Development of methylation-specific PCR (MSP) and multiplex methylation SNaPshot assay for efficient identification of human body fluids

Sumina Haripersad

214548055

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: 214548055@stu.ukzn.ac.za

Master of Science - Genetics
Supervisor: Dr Meenu Ghai
(UKZN - Westville Campus)
Development of methylation-specific PCR (MSP) and multiplex methylation SNaPshot assay for efficient identification of human body fluids

BY

SUMINA HARIPERSAD

2020

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 February 2018 to January 2020, under the supervision of Dr. M. Ghai.

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: _______________________

I, ………………………………………………………………………………………….declare that

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

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

3. This thesis does not contain other persons’ data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written 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 and inside quotation marks and referenced.

5. This thesis 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: ……………………………………………………………..
# Table of Contents

Acknowledgments ....................................................................................................................... i

Abbreviations ............................................................................................................................. ii

Abstract .................................................................................................................................... viii

Organisation of dissertation ........................................................................................................ x

List of Figures ............................................................................................................................ xi

List of Tables ........................................................................................................................... xvi

CHAPTER 1: Introduction........................................................................................................1

CHAPTER 2: Literature Review .............................................................................................5

2.1 The phenomenon of epigenetics .........................................................................................6

2.2 DNA methylation ...............................................................................................................6

2.3 Association between DNA methylation and gene expression (transcription) .................8

2.4 Gene body methylation and gene expression: The DNA methylation Paradox .............10

2.5 Effect of environmental factors on DNA methylation .....................................................11

2.6 Tissue-specific differentially methylated regions (tDMRs) ...........................................12

2.8 Application of differential DNA methylation in forensics ...........................................14

2.8.1 Sex determination studies ..............................................................................................15

2.8.2 Age estimation ..............................................................................................................15

2.8.3 Distinguishing between artificially synthesised and naturally occurring DNA .......16

2.8.4 Differentiation between monozygotic/dizygotic twins .................................................16

2.8.5 Human body fluid identification ..................................................................................17

2.9 Analysis of body fluid mixtures in forensics ..................................................................24

2.10 DNA methylation analysis techniques .........................................................................27

2.10.1 Affinity enrichment ....................................................................................................28

2.10.1.1 Methyl-DNA immunoprecipitation (MeDIP) ......................................................29

2.10.1.2 Methyl-CpG immunoprecipitation (MCIp) .........................................................29

2.10.2 Endonuclease digestion .............................................................................................31

2.10.2.1 Restriction landmark genomic scanning (RLGS) ...............................................31

2.10.2.2 HpaII tiny fragment enrichment by ligation-mediated PCR (HELP) ...............32

2.10.3 Bisulfite conversion ....................................................................................................32

2.10.3.1 Methylation-specific polymerase chain reaction (MSP) .................................33
4.2 Materials and methods ................................................................. 71
  4.2.1 Sequencing of MSP products - bisulfite sequencing data .......... 71
  4.2.2 SNaPshot primer design .......................................................... 71
  4.2.3 Multiplex SNaPshot assay ....................................................... 73
  4.2.4 Mixture analysis ................................................................. 76

4.2.5 Data analyses ................................................................. 76

4.3 Results ................................................................................... 77

  4.3.1 Bisulfite sequencing .............................................................. 77
    4.3.1.1 HPCAL1 (saliva) MSP amplicon ........................................ 77
    4.3.1.2 cg09765089-231d (vaginal fluid) MSP amplicon ............... 78
    4.3.1.3 cg08792630 (blood) MSP amplicon ................................ 79
    4.3.1.4 ZNF282 (semen) MSP amplicon ...................................... 80

  4.3.2 Optimisation of multiplex SNaPshot assay .............................. 82

  4.3.3 Identification of body fluid identification by multiplex MSP-SNaPshot assay ........................................... 83

  4.3.4 Mixture analysis ................................................................. 86
    4.4.1.1 Identification of semen .................................................. 89
    4.4.1.2 Identification of vaginal fluid ......................................... 90
    4.4.1.3 Identification of blood .................................................. 91
    4.4.1.4 Identification of saliva .................................................. 92

  4.4.2 Mixture Analysis ................................................................. 92

4.5 Conclusion ............................................................................. 94

CHAPTER 5: General Discussion and Conclusion ............................. 96
References .................................................................................. 102
APPENDIX A ............................................................................. 123
APPENDIX B ............................................................................. 129
APPENDIX C ............................................................................. 132
APPENDIX D ............................................................................. 136
APPENDIX E ............................................................................. 137
APPENDIX F ............................................................................. 139
APPENDIX G ............................................................................. 140
APPENDIX H ............................................................................. 141
Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisor, Dr. Meenu Ghai. Thank you for your continuous support, motivation and overall, academic assistance during the course of this study. Thank you for always being understanding and encouraging me during the most difficult times, throughout this study. I would like to thank Dr. Ghai for her invaluable critique and insight in all writing material, as it has significantly contributed to my improved writing skills. I will remain eternally grateful for all that you have done for me.

I would like to thank my co-supervisor, Dr. Abraham Niehaus, for his support and assistance in this study.

I would like to thank Dr. Oliver Zishiri for his support and motivation. Thank you for constantly encouraging me to pursue all future endeavours and reminding me to remain grounded even as I aim for the stars.

I am extremely thankful to Ms. Farzeen Kader for her everlasting support, encouragement, and assistance whenever it was required. I consider myself extremely lucky to have had the opportunity of working alongside her, as it has been a valuable learning experience.

Thank you to all the volunteers who participated in the present study.

Finally, I would like to thank my mother Amy and sister Aarti for their motivation, understanding, patience and immense support. Thank you for helping me in whatever way you possibly could and standing by me during the bad days. I would like to especially thank my sister, Aarti, for her unprecedented critique and input during the write up of this study.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>percent</td>
</tr>
<tr>
<td>~</td>
<td>approximately</td>
</tr>
<tr>
<td>±</td>
<td>plus/minus</td>
</tr>
<tr>
<td>≤</td>
<td>less than or equal to</td>
</tr>
<tr>
<td>≥</td>
<td>greater than or equal to</td>
</tr>
<tr>
<td>&lt;</td>
<td>less than</td>
</tr>
<tr>
<td>&gt;</td>
<td>greater than</td>
</tr>
<tr>
<td>µL</td>
<td>microliters</td>
</tr>
<tr>
<td>µM</td>
<td>micromolar</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>1°</td>
<td>primary</td>
</tr>
<tr>
<td>2°</td>
<td>secondary</td>
</tr>
<tr>
<td>3’</td>
<td>3 prime</td>
</tr>
<tr>
<td>5’</td>
<td>5 prime</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methylcytosine/5-methylcytidine</td>
</tr>
<tr>
<td>A</td>
<td>adenine</td>
</tr>
<tr>
<td>Ab</td>
<td>Antibodies</td>
</tr>
<tr>
<td>ACCN4</td>
<td>amiloride-sensitive cation channel, neuronal 4</td>
</tr>
<tr>
<td>AIMS</td>
<td>amplification of inter-methylated sites</td>
</tr>
<tr>
<td>BCAS4</td>
<td>breast carcinoma amplified sequence 4</td>
</tr>
<tr>
<td>BL</td>
<td>Blood</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BREC</td>
<td>Biomedical Research Ethics Committee</td>
</tr>
<tr>
<td>BSPP</td>
<td>bisulphite padlock probes</td>
</tr>
<tr>
<td>C</td>
<td>cytosine</td>
</tr>
<tr>
<td>C/T</td>
<td>cytosine/thymine ratios</td>
</tr>
<tr>
<td>C20orf117</td>
<td>chromosome 20 open reading frame 117</td>
</tr>
<tr>
<td>CAGE</td>
<td>cap analysis gene expression</td>
</tr>
<tr>
<td>CD4⁺</td>
<td>cluster of differentiation 4</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>CD8+</td>
<td>cluster of differentiation 8</td>
</tr>
<tr>
<td>CDM</td>
<td>cell type-specific differentially methylated gene region</td>
</tr>
<tr>
<td>CGI</td>
<td>CpG islands</td>
</tr>
<tr>
<td>-CH₃</td>
<td>methyl group</td>
</tr>
<tr>
<td>CHARM</td>
<td>comprehensive high-throughput arrays for relative methylation</td>
</tr>
<tr>
<td>Chr</td>
<td>Chromosome</td>
</tr>
<tr>
<td>COBRA</td>
<td>combined bisulphite restriction analysis</td>
</tr>
<tr>
<td>CoRep</td>
<td>co-repressors</td>
</tr>
<tr>
<td>CYTH4</td>
<td>cytohesin 4</td>
</tr>
<tr>
<td>d</td>
<td>downstream</td>
</tr>
<tr>
<td>Da</td>
<td>Daltons</td>
</tr>
<tr>
<td>DACTI</td>
<td>dishevelled binding antagonist of beta catenin 1</td>
</tr>
<tr>
<td>ddATP</td>
<td>2'-3'-dideoxyadenosine triphosphate</td>
</tr>
<tr>
<td>ddGTP</td>
<td>2'-3'-dideoxyguanosine triphosphate</td>
</tr>
<tr>
<td>ddNTP</td>
<td>dideoxynucleotide triphosphate</td>
</tr>
<tr>
<td>DMH</td>
<td>differential methylation hybridisation</td>
</tr>
<tr>
<td>DMRTA2</td>
<td>DMRT-like family A2</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNMT</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>DNTP</td>
<td>deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>e.g.</td>
<td>exempli gratia (Example)</td>
</tr>
<tr>
<td>EDARADD</td>
<td>EDAR associated death domain</td>
</tr>
<tr>
<td>EFS</td>
<td>embryonal Fyn-associated substrate</td>
</tr>
<tr>
<td>ESCCC</td>
<td>oesophageal squamous cell carcinoma</td>
</tr>
<tr>
<td>et al.</td>
<td>et alia (and others)</td>
</tr>
<tr>
<td>etc.</td>
<td>et cetera (and the rest)</td>
</tr>
<tr>
<td>Exo1</td>
<td>exonuclease 1</td>
</tr>
<tr>
<td>FASTA</td>
<td>Fast-All</td>
</tr>
<tr>
<td>FGF7</td>
<td>fibroblast growth factor 7</td>
</tr>
<tr>
<td>FNDC1</td>
<td>fibronectin type III domain containing 1</td>
</tr>
<tr>
<td>FOXO3</td>
<td>forkhead box O3</td>
</tr>
</tbody>
</table>
G  guanine
GAS2LI  growth arrest specific 2 like 1
GBM  gene-body methylation
GR  glucocorticoid receptor
GRCh37  Genome Reference Consortium Human Build 37
GRCh38  Genome Reference Consortium Human Build 38
GTEx  genotype-tissue expression consortium
H19  H19 imprinted maternally expressed transcript
HDAC  histone deacetylases
HELP  HpaII tiny fragment enrichment by ligation-mediated polymerase chain reaction
HLP2  Hippocalcin-like protein 1
HOTAIR  Homeobox transcript antisense RNA
HOXA4  homeobox A4
HPCAL1  Hippocalcin-like 1
HPCE  high performance capillary electrophoresis
HTLV-1  human T-cell leukemia virus type 1
*i.e.*  id est (in other words)
IAP  intracisternal A particle
IGF2  insulin-like growth factor 2
IgG1  immunoglobulin G1
IP  immunoprecipitation
Line1  long interspersed repeat sequence 1
LTR  long terminal repeat
MAF  minor allele frequency
MBD  methyl-CpG binding domain
MBD2  methyl-CpG-binding domain protein 2
MBPs  methyl-CpG-binding proteins
mCIP/MeCIP  methyl-CpG immunoprecipitation
mDIP/MeDIP  methyl-DNA immunoprecipitation
MeCP2  methyl-CpG-binding proteins 2
MF  methylated forward primer
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIRA</td>
<td>methylated CpG island recovery assay</td>
</tr>
<tr>
<td>miRNA</td>
<td>micro ribonucleic acid</td>
</tr>
<tr>
<td>MIV</td>
<td>methylation indicative value</td>
</tr>
<tr>
<td>MMASS</td>
<td>microarray-based methylation assessment of single samples</td>
</tr>
<tr>
<td>MR</td>
<td>methylated reverse primer</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>MS-AP-PCR</td>
<td>methylation-sensitive arbitrarily primed PCR</td>
</tr>
<tr>
<td>MS-HRM</td>
<td>methylation-sensitive high resolution melting</td>
</tr>
<tr>
<td>MSP</td>
<td>methylation-specific polymerase chain reaction</td>
</tr>
<tr>
<td>MSP-ISH</td>
<td>methylation specific polymerase chain reaction in situ hybridisation</td>
</tr>
<tr>
<td>MS-RDA</td>
<td>methylation-sensitive representational difference analysis</td>
</tr>
<tr>
<td>MSRE</td>
<td>methylation-sensitive restriction enzyme</td>
</tr>
<tr>
<td>MSRE-PCR</td>
<td>methylation-sensitive restriction enzyme polymerase chain reaction</td>
</tr>
<tr>
<td>Ms-SNuPE</td>
<td>methylation-sensitive single nucleotide primer extension</td>
</tr>
<tr>
<td>MS-SSCA</td>
<td>methylation-specific single-strand conformation analysis</td>
</tr>
<tr>
<td>MVP</td>
<td>methylation variable position</td>
</tr>
<tr>
<td>MYOD</td>
<td>myoblast determination protein</td>
</tr>
<tr>
<td>n</td>
<td>number of sample size</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium chloride</td>
</tr>
<tr>
<td>NEB</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>NPTX2</td>
<td>neuronal pentraxin II</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotides</td>
</tr>
<tr>
<td>NTC</td>
<td>no template control</td>
</tr>
<tr>
<td>p</td>
<td>chromosome short arm</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PFN3</td>
<td>profilin-3</td>
</tr>
<tr>
<td>Pg</td>
<td>page</td>
</tr>
<tr>
<td>PRMT2</td>
<td>protein arginine N-methyltransferase 2</td>
</tr>
<tr>
<td>pTPM</td>
<td>protein-coding transcripts per million</td>
</tr>
<tr>
<td>q</td>
<td>chromosome long arm</td>
</tr>
<tr>
<td>qMSP</td>
<td>quantitative methylation specific polymerase chain reaction</td>
</tr>
</tbody>
</table>
QUMA quantification tool for methylation analysis
R reverse
RFU relative fluorescence unit
RLGS restriction landmark genomic scanning
RNA ribonucleic acid
RRBS reduced representation bisulphite sequencing
SAM S-Adenosylmethionine
SAP shrimp alkaline phosphatase
SAPS South African Police Services
SBE single base extension
SERPINb5 serpin family B member 5
SHANK3 SH3 and multiple ankyrin repeat domains protein 3
SNP single nucleotide polymorphism
SNuPE single nucleotide primer extension
STR short tandem repeats
T thymine
tDMR tissue-specific differentially methylated region
tDMS tissue-specific differentially methylated site
TiGER tissue-specific gene expression and regulation
TOMIL1 target of myb1 (chicken)-like 1
U units
u upstream
UCSC University of California, Santa Cruz
UF unmethylated forward primer
UKZN University of KwaZulu-Natal
UR unmethylated reverse primer
USP49 ubiquitin specific peptidase 49
UTR untranslated region
VF Vaginal Fluid
VILIP-3 visinin-like protein 3
viz namely
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>WBCs</td>
<td>white blood cells</td>
</tr>
<tr>
<td>WGSBS</td>
<td>whole-genome shotgun bisulphite sequencing</td>
</tr>
<tr>
<td>Xa</td>
<td>active X-chromosome</td>
</tr>
<tr>
<td>Xi</td>
<td>inactive X-chromosome</td>
</tr>
<tr>
<td>ZC3H12D</td>
<td>zinc finger CCCH-type containing 12D</td>
</tr>
<tr>
<td>ZNF282</td>
<td>Zinc finger protein 282</td>
</tr>
</tbody>
</table>
Abstract

The information gathered from crime scenes have the unique ability to direct the course of forensic investigations. In terms of criminal cases, the correct categorisation of human body fluids can provide significant leads, e.g., alleged sexual offences can be supported by the presence of vaginal fluid and semen, and signs of probable physical struggle can be inferred by blood at the crime scene. Methods for body fluid identification were initially developed using catalytic, enzymatic, chromatographic, and immunological procedures. Conventional methods, however, do not facilitate multiplexing and are not deoxyribonucleic acid (DNA) based. Tissue-specific differentially methylated regions/sites (tDMRs/tDMSs) are locations/sites on the human genome which display differential methylation patterns in tissues or cells, thus, enabling their use in body fluid identification. Techniques evaluating differential methylation in tDMRs have been particularly beneficial as only extracted DNA is processed, conserving additional physical evidence. Other benefits of DNA methylation-based methods include efficiency and multiplexing to evaluate multiple tissues in a single assay, therefore, is time-saving. The present study aimed to develop two DNA methylation-based assays: methylation-specific polymerase chain reaction (MSP) and methylation SNaPshot, in order to efficiently identify four different body fluids viz. blood, saliva, semen, and vaginal fluid, simultaneously and in a mixture.

Methylation profiling of novel DNA methylation markers was carried out by a qualitative MSP assay. The in-house developed ZNF282 and HPCAL1 gene-based tDMRs were employed for analysis of semen and saliva, respectively. The novel MSP primers were designed to target CpG islands in genes ZNF282 and HPCAL1. Two previously reported tDMS markers for blood (cg08792630; Park et al., 2014) and vaginal fluid (cg09765089-231d; Lee et al., 2015) were modified, to design MSP primers which targeted CpG sites flanking the reported sites for blood and vaginal fluid.

MSP analysis showed that saliva, semen, and vaginal fluid were correctly identified and differentiated from one another by HPCAL1, ZNF282 and cg09765089-231d markers, respectively. The complete unmethylation of HPCAL1 and ZNF282 in saliva and semen, respectively; and complete methylation in non-target body fluids demonstrated the immense potential both markers have for forensic application. Vaginal fluid marker revealed complete methylation in vaginal fluid. MSP analysis of the blood-specific MSP marker indicated that blood could neither be efficiently identified, nor differentiated from other body fluids of the study, due to the presence of both methylation and unmethylation.
A total of four SNaPshot primers were designed based on MSP amplicons, to target a single site which showed differential methylation between the body fluids. All SNaPshot markers were able to identify and differentiate their respective body fluids from others in SNapShot simplex reactions, by generating green (unmethylated) or blue (methylated) peaks only in target body fluids. Excluding the saliva marker, all others displayed high levels of specificity in the multiplex SNaPshot reaction. The clear observation of either complete methylation or complete unmethylation, depicted the high specificity of the markers for criminal investigations.

However, obtaining single-body fluid samples is a luxury that crime scenes do not afford. Most evidence is severely degraded, available in low quantities, and present in the form of mixtures. To assess the sensitivity of the designed multiplex MSP-SNaPshot assay, an assortment of body fluid mixtures in varying ratios were subjected to the MSP-SNaPshot assay. The high sensitivity of the semen and saliva marker was validated by the unambiguous identification of semen and saliva, even when target body fluids were minor constituents of the mixture. Vaginal fluid could not be identified when present in lower concentrations. The newly designed blood marker showed significant blood-methylation specificity and sensitivity by efficiently identifying the body fluid in low quantities. The present study reports novel MSP and SNaPshot markers for the identification of blood, saliva, semen, and vaginal fluid in single samples as well as mixtures. Future research would involve the use of the described markers, under different forensic conditions which will enhance their applicability in forensic analysis.
Organisation of dissertation

In order to achieve the stated objectives, the present 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 One
This chapter entails an introduction to the study.

Chapter Two
This chapter is a literature review, entailing a detailed explanation of the various aspects of this study.

Chapter Three
This chapter involved the design of primers for the determination of the methylation profile of candidate tDMRs in blood, saliva, semen, and vaginal fluid, using MSP. Specific primers targeting the gene body CpG islands, or regions flanked by CpG sites, were designed for methylation profiling.

Chapter Four
This chapter involved the design and optimisation of a multiplex methylation SNaPshot assay for the identification of blood, saliva, semen, and vaginal fluid, by novel markers. The sensitivity of the markers designed, and applicability of the multiplex MSP-SNaPshot method in forensic science, was investigated by the analysis of simulated mixtures.

Chapter Five
The final chapter includes the general discussion and conclusion, which provides an overview of the main objectives and findings described in each of the chapters of the dissertation. Possible limitations are acknowledged and a scope for future improvement and development is provided.
List of Figures

Figure 2.1: Diagrammatic representation of the DNA methylation mechanism. A methyl group (-CH$_3$) derived from SAM is transferred and covalently bound to the carbon five of the cytosine residue. This reaction resulting in the formation of 5mC is catalysed by DNMT enzymes. (Wong, 2003) 7

Figure 2.2: Association of DNA methylation with gene expression. (A) In the absence of methylated DNA, transcription factors are able to bind to the necessary recognition sequences, resulting in increased gene expression. The presence of elevated DNA methylation inhibits efficient binding of transcription factors, resulting in reduced gene expression. (B) Methyl-CpG-binding-proteins bind to methylated cytosine residues in the DNA thereby recruiting histone deacetylases (HDAC) and co-repressors (CoRep) to maintain an inactive chromatin state and bring about reduced gene expression. (Ling and Groop, 2009) 9

Figure 2.3: Summarized protocol of MeDIP and MCIp. Sheared and denatured DNA fragments are split into input DNA and immunoprecipitated (IP) DNA. Methylated DNA fragments are captured by antibodies (1° Ab) and recombinant proteins (MBD) in MeDIP and MCIp, respectively. Incubation is carried out with secondary antibodies (2° Ab) to bound beads that conjugate with the DNA-antibody or DNA-protein complex. Purification steps of repeatedly washing and digestion of the antibody/protein results in immunoprecipitated DNA. (Thu et al., 2010) 30

Figure 2.4: Mechanism of bisulfite conversion. Upon bisulfite treatment of DNA, methylated cytosines remain intact while unmethylated cytosines are deaminated to uracil. Subsequent amplification results in uracil amplified to thymine, and methylated cytosines amplified to cytosine. Sequencing will provide the methylation status of each cytosine on the DNA. (Yu and Snyder, 2016) 33

Figure 2.5: Representation of the MSP procedure. Isolated DNA is bisulfite-treated, converting unmethylated cytosines to uracil while methylated cytosines remain intact. Primer pairs specific for methylation and unmethylation are applied to PCR amplification. These primers will only amplify the sequence they are specific to. DNA methylation state of the sequence can be detected by simple gel electrophoresis. Completely unmethylated samples will produce a clear band in the unmethylated (U) reactions, and fully methylated samples show a distinct band in the methylated
Heterogenous tissue samples will produce amplification in both methylated and unmethylated reaction. MSP results can be quantified by methylation specific Polymerase chain reaction in situ hybridisation (MSP-ISH) or quantitative methylation specific Polymerase chain reaction (qMSP), with the inclusion of fluorescently labelled probes (*) (Derks et al., 2004).

**Figure 2.6:** Diagrammatic representation of multiplex SNaPshot reaction by SBE, and resultant methylation signals. Primers anneal one base pair upstream of the targeted CpG (SNP) site and are extended on the 3’ end by a fluorescently labelled ddNTP, complementary to the DNA sequence. After purification with SAP (Shrimp Alkaline Phosphatase), SNaPshot PCR products are separated via capillary electrophoresis according to size (nucleotides). Single peaks represent a homozygous SNP, and heterozygous SNPs are represented by the presence of two peaks at the same position. (Podini and Vallone, 2009)

**Figure 2.7:** Chromosomal location of ZNF282 gene (http://www.genecards.org/cgi-bin/carddisp.pl?gene=znf282; Fishilevich et al., 2016)

**Figure 2.8:** Chromosomal location of HPCAL1 gene (http://www.genecards.org/Search/Keyword?Query String=hpcal1; Fishilevich et al., 2016).

**Figure 2.9:** Chromosomal location of the FOXO3 gene (https://www.genecards.org/cgi-bin/carddisp.pl?gene=FOXO3&keywords=FOXO3; Fishilevich et al., 2016)

**Figure 2.10:** Location of the cg09765089-231d CpG site on chromosome 7 (UCSC Genome Browser; Kent et al., 2001)

**Figure 3.1:** Design of MSP primer targeting CGI of ZNF282 gene. a) UCSC genome browser view of chromosomal location of the target CGI within the gene indicated by the green bar, and the position of the CGI in relation to the introns and exons indicated by the blue horizontal line and box. b) MSP: Visual representation of the location of the primers designed within the CGI (blue coloured area). Two primer sets are indicated by the purple boxes (MF- methylated forward primer and MR- methylated reverse primer) and green boxes (UF- unmethylated forward primer and UR- unmethylated reverse primer). Horizontal red line indicates the input sequence and vertical red lines represent CpG sites.

**Figure 3.2:** Design of MSP primer targeting CGI of HPCAL1 gene. a) UCSC genome browser view of chromosomal location of the target CGI within the gene indicated by the green bar and
the position of the CGI in relation to the introns and exons indicated by the blue horizontal line and box. b) MSP: Visual representation of the location of the primers designed within the CGI (blue coloured area). Two primer sets are indicated by the purple boxes (MF and MR) and green boxes (UF and UR). Horizontal red line indicates the input sequence and vertical red lines represent CpG sites.

Figure 3.3: Design of MSP primer targeting region on the human genome for the cg09765089-231d tDMR. a) Visual representation of the location of the MSP primers designed. The absence of a CGI is indicated by no blue coloured area. Two primer sets are indicated by the purple boxes (MF and MR) and green boxes (UF and UR). Horizontal red line indicates the input sequence and vertical red lines represent CpG sites. b) UCSC genome browser view of chromosomal location of the region targeted within the human genome (chr7: 27291577).

Figure 3.4: Design of MSP primer targeting a single CpG site on FOXO3 gene. UCSC genome browser view of chromosomal location of the region (chr6: 10888910) targeted by the cg08792630 marker. No CGIs (usually indicated by a green box) are present in this region of the gene.

Figure 3.5: MSP based methylation profile of all four body fluids for the ZNF282 marker primer set. Lane 1: 100 bp ladder (Thermo Scientific). (A), (B), (C) and (D) represent results of semen, saliva, vaginal fluid, and blood, respectively. In A and B, lane 2 – lane 6: Products with primers specific for methylated cytosines (210 bp) and lane 9 – lane 13: Products with primers specific for unmethylated cytosines (210 bp). In C and D, lane 2 – lane 6: Products with primers specific for methylated cytosines (210 bp) and lane 10 – lane 14: Products with primers specific for unmethylated cytosines (210 bp). Lane 7 and lane 15 contain the no template control for the methylated and unmethylated primer set, respectively.

Figure 3.6: MSP based methylation profile of all body fluids for the HPCAL1 marker primer set. Lane 1: 100 bp ladder (Thermo Scientific). (A), (B), (C) and (D) represent results of saliva, semen, vaginal fluid, and blood, respectively. In A, lane 2 – lane 6: Products with primers specific for methylated cytosine (113 bp) and lane 9 – lane 13: Products with primers specific for unmethylated cytosine (114 bp). In B, lane 2 – lane 6: Products with primers specific for methylated cytosine (113 bp) and lane 8 – lane 12: Products with primers specific for unmethylated cytosine (114 bp). In C and D, lane 2 – lane 6: Products with primers specific for methylated cytosine (113 bp) and lane 9 – lane 13: Products with primers specific for
unmethylated cytosine (114 bp). Lane 7 and lane 14 contain the no template control for the methylated and unmethylated primer set, respectively.

**Figure 3.7:** MSP based methylation profile of all body fluids for the cg09765089-231d marker primer set. Lane 1: 100 bp ladder (Thermo Scientific). (A), (B), (C) and (D) represent results of vaginal fluid, saliva, semen, and blood, respectively. In A and D, lane 2 – lane 6: Products with primers specific for methylated cytosine (259 bp) and lane 9 – lane 13: Products with primers specific for unmethylated cytosine (262 bp). Lane 7 and lane 14 contain the no template control for the methylated and unmethylated primer set, respectively. In B and C, lane 2 – lane 6: Products with primers specific for methylated cytosine (259 bp) and lane 10 – lane 14: Products with primers specific for unmethylated cytosine (262 bp). Lane 7 and lane 15 contain the no template control for the methylated and unmethylated primer set, respectively.

**Figure 3.8:** MSP based methylation profile of all body fluids for the cg08792630 marker primer set. Lane 1: 100 bp ladder (Thermo Scientific). (A), (B), (C) and (D) represent results of blood, saliva, semen, and vaginal fluid, respectively. In A, lane 2 – lane 6: Products with primers specific for methylated cytosine (280 bp) and lane 9 – lane 13: Products with primers specific for unmethylated cytosine (280 bp). Lane 7 and lane 14 contain the no template control for the methylated and unmethylated primer set, respectively. In B – D, lane 2 – lane 6: Products with primers specific for methylated cytosine (280 bp) and lane 8 – lane 12: Products with primers specific for unmethylated cytosine (280 bp). Lane 7 and lane 13 contain the no template control for the methylated and unmethylated primer set, respectively.

**Figure 4.1:** Diagrammatic representation of forward and reverse SNaPshot primers annealing to DNA and resultant methylation signals. Primers anneal one bp upstream of the targeted CpG site and are extended on the 3’ end by a fluorescently labelled ddNTP, complementary to the DNA sequence. Primers which are synthesised based on the forward strand present black (methylated) and red (unmethylated) peaks, while reverse strand-based primers display blue (methylated) and green (unmethylated) peaks. (Forat *et al*., 2016)

**Figure 4.2:** Overview of methylation SNaPshot protocol carried out in the present study

**Figure 4.3:** Sequencing results of HPCAL1 MSP amplicon in all body fluids. Black and clear circles indicate a methylated and unmethylated CpG site, respectively. The black line represents the reference amplicon sequence to which sequencing products were aligned to. Blank spaces
indicate an unknown status at that site. The target CpG site is illustrated by the red rectangle. Numbers above represent the CpG site number and numbers below represent the nucleotide position of that specific CpG site in the amplicon.

