	100.00 [Alignment]
12:109422767-109422787 [Sequence]	MYO1H	Forward	220	240	21 [Sequence]	42.1	4.7	100.00 [Alignment]
12:109769678-109769698 [Sequence]	FAM222A, FAM222A-AS1	Forward	481	501	21 [Sequence]	42.1	4.7	100.00 [Alignment]
11:2065060-2065080 [Sequence]		Reverse	511	531	21 [Sequence]	42.1	4.7	100.00 [Alignment]
1:45843122-45843142 [Sequence]	MAST2	Reverse	220	240	21 [Sequence]	42.1	4.7	100.00 [Alignment]
1:160474723-160474743 [Sequence]		Reverse	472	492	21 [Sequence]	42.1	4.7	100.00 [Alignment]

Figure 2: The CGI sequence specific to gene *DYSF*.

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
19:3977535-3978039 [Sequence]	EEF2	Forward	1	505	505 [Sequence]	1001	0.0	100.00 [Alignment]

Figure 3: The CGI sequence specific to gene *EEF2*.

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
2:10419626-10419743 [Sequence]	HPCAL1	Forward	1	118	118 [Sequence]	234	1e-58	100.00 [Alignment]
21:37758974-37758993 [Sequence]	KCNJ6	Forward	46	65	20 [Sequence]	40.1	3.5	100.00 [Alignment]
19:810675-810694 [Sequence]	PTBP1	Forward	90	109	20 [Sequence]	40.1	3.5	100.00 [Alignment]

Figure 4: The CGI sequence specific to gene *HPCAL1*.

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
3:131951779-131952272 [Sequence]	CPNE4	Forward	1	494	494 [Sequence]	979	0.0	100.00 [Alignment]
8:21712315-21712808 [Sequence]	GFRA2	Reverse	1	494	494 [Sequence]	876	0.0	97.37 [Alignment]
22:29644229-29644722 [Sequence]	NF2	Reverse	1	494	494 [Sequence]	876	0.0	97.37 [Alignment]
2:27358971-27359464 [Sequence]	AC074117.10	Forward	1	494	494 [Sequence]	876	0.0	97.37 [Alignment]
X:77766529-77767021 [Sequence]	ATRX	Reverse	1	494	494 [Sequence]	868	0.0	97.37 [Alignment]
9:125459123-125459616 [Sequence]	MAPKAP1	Reverse	1	494	494 [Sequence]	868	0.0	97.17 [Alignment]
7:47959038-47959531 [Sequence]	HUS1	Forward	1	494	494 [Sequence]	868	0.0	97.17 [Alignment]
5:138777166-138777659 [Sequence]	CTNNA1	Reverse	1	494	494 [Sequence]	868	0.0	97.17 [Alignment]
8:37927337-37927830 [Sequence]		Forward	1	494	494 [Sequence]	860	0.0	96.96 [Alignment]
1:231128116-231128609 [Sequence]		Forward	1	494	494 [Sequence]	844	0.0	96.56 [Alignment]

Figure 5: The CGI sequence specific to gene *CPNE4*.

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
X:68217045-68217552 [Sequence]	OPHN1	Forward	1	508	508 [Sequence]	1007	0.0	100.00 [Alignment]
9:97586395-97586891 [Sequence]	TMOD1	Reverse	12	508	497 [Sequence]	858	0.0	96.78 [Alignment]
8:77088551-77089048 [Sequence]		Forward	11	508	498 [Sequence]	876	0.0	97.19 [Alignment]
7:104563072-104563576 [Sequence]	LHFPL3	Reverse	4	508	505 [Sequence]	882	0.0	97.03 [Alignment]
6:83803133-83803637 [Sequence]		Reverse	4	508	505 [Sequence]	874	0.0	96.83 [Alignment]
3:15186798-15187302 [Sequence]	COL6A4P1	Forward	4	508	505 [Sequence]	858	0.0	96.44 [Alignment]
18:60035071-60035568 [Sequence]		Reverse	11	508	498 [Sequence]	860	0.0	96.79 [Alignment]
13:95740201-95740705 [Sequence]	DNAJC3	Reverse	4	508	505 [Sequence]	842	0.0	96.04 [Alignment]
13:94372333-94372837 [Sequence]	GPC6	Forward	4	508	505 [Sequence]	858	0.0	96.44 [Alignment]
12:108997081-108997584 [Sequence]	SVOP	Forward	4	508	505 [Sequence]	850	0.0	96.44 [Alignment]

Figure 6: The CGI sequence specific to gene *ESR1*.

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
1:43591527-43591555 [Sequence]	PTPRF	Reverse	33	61	29 [Sequence]	58.0	7e-05	100.00 [Alignment]
1:43591279-43591374 [Sequence]	PTPRF	Reverse	214	309	96 [Sequence]	56.0	3e-04	82.29 [Alignment]
19:5243867-5244323 [Sequence]	PTPRS	Forward	1	457	457 [Sequence]	906	0.0	100.00 [Alignment]

Figure 7: The CGI sequence specific to gene *PTPRS*.

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
8:143296212-143296248 [Sequence]	ZNF696	Forward	320	356	37 [Sequence]	42.1	3.8	89.19 [Alignment]
7:149432531-149432592 [Sequence]	ZNF777	Reverse	368	429	62 [Sequence]	75.8	3e-10	90.32 [Alignment]
7:149432078-149432117 [Sequence]	ZNF777	Reverse	399	438	40 [Sequence]	48.1	0.062	90.00 [Alignment]
7:149224516-149224548 [Sequence]	ZNF282	Forward	321	353	33 [Sequence]	42.1	3.8	90.91 [Alignment]
7:149224342-149224375 [Sequence]	ZNF282	Forward	399	432	34 [Sequence]	44.1	0.97	91.18 [Alignment]
7:149223860-149224303 [Sequence]	ZNF282	Forward	1	444	444 [Sequence]	880	0.0	100.00 [Alignment]
1:119622838-119622872 [Sequence]	ZNF697	Reverse	318	352	35 [Sequence]	46.1	0.24	91.43 [Alignment]
19:58471321-58471355 [Sequence]	ZNF324	Forward	384	418	35 [Sequence]	46.1	0.24	91.43 [Alignment]
19:58455773-58455805 [Sequence]	ZNF324B	Forward	384	416	33 [Sequence]	42.1	3.8	90.91 [Alignment]
19:58368420-58368447 [Sequence]	ZNF837	Reverse	399	426	28 [Sequence]	48.1	0.062	96.43 [Alignment]
19:58368139-58368168 [Sequence]	ZNF837	Reverse	318	347	30 [Sequence]	52.0	0.004	96.67 [Alignment]
16:57997463-57997503 [Sequence]	ZNF319	Reverse	399	439	41 [Sequence]	42.1	3.8	87.80 [Alignment]
16:3119995-3120031 [Sequence]	ZNF205, ZNF213-AS1	Forward	317	353	37 [Sequence]	50.1	0.016	91.89 [Alignment]
16:30783260-30783298 [Sequence]	ZNF629	Reverse	318	356	39 [Sequence]	46.1	0.24	89.74 [Alignment]
16:30783090-30783128 [Sequence]	ZNF629	Reverse	320	358	39 [Sequence]	46.1	0.24	89.74 [Alignment]
16:30782843-30782878 [Sequence]	ZNF629	Reverse	318	353	36 [Sequence]	56.0	3e-04	94.44 [Alignment]
16:30417759-30417787 [Sequence]	ZNF771	Forward	399	427	29 [Sequence]	42.1	3.8	93.10 [Alignment]
15:84621192-84621223 [Sequence]	ZSCAN2	Forward	321	352	32 [Sequence]	48.1	0.062	93.75 [Alignment]
12:70366242-70366263 [Sequence]		Forward	286	307	22 [Sequence]	44.1	0.97	100.00 [Alignment]

Figure 8: The CGI sequence specific to gene *ZNF282*.

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
7:23415400-23415593 [Sequence]	IGF2BP3	Forward	3	196	195 [Sequence]	228	4e-56	90.77 [Alignment]
7:23415364-23415562 [Sequence]	IGF2BP3	Forward	3	201	200 [Sequence]	278	4e-71	93.50 [Alignment]
7:23415323-23415522 [Sequence]	IGF2BP3	Forward	68	267	201 [Sequence]	232	2e-57	90.55 [Alignment]
7:23415323-23415527 [Sequence]	IGF2BP3	Forward	32	236	206 [Sequence]	281	3e-72	93.20 [Alignment]
7:23415327-23415850 [Sequence]	IGF2BP3	Forward	1	524	524 [Sequence]	1039	0.0	100.00 [Alignment]
7:23415202-23415439 [Sequence]	IGF2BP3	Forward	15	256	243 [Sequence]	184	5e-43	86.01 [Alignment]
17:29249441-29249461 [Sequence]	CRYBA1	Forward	313	333	21 [Sequence]	42.1	4.5	100.00 [Alignment]

Figure 9: The CGI sequence specific to gene *IGF2BP3*.

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
9:110791185-110791652 [Sequence]	MUSK	Forward	1	468	468 [Sequence]	928	0.0	100.00 [Alignment]
4:78348067-78348534 [Sequence]	FRAS1	Forward	1	468	468 [Sequence]	912	0.0	99.57 [Alignment]
4:21159477-21159944 [Sequence]	KCNIP4	Forward	1	468	468 [Sequence]	896	0.0	99.15 [Alignment]
1:84057854-84058321 [Sequence]		Reverse	1	468	468 [Sequence]	896	0.0	99.15 [Alignment]
1:80944711-80945178 [Sequence]		Reverse	1	468	468 [Sequence]	904	0.0	99.36 [Alignment]
16:33958059-33958526 [Sequence]		Reverse	1	468	468 [Sequence]	896	0.0	99.15 [Alignment]
13:29647184-29647651 [Sequence]		Reverse	1	468	468 [Sequence]	896	0.0	99.15 [Alignment]
11:95436302-95436769 [Sequence]		Forward	1	468	468 [Sequence]	896	0.0	99.15 [Alignment]
11:93136725-93137192 [Sequence]		Forward	1	468	468 [Sequence]	896	0.0	99.15 [Alignment]
11:109182972-109183439 [Sequence]		Reverse	1	468	468 [Sequence]	896	0.0	99.15 [Alignment]

Figure 10: The CGI sequence specific to gene *MUSK*.

Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
X:153951536-153951556 [Sequence]	HCFC1	Reverse	435	455	21 [Sequence]	42.1	4.5	100.00 [Alignment]
CHR_HSCHR16_1_CTG1:15882124-15882148 [Sequence]	MYH11	Reverse	355	379	25 [Sequence]	42.1	4.5	96.00 [Alignment]
9:93912123-93912145 [Sequence]		Forward	383	405	23 [Sequence]	46.1	0.29	100.00 [Alignment]
20:57652333-57652852 [Sequence]	PMEPA1	Forward	1	520	520 [Sequence]	1031	0.0	100.00 [Alignment]
18:13645177-13645386 [Sequence]	LDLRAD4, RP11-701H16.4	Reverse	1	210	210 [Sequence]	131	6e-27	82.86 [Alignment]
16:15797000-15797024 [Sequence]	MYH11	Reverse	355	379	25 [Sequence]	42.1	4.5	96.00 [Alignment]
10:110481828-110481848 [Sequence]	RP11-525A16.4	Reverse	380	400	21 [Sequence]	42.1	4.5	100.00 [Alignment]

Figure 11: The CGI sequence specific to gene *PMEPA1*.

APPENDIX B

UNIVERSITY OF KWAZULU-NATAL

INYUVESI YAKWAZULU-NATALI

25 March 2015

Miss Farzeen Kader
University Road
Westville
Private Bag X 54001
Durban
4000
farzeenkader68@gmail.com

PROTOCOL: Development and validation of DNA Methylation Specific Markers for Body Fluid and Tissue Identification in the South African population for Forensic Application.
REF: BE221/14

EXPEDITED APPLICATION

A sub-committee of the Biomedical Research Ethics Committee has considered and noted your application received on 10 April 2014.

The study was provisionally approved pending appropriate responses to queries raised. Your responses received on 12 March 2015 to queries raised on 10 March 2015 have been noted by a sub-committee of the Biomedical Research Ethics Committee. The conditions have now been met and the study is given full ethics approval.

This approval is valid for one year from **25 March 2015.** To ensure uninterrupted approval of this study beyond the approval expiry date, an application for recertification must be submitted to BREC on the appropriate BREC form 2-3 months before the expiry date.

Any amendments to this study, unless urgently required to ensure safety of participants, must be approved by BREC prior to implementation.

Your acceptance of this approval denotes your compliance with South African National Research Ethics Guidelines (2004), South African National Good Clinical Practice Guidelines (2006) (if applicable) and with UKZN BREC ethics requirements as contained in the UKZN BREC Terms of Reference and Standard Operating Procedures, all available at http://research.ukzn.ac.za/Research-Ethics/Biomedical-Research-Ethics.aspx.

BREC is registered with the South African National Health Research Ethics Council (REC 290408-009). BREC has US Office for Human Research Protections (OHRP) Federal-wide Assurance (FWA 678).

The sub-committee's decision will be **RATIFIED** by a full Committee at its meeting taking place on **14 April 2015.**

We wish you well with this study. We would appreciate receiving copies of all publications arising out of this study.

Yours sincerely

Professor J Tsoka-Gwegweni
Chair: Biomedical Research Ethics Committee

UKZN BIOMEDICAL RESEARCH ETHICS COMMITTEE

APPLICATION FOR ETHICS APPROVAL
For research with human participants (Biomedical)

INFORMED CONSENT RESOURCE FORM

Information Sheet and Consent to Participate in Research - Volunteers

Date: Jan 2014-Dec 2015

Good day

My name is Farzeen Kader from UKZN – Westville, School of Life Sciences (Genetics). Contact Number: 031 260 8617. Email address: farzeenkader68@gmail.com / 209503949@stu.ukzn.ac.za.

You are being invited to consider participating in a study that involves research. The title of the project is "The Identification of Tissue Specific Differential Methylation in Human Body Fluids and Its Potential Application in Forensics." This research study is towards the Masters of Science Degree of the Principal Investigator Ms. Farzeen Kader.

The aim and purpose of this research is to study the DNA methylation changes in different racial groups of the South African population, with respect to identified tissue-specific differentially methylated regions (tDMRs).
The study aims to develop new primers that target differentially methylated CpGs and test these primers on blood, semen and vaginal fluid and saliva (from volunteers).
The study is expected to enroll 80 participants i.e. 10 males and 10 females of 4 ethnic groups (Black, White, Indian, and Coloured). The study requires participants to provide saliva samples only. The duration of your participation if you choose to enroll is minimal, as all that is required of you is to fill in a questionnaire and provide only saliva samples. This is not a diagnostic study and all information disclosed is confidential. The study is funded by the National Research Foundation.
The samples will stored at – 20 ˚C in UKZN - Westville. The study will provide no direct benefits to participants; however it may enable development of new markers within human body fluids that may be used in forensic analysis.

There are no risks involved if you participate in this study.

This study has been ethically reviewed and approved by the UKZN Biomedical research Ethics Committee (Approval Number: BE221/14).

In the event of any problems or concerns/questions you may contact the researcher or the UKZN Biomedical Research Ethics Committee, contact details as follows:

BIOMEDICAL RESEARCH ETHICS ADMINISTRATION
Research Office, Westville Campus
Govan Mbeki Building
Private Bag X 54001
Durban
4000
KwaZulu-Natal, SOUTH AFRICA
Tel: 27 31 2604769 - Fax: 27 31 2604609
Email: BREC@ukzn.ac.za

Participation in this research is voluntary and you may withdraw participation at any point. In the event of refusal/withdrawal of participation you will not incur any penalty or loss.
Everything possible will be done to keep your details confidential and that they are not revealed to anyone else. We will provide you with a code name such as Volunteer 1, and this code will be written onto the tubes used for the sample collection as well as the accompanying body fluid data collection form.
The above-mentioned body fluid samples will be collected and stored at -20˚C at the Department of Genetics, UKZN - Westville. DNA will be isolated from the body fluids and stored at -20˚C until the end of the study. Thereafter it will be destroyed.
You will incur no losses or cost for participating in the study. There is no monetary gain for participation.

CONSENT

I (Full Name/Sample Number ……) have been informed about the study entitled "Identification of Tissue Specific Differential Methylation in Human Body Fluids and Its Potential Application in Forensics" by Farzeen Kader.

I understand the purpose and procedures of the study.

I have been given an opportunity to answer questions about the study and have had answers to my satisfaction.

I declare that my participation in this study is entirely voluntary and that I may withdraw at any time without affecting any treatment or care that I would usually be entitled to.

I have been informed about any available compensation or medical treatment if injury occurs to me as a result of study-related procedures.

If I have any further questions/concerns or queries related to the study I understand that I may contact the researcher at UKZN – Westville, School of Life Sciences (Genetics). Contact Number: 031 260 8617. Email address: farzeenkader68@gmail.com / 209503949@stu.ukzn.ac.za.

If I have any questions or concerns about my rights as a study participant, or if I am concerned about an aspect of the study or the researchers then I may contact:

BIOMEDICAL RESEARCH ETHICS ADMINISTRATION
Research Office, Westville Campus
Govan Mbeki Building
Private Bag X 54001
Durban
4000
KwaZulu-Natal, SOUTH AFRICA
Tel: 27 31 2604769 - Fax: 27 31 2604609
Email: BREC@ukzn.ac.za

_____ _____

Signature of Participant **Date**

_____ _____

Signature of Witness **Date**
(Where applicable)

_____ _____

Signature of Translator **Date**
(Where applicable)

QUESTIONNAIRE FOR STUDY ENTITLED:
Identification of Tissue Specific Differential Methylation in Human Body Fluids and Its Potential Application in Forensics.

Please tick the appropriate boxes where necessary. SAMPLE NUMBER :

1. Age

   Below 30 ☐ Between 30 – 50 ☐ Above 50 ☐

2. Sex

   Male ☐ Female ☐

3. Weight

   Below 40 kg ☐ Between 40 – 70 kg ☐ Above 70 kg ☐

4. Are you a citizen of South Africa?

   Yes ☐ No ☐

5. Race

   Black ☐ White ☐ Indian ☐ Coloured ☐ Other ☐

6. Have you lived outside South Africa for longer than 5 years?

   Yes ☐ No ☐

7. Do you smoke?

   Yes ☐ No ☐

8. Do you consume alcohol?

   Yes ☐ No ☐

9. Have you had blood transfusion?

   Yes ☐ No ☐

10. Do you suffer from any chronic illnesses? If yes, please elaborate.

    ……

11. Do you suffer from any cardiovascular illnesses? If yes, please elaborate.

    ……

12. Have you undergone any medical procedures? If yes, please elaborate.

    ……

13. Do you take any medication? This includes steroids. If yes, please elaborate.

    ……

14. Your assistance is well appreciated. Kindly note that the present study is not for diagnostic purposes. All information disclosed will be kept confidential.

    …… Farzeen Kader (Principal Investigator)

Table 1: Concentration of DNA in saliva obtained from 80 study subjects.

Sample Number	Ethnic Group	Age (Years)	DNA Concentration (ng/µL)
1	Black	Below 30	7.3
2	Black	Below 30	47.7
3	Black	Below 30	2.5
4	Black	Below 30	110
5	Black	Below 30	91.6
6	Black	Below 30	54.2
7	Black	30-50	38.8
8	Black	30-50	40
9	Black	30-50	2.2
10	Black	30-50	7.1
11	Black	30-50	469.5
12	Black	Below 30	6
13	Black	Below 30	8
14	Black	Below 30	146.4
15	Black	Below 30	81.5
16	Black	Below 30	230.3
17	Black	Below 30	73.9
18	Black	Below 30	26.1
19	Black	30-50	25.3
20	Black	30-50	46.2
21	Black	30-50	28.6
22	Indian	30-50	64.4
23	Indian	30-50	28.5
24	Indian	30-50	48.7
25	Indian	30-50	52.8
26	Indian	30-50	27.3
27	Indian	30-50	67.9
28	Indian	30-50	48.3
29	Indian	30-50	119.2
30	Indian	Above 50	13.2
31	Indian	Below 30	100.7
32	Indian	Below 30	15.1

33	Indian	Below 30	241.1
34	Indian	Below 30	373.6
35	Indian	Below 30	0.6
36	Indian	Below 30	33.6
37	Indian	Below 30	29.9
38	Indian	Below 30	19.9
39	Indian	Below 30	13.8
40	Indian	Below 30	98.3
41	White	Below 30	42.3
42	White	Below 30	114.9
43	White	Below 30	2.9
44	White	Below 30	76.5
45	White	Below 30	55.1
46	White	30-50	62.1
47	White	30-50	91.2
48	White	30-50	24.5
49	White	30-50	21.8
50	White	30-50	29.3
51	White	30-50	36.5
52	White	Above 50	78
53	White	Below 30	65.2
54	White	Below 30	94.3
55	White	Below 30	37.2
56	White	Below 30	2.5
57	White	Below 30	192.8
58	White	30-50	158.3
59	White	30-50	1.7
60	White	30-50	54.3
61	Coloured	30-50	96
62	Coloured	30-50	49.1
63	Coloured	30-50	82.3
64	Coloured	30-50	44.6
65	Coloured	30-50	14.2
66	Coloured	30-50	8.6
67	Coloured	30-50	7.1
68	Coloured	30-50	0.9

69	Coloured	30-50	5.6
70	Coloured	30-50	8.3
71	Coloured	Below 30	41.3
72	Coloured	Below 30	29.3
73	Coloured	Below 30	12.4
74	Coloured	Below 30	74.9
75	Coloured	30-50	40.1
76	Coloured	30-50	134.5
77	Coloured	30-50	28.8
78	Coloured	30-50	1.6
79	Coloured	30-50	3.9
80	Coloured	Above 50	12.5

Methylation profiling of restricted saliva obtained from Blacks as analysed by GeneMapper ID Software 5.

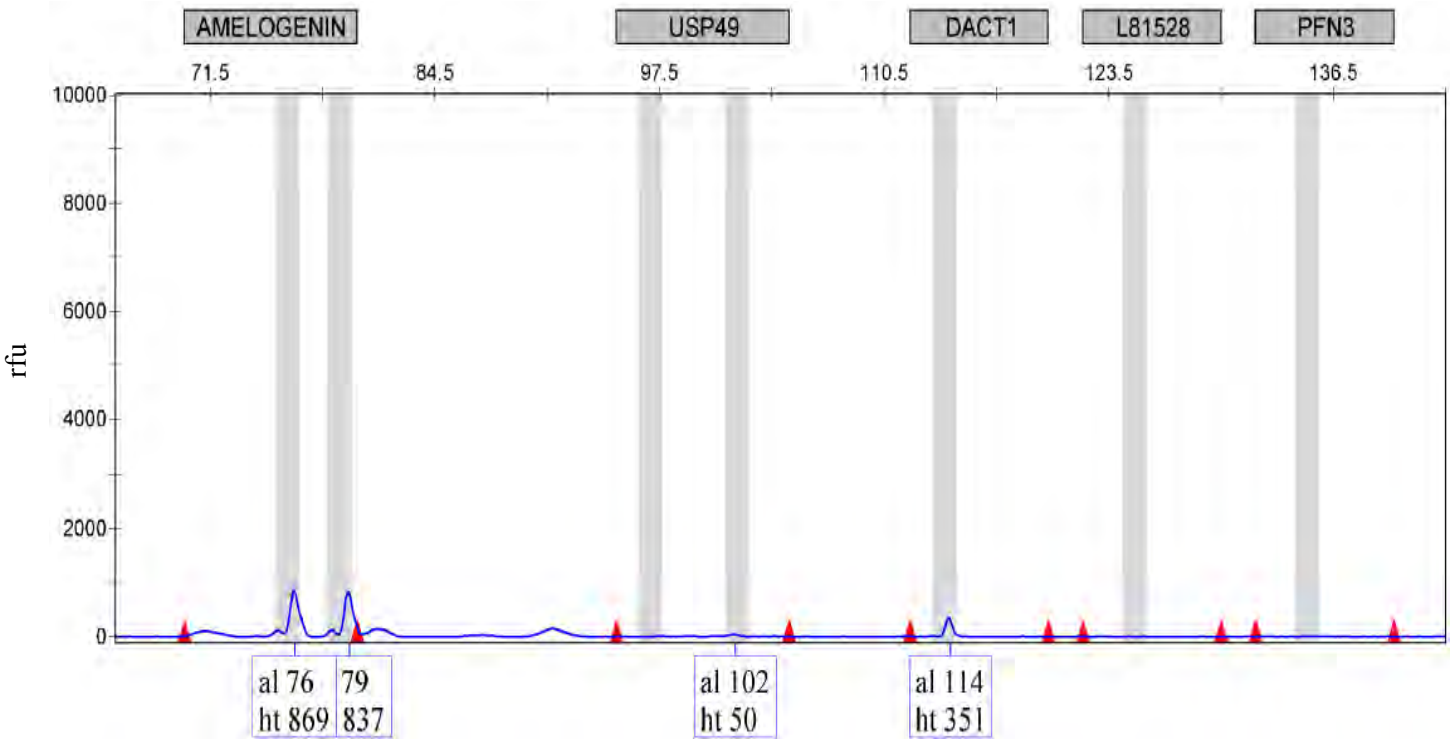

Figure 1: Methylation Profile of Saliva from Black Male (3).

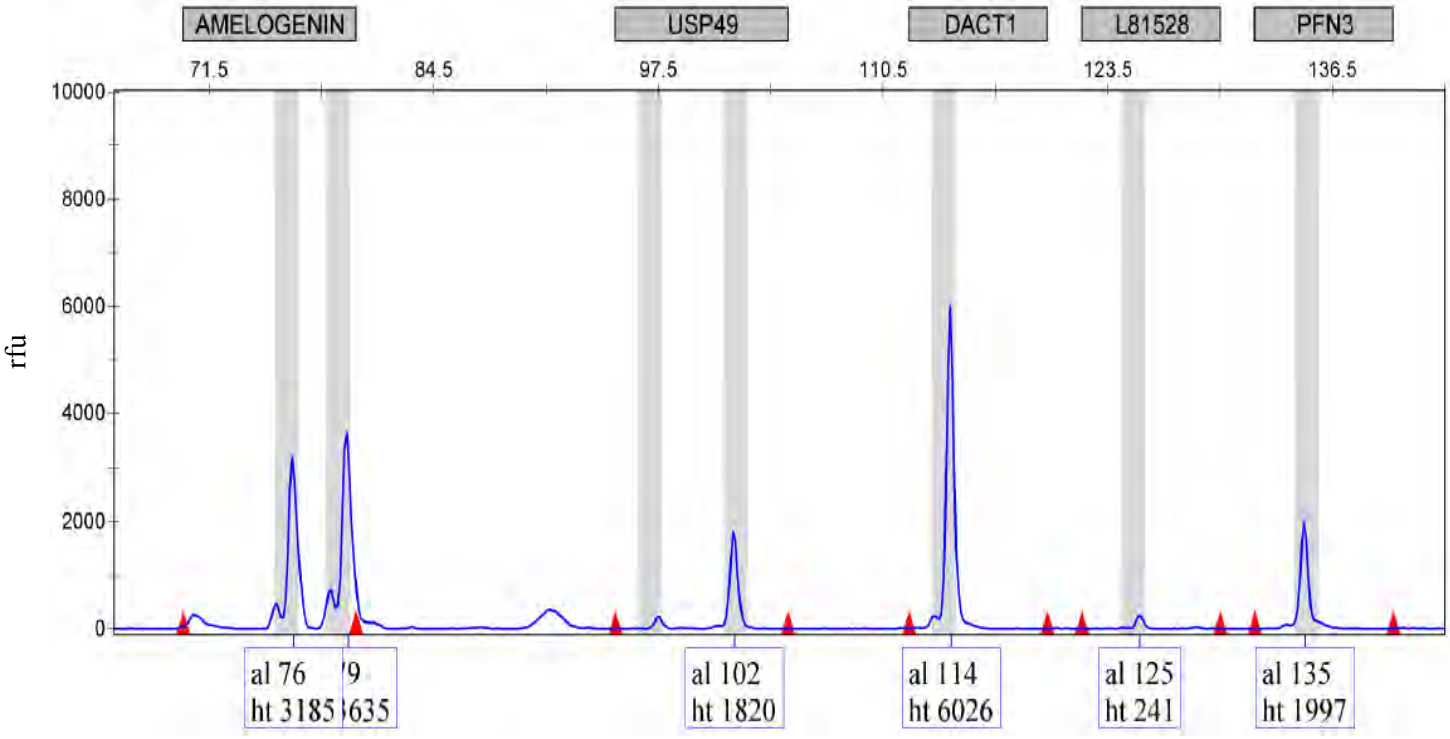

Figure 2: Methylation Profile of Saliva from Black Male (4).

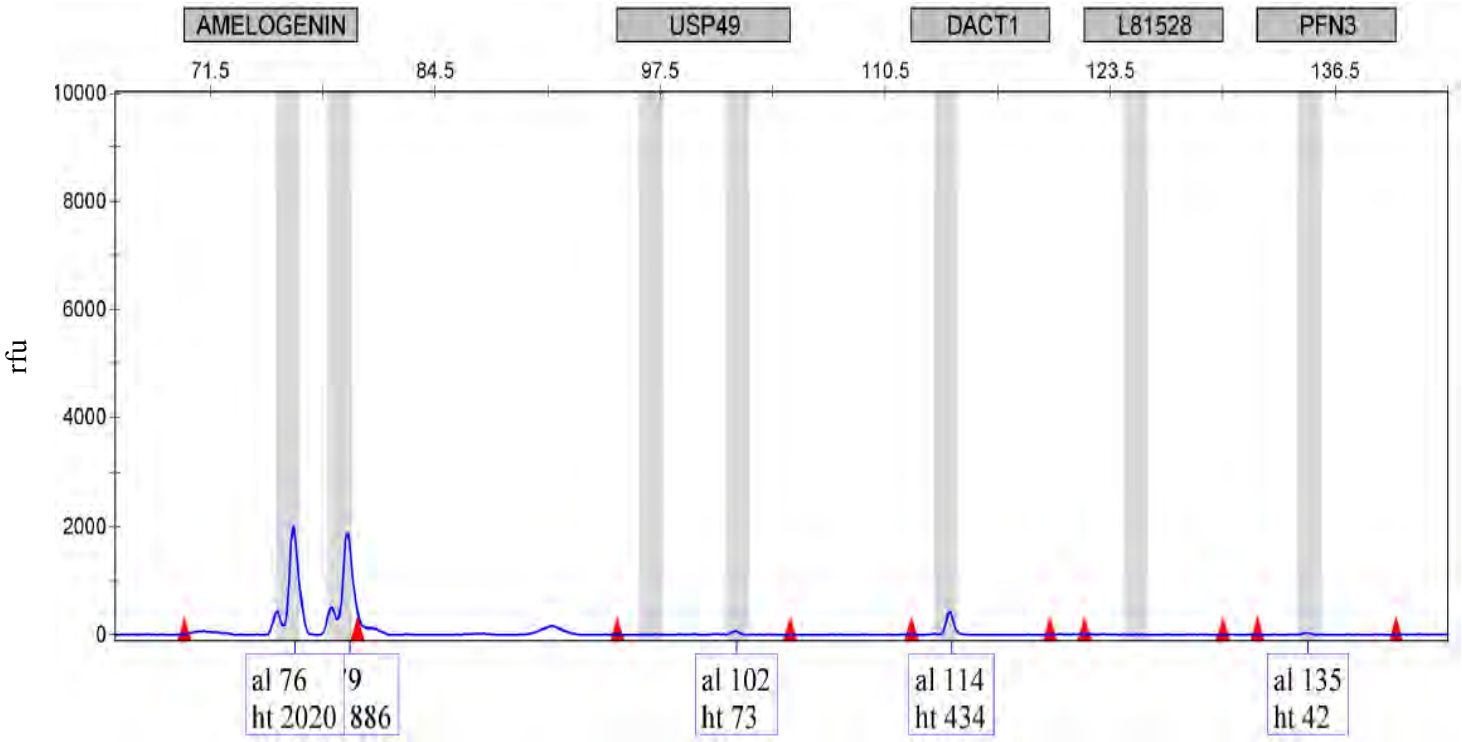

Figure 3: Methylation Profile of Saliva from Black Male (5).

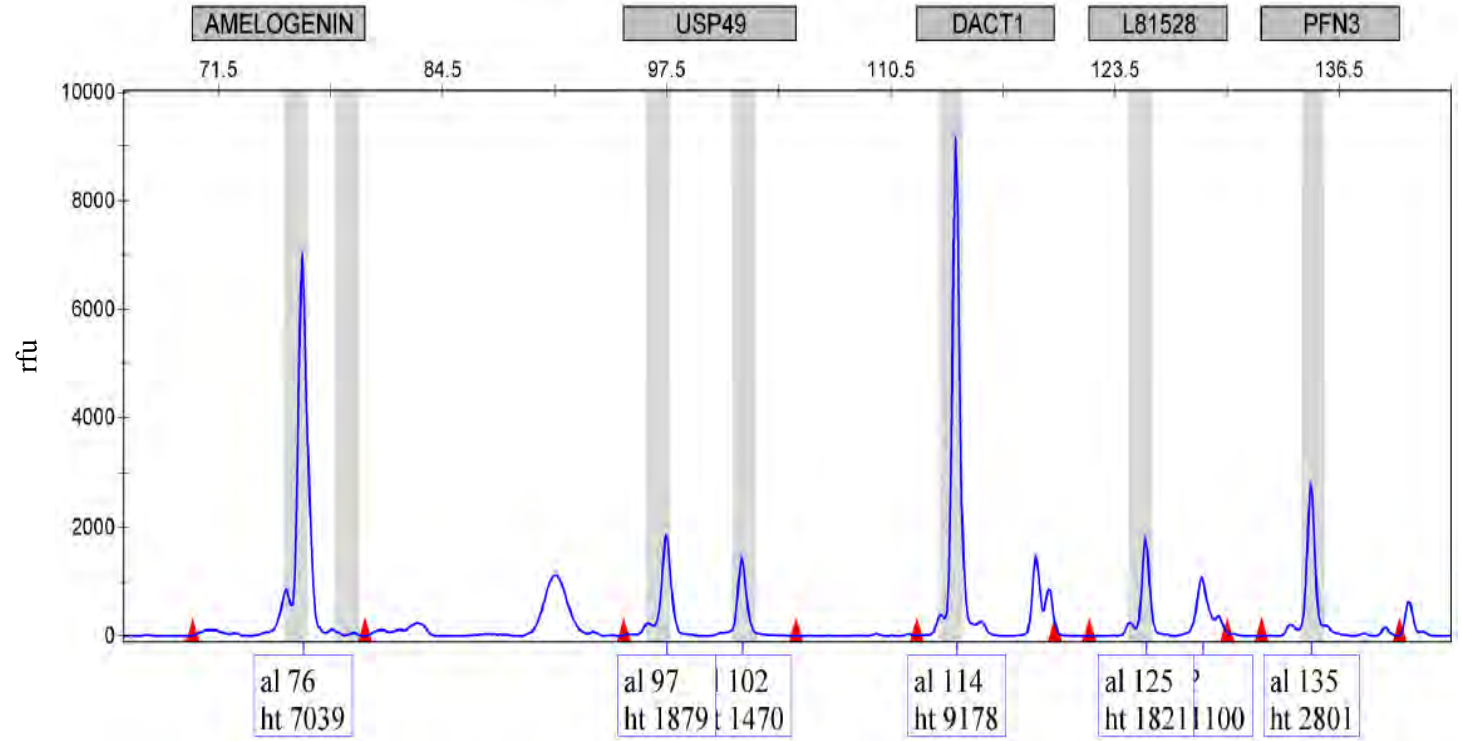

Figure 4: Methylation Profile of Saliva from Black Female (11).

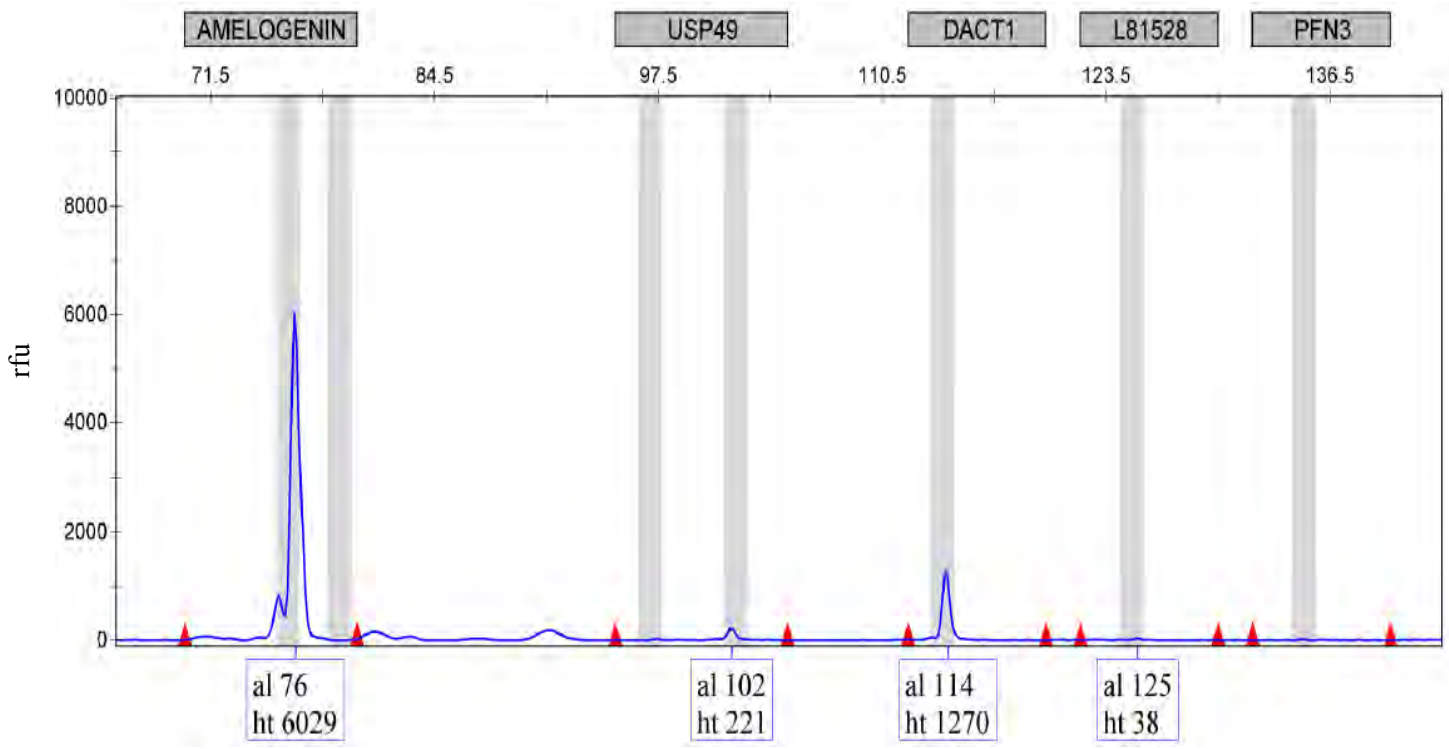

Figure 5: Methylation Profile of Saliva from Black Female (12).

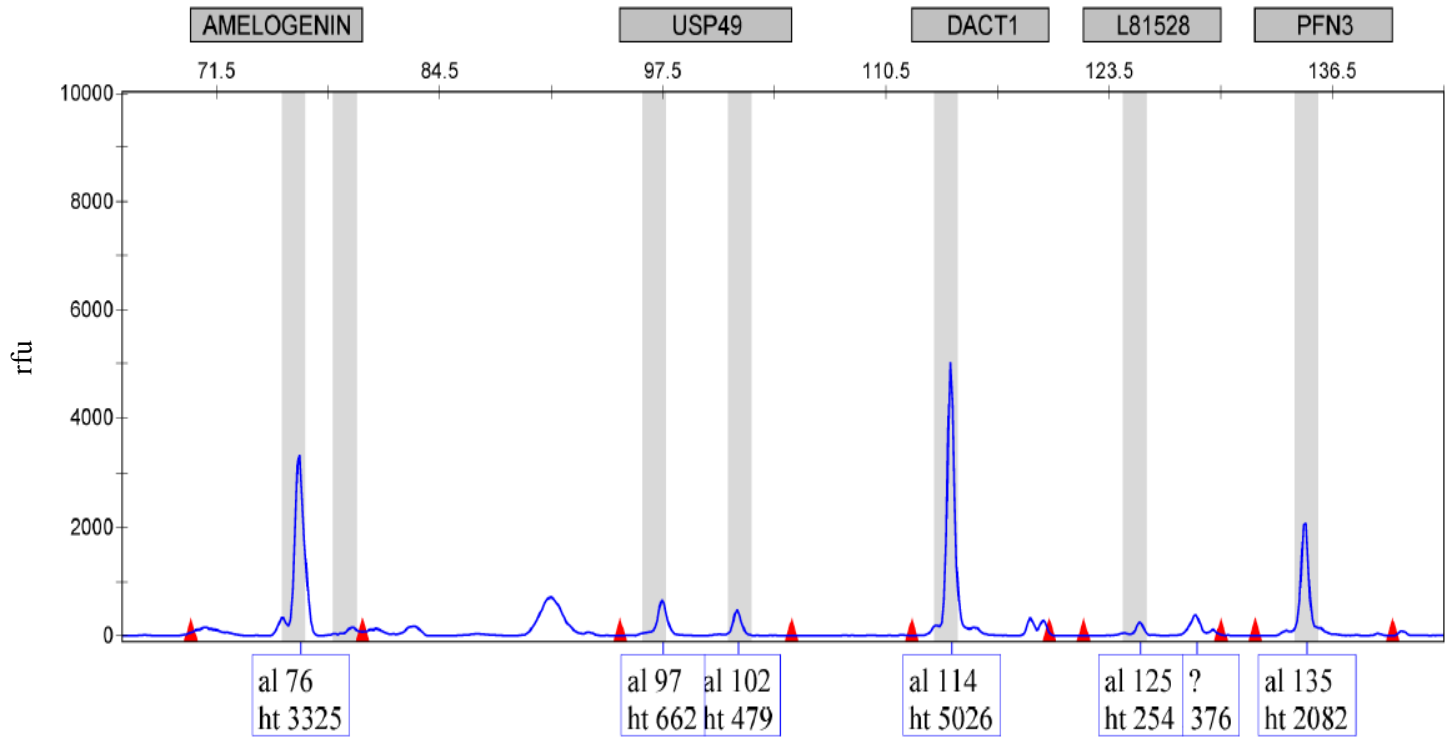

Figure 6: Methylation Profile of Saliva from Black Female (14).

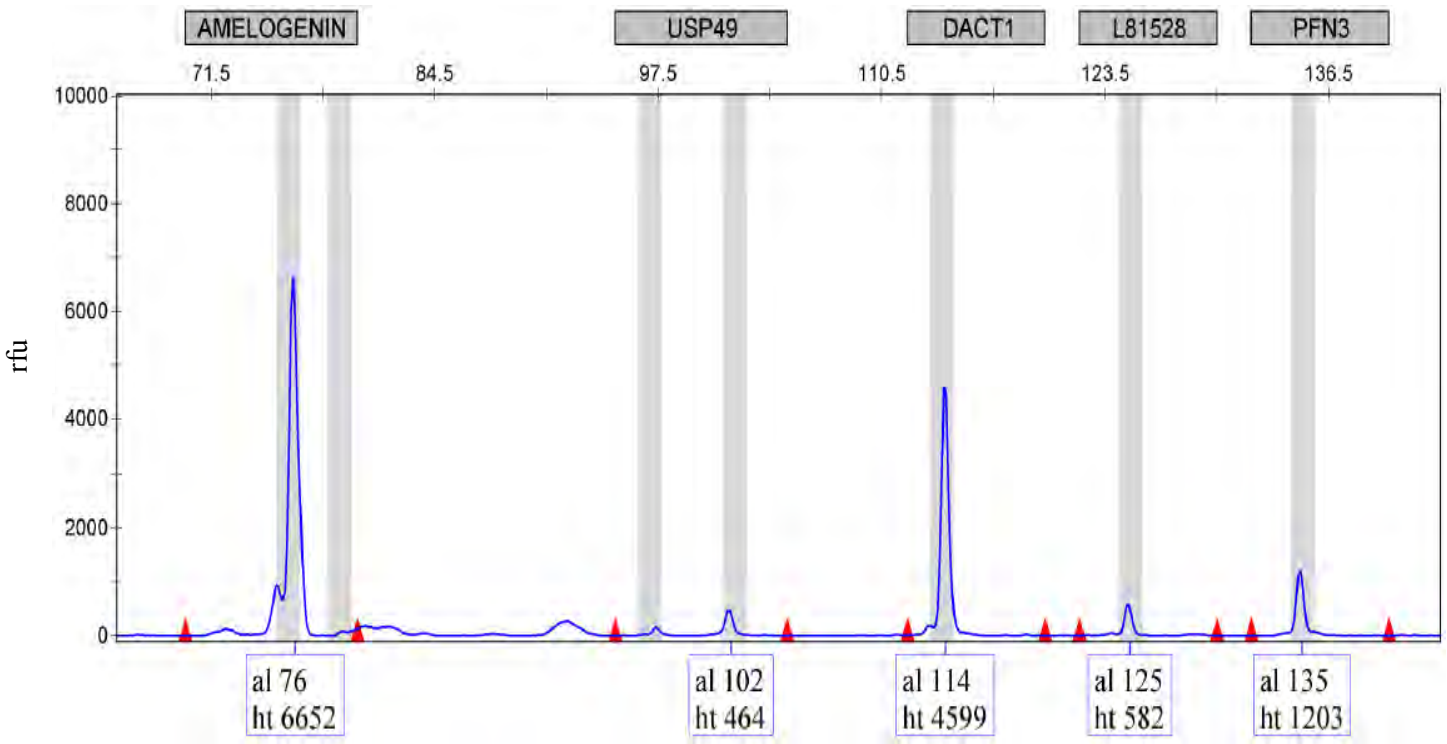

Figure 7: Methylation Profile of Saliva from Black Female (15).

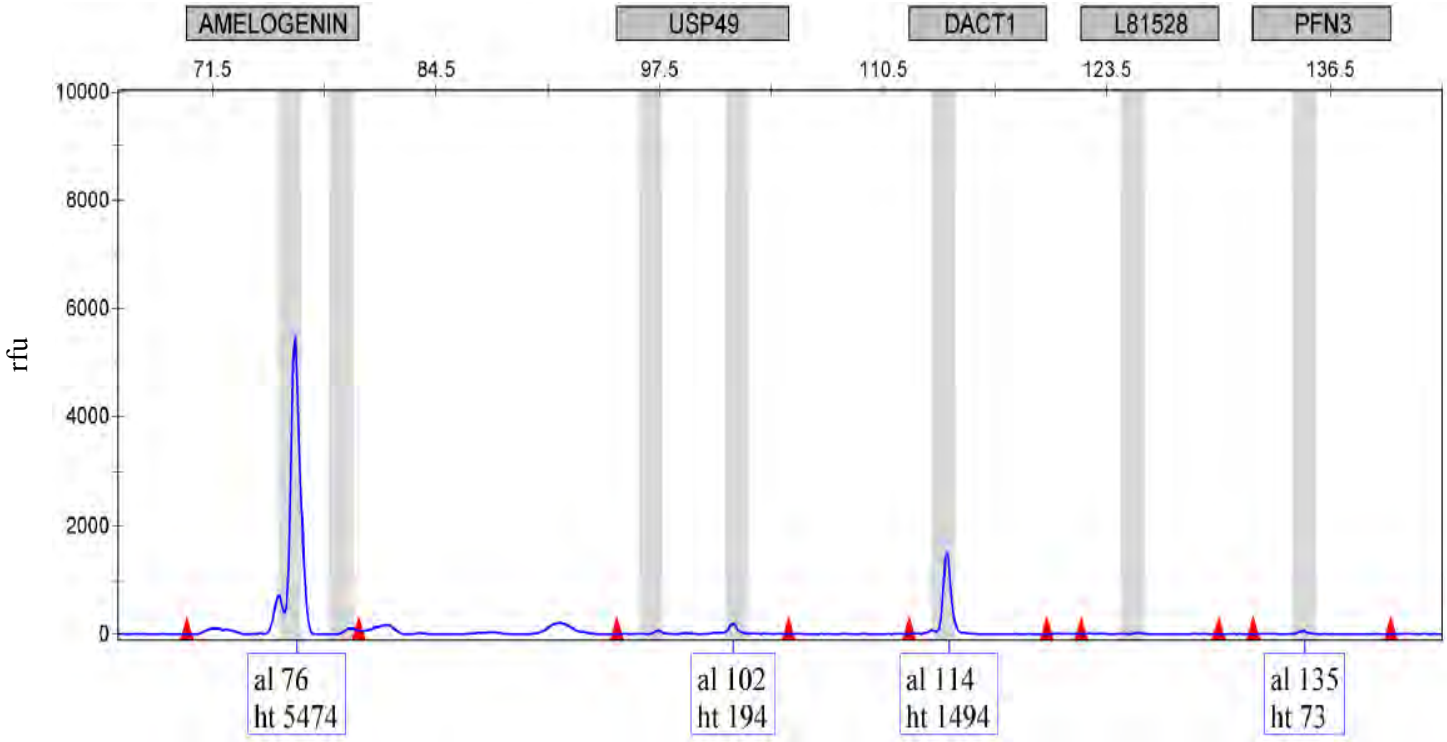

Figure 8: Methylation Profile of Saliva from Black Female (17).