**Figure 4.4:** Sequencing results of cg09765089-231d (vaginal fluid-specific) MSP amplicon in all body fluids. Black and clear circles indicate a methylated and unmethylated CpG site, respectively. The black line represents the reference amplicon sequence which sequencing products were aligned to. Blank spaces indicate an unknown status at that site. The target CpG site is illustrated by the red rectangle. Numbers above represent the CpG site number and numbers below represent the nucleotide position of that specific CpG site in the amplicon.

**Figure 4.5:** Sequencing results of cg08792630 (blood-specific) MSP amplicon in all body fluids. Black and clear circles indicate a methylated and unmethylated CpG site, respectively. The black line represents the reference amplicon sequence which sequencing products were aligned to. Blank spaces indicate an unknown status at that site. The target CpG site is illustrated by the red rectangle. Numbers above represent the CpG site number and numbers below represent the nucleotide position of that specific CpG site in the amplicon.

**Figure 4.6:** Sequencing results of ZNF282 (semen-specific) MSP reaction on all body fluids. Black and clear circles indicate a methylated and unmethylated CpG site, respectively. The black line represents the reference amplicon sequence which sequencing products were aligned to. Blank spaces indicate an unknown status at that site. The target CpG site is illustrated by the red rectangle. Numbers above represent the CpG site number and numbers below represent the nucleotide position of that specific CpG site in the amplicon.

**Figure 4.7:** Representative electropherograms of body fluids in simplex and multiplex methylation SNaPshot. Results for simplex reactions in semen, BL, VF, and saliva are represented in A, B, C and D, respectively; as well as the multiplex SNaPshot reaction represented in E. The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green and blue peaks indicated an unmethylated and methylated CpG site (reverse primer), respectively.

**Figure 4.8:** Percentage methylation or unmethylation (MIV) of each marker with its specific body fluid in the simplex reaction. Percentage methylation values were calculated by dividing nucleotide G intensity (blue peak) by nucleotide G plus nucleotide A (green peak) intensities.
Percentage unmethylation values were calculated by dividing nucleotide A intensity by nucleotide G plus nucleotide A intensities. 85

**Figure 4. 9:** Percentage methylation or unmethylation (MIV) of each marker in the multiplex SNaPshot reaction. Percentage methylation values were calculated by dividing nucleotide G intensity (blue peak) by nucleotide G plus nucleotide A (green peak) intensities. Percentage unmethylation values were calculated by dividing nucleotide A intensity by nucleotide G plus nucleotide A intensities. 85

**Figure 4. 10:** Representative electropherograms of vaginal fluid and semen mixtures in varying ratios. Vaginal fluid and semen mixture analysis is indicated by 9:1 ratio (A), 1:1 ratio (B) and 1:9 ratio (C). The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green peaks indicated an unmethylated CpG site (reverse primer), while blue peaks represent a methylated CpG site (reverse primer). 86

**Figure 4. 11:** Representative electropherograms of blood and semen mixtures in varying ratios. Blood and semen mixture analysis is indicated by 9:1 ratio (A), 1:1 ratio (B) and 1:9 ratio (C). The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green peaks indicated an unmethylated CpG site (reverse primer), while blue peaks represent a methylated CpG site (reverse primer). 87

**Figure 4. 12:** Representative electropherograms of saliva and semen mixtures in varying ratios. Saliva and semen mixture analysis is indicated by 9:1 ratio (A), 1:1 ratio (B) and 1:9 ratio (C). The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green peaks indicated an unmethylated CpG site (reverse primer), while blue peaks represent a methylated CpG site (reverse primer). 88

**Figure 4. 13:** Representative electropherogram of all four body fluids (blood, saliva, semen and vaginal fluid) mixed in an equal ratio. All four body fluids were mixed in a equal ratio of 1:1:1:1. The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green peaks indicated an unmethylated CpG site (reverse primer), while blue peaks represent a methylated CpG site (reverse primer). 88

**List of Tables**
Table 2.1: Different DNA methylation analysis methods

Table 3.1: Genomic location and target CpG sites of the MSP primers in this study

Table 3.2: MSP primers designed for the analysis of differential methylation patterns in human body fluids (blood, saliva, semen, and vaginal fluid)

Table 3.3: Summarised results for all MSP reactions on body fluids by the candidate markers

Table 4.1: Multiplex SNaPshot reaction setup

Table 4.2: Preparation of body fluid mixtures for MSP-SNaPshot multiplex assay

Table 4.3: Primers designed in the present study for SNaPshot methylation analysis
CHAPTER 1

Introduction
The analysis of crime scene exhibits provides insightful information to a forensic investigation in the criminal justice system. Identification of human body fluids in crime scene samples can indicate the nature of the crime and help reconstruct the course of events e.g., vaginal fluid and semen often indicate sexual offence (Vidaki et al., 2013), and blood may suggest physical struggle (Virkler and Lednev, 2009). The existing presumptive and confirmatory tests for body fluid identification are based on catalytic, enzymatic, and immunological techniques (An et al., 2012; Lee et al., 2016a). For decades, these techniques successfully identified body fluids and are sensitive enough however, the protocols are not always specific and do not allow for multiplexing or automation (Frumkin et al., 2011; Lee et al., 2016a). In addition, each test can detect the presence of only a single biological fluid and need large amounts of initial deoxyribonucleic acid (DNA) – a rarity in forensic cases (Kader, 2015). These restrictions have resulted in the continuous development of novel methods to overcome the shortcomings of existing methods (Kader and Ghai, 2015). One of the main objectives of new techniques is the identification of body fluids in a non-destructive manner, thereby, preserving DNA for use in downstream analyses e.g., short tandem repeat (STR) profiling (Zahra et al., 2018).

In the last decade, body fluid identification methods based on DNA methylation have gained the most attention from the scientific community. These developed methods facilitated improved and specific body fluid identification by demonstrating higher specificity and sensitivity (Rana, 2018). DNA methylation-based assays only target extracted DNA thereby preserving additional physical material (Frumkin et al., 2011; Gomma et al., 2017) – an essential factor in forensic cases wherein only minute quantities of the original sample are available. Other advantages include efficiency, convenience, and the ability to analyse multiple tissues in a single assay through multiplexing, thereby saving cost, time, and labour (Rana, 2018; Richards et al., 2018). Another benefit of using DNA methylation tests as opposed to RNA-based tests is that the equipment needed for the tests are standardly available in forensic DNA laboratories, as the same equipment is used for DNA profiling.

DNA methylation is by far the most widely researched and best characterised of all epigenetic phenomena. DNA methylation involves the covalent attachment of a methyl group to a cytosine residue in the DNA sequence, producing a 5-methylcytosine (5mC). This reaction is mediated by DNA methyltransferases (viz. DNMT1, DNMT3A and DNMT3B) (Gomma et al., 2017). In the human genome, a hotspot for DNA methylation is CpG dinucleotides with 60 % - 90 % being methylated (Richards et al., 2018). Unmethylated CpG regions are frequently clustered together and referred to as CpG islands (CGIs) with a base pair (bp) length of 300 bp - 3000 bp. CGIs are located in the promoter region of many genes (Kader and Ghai, 2015; Tammen et al., 2013). A
common observation of CGIs effect on gene expression is an inversely proportional relationship, *i.e.*, a decreased expression or inactivation of a gene when CGIs in promoters display increased methylation (Moore *et al.*, 2013).

Several genome-wide studies have shown DNA methylation profiles, specific to a particular tissue, are attributed by chromosomal regions known as tissue-specific differentially methylated regions (tDMRs) (Forat *et al.*, 2016; Holtkötter *et al.*, 2018; Lin *et al.*, 2016). Tissue-specific differentially methylated sites (tDMSs) are CpG sites on the human genome which display differential methylation patterns in tissues/cells. The differential methylation patterns provide a distinguishing characteristic between tissues and cells thus enabling their use in body fluid identification; however, these patterns are susceptible to change in response to environmental stimuli, *e.g.*, diet and smoking (Kader and Ghai, 2015; Vidaki and Kayser, 2018). The DNMT enzymes are involved in the maintenance of tDMR methylation patterns.

Identification of tDMRs/tDMSs in the human genome (Eckhardt *et al.*, 2006) resulted in the emergence of epigenetics in the field of forensics. Numerous studies have reported successful body fluid identification individually and in mixtures, by the employment of tDMRs and DNA methylation analysis techniques (An *et al.*, 2013; Frumkin *et al.*, 2011; Holtkötter *et al.*, 2018; Lee *et al.*, 2012, 2016b; Lin *et al.*, 2016; Madi *et al.*, 2012). **Bisulfite conversion is one of the most important preliminary steps in DNA methylation analysis.** This DNA modification technique uses sodium bisulfite which deaminates unmethylated cytosines to uracil, and methylated cytosines remain intact (Richards *et al.*, 2018). Polymerase chain reaction (PCR) amplifies uracils to thymine and cytosines to cytosine, therefore, the ratio of cytosines to thymines (C/T) represents the methylation status. A vast number of DNA methylation analysis methods based on bisulfite conversion are available. These include direct, or cloned bisulfite genomic sequencing (Frommer *et al.*, 1992), **methylation-specific polymerase chain reaction (MSP)** (Herman *et al.*, 1996), methylation-sensitive high resolution melting (MS-HRM) (Wojdacz and Dobrovic, 2007), and methylation SNaPshot (Kaminsky and Petronis, 2009).

This study aimed to develop two DNA methylation-based assays in order to efficiently identify four different body fluids *viz.* blood, saliva, semen, and vaginal fluid. **MSP** was developed to identify potential tDMRs/tDMSs for differentiation between body fluids. A differentially methylated CpG site on MSP amplicons was targeted to design primers for the multiplex methylation SNaPshot assay. The MSP-SNaPshot assay allowed for multiplexing of four primers and body fluid mixture analysis. The successful application of the developed multiplex MSP-SNaPshot assay would allow simultaneous identification of saliva, blood, semen, and vaginal fluid in a single reaction. The ability to efficiently categorise body fluids in mixtures, especially
semen and vaginal fluid, would provide invaluable insight into criminal cases as these are specifically significant in a South African context. The social connotations of confirming such sensitive information would be historic.
CHAPTER 2

Literature Review
2.1 The phenomenon of epigenetics

In 1942, the term “epigenetics” was coined by a British embryologist known as Conrad Waddington. Waddington defined the term as “the branch of biology which studies the casual interactions between genes and their products, which bring the phenotype into being” (Waddington, 1942). The coined term joined two disciplines which were until then considered separate i.e., genetics and developmental biology. With an increase in numerous unanswered and unexplained essential developmental features, the importance of collaboration between the fields of developmental biology and genetics, became more necessary. One such feature involved the phenotypic stability of differentiated cells (fibroblasts or lymphocytes) through the process of cellular division. This highlighted that some genes responsible for the phenotype of differentiated cells were constantly switched on while other active genes in another cell were constantly switched off (Holliday, 2006). Arthur Riggs and colleagues updated this definition in 1996 to “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence” (Riggs et al., 1996). A final consensus of the definition was decided in 2008 stating that an epigenetic trait is a “stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence” (Berger et al., 2009).

The four major classes of epigenetic modifications include: DNA methylation, posttranslational histone modification, chromatin remodelling and non-coding ribonucleic acids (RNAs). These heritable modifications affect gene expression devoid of any changes to the associated DNA sequence. The inheritance of the epigenome has several differences from the traditional genetic inheritance, with a lot left to be understood about the epigenetic system. Epigenetic modifications are more likely to occur in groups of cells, unlike classical genetics wherein the occurrence of a somatic mutation, all descendants are expected to have the same genotype (Holliday, 2006). Genetic mutations are stable and often irreversible whereas epigenetic changes are usually reversed e.g., epigenetic modifications made in genomic imprinting could be lost during development or erased during gametogenesis (Wong et al., 2010). In the last decade, there has been a surge of available literature in the field of epigenetic inheritance over subsequent generations (transgenerational and intergenerational). The fascination stems from the idea that biological, physiological, and psychological traits resulting in epigenetic changes can be passed on from parent to offspring – thereby filling in gaps of research that have been present for an extremely long time.

2.2 DNA methylation

DNA methylation is by far the most widely researched and best characterised of all previously mentioned epigenetic phenomena. DNA methylation is a reaction which produces a 5mC (also
known as the fifth base) by covalently attaching a methyl group (CH$_3$) to a cytosine residue in the DNA sequence. The methyl group, derived from S-Adenosylmethionine (SAM), is transferred onto the C5 position of a cytosine residue (Figure 2.1). This reaction is mediated by DNMTs (viz. DNMT1, DNMT3A and DNMT3B) (Kader, 2015). DNMT enzymes are classified into two groups governed by their function viz. maintenance and de novo. DNMT1 is the maintenance methyltransferase, as it is responsible for mimicking the methylation pattern of the parent strand to the daughter strand during DNA replication (Fukagawa et al., 2014) as well as, repairing methylation (Ha et al., 2010). De novo methyltransferases are covered by DNMT3a and DNMT3b since these play a role in early development to introduce methylation patterns (Moore et al., 2013). DNA methylation plays a fundamental role in mammalian development, cellular differentiation, and maintenance of different characteristics of cells, by controlling and regulating gene expression through chromatin structure (Li and Zhang, 2014).

Proceeding the discovery of DNA as genetic material (Avery et al., 1944; McCarty and Avery, 1946), the modified cytosine residue was discovered after conducting a paper chromatography protocol in preparation of calf thymus (Hotchkiss, 1948). Rollin Hotchkiss observed that the residue separated from cytosine in the same manner that thymine (methyluracil) separated from uracil. Based on these observations it was hypothesized that the residue was 5mC (Hotchkiss, 1948).

**Figure 2.1:** Diagrammatic representation of the DNA methylation mechanism. A methyl group (-CH$_3$) derived from SAM is transferred and covalently bound to the carbon five of the cytosine residue. This reaction resulting in the formation of 5mC is catalysed by DNMT enzymes. (Wong, 2003)
Methylated cytosines only account for ~1% of the entire human genome (Tost, 2010) and are predominantly found in CpG dinucleotides within the genome. Situated within the human genome are approximately 30 million CpG dinucleotides of which 60% - 90% are present in a methylated state (Tost, 2010). Unmethylated CpG regions are frequently found in groups and referred to as CpG islands (CGIs). Gardiner-Garden and Frommer (1987) carried out the first computational analysis of CGIs. CGIs were defined as regions with a base pair (bp) length of > 200 bp, GC content > 50%, as well as > 0.6 ratio of observed to expected (Illingworth and Bird, 2009; Kakumani et al., 2012).

Approximately 50% of CGIs are located in the promoter region of many genes, specifically housekeeping genes (Kader and Ghai, 2015; Ohgane et al., 2008; Tammen et al., 2013). Illingworth et al. (2010) reported that these promoter-based CGIs display high levels of conservation between mice and humans. CGIs comprise of fewer nucleosomes (histone octamer wrapped by a stretch of DNA) compared to other regions of the DNA (Choi, 2010; Tazi and Bird, 1990). This is an important feature as these nucleosomes contain modified histone molecules improving gene expression. Methylated CpG sites are a mutational hotspot as spontaneous deamination (removal of an amino group) occurs to methylated cytosines, thus forming thymine bases (CpG → TpG) (Xia et al., 2012). This explains the low frequency of methylated CpG dinucleotides present in the genome. Conversely, spontaneous deamination on unmethylated cytosines results in the conversion to uracil. The cell rapidly recognises this error and rectifies it accordingly (Lander et al., 2001).

2.3 Association between DNA methylation and gene expression (transcription)

A recurring observation is the inversely proportional relationship between methylation status of CGIs in promoter regions and the effect on gene expression. A decreased expression or inactivation of a gene occurs when CGIs in promoters display increased methylation levels (Zemach et al., 2010). Genes which are actively transcribed usually contain unmethylated CpG sites in their promoters and first exons whilst promoters of transcriptionally silent genes often exhibit a high number of methylated CpG sites. The first gene exons of transcriptionally active chromatin (euchromatin) are frequently unmethylated CpG dinucleotides however, methylated DNA is associated with transcriptionally inactive heterochromatin (Siegfried and Simon, 2010).

Gene transcriptional repression could be caused by two mechanisms: physical obstruction or binding of methyl-CpG-binding proteins (MBPs) (Figure 2.2). The former mechanism is due to the physical obstruction by methyl groups attached to the DNA thereby inhibiting the binding of necessary transcription factors to the gene (Tampe and Zeisberg, 2013). The latter mechanism
recognises and binds MBPs to methylated cytosines on the DNA sequence, prompting gene repression (Tampe and Zeisberg, 2013). It has been demonstrated that methylated cytosines bound to MBPs further recruit chromatin remodeling proteins (e.g., histone deacetylases), which assist in maintaining an inactive chromatin state i.e., heterochromatin. Along with maintenance of this state, MBPs utilise transcriptional co-repressor molecules leading to a transcriptionally silent gene (Gibney and Nolan, 2010).

Restriction enzymes coupled with blot hybridisation on a variety of tissues was conducted to initially study the DNA methylation patterns of genes and the relation to gene activity. These experiments were carried out on the globin gene of rabbit (Waalwijk and Flavell, 1978), chicken (McGhee and Ginder, 1979), and eventually humans (van der Ploeg and Flavell, 1980), by employing enzymes \textit{HpaII} (CCGG) and \textit{HhaI} (GCGC). These enzymes were incorporated into the study due to its inhibition that occurred if the respective restriction sites contained methylated cytosines (Engel \textit{et al}., 2009; Razin and Cedar, 1991). All studies clearly indicated unmethylated cytosines were found in examined tissues that expressed the globin gene.

\textbf{Figure 2.2: Association of DNA methylation with gene expression.} (A) In the absence of methylated DNA, transcription factors are able to bind to the necessary recognition sequences, resulting in increased gene expression. The presence of elevated DNA methylation inhibits efficient binding of transcription factors, resulting in reduced gene expression. (B) Methyl-CpG-binding-proteins bind to methylated cytosine residues in the DNA thereby recruiting histone deacetylases (HDAC) and co-repressors (CoRep) to maintain an inactive chromatin state and bring about reduced gene expression. (Ling and Groop, 2009)

Kuroda \textit{et al}. (2009) studied the effect of DNA methylation on insulin and documented that partial or complete methylation of CpG sites was present in promoter regions of non-insulin cells; furthermore, insulin-producing cells displayed complete demethylation. This study provided data which revealed that complete methylation of promoters in the insulin gene of mice and humans
repressed gene expression by 79 % and 85 %, respectively. Additionally, the experimental findings concluded that higher methylation levels facilitate the attachment of MBPs (MeCP2 in this case) which prohibits the binding of transcription factors along with the recruitment of HDACs, aiding in sustaining the condensed form of chromatin.

Methylated DNA inhibits 45 % of transposable and viral elements. The expression of these elements could be detrimental – leading to DNA mutations and gene interference (Moore et al., 2013). For example, the destructive mouse genome retrovirus (intracisternal A particle – IAP) retains high methylation levels from the formation of gametes to mature adulthood through the DNMT1 maintenance enzyme (Li et al., 2015). This validates the role of DNA methylation in suppressing harmful genetic components.

2.4 Gene body methylation and gene expression: The DNA methylation Paradox

A vast number of studies have reported a negative correlation between DNA methylation in the gene promoter and expression of the gene. Contrary to this, fewer studies have reported a positive correlation between DNA methylation in gene bodies and expression of the gene i.e., high levels of DNA methylation in gene bodies have been associated with an overexpressed gene (Aran et al., 2011; Hellman and Chess, 2007; Lister et al., 2009). This disparity is referred to as the DNA methylation paradox (Jones, 1999). One of the more popular explanations that have been suggested to unravel this contradiction, is that increased DNA methylation drives the suppression of spurious transcription in genes (Jjingo et al., 2012). However, it is important to note that methylation in these regions should not prevent the transcriptional process of the gene, simply repress expression of spurious promoters. Based on data received from Cap Analysis Gene Expression (CAGE), it can be assumed that this paradox functions in intragenic transcription repression (Jjingo et al., 2012). Improved transcriptional elongation could take place by inhibiting intragenic promoters via DNA methylation, thereby elucidating the positive correlation relationship between gene body methylation and gene expression.

Maunakea et al. (2010) identified a negative correlation between intragenic methylation and intragenic promoter activity in the human SHANK3 locus. This finding revealed that increased methylation in the intragenic promoter region displayed repression of transcription initiation in the promoter. A monotonic positive correlation was demonstrated between gene-body methylation (GBM) and gene expression from the analysis of various human tissues i.e., an increase in one factor was supported by an increase in the other (Aran et al., 2011).

Hellman and Chess (2007) performed an allele-specific analysis to determine the methylation pattern on active (Xa) and inactive (Xi) X-chromosomes. Xa contains methylation in double
dosage as compared to the Xi, with these methylation-rich regions being the gene bodies. The epigenetic mechanism of DNA methylation is integral to preserve the silent form of Xi (brought about by X-chromosome inactivation) (Dupont et al., 2009). Results of cytogenetic experiments (Viegas-Pequignot et al., 1988) had shed light on the global hypomethylation of the Xi even though inactivated regions generally display increased methylation levels. The results of this study (Hellman and Chess, 2007) unveiled large-scale methylation of gene bodies on Xa (irrespective of gender).

Jjingo et al. (2012) evaluated this paradoxical relationship by carrying out several analyses on existing genome-scale datasets. While some of the findings supported existing explanations, some revealed a clear distinction from previous reports. The analysis of the relationship between GBM and intragenic transcriptional repression had shown that all considered cell lines exhibited a negative correlation between intragenic promoter methylation and corresponding transcriptional initiation levels. Surprisingly, another data analysis on five different cell lines provided a coherent non-monotonic, bell-shaped relationship between GBM and gene expression. Mid-level expressed genes showed the greatest methylation levels while the lowest and highest expressed genes showed substantially low methylation levels.

2.5 Effect of environmental factors on DNA methylation

Changes in the environment e.g., age, diet, and lifestyle, have shown to result in changes in DNA methylation patterns. Monozygotic twins contain identical methylation patterns until they are exposed to, or affected by environmental differences (Bocklandt et al., 2011). For this reason, they provide the perfect model to study the effects of these changes on DNA methylation. Pirazzini et al. (2012) studied two regions in monozygotic twins (60 years old) and discovered a significantly higher intra-twin difference in methylation levels. An increase in age is associated with a decrease in the efficiency of internal machinery to sustain DNA methylation patterns during cell division (Aunan et al., 2016). Myoblast determination protein 1 (MYOD) gene and insulin-like growth factor 2 (IGF2) gene encode an oestrogen receptor. At a young age, methylation in the CGIs of these genes are not detectable however, aging brings about an observable methylation signal (Christensen et al., 2009).

The one-carbon metabolism pathway functions with the use of vitamins B6 and B12, betaine, folate, and choline as well as amino acids. In the absence of any of these components, an altered DNA methylation pattern is noticed. This is due to the involvement of the pathway in determining the availability of the SAM which provides the methyl moiety (Maddocks et al., 2016). Alcohol intake decreases the availability of SAM by modifying the amount of Vitamin B as well as, depleting choline and methionine. A decrease in SAM brings about a subsequent methylation
variation (Zakhari, 2013). Consumption of green tea has reported a reduction of DNA methylation shown in cancer cells due to epigallocatechin-3-gallate (in green tea) inhibiting DNMTs thus, removing the repression placed on tumour suppressor genes (Yiannakopoulou, 2015).

DNA methylation has also been related to psycho-social factors (cortisol output and stress) and early-life socio-economic status (Christensen et al., 2009). Radtke et al. (2011) examined whether a prolonged effect was present in DNA methylation of offspring whose mothers experienced gestational maternal adverse experiences e.g., physical abuse. A positive correlation was noted between violence exposure during pregnancy and increased methylation of the children’s glucocorticoid receptor (GR) promoter. The results indicated that changes in the intrauterine environment have an effect on GR methylation in children.

2.6 Tissue-specific differentially methylated regions (tDMRs)

TDMRs are regions on the chromosome which display differential DNA methylation patterns (hypomethylation or hypermethylation) according to tissue/cell type (Kader and Ghai, 2015). TDMS is a single CpG site which displays variation in methylation profiles among different tissues. TDMS is also referred to as a methylation variable position (MVP) (Lee et al., 2016a). The margins of CGIs have been found to be rich in tDMRs/tDMSs. These tDMRs have a lower GC and CpG content than surrounding regions. TDMRs are present with transcription-factor binding sites (Anastasiadi et al., 2018; Wan et al., 2015) which have been reported to facilitate or inhibit the binding of specific factors in a methyl-dependent manner (Cohen et al., 2011; Vidaki et al., 2013). The DNMT enzymes are involved in the maintenance of tDMR methylation patterns.

2.7 Identification of tDMRs/tDMSs

The stability of tDMRs/tDMSs and their capability to differentiate between various tissues had ignited a surge in research towards their identification in the genome. Eckhardt et al. (2006) conducted a study to identify the methylation profile of chromosomes 6, 20 and 22 in humans. A bisulfite sequencing PCR assay was used to determine differentially methylated regions/sites in twelve different tissues. The analysis of amplicons associated with 873 genes resulted in the identification of the methylation status at 1.88 million CpG sites. Each specific type of tissue was grouped together in the hierarchical clustering results thus, indicated tissue-specific methylation patterns. Of the amplicons analysed, 22 % were differentially methylated in the various tissues i.e., tDMRs. Some of the genes, e.g., SERPINB5, which are involved in the regulation of methylation at promoter sites corresponded to messenger RNA (mRNA) silencing when the 5’ UTR (untranslated region) of the genes were methylated i.e., transcriptional silencing by methylation of the promoter region.
Baron et al. (2006) designed an experimental assay to determine if DNA methylation patterns could be used in the identification of different cell types. DNA from cells of different tissues were subjected to a designed protocol to identify markers which displayed differential methylation among these tissues. Out of 21 cell type-specific differentially methylated gene regions (CDMs) identified, eight were selected which showed the highest discriminatory power. The markers identified and employed by this study were able to unambiguously identify cells derived of the ectoderm and mesoderm. The differential methylation levels of the markers in each tissue allowed for successful differentiation of the ectoderm from the mesoderm as well as, different cells within both regions.

Slieker et al. (2013) reported the identification and annotation of tDMRs by using Illumina 450k DNA methylation chip data in conjunction with an algorithm. Data were collected for four peripheral tissues and six internal tissues. From the tDMRs detected in peripheral and internal tissues, 40% showed no overlapping between the two datasets i.e., markers were unique, and 23% of markers were in both peripheral and internal tissues. As expected, the markers clustered together according to germ layer, which further validated their tissue specificity. The identified tDMRs were mapped to nearest genes to corroborate if set genes are expressed in the examined tissues. Tissue-specific hypomethylation was observed in the preferentially expressed gene when compared with other tissues (Slieker et al., 2013).

Ma et al. (2013) conducted a methylation-sensitive representational difference analysis (MS-RDA) assay to isolate differentially methylated fragments from venous blood, saliva, semen, and vaginal fluid. A quantitative DNA analysis was carried out by Sequenom Massarray®. The study aimed to identify tDMRs for application in forensic casework. A total of six blood-specific differentially methylated markers were identified of which two displayed blood-specific hypomethylation, and the remaining four displayed blood-specific hypermethylation. Ma et al. (2013) stated that “obvious” differences were noted between blood-specific methylation and methylation in other body fluids; but failed to observe methylation differences among different individuals. However, with alternative testing techniques e.g., effective clone-based bisulfite sequencing analysis by Lee et al. (2012), the aforementioned markers could potentially differentiate blood from other body fluids.

Genome-wide methylation study performed by Lokk et al. (2014) profiled tDMRs in seventeen somatic tissues to analyse the methylation levels of these regions and their association with gene regulation. Many similar tissues were grouped together in the hierarchical clustering analysis, indicating tissue-specific methylated profiles. Highly methylated CpG sites were located in regions with low CpG presence, while hypomethylated CpG sites were largely found in gene
promoters. This supported previous reports which elucidated that gene promoters contain frequently expressed/transcribed genes and therefore, are unmethylated for better accessibility to transcription factors. CGI-promoter regions consisted of hypermethylated genes involved in the reproductive system and, hypomethylated genes in these regions were responsible for housekeeping processes. Regions which revealed greater variation between tissues were gene bodies, 3’-UTR and sites not associated with genes. Lokk et al. (2014) determined that variation in DNA methylation patterns were affected more by tissues than individuals, thus, inter-individual variation was suggested to be “insignificant” in this study. Lastly, results of the study supported previous reports of a more negative and/or inverse relationship than positive, between gene expression and DNA methylation.

2.8 Application of differential DNA methylation in forensics

Forensic science plays a vital role in the criminal justice system by providing scientifically-based information through the analysis of physical evidence, legally known as exhibits, from crime scenes. Body fluid identification of samples collected from crime scenes contribute towards solving criminal cases by revealing essential insights. One of the most important steps in investigations is the reconstruction of crime scenes, which is enabled by the analysis of evidence, resulting in a link between sample donors and the actual crime exhibits (An et al., 2012).

Previously applied conventional, presumptive, and confirmatory methods for body fluid identification utilise catalytic, enzymatic, and immunological-based techniques (Virkler and Lednev, 2009). Although these were successful and are still used to an extent, the methodologies are labour-intensive, time-consuming, as well as lack specificity, sensitivity, and stability. Furthermore, the methods are only applicable to specific biological fluids and require large amounts of DNA which generally is not possible in forensic cases (Kader and Ghai, 2015). These disadvantages may generate false positives or false negative results. Body fluid identification in forensic investigations could result in the probable conviction or exoneration of an individual, thus identification methods with many limitations in important characteristics are a major concern (Frumkin et al., 2011).

Scientists are therefore, constantly developing novel techniques for body fluid identification. DNA is the most frequently used biological source for profiling individuals, therefore it is crucial to ensure the preservation of samples. One of the most vital objectives of all newer methods is to maintain the integrity of crime scene evidence. This can be achieved through identification of samples in a non-destructive manner to provide the best possible sample quality for downstream analyses e.g., STR analysis for individual identification (Zahra et al., 2018). The differential DNA methylation patterns of tDMRs/tDMSs has been indispensable to forensic casework due to
their ability to differentiate tissues/cells based on specific hypomethylation and/or hypermethylation patterns.