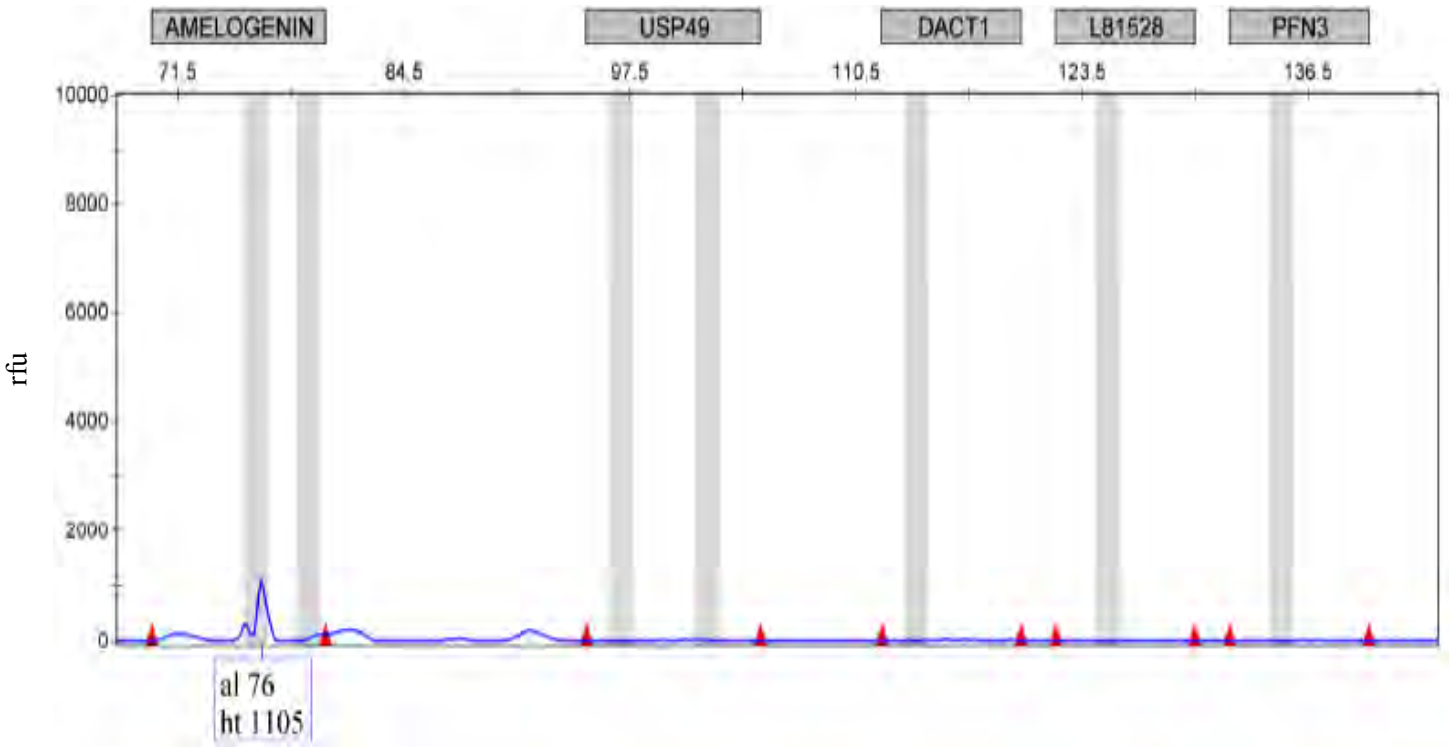

Figure 9: Methylation Profile of Saliva from Black Female (19).

Methylation profiling of restricted saliva obtained from Indians as analysed by GeneMapper ID Software 5.

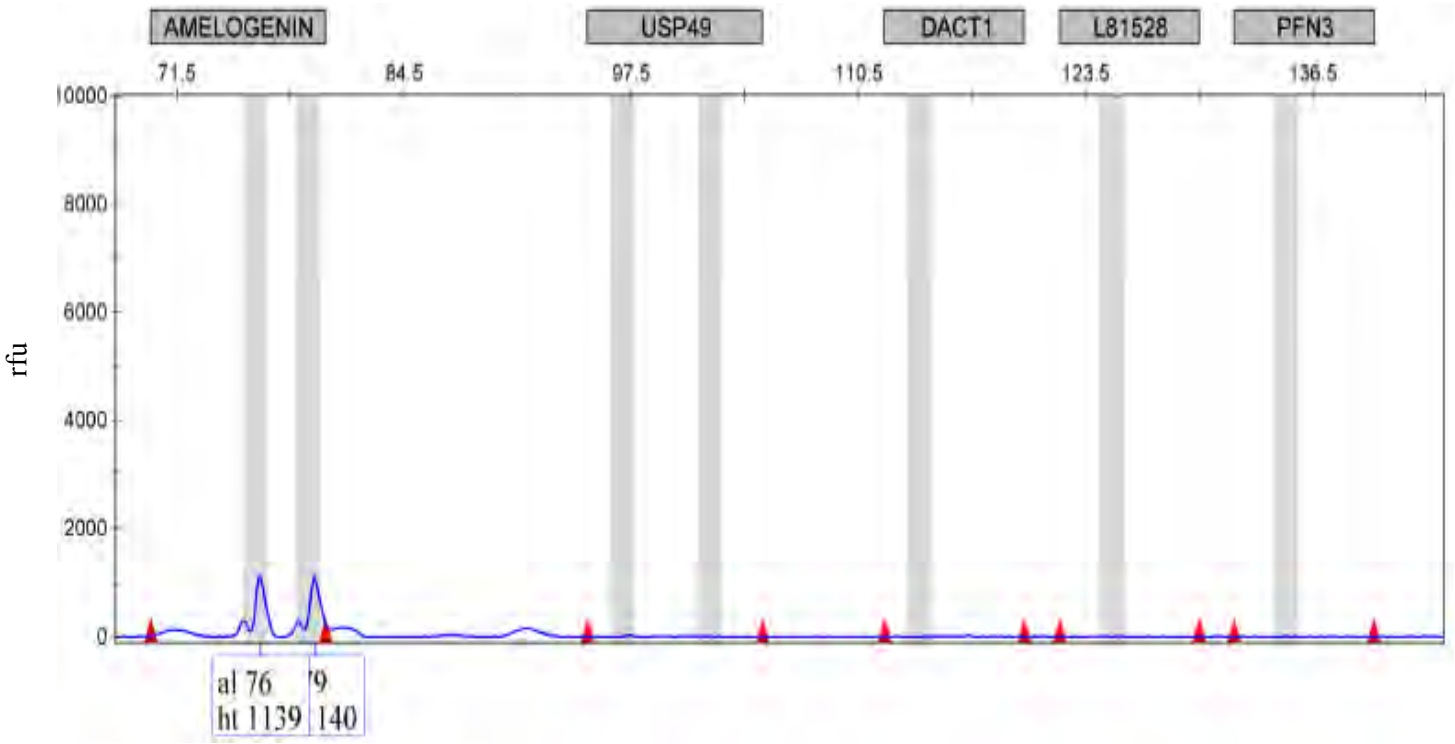

Figure 10: Methylation Profile of Saliva from Indian Male (24).

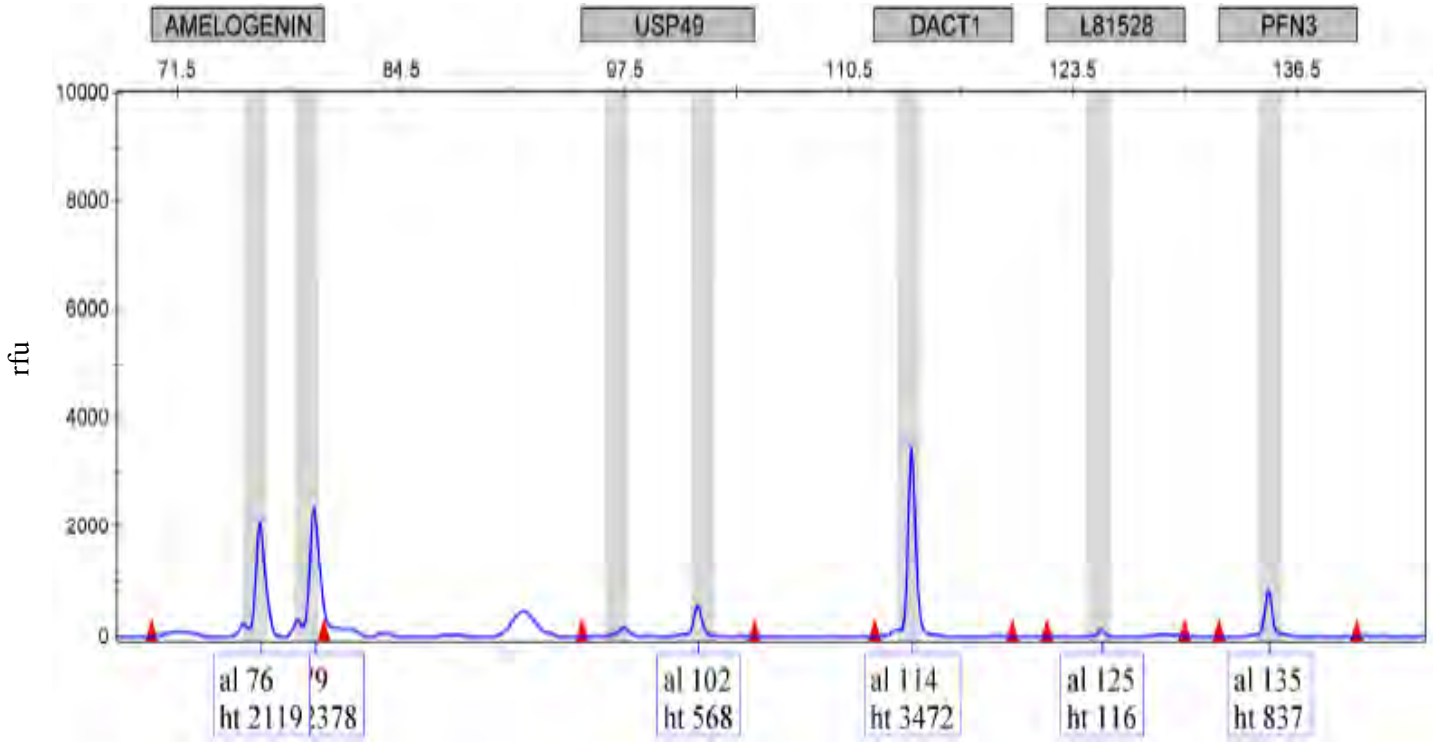

Figure 11: Methylation Profile of Saliva from Indian Male (29).

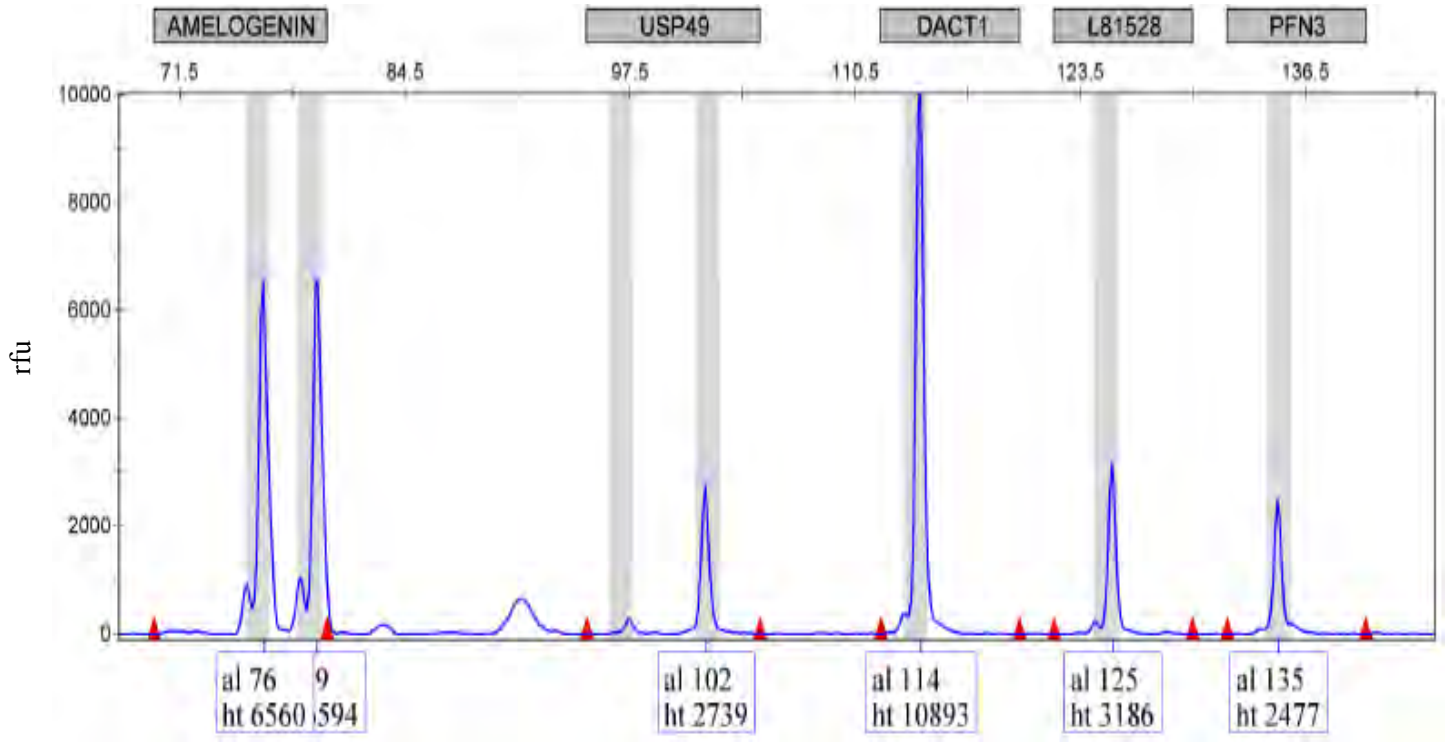

Figure 12: Methylation Profile of Saliva from Indian Male (30).

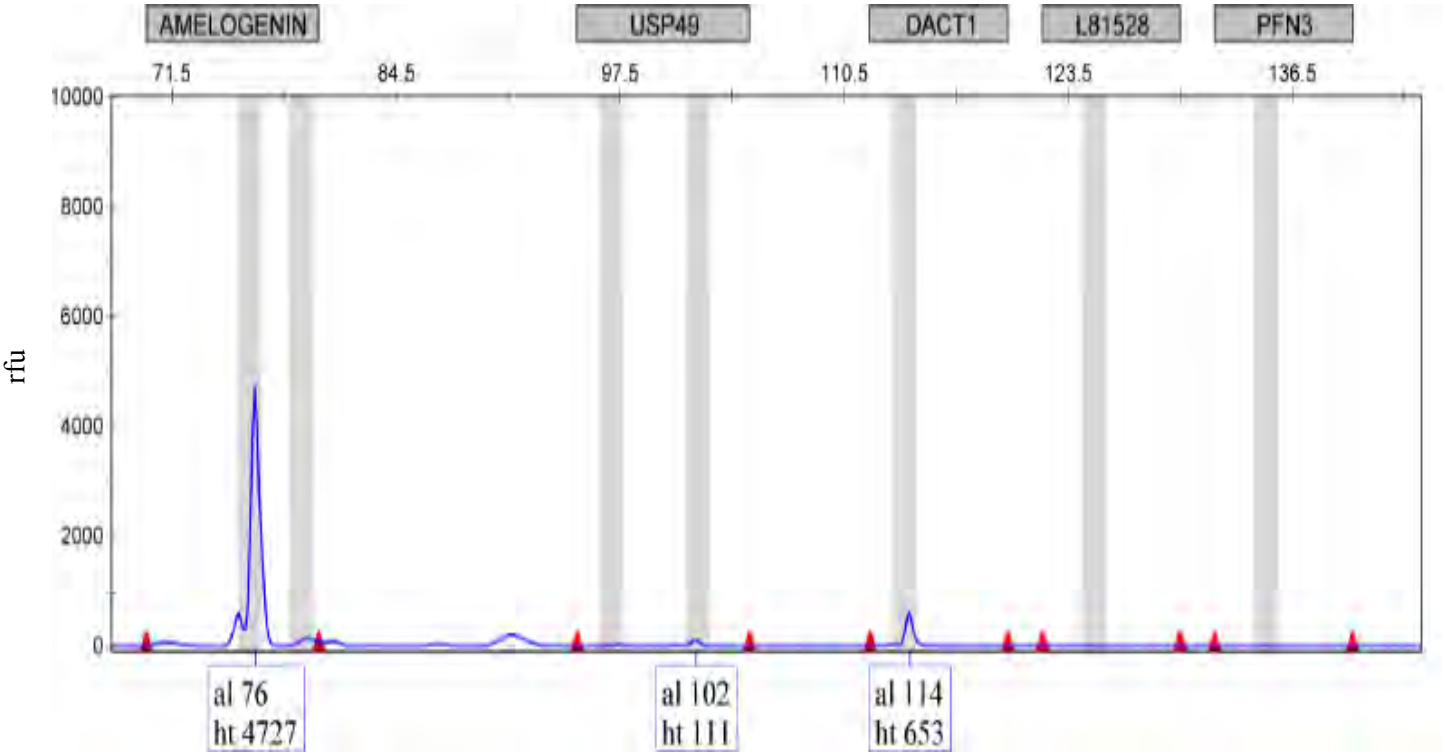

Figure 13: Methylation Profile of Saliva from Indian Female (31).

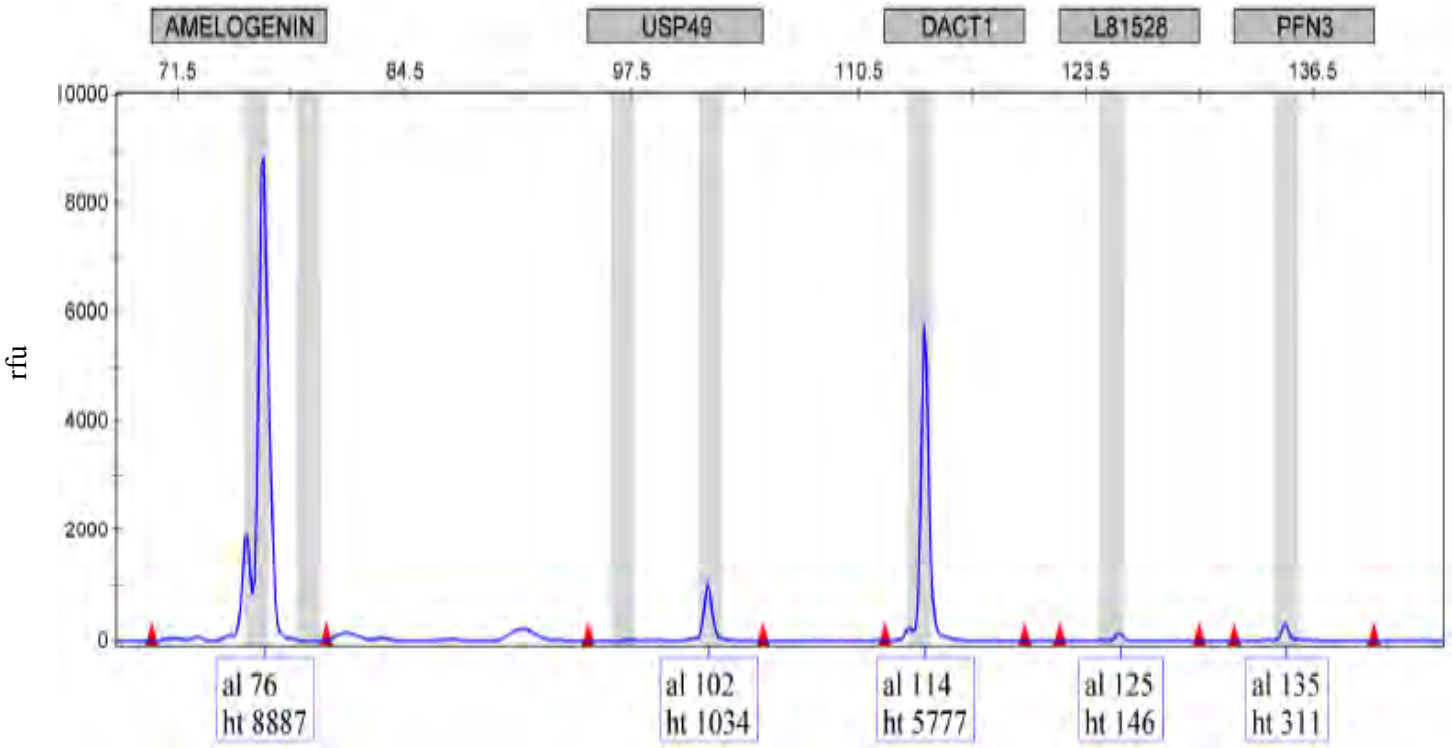

Figure 14: Methylation Profile of Saliva from Indian Female (32).

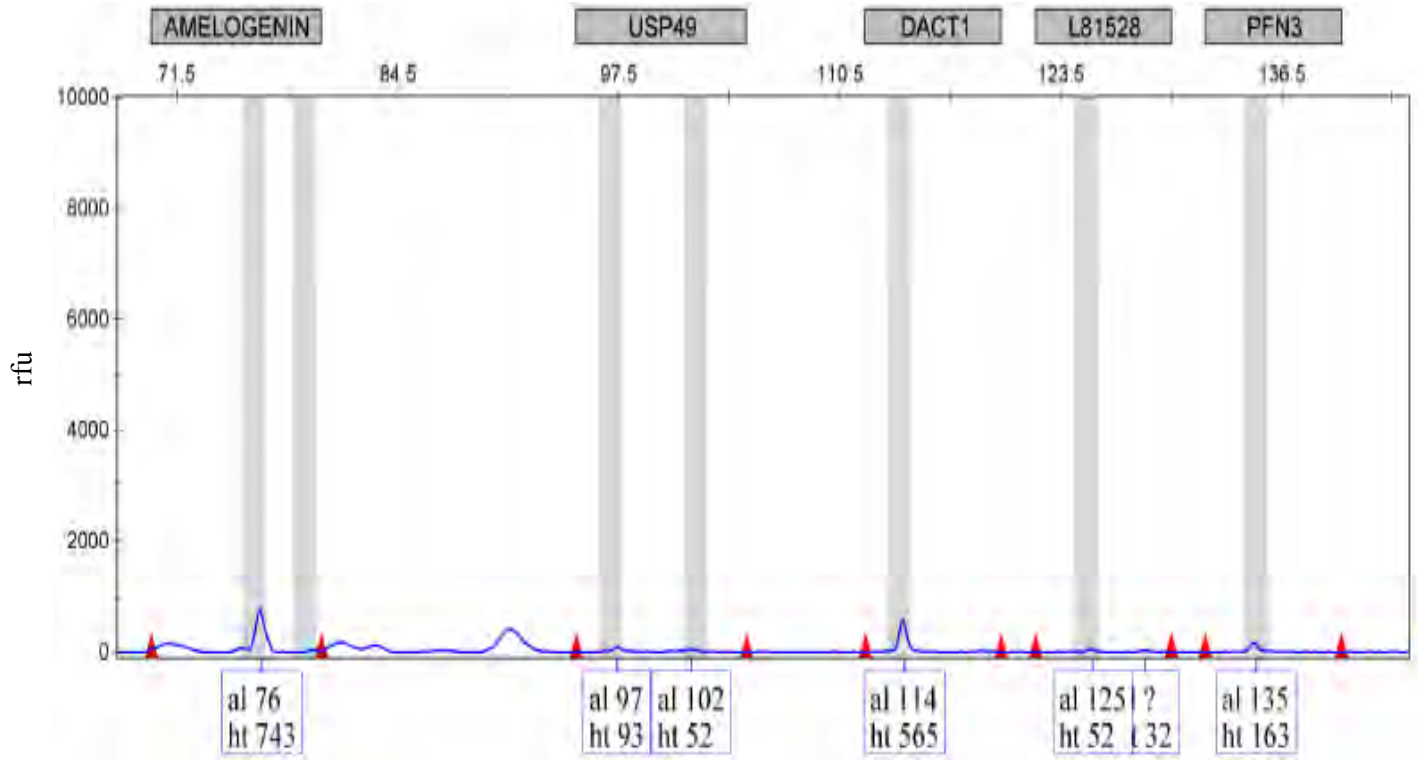

Figure 15: Methylation Profile of Saliva from Indian Female (33).

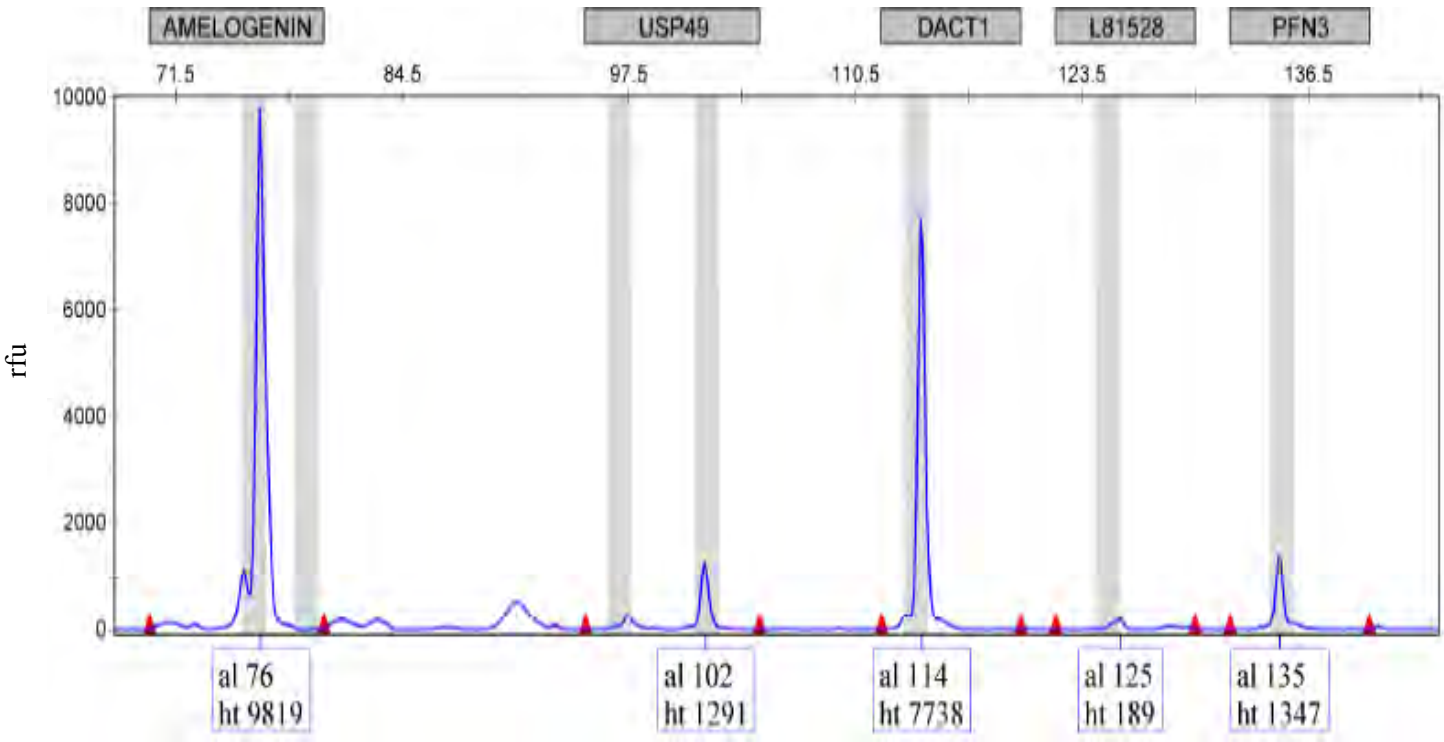

Figure 16: Methylation Profile of Saliva from Indian Female (34).

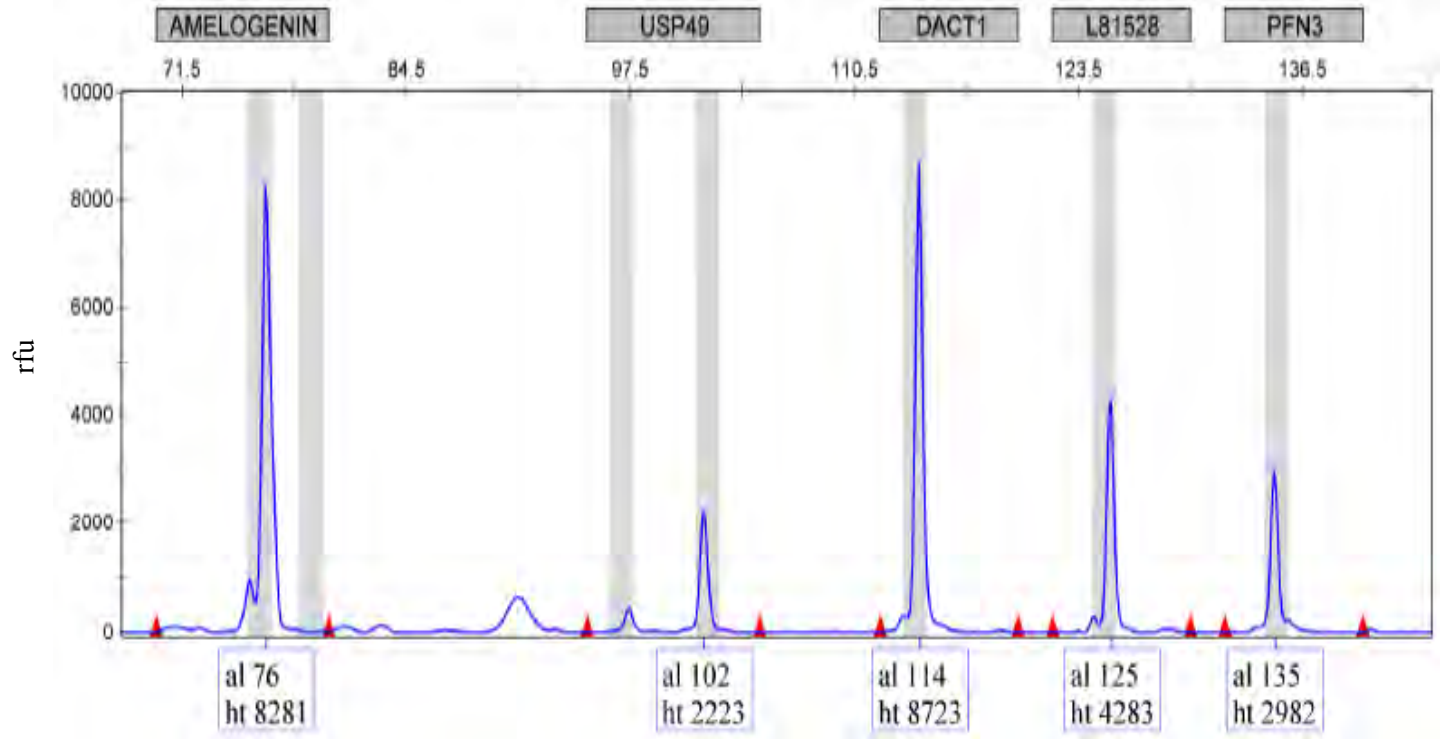

Figure 17: Methylation Profile of Saliva from Indian Female (37).

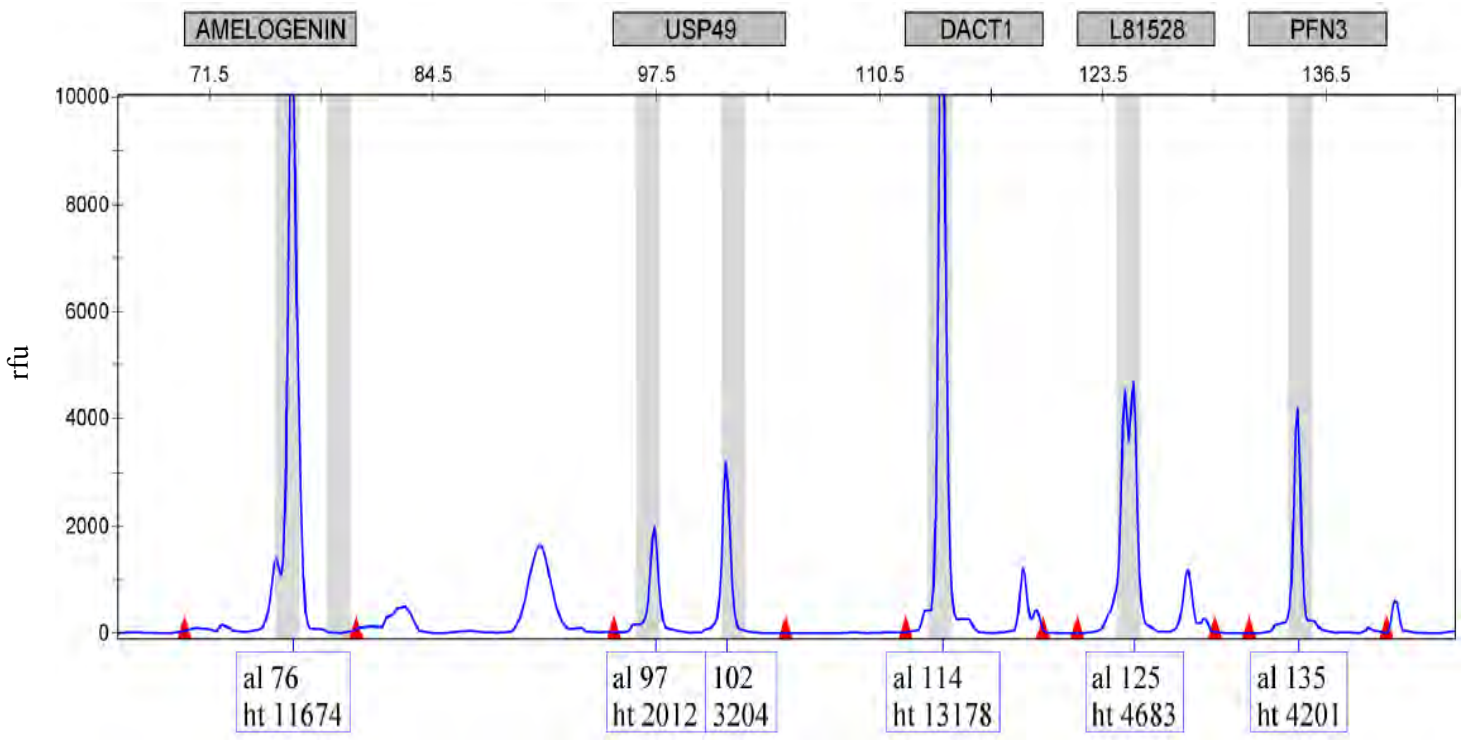

Figure 18: Methylation Profile of Saliva from Indian Female (38).

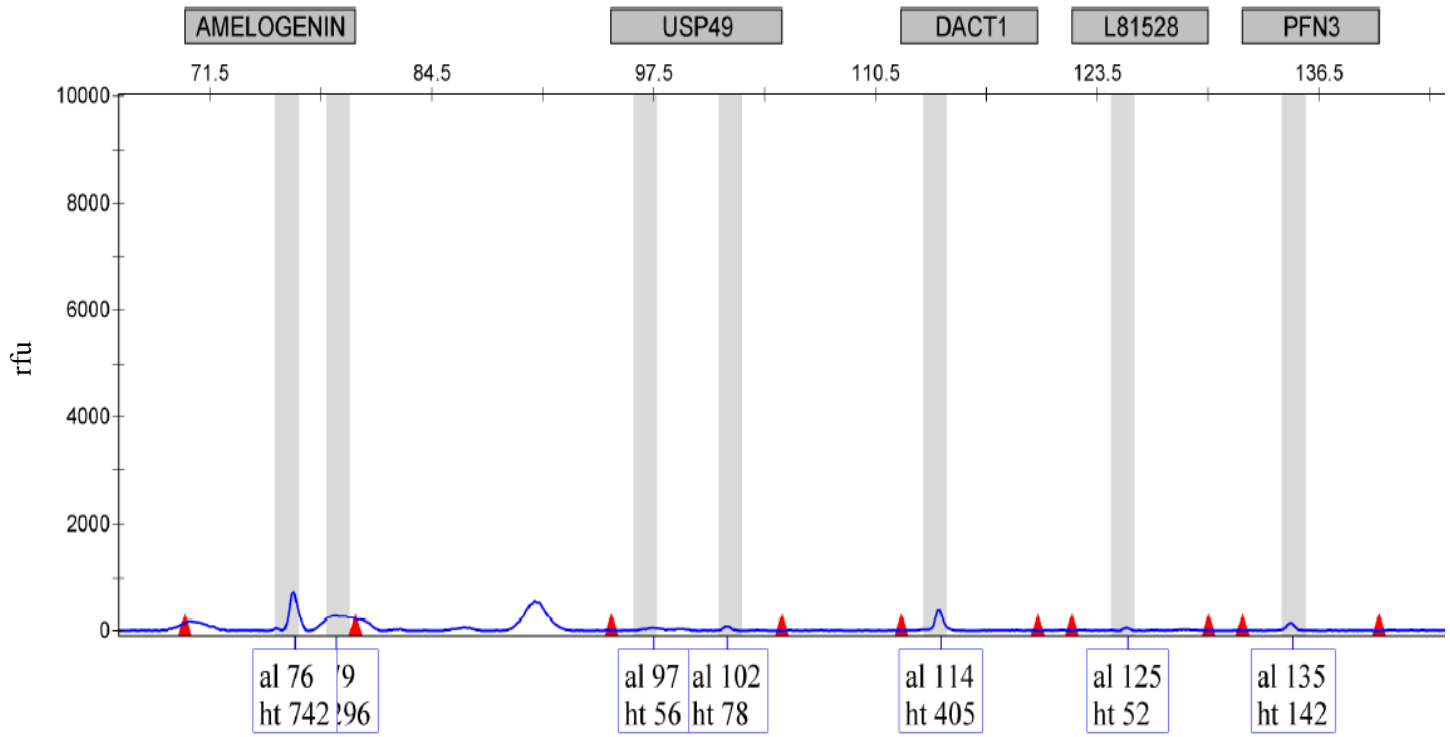

Figure 19: Methylation Profile of Saliva from Indian Female (40).

Methylation profiling of restricted saliva obtained from Whites as analysed by GeneMapper ID Software 5.

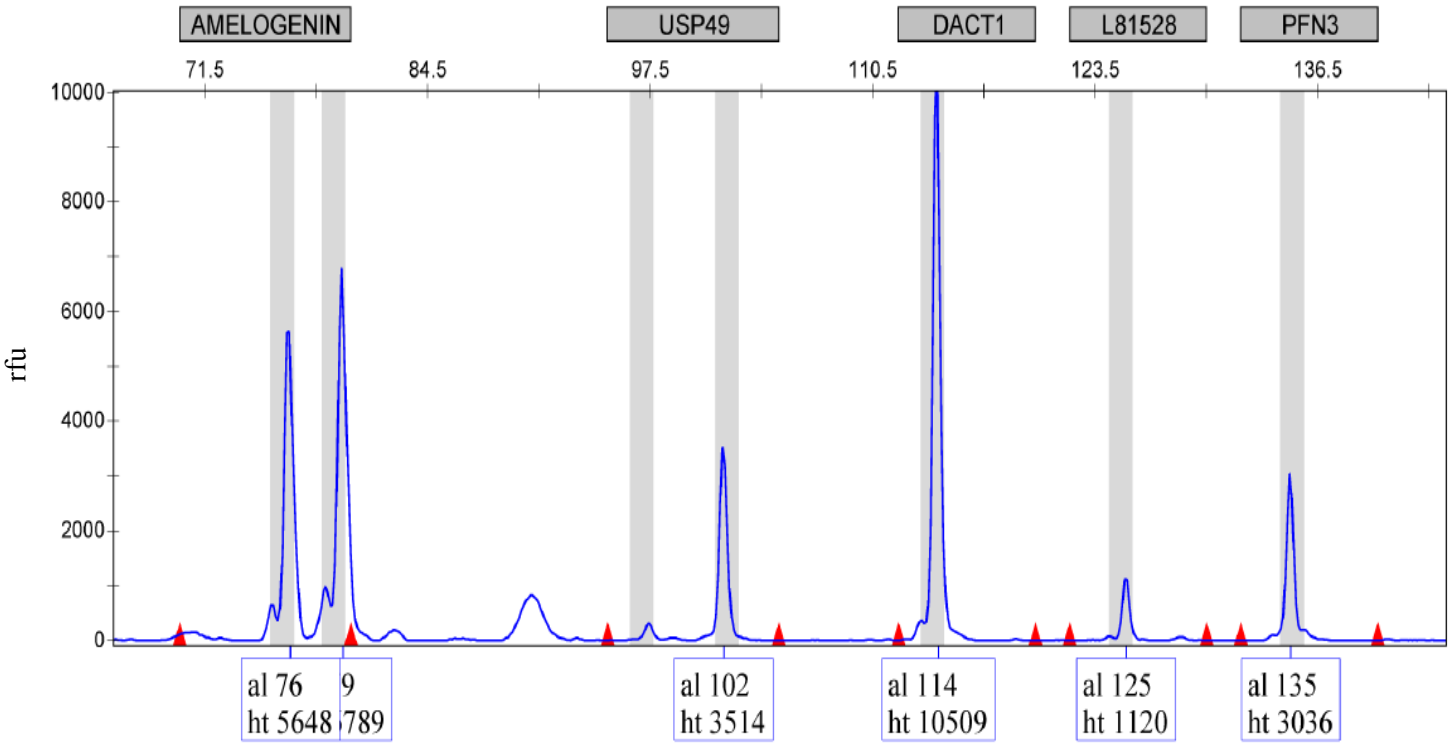

Figure 20: Methylation Profile of Saliva from White Male (41).

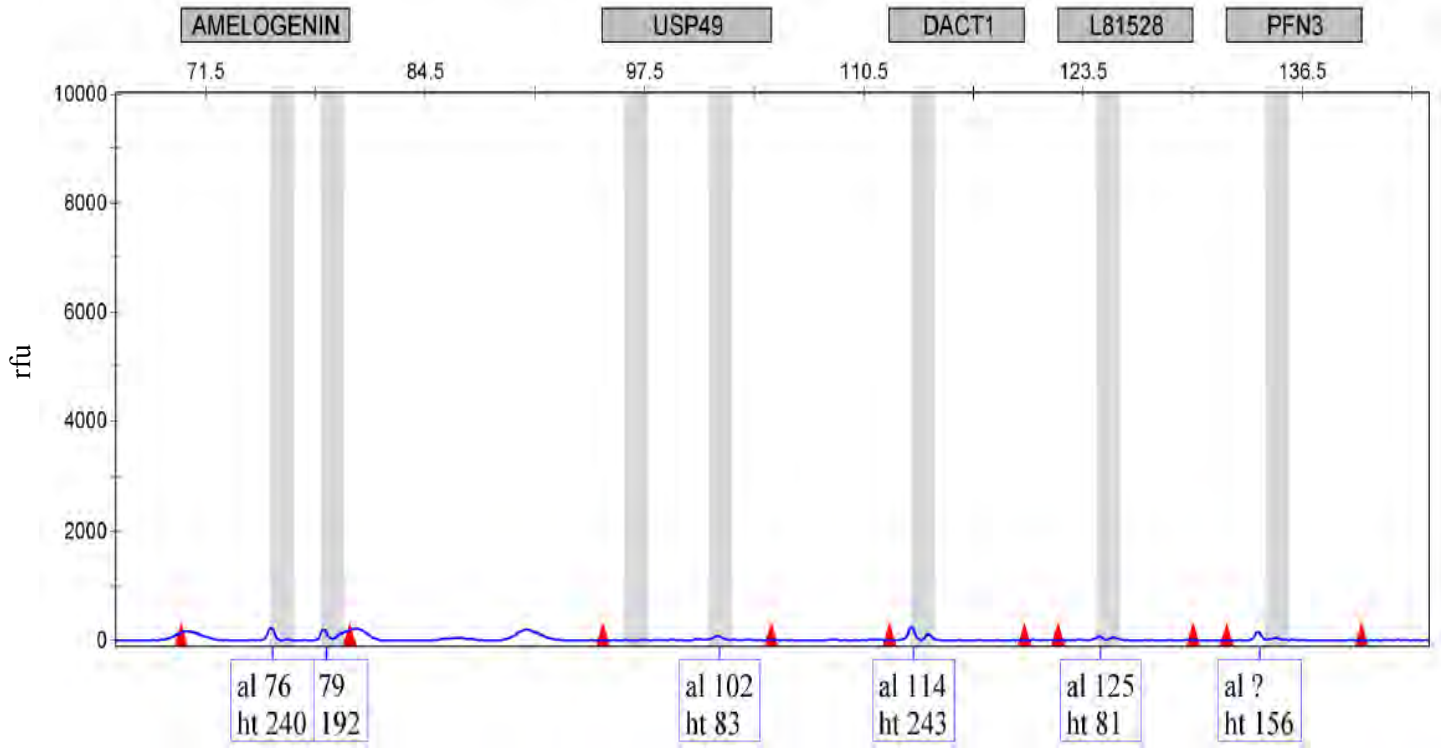

Figure 21: Methylation Profile of Saliva from White Male (42).

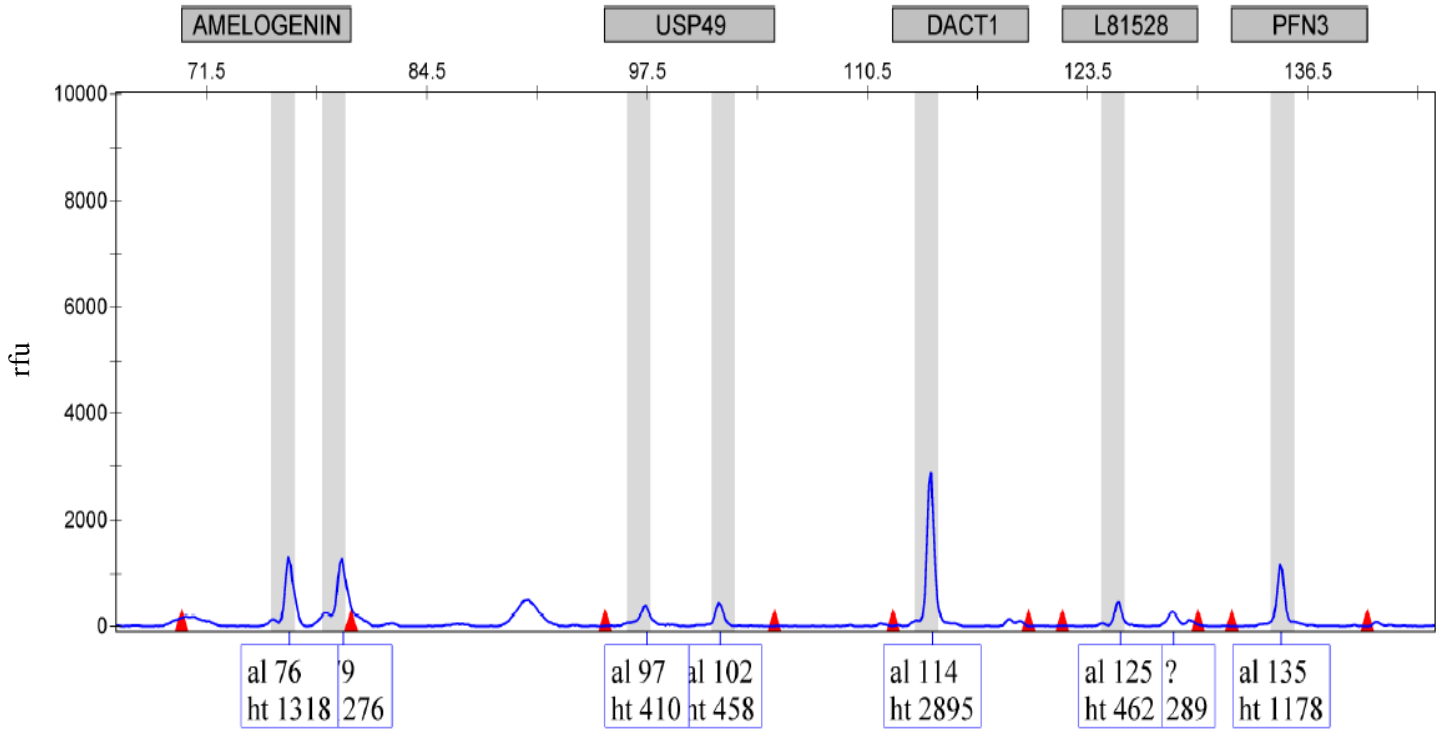

Figure 22: Methylation Profile of Saliva from White Male (43).

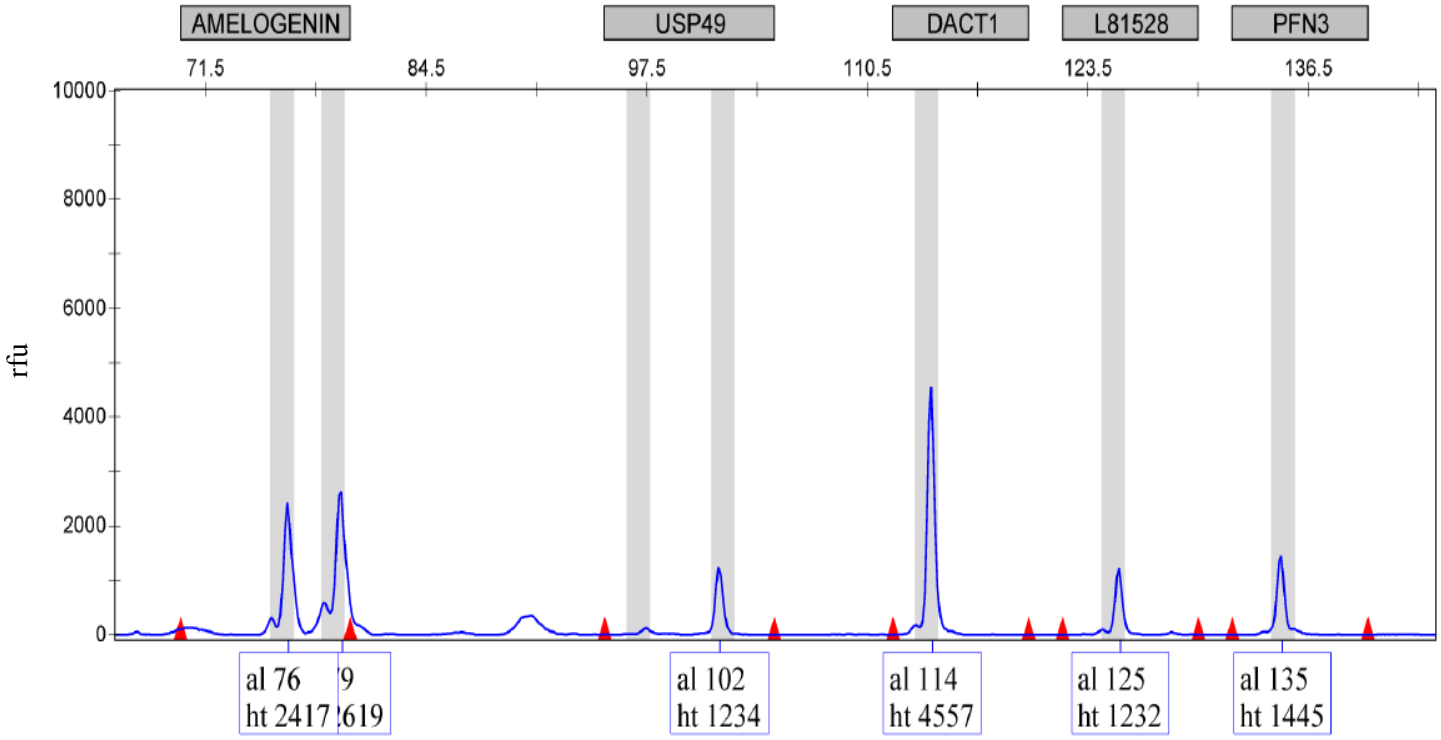

Figure 23: Methylation Profile of Saliva from White Male (44).