2.8.1 Sex determination studies

One of the more prominent uses of gender identification in forensics is to differentiate between the victims and perpetrators in sexual offence cases. However, it is also applicable in casework regarding the analysis of remains from mass disasters and missing persons cases (Butler, 2012). The study by Naito et al. (1993) proposed the application of DNA methylation in sex determination by reporting that the DXZ4 region was hypermethylated on Xa and hypomethylated on Xi. Boks et al. (2009) reported that even the most significant locus of the study, could not completely differentiate between females (0.75) and males (0.79) but noted an overall good correlation between gender and DNA methylation. More than 50% of the autosomal probes displaying substantial correlation with gender were located in non-CGIs (Boks et al., 2009). This result supported previous reports made by Igarashi et al. (2008), that stated differential DNA methylation is not restricted to promoter regions or CGIs. Zhang et al. (2011) discovered a global methylation difference in the Long Interspersed Repeat Sequence 1 (LINE1) transposable element between men and women to be 1.8%, with women being lower. This finding was corroborated by Hsuing et al. (2007), El-Maari et al. (2007) and Zhu et al. (2012) as well as, suggested that the decreased levels of methylation seen in females could be as a result of consuming differing amounts of one-carbon nutrients e.g. methionine and B vitamins, or dietary folate. (Zhang et al., 2011)

2.8.2 Age estimation

Chemical modifications, such as DNA methylation, have a partial but significant effect on aging (Bocklant et al., 2011), therefore DNA methylation analysis can be useful in determining the age of an individual. Age estimation is a valuable tool in cases involving unknown persons and disaster victim identification (Zubakov et al., 2016). A relationship displaying the linear correlation between aging and differential DNA methylation in CpG sites had been reported (Christensen et al., 2009). Bocklandt et al. (2011) studied this correlation by analysis of saliva samples from identical twins, aged between 21 and 55 years. Differential methylation patterns at promoter-sites in three genes (EDARADD, TOM1L1 and NPTX2) were shown to be associated with age-related diseases e.g. cardiovascular and neurological disease. Zbieć-Piekarska et al. (2015) analysed DNA methylation in eight genes previously reported by Hannum et al. (2013). The study created a linear regression age prediction model which explained 94% of age variance with a 4.5 years precision and correctly predicted the age of 72% of individuals. The age of children and teenagers were the most correctly predicted while participants aged between 60 years
-75 years were the least correctly predicted. These results not only demonstrated the accuracy of the age prediction model but also provided a clear deterministic between children and older individuals.

2.8.3 Distinguishing between artificially synthesised and naturally occurring DNA

There is an existing assumption that DNA recovered from crime scenes is authentic and a naturally occurring material. However, in vitro synthesised, or artificial DNA can be easily generated, incorporated with legitimate human samples, and planted at the crime scene (Frumkin et al., 2010). This may result in the false conviction of individuals (Frumkin et al., 2010) therefore, the successful detection of artificial DNA and its differentiation from natural DNA, is extremely beneficial to forensic science and the judicial system. A study by Frumkin et al. (2010) demonstrated that artificially synthesised DNA provided an identical profile to natural DNA, despite undergoing forensic procedures. For validation studies, the developed authentication assay based on different methylation patterns of in vivo (natural) and in vitro (artificial) DNA, successfully determined the methylation status of twenty mock forensic samples. Wang et al. (2015) successfully identified artificial DNA from natural DNA by employing an assay involving the use of the methylation-sensitive restriction enzyme (MSRE), HhaI, on seven loci, which displayed differential DNA methylation patterns. Blood and saliva natural DNA samples were correctly identified by efficient amplification only at the loci specific for those body fluids. The in vitro synthesised bloodstains did not amplify at any loci of the study, indicating that bloodstains (natural and artificial) could be differentiated from one another.

2.8.4 Differentiation between monozygotic/dizygotic twins

Monozygotic twins are formed by the splitting of a zygote (single egg fertilized by a single sperm cell) which gives rise to two foetuses that are genetically identical, thus have identical methylation patterns. The appearance of variation in DNA methylation between twins only occurred when exposed to different environmental factors (Bocklandt et al., 2011). For this reason, monozygotic twins are a favourable model in studying how several environmental factors cause phenotypic differences (Fraga et al., 2005; Hannon et al., 2018). Coolen et al. (2011) performed quantitative mass spectrometry to analyse differential DNA methylation patterns between monozygotic and dizygotic twin pairs. Analysis of four imprinted control region genes and one non-imprinted gene displayed similar methylation patterns between each twin group. Statistical analysis indicated that methylation patterns in H19 and IGF2 were significantly different between monozygotic and dizygotic twin pairs. Fewer differences in the methylation pattern of H19 were noticed in monozygotic twins as opposed to dizygotic twins. Peripheral blood samples from female and
male monozygotic twins were examined by Li et al. (2013). A total of 92 differentially methylated CpG sites were identified and successful in the differentiation of each twin pair. However, it was hypothesised that the identified CpG sites may only be applicable to blood samples as only that body fluid was analysed in this study.

2.8.5 Human body fluid identification

DNA has been noted as the more predominantly analysed biological source for identifying individuals. Therefore, maintaining the quality and quantity of DNA for body fluid identification in crime scene evidence is imperative, specifically for downstream analyses e.g., DNA profiling by STRs (Liu et al., 2019). Molecular markers derived from mRNA, microRNA (miRNA) and microbes also have potential in cell-type identification however, disadvantages of their application include necessity of additional procedures, spurious transcription causing background signals, and are not completely human-specific. Additionally, the methylation pattern exhibits inter- and intra-individual variation, and is likely to alter with diseases (Sijen, 2015). Body fluid identification methods based on differential DNA methylation patterns are advantageous as they demonstrate the greater specificity, sensitivity, and stability necessary for forensic samples (Kader and Ghai, 2015). The present study focuses on efficient identification of body fluids using DNA methylation analysis.

Frumkin et al. (2011) reported successful differentiation between blood, saliva, semen, and epidermis samples by 38 selected loci. MSRE-PCR with subsequent capillary electrophoresis distinguished body fluids from one another by the calculation of a methylation ratio for every differentially methylated co-amplified loci. L91762 and L68346 loci presented with hypomethylation and hypermethylation patterns, respectively, in semen samples; thereby enabling their use in accurately identifying semen. The high methylation levels displayed by L91762 in blood, saliva and epidermis allowed for the discrimination of semen from these body fluids, with this marker. The methylation ratio of L91762/L68346 in semen (0.04 – 0.53) and other body fluid ratios (ranged between 2.15 – 18.28) further verified the distinction of semen. Identification of skin epidermis and its differentiation from other body fluids, was revealed by higher methylation ratios in L91762/L68346 and L76138/L26688. In order to detect and identify semen, two loci (L68346 and L16264) were tested on pure semen samples, urine from males, venous blood, menstrual blood, vaginal secretion, and saliva. L68346 only amplified in semen and L16264 amplified all other body fluids, except semen. The tissue identification algorithm precisely identified the presence or absence of semen in all samples of this study.
Madi et al. (2012) applied four differentially methylated markers (C20orf117, ZC3H12D, BCAS4, and FGF7) in a pyrosequencing assay, to efficiently identify body fluids (blood, saliva, sperm and epithelial cells). The results for C20orf117 were in line with prior reports by Eckhardt et al. (2006) seeing as this tDMR was hypermethylated in white blood cells (CD4⁺ and CD8⁺ lymphocytes) when compared to the other body fluids. Methylation analysis of ZC3H12D showed that all CpG sites of the marker were hypomethylated in all of the sperm samples, and hypermethylated in all blood, saliva, and epithelial samples – providing differentiation of semen from the other body fluids. Madi et al. (2012) found that results for BCAS4 contradicted those previously reported by Eckhardt et al. (2006), as the tDMR was found hypermethylated in sperm instead of saliva. Methylation levels of FGF7 were higher in sperm when compared to the other body fluids in the study. Accessing the reproducibility of results was an essential step; therefore, one saliva sample was processed five times, modifying the DNA concentration for bisulfite conversion each time. All replicates presented with similar results and no considerable changes from previously observed methylation values.

A panel of candidate markers (DACT1, USP49, HOXA4, PFN3, and PRMT2) were used in body fluid identification of venous blood, saliva, semen, menstrual blood, and vaginal fluid (Lee et al., 2012). Differential DNA methylation was displayed in all five tDMR markers specific to the body fluids of the study. DACT1 and USP49 tDMRs exhibited hypomethylation in 93% and 97% of semen clones, respectively; as well as hypermethylation in all clones of the other body fluids. DACT1 and USP49 validated the successful identification and discrimination of semen from other bodily fluids. Lee et al. (2012) confirmed the stability of the semen-specific markers (DACT1 and USP49) with significantly successful results from 30-days old samples. HOXA4 displayed hypomethylation in all semen samples, 50% of saliva from males, menstrual blood and vaginal fluid, while hypermethylation was present in all blood and female saliva samples. PRMT2 was hypomethylated in all semen samples and hypermethylated in ≥ 50% of menstrual blood and vaginal fluid samples. Despite differential methylation patterns of HOXA4 and PRMT2 markers, the low variation in methylation between body fluids created difficulty in complete identification of a single body fluid in the study. PFN3 marker was hypermethylated in relatively all body fluids (≥ 80% of CpG loci were methylated); with approximately 65% of the CpG sites in the marker being methylated in vaginal fluid and statistically different from other body fluids. These results established the potential of the PFN3 tDMR in the identification of vaginal fluid; provided that the marker is coupled with DNA methylation analysis techniques, which are site-specific e.g., methylation SNaPshot.
A validation study of the DSI-Semen™ kit (Nucleix, Tel Aviv, Israel) for the identification of semen was conducted by LaRue et al. (2012). The kit’s chemistry is based on digestion by restriction endonucleases at methylated CpG sites. The semen samples subjected to the procedure were all successfully identified as peaks were only present at the semen-specific markers; and the absence of semen was clearly indicated by peaks only at non-semen specific markers. This study further investigated the efficiency of the kit by including tests which assessed the limitation of detection, reproducibility, effect of environmental inhibitors, and mixtures. Majority of the tests were successful (except DNA concentrations lower than 31 pg and hematin concentration greater than 5 µM), and visibly identified semen thus, validating the use of the kit in forensic science. However, LaRue et al. (2012) suggested that visual/manual interpretation of results were superior to interpretation by software analysis. It was cautioned that the conservative nature of the software designed included stringent parameters which presented a large number of the positive results as inconclusive.

An et al. (2013) investigated the methylation status of four tDMRs (DACT1, USP49, PRMT2 and PFN3) by two independent assays: MSRE-PCR, and multiplex methylation SNaPshot. Blood and saliva results of the MSRE-PCR method had shown differential methylation levels in all four markers. The highest and lowest methylation levels were reported in DACT1 and PRMT2, respectively. Methylation values for vaginal fluid and menstrual blood were reported as significantly higher in DACT1 and USP49, than PFN3 and PRMT2. Methylation SNaPshot reciprocated the results of MSRE-PCR, seeing as ≥ 90% methylation values were noted for USP49, DACT1 and PFN3, in blood and saliva samples. Similar to Lee et al. (2012), higher methylation levels at USP49 and DACT1 were also seen in menstrual blood and vaginal fluid samples of this study. Unlike Lee et al. (2012), both assays of An et al. (2013) displayed complete unmethylation at the PFN3 tDMR in semen.

Choi et al. (2014) carried out MSRE-PCR to identify and differentiate body fluids with tDMRs USP49, DACT1, PFN3 and L81528. As previously reported by An et al. (2013), low methylation levels in vaginal fluid and menstrual blood at PFN3 were observed in this study as well. Amplification of the semen-specific, L81528, marker was found only in 55.56 % non-vasectomized semen samples. Sensitivity analysis determined that the minimum concentration required to generate significant results was 500 pg of DNA (for saliva and semen), and 250 pg of DNA (for vaginal fluid). Amplification of L81528 distinctly differentiated saliva and semen present in mixtures of 1:1 and 1:2 ratios.

The aim of Park et al. (2014) to identify potential markers that displayed body fluid-specific DNA methylation was achieved by identifying candidate markers through genome-wide profiling
thereafter, exclusion of markers associated with aging. The selected eight novel markers in the study were subjected to pyrosequencing analysis and showed hypermethylation in their target body fluids as compared to others. Hypermethylation of all CpG sites of the selected markers in their target body fluids validated their application for body fluid identification. Additionally, the markers exhibited good sensitivity (successful analysis ≥ 10 ng of DNA) and specificity (strongly discriminated body fluids from each other).

Lee et al. (2015) selected CpG sites based on microarray data, for application in body fluid identification. Seven CpG sites displayed body fluid-specific hypermethylation and one site presented semen-specific hypomethylation. The multiplex methylation SNaPshot assay quantified DNA methylation across all CpG sites in a single reaction. Methylated CpG targets were illustrated by blue peaks on the electropherogram, and green peaks indicated the amplification of an unmethylated CpG target (Lee et al., 2015). The specificity of the markers was confirmed by body fluid-specific hypermethylation patterns in seven markers and relatively complete unmethylation in non-target body fluids. The hypomethylated semen-specific marker displayed relatively full methylation patterns in all other body fluids, excluding semen. Majority of the menstrual blood samples were methylated at vaginal fluid and blood-specific markers. However, two menstrual blood samples showed methylation profiles corresponding to vaginal fluid only. Semen samples lacking spermatozoa from vasectomized males were completely methylated at the semen-specific hypomethylated marker and unmethylated in the other semen-specific hypermethylated markers.

Lee et al. (2015) subsequently tested various forensic casework samples to validate the markers’ applicability in forensic laboratories. Analysis of the post-coital samples showed that the assay efficiently identified both semen and vaginal fluid from one, while another only identified vaginal fluid. Lee et al. (2015) suggested that this discrepancy may be due to a difference in methods used for vaginal fluid collection as the incomplete profiled samples were collected on tissue paper. The saliva sample and 75 % of skin samples were correctly identified. The sensitivity test confirmed that as little as 0.5 ng of bisulfite modified DNA would retrieve successful and accurate body fluid-specific methylation profiles.

Vidaki et al. (2016) conducted a study aimed at the identification of novel differentially methylated sites by evaluating previous data reported by Rakyan et al. (2008, 2010), and validation studies of methylation levels by pyrosequencing. A few of the markers selected showed lower methylation levels in their specific body fluids than reported and could not confirm former findings as well as, a large variation in inter-individual methylation patterns was observed. Nevertheless, the methylation ratios of the semen-specific markers (SEU1, SEU2, SEM1 and
SEM2) were confirmed and supported data published by Rakyan et al. (2008). SEU1 and SEU2, specifically, were chosen for further validation studies due to their robustness and high specificity for semen. Validation of the markers were verified by successful amplification with as little as 50 pg of DNA, correct identification in aged stains, and the absence of amplification bias.

The PFN3 tDMR previously reported (Lee et al., 2012) to have shown differential DNA methylation patterns in vaginal fluid as compared to other body fluids, was considered a potential vaginal fluid identification marker for forensic use. Antunes et al. (2016) subjected bisulfite converted DNA samples to pyrosequencing of the PFN3A marker (sub-region of PFN3) – a region previously reported to display the greatest differential methylation patterns between vaginal fluid and other body fluids. Results had shown that nine of the ten CpG sites identified, had an intermediate methylation level which was significantly different in vaginal fluid than blood, saliva, and semen (p-value ≤ 0.05).

Antunes et al. (2016) conducted several sensitivity and specificity tests to determine if the marker could successfully identify vaginal fluid under unfavourable conditions (e.g., degradation, minute quantity and mixtures). As low as 5 ng of DNA displayed significantly differential methylation across all analysed CpG sites and was recommended as the minimal DNA concentration required for reliable results. The species-specificity test indicated that the marker was able to amplify non-primate samples, therefore was not specific to human DNA. These results exhibited that PFN3A had immense potential as a marker for vaginal fluid identification, and to distinguish it from blood, saliva, and semen. In saying that, Antunes et al. (2016) recommended that thorough validation studies of this marker be completed before implementation in forensic procedures.

Lin et al. (2016) selected a total of eight candidate markers, displaying body fluid-specific methylation profiles (two for each body fluid viz. blood, saliva, semen, and vaginal fluid). Preliminary MSP reactions (methylated and unmethylated reaction) were carried out on each individual candidate marker to assess the methylation status at CpG sites. MSP results showed successful amplification of methylation reactions only in the respective body fluid-specific markers. However, traces of venous blood were present at the vaginal fluid-specific methylation markers. A multiplex MSP reaction was performed using only the methylated primer pair for each body fluid-specific marker, followed by multiplex methylation SNaPshot on the PCR products of the 10-plex MSP reaction.

Lin et al. (2016) determined the methylation status of the body fluids using a “methylation indicative value” (MIV) calculated by the formula \[ \frac{M}{M+U} \]. “M” and “U” in the equation...
represent peak heights of methylated and unmethylated signals on the electropherogram. Resultant electropherograms of the multiplex SNaPshot assay showed methylated peaks were only present in the target body fluid. Methylated peaks found at both vaginal fluid and blood-specific markers represented the menstrual blood samples, highlighting that it contained both body fluids. All samples (n=65) analysed were unambiguously identified. To test the applicability and advantage over existing identification assays, the multiplex MSP-SNaPshot assay tested five non-probative forensic samples which had previously reported inconclusive results with catalytic, immunological, and microscopic tests. As expected, all the samples were correctly identified. Furthermore, successful, and accurate results were possible with an input of 0.25 ng of bisulfite converted DNA.

For the identification of blood, Watanabe et al. (2016b) applied a previously developed real-time PCR-based method (Watanabe et al. 2016a) to cg06379435 and its neighbouring CpG sites, reported to display methylation in blood (Lee et al., 2015). The methylation ratios in blood samples were substantially greater than other body fluids, which indicated the successful identification of blood from others (Watanabe et al., 2016b). All aged samples studied were correctly identified as blood, and samples with a DNA amount up to 1 ng were also successful in the identification of blood. To combat the unsuccessful analysis of mixtures in the study, bisulfite sequencing of cg06379435 and its neighbouring single nucleotide polymorphism (SNP), rs7359943, were used for allele-specific blood identification. This preliminary mixture analysis correctly identified blood, indicative by a greater methylation ratio of allele clones from blood DNA.

In order to validate possible markers for human body fluid identification, Holtkötter et al. (2017) evaluated thirteen markers from literature in a simplex SNaPshot reaction. Except one, all other markers confirmed the DNA methylation patterns previously reported in various body fluids. Out of the twelve assessed, four markers which displayed the greatest level of differentiation between target body fluids and others, were selected for the multiplex SNaPshot assay. Successful identification of each body fluid by their respective markers were distinctly shown by blue peaks (indicated hypermethylation of the marker) on the electropherograms, only at their target body fluid. However, the blood-specific marker (BL1) was methylated in blood and semen and was advised to use in conjunction with the semen-specific marker (SE2) to eliminate the possible presence of semen. The specificity, accuracy and robustness of the designed multiplex SNaPshot assay was further validated by the correct identification of body fluids present in mixtures as well as, synthesised stains (single source and mixed samples).
Blood-specific methylation patterns of the embryonal Fyn-associated substrate (EFS) gene was reported by Neumann et al. (2011) i.e., highly methylated in blood but completely unmethylated and partially methylated in semen and buccal cells, respectively. Vidaki et al. (2017) designed a bisulfite pyrosequencing assay on ten CpG sites in EFS to assess its potential as a marker for the identification of blood. Previous results were supported with methylation ratios of blood being higher than semen (almost completely unmethylated) and other tissues of the study (partially methylated). Methylation levels in blood compared to other tissues were found statistically different (p < 0.001) for all CpG sites thus, confirmed this region as a potential marker for the identification of blood. In particular, CpG site four was discovered as the greatest potential for a blood-specific marker by showing the largest difference between blood and other body fluids as well as, the lowest inter-individual variation. This CpG site also showed great discriminatory power between blood and menstrual blood. No significant differences were noted by tests assessing the effect on DNA methylation ratios, by age, ethnicity, and gender. The stability of DNA methylation in the marker was confirmed as blood was successfully identified in aged stains, mixtures, samples stored at different temperatures, and UV-degraded samples (exposure to UV light up to 90 minutes).

McCord and Kuppareddi (2018) reported the identification of ten new loci, each displaying methylation-specific patterns for semen, saliva, blood, and vaginal epithelia. The identified genomic markers were suggested to show great potential for their use in forensic science by successful body fluid identification in all validation tests. Body fluids in a mixture were correctly identified by markers specific to the constituents of the samples, exposure to heat and UV light did not change the DNA methylation pattern, and post-bisulfite treatment steps eliminated PCR inhibitors. Reproducibility of the loci were confirmed, and sensitivity tests indicated that successful amplification was possible down to 0.1 ng of DNA.

SNP genotyping was carried out by a novel in-house developed SNaPshot single base extension (SBE) multiplex assay (Zar et al., 2018). The study focused on analysing fresh (blood) and highly degraded (bone) samples with the use of nine SNPs, located in genes that are routinely used in forensic science for phenotyping. The analysis of the blood samples validated the SNaPshot SBE multiplex system, by determining the minor allele frequency (MAF) across different populations. The sensitivity of the developed assay was indicated by the retrieval of 38 % full STR profiles and 46 % partial profiles, from bone samples (aged 200- to 500- years old) obtained from mass graves. The results of comparing MAFs of each bone sample with MAFs of each population, in the study, elucidated that the skeletal remains were the most closely related (genetically) to the Pathan population.
2.9 Analysis of body fluid mixtures in forensics

Most forensic samples are not found in singularity but rather in the form of mixtures, at the crime scenes. One of the more predominant mixtures is vaginal fluid and semen, usually seen in sexual encounters and sexual assaults. Statistics South Africa found that sexual offence cases reported to the police were 97 938 for the period of 2018/2019 (Governance, Public Safety and Justice Survey). South African Police Service (SAPS) reported 52 420 sexual offences of which 76 % were rape cases (SAPS Crime Statistics). Thus, a reliable, stable, sensitive, and powerful discriminatory assay would contribute greatly to forensic science. The present study aimed to efficiently identify human body fluid samples in mixtures, indicating the sensitivity of the markers.

Frumkin et al. (2011) designed a MSRE-PCR assay for the identification of tissues with simultaneous DNA profiling. This assay was conducted on mixtures containing saliva and semen DNA in varying percentages/ratios. Results showed that the percentage of semen in the mixtures were accurately identified, with at most a 10 % difference. Furthermore, in all mixtures the presence and absence of semen was correctly identified, even a sample which contained only 13 % of semen. This would be especially applicable in sexual offences evidence wherein the semen constituent of samples is often present in minor quantities.

Antunes et al. (2016) facilitated mixture analysis by combining blood, vaginal epithelial and semen, in varying ratios. Each mixture was composed of only two body fluids. Initially, one randomly selected sample from the fore-mentioned three body fluids was quantified to determine their methylation levels at the PFN3A tDMR. These samples were then mixed together according to different ratios, proceeding with DNA methylation analysis. A similar pattern of methylation was observed for mixtures of blood and vaginal fluid, blood, and semen, as well as vaginal fluid and semen. All mixed samples displayed an intermediate methylation level when compared to percentage methylation of singular body fluid. A decrease or increase in percent methylation was reported as the mixture ratios of body fluids decreased or increased, respectively. Antunes et al. (2016) concluded that even though the results were promising, unknown crime scene samples could be incorrectly identified, and suggested that more body fluid-specific markers be analysed and incorporated in future studies.

Mixture studies was one of the aspects carried out by Silva et al. (2016) when analysing the efficiency of identifying semen, saliva, and blood, using ZC3H12D, BCAS4 and cg06379435, respectively. Pyrosequencing enabled the quantification of methylation levels at CpG sites in each body fluid. Mixtures were prepared by combining body fluids in various ratios. Single body fluid
(pure) samples were concurrently run as positive controls for comparison. The presence of a mixture was evident for all markers irrespective of whether the mixture contained two or three body fluids. As compared to singular samples, the methylation levels for mixtures were lower; with the exception of ZC3H12D on pure semen samples, which displayed significantly lower methylation values than mixed samples. The methylation analysis of the study did not provide results which displayed complete identification of constituents in each mixture, however, did indicate the detection of mixtures for sample discrimination.

Choi et al. (2014) mixed saliva and semen samples in 1:1 and 1:2 ratios. These mixtures were accurately identified in the L81528, semen-specific methylation marker. The use of general DNA extraction methods resulted only in the identification of vaginal fluid in semen and vaginal fluid mixtures. Differential extraction methods allowed for the identification of vaginal fluid and semen from the supernatant and pellet, respectively. Post-sexual intercourse samples were collected (one penile sample and four vaginal fluid) to evaluate methylation level differences that would be present for sexual offence cases. Similarly, results differed based on whether DNA was extracted using general or differential methods. Mixed samples extracted by general protocols generated low amplification peaks at PFN3 and L81528 (semen-specific marker). Methods of differential extraction revealed profiles consistent with individual vaginal fluid profiles. Similar to Antunes et al. (2016), this study also recommended more tDMR markers be analysed for the accurate differentiation between body fluids (especially in mixtures).

Forat et al. (2016) conducted a mixture analysis using the methylation-sensitive single nucleotide primer extension-based (SNuPE-based) methylation SNaPshot assay. The method employed a total of nine body fluid-specific markers that displayed methylation/unmethylation signals only in the target body fluid, and unmethylation/methylation, respectively, in the remaining body fluids. Mixtures were prepared by combining the marker-specific body fluid with a mixture that contained equal volumes of the remaining body fluids. All admixed samples were bisulfite converted and subjected to the multiplex SNaPshot assay. Observed and expected methylation values for each marker were plotted on a single graph to determine if overlapping occurred i.e., if observed values matched the expected. Minor variation from theoretical values (20 % greater than expected) were noted for five of the nine markers viz. Blut1 (venous blood), Vag1 and Vag2 (vaginal fluid), Spei2 (saliva), and Sperm2 in semen. Methylation analysis of the mixtures showed that both markers for the identification of venous blood allowed for the discrimination of blood from other body fluids in a mixture. The marker, Mens1, clearly indicated the presence of menstrual blood in a mixture, as it was methylated only in the target body fluid and unmethylated
in the others. The saliva-specific, Spei1, marker identified saliva in mixture samples with saliva quantities greater than 20%. Similar results were found in the markers for vaginal fluid (Vag2) and semen (Sperm2), in mixtures containing more than 20% of the respective target body fluid.

Watanabe et al. (2018) suggested that analysing CpG sites which display body fluid-specific methylation or unmethylation, as well as its neighbouring SNPs, could identify a single body fluid in mixtures. Randomly selected regions which were successfully amplified in preliminary MSP reactions were applied in mixture tests. SE-SNP 1 and SE-SNP4 displayed semen-specific methylation, while SE-SNP2 displayed semen-specific unmethylation. The PCR products were then subjected to pyrosequencing for SNP genotyping. This semen-specific SNP typing protocol was tested on mixtures containing blood and semen DNA, of the same concentration. Results of the blood-semen mixture showed amplification of SE-SNP 1 and SE-SNP4 semen-specific methylation markers, as well as amplification of SE-SNP2 (semen-specific unmethylated marker). Therefore, demonstrated that semen could be successfully amplified and genotyped in mixtures. Pure saliva and semen samples, along with, saliva-semen mixtures in ratios of 1:1 and 1:4, were also evaluated. The presence of saliva in mixed samples was detected by primers designed to target methylated sites in SE-SNP2 (semen-specific unmethylated). SE-SNP1 successfully detected semen in a stand-alone sample and in mixtures, while no peak for the pure saliva sample was visible. The results for SE-SNP2 and SE-SNP4 were in line with SE-SNP1 by correctly identifying semen in the 1:1 mixture, however, inconclusive results were reported in the 1:4 mixture. Watanabe et al. (2018) reasoned that this may have been as a result of SE-SNP1 displaying greater sensitivity than SE-SNP2 and SE-SNP4 but suggested that this could be improved by altering PCR conditions.

Lin et al. (2016) conducted mixture analysis by individually combining bisulfite modified DNA from vaginal fluid, blood, and saliva of one female with semen DNA in ratios of 9:1, 1:1, and 1:9. A mixture of all four body fluids was also evaluated by an equal combination in a 1:1:1:1 ratio. The analysis of these various mixtures by the developed 10-plex MSP-SNaPahot assay (Lin et al., 2016) revealed specific methylation peaks for both vaginal fluid and semen markers. This proved significant to forensic sample analysis in sexual assault cases, as even a mixture containing just 10% of semen was accurately identified. Peak heights for vaginal fluid at the VG-B marker fell below the threshold in vaginal fluid and semen mixtures (1:9), which indicated false-negatives when vaginal fluid was present in lower volumes. However, peak heights above the threshold were illustrated for semen and vaginal fluid when mixed in a 1:1 and 9:1 ratio. Mixtures of blood and semen or saliva and semen for all three ratios had shown peaks that were above the threshold.
at respective body fluid-specific markers. The mixture of all body fluids in equal proportion (1:1:1:1) produced clearly visible methylated peaks in all eight candidate markers of the study.

2.10 DNA methylation analysis techniques

DNA methylation is an integral epigenetic mark responsible for maintaining cellular function (Moore et al., 2013). Spontaneous deamination of methylated cytosines in CpG sites result in the conversion to thymine bases (CpG → TpG) (Cooper et al., 2010). The frequent occurrence of spontaneous deamination explains the low frequency (~1%) of methylated CpG dinucleotides in the genome (Xia et al., 2012). DNMTs play an important role in the maintenance of DNA methylation (Moore et al., 2013), therefore, DNA methylation is removed during PCR and biological cloning due to the absence of DNMTs. For this reason, the plethora of DNA analysis techniques (Table 2.1) requires methylation-dependent treatment prior to hybridisation or amplification. The DNA can be treated by three main procedures: endonuclease digestion, affinity enrichment or bisulfite conversion. (Laird, 2010)

Table 2.1: Different DNA methylation analysis methods

<table>
<thead>
<tr>
<th>DNA methylation analysis method</th>
<th>DNA treatment prior to analysis</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPCE</td>
<td>endonuclease digestion</td>
<td>Fraga et al., 2002</td>
</tr>
<tr>
<td>MSRE-PCR</td>
<td>endonuclease digestion</td>
<td>Singer-Sam et al., 1990</td>
</tr>
<tr>
<td>RLGS</td>
<td>endonuclease digestion</td>
<td>Costello et al., 2000</td>
</tr>
<tr>
<td>DMH</td>
<td>endonuclease digestion</td>
<td>Huang et al., 1999</td>
</tr>
<tr>
<td>CHARM</td>
<td>endonuclease digestion</td>
<td>Irizarry et al., 2008</td>
</tr>
<tr>
<td>HELP</td>
<td>endonuclease digestion</td>
<td>Khulan et al., 2006</td>
</tr>
<tr>
<td>MS-AP-PCR</td>
<td>endonuclease digestion</td>
<td>Gonzalgo et al., 1997</td>
</tr>
<tr>
<td>AIMS</td>
<td>endonuclease digestion</td>
<td>Frigola et al., 2002</td>
</tr>
<tr>
<td>MMAS</td>
<td>endonuclease digestion</td>
<td>Ibrahim et al., 2006</td>
</tr>
<tr>
<td>MethylScope</td>
<td>endonuclease digestion</td>
<td>Ordway et al., 2006</td>
</tr>
<tr>
<td>Method</td>
<td>Technique</td>
<td>Reference</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>mDIP/ MeDIP</td>
<td>affinity enrichment</td>
<td>Weber et al., 2005, 2007</td>
</tr>
<tr>
<td>mCIP/ MeCIP</td>
<td>affinity enrichment</td>
<td>Gebhard et al., 2006</td>
</tr>
<tr>
<td>MIRA</td>
<td>affinity enrichment</td>
<td>Rauch et al., 2006, 2009</td>
</tr>
<tr>
<td>MethyLight</td>
<td>bisulfite conversion</td>
<td>Eads et al., 2000</td>
</tr>
<tr>
<td>Pyrosequencing</td>
<td>bisulfite conversion</td>
<td>Korshunova et al., 2008</td>
</tr>
<tr>
<td>Direct Bisulfite genomic sequencing</td>
<td>bisulfite conversion</td>
<td>Frommer et al., 1992</td>
</tr>
<tr>
<td>Cloned Bisulfite genomic sequencing</td>
<td>bisulfite conversion</td>
<td>Frommer et al., 1992</td>
</tr>
<tr>
<td>MSP</td>
<td>bisulfite conversion</td>
<td>Herman et al., 1996</td>
</tr>
<tr>
<td>MS-SNuPE</td>
<td>bisulfite conversion</td>
<td>Gonzalgo and Jones, 1997</td>
</tr>
<tr>
<td>COBRA</td>
<td>bisulfite conversion</td>
<td>Xiong and Laird, 1997</td>
</tr>
<tr>
<td>RRBS</td>
<td>bisulfite conversion</td>
<td>Meissner et al., 2005</td>
</tr>
<tr>
<td>BSPP</td>
<td>bisulfite conversion</td>
<td>Ball et al., 2009; Deng et al., 2009; Berman et al., 2009</td>
</tr>
<tr>
<td>WGSBS</td>
<td>bisulfite conversion</td>
<td>Cokus et al., 2008; Lister et al., 2008; Lister et al., 2009</td>
</tr>
<tr>
<td>MS-HRM</td>
<td>bisulfite conversion</td>
<td>Wojdacz and Dobrovic, 2007</td>
</tr>
<tr>
<td>MS-SSCA</td>
<td>bisulfite conversion</td>
<td>Bianco et al., 1999</td>
</tr>
<tr>
<td>Methylation SNaPshot</td>
<td>bisulfite conversion</td>
<td>Kaminsky and Petronis, 2009</td>
</tr>
</tbody>
</table>

HPCE, high performance capillary electrophoresis; MSRE-PCR, methylation-sensitive restriction endonuclease polymerase chain reaction; RLGS, restriction landmark genome scanning; DMH, differential methylation hybridisation; CHARM, comprehensive high-throughput arrays for relative methylation; HELP, HpaII tiny fragment enrichment by ligation-mediated PCR; MS-AP-PCR, methylation-sensitive arbitrarily primed PCR; AIMS, amplification of inter-methylated sites; MMASS, microarray-based methylation assessment of single samples; MeDIP and mDIP, methylated DNA immunoprecipitation; MeCIP and mCIP, methyl-CpG immunoprecipitation; MIRA, methylated CpG island recovery assay; MSP, methylation-specific PCR; MS-SNuPE, methylation-sensitive single nucleotide primer extension; COBRA, combined bisulphite restriction analysis; RRBS, reduced representation bisulphite sequencing; BSPP, bisulphite padlock probes; WGSBS, whole-genome shotgun bisulphite sequencing; MS-HRM, Methylation-sensitive high resolution melting; MS-SSCA, methylation-specific single-strand conformation analysis

2.10.1 Affinity enrichment

The affinity enrichment is also known as immunoprecipitation techniques, which involve specific interactions between proteins and methylated DNA (Nair et al., 2011; Vucic et al., 2009). The two main ways to “enrich” the methylated DNA are: a) monoclonal antibodies specific to 5-methylcytidine; and b) precipitating DNA with methyl-CpG binding proteins 2 (MeCP2 and MBD2) (Laird, 2010).
2.10.1.1 Methyl-DNA immunoprecipitation (MeDIP)

Weber et al. (2005) established a DNA methylation analysis based on direct immunoprecipitation of methylated DNA. The assay, MeDIP, purifies methylated DNA by raising monoclonal antibodies against 5mC. The procedure is initiated by brief sonication of genomic DNA in order to produce smaller, random fragments. The shearing of DNA into shorter fragments (ideally between 300 bp - 600 bp) ensured achieving adequate resolution, better downstream immunoprecipitation as well as eliminated fragment-length biases (Brebi-Mieville et al. 2012; Jacinto et al., 2008). The monoclonal antibody (anti-5mC) has greater attraction for single-stranded DNA (created by DNA denaturation at 95°C). Subsequent incubation of the monoclonal anti-5mC and the single-stranded DNA created an antibody-methylated DNA complex (Rana, 2018).

Standard immunoprecipitation steps occur; paramagnetic beads bound to anti-mouse-IgG conjugate to the antibody-methylated DNA complex. After numerous washing steps, DNA which has not been bound (unmethylated 5mC or non-specific DNA fragments) will be discarded in the supernatant. Proteinase K is added to the sample resulting in antibody degradation. The free immunoprecipitated DNA can then be used to study the methylation status of a particular gene/s by MSP, MSRE, sequencing or microarray hybridisation (Borgel et al., 2012). Most studies couple MeDIP with array hybridisation wherein the input and immunocaptured enriched DNA are fluorescently labelled with Cy5 and Cy3 dyes, to evaluate methylation levels (Laird, 2010). The proportion of green to red fluorescence emitted by the dyes indicate the respective hypermethylation or hypomethylation levels (Harrison and Parle-McDermott, 2011). MeDIP-PCR successfully generated methylation profiles from as little as ~1 ng DNA extracted from blood, bone, and hair forensic samples (Rana, 2018; Zhao et al., 2014). One major disadvantage of the analysis technique is that methylation differences cannot be determined at single base-pair resolution (Jeong et al., 2016; Taiwo et al., 2012).

2.10.1.2 Methyl-CpG immunoprecipitation (MCIP)

MCIP follows a similar protocol to MeDIP with the main difference being the protein used to capture the methylated DNA. MCIP employs a recombinant protein (MBD-Fc) composed of the human methyl-CpG-binding-domain protein 2 (MBD2) and the Fc tail of human immunoglobulin G (IgG1) (Thu et al., 2010). This recombinant protein has similar properties to the anti-5mC antibody used in MeDIP, e.g., a high affinity for single-stranded methylated DNA (Gebhard et al., 2006). After genomic DNA sonication and denaturation, single-stranded DNA is incubated with the MBD-Fc protein, which is bound to Sepharose beads. Centrifugation of the beads eliminating unbound DNA is proceeded with repeatedly washing using sodium chloride (NaCl),
increasing the salt concentration in each wash step, to elute DNA fragments containing methylated CpG sites. Elution with differing salt volumes separates the methylated fragments into fractions of increasing methylation quantity; thus, permitting quantification of hypermethylated and hypomethylated DNA (Nair et al., 2011).

A major advantage of MCIp is that it can be effectively applied in locus-specific and genome-wide studies. Higher sensitivity to CpG sites by the protein as compared to the antibody could explain the determination of more differentially methylated genes using MCIp, as opposed to MeDIP (Gebhard et al., 2006). A drawback of this application is the directly proportional relationship between the affinity of the protein with methylated DNA, and the proportion of methylation CpG sites. DNA fragments with an increased number of methylated cytosines have a higher affinity with the protein. DNA with limited methylated CpG sites may be insufficiently bound by the protein therefore, producing incomplete or underrepresented methylation results (Thu et al. 2010).

**Figure 2.3:** Summarized protocol of MeDIP and MCIp. Sheared and denatured DNA fragments are split into input DNA and immunoprecipitated (IP) DNA. Methylated DNA fragments are captured by antibodies (1° Ab) and recombinant proteins (MBD) in MeDIP and MCIp, respectively. Incubation is carried out with secondary antibodies (2° Ab) to bound beads that conjugate with the DNA-antibody or DNA-protein complex. Purification steps of repeatedly washing and digestion of the antibody/protein results in immunoprecipitated DNA. (Thu et al., 2010)
2.10.2 Endonuclease digestion
Methylation analysis can be performed using endonuclease digestion by MSRE. All restriction enzymes which cleave DNA at a specific sequence has a corresponding DNMT, which maintains methylation at the recognition site to protect endogenous DNA (Vasu and Nagaraja, 2013). Most restriction enzymes are only able to cleave DNA at recognition sites which are unmethylated. Methylation present at recognition sites prohibits digestion by restriction enzymes. These features of the enzymes allow for the generation of differential hypomethylated and hypermethylation profiles (Laird, 2010). Although, certain exceptions do exist such as, methylation insensitive enzymes which are able to cleave methylated DNA (Dahl and Guldberg, 2003).

2.10.2.1 Restriction landmark genomic scanning (RLGS)
RLGS is an example of a genome-scale methylation analysis technique, detecting large sets of unselected genome-wide CGIs in simply one gel (Costello et al. 2000). Digestion of genomic DNA is brought about by MSRE, such as NotI (restriction site: 5′-GC↓GGCCGC-3′). The recognition sites are often found in CGIs and cut the DNA at unmethylated CpG dinucleotides. The restricted sites of the DNA are radioactively labelled with phosphorous isotopes [α-32P]dCTP and [α-32P]dGTP, catalysed by DNA polymerase. The labelled DNA is further digested with another restriction enzyme (e.g., EcoRV) and subjected to first-dimensional electrophoresis on an agarose gel. Hinfl digestion buffer is used to equilibrate the gel, and the DNA digested in the gel with Hinfl. After the second-dimensional electrophoresis separation, the gel is exposed to a large X-ray film which eventually develops into an autoradiograph (Harrison and Parle-McDermott, 2011).

Every spot on the autoradiograph depicts a labelled DNA restriction landmark. Missing spots denote a methylated recognition site for NotI; given that the site will not be cleaved nor radioactively labelled, therefore absent. RLGS with subsequent sequencing or PCR determines the methylation status of RGLS profiles. Methylation differences in profiles generated have been especially significant in the identification of imprinted and diseasing-causing genes. This method does, however, present a fair number of disadvantages; it is labour-intensive, even slightly degraded DNA reduces the quality of the profile, large amounts of DNA is initially required, time-consuming and no reports on whether the CGI is present in a promoter (Dahl and Guldberg, 2003). Further drawbacks include the requirement of highly specific instrumentation and interpretation difficulty (Georgiev and Pavlov, 2018).
2.10.2.2 *HpaII* tiny fragment enrichment by ligation-mediated PCR (HELP)

The HELP assay is also applicable in determining gene-specific and genome-wide methylation patterns of different tissues and cells (Oda and Greally, 2009). HELP utilizes a pair of isoschizomer restriction enzymes, *HpaII* (methylation sensitive) and *MspI* (methylation insensitive). Despite being structurally different, *HpaII* and *MspI* have the same recognition site and cleavage position (Laird, 2010). The digested genomic DNA is subjected to a ligation-mediated PCR and hybridised to a custom microarray with independent fluorochromes (Khulan et al., 2006). Unmethylated restriction sites (5′-CCGG-3′) will be cleaved by *HpaII* thus, providing enrichment in regions with lower methylation levels. Comparing *HpaII* and *MspI* results determine the methylation at each locus point, i.e., whether it is unmethylated or methylated (Shaknovich et al., 2010).

2.10.3 Bisulfite conversion

Modification of DNA with sodium bisulfite is a vital precursor step in DNA methylation analysis, as this conversion permits the differentiation and detection of unmethylated and methylated cytosine within the DNA sequence (Birkner et al., 2009). Without bisulfite treatment, downstream amplification will remove the methylation marks (Madi et al., 2012) thus, disabling DNA methylation identification. Treating single-stranded DNA molecules with sodium bisulfite drives this mechanism by bringing about deamination of unmethylated cytosines to uracil while methylated cytosine residues remain protected and unchanged (Frommer et al., 1992; Yu and Snyder, 2016). Post PCR, uracil is amplified to thymine and methylated cytosines are amplified to cytosines (Figure 2.4). Bisulfite conversion results in single nucleotide resolution of methylated cytosines thus, indicating the methylation status of all cytosines in a specific gene or the entire genomic region (Hernández et al., 2013). Methylation patterns of the gene studied can be elucidated by comparing the modified sequence to reference sequences.
Figure 2. 4: Mechanism of bisulfite conversion. Upon bisulfite treatment of DNA, methylated cytosines remain intact while unmethylated cytosines are deaminated to uracil. Subsequent amplification results in uracil amplified to thymine, and methylated cytosines amplified to cytosine. Sequencing will provide the methylation status of each cytosine on the DNA. (Yu and Snyder, 2016)

Bisulfite treatment coupled with various available techniques has shown to be successful in determining the percent methylation at every CpG site, as well as overall methylation status (An et al., 2013; Lee et al., 2012, 2016b; Lin et al., 2016). This method determines DNA methylation in real-time – globally and locally (Delaney et al., 2015; Madi et al., 2012). CpG dinucleotides are predominantly unmethylated in mammals therefore, higher levels of observed DNA methylation may indicate incomplete bisulfite conversion of unmethylated CpGs (Bock, 2012; Ziller et al., 2017). Although quite rarely occurring, over-conversion of methylated cytosines could also present false results when analysing DNA methylation. This error is extremely difficult to differentiate from a methylation error by (Bock, 2012). The present study will focus on methylation analysis techniques that involve this chemical modification.

2.10.3.1 Methylation-specific polymerase chain reaction (MSP)

MSP is a qualitative method developed by Herman et al. (1996) that assesses the DNA methylation status of CGIs in the human genome. It is an application of the bisulfite sequencing method excluding the cloning and sequencing steps, therefore less time-consuming (Lizardi et al., 2016). This method involves designing primers to distinguish methylated from unmethylated CpGs in bisulfite converted DNA (Figure 2.5). Sodium bisulfite deaminates unmethylated cytosines to uracil while methylated cytosine residues remain unchanged; uracil and methylated cytosine are subsequently amplified by the PCR reaction to thymine and cytosine, respectively. A primer set is designed which is complementary to the sequence with methylated CpGs, but not to the initial sequence of DNA with unmethylated CpGs thus, resulting in the amplification of only CpG sites that display methylation. The same would be applicable for primer pairs designed to amplify only unmethylated CpGs (Hernández et al., 2013). The specific methylated and unmethylated primer pair is generally used for the analysis of methylation in the same gene; therefore, amplified PCR products for both reactions can run alongside one another in electrophoresis for comparative purposes (Figure 2.5). This gene-specific methylation screening method is advantageous as it requires significantly less amounts of DNA, interpretation of results is simple, elimination of false positives, and the method is sensitive to 0.1 % methylated alleles of a given locus (Häfner et al., 2011).
Figure 2.5: Representation of the MSP procedure. Isolated DNA is bisulfite-treated, converting unmethylated cytosines to uracil while methylated cytosines remain intact. Primer pairs specific for methylation and unmethylation are applied to PCR amplification. These primers will only amplify the sequence they are specific to. DNA methylation state of the sequence can be detected by simple gel electrophoresis. Completely unmethylated samples will produce a clear band in the unmethylated (U) reactions, and fully methylated samples show a distinct band in the methylated (M) reaction. Heterogenous tissue samples will produce amplification in both methylated and unmethylated reaction. MSP results can be quantified by methylation specific Polymerase chain reaction in situ hybridisation (MSP-ISH) or quantitative methylation specific Polymerase chain reaction (qMSP), with the inclusion of fluorescently labelled probes (*) (Derks et al., 2004).

2.10.3.2 Methylation SNaPshot

Methylation SNaPshot is the second DNA methylation-based assay that will be conducted in this study. Unlike MSP, this is a quantitative method which investigates DNA methylation levels by the calculation of cytosine/thymine (C/T) ratios (Kaminsky and Petronis, 2009). The reaction is based on repeated hybridisation of an oligonucleotide probe (primer) exactly one base pair upstream of a target CpG, with the primer’s 3’ end directly adjacent to the nucleotide base to be identified. The primer is enzymatically extended a single base with a fluorescently labelled nucleotide terminator (2’,3’-dideoxynucleotide triphosphate, ddNTP) (Figure 2.6) (Podini and Vallone, 2009). The fluorescent labelling allows measurable signals to be detected. The samples are subsequently separated via capillary electrophoresis, and the proportion of fluorescent signal
is determined by analysing the electropherogram output by an ABI Genetic Analyzer (Kaminsky and Petronis, 2009; Lin et al., 2016; Rana, 2018).

Methylated and unmethylated CpG sites are present as blue and green peaks, respectively, when primers are designed based on the reverse strand. Black and red peaks are generated by amplification of methylated and unmethylated sites, respectively, when primers are designed based on the forward strand (Forat et al., 2016). Unwanted bias of results can be eliminated by ensuring that the primers anneal to regions containing a single CpG site; though, this is especially difficult in gene regions which are GC-rich. This method is rapid, accurate, feasible, results are easily reproduced, and multiplexing enables large-scale methylation screening in a single reaction (Gomma et al., 2017). Furthermore, this method is quantitative, thereby eradicates the need for restriction enzymes, and is efficient with even significantly low volumes of DNA (Dhingra et al., 2014).

**Figure 2.6:** Diagrammatic representation of multiplex SNaPshot reaction by SBE, and resultant methylation signals. Primers anneal one base pair upstream of the targeted CpG (SNP) site and are extended on the 3’ end by a fluorescently labelled ddNTP, complementary to the DNA sequence. After purification with SAP (Shrimp Alkaline Phosphatase), SNaPshot PCR products are separated via capillary electrophoresis according to size (nucleotides). Single peaks represent a homozygous SNP, and heterozygous SNPs are represented by the presence of two peaks at the same position. (Podini and Vallone, 2009)
2.11 TDMRs/tDMSs of interest studied in the present research

2.11.1 ZNF282 gene (Zinc finger protein 282)

The ZNF282 gene is also known as HUB1. It is a protein-coding gene that contains eight exons and is located on chromosome seven on the long arm (q) on band three, sub-band six, sub-sub-band one (7q36.1) (Figure 2.7) (https://www.ncbi.nlm.nih.gov/gene/8427; ZNF282). The genomic size of the ZNF282 gene is 30,787 bases, and the protein size is 671 amino acids with a molecular mass of 74,295 daltons (Da) (http://www.genecards.org/cgi-bin/carddisp.pl?gene=znf282; Fishilevich et al., 2016). The proteins encoded by the ZNF282 gene belong to the krueppel C2H2-type zinc finger protein family (https://www.uniprot.org/uniprot/Q9UDV7; Breuza et al., 2016). There are four transcripts of ZNF282, of which two are protein-coding. (http://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000170265;r=7:149195546-149226238; Aken et al., 2016). A repressive effect of ZNF282 on human T-cell leukemia virus type 1 (HTLV-1) was reported (Okumura et al. 1997) by U5 repressive element of its long terminal repeat (LTR). ZNF272 recognized the sequence 5’-TCCACCCC-3’ as a core motif and exert a repressive effect on the LTR of HTLV-1. Yeo et al. (2014) reported the overexpression of ZNF282 in oesophageal squamous cell carcinoma (ESCC), and depletion of ZNF282 reduced tumorigenicity of ESCC. It was further reported that ZNF282 was a co-activator in E2F1-mediated expression resulting in ESCC. According to the Genotype-Tissue Expression Consortium (GTEx), this gene is highly expressed in the prostate showing 25.5 pTPM (protein-coding transcripts per million), based on RNA expression. In the vagina and cervix, the GTEx showed 15.8 pTPM and 22.4 pTPM, respectively, for RNA expression. In the salivary glands, the GTEx showed 14.2 pTPM for RNA expression, and no data was available for peripheral blood. (https://www.proteinatlas.org/ENSG00000170265-ZNF282/tissue; Uhlén et al., 2015). Based on the tissue-specific (Tissue-specific Gene Expression and Regulation) gene expression profiles of the TiGER database, ZNF282 is highly expressed in the cervix (http://bioinfo.wilmer.jhu.edu/tiger/db_gene/ZNF282- index.html; Liu et al., 2008).

![Figure 2.7: Chromosomal location of ZNF282 gene](http://www.genecards.org/cgi-bin/carddisp.pl?gene=znf282; Fishilevich et al., 2016)
2.11.2 HPCAL1 gene (Hippocalcin-like 1)

The HPCAL1 gene is also known as calcium-binding protein (BDR1), hippocalcin-like protein 1 (HLP2) and visinin-like protein 3 (VILIP-3). It is a protein-coding gene that contains seventeen exons and is located on chromosome two on the short arm (p) on band two, sub-band five, sub-sub-band one (2p25.1) (Figure 2.8) (https://www.ncbi.nlm.nih.gov/gene/?term=hpcal1; HPCAL1). The genomic size of the HPCAL1 gene is 124 729 bases and has a protein size of 193 amino acids with a molecular mass of 22 313 Da (http://www.genecards.org/Search/Keyword?Query String=hpcal1; Fishilevich et al., 2016). There are eight transcript variants of the gene (https://www.ensembl.org/Homosapiens/Gene/Summary?db=core;g=ENSG00000115756;r=2:1 0302889-10427617; Aken et al., 2016). The proteins that are encoded by the HPCAL1 gene belong to the neuron-specific calcium-binding proteins family (recoverin) found in the brain and retina and are involved in the visual cycle of retinal rods. There is a possibility that the HPCAL1 gene is involved in calcium-dependent regulation of rhodopsin phosphorylation and calcium-ion binding (http://www.uniprot.org/uniprot/? query=hpcal1&sort=score; Breuza et al., 2016). The expression of HPCAL1 was seen predominantly in the cerebellum (specifically cerebellar Purkinje cells) at late developmental stages (Spilker et al., 2000). According to the GTEx, this gene is highly expressed in the cerebellum showing 576.4 pTPM, based on the RNA expression. In the vagina and cervix, the GTEx showed 29.6 pTPM and 67.4 pTPM, respectively for RNA expression. In the prostate and salivary glands, the GTEx showed 27.8 pTPM and 20.9 pTPM, respectively, for RNA expression ((https://www.proteinatlas.org/ENSG00000115756-HPCAL1/tissue; Uhlén et al., 2015). There was no available data on peripheral blood. Based on the tissue-specific gene expression profiles of the TiGER database, HPCAL1 is highly expressed in the cervix (http://bioinfo.wilmer.jhu.edu/ tiger/db_gene/HPCAL1-index.html; Liu et al., 2008).

Figure 2.8: Chromosomal location of HPCAL1 gene (http://www.genecards.org/Search/Keyword?Query String=hpcal1; Fishilevich et al., 2016).
The cg08792630 tDMS was reported by Park et al. (2014) displaying blood-specific hypermethylation. The CpG site studied is located on chromosome six at nucleotide position 10883910 and present in the FOXO3 gene, also known as forhead box O3. It is a protein-coding gene that contains nine exons and is located on chromosome six on the long arm (q) on band two, sub-band one (6q21) (Figure 2.9). The genomic size of the FOXO3 gene is 124952 bases and has a protein size of 673 amino acids with a molecular mass of 71277 Da (https://www.genecards.org/cgi-bin/carddisp.pl?gene=FOXO3&keywords=FOXO3; Fishilevich et al., 2016). The proteins encoded by the FOXO3 gene belong to the forkhead family. There are three alternatively spliced transcript variants that have been reported which encode unique proteins (http://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000118689;r=6:108559835-108684774; Aken et al., 2016). Proteins encoded by the gene have a DNA-binding transcription factor activity, wherein it recognizes and binds to the DNA sequence 5'-[AG]TAAA[TC]A-3' as well as, involved in regulation of different processes e.g., apoptosis and autophagy (Brunet et al., 1999; Lehtinen et al., 2006). The gene also functions in the positive regulation of autophagy in skeletal muscle. In starved cells, the proteins encoded by FOXO3 enter the nucleus after dephosphorylation and bind to the promoters of autophagy genes. Thus, enabling their expression and results in proteolysis of skeletal muscle proteins. Gao et al. (2018) found that human gastric adenocarcinoma cell growth was inhibited by FOXO3 promotion of autophagy in an acidic environment. According to the GTEx, this gene is highly expressed in the cerebellum showing 56.0 pTPM, based on the RNA expression. In the vagina and cervix, the GTEx showed 35.7 pTPM and 40.0 pTPM, respectively for RNA expression. In the prostate and salivary glands, the GTEx showed 25.1 pTPM and 29.8 pTPM, respectively for RNA expression (https://www.proteinatlas.org/ENSG00000118689-FOXO3; Uhlén et al., 2015). Overexpression of this gene is present in bone marrow mesenchymal stem cell (27.6) and cervix (21.2) (https://www.genecards.org/cgi-bin/carddisp.pl?gene=FOXO3&keywords=FOXO3; Fishilevich et al., 2016). Based on the tissue-specific gene expression profiles of the TiGER database, FOXO3 is highly expressed in the spleen (http://bioinfo.wilmer.jhu.edu/tiger/db_gene/FOXO3-index.html; Liu et al., 2008).
2.11.4 cg09765089-231d tDMS

The cg09765089-231d tDMS has been reported to display vaginal fluid-specific hypermethylation in several body fluid identification studies (Lee et al., 2015, 2016b; Lin et al., 2016). The CpG site studied is located at nucleotide 27291577 on chromosome seven on the short arm (p) on band one, sub-band five, sub-sub-band two (7p15.2) (Figure 2.10). This site is not present in any genes on the human genome.

2.12 Rationale of the study

The analysis of crime scene evidence in forensic science provides crucial information in criminal investigations, such as possible suspects and the nature of the crime. Identification of body fluids
reveal important aspects which aid in reconstructing events of the crime (Vidaki et al., 2013). Previously applied conventional, presumptive, and confirmatory methods for body fluid identification utilise catalytic, enzymatic, and immunological-based techniques. Till date, none of the body fluid identification methods are DNA based. Although conventional methods are still used and are sensitive, they lack specificity, are only applicable to specific biological fluids and require large amounts of starting material, which generally isn’t possible in forensic casework (Kader, 2015). Due to the fore mentioned limitations it would be worthwhile to develop DNA-based techniques to complement conventional methods of body fluid identification (Kader and Ghai, 2015). TDMRs have demonstrated varying methylation levels according to specific tissues and cells. The application of this feature in DNA methylation analysis allows for the identification and differentiation of body fluids (Lee et al., 2012, 2015; Madi et al., 2012).

At the University of KwaZulu-Natal, Kader (2015, MSc Genetics dissertation) identified novel candidate gene-based tDMR markers to identify and differentiate blood, saliva, semen, and vaginal fluid, based on their body fluid-specific methylation patterns. Using these results, Naidoo (2017, MSc Genetics dissertation) carried out validation studies of the candidate markers. The present study aims to develop a multiplex MSP-SNaPshot assay for efficient identification of human body fluids using markers discovered in our laboratory (ZNF282 for semen and HPCAL1 for saliva), and markers reported in literature (blood and vaginal fluid). To assess the sensitivity of the markers, the developed MSP-SNaPshot assay was also tested for the identification of body fluids in mixtures. To the best of my knowledge, these markers have not been multiplexed into a SNaPshot assay before.

2.13 Hypothesis

- It can be hypothesised that the selected tDMRs markers evaluated by an MSP assay, will be differentially methylated in human blood, saliva, semen, and vaginal fluid.

- It can be hypothesised that the developed multiplex MSP-SNaPshot assay will successfully allow for the simultaneous identification all body fluids (blood, saliva, semen, and vaginal fluid).

- It can be hypothesised that the validated tDMRs markers will allow specific identification of all body fluids in a mixture.
2.14 Aims

- Design and optimisation of an MSP reaction for profiling of tissue specific differential regions/sites, specific to blood, saliva, semen, and vaginal fluid.
- To design a multiplex methylation SNaPshot assay for simultaneous identification of blood, saliva, semen, and vaginal fluid.
- To test the sensitivity of multiplex MSP-SNaPshot assay for identification of body fluids in a mixture.

2.15 Objectives

- To design MSP primers for identification of blood and vaginal fluid.
- Methylation profiling of candidate tDMRs in blood, saliva, semen, and vaginal fluid by MSP.
- Design SNaPshot primers targeting differentially methylated CpG sites identified following MSP.
- Optimization of multiplex methylation SNaPshot assay.
- Application of the multiplex MSP-SNaPshot system for mixture analysis, to test the applicability of the system and ensure the stability and sensitivity of the marker under forensic conditions.