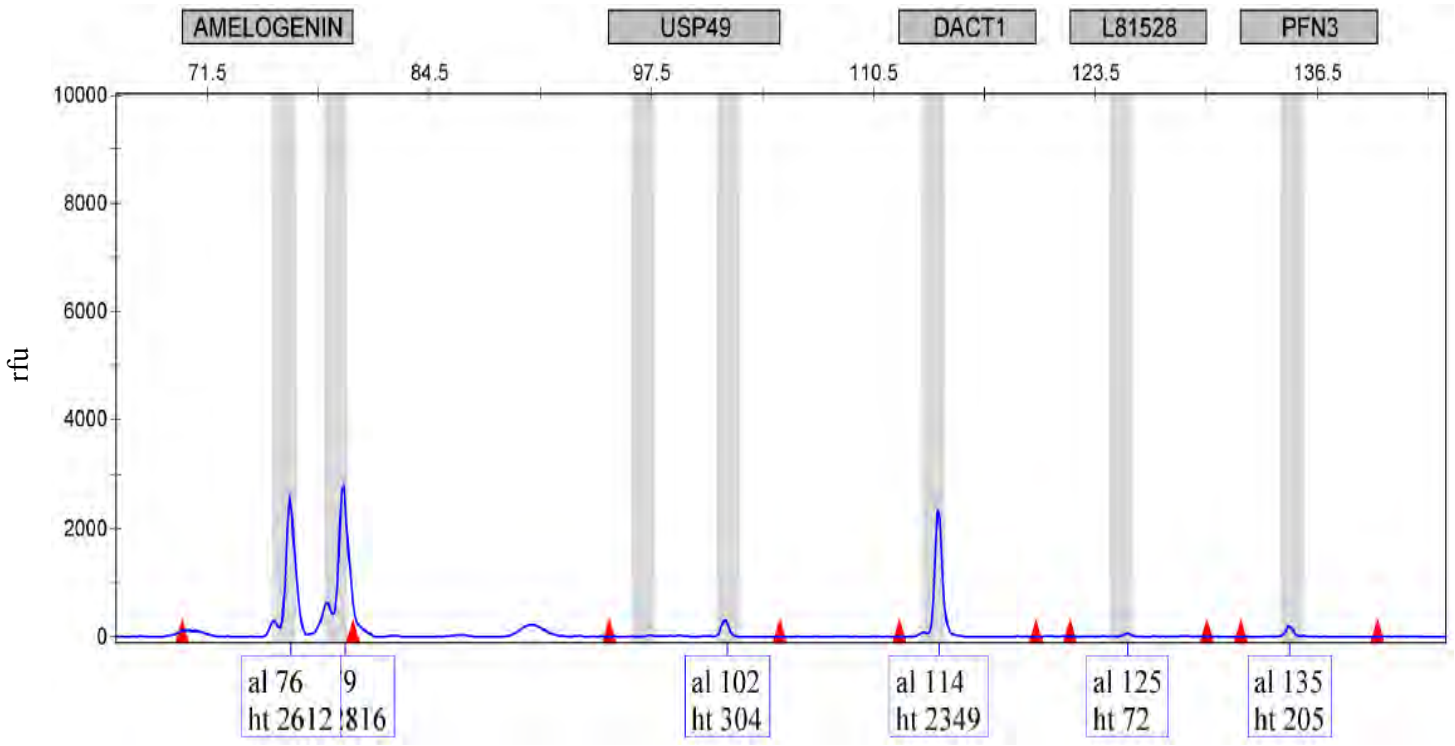

Figure 24: Methylation Profile of Saliva from White Male (45).

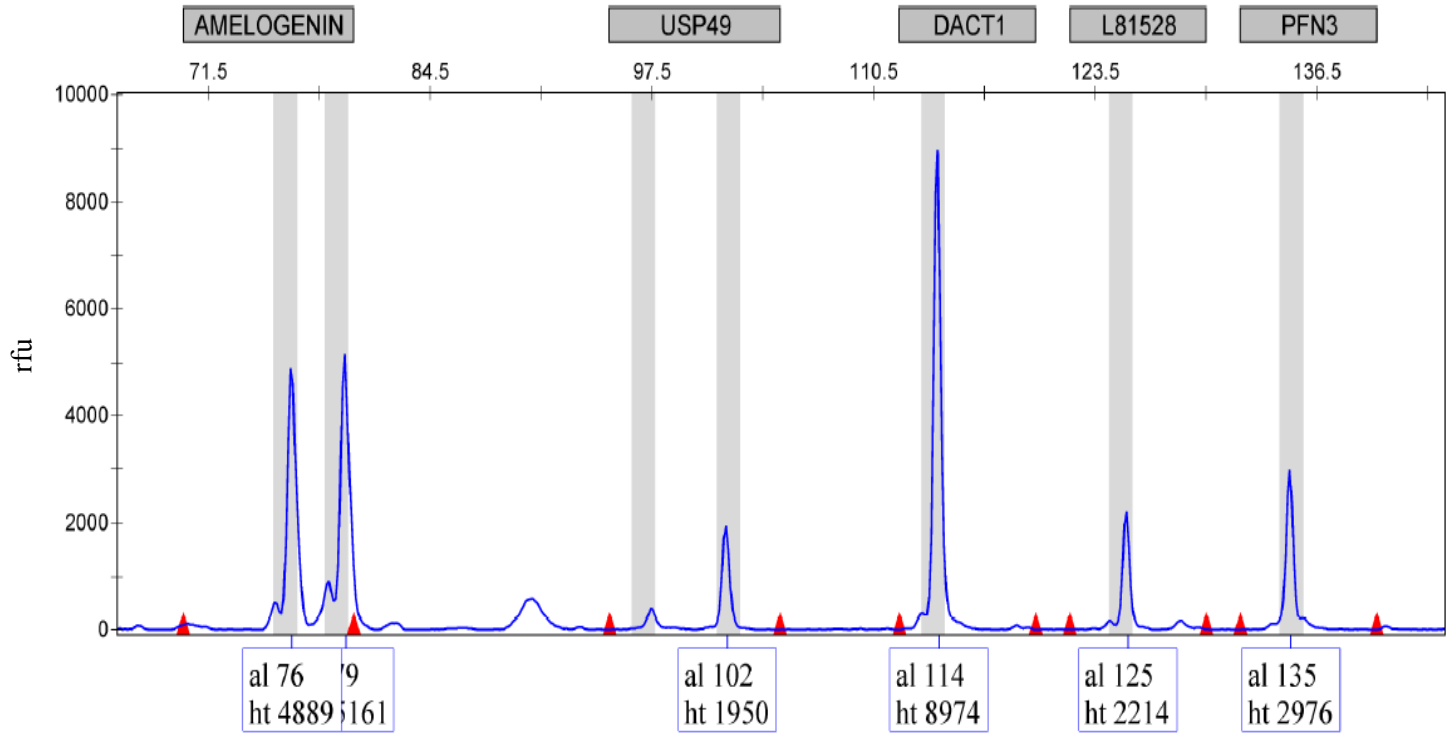

Figure 25: Methylation Profile of Saliva from White Male (46).

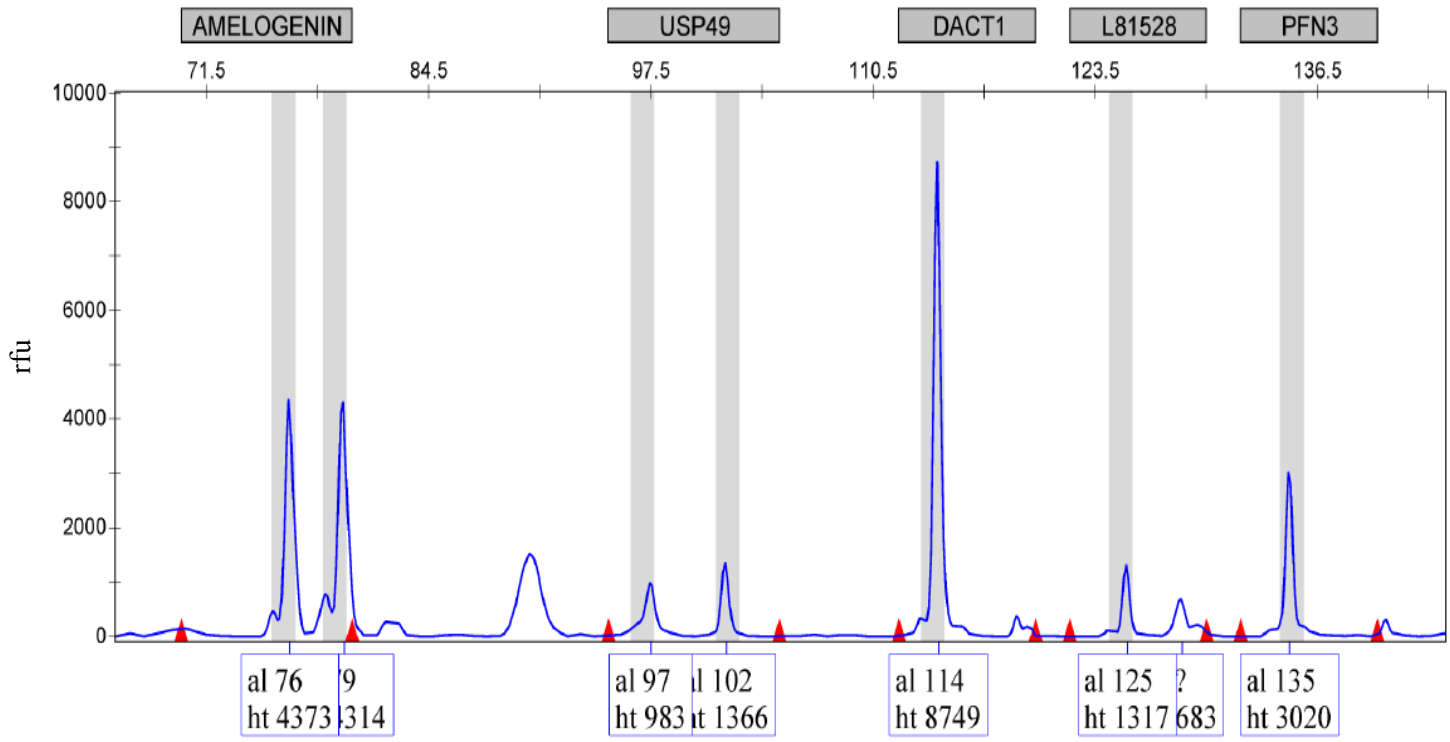

Figure 26: Methylation Profile of Saliva from White Male (47).

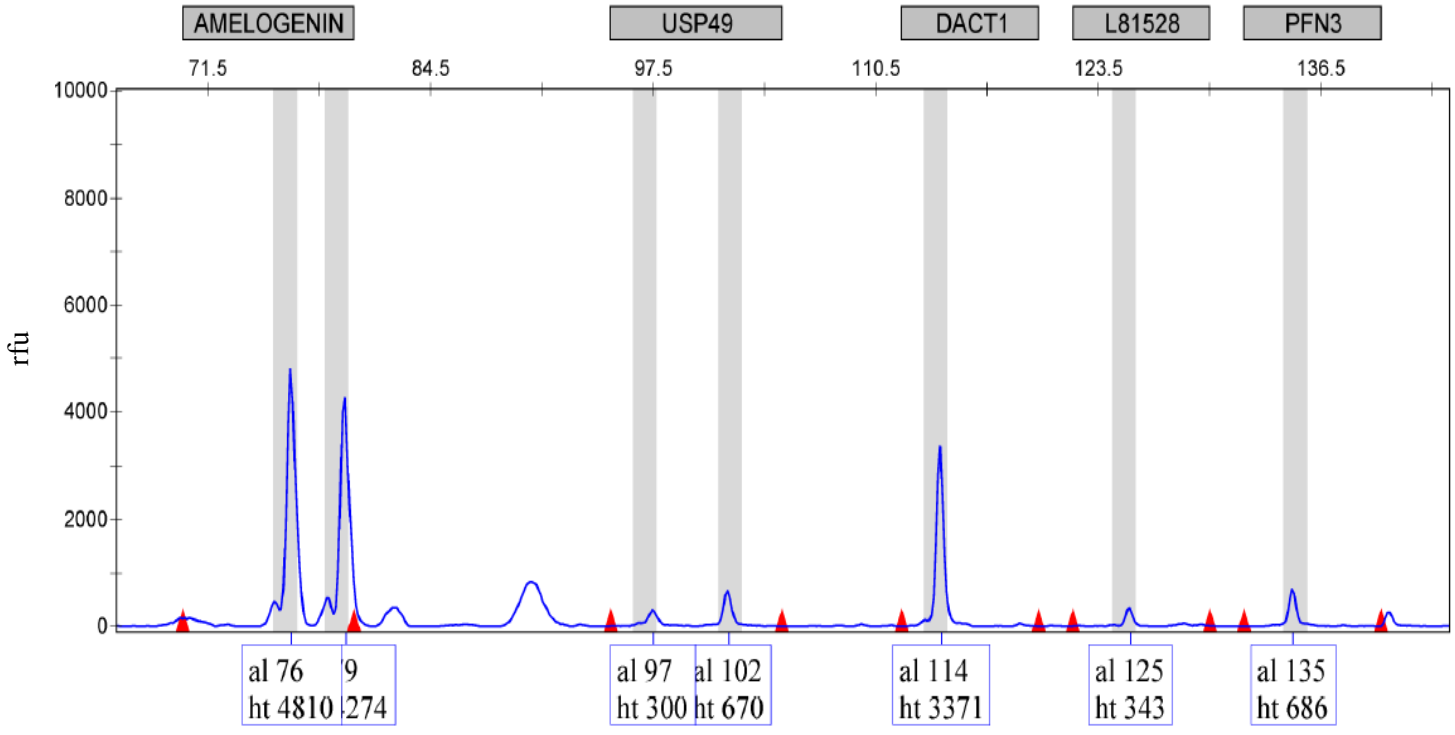

Figure 27: Methylation Profile of Saliva from White Male (48).

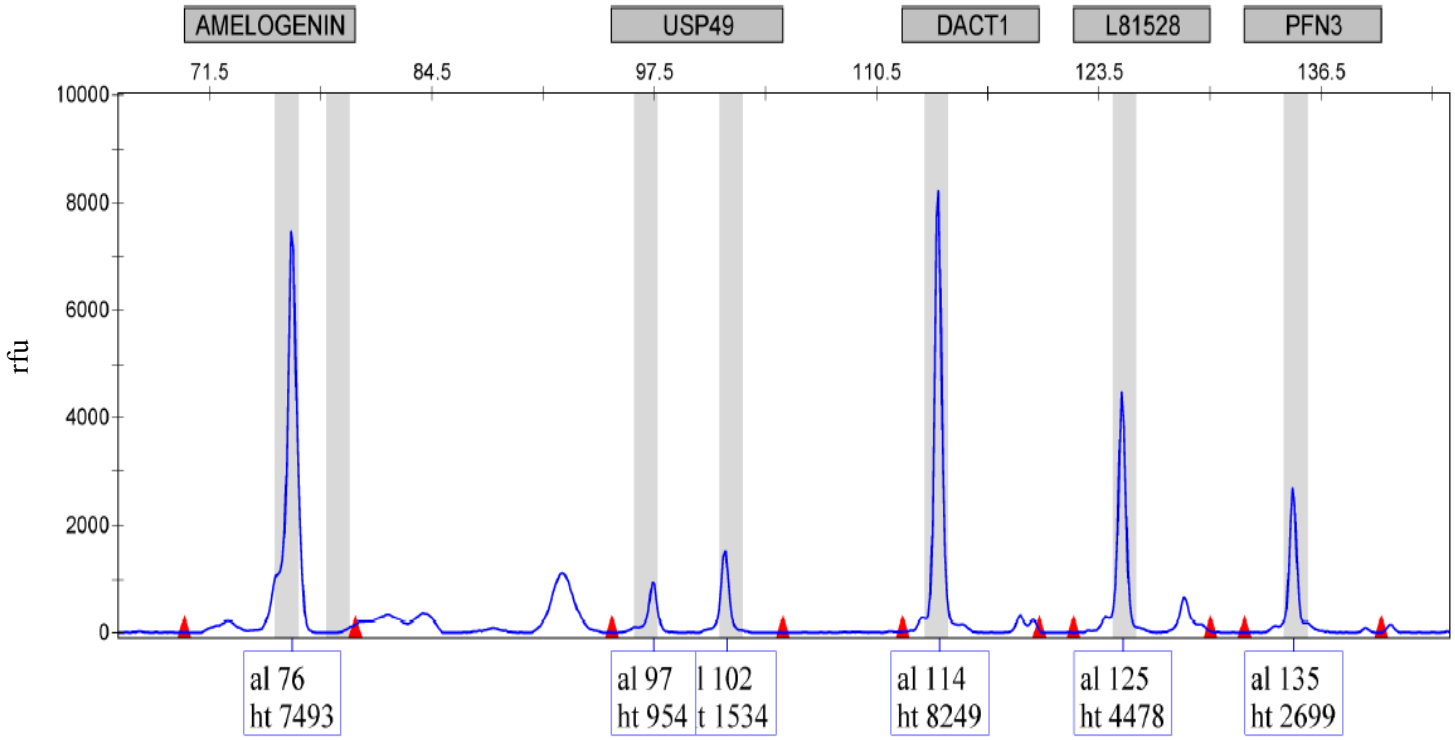

Figure 28: Methylation Profile of Saliva from White Male (49).

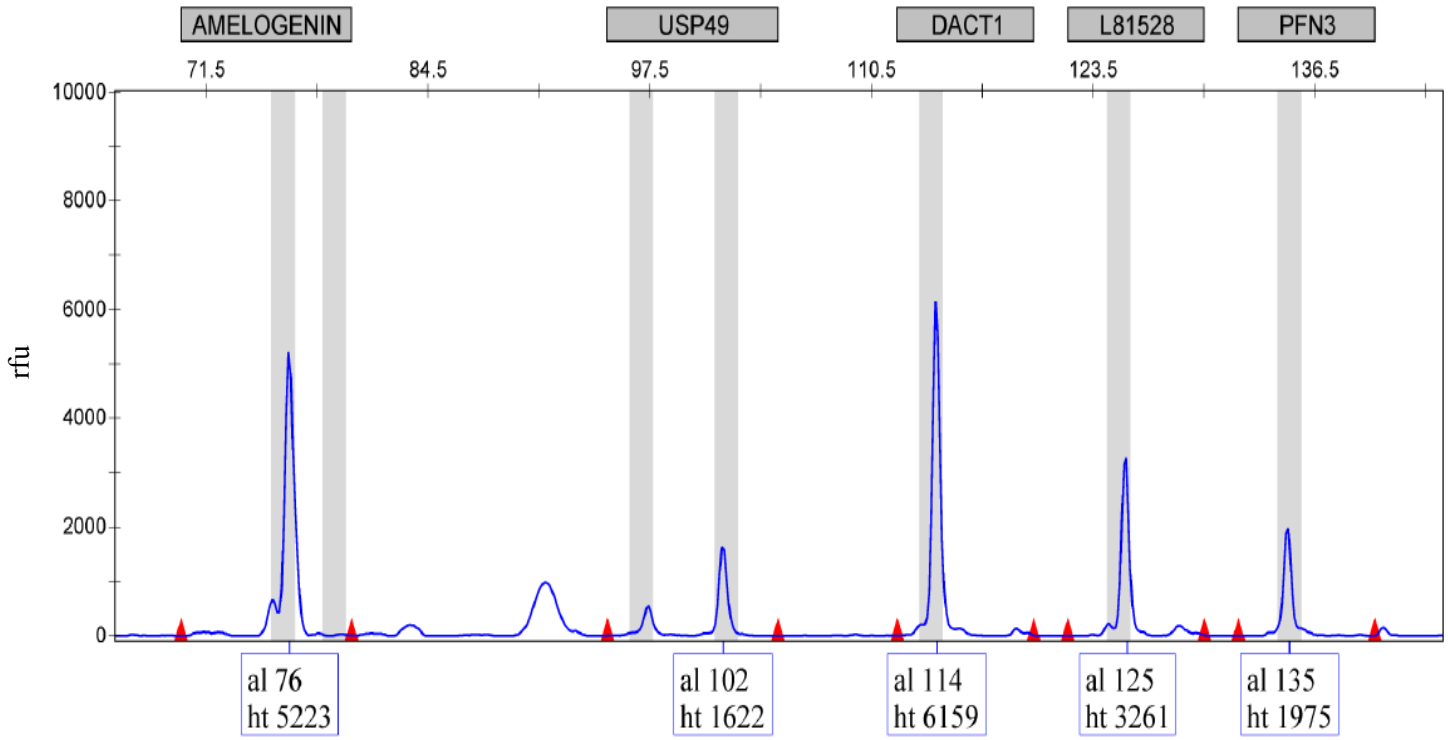

Figure 29: Methylation Profile of Saliva from White Female (53).

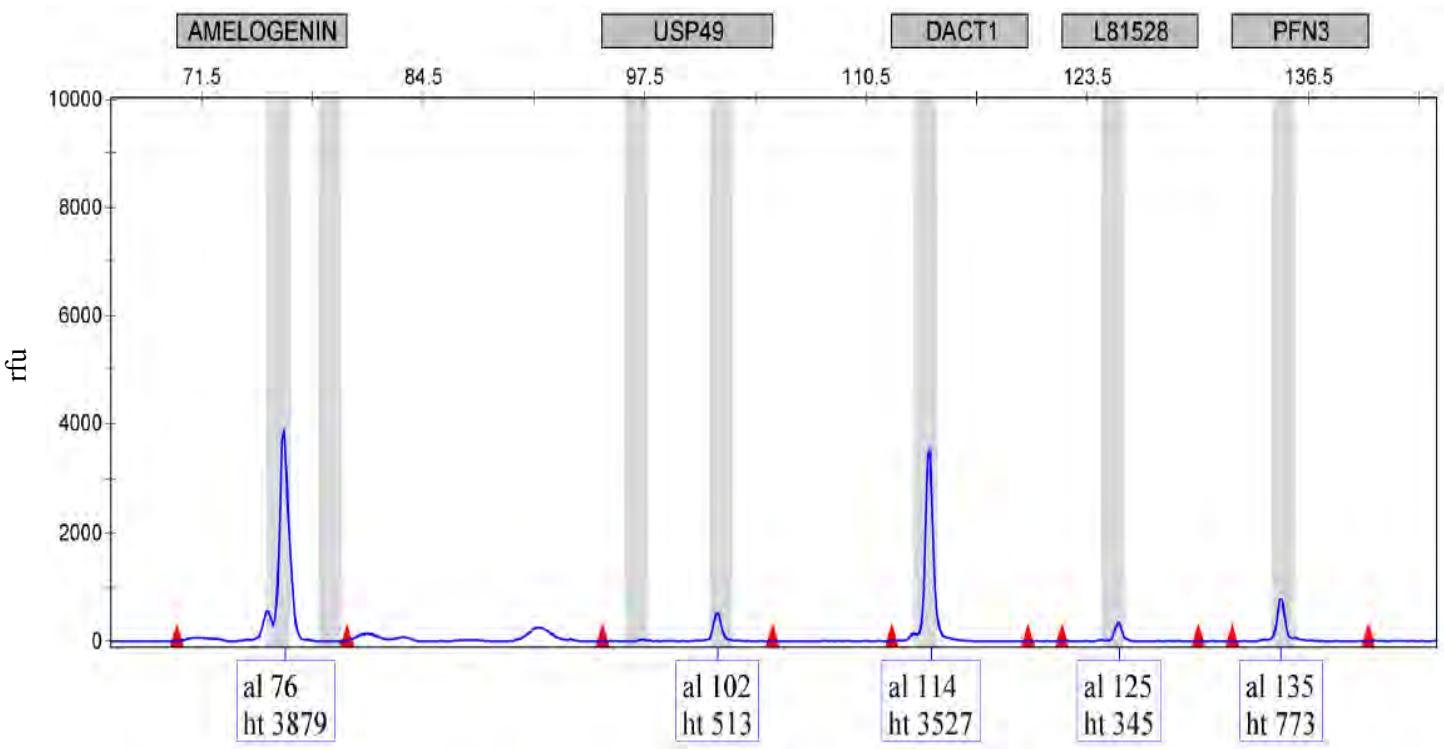

Figure 30: Methylation Profile of Saliva from White Female (56).