2.16 Key Questions

- Will the MSP-SNaPshot assay be able to effectively identify and differentiate the four body fluids (blood, saliva, semen, and vaginal fluid)?
- Is the methylation profile of identified tDMRs markers sensitive with sample mixtures?
CHAPTER 3

Development of methylation-specific PCR (MSP) assay for human body fluid identification
Abstract

Body fluid identification is the primary crucial step in a criminal investigation. The application of DNA methylation analysis techniques in identification and differentiation of body fluids has emerged as a robust tool for forensic science. TDMRs/iDMRs are chromosomal segments/locations which display different methylation patterns according to tissue/cell type; a characteristic which enables their use in body fluid identification. The present study aimed to design primers specific to blood, saliva, semen and vaginal fluid, as well as develop an MSP assay for body fluid identification. Primers for saliva and semen targeted CpG islands in genes ZNF282 and HPCAL1, respectively. Primers for blood and vaginal fluid flanked a single CpG site: cg08792630 and cg09765089-231d, respectively. A total of 20 samples per body fluid (n=80) were tested with each methylated and unmethylated primer set. ZNF282 primer displayed hypomethylation in semen, and complete methylation in other three body fluids. Complete unmethylation was also observed in saliva with HPCAL1 primer, and methylation was seen in the other three body fluids. Vaginal fluid primer exhibited hypermethylation only in vaginal fluid and unmethylation in the remaining body fluids. The blood primer displayed hypomethylation in saliva, semen, and vaginal fluid, however exhibited a methylation and unmethylation pattern in blood. Hence, the designed primers, except the blood primer could efficiently distinguish the target body fluid from the rest. To the best of my knowledge, the primers designed in this study are novel and have not been reported before. The MSP amplicons will be targeted to develop a methylation SNaPshot assay in the next chapter to allow multiplexing of markers for body fluid identification and mixture analysis.
3.1 Introduction

Identification of body fluid samples which are retrieved from crime scenes is a major backbone of forensic science. This process enables crime scene reconstruction by linking samples obtained from suspects and victims to actual crime scene samples - an indispensable tool to the judicial system of a country (Virkler and Lednev, 2009). DNA profiling is the most prevalent method for individual identification therefore, crime scene samples require analysis techniques which preserve the quality and quantity of crime scene evidence, without destroying or diminishing its integrity (Kader and Ghai, 2015). Crime scenes are saturated with substances which could be mistaken for body fluids; therefore, it is imperative that these substances/compounds are distinguished from body fluids at the crime scene.

Earlier developed presumptive and confirmatory body fluid identification tests are based on catalytic, enzymatic and immunochromatographic techniques (Gršković et al., 2013; Frumkin et al., 2011). These tests have displayed successful results for several decades, however, with new technologies and techniques constantly emerging, their disadvantages have become more apparent. The time-consuming protocols, applicability to only specific body fluids, a lack of specificity, sensitivity, and stability, are just a few of the limitations with these techniques (Lee et al., 2016b). One of the major disadvantages, is the requirement of large amounts of DNA which creates an obstacle seeing as, minimal volumes of DNA are normally encountered in forensic casework (Kader, 2015).

Despite the lack of stability in RNA (An et al., 2012; Choi et al., 2014), mRNA and miRNA have been utilised in body fluid identification methods (Hanson et al., 2009; Zubakov et al., 2010). The application of miRNA and mRNA prevents the consumption of DNA extracted. MiRNA not only has a longer half-life and better stability than mRNA, but also is protected against \textit{in vivo} degradation, and has high levels of expression (Sijen, 2015). However, these biological molecules display background signals and variation in expression, entail further processes, and require immense volumes of the initial sample (Sijen, 2015; Zubakov et al., 2010). In order to overcome the shortcomings of the previously applied body fluid identification tests, DNA methylation-based methods are being continuously developed and improved on for body fluid identification. DNA methylation-based methods have numerous advantages over conventionally used techniques, such as greater efficiency, higher specificity, greater convenience, and the ability to analyse multiple tissues/body fluids in a single assay (Rana, 2018; Richards et al., 2018). One of the most important advantages, is the successful analysis of crime scene evidence by using very little quantity of the sample. Considering that degraded samples found in small quantities is
relatively common in forensic casework, the sensitivity of these methods are extremely beneficial (Kader and Ghai, 2015).

DNA methylation (covalent attachment of a methyl group to the C5 position of a cytosine) is present in ~1% of the entire human genome (Tost, 2010). The genome consists of DNA methylation hotspots, known as, CpG dinucleotides. An average of 60% - 90% of approximately 30 million CpG dinucleotides, exist in a methylated state (Tost, 2010). CGIs, clustered unmethylated CpG regions, are defined by regions with a length of > 200 bp, GC content > 50%, as well as > 0.6 ratio of observed to expected (Gardiner-Garden and Frommer, 1987; Kakumani et al., 2012). Roughly 50% of CGIs in the genome are found in the vicinity of gene promoter regions (Tammen et al., 2013). An inversely proportional relationship has been reported, between gene expression and the methylation state of CGIs in promoters – reduced/non-expression of genes have been observed when promoter-located CGIs exhibit increased methylation levels (Zemach et al., 2010). In contrast to this, the DNA methylation paradox explains a positive correlation between gene body DNA methylation and gene expression – overexpressed genes display increased DNA methylation levels in gene bodies (Aran et al., 2011; Lister et al., 2009).

TDMRs/tDMSs have been reported by many genome-wide studies (Lee et al., 2012; Lin et al., 2016). The margins of CGIs have shown to be rich in tDMRs, however, tDMRs are also present in non-CGIs (Song et al., 2009) and CpG shores (Irizarry et al., 2009). A lower GC and CpG content have been noted in tDMRs than its surrounding regions. TDMRs/tDMSs display different methylation patterns (hypo- and hypermethylation patterns) according to tissue or cell type (Lee et al., 2016a). This feature of tDMRs permits its application as biomarkers in order to differentiate between human body fluids, based on differential methylation patterns (Ghosh et al., 2010). TDMSSs, similar to tDMRs, are specific CpG sites present on the human genome which displays a differential methylation status in different tissues/cells; therefore, can also be employed in the identification of body fluids. The human genome project prompted the study of the epigenome, as well as identification of tDMRs by bisulfite sequencing (Eckhardt et al. 2006), RLGS (Ghosh et al., 2010), and an algorithm-based search on 450K methylation data (Slicker et al., 2013). Innumerable studies have used tDMRs or tDMSs to identify and differentiate between forensically relevant body fluids (An et al., 2013; Antunes et al., 2016; Choi et al., 2014; Frumkin et al., 2011; Lee et al., 2015; Lin et al., 2016; Madi et al., 2012).

The methylation level of DNA at tDMRs/tDMSs can be estimated and analysed by several techniques. These techniques are broadly divided into two categories: those which do not require bisulfite conversion (functioning on the principle of endonuclease digestion and affinity enrichment) and those which require bisulfite conversion (e.g., MSP and methylation SNaPshot).
DNA methylation analysis by sodium bisulfite conversion is an extremely common method to determine methylated and unmethylated cytosines (Birkner et al., 2009; Herman et al., 1996). The treatment of DNA with sodium bisulfite results in deamination of unmethylated cytosines to uracil, while methylated cytosines remain protected and unchanged (Sasaki et al., 2003; Yu and Snyder, 2016). After PCR amplification, uracils are amplified to thymine and methylated cytosines are amplified to cytosines. Evaluating the C/T ratio determines the methylation status of the sample.

MSP is a qualitative DNA methylation analysis method developed by Herman et al. (1996). Primers are designed to differentiate between methylated and unmethylated CpG sites in bisulfite treated DNA. One primer pair is complementary to the sequence in a methylated state, and not to the sequence of DNA with unmethylated CpGs therefore, the amplification of only CpG sites that display methylation will take place. The same principle is applicable for primer pairs designed to amplify only unmethylated CpGs (Hernández et al., 2013). This gene-specific methylation analysis is advantageous over traditional bisulfite sequencing as it does not include the cloning procedure, resulting in a less time-consuming technique. Furthermore, it requires significantly less amounts of DNA, is cost effective, results are simple to interpret, and is extremely sensitive (a single methylated allele can be detected among countless unmethylated alleles) (Häfner et al., 2011).

Differential DNA methylation profiles exhibited by tDMRs, and the function of tDMRs in body fluid identification, is gaining rapid attention from the scientific community. Thus, increasing the need for the discovery of additional methylation-specific markers for human body fluid identification. At the University of KwaZulu-Natal (Westville), Naidoo (2017, MSc Genetics dissertation) found that a CGI region on the ZNF282 gene displayed semen-specific hypomethylation, while the other forensically relevant body fluids were methylated; therefore, this tDMR was able to differentiate between semen and all other body fluids of the study. Naidoo (2017) also discovered that a CGI region on HPCAL1 had shown saliva-specific hypomethylation patterns, and methylation for the rest of the body fluids.

The same CGIs reported by Naidoo (2017) were targeted in the present study to design MSP primers. Two additional markers were applied from literature; a blood-specific CpG site on the FOXO3 gene, cg08792630 (Park et al., 2014) and a vaginal fluid-specific CpG site, cg09765089-231d (Lee et al., 2015). CpG sites for blood and vaginal fluid were selected from previous reports as they displayed complete hypermethylation in the target body fluid (blood/vaginal fluid) and unmethylation in other three body fluids. The markers chosen for interrogation were selected only if they previously showed significant differences in DNA methylation between the target body
fluids and the other tested body fluids. MSP primers for blood and vaginal fluid were designed flanking the above-mentioned CpG sites. Hence, in the present study, four MSP primers were designed to develop an MSP assay for identification and differentiation of blood, saliva, semen, and vaginal fluid. To the best of my knowledge, the designed markers have not been employed for body fluid identification by the MSP methylation analysis technique before. Thereby, indicating that the developed assay would be extremely beneficial to the science and forensic communities.
3.2 Materials and methods

3.2.1 Ethics approval

The study was conducted according to the methods specified by the Biomedical Research Ethics Committee (BREC) at the University of KwaZulu-Natal. The BREC reference number is BE115/19 (sub-study of BE434/17) (Appendix A). Blood, saliva, semen, and vaginal fluid samples were collected from volunteers after a written consent was signed and a questionnaire answered by each participant (Appendix A).

3.2.2 Sample collection and storage

A total of 80 samples (20 blood, 20 saliva, 20 semen and 20 vaginal fluid) were collected from healthy participants (male and female), 18 years – 60 years, after informed consent. It was not necessary for each participant to provide all three body fluid samples. No other variable like smoking status, sex, blood transfusion, intake of chronic medication etc., was considered in the study. Blood, saliva (buccal cells), and vaginal fluid were collected from volunteers at the King Dinuzulu Hospital in Durban. Blood was collected as dried blood spots using the Dried Blood Spot (DBS) kit (Lasec), while saliva and vaginal fluid were collected on sterile cotton swabs. Freshly ejaculated semen was collected in plastic cups from volunteers at the Dr. Aevitas Fertility Clinic in Cape Town and transported to Laboratory 1 (Genetics Department, University of KwaZulu-Natal, Westville) in ice boxes with frozen ice packs. Saliva and vaginal fluid samples which were deposited on sterile cotton swabs were initially air dried at room temperature prior to DNA extraction procedures. The dried blood spot cards also air dried at room temperature for three hours before DNA extraction was conducted.

3.2.3 DNA extraction and quantification

DNA was extracted from each sample of blood, saliva, semen, and vaginal fluid using the Invitrogen™ PureLink™ Genomic DNA Mini Kit (Thermo Fisher Scientific Inc, Waltham, MA, USA) according to manufacturer’s instructions. Extracted DNA was eluted in 50 μL nuclease free water. The DNA extracted was quantified using a NanoDrop (NanoDrop™ 2000 Spectrophotometer, Waltham, Massachusetts, United States). The concentration of blood, saliva, semen, and vaginal fluid DNA of each volunteer is listed in Table 1, Appendix B. All DNA samples were stored at -20 ºC after quantification. DNA concentration of the samples was adjusted to the optimal DNA input for bisulfite treatment of 200 ng to 500 ng.
3.2.4 Bisulfite conversion

In the absence of treatment with sodium bisulfite, methylation marks are removed from DNA by downstream PCR amplification; thereby, preventing the differentiation between methylated and unmethylated CpG sites (Madi et al., 2012). Bisulfite conversion of isolated DNA enables the preservation of methylated cytosines while deaminating unmethylated cytosines to uracil (Hernández et al., 2013). Bisulfite conversion was carried out by using the EpiJET Bisulfite Conversion Kit (Thermo Fisher Scientific Inc, Waltham, MA, USA). This method replaces the converted uracil to thymine during the PCR amplification process. The bisulfite modified DNA was amplified by site-specific PCR to amplify the bisulfite modified regions.

3.2.5 Selection of candidate genes and MSP primer design

In the genes ZNF282 and HPCAL1, CpG sites were previously identified as being differentially methylated in human body fluids by bisulphite sequencing at the University of KwaZulu-Natal (Genetics department) by Naidoo (2017, MSc Genetics dissertation, UKZN). In the present study, MSP primers were designed to target the same CpG sites reported by Naidoo (2017). The previously reported blood-specific marker, cg08792630 (Park et al., 2014) and vaginal fluid-specific marker, cg09765089-231d (Lee et al., 2015) target a single CpG site, which displayed differential methylation patterns between body fluids.

This study used the genomic regions flanking the CpG sites reported in literature to design the blood and vaginal fluid marker (Table 3.1). Primers targeting CpG sites of genes (ZNF282 and HPCAL1) and a single CpG site in cg09765089-231d, were designed using the MethPrimer program (Li and Dahiya, 2002) (https://www.urogene.org/methprimer/). MSP primers targeting cg08792630 and its neighbouring CpG sites, were manually designed, by following the guidelines of Davidović et al. (2014) and Li and Dahiya (2002). The amplicon sizes and annealing temperatures of the MSP primers are presented in Table 3.2. The number of CpG sites analysed by the MSP primer set is shown in Figure 3.1b for ZNF282 tDMR, Figure 3.2b for HPCAL1 tDMR, and Figure 3.3a for cg09765089-231d tDMS. The chromosomal location of the target CGI within ZNF282 and HPCAL1 is shown in Figure 3.1a (ZNF282) and Figure 3.2a (HPCAL1). The CpG site targeted by the cg08792630 blood-specific marker and cg09765089-231d vaginal fluid-specific marker is not present in a CGI.

The target region amplified by the primers will be referred to as tDMR/tDMS or markers in the rest of the chapter, to provide easier understanding for the reader.
Figure 3.1: Design of MSP primer targeting CGI of ZNF282 gene. a) UCSC genome browser view of chromosomal location of the target CGI within the gene indicated by the green bar, and the position of the CGI in relation to the introns and exons indicated by the blue horizontal line and box. b) MSP: Visual representation of the location of the primers designed within the CGI (blue coloured area). Two primer sets are indicated by the purple boxes (MF- methylated forward primer and MR- methylated reverse primer) and green boxes (UF- unmethylated forward primer and UR- unmethylated reverse primer). Horizontal red line indicates the input sequence and vertical red lines represent CpG sites.
Figure 3.2: Design of MSP primer targeting CGI of *HPCAL1* gene. a) UCSC genome browser view of chromosomal location of the target CGI within the gene indicated by the green bar and the position of the CGI in relation to the introns and exons indicated by the blue horizontal line and box. b) MSP: Visual representation of the location of the primers designed within the CGI (blue coloured area). Two primer sets are indicated by the purple boxes (MF and MR) and green boxes (UF and UR). Horizontal red line indicates the input sequence and vertical red lines represent CpG sites.
Figure 3.3: Design of MSP primer targeting region on the human genome for the cg09765089-231d tDMR. a) Visual representation of the location of the MSP primers designed. The absence of a CGI is indicated by no blue coloured area. Two primer sets are indicated by the purple boxes (MF and MR) and green boxes (UF and UR). Horizontal red line indicates the input sequence and vertical red lines represent CpG sites. b) UCSC genome browser view of chromosomal location of the region targeted within the human genome (chr7: 27291577).

Figure 3.4: Design of MSP primer targeting a single CpG site on FOXO3 gene. UCSC genome browser view of chromosomal location of the region (chr6: 108883910) targeted by the cg08792630 marker. No CGIs (usually indicated by a green box) are present in this region of the gene.
Table 3.1: Genomic location and target CpG sites of the MSP primers in this study

<table>
<thead>
<tr>
<th>Name of tDMR/tDMS</th>
<th>Chromosomal Location</th>
<th>Genomic co-ordinates of the gene or CpG site on UCSC Genome Browser assembly</th>
<th>Genomic coordinates of CpG island</th>
<th>CGI Number</th>
<th>Number of CpG sites on amplicon</th>
<th>Gene Symbol</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZNF282</td>
<td>7</td>
<td>149 195 546 – 149 226 238 (Human GRCh38/hg38)</td>
<td>149224164-149224745</td>
<td>58</td>
<td>18</td>
<td>ZNF282</td>
<td>Natalie Naidoo (2017)</td>
</tr>
<tr>
<td>HPCAL1</td>
<td>2</td>
<td>10 302 904 -10 427 604 (Human GRCh38/hg38)</td>
<td>10419591-10420168</td>
<td>54</td>
<td>10</td>
<td>HPCAL1</td>
<td>Natalie Naidoo (2017)</td>
</tr>
<tr>
<td>cg09765089-231d</td>
<td>7</td>
<td>27 291 577 (Human GRCh37/hg19)</td>
<td>Not present in a CGI</td>
<td>No CGI</td>
<td>16</td>
<td>Not present in a gene</td>
<td>Lee et al. (2015)</td>
</tr>
<tr>
<td>cg08792630</td>
<td>6</td>
<td>108 883 910 (Human GRCh37/hg19)</td>
<td>Not present in a CGI</td>
<td>No CGI</td>
<td>16</td>
<td>FOXO3</td>
<td>Park et al. (2014)</td>
</tr>
</tbody>
</table>

3.2.6 Methylation-specific PCR amplification

The markers were evaluated by two independent MSP amplifications, referred to as methylated (M) and unmethylated (U) reactions, on the samples. The specificity of both forward and reverse methylated/unmethylated primers enabled the selective amplification of methylated/unmethylated CpG sites, respectively. The primers used in MSP reactions are shown in Table 3.2. PCR amplification was carried out in a 25 μL reaction volume containing 10 uL of nuclease free water, 1X Promega GoTaq® G2 Hot Start Green Master Mix (made up of GoTaq® G2 Hot Start Polymerase, buffer, deoxynucleotide triphosphates *i.e.*, dNTPs, and optimised magnesium), 0.15 μM of the forward methylated/unmethylated primer (Inqaba Biotec/Metabion), 0.15 μM of the reverse methylated/unmethylated primer (Inqaba Biotec/Metabion) and 1 μL of bisulfite converted DNA. A no template control (NTC) was included with each set of PCR reactions. The PCR was conducted in a BIORAD T100™ Thermal Cycler (Bio-Rad Laboratories, Hercules, California, United States) under the following conditions: 95 °C for 10 minutes, 35 cycles of 95 °C for 30 seconds, varying annealing temperatures (Table 3.2) for 40 seconds, and 72 °C for 60 seconds, and a final extension of 72 °C for 7 minutes. Annealing temperatures were calculated by New England Biolabs (NEB) Tm calculator (https://tmcalculator.neb.com). The amplified products were analysed by agarose gel electrophoresis using 2 % agarose gel stained with 10 mg/ml ethidium bromide at 60 volts for 60 minutes. The gel was then viewed on a BIORAD ChemiDoc™ MP Imaging System (Bio-Rad Laboratories, Hercules, California, United States).
States), using the Image Lab software (version 6.1.0. build 7) (Bio-Rad Laboratories, Hercules, California, United States).

Table 3.2: MSP primers designed for the analysis of differential methylation patterns in human body fluids (blood, saliva, semen, and vaginal fluid)

<table>
<thead>
<tr>
<th>Marker Name</th>
<th>Primer Sequence</th>
<th>Amplicon Size (bp)</th>
<th>Annealing Temperature (˚C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZNF282</td>
<td>³FM  GGAGTGCGAGAAGATTTATAGTC</td>
<td>210</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>⁴RM  TCCTAAAACGACTCTTTATAACGAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>⁴FU  GGAGTGCTGAGAAGATTTATAGTTGT</td>
<td>210</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>⁴RU  TCCTTAAACACACTCTTTATAACAAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPCAL1</td>
<td>FM  GTTTTGTAGGGTGTAGTCTGC</td>
<td>113</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>RM  AACTCGTAAATCGATAACCTCCGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FU  GTTTTGTAGGGTGTAGTTGTTTGT</td>
<td>114</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>RU  CAACTCATATAATCAAAACTCCATA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg09765089-231d</td>
<td>FM  GTCGTTTTGTTATTTTGAGAC</td>
<td>259</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>RM  ACTACACGAAATCTCAGGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FU  TTGTTTTGTATTTTGAGGATGG</td>
<td>262</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>RU  CTCAACTACACAAAATCTCAGAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg08792630</td>
<td>FM  TGTTTTGGGGTTTTCCAATCGGTGC</td>
<td>280</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>RM  TCCACAACCTCTCTCCCTACCATGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FU  TGTTTTGGGGTTTTGAATTGGTTGTTT</td>
<td>280</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>RU  TCCACAACCTCTCTCCCTACCATATC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹FM – Forward Methylated
²RM – Reverse Methylated
³FU – Forward Unmethylated
⁴RU – Reverse Unmethylated
3.3 Results

A total of 80 body fluids, 20 each of blood, saliva, semen, and vaginal fluid were collected to validate each MSP primer set. The following results were obtained for each candidate marker primer set.

3.3.1 ZNF282 semen marker

Representative gel image of results obtained is depicted in (Figure 3.5). Semen was unmethylated for the ZNF282 marker as amplification was observed in all 20 samples which were amplified with the unmethylated primer set (210 bp) (Figure 3.5a). No amplification was observed with the methylated primers (210 bp) (Table 3.3). All 20 saliva samples displayed methylation which was inferred by successful amplification with the methylated primer pair and no amplification seen with the unmethylated primer pair (Figure 3.5b). Vaginal fluid (Figure 3.5c) and blood samples (Figure 3.5d) presented the same results, with methylation in all 20 samples and no unmethylation in any sample. Faint amplification was observed with unmethylated primers; however, instead of clear bands, smudges were seen. The NTC displayed no amplification, indicating no contamination present in the reaction.

3.3.2 HPCAL1 saliva marker

Representative gel image of results obtained is depicted in (Figure 3.6). All 20 saliva samples were consistently unmethylated as amplification was seen only with the unmethylated primer pair (114 bp), while no amplification was observed with the methylated primer pairs (113 bp) (Figure 3.6a) (Table 3.3). Conversely, blood was methylated in the HPCAL1 marker as successful amplification was observed for all 20 samples (Figure 3.6d) and an absence of amplification was noted in all samples when amplified by the unmethylated primer set. The same results were obtained for all 20 vaginal fluid (Figure 3.6c) and 20 semen samples (Figure 3.6b), wherein both body fluids were amplified successfully with the methylated primer set but showed no amplification with the unmethylated primers.
Figure 3.5: MSP based methylation profile of all four body fluids for the ZNF282 marker primer set. Lane 1: 100 bp ladder (Thermo Scientific). (A), (B), (C) and (D) represent results of semen, saliva, vaginal fluid, and blood, respectively. In A and B, lane 2 – lane 6: Products with primers specific for methylated cytosines (210 bp) and lane 9 – lane 13: Products with primers specific for unmethylated cytosines (210 bp). In C and D, lane 2 – lane 6: Products with primers specific for methylated cytosines (210 bp) and lane 10 – lane 14: Products with primers specific for unmethylated cytosines (210 bp). Lane 7 and lane 15 contain the no template control for the methylated and unmethylated primer set, respectively.
Figure 3. 6: MSP based methylation profile of all body fluids for the HPCAL1 marker primer set. Lane 1: 100 bp ladder (Thermo Scientific). (A), (B), (C) and (D) represent results of saliva, semen, vaginal fluid, and blood, respectively. In A, lane 2 – lane 6: Products with primers specific for methylated cytosine (113 bp) and lane 9 – lane 13: Products with primers specific for unmethylated cytosine (114 bp). In B, lane 2 – lane 6: Products with primers specific for methylated cytosine (113 bp) and lane 8 – lane 12: Products with primers specific for unmethylated cytosine (114 bp). In C and D, lane 2 – lane 6: Products with primers specific for methylated cytosine (113 bp) and lane 9 – lane 13: Products with primers specific for unmethylated cytosine (114 bp). Lane 7 and lane 14 contain the no template control for the methylated and unmethylated primer set, respectively.
3.3.3 cg09765089-231d vaginal fluid marker

Representative gel image of results obtained is depicted in (Figure 3.7). Of the 20 samples, nineteen saliva samples were unmethylated as no amplification was observed with the methylated primer (259 bp) however, one sample had shown slight amplification with the methylated primer set. All saliva samples displayed a uniform unmethylated status which was indicated by the amplification using the unmethylated primers (262 bp) (Figure 3.7b). The results of semen (Figure 3.7c) and blood samples (Figure 3.7d) were in consensus with that of saliva, by exhibiting amplification only with the unmethylated primer pair, while no amplification was observed with the methylated primer pair. On the other hand, all 20 vaginal fluid samples were methylated in the cg09765089-231d marker as successful amplification was observed when amplified by the methylated primers (Figure 3.7a) and no amplification was detected in all samples when amplified by the unmethylated primer set (Table 3.3).

3.3.4 cg08792630 blood marker

All 20 saliva samples successfully amplified with the unmethylated primer pair (280 bp) for the cg08792630 marker. Slight amplification was observed with the methylated primers (280 bp) in three samples (Figure 3.8b). Semen (Figure 3.8c) and vaginal fluid samples (Figure 3.8d) demonstrated similar results; showing unmethylation in all 20 samples when amplified with the unmethylated primers and, significantly low methylation in two and three samples, respectively, when amplified by the methylated primers. All 20 blood samples displayed methylation which was shown by successful amplification with the methylated primer pair. However, unmethylation was also noted in eighteen out of the 20 samples amplified with the unmethylated primer pair (Figure 3.8a) (Table 3.3).
Figure 3.7: MSP based methylation profile of all body fluids for the cg09765089-231d marker primer set. Lane 1: 100 bp ladder (Thermo Scientific). (A), (B), (C) and (D) represent results of vaginal fluid, saliva, semen, and blood, respectively. In A and D, lane 2 – lane 6: Products with primers specific for methylated cytosine (259 bp) and lane 9 – lane 13: Products with primers specific for unmethylated cytosine (262 bp). Lane 7 and lane 14 contain the no template control for the methylated and unmethylated primer set, respectively. In B and C, lane 2 – lane 6: Products with primers specific for methylated cytosine (259 bp) and lane 10 – lane 14: Products with primers specific for unmethylated cytosine (262 bp). Lane 7 and lane 15 contain the no template control for the methylated and unmethylated primer set, respectively.
Figure 3. 8: MSP based methylation profile of all body fluids for the cg08792630 marker primer set. Lane 1: 100 bp ladder (Thermo Scientific). (A), (B), (C) and (D) represent results of blood, saliva, semen, and vaginal fluid, respectively. In A, lane 2 – lane 6: Products with primers specific for methylated cytosine (280 bp) and lane 9 – lane 13: Products with primers specific for unmethylated cytosine (280 bp). Lane 7 and lane 14 contain the no template control for the methylated and unmethylated primer set, respectively. In B – D, lane 2 – lane 6: Products with primers specific for methylated cytosine (280 bp) and lane 8 – lane 12: Products with primers specific for unmethylated cytosine (280 bp). Lane 7 and lane 13 contain the no template control for the methylated and unmethylated primer set, respectively.
### Table 3.3: Summarised results for all MSP reactions on body fluids by the candidate markers

<table>
<thead>
<tr>
<th>Marker</th>
<th>Body Fluid</th>
<th>Number of Samples</th>
<th>Methylated</th>
<th>Unmethylated</th>
<th>No amplification</th>
<th>No amplification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Methylated</td>
<td>No amplification</td>
<td>Unmethylated</td>
<td>No amplification</td>
</tr>
<tr>
<td>ZNF282</td>
<td>Blood</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Saliva</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Semen</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Vaginal Fluid</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>HPCAL1</td>
<td>Blood</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Saliva</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Semen</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Vaginal Fluid</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>cg09765089-231d</td>
<td>Blood</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Saliva</td>
<td>20</td>
<td>1</td>
<td>19</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Semen</td>
<td>20</td>
<td>0</td>
<td>20</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Vaginal Fluid</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>cg08792630</td>
<td>Blood</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Saliva</td>
<td>20</td>
<td>3</td>
<td>17</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Semen</td>
<td>20</td>
<td>2</td>
<td>18</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Vaginal Fluid</td>
<td>20</td>
<td>3</td>
<td>17</td>
<td>20</td>
<td>0</td>
</tr>
</tbody>
</table>
3.4 Discussion

The application of DNA methylation analysis techniques for human body fluid identification has shown great benefit to the field of forensic science. To increase the application of methylation markers for forensics, additional novel markers need to be continually identified along with the optimisation of robust and simple DNA methylation analysis methods. The aim of the present study was to design and develop MSP assays to analyse differential DNA methylation patterns in novel and previously reported tDMRs and tDMSs, for the accurate identification of four forensically relevant body fluids: blood, saliva, semen, and vaginal fluid. Novel tDMRs for semen (ZNF282) and saliva (HPCAL1) were identified in-house at the University of KwaZulu-Natal and two tDMSs for blood (cg08792630) and vaginal fluid (cg09765089-231d) were adopted from literature. The cg08792630 marker had displayed a blood-specific hypermethylated pattern (Lee et al., 2015, 2016b; Park et al., 2014), and a vaginal fluid-specific hypermethylated pattern was observed with the cg09765089-231d marker (Lee et al., 2015, 2016b; Lin et al., 2016). Novel MSP primers were designed for all markers of the present study.

MSP reactions were designed to differentiate between methylated and unmethylated cytosines in four markers; ZNF282, HPCAL1, cg08792630 and cg09765089-231d. MSP is an inexpensive analysis method which enables the simultaneous detection of methylation and unmethylation patterns of a single sample as well as, sensitive enough to detect methylation in small quantities of DNA (Herman et al., 1996; Lizardi et al., 2016). The MSP analysis technique is less time-consuming and labour intensive, as it eliminates the need to validate results by subsequent sequencing reactions or cloning (as seen with bisulfite sequencing PCR). Denatured single-stranded DNA is treated with sodium bisulfite and amplified by two independent primer sets (methylated and unmethylated). For the best possible outcome, the primer design is a crucial aspect of an efficient MSP reaction.

The novel ZNF282 marker’s complete and exclusive unmethylated pattern in semen and contrasting methylated pattern in blood, saliva, and vaginal fluid, clearly allowed for identification of semen and its differentiation from other body fluids in this study. The differential methylation profiles of the ZNF282 marker have provided significant evidence for the application as a novel unmethylated marker for the identification of semen. A marker (ZC3H12D) was also reported by Madi et al. (2012) which successfully differentiated semen from other forensically relevant body fluids. However, this marker only included five CpG sites which displayed hypomethylation in semen and hypermethylation in the others. Lee et al. (2012) used MSP analysis to successfully differentiate semen from blood, saliva, menstrual blood, and vaginal fluid using tDMRs in the DACT1 and USP49 genes. However, Lee et al. (2012) reported that DACT1
tDMR showed 93% unmethylation and USP49 tDMR showed 97% hypomethylation in all semen samples; and hypermethylation in the other body fluids. The present study reports complete (100%) unmethylation in all semen samples and complete methylation in the remaining body fluids. Hypermethylated semen-specific markers which are differentially methylated at a single CpG site (tDMS) have also been reported, namely, cg05261336 (Lin et al., 2016), cg17621389 (Lee et al., 2015), cg26763284 (Lee et al., 2015) and cg17610929 (gene ACCN4) (Park et al., 2014). The semen marker reported in the present study shows differential methylation at eighteen CpG sites, as opposed to just one site targeted by the markers mentioned above; thus, rendering it more reliable (Kader et al., 2020).

The HPCAL1 marker successfully differentiated saliva from blood, semen, and vaginal fluid by displaying an unmethylated pattern in saliva and methylated pattern in all other body fluids. Further analyses of the HPCAL1 tDMR marker using a larger sample size will validate its great potential as a novel saliva-specific hypomethylation marker. Studies which have previously reported saliva-specific markers have identified the methylation status of these markers as hypermethylated instead of hypomethylated. These markers targeting a single CpG site include, cg09107912 (gene FNDC1) and cg16732616 (gene DMRTA2) (Lin et al., 2016) and cg09652652-2d (Lee et al., 2015). The BCAS4 marker, spanning seven CpG sites of which five displayed significant differential methylation between body fluids, was reported as hypermethylated in saliva (Madi et al., 2012). Hence, the HPCAL1 tDMR seems to be the only hypomethylation marker reported for the identification of saliva. Even though, hypomethylated saliva-specific markers are uncommon, the HPCAL1 tDMR targets ten CpG sites as opposed to just one targeted by several previous markers. This strengthens its reliability and potential applicability as a novel marker for the identification of saliva.

The cg09765089-231d tDMS had been previously reported as hypermethylated in vaginal fluid and hypomethylated in blood, saliva, and semen (Lee et al., 2015, 2016; Lin et al., 2016). In the present study, not only was the cg09765089-231d CpG site targeted but, also neighbouring CpG sites, to allow for methylation analysis of sixteen CpG sites. The MSP analysis confirmed the specificity of the primer to differentiate between vaginal fluid samples and the other body fluids (blood, saliva, and semen). Previous studies have also reported vaginal fluid-specific hypermethylation markers, such as PFN3 (Lee et al., 2012), cg09765089 and cg26079753 (gene HOTAIR) (Lee et al., 2015), and cg25416153 (Lin et al., 2016). PNF3 was hypermethylated in all body fluids analysed (Lee et al., 2012) but 65% of the markers CpG sites were methylated in vaginal fluid. DNA methylation was found statistically different at six of the CpG sites which suggested the application of the marker in site-specific DNA methylation analyses. Antunes et al.
(2016) found that a sub-region of PFN3, PFN3A (consisting of ten CpG sites), displayed intermediate methylation levels in vaginal fluid (neither hypo- nor hypermethylation). However, the results were significantly different from other body fluids of the study and indicated that PFN3A is a promising marker for vaginal fluid identification. Lee et al. (2015) found that at times, the cg09765089-231d marker displayed low yet detectable methylation signals in other body fluids apart from vaginal fluid. However, the marker designed in the present study differed from previous findings by exhibiting hypermethylation and/or complete methylation in vaginal fluid and complete unmethylation in blood, semen, and saliva.

The cg08792630 marker’s CpG site was expected to be hypermethylated in blood and hypomethylated in saliva, semen, and vaginal fluid, as reported by Park et al. (2014). In the present study we designed primers flanking the reported site to include additional CpG locations. Despite a uniform hypomethylation pattern of the marker in saliva, semen, and vaginal fluid; the marker displayed both methylation and unmethylation in 90% of the blood samples. Methylation in conjunction with unmethylation in the same body fluid is possibly an indication of cell type or methylation heterogeneity within a tissue (Song et al., 2005). Therefore, the cg08792630 marker was unable to efficiently identify blood and differentiate blood from the other body fluids. However, the varying methylation levels between blood and other body fluids could be further explored to determine if the varying degree of methylation could aid in identifying blood from the remaining body fluids of the study. The higher methylation of blood and lower methylation of saliva, semen, and vaginal fluid when amplification is carried out by the methylated primer pair could be utilised to exclude or eliminate the possibility of saliva, semen, or vaginal fluid.

Lee et al. (2012) reported a similar finding in which HOXA4, PFN3 and PRMT2 (gene-specific CpG sites) were selected as blood-specific markers, however, were methylated in blood as well as other body fluids (in varying degrees). Markers which display a hypermethylated/methylated blood-specific pattern include: C20orf117 (Eckhardt et al., 2006; Madi et al., 2012), cg06379435 (Lee et al., 2016b) and GAS2L1 (Fu et al., 2015). Madi et al. (2012) showed that C20orf117 (consisting of seven CpG sites) was hypermethylated in blood with methylation values ranging between 31.8% - 81.6%. The GAS2L1 marker which spanned eight CpG sites was only able to identify blood from other body fluids at three of these sites (Fu et al., 2015). Hypomethylated or unmethylated markers are not as common for the identification of blood; some include, cg26285698 (Forat et al., 2016) and CYTH4 (Fu et al., 2015). A total of eight CpG sites were analysed in the CYTH4 marker by Fu et al. (2015), of which only a single CpG site displayed hypomethylation in blood and hypermethylation in other body fluids. In comparison to these previous studies, the designed MSP primers targeting the cg08792630 marker in the present study
spanned sixteen CpG sites, therefore, methylation analysis by this marker may prove more beneficial as opposed to just single CpG site analysis.

3.5 Conclusion
The present study designed an **MSP** assay to identify differences in methylation profiles of four forensically relevant body fluids (blood, saliva, semen, and vaginal fluid). The study aimed to identify the body fluids based on differential methylation patterns of each body fluid-specific marker (tDMR/tDMS). The markers examined include two in-house (University of KwaZulu-Natal, Westville) developed novel tDMRs, ZNF282 (semen-specific hypomethylated) and HPCAL1 (saliva-specific hypomethylated) (Naidoo, 2017). Additional markers from literature were applied, cg09765089-231d (Lee et al., 2015) and cg08792630 (Park et al., 2014), which reported to exhibit a vaginal fluid-specific hypermethylation and blood-specific hypermethylation pattern, respectively.

MSP primers amplifying ZNF282, HPCAL1 and cg09765089-231d generated results which supported our expectations. ZNF282 and HPCAL1 tDMRs produced hypomethylation patterns only in the respective body fluids, i.e., semen and saliva, respectively. The clear differential methylation patterns permit the application of these novel markers in body fluid identification and differentiation for forensic casework. The vaginal fluid-specific, cg09765089-231d, marker produced hypermethylated patterns only in vaginal fluid therefore, the MSP primers designed for this marker are also employable in criminal investigations for vaginal fluid identification. The blood-specific marker, cg08792630, was unable to identify blood efficiently due to the presence of methylation and unmethylation. Despite the probable heterogeneity of methylation within the tissue, the marker could be used for exclusion of saliva, semen, and vaginal fluid. This is possible seeing as varying degrees of methylation (methylated primer pair) exists between the four body fluids: high methylation (blood) and low methylation (saliva, semen, and vaginal fluid).

MSP primers are designed to amplify and generate small amplicons (< 300 bp) and conduct DNA methylation analyses using bisulfite converted DNA; therefore, the assay is sensitive and works well on degraded samples which is regularly encountered in crime scene investigations (Lin et al., 2016). Future research will involve the identification of more methylation-specific markers as well as, their application in MSP analyses for body fluid identification. Multiplex MSP (also known as nested MSP) is advantageous as it is less time-consuming and uses less quantity of DNA (Lee et al., 2015). Therefore, moving forward, **multiplex MSP** assays should be developed with the use of the novel in-house developed markers to evaluate their sensitivity and stability, which is essential for forensic investigations.
CHAPTER 4

Development of multiplex MSP-SNaPshot assay for body fluid identification and mixture analysis
Abstract

Methylation SNaPshot is a quantitative technique for DNA methylation analysis, commonly used for body fluid identification. The present study aimed to develop a multiplex methylation SNaPshot assay, for the simultaneous detection of four body fluids in a single reaction, and for mixture analysis. The MSP amplicons reported in chapter 3 were sequenced to obtain the methylation status of each CpG site. Based on bisulfite sequencing results, a single differentially methylated CpG site was targeted for SNaPshot primer design. A total of four SNaPshot primers specific to each body fluid were designed i.e., ZNF282 (semen), HPCAL1 (saliva), cg09765089-54d (VF) (vaginal fluid) and cg08792630-90u (BL) (blood). The multiplex SNaPshot assay was optimised with varying primer concentration, cycling conditions, and enzyme concentrations. Simplex reactions with each body fluid and SNaPshot markers were conducted; followed by a multiplex reaction, with all body fluids and all SNaPshot markers. Mixture analysis with varying ratios of body fluids were performed to assess the sensitivity of detection. All simplex reactions showed successful and specific identification of body fluids by their respective markers as methylation or unmethylation peaks for markers were present only in their target body fluid. All markers, except saliva marker, were successfully multiplexed and clearly identified blood, semen, and vaginal fluid. Even though the saliva marker did not produce results in the multiplex reaction, mixture studies validated the specificity and sensitivity of the saliva-specific unmethylated marker. The semen marker demonstrated high levels of sensitivity by effectively identifying the presence of semen even in mixtures where semen was the minor component – usually found in sexual offence cases. All the markers developed in the present study are novel markers and reveal complete methylation/unmethylation, facilitating easier analysis. The multiplex assay allowed for the identification and differentiation of body fluids that are encountered on a daily basis at crime scenes. The implementation of the markers and multiplex SNaPshot assay, developed in the present study, would be a great addition to the field of forensic science in South Africa.
4.1 Introduction

Conventional methods of body fluid identification, allow for the identification of only one body fluid at a time, hence leads to consumption of precious evidence and takes time to obtain conclusive results (Holkötter et al., 2018; Lee et al., 2015). A multiplex detection system could permit the simultaneous identification of several body fluids.

Kaminsky and Petronis (2009) described a quantitative method of measuring site-specific DNA methylation levels, referred to as “methylation SNaPshot”. Methylation SNaPshot, based on the principle of the SBE chemistry, is a quantitative DNA methylation-based assay that measures DNA methylation levels by investigating C/T ratios. Methylation SNaPshot involves the hybridisation of a primer exactly one base pair upstream of the target CpG site, with the 3’ end of the primer directly adjacent to the nucleotide that the assay aims to identify. The primer is extended by a single fluorescently labelled nucleotide terminator (2’,3’-ddNTP) (Kaminsky and Petronis, 2009; Rana, 2018). The fluorescent dyes allow for the detection and measurement of signals emitted. Methylated CpG targets are shown as blue peaks on the resulting electropherogram, and green peaks imply the successful amplification of an unmethylated CpG site (Forat et al., 2016). Based on the dyes emitted, the methylation status of a sample can be determined, thereby enabling body fluid identification. A multiplex methylation SNaPshot reaction involves the interaction of a single PCR product with more than one SNaPshot primer. The multiplexing feature of this technique enables the analysis of multiple tissues in a single reaction thus, ensuring quicker identification of different body fluids (Gomma et al., 2017). The analysis of DNA methylation by the SNaPshot method is also beneficial as it eliminates the requirement of restriction enzymes for quantification, is cost-effective, able to easily reproduce results, and successful with low DNA concentrations (Rana, 2018).

Several methylation SNaPshot studies have identified body fluids with the use of tDMRs/tDMSs. Lee et al. (2015) employed a panel of eight markers to a multiplex methylation SNaPshot assay in order to identify different bodily fluids frequently encountered at crime scenes i.e., blood, saliva, semen, vaginal fluid, and menstrual blood. Seven markers were hypermethylated in their respective body fluids, and one marker displayed semen-specific hypomethylation. The specificity of the markers was validated by being methylated or unmethylated in the target body fluid only, and completely unmethylated or methylated in non-target body fluids, respectively. The differential methylation profiles for the markers studied enabled the identification of body fluids, as well as differentiation from one another. Lin et al. (2016) carried out a multiplex SNaPshot assay using eight markers (two for each body fluid viz. blood, saliva, semen, and
vaginal fluid) for body fluid identification. Blue peaks (indicating methylated sites) were present only in the target body fluids therefore, the method unambiguously identified all samples.

Prior to the incorporation of DNA methylation markers in forensic casework, extensive validation studies are necessary for their implementation in identifying and differentiating human body fluids. Encountering samples which contain more than a single body fluid (mixture) is a frequent occurrence in crime scene investigations. Identification of the composition of mixtures could aid in the investigation e.g., sexual offence cases are usually indicated by the presence of vaginal fluid and semen mixtures. Many studies have been conducted using various techniques to identify the body fluid composition of a mixture. Frumkin et al. (2011) used PCR coupled with a restriction enzyme to analyse saliva and semen mixtures, while Antunes et al. (2016) and Silva et al. (2016) carried out mixture analysis via pyrosequencing. Watanabe et al. (2018) incorporated SNP genotyping with pyrosequencing reactions to analyse blood and saliva samples individually mixed with semen. Lin et al. (2016) executed an assay including multiplex MSP in combination with a multiplex SNaPshot assay, to identify body fluids in mixtures of varying ratios. With the high likelihood of coming across mixed samples in forensic casework, an assay possessing high discriminatory power as well as, increased stability and sensitivity would be extremely beneficial to the forensic community.

The development of methylation-specific SNaPshot markers is based on the identification of differentially methylated DNA regions/sites in human body fluids. This chapter aimed to design SNaPshot primers based on a single differentially methylated site obtained via MSP bisulfite sequencing. The designed SNaPshot primers were multiplexed in a methylation SNaPshot assay for the identification of human blood, saliva, semen, and vaginal fluid. This chapter also aimed to efficiently identify human body fluids in mixtures by application of the multiplex MSP-SNaPshot assay, to evaluate the sensitivity of the markers. The present study reported four methylation markers specific to blood, saliva, semen, and vaginal fluid. The saliva and semen markers were developed at the Genetics laboratory (University of KwaZulu-Natal, Westville), while vaginal fluid and blood markers were modified from literature. To the best of my knowledge, the markers designed have not been reported to be multiplexed into a methylation SNaPshot assay before, therefore, the developed assay would be extremely valuable to the field of forensic science.
4.2 Materials and methods

4.2.1 Sequencing of MSP products - bisulfite sequencing data

According to results obtained in chapter 3, PCR amplicons of MSP reactions (amplified by the respective methylated/unmethylated primers) were sent for sequencing to the Central Analytical Facilities (University of Stellenbosch, Western Cape). Sequencing was performed unidirectionally using the forward primer. The sequencing data received in an ABI format was initially converted to a FASTA file on Chromas (https://technelysium.com.au/wp/chromas/). Quantification tool for Methylation Analysis (QUMA) is an online tool used for DNA methylation analysis of bisulfite sequencing reactions (Kumaki et al., 2008) (http://quma.cdb.riken.jp/). In order to analyse and determine the methylation status of each marker in the four different body fluids, the sequencing data obtained were aligned to unconverted reference genomic sequences from the UCSC genome browser (hg38) (Appendix C) in QUMA. The results identified the methylation status of each CpG site present on the amplicon, produced from the amplification of each body fluid by their respective body fluid-specific markers.

4.2.2 SNaPshot primer design

The SNaPshot primers were manually designed using Kaminsky and Petronis (2009), Lee et al. (2016b) and Forat et al. (2016) as references. Primers for methylation SNaPshot were designed complementary to the amplicon product of the MSP reaction, and according to the following guidelines:

a. **General:** All primers were designed so that the 3’ end of the primer terminates exactly 1 bp upstream (5’) of the target C of the CpG dinucleotide (Figure 4.1). Based on the bisulfite sequencing data of the MSP amplified products, a single CpG site on the marker’s amplicon was selected which displayed differential methylation patterns between the target body fluid and other body fluids e.g., a CpG site on the ZNF282 amplicon which displayed unmethylation in semen but was methylated in blood, saliva, and vaginal fluid.

b. **Primer Orientation:** All primers were designed in the reverse orientation i.e., complementary to the sense strand (Figure 4.1). The mobility of nucleotides in electrophoresis is in the order C >> A > T > G. The mobility difference between degenerate primers for a target tDMS can be minimised by the use of primers designed in the reverse direction (Lee et al., 2015).
c. **Primer Size:** When multiplexing, it was extremely vital to have primers with different sizes as this brought about different electrophoretic mobilities between the primers. This allowed for the separation of products and placement on the resulting electropherogram to be clearly indicated with no overlapping peaks. **Primers to be multiplexed were designed to have different lengths of at least 8 nucleotides (nt) (Table 4.1).** For example, if one of the primers designed had a total length of 45 nt and was to be multiplexed with another primer, the other primer would be designed to have a length of either 37 nt or 53 nt. All SNaPshot primers designed for this study were between 16 nt -26 nt. To bring about different primer sizes, T-tails of differing lengths were added to the 5’ end of the primer sequence (Table 4.1).

d. **Primer Annealing Temperature:** All primers of the multiplex were designed to have a similar annealing temperature (± 3 °C) (Table 4.1). The annealing temperature of each primer was estimated using an online tool, New England Biolabs (NEB) Tm Calculator (https://tmcalculator.neb.com/#!/main). From the melting temperature (Tm) obtained online, the annealing temperature was determined by subtracting 5 °C off the value. According to the ABI PRISM® SNaPshot™ Multiplex Kit (Applied Biosystems) protocol, an ideal annealing temperature is 50 °C, however, this was manipulated to better suit the primers designed in this study. Given that the primers have to display varying lengths but require similar annealing temperatures, only the sequence of the primer complementary to the target amplicon was used when calculating the approximate annealing temperature online. Noncomplementary T-tails added to the 5’ end was not included in this sequence.

e. **Cytosines not in a CpG dinucleotide:** In the design of the SNaPshot primers, all non-CpG cytosines present on the amplicon sequence were converted to adenines (A) in the primer.
Figure 4.1: Diagrammatic representation of forward and reverse SNaPshot primers annealing to DNA and resultant methylation signals. Primers anneal one bp upstream of the targeted CpG site and are extended on the 3’ end by a fluorescently labelled ddNTP, complementary to the DNA sequence. Primers which are synthesised based on the forward strand present black (methylated) and red (unmethylated) peaks, while reverse strand-based primers display blue (methylated) and green (unmethylated) peaks. (Forat et al., 2016)

4.2.3 Multiplex SNaPshot assay

A major component of the present study was the optimisation of the multiplex MSP-SNaPshot assay. The assay was optimised by using different concentrations of the primer (0.4 μM or 0.5 μM), type of template DNA (diluted or undiluted); and varying concentration of SAP (4 units, U, and 5 U) and exonuclease I (ExoI) (40 U and 8 U). Different cycling conditions were also tested in the optimisation phase such as, purification steps (37 °C for 60 minutes, 75 °C for 15 minutes), thermal cycling annealing time (5 seconds and 10 seconds) as well as, thermal cycling extension time (30 seconds, 1 minute and 2 minutes).

Positive and negative control reactions (Appendix E) were made up on ice according to manufacturer’s instructions of the ABI PRISM® SNaPshot™ Multiplex Kit (Applied Biosystems), and concurrently run with the multiplex SNaPshot reaction (Table 4.1). The MSP PCR products consisted of the following: amplicons from the methylated primer set (for vaginal fluid and blood), and amplicons from the unmethylated primer set (for saliva and semen). Multiplex SNaPshot was carried out on the purified and pooled MSP PCR products and using the ABI PRISM® SNaPshot™ Multiplex Kit (Applied Biosystems) in a total volume of 10 μL, containing 4 μL of SNaPshot Multiplex Ready Reaction Mix, 2 μL of purified MSP product and 0.5 μM of each of the four SNaPshot primers (Table 4.1). Each sample of the four body fluids was number-coded from one to 20. For example, all four body fluids coded 1 (saliva 1, semen 1,
vaginal fluid 1 and blood 1), were used for the multiplex SNaPshot reaction. Similarly, the other nineteen body fluids were utilised for the multiplex reaction.

Figure 4.2: Overview of methylation SNaPshot protocol carried out in the present study
Table 4.1: Multiplex SNaPshot reaction setup

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Positive control (μL)</th>
<th>Negative Control (μL)</th>
<th>SNaPshot reaction (μL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNaPshot Multiplex Ready Reaction Mix</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>SNaPshot Multiplex Control Template\textsuperscript{a}/MSP PCR product\textsuperscript{b}</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>SNaPshot Multiplex Control Primer Mix\textsuperscript{a}/ SNaPshot primers\textsuperscript{b}</td>
<td>1</td>
<td>1</td>
<td>4 (1 μL each)</td>
</tr>
<tr>
<td>Nuclease Free Water</td>
<td>2</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10</strong></td>
<td><strong>10</strong></td>
<td><strong>10</strong></td>
</tr>
</tbody>
</table>

\textsuperscript{a}SNaPshot multiplex control template and control primer mix were used in the negative and positive control reactions.  
\textsuperscript{b}MSP PCR product and four SNaPshot primers were used in the multiplex SNaPshot reaction.

The multiplex SBE reaction was conducted in a BIORAD T100™ Thermal Cycler (Bio-Rad Laboratories, Hercules, California, United States) under the following conditions: 25 cycles of 96 °C for 10 seconds, 44 °C for 10 seconds and 60 °C for 30 seconds; with a final extension of 60 °C for 10 minutes. After final purification, the fluorescently labelled SNaPshot products were separated by capillary electrophoresis on an ABI PRISM 3500 Genetic Analyzer (Life Technologies) using GS120LIZ (Life Technologies) as the internal standard (Appendix F). The data was analysed on an online interface, Thermo Fisher Connect™ (also known as, Thermo Fisher Cloud) (https://apps.thermofisher.com/apps/spa/#/dashboard). The online tool performs fragment analysis using Peak Scanner™ Software v1.0 (Life Technologies). The methylation status of all four body fluid-specific markers was determined by the incorporation of either a labelled ddGTP or ddATP on the reverse strand. A labelled ddGTP, complementary to a methylated cytosine would result in a blue peak and indicate that the CpG site studied is methylated (Forat et al., 2016). Whereas the addition of a labelled ddATP on the reverse strand would be complementary to an unmethylated thymine and this unmethylated status would be indicated by the presence of a green peak (Forat et al., 2016). The peak threshold was set at 100 RFU (relative fluorescence unit), and peaks less than 100 RFU were excluded in the analysis of the output electropherograms.
4.2.4 Mixture analysis

To test the applicability of the developed (multiplex MSP-SNaPshot) system for the identification of body fluid mixtures, blood, saliva, and vaginal fluid was individually mixed with semen in ratios of 9:1, 1:1 and 1:9 (Table 4.2), followed by DNA extraction. Furthermore, DNA from the four body fluids were mixed equally (1:1:1:1) prior to DNA extraction. DNA concentration of each body fluid was used in the mixture was 20 ng/μL.

Table 4. 2: Preparation of body fluid mixtures for MSP-SNaPshot multiplex assay

<table>
<thead>
<tr>
<th>Mixture</th>
<th>Body fluid mixture</th>
<th>Ratio</th>
<th>Volume of body fluids used to make mixtures*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>vaginal fluid : semen</td>
<td>9:1</td>
<td>4.5 μL of vaginal fluid + 0.5 μL of semen</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1:1</td>
<td>3 μL of vaginal fluid + 3 μL of semen</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1:9</td>
<td>0.5 μL of vaginal fluid + 4.5 μL of semen</td>
</tr>
<tr>
<td>4</td>
<td>blood : semen</td>
<td>1:1</td>
<td>4.5 μL of blood + 0.5 μL of semen</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>9:1</td>
<td>3 μL of blood + 3 μL of semen</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1:9</td>
<td>0.5 μL of blood + 4.5 μL of semen</td>
</tr>
<tr>
<td>7</td>
<td>saliva : semen</td>
<td>1:1</td>
<td>4.5 μL of saliva + 0.5 μL of semen</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>9:1</td>
<td>3 μL of saliva + 3 μL of semen</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>1:9</td>
<td>0.5 μL of saliva + 4.5 μL of semen</td>
</tr>
<tr>
<td>10</td>
<td>blood : saliva : semen : vaginal fluid</td>
<td>1:1:1:1</td>
<td>2 μL of blood + 2 μL of saliva + 2 μL of semen + 2 μL of vaginal fluid</td>
</tr>
</tbody>
</table>

*Mixtures were prepared by mixing the specified volumes of different body fluids however, only 1 μL of the prepared mixture was subjected to the multiplex MSP-SNaPshot assay.

4.2.5 Data analyses

Quantitative analysis of site-specific DNA methylation was carried out by determining MIVs. MIVs are calculated by the formula, M/(M+U), wherein M and U represent the height of methylated and unmethylated peaks on the electropherogram. Percentage methylation values were calculated by dividing nucleotide G intensity (blue peak) by nucleotide G plus nucleotide A (green peak) intensities i.e., G/(G+A). Percentage unmethylation values were calculated by dividing nucleotide A intensity by nucleotide G plus nucleotide A intensities i.e. A/(G+A).
4.3 Results

4.3.1 Bisulfite sequencing
MSP primers were designed in our laboratory (Genetics Department, University of KwaZulu-Natal) (chapter 3) to obtain methylation profiles of saliva, semen, blood, and vaginal fluid. Most of the primers (saliva, semen, and vaginal fluid) showed clear methylation/unmethylation in their respective body fluid. The MSP reaction products from all 80 body fluids samples were sequenced to identify the methylation status at each CpG site on the amplicon in all body fluids. The results of the sequencing reactions are depicted below (Figure 4.3-4.6).

4.3.1.1 HPCAL1 (saliva) MSP amplicon
A total of ten CpG sites were amplified by the HPCAL1 MSP marker. Sequencing results showed differential methylation profiles of saliva from the other body fluids at most of the CpG sites (CpG five, six, seven, eight, and ten) (Figure 4.3). CpG site eight was found unmethylated in at least ten saliva samples during preliminary sequencing reactions (Appendix D). This CpG site was selected to be targeted for SNaPshot primer design, because of the consistent unmethylation in saliva and methylation in other body fluids. Few CpG sites yielded unknown results and were not included in the selection process. To provide easier understanding for the reader, SNaPshot marker HPCAL1 will be referred to as saliva marker, for the rest of the chapter.
Figure 4.3: Sequencing results of HPCAL1 MSP amplicon in all body fluids. Black and clear circles indicate a methylated and unmethylated CpG site, respectively. The black line represents the reference amplicon sequence to which sequencing products were aligned to. Blank spaces indicate an unknown status at that site. The target CpG site is illustrated by the red rectangle. Numbers above represent the CpG site number and numbers below represent the nucleotide position of that specific CpG site in the amplicon.

4.3.1.2 cg09765089-231d (vaginal fluid) MSP amplicon
A total of sixteen CpG sites were present on the cg09765089-231d amplicon of which quite a few sites with an unknown status were observed in the sequencing results. CpG site seven was selected to be targeted for SNaPshot primer design (Figure 4.4). This site was 54 bp downstream from a previous reported marker cg09765089 (Lee et al., 2015) hence, the SNaPshot primer targeting this site was named cg09765089-54d in the present study. CpG seven was found only methylated in vaginal fluid and unmethylated in the other body fluids, thus, enabling the site/marker as a potential tool in the identification of vaginal fluid, and differentiation of vaginal fluid from others in the present study. To provide easier understanding for the reader, the SNaPshot marker cg09765089-54d will be referred as a vaginal fluid (VF) marker for the rest of the chapter.
Figure 4.4: Sequencing results of cg09765089-231d (vaginal fluid-specific) MSP amplicon in all body fluids. Black and clear circles indicate a methylated and unmethylated CpG site, respectively. The black line represents the reference amplicon sequence which sequencing products were aligned to. Blank spaces indicate an unknown status at that site. The target CpG site is illustrated by the red rectangle. Numbers above represent the CpG site number and numbers below represent the nucleotide position of that specific CpG site in the amplicon.

4.3.1.3 cg08792630 (blood) MSP amplicon

A total of sixteen CpG sites were present on the cg08792630 amplicon of which few sites with an unknown status were observed in the sequencing results. Differential methylation profiles of blood from other body fluids were noted at six CpG sites in this study (CpG six, seven, eleven, twelve, thirteen, and sixteen) (Figure 4.5). CpG site seven was selected to be targeted for SNaPshot primer design. This site was 90 bp upstream from a previous reported marker cg08792630 (Park et al., 2014) hence, the SNaPshot primer targeting this site was named cg08792630-90u in the present study. The novel site was found methylated in blood and unmethylated in the others therefore, facilitated the application of this marker in identification and differentiation of blood from other body fluids. To provide easier understanding for the reader, SNaPshot marker cg08792630-90u will be referred to as blood (BL) marker, for the rest of the chapter.
4.3.1.4 ZNF282 (semen) MSP amplicon

A total of eighteen CpG sites were present on the ZNF282 amplicon. Sequencing results of the ZNF282 MSP reaction showed differential methylation profiles of semen from the other body fluids at six CpG sites in the study (CpG two, three, five, seven, eight, and ten) (Figure 4.6). CpG site eight was selected to be targeted for SNaPshot primer design, because of the unmethylation profile of semen at this site, and complete methylation in other body fluids. To provide easier understanding for the reader, SNaPshot marker ZNF272 will be referred to as semen marker, for the rest of the chapter.
Figure 4.6: Sequencing results of ZNF282 (semen-specific) MSP reaction on all body fluids. Black and clear circles indicate a methylated and unmethylated CpG site, respectively. The black line represents the reference amplicon sequence which sequencing products were aligned to. Blank spaces indicate an unknown status at that site. The target CpG site is illustrated by the red rectangle. Numbers above represent the CpG site number and numbers below represent the nucleotide position of that specific CpG site in the amplicon.