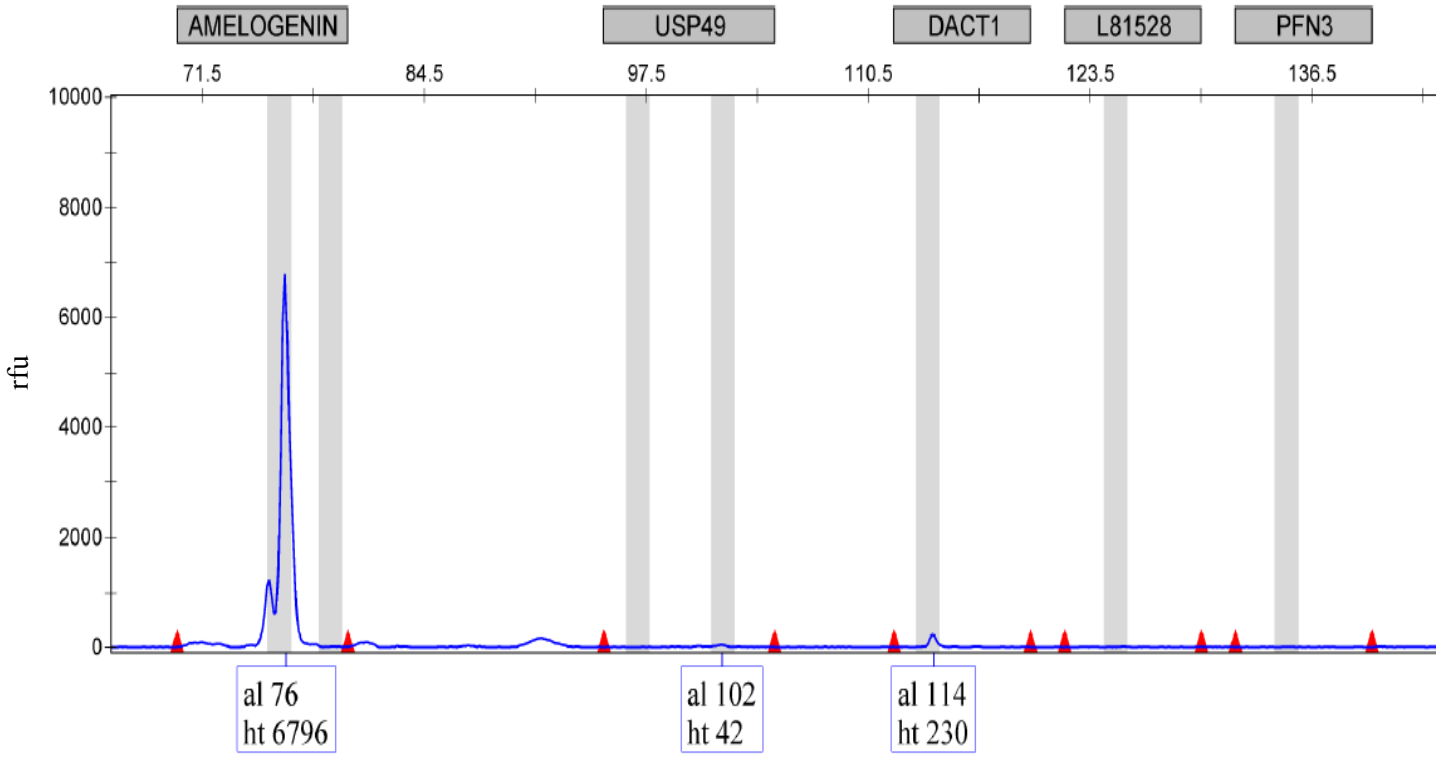

Figure 31: Methylation Profile of Saliva from White Female (57).

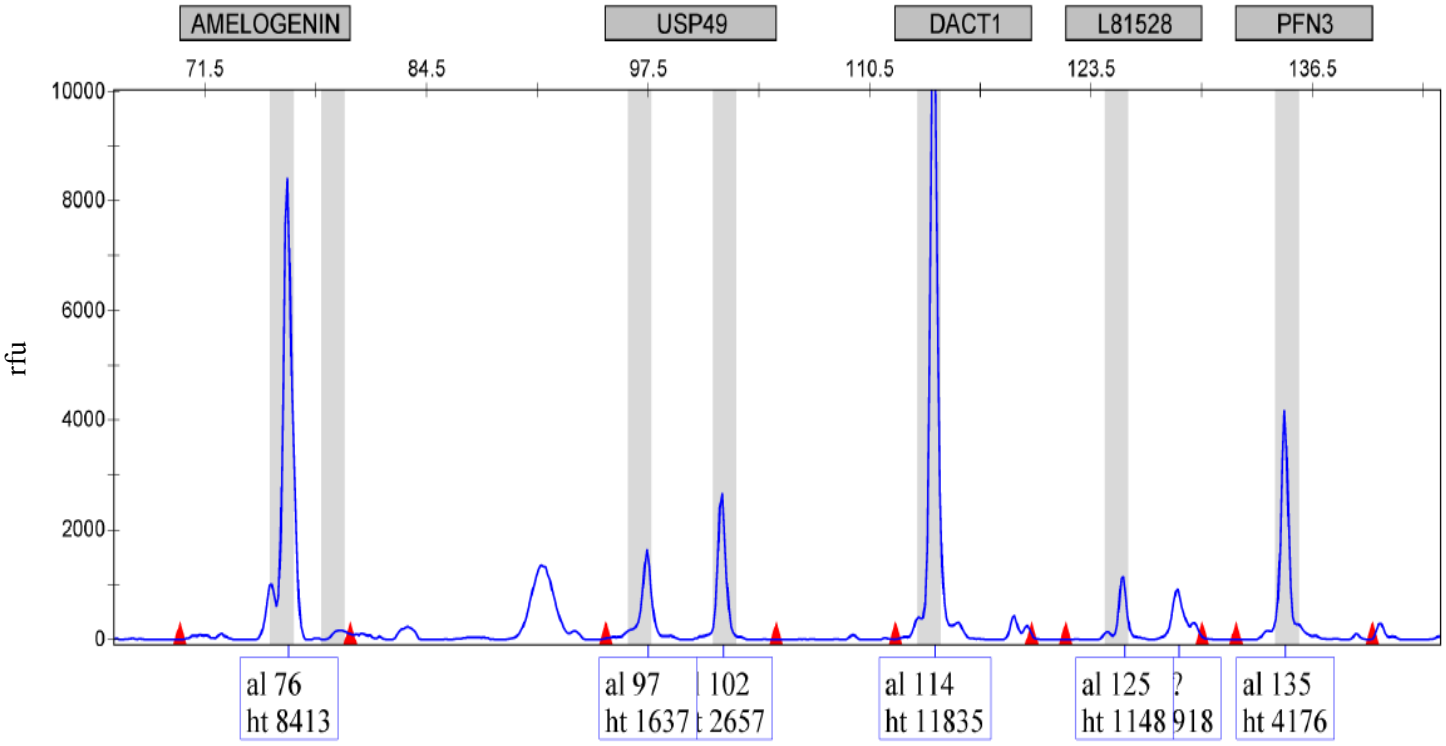

Figure 32: Methylation Profile of Saliva from White Female (58).

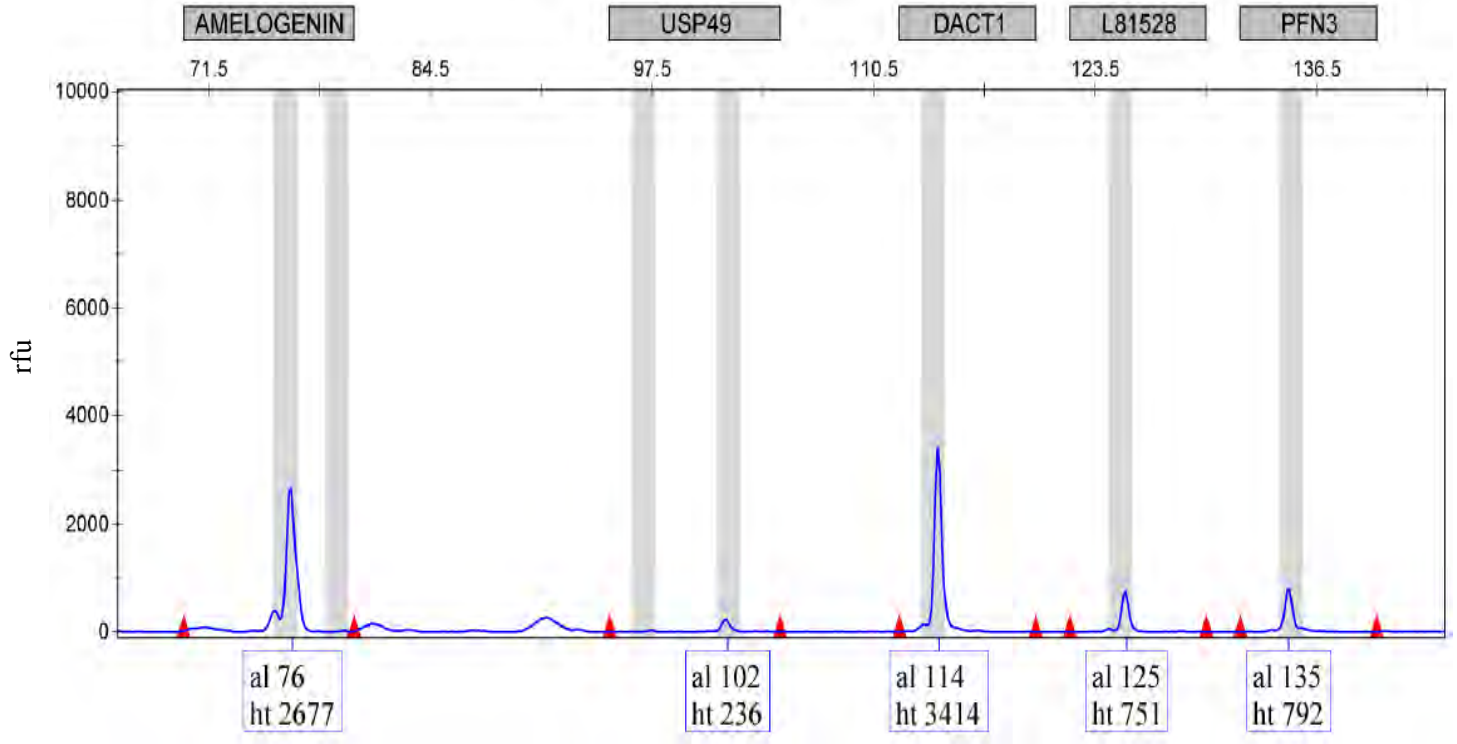

Figure 33: Methylation Profile of Saliva from White Female (59).

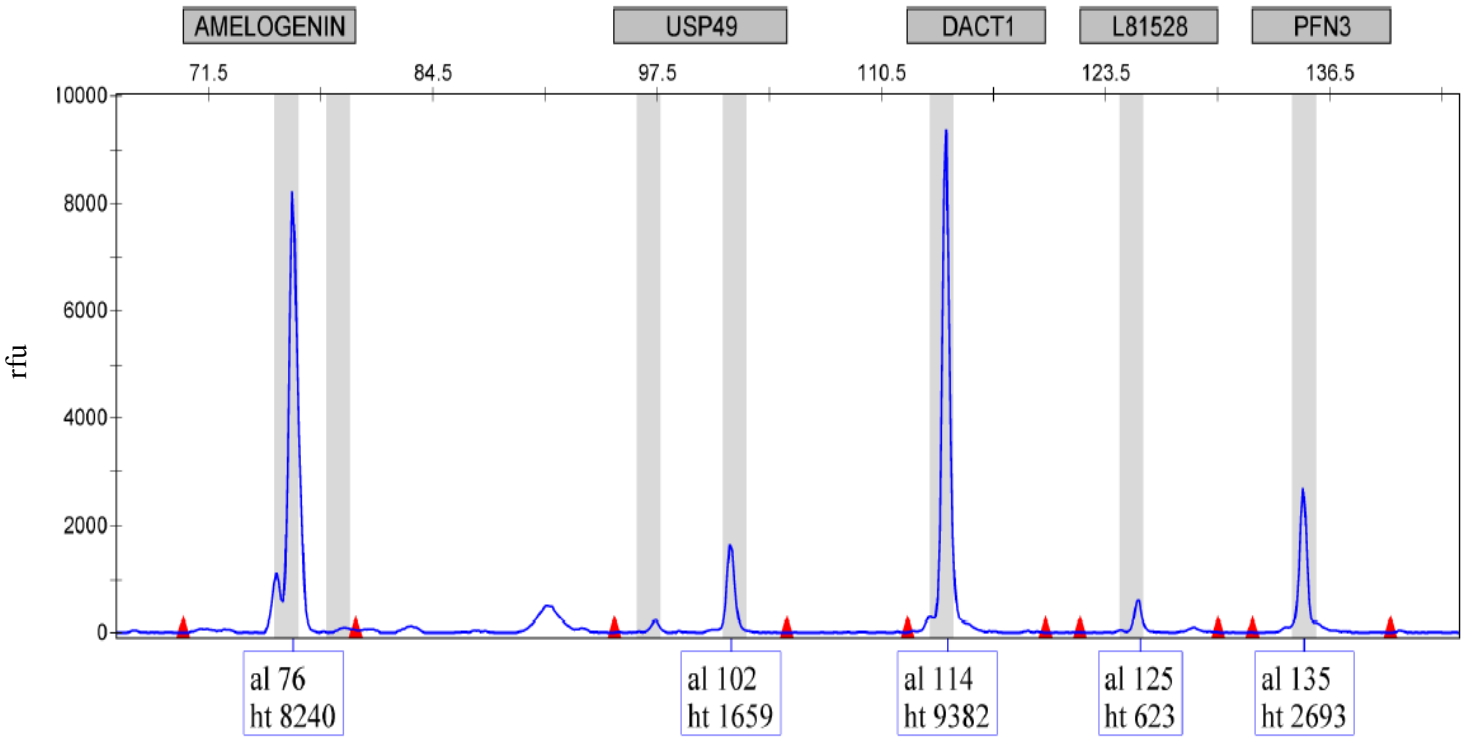

Figure 34: Methylation Profile of Saliva from White Female (60).

Methylation profiling of restricted saliva obtained from Coloureds as analysed by GeneMapper ID Software 5.

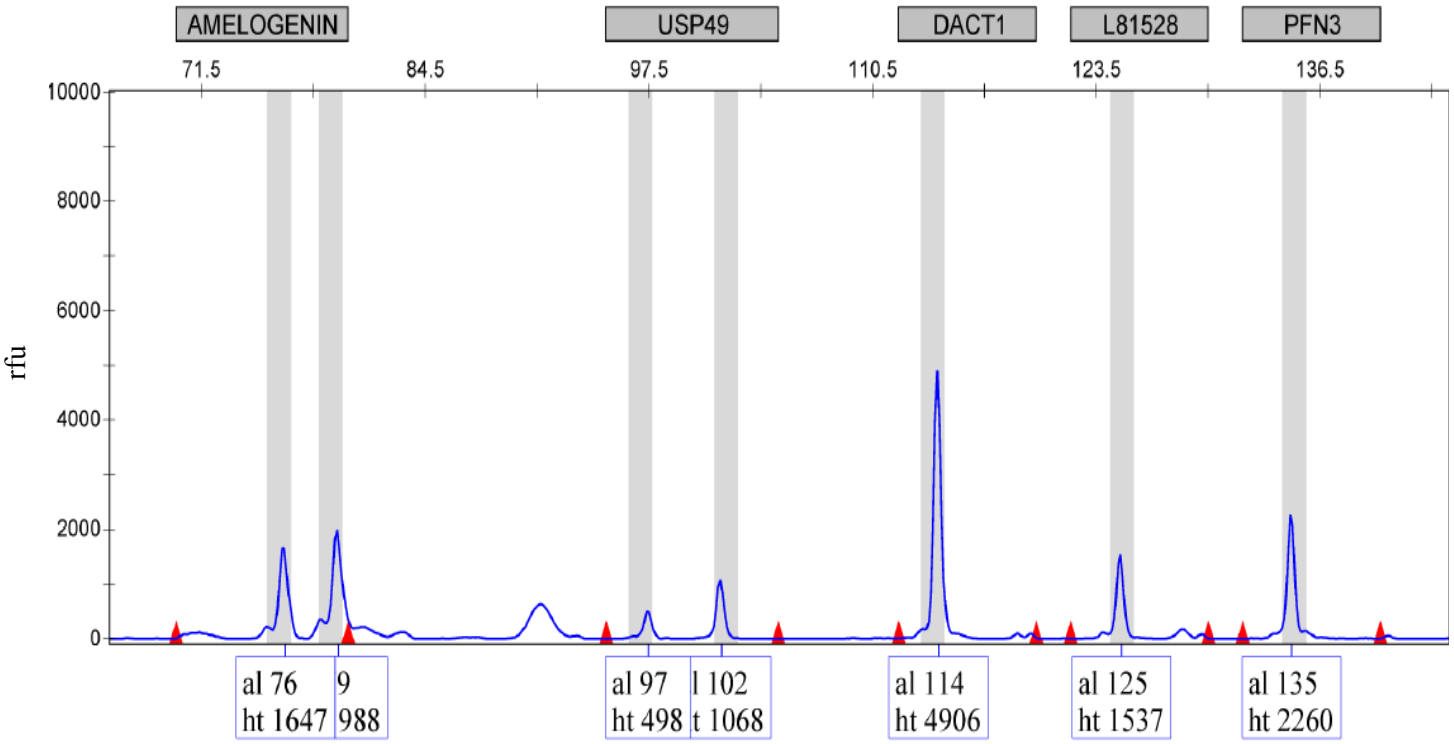

Figure 35: Methylation Profile of Saliva from Coloured Male (61).

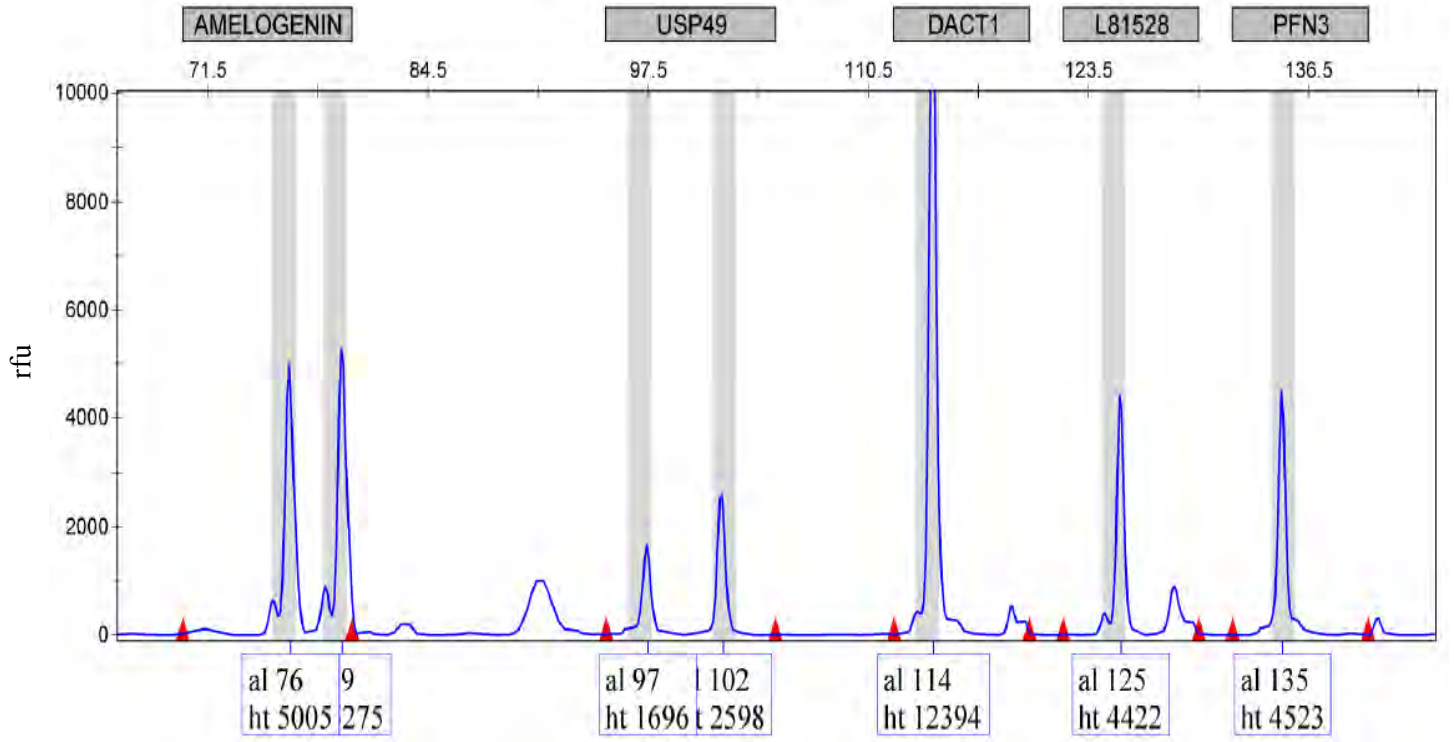

Figure 36: Methylation Profile of Saliva from Coloured Male (64).

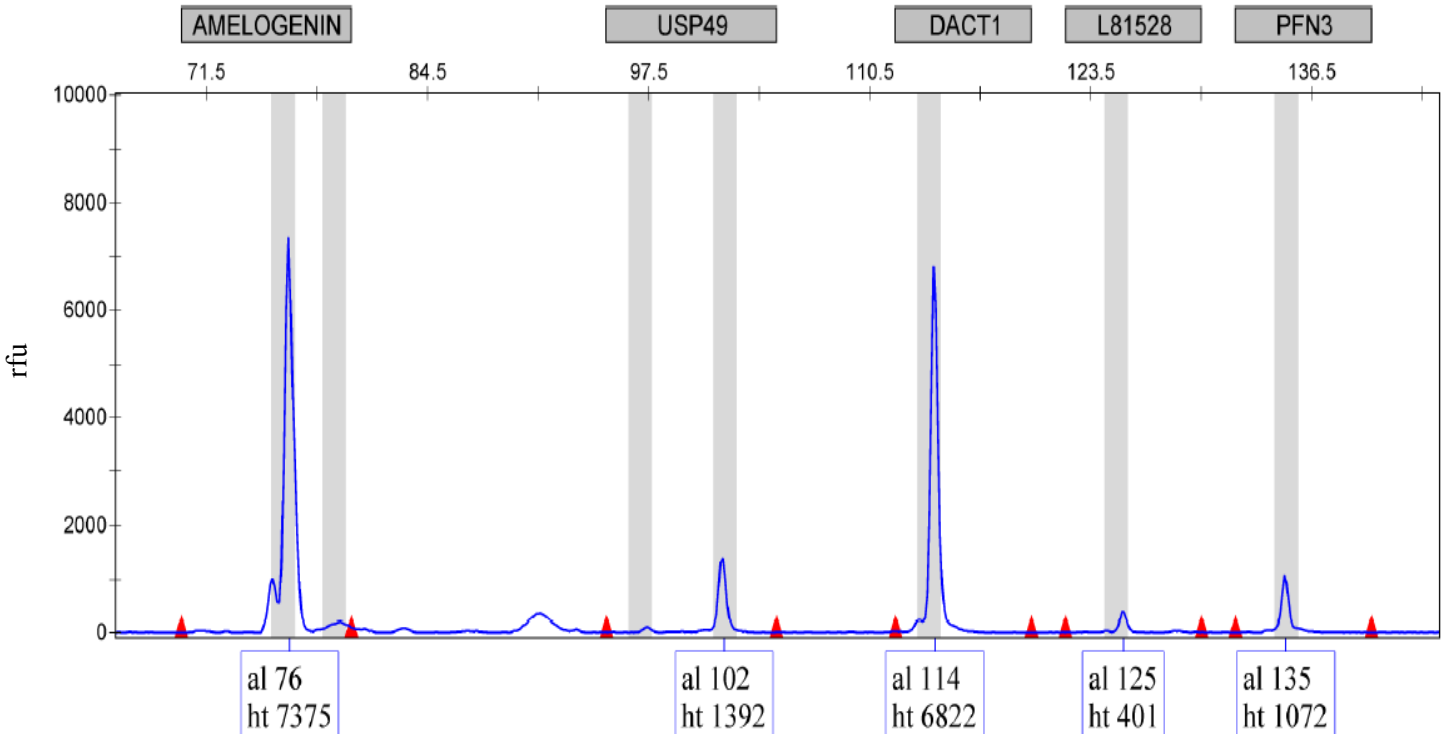

Figure 37: Methylation Profile of Saliva from Coloured Female (71).