Table 4.3: Primers designed in the present study for SNaPshot methylation analysis

<table>
<thead>
<tr>
<th>Marker Name</th>
<th>Primer Sequence (5’ → 3’)</th>
<th>Orientationª</th>
<th>Lengthb (nucleotides)</th>
<th>Annealing temperature (°C)</th>
<th>Methylation status of the target CpG site</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZNF282 (semen)</td>
<td>(T)₉₄ACTTTAACAATTCTACTTAC</td>
<td>R</td>
<td>35</td>
<td>41</td>
<td>Unmethylated</td>
</tr>
<tr>
<td>HPCAL1 (saliva)</td>
<td>(T)₉₄TCAACTCATAATCAATACCTCC</td>
<td>R</td>
<td>66</td>
<td>44</td>
<td>Unmethylated</td>
</tr>
<tr>
<td>cg09765089-54d (vaginal fluid)</td>
<td>(T)₉₀TCCGACCCCTCTCTCC</td>
<td>R</td>
<td>46</td>
<td>44</td>
<td>Methylated</td>
</tr>
<tr>
<td>cg08792630-90u (blood)</td>
<td>(T)₂₀AAAAAAAAAATCTATATATACCTCT</td>
<td>R</td>
<td>55</td>
<td>45</td>
<td>Methylated</td>
</tr>
</tbody>
</table>

ª Indicates the direction that the SNaPshot primer anneals to the target sequence (R – reverse)
4.3.2 Optimisation of multiplex SNaPshot assay

Initially, the PCR purification step was carried out with the use of 4 U of SAP and 40 U of Exol (New England Biolabs, MA, USA). These enzyme concentrations produced significant background noise and non-specific peaks; therefore, the enzyme concentration was changed to, and maintained at 5 U of SAP and 8 U of Exol for the rest of the optimisation reactions. The decrease in enzyme concentration for purification, had a positive effect by displaying a drastic reduction in extraneous peaks (Appendix F).

During the initial stages, purified PCR products were diluted prior to SNaPshot reactions according to stipulations in the ABI PRISM® SNaPshot™ Multiplex Kit protocol (Applied Biosystems). The suggested range of concentration of the diluted products was 0.01 pmol – 0.40 pmol. Dilution of the products did not yield results of high specificity and, elimination of the dilution step resulted in sharper peaks as well as, overall, less noise on the electropherogram (Appendix G). The improved results were attributed to not diluting PCR products as all other reaction conditions remained the same. Even though, the SNaPshot thermal cycling reaction was optimised by increasing the annealing time from five seconds to ten seconds, the effect of this change is difficult to determine. This is because, the annealing time of five seconds was always applied in conjunction with the increased enzyme concentration discussed above; therefore, it is difficult to deduce if the change in annealing time substantially contributed to the improved results.

By standardisation of other conditions (primer and enzyme concentration, as well as template type), besides the extension time, significantly different results were reported in this study. As expected, SNaPshot products with an extension time of one minute generated a peak 6-fold greater in height than the reaction with a 30 second extension time (Appendix G). Primer concentrations varied between 0.5 μM and 0.4 μM. During preliminary tests of fewer primers 0.5 μM was used however, with the increase in the number of primers multiplexed, a 0.4 μM concentration was tested. Electropherograms show that both concentrations generated specific and clear results at the correct position (Appendix H). Thermal cycling conditions for every step of the multiplex SNaPshot reaction were amended. A final optimised protocol, similarly, reported by Holtkötter et al. (2018), was reported in the present study to generate highly specific results. The main improvement of the finalised protocol in comparison to preliminary reactions, is the marked low presence of non-specific peaks and background noise. The results are visibly clearer and more
comprehensible in the optimised reaction (Figure 4.7) when compared to the trial runs (Appendix F-H). This is possibly due to the increased thermal cycling purification time, for initial and final purification steps.

4.3.3 Identification of body fluid identification by multiplex MSP-SNaPshot assay

MSP reactions were carried out, followed by simplex and multiplex SNaPshot reactions. Simplex SNaPshot reactions were conducted with all four SNaPshot primers, and PCR products from MSP amplification of a single body fluid with its methylation-specific MSP primer pair. Multiplex SNaPshot reactions were conducted with all four SNaPshot primers, and PCR products from amplification of all four body fluids with all four methylation-specific MSP primers. The positive and negative control reactions are shown in Appendix E.

All simplex SNaPshot reactions showed that markers generated either methylated (blue) or unmethylated (green) peaks only in their target body fluids (Figure 4.7A-D). A green peak indicated the addition of an ddATP, thus represented an unmethylated CpG site. Green peaks represented the successful amplification of the semen marker (Figure 4.7A) and saliva marker (Figure 4.7D) located at 40 nt and 69 nt, respectively. Blue peaks were generated as a result of the addition of a ddGTP, therefore, demonstrated a methylated CpG site present. Blue peaks represented the amplification of VF (vaginal fluid) (Figure 4.7C) and BL (blood) (Figure 4.7B) at 51 nt and 59 nt, respectively.

The presence of blood, semen and vaginal fluid were clearly indicated by peaks at BL, semen and VF markers, respectively, in the multiplex reaction (Figure 4.7E). The semen and VF marker presented with methylated and unmethylated peaks at their respective positions in the multiplex reaction. The stacked bar graph (Figure 4.8) illustrating MIVs for all markers in their respective simplex reactions, indicated that the SNaPshot semen and saliva markers were completely unmethylated (100% unmethylation) in the body fluids which they are specific for. VF and BL markers were completely methylated (100% methylation) in vaginal fluid and blood, respectively. The bar graph representing MIVs of each marker in the multiplex reaction indicated that the semen and VF markers were hypomethylated; 53% and 54% unmethylation, respectively. BL was 100% methylated (Figure 4.9) in the multiplex reaction. MIVs could not be calculated for the saliva marker, seeing as unsuccessful SBE indicated by the absence of a peak on the electropherogram (Figure 4.7E).
Figure 4.7: Representative electropherograms of body fluids in simplex and multiplex methylation SNaPshot. Results for simplex reactions in semen, BL, VF, and saliva are represented in A, B, C, and D, respectively, as well as the multiplex SNaPshot reaction represented in E. The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green and blue peaks indicated an unmethylated and methylated CpG site (reverse primer), respectively.
Figure 4. 8: Percentage methylation or unmethylation (MIV) of each marker with its specific body fluid in the simplex reaction. Percentage methylation values were calculated by dividing nucleotide G intensity (blue peak) by nucleotide G plus nucleotide A (green peak) intensities. Percentage unmethylation values were calculated by dividing nucleotide A intensity by nucleotide G plus nucleotide A intensities.

Figure 4. 9: Percentage methylation or unmethylation (MIV) of each marker in the multiplex SNAPshot reaction. Percentage methylation values were calculated by dividing nucleotide G intensity (blue peak) by nucleotide G plus nucleotide A (green peak) intensities. Percentage unmethylation values were calculated by dividing nucleotide A intensity by nucleotide G plus nucleotide A intensities.
4.3.4 Mixture analysis

The body fluid mixtures were carried out by mixing blood, saliva, and vaginal fluid individually with semen in ratios of 9:1, 1:1 and 1:9. Bisulfite-converted DNA from the above four body fluids were also mixed equally (1:1:1:1). These mixed bisulfite-treated DNA samples were subjected to the designed multiplex MSP-SNaPshot assay.

The presence of semen in the mixtures with vaginal fluid (9:1) was clearly indicated by the unmethylated status (green peak) of the semen marker (Figure 4.10A). The presence of vaginal fluid in the mixtures was indicated by the methylation status (blue peak) of the VF marker. No results were generated in the mixture consisting of vaginal fluid and semen in a 1:1 ratio (Figure 4.10B). A high-intensity methylated peak at the BL marker indicated the presence of blood in the vaginal fluid and semen mixture (9:1 ratio) (Figure 4.10A). The semen and vaginal fluid mixture (1:9) showed the presence of semen, however no peak was observed for vaginal fluid (Figure 4.10C).

![Figure 4.10: Representative electropherograms of vaginal fluid and semen mixtures in varying ratios. Vaginal fluid and semen mixture analysis is indicated by 9:1 ratio (A), 1:1 ratio (B) and 1:9 ratio (C). The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green peaks indicated an unmethylated CpG site (reverse primer), while blue peaks represent a methylated CpG site (reverse primer).]
The presence of blood and semen was successfully detected in mixture ratios of 9:1 and 1:9 (Figure 4.11A and C), by the presence of methylated and and unmethylated peaks only at BL and semen, respectively. No peaks were generated in the mixture consisting of blood and semen in a 1:1 ratio (Figure 4.11B).

**Figure 4.11:** Representative electropherograms of blood and semen mixtures in varying ratios. Blood and semen mixture analysis is indicated by 9:1 ratio (A), 1:1 ratio (B) and 1:9 ratio (C). The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green peaks indicated an unmethylated CpG site (reverse primer), while blue peaks represent a methylated CpG site (reverse primer).

Body fluids, saliva and semen, were unambiguously identified in all mixtures (Figure 4.12A-C) by the presence of unmethylated peaks only at saliva and semen markers, respectively. The methylated status of BL in the 9:1 ratio (Figure 4.12A) indicated the presence of blood in this mixture. The peak for the saliva marker displayed a consistent low intensity, even if saliva was the major constituent of the mixture (9:1) (Figure 4.12A).
Figure 4.12: Representative electropherograms of saliva and semen mixtures in varying ratios. Saliva and semen mixture analysis is indicated by 9:1 ratio (A), 1:1 ratio (B) and 1:9 ratio (C). The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green peaks indicated an unmethylated CpG site (reverse primer), while blue peaks represent a methylated CpG site (reverse primer).

No significant results were obtained on the electropherogram for body fluid mixture (Mixture 10) comprising all body fluids mixed in an equimolar concentration (1:1:1:1) (Figure 4.13). Therefore, the methylation status at all markers was unknown.

Figure 4.13: Representative electropherogram of all four body fluids (blood, saliva, semen and vaginal fluid) mixed in an equal ratio. All four body fluids were mixed in a equal ratio of 1:1:1:1. The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green peaks indicated an unmethylated CpG site (reverse primer), while blue peaks represent a methylated CpG site (reverse primer).
4.4 Discussion

Efficient body fluid identification from crime scene exhibits is a pivotal step in criminal investigations. There has been a recent surge in the use of tDMRs and tDMSs in identifying the source of biological evidence found at crime scenes (Antunes et al., 2016; Frumkin et al., 2011; Holtkötter et al., 2018; Madi et al., 2012). The differing methylation patterns in varying tissues/cells allow for specific identification of forensically relevant body fluids. The objective of the present study was to develop a multiplex methylation SNaPshot assay using novel markers to assess the difference in methylation profiles between four body fluids, and subsequently result in the simultaneous identification of blood, semen, saliva, and vaginal fluid. The study also aimed to apply the multiplex MSP-SNaPshot assay to identify the body fluid composition of mixtures.

SNaPshot primers were designed based on MSP bisulfite sequencing data. A single differentially methylated CpG site on the MSP amplicon was chosen, and SNaPshot primers were developed to target that specific CpG site. The two novel MSP markers, ZNF282 (semen) and HPCAL1 (saliva) (identified and developed in our lab) were used to design SNaPshot primers for semen and saliva, respectively. The two markers for blood (Park et al., 2014) and vaginal fluid (Lee et al., 2015) were initially adopted from the reported studies, but target CpG sites were modified in the present study to design novel MSP and SNaPshot markers.

4.4.1 Identification of body fluids by multiplex SNaPshot methylation analysis

4.4.1.1 Identification of semen
The semen SNaPshot primer targeted an unmethylated CpG site on the ZNF282 MSP amplicon; and would ideally yield a green peak at 36 nt (Table 4.3). Identification of semen and differentiation from other body fluids by the semen marker was clearly indicated by the presence of a green peak at 40 nt only in the semen reaction. This highlighted the specificity of the in-house designed semen marker for the identification of semen and showed great potential of this marker as a novel semen-specific unmethylation marker for forensic science. According to the ABI PRISM® SNaPshot™ Multiplex Kit protocol (Applied Biosystems), the incorporation of fluorescently labelled ddNTPs during the SNaPshot reaction could impact the mobility of the products during capillary electrophoresis. Therefore, it is possible that shorter fragments may appear approximately 4 nt larger than their expected size. This explained why the electropherogram peaks in this study were approximately 5 nt larger than the expected size (Table 4.3).
Results of the multiplex reaction, which included PCR amplification of single body fluid followed by methylation SNaPshot, showed that the semen marker could be successfully multiplexed to identify the presence of semen. However, mixture analysis was unable to identify semen in 1:1 ratio mixture, which needs to be optimised. Other semen-specific unmethylation markers which were unmethylated in semen only and methylated in other body fluids were reported by Lee et al. (2015) (cg17621389) and An et al. (2013) (DACT1 and USP49). These markers were able to efficiently identify and differentiate semen from others in the study. Similar to the semen marker designed in our laboratory, these SNaPshot markers displayed complete unmethylation only in semen and complete methylation in the rest of the body fluids. Semen-specific methylation markers for the identification of semen have also been reported, cg05261336 (Lin et al., 2016), cg17610929 (Lee et al., 2015, 2016b; Lin et al., 2016), and cg26763284-138d (Lee et al., 2015, 2016b).

4.4.1.2 Identification of vaginal fluid

The cg09765089-54d (also referred to as VF in the present study) SNaPshot primer was designed based on the reverse sequence, targeting a differentially methylated CpG site on the cg09765089-231d MSP amplicon. This site displayed methylation in vaginal fluid and unmethylation in the others and would ideally yield a blue peak at 47 nt (Table 4.3). The identification and differentiation of vaginal fluid from other body fluids by the VF marker was clearly indicated by the presence of a blue peak at 51 nt only in vaginal fluid. However, identification of vaginal fluid was unsuccessful, when in low quantity during mixture analysis, and requires further optimisation.

Most SNaPshot markers designed to identify vaginal fluid display a methylated state, such as cg26079753-7d (Lee et al., 2015), cg14991487-85d and cg03874199-212d (Forat et al., 2016), and cg03874199 (Holtkötter et al., 2018). Lee et al. (2015) reported that the cg09765089-231d marker had displayed hypermethylation in vaginal fluid and menstrual blood, and hypomethylation in the others. However, low methylation values of this marker were also seen in blood, saliva, and semen (Lee et al., 2015). Lee et al. (2016b) also reported similar findings as Lee et al. (2015), excluding the presence of detectable methylation in blood, saliva, and semen. Greater peak heights shown by Lee et al. (2016b) in comparison to the present study may be due to the increased primer concentration (0.9 μM) used in that study. Selection of a different CpG site on the cg09765089-231d MSP amplicon, as opposed to the site targeted by Lee et al. (2015, 2016b), could be the reason a higher specificity was noted for the marker in the present study (cg09765089-54d). In contrast to the present study (100% methylation), both previously mentioned studies had shown slight unmethylated peaks with methylated peaks at the vaginal
fluid markers. Vaginal fluid-specific hypermethylation markers (cg25416153 and cg09765089) utilised by Lin et al. (2016) had shown hypermethylation in vaginal fluid (52 % - 65 %) only, and hypomethylation in the others. The specificity of the VF marker designed in the present study for the identification of vaginal fluid suggests the great potential of this marker as a novel vaginal fluid-specific methylation marker.

4.4.1.3 Identification of blood

The cg08792630-90u (also referred to as BL in the present study) SNaPshot primer was designed based on the reverse sequence, targeting a methylated CpG site on the cg08792630 MSP amplicon; and would ideally yield a blue peak at 56 nt (Table 4.3). The identification and differentiation of blood from other body fluids by the BL marker was clearly indicated by the presence of a blue peak at 59 nt, only in the blood reaction. Therefore, suggested the applicability of it as a potential blood-specific hypermethylation marker. The multiplex SNaPshot reaction showed that BL could be successfully multiplexed to identify the presence of blood; however, mixture analysis did not display consistent results when a low quantity of blood was present in the mixture.

Several studies have reported SNaPshot markers designed to identify blood in a methylated state, such as cg03363565-59d (Forat et al., 2016), as well as cg13763232, cg03363565 and EFS (Holtkötter et al., 2018). Blood-specific markers reported by Lin et al. (2016) displayed hypermethylation in blood (48 % - 50 %) and menstrual blood (25 % - 30 %) only, and hypomethylation in the rest. Selection of a different CpG site on the cg08792630 MSP amplicon in the present study, as opposed to the site targeted by Lee et al. (2015, 2016b), showed significant difference. The cg08792630-90u marker (in the present study) displayed complete methylation in blood, while cg08792630 (Lee et al., 2016b) displayed hypermethylation. CpG site thirteen on the cg08792630 MSP amplicon of the present study was targeted by Lee et al. (2016b) for blood identification. Hypermethylation was illustrated by the presence of both green and blue peaks for the cg08792630 marker Lee et al. (2015, 2016b). However, the blood-specific marker designed in the present study showed only methylated (blue) peaks in blood, thus, indicated the higher specificity of the marker.

The cg01543184 marker had shown hypermethylation in blood and low levels of methylation in semen as well (Lee et al., 2015). This was explained by the presence of white blood cells (WBCs) in semen that contributes to the amplification of semen by this marker. Along with hypermethylation in blood, cg06379435 was found methylated in semen and saliva (Lee et al., 2016b). There are many studies which have displayed blood-specific identification markers in an
unmethylated state. These include cg26285698-14d (Forat et al., 2016) and cg26285698 (Holtkötter et al., 2018) which is hypomethylated in blood, while cg03363565 was reported unmethylated in blood (Holtkötter et al., 2018).

4.4.1.4 Identification of saliva

The saliva SNaPshot marker was designed in our laboratory based on the reverse sequence, targeting an unmethylated CpG site on the HPCAL1 MSP amplicon; thus, expected to yield a green peak at 67 nt (Table 4.3). Saliva was correctly identified and differentiated from other body fluids of the study by the saliva marker; indicated by the green peak at 69 nt, only in saliva. This clearly depicted the specificity of the marker and provided evidence on its application in forensic science as a novel saliva-specific unmethylation identification marker. However, results showed that saliva was not amplified by the saliva marker in the multiplex reaction. The saliva marker was the longest marker in the study (67 nt) and a reasonable explanation for the lack of amplification in the multiplex assay could be the length of the primer, which could have resulted in a low annealing efficiency. Lee et al. (2015) included a saliva-specific primer in their multiplex assay, which was 74 nt in length, however, variable concentrations of each primer was used as opposed to the standardised concentration, in the present study. Kaminsky and Petronis (2009) have stated that in a multiplex MS-SNuPE assay, combining ten primers, the most extended primer should ideally be around 65 nt.

Lin et al. (2016) reported a saliva-specific markers cg09107912 and cg16732616 which displayed hypermethylation (~55 %) in saliva and hypomethylation in other body fluids. Lee et al. (2015) reported that all saliva samples of the study were hypermethylated in cg09652652-2d and hypomethylated in the others. The in-house saliva-specific marker developed in the present study differed from cg09652652-2d, as the marker in the present study was completely unmethylated in saliva only. However, cg09652652-2d displayed methylation in vaginal fluid, menstrual blood, and semen from vasectomised males (Lee et al., 2015). Lee et al. (2016b) carried out SNaPshot reactions with the same marker and found hypermethylation in saliva samples only, and complete unmethylation in other body fluids.

4.4.2 Mixture Analysis

Mixtures containing vaginal fluid and semen are one of the more routinely encountered admixed samples in forensics – usually indicative that a sexual offence has taken place. Vaginal fluid and semen were correctly identified in the mixture ratio of 9:1 only. The inability to detect vaginal fluid in other mixtures (1:1 and 1:9) suggested that the VF marker can not identify vaginal fluid when present as the minor constituent in a mixture. Lin et al. (2016) reported low intensity peaks for the cg09765089, vaginal fluid-specific, marker as well as, false negatives for the marker in
mixtures containing low quantities of vaginal fluid (i.e., 10%). The identification of semen in the mixture (9:1) clearly indicated that even a mixture, in which semen is the minor component, the semen SNaPshot marker was able to detect and identify the specific body fluid. This is an extremely vital feature in the marker’s potential for application in forensic science as more often than not, in sexual offence cases, semen is the minor component of the evidence obtained.

The high intensity peak at BL suggesting the presence of blood in the semen and vaginal fluid mixture (9:1), is a relatively common find. Lee et al. (2015) reported a high intensity methylated (blue) peak corresponding to a blood marker (cg01543184) when semen samples were analysed in a multiplex SNaPshot reaction. The presence of blood may be as a result of semen containing WBCs therefore, the detection of WBCs contributes to the BL peak on the electropherogram (Lee et al., 2015). The presence of WBCs in semen usually occurs among males with infertility problems (Pentyala et al., 2007). There is a high likelihood of the semen samples in the present study containing WBCs as these samples were obtained from a fertility clinic.

Blood and semen were unambiguously identified in mixtures (9:1 and 1:9) by methylated and unmethylated peaks only at the BL and semen markers, respectively. The results validated the specificity of both markers in blood and semen mixtures as the successful detection of both body fluids was possible even when present as minor components. The greater peak height at BL for the mixture (1:9) could be due the amplification of the marker in blood and WBCs in semen.

Despite unsuccessful amplification of saliva in the multiplex SNaPshot reaction, semen and saliva were successfully identified in all mixtures by the presence of unmethylated peaks only at the saliva and semen marker, respectively. This further validated the ability of the in-house designed saliva and semen marker to unambiguously identify the presence of their specific body fluids, even in low concentrations. Further research, regarding the saliva-specific novel marker, should include applying higher concentration of primers. This was not possible to accomplish in the current study due to time constraints. An increased concentration (300 μM) of each SNaPshot primer reportedly displayed high specific and sensitive results (Lin et al., 2016). An et al. (2013) and Lee et al. (2015) reported that the concentration of SNaPshot primer used in their reactions ranged between 0.2 μM and 0.8 μM, similar to that of the present study (0.5 μM).

The peak at BL in saliva and semen mixture (9:1) (similar to vaginal fluid and semen 9:1 mixture) may have resulted by amplification of the marker in WBCs of semen. Silva et al. (2016) reported a pyrosequencing assay that was only able to detect that a mixture is present, however, unable to identify specific body fluids of the mixture. The developed multiplex MSP-SNaPshot assay in the present study was able to clearly identify different body fluids components in 60 % of the stimulated mixtures. SNaPshot markers employed by Forat et al. (2016) were able to identify their
respective specific body fluids (≥ 20 %) in mixtures with other body fluids of the study. The present study, using in-house developed semen, saliva and BL markers, was able to detect the presence of their respective body fluids (≥ 10 %) in various mixtures. The unambiguous identification of all body fluids mixed in equal ratios was reported by Lin et al. (2016), however, the present study generated no peaks in the equimolar mixture of 1:1:1:1.

Holtkötter et al. (2018) described the importance of reporting markers that were able to differentiate blood from menstrual blood and vaginal fluid, by differentially methylation profiles. The study identified a single marker which was completely unmethylated in blood but hypermethylated in the other three body fluids tested. This is an valuable tool in forensic investigations as it is often extremely difficult to differentiate blood from menstrual blood and vaginal fluid, individually and in mixtures. The validity of the novel BL marker, developed in the present study, could be strengthened by investigating its applicability in differentiating blood from menstrual blood in future studies.

### 4.5 Conclusion

The objective of the present study was to develop a multiplex SNaPshot assay using MSP markers to assess the difference in methylation profiles between four body fluids. Analysis of the resultant methylation profiles were employed in the identification of blood, semen, saliva, and vaginal
Semen and saliva-specific in-house developed markers displayed high specificity and successfully identified the two body fluids. The additional two SNaPshot markers (BL and VF) displayed methylation profiles which permitted their use in the identification of blood and vaginal fluid, respectively. VF and BL markers were not only able to clearly detect the presence of their respective body fluids, but also showed higher body fluid specificity than the previously reported markers which their design was based on.

Sensitivity of the designed SNaPshot markers were tested by application of the developed multiplex MSP-SNaPshot system on mixtures. All markers, except VF, were able to detect their respective body fluids in a mixture, even when the specific body fluid was present as a minor component. Multiplexing requires a smaller amount of DNA for the reaction which is an extremely beneficial feature since forensic science usually involves degraded samples in lesser quantity. The analysis of many samples in a single reaction also ensures a quicker and simpler protocol. Furthermore, the completely methylated or unmethylated profiles in the markers’ target body fluid facilitate easier visual interpretation of electropherograms, without the requirement of supplementation by extensive statistical analysis. Future research will involve testing the applicability of this developed multiplex SNaPshot assay with real crime scene samples to enable the efficient identification of human body fluids for forensic purpose.
5.1 Purpose of this study

The crime statistics of South Africa (South African Police Services, Crime Statistics) exemplify the importance and requirement of a judicial system that works alongside forensic science in investigations, to ensure the conviction of perpetrators and exoneration of the innocent. Apart from individual identification, detection of human body fluids in crime scene exhibits presents
valuable information to the investigation, for example, underlining the nature of the crime and probable course of events (Vidaki et al., 2013). The chemistry of catalytic, enzymatic, spectroscopic, and immunological techniques forms the basis of traditional presumptive and confirmatory tests for body fluid identification (Lee et al., 2016a; Virkler and Lednev, 2009). Despite being sensitive, the presumptive tests have low specificity, conventional methods cannot be multiplexed, and specific tests are not available for body fluids like vaginal fluid and menstrual blood (Gomma et al., 2017; Frumkin et al., 2011; Lee et al., 2016a).

Most samples are also degraded and not present in their natural texture due to drying out over a long time or by being washed with bleach (Rana, 2018). Each presumptive or confirmatory test is able to detect only a single body fluid and does not enable multiplexing (detection of several body fluids in one reaction) (Gomma et al., 2017; Lee et al., 2016a). Finally, the inability to automate these methods make them time-consuming and labour-intensive. RNA-based methods have also been employed for body fluid identification (Hanson et al., 2009; Zubakov et al., 2008), however, RNA is unstable and prone to degradation.

The discovery of differentially methylation regions on the human genome (tDMRs/tDMSs) that display varying methylation patterns according to tissue type, presented a promising marker for body fluid identification to the forensic community (Eckhardt et al., 2006; Slieker et al., 2013). DNA methylation-based markers showed advantages over previous tests as they are sensitive and specific, are automatable thereby saving time, and enable multiplexing to analyse several samples in a single reaction. As opposed to RNA-based tests, the equipment needed for DNA methylation-based methods are standardly available in forensic DNA laboratories, as the same equipment is used for DNA profiling (Richards et al., 2018). These advantages ensure the preservation of precious material for downstream analyses e.g., STR profiling for individual identification (Gomma et al., 2017; Zahra et al., 2018).

Several studies have used tDMR/tDMS markers to successfully identify body fluids in single samples and mixtures (An et al., 2013; Frumkin et al., 2011; Holkötter et al., 2018; Lee et al., 2012, 2015; Lin et al., 2016). The significant need for methods which correctly identify body fluids in criminal investigations, in addition to the limited number of available markers which can display differential methylation among tissues, led to our attempt in narrowing this gap in research. The present study aimed to design DNA methylation-based assays using novel markers for the efficient and accurate identification of forensically relevant body fluids (blood, saliva, semen and vaginal fluid), individually and in mixtures.

A total of 60 samples (20 each of blood, saliva, and vaginal fluid) were collected from volunteers at the King Dinuzulu Hospital, Durban, and 20 semen samples were received from volunteers at
the Dr. Aevitas Fertility Clinic, Cape Town. Potential differentially methylation regions on ZNF282 and HPCAL1 genes had been previously identified in our laboratory (Genetics Department, University of KwaZulu-Natal) by Naidoo (2017). To determine the methylation status of these tDMRs in each body fluid analysed, the present study developed an MSP assay. MSP primers specific to semen and saliva were designed targeting CGIIs in the tDMRs of genes ZNF282 and HPCAL1, respectively. For blood and vaginal fluid primers, single CpG sites each, were chosen in two previously reported markers from literature; cg08792630 (Park et al., 2014) and cg09765089-231d (Lee et al., 2015). MSP primers for blood and vaginal fluid were designed to include additional CpG sites flanking the reported CpG site.

MSP requires a significantly less volume of DNA, is cost-effective, involves simple and easier interpretation of results, and displays high levels of sensitivity i.e., the detection of even one methylated CpG site among several unmethylated sites (Dahl and Guldberg, 2003; Herman et al., 1996). The elimination of cloning or restriction enzymes results in a faster and less laborious protocol than other DNA methylation-based methods (Dhingra et al., 2014). MSP is a qualitative method therefore, to validate the results and allow multiplexing of primers, a methylation SNaPshot assay was designed for body fluid identification in singularity and mixtures.

Sequencing of the MSP products indicated the methylation status of every CpG site on the marker’s amplicon in all body fluids. A single CpG site which displayed a differential methylation status among body fluids was targeted for SNaPshot design, according to guidelines by Kaminsky and Petronis (2009), Forat et al. (2016) and Lee et al., (2016b). Methylation SNaPshot measures site-specific DNA methylation levels by evaluating C/T ratios (Kaminsky and Petronis, 2009). Amplification of methylated CpG sites generate blue peaks on the electropherogram, and unmethylated CpG sites are represented as green peaks when the SNaPshot primer design is based on the complementary reverse sequence (Forat et al., 2016). Quantitative analysis is conducted by calculation of MIVs – M/(U+M), in which M and U represent the methylated and unmethylated peak heights, respectively (Lin et al., 2016). Multiplex SNaPshot simultaneously analyses multiple samples in a single reaction (Gomma et al., 2017). The multiplexing ability of this method enabled the analysis of mixtures to determine which individual body fluids were present. Methylation SNaPshot reproduces results easily, is cost-effective, uses very little DNA, and is less time-consuming. A multiplex SNaPshot assay was designed to correctly identify blood, saliva, semen, and vaginal fluid. A frequent encounter in forensic casework is the presence of evidence in the form of mixtures rather than pure samples (single body fluid). To test the sensitivity of the novel markers and applicability of the designed multiplex MSP-SNaPshot assay
in forensic casework, body fluid mixtures prepared in varying ratios were subjected to the assay, to accurately identify individual components of the mixture.

5.2 Summary of findings

Novel tDMRs for semen (ZNF282) and saliva (HPCAL1) were identified in-house at the University of KwaZulu-Natal. Two previously reported tDMSs, cg08792630 and cg09765089-231d, were targeted and modified to design MSP primers for blood and vaginal fluid, respectively. MSP primers were designed in-house for all markers of the present study. The first objective of the study was to conduct methylation profiling of tDMRs/tDMSs in blood, saliva, semen, and vaginal fluid by MSP. The second objective was to design SNaPshot primers targeting a differentially methylated site on MSP amplicons, as well as optimisation of a multiplex methylation SNaPshot assay, for body fluid identification and mixture analysis.

The novel ZNF282 MSP marker was completely unmethylated in semen and methylated in blood, saliva, and vaginal fluid; therefore, clearly allowed for the identification of semen and its discrimination from other body fluids, in this study. The ZNF282 SNaPshot marker was successful in a multiplex reaction, however, could not efficiently identify semen during mixture analysis. Future studies should focus on optimisation of mixture SNaPshot assays with varying primer concentrations. SNaPshot results for the ZNF282 marker provided significant evidence to permit its application as a novel unmethylated marker for the identification of semen.

The HPCAL1 marker successfully differentiated saliva from other body fluids by displaying a saliva-specific unmethylated profile and methylated profile in all other body fluids. Saliva-specific unmethylated markers are uncommon, with HPCAL1 possibly being the only hypomethylated tDMR reported for identification of saliva. Even though saliva-specific hypomethylated markers are rare, HPCAL1 MSP marker spans ten CpG sites as opposed to just one site reported by previous studies – strengthening the reliability of the marker. Despite the unsuccessful detection of saliva in the multiplex SNaPshot reaction, the simplex SNaPshot reaction correctly identified and differentiated saliva from other body fluids by the saliva SNaPshot marker. The length of the primer (67 nt) could be the reason for failed amplification in the multiplex reaction. Kaminsky and Petronis (2009) have stated that in a multiplex MS-SNuPE assay combining ten primers, the most extended primer should ideally be around 65 nt. Due to time constraints, the multiplex SNaPshot reaction could not be further optimised in the present study. However, future research should use variable concentration of primers, and possibly reduce the primer length for the HPCAL1 SNaPshot primer.
The cg09765089-231d tDMS was previously reported as hypermethylated in vaginal fluid and hypomethylated in blood, saliva, and semen (Lee et al., 2015). In the present study, not only was the cg09765089-231d CpG site targeted for MSP analysis but, other CpG sites neighbouring it were also amplified to allow for methylation profiling of **sixteen** CpG sites. The MSP results confirmed the specificity of the primer to identify and differentiate between vaginal fluid samples and other body fluids, by displaying vaginal-fluid specific hypermethylation and complete unmethylation in others. The present study reported a novel SNaPshot marker targeting a differentially methylated CpG site based on the cg09765089-231d MSP amplicon i.e., cg09765089-54d SNaPshot marker (also referred to as VF in the present study). The previously reported marker, cg09765089-213d, targeted a CpG site 231 bp downstream from cg09765089; whereas VF designed in the present study, targeted a CpG site 54 bp downstream from cg09765089. Results of the multiplex SNaPshot reaction showed that VF could be successfully multiplexed to correctly identify the presence of vaginal fluid by complete methylation (100%).

The cg08792630 marker was reported as hypermethylated in blood and hypomethylated in saliva, semen, and vaginal fluid by Park et al. (2014). In the present study, we designed MSP primers for methylation profiling that flanked the reported site, to include other CpG sites for analysis. Despite a uniform hypomethylation pattern of the MSP marker in saliva, semen, and vaginal fluid, the majority of blood samples showed both methylation as well as unmethylation. Methylation and unmethylation in the same body fluid possibly indicated methylation heterogeneity within a tissue or cell. The results showed that the MSP primers designed for the cg08792630 marker were unable to efficiently identify blood or differentiate blood from the other body fluids. However, amplification of body fluids with the methylated primer set produced significantly different methylation levels between blood and other body fluids. The varied methylation levels could be further explored to exclude or eliminate the presence of saliva, semen, or vaginal fluid. The BL marker in the present study targeted a CpG site 90 bp upstream from the CpG site targeted in previous studies, cg09765089. The designed cg08792630-90u SNaPshot marker (also referred to as BL in the present study) could be successfully multiplexed to identify the presence of blood. The heterogeneity of blood reported in MSP reactions of this study was not visible by multiplex SNaPshot analysis, as only a single methylated (blue) peak represented the presence of blood. This validated the greater specificity of the multiplex methylation SNaPshot assay in comparison to DNA methylation analysis by MSP only.

Semen and vaginal fluid were correctly identified in the mixture ratio of 9:1. BL and semen unambiguously identified blood and semen, respectively, in mixtures (9:1 and 1:9) therefore, provided validation of both marker’s specificity in the detection of their respective body fluids.
Semen and saliva were successfully identified in all mixture ratios by semen and saliva markers. A detectable peak at BL in mixtures not containing blood DNA may be due to semen containing white blood cells (WBCs). Semen usually comprises of WBCs among males with infertility problems. Seeing as the semen samples of this study were obtained from a fertility clinic, there is a higher possibility of WBCs present in the semen samples analysed. Successful identification of semen in the mixture, 9:1, showed that even in a mixture containing semen as the minor component, ZNF282 was able to detect and identify the specific body fluid. This high sensitivity is an extremely vital feature in the marker’s potential for application in forensic science as more often than not, in sexual offences, semen is the minor component of evidence. Herein, the developed multiplex MSP-SNaPshot assay of the present study was able to clearly identify different body fluids in 60 % of the simulated mixtures. SNaPshot markers described in the present study were able to identify their respective body fluids in a mixture containing more than 10 % of that body fluid. Unsuccessful amplification of all SNaPshot markers was reported in the mixture analysis of all body fluids present in equal concentration (1:1:1:1). The effect of increased primer concentration should be investigated in the future studies.

5.3 Conclusion and future research

In-house developed markers, ZNF282 and HPCAL1, displayed semen and saliva-specific unmethylated profiles, respectively. Novel VF and BL SNaPshot markers targeting different CpG sites than those reported in literature, were found methylated in vaginal fluid and blood, respectively. To our knowledge, these tDMRs/tDMSs have not been previously reported for the identification and differentiation of forensically relevant body fluids. The results reported by the study do suggest great potential of the novel markers and developed multiplex MSP-SNaPshot assay in routine forensic casework. However, further research involving validation studies is required prior to their implementation in forensic laboratories.

Future research would involve extensive validation studies assessing the marker on a larger sample size including: different ethnic and age groups, diverse geographical locations, as well as evaluating the methylation profiles of the marker in diseased individuals. Studies going forward will help elucidate any changes/differences in methylation patterns of semen, saliva, VF, and BL markers that may occur as a result of age, ethnicity or disease. The present study applied the developed multiplex MSP-SNaPshot assay to analyse the sensitivity of the markers only on mixtures, which are one of the many unfavourable conditions in forensic casework. The impact of additional adverse conditions need to be studied by conducting stability tests to evaluate the effect of prolonged exposure to various environmental conditions (different surfaces, temperatures, soil microbes).
Mixture analysis by combination with artificially synthesised DNA (specificity test) and DNA of non-primates species (species-specificity test) will also be beneficial. The results of such stringent tests may supplement the methylation data reported in the present study. The application of more identified tDMRs in addition to their methylation profile in various body fluids, to the assay developed in this study, will validate and corroborate the efficiency and reliability of this multiplex MSP-SNaPshot method for identification of body fluids in forensic investigations.

References


Coolen, M., Statham, A., Qu, W., Campbell, M., Henders, A. & Montgomery, G. et al. (2011). Impact of the genome on the epigenome is manifested in DNA methylation patterns of imprinted regions in monozygotic and dizygotic twins. Public Library of Sciences One, 6(10), e25590.


locus in different Italian populations and in mono- and di-zygotic twins of different age. Ageing, 4(7), 509–520.


cerebellar Purkinje cells and shows a calcium-dependent membrane association. *Neuroscience*, 96(1), 121-129.


URL: http://www.statssa.gov.za/?p=12620

Data accessed: 09/10/2019


Data accessed: 28/09/2019


APPENDICES
APPENDIX A
05 July 2019

Ms S Haripersad (214548055)
School of Life Science
College of Agriculture, Engineering and Sciences
haripersadsumina@gmail.com

Dear Ms Haripersad

Protocol: Development of multiplex Methylation-specific PCR assay and methylation SNaPshot assay for efficient identification of human body fluids
Degree: MSc
BREC Ref No: BE115/19

EXPEDITED APPLICATION: APPROVAL LETTER

A sub-committee of the Biomedical Research Ethics Committee has considered and noted your application received 22 February 2019.

The study was provisionally approved pending appropriate responses to queries raised. Your response received on 10 June 2019 to BREC letter dated 28 May 2019 has been noted by a sub-committee of the Biomedical Research Ethics Committee. The conditions have been met and the study is given full ethics approval and may begin as from 05 July 2019. Please ensure that site permissions are obtained and forwarded to BREC for approval before commencing research at a site.

This approval is valid for one year from 05 July 2019. 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.


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 noted by a full Committee at its next meeting taking place on 13 August 2019.

Yours sincerely,

[Signature]

Prof V Rambiritch
Chair: Biomedical Research Ethics Committee

cc: Postgrad administrator: manjoon@ukzn.ac.za  Supervisor: ghaid@ukzn.ac.za  nichaua@ukzn.ac.za
UKZN BIOMEDICAL RESEARCH ETHICS COMMITTEE

APPLICATION FOR ETHICS APPROVAL
For research with human participants (Biomedical)

INFORMED CONSENT FORM – BLOOD, SALIVA AND VAGINAL FLUID

Title of the research project:
Development of multiplex MSP and methylation SNaPshot assay for efficient identification of human body fluids

Information Sheet and Consent to Participate in Research

Date: January 2019 – January 2020

Good-day Potential Volunteer

My name is Sumina Haripersad. I am an MSc Student from the Department of Genetics at UKZN, Westville Campus. My contact number is 084 533 7014 and email address is 214548055@stu.ukzn.ac.za.

You are invited to consider participating in a study that involves a scientific method which can help to identify different human body fluids. The research will involve first collecting body fluids from participants, and then laboratory work to identify human body fluids such as blood, saliva, semen and vaginal fluid. The study is based on identifying the different body fluids and using the method to identify the fluids in mixtures. The study requires 80 samples (20 of each body fluid i.e. blood, saliva, semen and vaginal fluid). Blood, saliva and vaginal fluid samples are required from volunteers at the King Dinizulu Hospital in Overport, Durban however, semen samples will be obtained from the Aevitas Fertility Clinic, Cape Town. The collection of body fluids will involve obtaining saliva and vaginal fluid using sterile cotton swabs which will be swabbed onto the area. You may perform the collection yourself or if you feel comfortable, a trained nurse will help you. Collection of blood will be by slight pricking of the finger using a new kit for each volunteer and dropping blood onto dried blood spot cards by trained nurses/hospital staff. There is minimal risk or pain with this process. If you do not wish to provide a sample, your treatment at the hospital will not be affected and there will be no consequences. Your participation is completely voluntary.

Should you choose to enroll and remain in the study, all that is expected of you will be to read and sign the Informed Consent and Tissue Storage forms, fill in a brief questionnaire, and provide the samples you wish to provide. You may choose to provide any/all body fluids. Your identity will not be maintained with the sample. No further participation is necessary.

The study does not involve any risks and chances of injury are minimal as sample collection is non-invasive. However, since collection of all samples will be performed at the King Dinizulu Hospital, there are trained nurses and doctors readily available if necessary. The study will provide no direct benefits to participants; however it will benefit forensic sciences to accurately identify human body fluids. This may assist forensic scientists in future during crime scene investigations. This study has been ethically reviewed and approved by the UKZN Biomedical Research Ethics Committee (BE115/19).

In the event of any problems or concerns/questions you may contact the researcher (Sumina Haripersad) at 084 533 7014 and email address is 214548055@stu.ukzn.ac.za. Alternatively, contact 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 penalty or loss of any sort. If you do wish to withdraw, you may contact the Principal Investigator and you will be withdrawn immediately.

You will not incur any costs because of participation in the study. There are no benefits or incentives for participation. Your sample will be labelled according to the sample identity which only indicates your gender and age. Your name will not be associated/stated on the sample, and all information will be kept confidential.

Samples will be stored only from January 2019 – January 2020 at the Genetics Department at UKZN Westville, and all results, data obtained, and electronic data will be password protected. The samples will not be used for any other study. All samples/data will be destroyed once the study is completed at the end of January 2020.

CONSENT (Edit as required)

I have been informed about the study entitled Development of multiplex MSP and methylation SNaPshot assay for Efficient Identification of Human Body Fluids by Sumina Haripersad.
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.
If I have any further questions/concerns or queries related to the study I understand that I may contact the researcher at 084 533 7014 and 214548055@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 (Day/Month/Year)

Signature of Witness Date (Day/Month/Year)
(Where applicable)

Signature of Translator Date (Day/Month/Year)
(Where applicable)

Name of Researcher/person taking the consent:
Signature of Researcher/person taking the consent:
Date: (Day/Month/Year)
QUESTIONNAIRE FOR THE STUDY:
Development of multiplex MSP and methylation SNAPSHOT assay for efficient identification of human body fluids

(Please tick the appropriate boxes ONLY)

1. What is your gender?
   - Male
   - Female
   - Other

2. What is your age?
   - 21-30
   - 31-40
   - 41-50
   - 51-60
   - 60 and over

3. What is your weight (in kilograms)?
   - 40 & below
   - 40-80
   - above 80

4. Are you a South African citizen?
   - Yes
   - No

Your assistance and time is well appreciated. Kindly note that the present study is not for diagnostic purposes. All information disclosed will be kept confidential. Your identity will NOT be maintained with samples you provide.

Miss Sumina Haripersad
MSc Candidate / Primary Investigator
The study is approved by the Biomedical Research Ethics Committee of UKZN. If you have any queries; please contact:

BIOMEDICAL RESEARCH ETHICS ADMINISTRATION
Research Office, Westville Campus
Govan Mbeki Building
University of KwaZulu-Natal
Private Bag X 54001, Durban, 4000
KwaZulu-Natal, SOUTH AFRICA
Tel: 27 31 2602486 - Fax: 27 31 2604609
Email: BREC@ukzn.ac.za
### APPENDIX B

**Table 1:** Concentration of DNA isolated from blood, saliva, semen and vaginal fluid

<table>
<thead>
<tr>
<th>Sample Identity</th>
<th>DNA concentration (ng/μL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood 1</td>
<td>24.2</td>
</tr>
<tr>
<td>Blood 2</td>
<td>24.7</td>
</tr>
<tr>
<td>Blood 3</td>
<td>53.4</td>
</tr>
<tr>
<td>Blood 4</td>
<td>16.8</td>
</tr>
<tr>
<td>Blood 5</td>
<td>42.3</td>
</tr>
<tr>
<td>Blood 6</td>
<td>25.2</td>
</tr>
<tr>
<td>Blood 7</td>
<td>17.1</td>
</tr>
<tr>
<td>Blood 8</td>
<td>40.1</td>
</tr>
<tr>
<td>Blood 9</td>
<td>44.2</td>
</tr>
<tr>
<td>Blood 10</td>
<td>63.9</td>
</tr>
<tr>
<td>Blood 11</td>
<td>62.8</td>
</tr>
<tr>
<td>Blood 12</td>
<td>92.9</td>
</tr>
<tr>
<td>Blood 13</td>
<td>40.1</td>
</tr>
<tr>
<td>Blood 14</td>
<td>140.2</td>
</tr>
<tr>
<td>Blood 15</td>
<td>41.9</td>
</tr>
<tr>
<td>Blood 16</td>
<td>31.8</td>
</tr>
<tr>
<td>Blood 17</td>
<td>45</td>
</tr>
<tr>
<td>Blood 18</td>
<td>18.1</td>
</tr>
<tr>
<td>Blood 19</td>
<td>203.8</td>
</tr>
<tr>
<td>Blood 20</td>
<td>107.2</td>
</tr>
<tr>
<td>Saliva 1</td>
<td>22.6</td>
</tr>
<tr>
<td>Saliva 2</td>
<td>31.2</td>
</tr>
<tr>
<td>Saliva 3</td>
<td>26.7</td>
</tr>
<tr>
<td>Saliva 4</td>
<td>50.5</td>
</tr>
<tr>
<td>Saliva 5</td>
<td>37.6</td>
</tr>
<tr>
<td>Saliva 6</td>
<td>41.2</td>
</tr>
<tr>
<td>Saliva 7</td>
<td>32.2</td>
</tr>
<tr>
<td>Sample</td>
<td>Value</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>Saliva 8</td>
<td>25.8</td>
</tr>
<tr>
<td>Saliva 9</td>
<td>24.9</td>
</tr>
<tr>
<td>Saliva 10</td>
<td>58.8</td>
</tr>
<tr>
<td>Saliva 11</td>
<td>57.5</td>
</tr>
<tr>
<td>Saliva 12</td>
<td>39.5</td>
</tr>
<tr>
<td>Saliva 13</td>
<td>42.6</td>
</tr>
<tr>
<td>Saliva 14</td>
<td>33.8</td>
</tr>
<tr>
<td>Saliva 15</td>
<td>29.6</td>
</tr>
<tr>
<td>Saliva 16</td>
<td>26.8</td>
</tr>
<tr>
<td>Saliva 17</td>
<td>29.7</td>
</tr>
<tr>
<td>Saliva 18</td>
<td>30.2</td>
</tr>
<tr>
<td>Saliva 19</td>
<td>59.0</td>
</tr>
<tr>
<td>Saliva 20</td>
<td>68.9</td>
</tr>
<tr>
<td>Semen 1</td>
<td>68.6</td>
</tr>
<tr>
<td>Semen 2</td>
<td>77.0</td>
</tr>
<tr>
<td>Semen 3</td>
<td>48.3</td>
</tr>
<tr>
<td>Semen 4</td>
<td>58.9</td>
</tr>
<tr>
<td>Semen 5</td>
<td>53.7</td>
</tr>
<tr>
<td>Semen 6</td>
<td>43.4</td>
</tr>
<tr>
<td>Semen 7</td>
<td>27.5</td>
</tr>
<tr>
<td>Semen 8</td>
<td>39.8</td>
</tr>
<tr>
<td>Semen 9</td>
<td>26.0</td>
</tr>
<tr>
<td>Semen 10</td>
<td>36.2</td>
</tr>
<tr>
<td>Semen 11</td>
<td>62.3</td>
</tr>
<tr>
<td>Semen 12</td>
<td>23.7</td>
</tr>
<tr>
<td>Semen 13</td>
<td>30.3</td>
</tr>
<tr>
<td>Semen 14</td>
<td>63.3</td>
</tr>
<tr>
<td>Semen 15</td>
<td>58.2</td>
</tr>
<tr>
<td>Semen 16</td>
<td>31.2</td>
</tr>
<tr>
<td>Semen 17</td>
<td>42.3</td>
</tr>
<tr>
<td>Semen 18</td>
<td>81.6</td>
</tr>
<tr>
<td></td>
<td>Value</td>
</tr>
<tr>
<td>----------------</td>
<td>-------</td>
</tr>
<tr>
<td>Semen 19</td>
<td>23.6</td>
</tr>
<tr>
<td>Semen 20</td>
<td>37.8</td>
</tr>
<tr>
<td>Vaginal fluid 1</td>
<td>29.5</td>
</tr>
<tr>
<td>Vaginal fluid 2</td>
<td>33.3</td>
</tr>
<tr>
<td>Vaginal fluid 3</td>
<td>17.0</td>
</tr>
<tr>
<td>Vaginal fluid 4</td>
<td>116.9</td>
</tr>
<tr>
<td>Vaginal fluid 5</td>
<td>85.9</td>
</tr>
<tr>
<td>Vaginal fluid 6</td>
<td>53.1</td>
</tr>
<tr>
<td>Vaginal fluid 7</td>
<td>162.6</td>
</tr>
<tr>
<td>Vaginal fluid 8</td>
<td>115.7</td>
</tr>
<tr>
<td>Vaginal fluid 9</td>
<td>89.5</td>
</tr>
<tr>
<td>Vaginal fluid 10</td>
<td>66.8</td>
</tr>
<tr>
<td>Vaginal fluid 11</td>
<td>99.2</td>
</tr>
<tr>
<td>Vaginal fluid 12</td>
<td>103.7</td>
</tr>
<tr>
<td>Vaginal fluid 13</td>
<td>67.9</td>
</tr>
<tr>
<td>Vaginal fluid 14</td>
<td>29.5</td>
</tr>
<tr>
<td>Vaginal fluid 15</td>
<td>36.7</td>
</tr>
<tr>
<td>Vaginal fluid 16</td>
<td>47.6</td>
</tr>
<tr>
<td>Vaginal fluid 17</td>
<td>52.2</td>
</tr>
<tr>
<td>Vaginal fluid 18</td>
<td>199.8</td>
</tr>
<tr>
<td>Vaginal fluid 19</td>
<td>44.9</td>
</tr>
<tr>
<td>Vaginal fluid 20</td>
<td>38.6</td>
</tr>
</tbody>
</table>
APPENDIX C

a. Chr7: 149224362-149224571
GGAGTGCGAGAAGACCTACAGCCGTAAGGAGCACCTGCAAGAACCACCCCGGC
TGACACACGGGGAGCGGGCCGCCTTTTCCAATGTGCACTGTGCGGCAAGAGCTTCATCC
GCAAGCAGAACCCTGCTCAAGCACCAGCGCATCCACACGGGCGAGCGGCCCTAC
ACGTGCGGGCGAGGTGCGGCAAGAGCTTCCGCTACAAGGAGTACAGCTCAAGAGGA

b. Chr2: 10419725-10419837
GTCTTGACGGTGATGTCGCCGCCGCCAGCCGCCATGGGCAAACAGAACAGCA
GAACCCGCCAGGGTGCTGCGGCCCGAGGTGCTGCAGGACCTGCGGGAGAACACGGAGTTCACCGGAC
GCGAGCT

c. Chr2: 10419725-10419838
GTCTTGACGGTGATGTCGCCGCCGCCAGCCGCCATGGGCAAACAGAACAGCA
GAACCCGCCAGGGTGCTGCGGCCCGAGGTGCTGCAGGACCTGCGGGAGAACACGGAGTTCACCGGAC
GCGAGCT

d. Chr7: 27291586-27291844
GCCGCTTGCATTTGGGGAGCGGAGCGCCGCCCAGGTGCTTCAAGTGCTCTCTG
GGCCAGGCCCCCTTCCATATGGCTGGCTCTCTGCTGGGGCCCTGGGCTGAGCA
GCGGCTACTTCTCAGAGAGAAATCTTTCTCTTCTCCTCCATTGCAAGATGTCTCTCT
CTTGCTCAGAAAATAACCTCTCTGGTGAGGCGCCGGTCGCTCTCTCTGCGCCCTCC
TCCCGCAGGGCCGCCGCTCGGAGGACCCCGGTGCAGT

e. Chr7: 27291587-27291848
CCCGCTTGGCCATTGGGAGACGCGGCCCAGGTGCTTCAAGTGCTCTCTG
GGCCAGGCCCCCTTCCATATGGCTGGCTCTCTGCTGGGGCCCTGGGCTGAGCA
GCGGCTACTTCTCAGAGAGAAATCTTTCTCTTCTCCTCCATTGCAAGATGTCTCTCT
CTTGCTCAGAAAATAACCTCTCTGGTGAGGCGCCGGTCGCTCTCTCTGCGCCCTCC
TCCCGCAGGGCCGCCGCTCGGAGGACCCCGGTGCAGTTGAG
Figure 1: Genomic reference sequences of MSP amplicons. Amplicon for ZNF282 (methylated and unmethylated primers) (a), HPCAL1 (methylated primers) (b), HPCAL1 (unmethylated primers) (c), cg09765089-231d (methylated primers) (d), cg09765089-231d (unmethylated primers) (e), and cg08792630 (methylated and unmethylated primers) (f). CpG sites analysed by the MSP assay are underlined. CpG sites on the MSP amplicons targeted for SNaPshot design are highlighted in yellow for the MSP primer specific to the body fluid.
Alignment of sequencing data to unconverted genomic reference sequence

Figure 2: Alignment of HPCAL1 genomic reference sequence (unmethylated primer pair) with sequenced saliva MSP amplicon. The single CpG site targeted for SNaPshot primer is underlined in red. The site is identified as unmethylated due to the presence of a T residue. Mismatches/gaps in the alignment are denoted by dashes (-).

Figure 3: Alignment of cg09765089-231d genomic reference sequence (methylated primer pair) with sequenced vaginal fluid MSP amplicon. The single CpG site targeted for SNaPshot primer is underlined in red. The site is identified as methylated due to the presence of a C residue. Mismatches/gaps in the alignment are denoted by dashes (-).
Figure 4: Alignment of cg08792630 genomic reference sequence (methylated primer pair) with sequenced blood MSP amplicon. The single CpG site targeted for SNaPshot primer is underlined in red. The site is identified as methylated due to the presence of a C residue. Mismatches/gaps in the alignment are denoted by dashes (-).

Figure 5: Alignment of ZNF282 genomic reference sequence (unmethylated primer pair) with sequenced semen MSP amplicon. The single CpG site targeted for SNaPshot primer is underlined in red. The site is identified as unmethylated due to the presence of a T residue. Mismatches/gaps in the alignment are denoted by dashes (-).
**APPENDIX D**

*Figure 6*: Methylation profiles of the unmethylated primer set on saliva. Black and clear circles indicate a methylated and unmethylated CpG site, respectively. The black line represents the reference amplicon sequence to which sequencing products were aligned to. Blank spaces indicate unknown status at that site. Numbers above represent the CpG site number and numbers below represent the nucleotide position of that specific CpG site in the amplicon. Representative of HPCAL1 reaction on ten saliva samples in preliminary tests.
APPENDIX E

**Figure 7:** Positive control according to ABI PRISM® SNaPshot™ Multiplex Kit (Applied Biosystems)

**Figure 8:** Negative control according to ABI PRISM® SNaPshot™ Multiplex Kit (Applied Biosystems)

**Figure 9:** GeneScan-120 LIZ internal sizing standard
<table>
<thead>
<tr>
<th>SNaPshot™ Multiplex Control Primer Mix</th>
<th>Length of Final Products (nt)</th>
<th>Signal Colour</th>
<th>Heterozygosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>20A</td>
<td>33</td>
<td>Green</td>
<td>Homozygote</td>
</tr>
<tr>
<td>28G/A</td>
<td>34</td>
<td>Blue/Green</td>
<td>Heterozygote</td>
</tr>
<tr>
<td>36G</td>
<td>44</td>
<td>Blue</td>
<td>Homozygote</td>
</tr>
<tr>
<td>44T</td>
<td>52</td>
<td>Red</td>
<td>Homozygote</td>
</tr>
<tr>
<td>52C/T</td>
<td>57</td>
<td>Black/Red</td>
<td>Heterozygote</td>
</tr>
<tr>
<td>60C</td>
<td>65</td>
<td>Black</td>
<td>Homozygote</td>
</tr>
</tbody>
</table>
APPENDIX F

Optimization of enzyme concentration for purification:

Figure 10: Electropherogram of SNaPshot reaction purified by 4 U SAP and 40 U ExoI. The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green and blue peaks indicated an unmethylated and methylated CpG site (reverse primer), respectively.

Figure 11: Electropherogram of SNaPshot reaction purified by 5 U SAP and 8 U ExoI. The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green and blue peaks indicated an unmethylated and methylated CpG site (reverse primer), respectively.
APPENDIX G

Optimization of template type – diluted and undiluted PCR products (Figure 8 and 9), as well as optimization of extension time (Figure 9 and 10).

**Figure 12:** Electropherogram of SNaPshot reaction with diluted PCR products. The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green and blue peaks indicated an unmethylated and methylated CpG site (reverse primer), respectively.

**Figure 13:** Electropherogram of SNaPshot reaction with undiluted PCR products and an extension time of 30 seconds. The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green and blue peaks indicated an unmethylated and methylated CpG site (reverse primer), respectively.

**Figure 14:** Electropherogram of SNaPshot reaction with undiluted PCR products and an extension time of one minute. The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green and blue peaks indicated an unmethylated and methylated CpG site (reverse primer), respectively.
APPENDIX H

Optimization of primer concentration in SNaPshot reaction

Figure 15: Electropherogram of SNaPshot reaction with 0.4 μM of each SNaPshot primer run with an extension time of 2 minutes. The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green and blue peaks indicated an unmethylated and methylated CpG site (reverse primer), respectively.

Figure 16: Electropherogram of SNaPshot reaction with 0.5 μM of each SNaPshot primer run with an extension time of 2 minutes. The y-axis represents the RFU of the emitted dye and the x-axis represents the SNaPshot amplicon size in nucleotides. Green and blue peaks indicated an unmethylated and methylated CpG site (reverse primer), respectively.
### Thesis

**Originality Report**

<table>
<thead>
<tr>
<th>Similarity Index</th>
<th>Internet Sources</th>
<th>Publications</th>
<th>Student Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>13%</td>
<td>0%</td>
<td>0%</td>
<td>13%</td>
</tr>
</tbody>
</table>

**Primary Sources**

1. **Submitted to University of KwaZulu-Natal**  
   Student Paper  
   13%

Exclude quotes: Off  
Exclude bibliography: On  
Exclude matches: < 8%