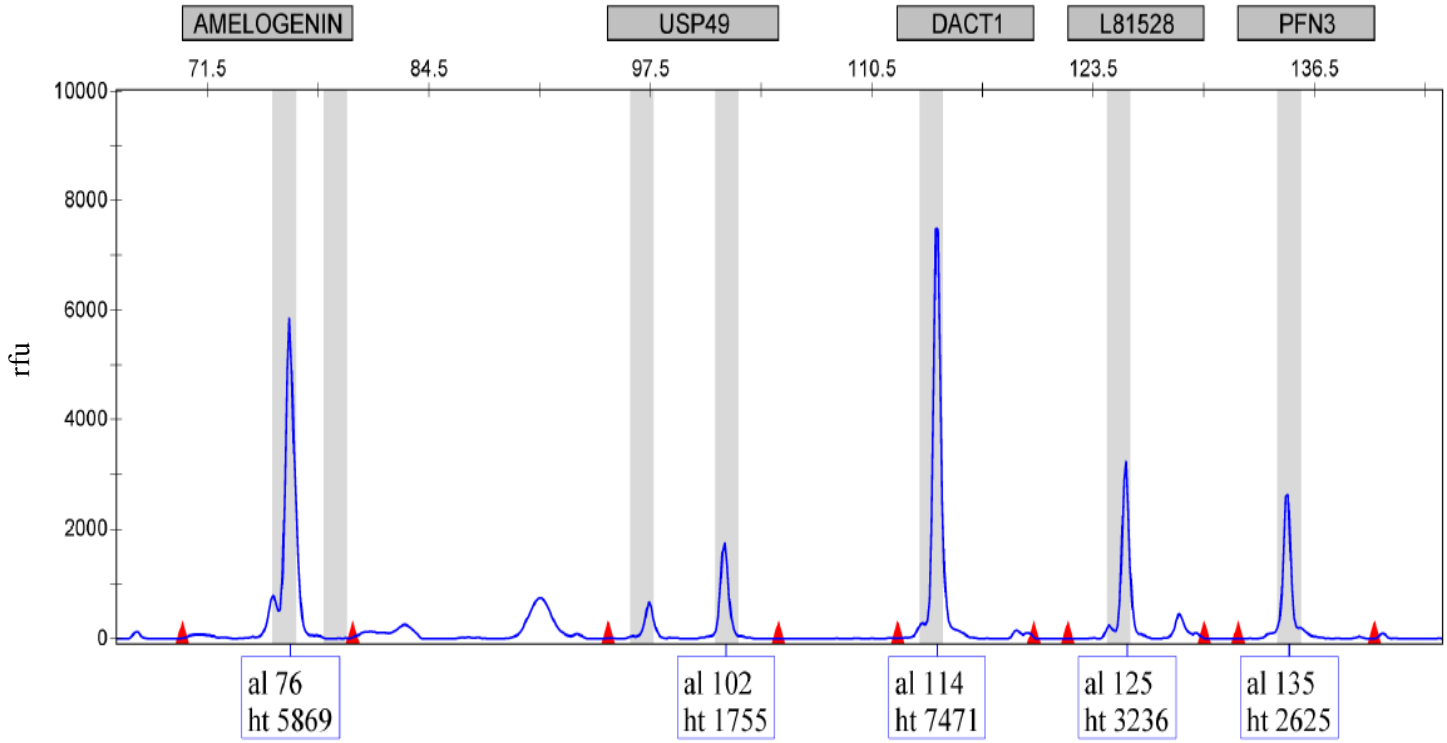

Figure 38: Methylation Profile of Saliva from Coloured Female (72).

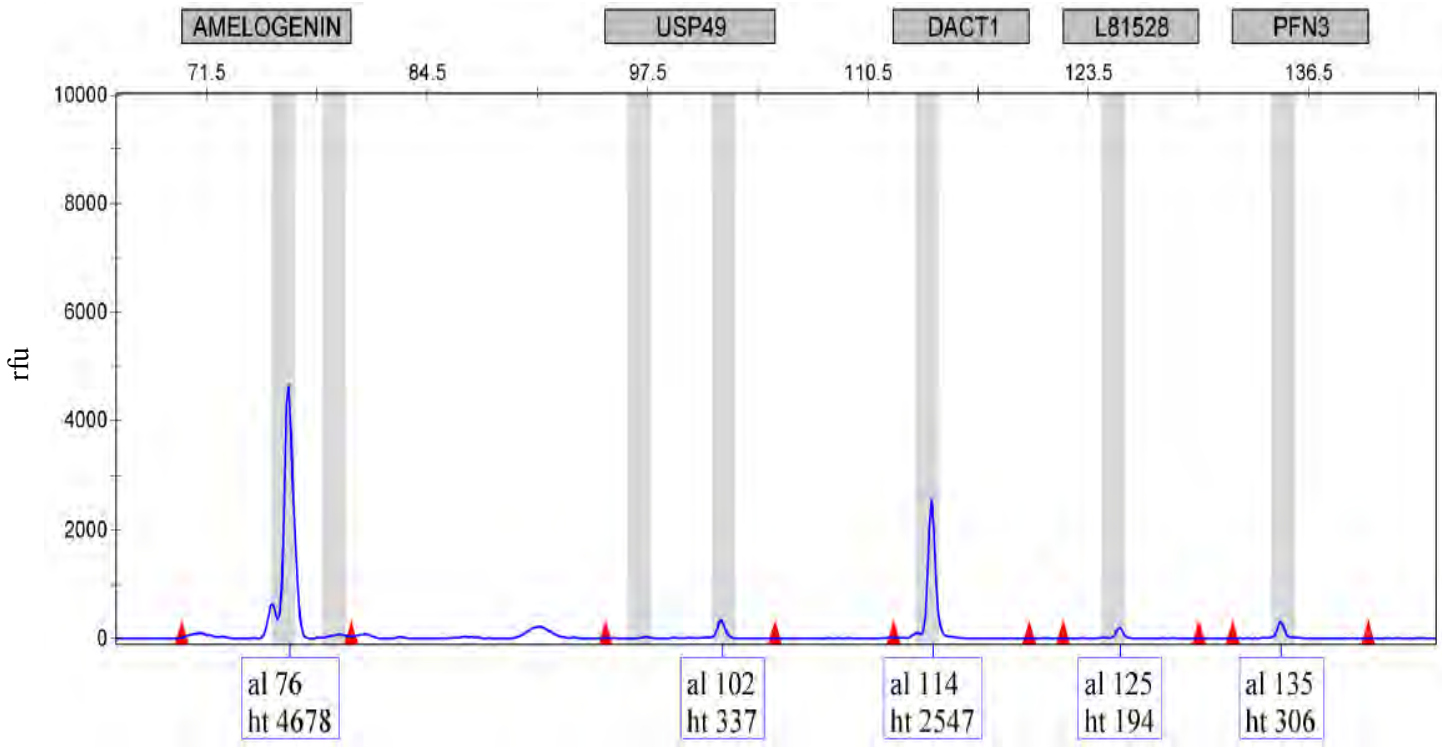

Figure 39: Methylation Profile of Saliva from Coloured Female (74).

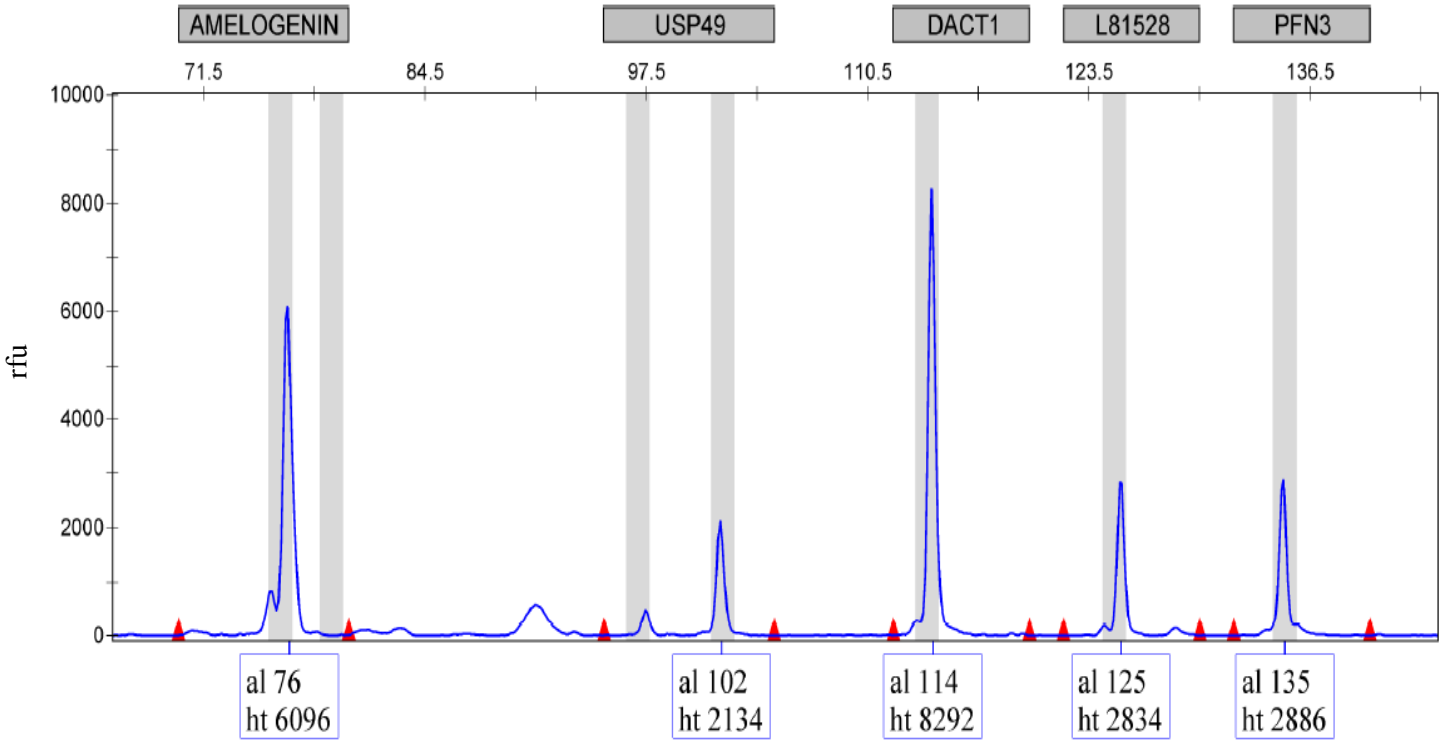

Figure 40: Methylation Profile of Saliva from Coloured Female (75).

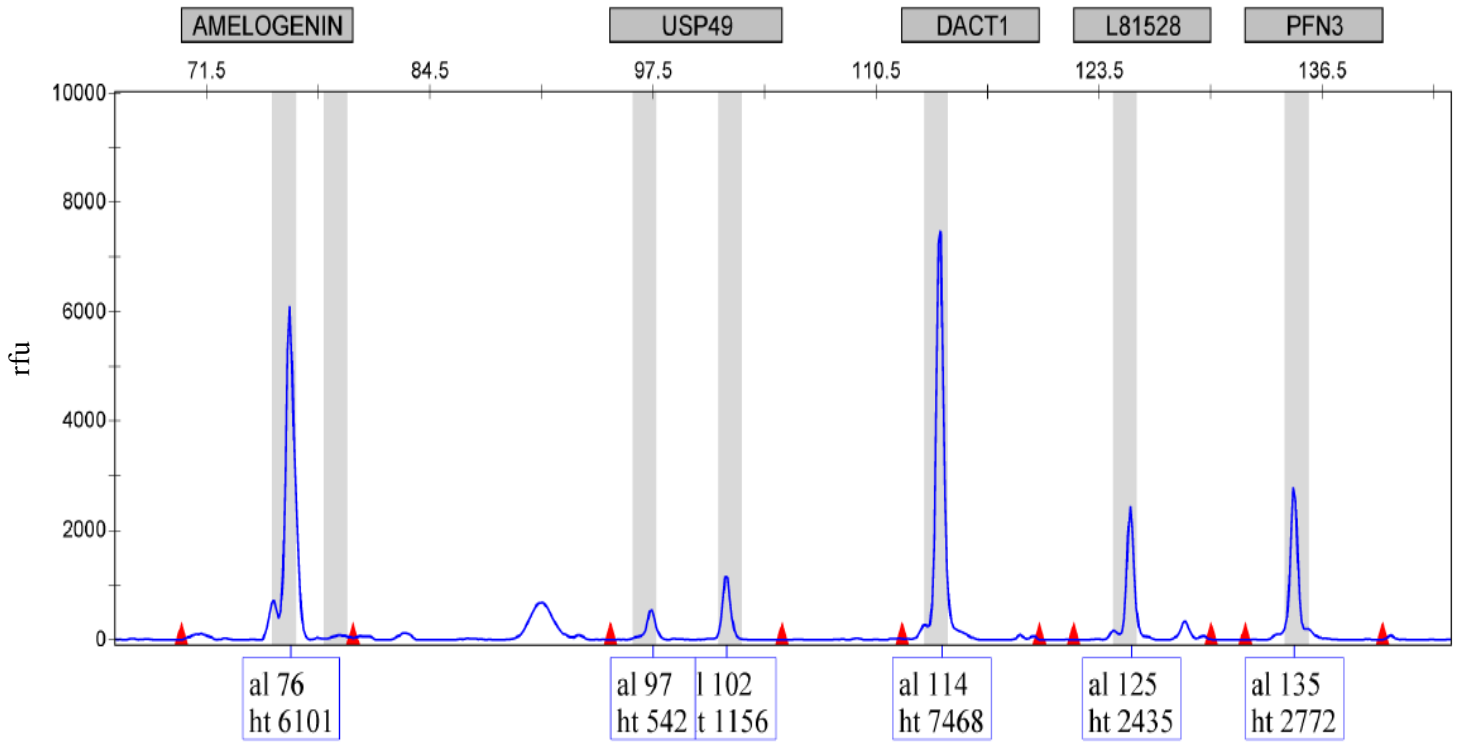

Figure 41: Methylation Profile of Saliva from Coloured Female (76).

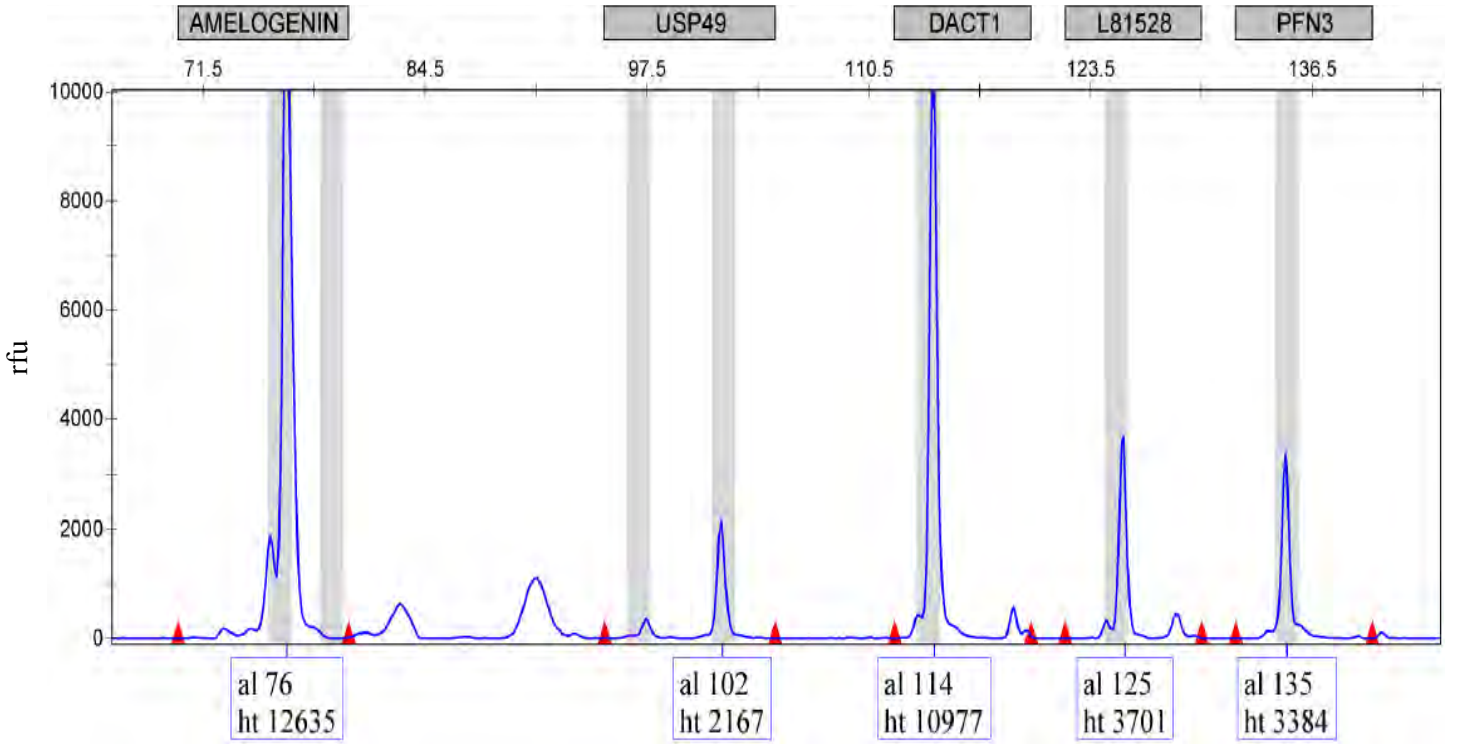

Figure 42: Methylation Profile of Saliva from Coloured Female (77).

Methylation profiles of unrestricted saliva (positive controls) obtained from all ethnicities as analysed by GeneMapper ID Software 5.

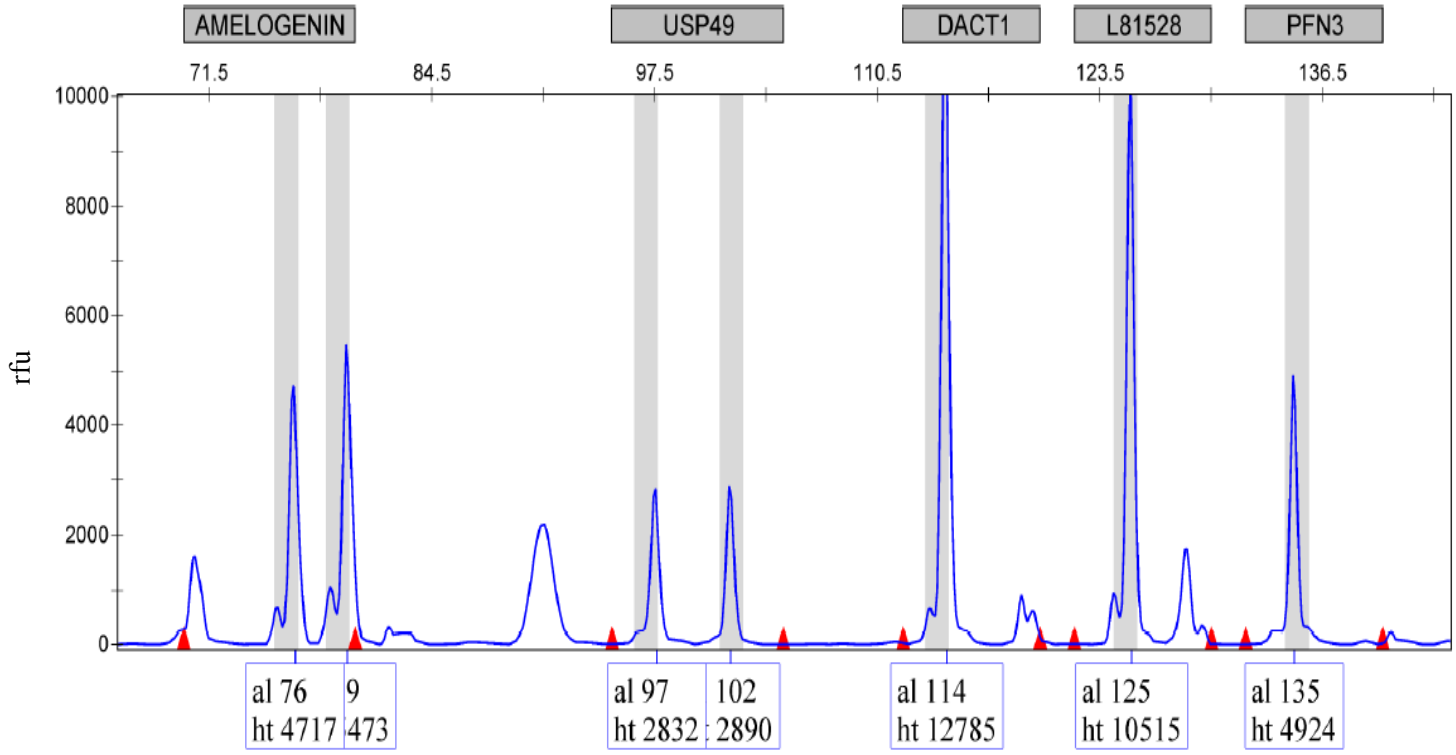

Figure 43: Unrestricted Saliva DNA of Black Male (4).

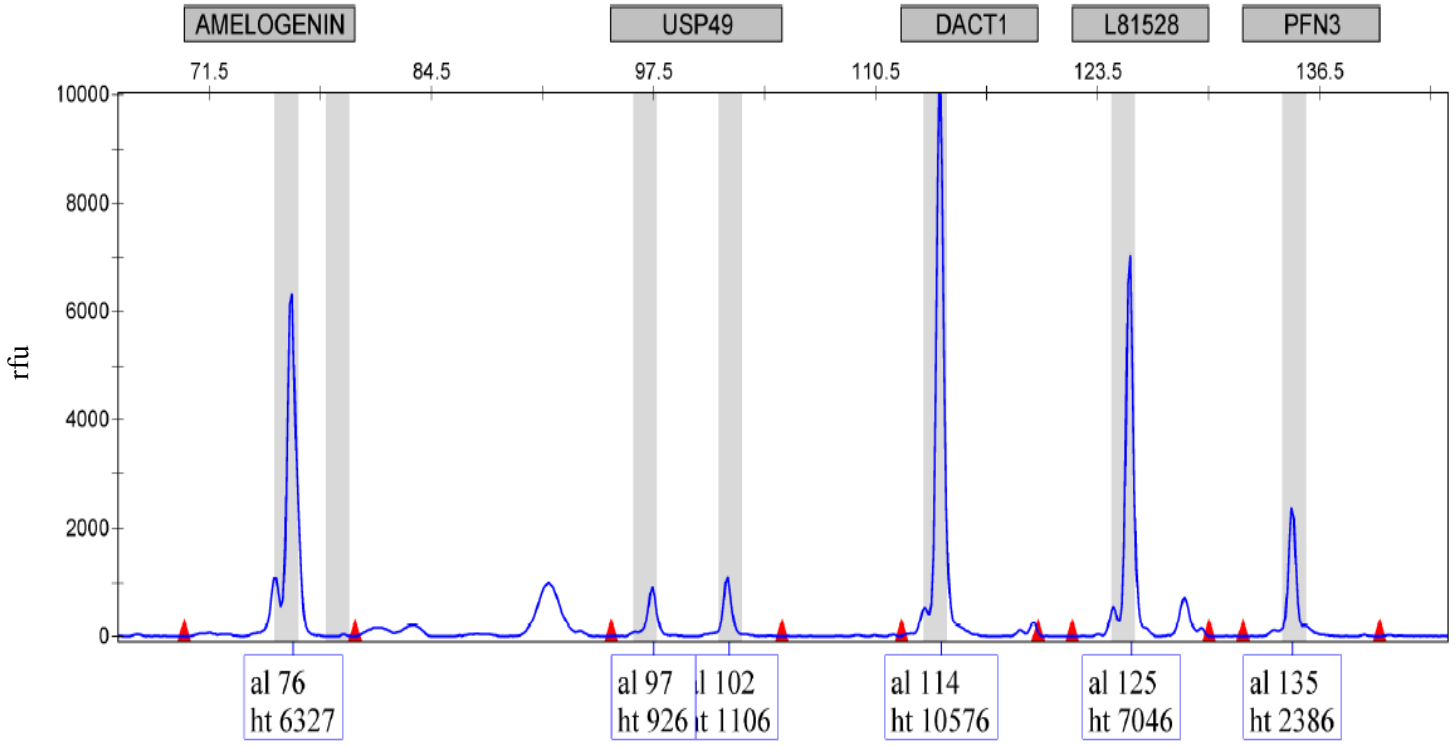

Figure 44: Unrestricted Saliva DNA of Coloured Female (77).

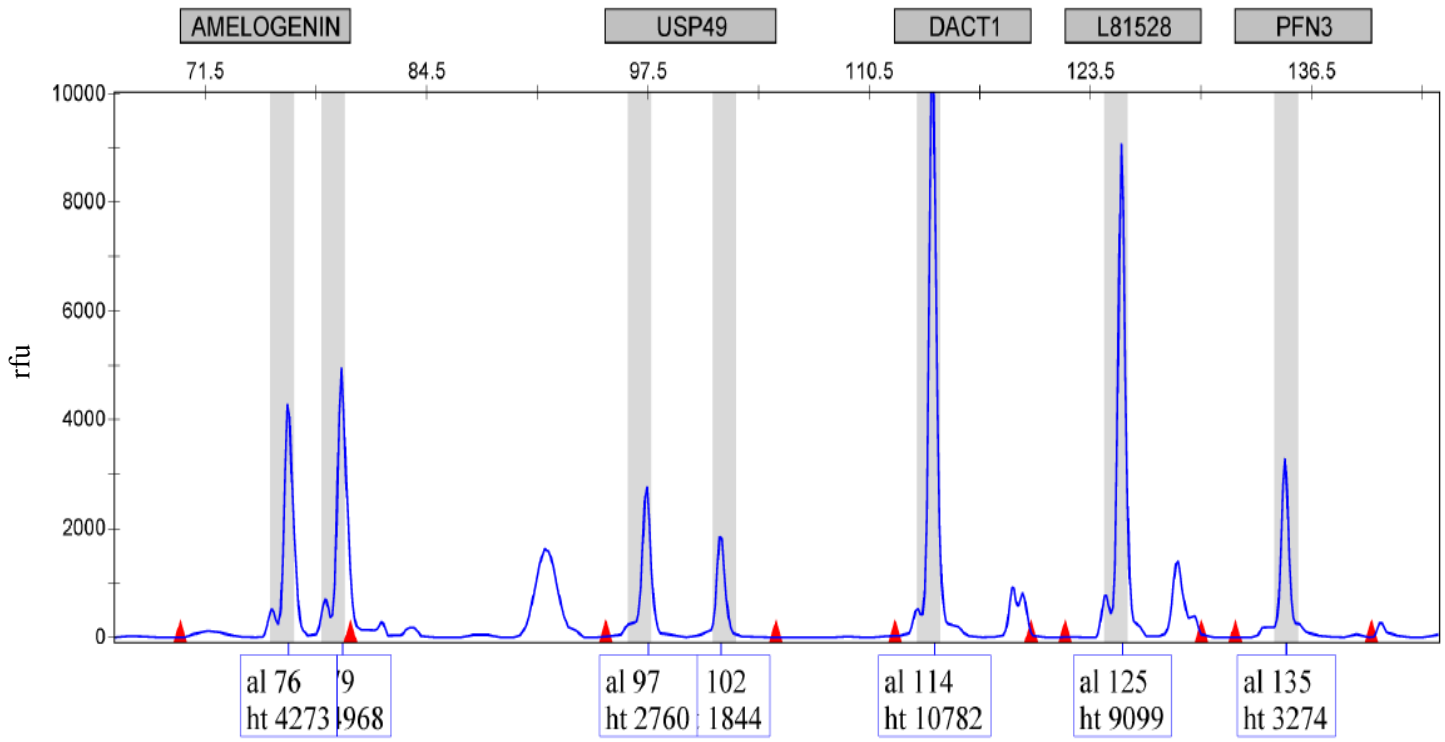

Figure 45: Unrestricted Saliva DNA of Indian Male (29).

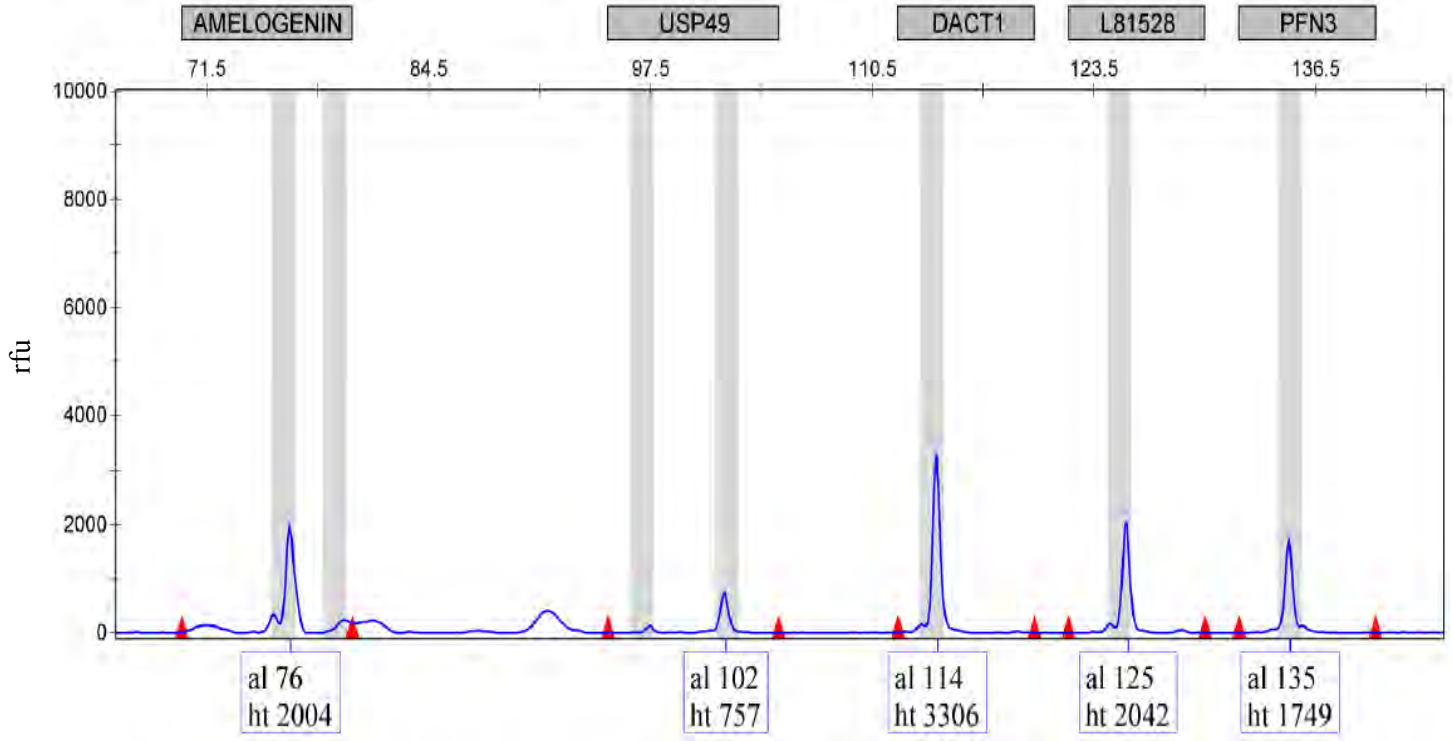

Figure 46: Unrestricted Saliva DNA of White Female (56).

Restricted Unmethylated Artificial DNA (pCR® 2.1 TOPO® vector) used as a positive control as analysed by GeneMapper ID Software 5.

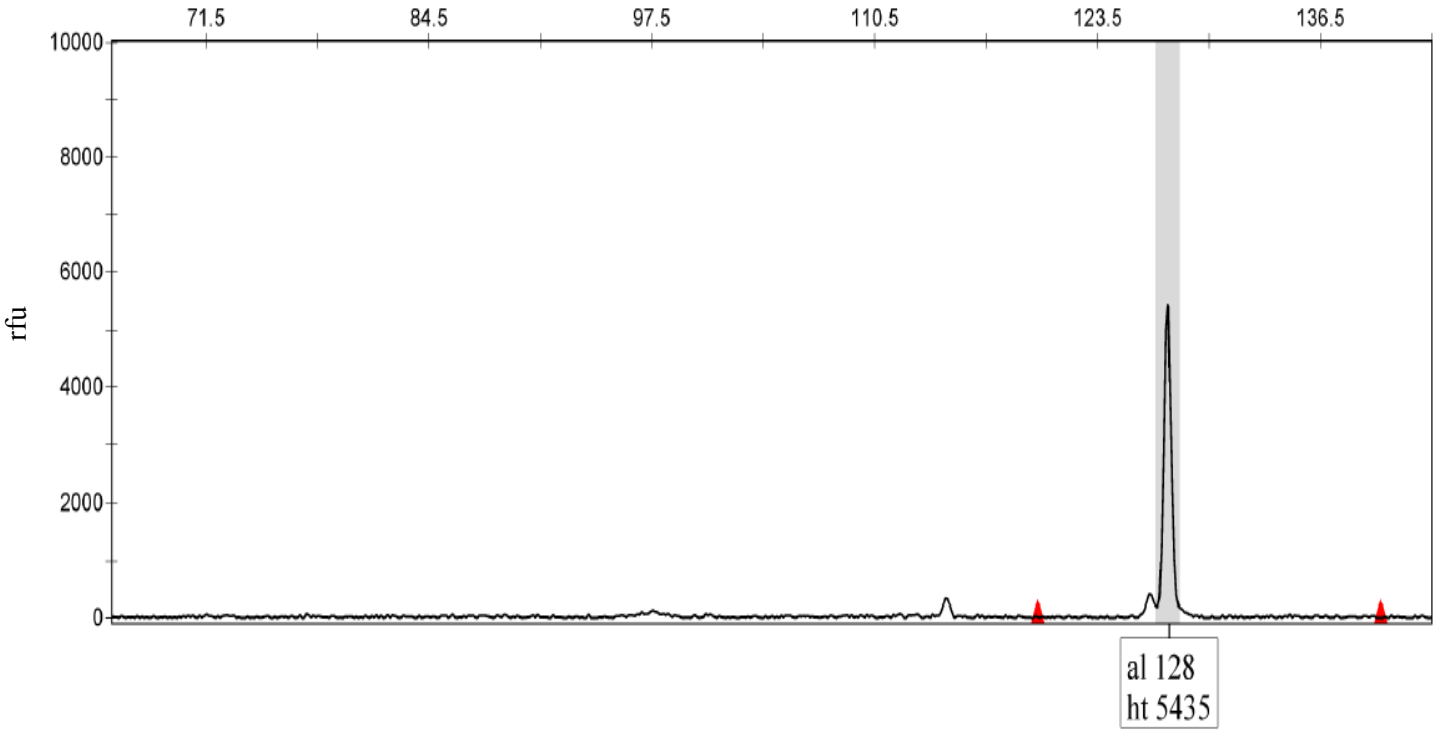

Figure 47: Methylation Profile of the Unmethylated Artificial DNA Control for proof of amplification success.

Statistical Analyses – Proof of normal distribution of the methylation data of each tDMR.

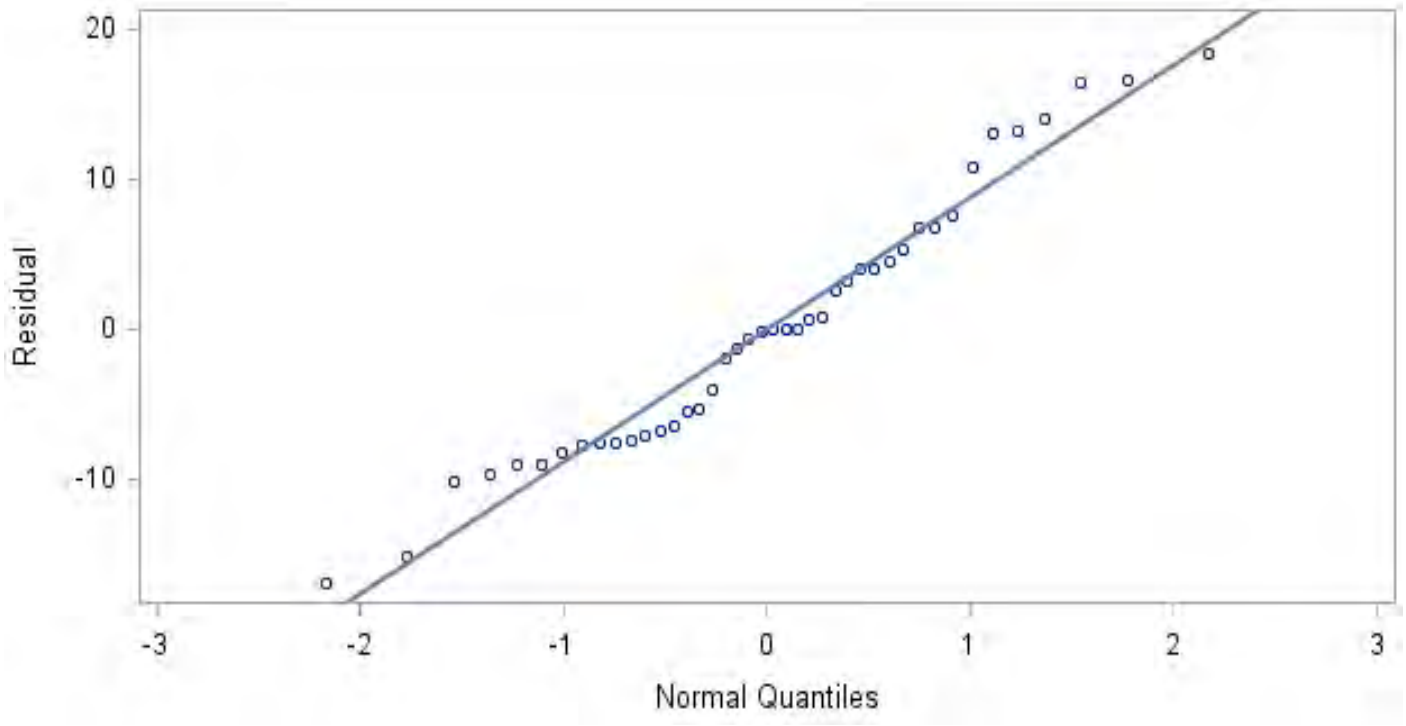

Figure 48: Graph showing normal distribution for the USP49 tDMR across all ethnic groups.

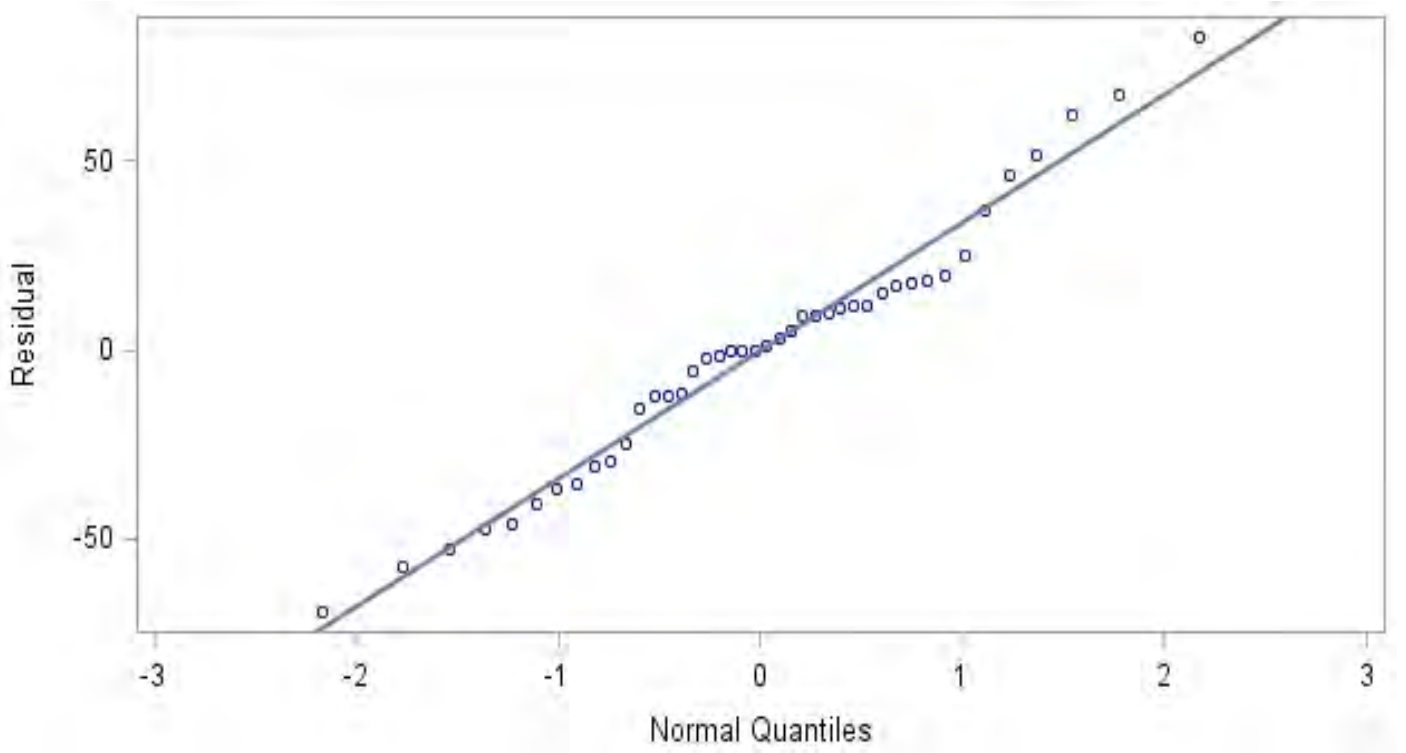

Figure 49: Graph showing normal distribution for the DACT1 tDMR across all ethnic groups.

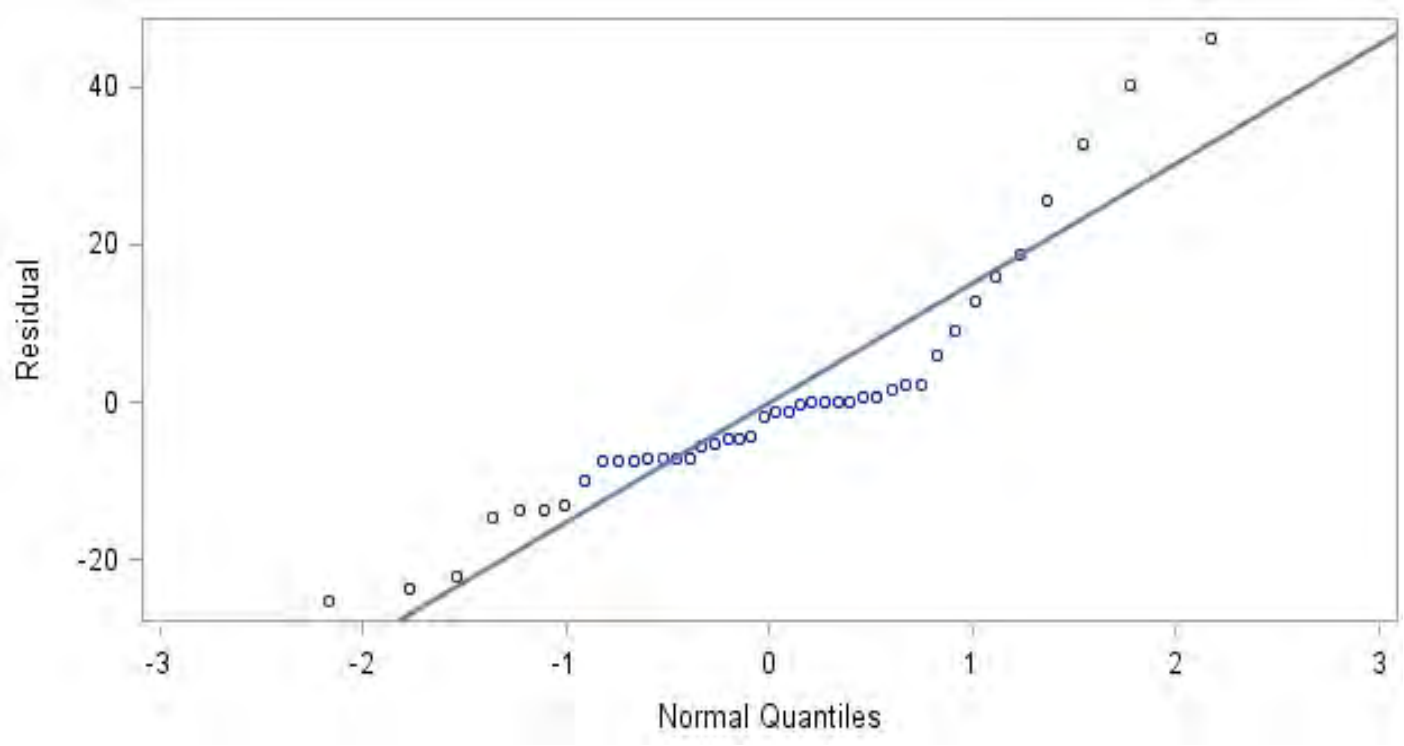

Figure 50: Graph showing normal distribution for the L81528 tDMR across all ethnic groups.

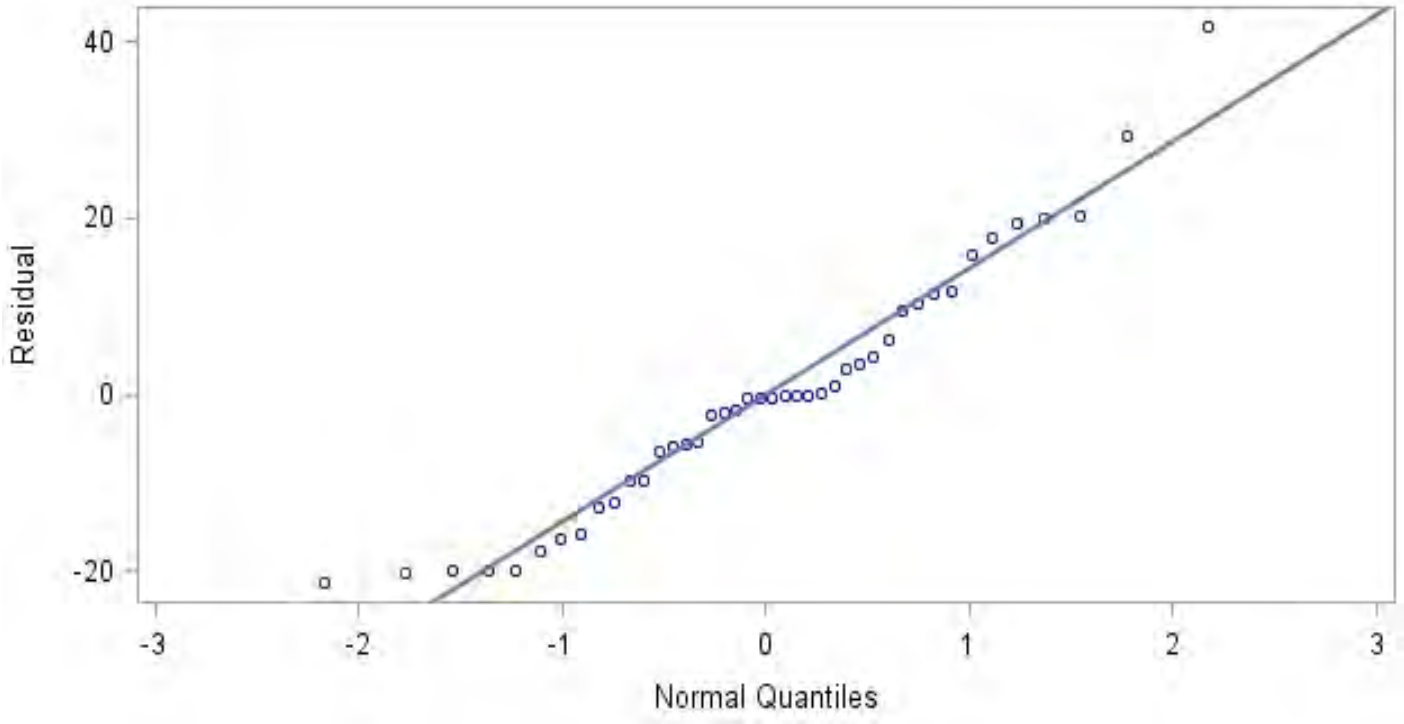

Figure 51: Graph showing normal distribution for the PFN3 tDMR across all ethnic groups.

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

25 March 2015

Miss Farzeen Kader
University Road
Westville
Private Bag X 54001
Durban
4000
farzeenkader68@gmail.com

PROTOCOL: Development and validation of DNA Methylation Specific Markers for Body Fluid and Tissue Identification in the South African population for Forensic Application.
REF: BE221/14

EXPEDITED APPLICATION

A sub-committee of the Biomedical Research Ethics Committee has considered and noted your application received on 10 April 2014.

The study was provisionally approved pending appropriate responses to queries raised. Your responses received on 12 March 2015 to queries raised on 10 March 2015 have been noted by a sub-committee of the Biomedical Research Ethics Committee. The conditions have now been met and the study is given full ethics approval.

This approval is valid for one year from **25 March 2015.** To ensure uninterrupted approval of this study beyond the approval expiry date, an application for recertification must be submitted to BREC on the appropriate BREC form 2-3 months before the expiry date.

Any amendments to this study, unless urgently required to ensure safety of participants, must be approved by BREC prior to implementation.

Your acceptance of this approval denotes your compliance with South African National Research Ethics Guidelines (2004), South African National Good Clinical Practice Guidelines (2006) (if applicable) and with UKZN BREC ethics requirements as contained in the UKZN BREC Terms of Reference and Standard Operating Procedures, all available at http://research.ukzn.ac.za/Research-Ethics/Biomedical-Research-Ethics.aspx.

BREC is registered with the South African National Health Research Ethics Council (REC-290408-009). BREC has US Office for Human Research Protections (OHRP) Federal-wide Assurance (FWA 678).

The sub-committee's decision will be **RATIFIED** by a full Committee at its meeting taking place on **14 April 2015.**

We wish you well with this study. We would appreciate receiving copies of all publications arising out of this study.

Yours sincerely

Professor J Tsoka-Gwegweni
Chair: Biomedical Research Ethics Committee