Title: Differentially methylated Embryonal Fyn-associated Substrate (EFS) gene as a blood-specific epigenetic marker and its potential application in forensic casework

Authors: Athina Vidaki, Cecilia Johansson, Federica Giangasparo

PII: S1872-4973(17)30094-7
DOI: http://dx.doi.org/doi:10.1016/j.fsigen.2017.04.010
Reference: FSIGEN 1706

To appear in: *Forensic Science International: Genetics*

Received date: 15-10-2016
Revised date: 8-4-2017
Accepted date: 14-4-2017

Please cite this article as: Athina Vidaki, Cecilia Johansson, Federica Giangasparo, Differentially methylated Embryonal Fyn-associated Substrate (EFS) gene as a blood-specific epigenetic marker and its potential application in forensic casework, Forensic Science International: Geneticshttp://dx.doi.org/10.1016/j.fsigen.2017.04.010

Differentially methylated Embryonal Fyn-associated Substrate (EFS) gene as a blood-specific epigenetic marker and its potential application in forensic casework

Athina Vidaki [1,*], Cecilia Johansson, Federica Giangasparo, Denise Syndercombe Court

Department of Pharmacy and Forensic Science, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London, UK

*Corresponding author: Dr Athina Vidaki, E-mail: athina.a.vidaki@kcl.ac.uk.

[1] Present address: Department of Genetic Identification, Erasmus MC University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands, a.vidaki@erasmusmc.nl

Highlights

- A novel blood-specific differentially methylated locus (EFS) was investigated.
- Most CpGs, especially CpG4, was found highly methylated in whole blood.
- EFS methylation has the potential to differentiate between whole and menstrual blood.
- The designed bisulfite pyrosequencing assay is both accurate and reproducible.
- The method's applicability to casework was demonstrated via aged and mixed stains.

Abstract

DNA methylation patterns have the ability to reveal the activities of genes within a certain tissue at a particular time point. Tissue-specific DNA methylation patterns have been previously investigated for their applicability in the identification of forensically relevant body fluids, however there is still a lack in robust markers. While following a genome-wide scale investigation has a great potential to reveal useful tissue-specific changes, a gene-targeted approach can also lead to significant outcomes, especially in genomic locations not included in the genome-wide experiments. In this study, the potential of the candidate embryonal Fyn-associated substrate (EFS) gene for the positive identification of whole blood was investigated. For this purpose, the methylation profile of a selected genomic region containing a total of 10 CpG sites was analysed in 124 individuals via bisulfite pyrosequencing. Volunteers donated various forensically relevant tissues, including whole blood, saliva, seminal fluid, vaginal fluid and menstrual secretion. Whole blood showed the highest levels of DNA methylation (mean=0.67), while semen samples were found to be very low methylated (mean=0.06). The remaining tissues demonstrated partial mean methylation levels; more specifically, saliva – 0.43, vaginal fluid – 0.22 and menstrual blood – 0.22. One out of the 10 analysed CpG sites, CpG4, showed to be more robust, resulting in not only the highest methylation difference between blood and the rest of the tissues, but also the lowest inter-individual methylation difference. The proposed pyrosequencing assay was found to be accurate, linear and reproducible. Lastly, the method's applicability to forensic casework was assessed via the analysis of very old bloodstains stored up to 18 years, blood DNA samples stored long-term up to 9 years, mixed stains as well as other 'forensic-like' samples. In the majority of cases the expected methylation ratios were obtained indicating a stable DNA

methylation pattern, however caution is necessary when analysing low quantity and/or quality samples due to potential stochastic effects. Future validation experiments can shed more light into the usefulness of EFS locus as a promising blood-specific epigenetic marker.

Keywords: forensic tissue identification; ; ; ; ; , blood, epigenetics, DNA methylation, bisulfite pyrosequencing, EFS gene

1. Introduction

Various cellular decisions including survival, growth and differentiation are controlled by particular gene expression patterns, which are further regulated by changes in the epigenetic state of 'key' genes. The possibility of quantifying the DNA methylation levels of specific genes is of interest in a broad range of scientific and medical disciplines. It is known that the epigenome acts as an interphase between the fixed genome and the dynamic environment, and can change according to the specific needs of a cell [1]. Therefore, DNA methylation patterns could reveal the activities of genes within a certain tissue at a particular point in time. Tissue-specific methylation patterns have been investigated either by analysing specific gene loci [2], entire chromosomes [3] or on a genome-wide scale [4]. There are numerous studies that have looked at differential methylation patterns across various tissues in order to identify a comprehensive genome-wide set of tissue-specific differentially methylated regions (tDMRs) that could play a role in cellular identity as well as the regulation of tissue-specific genome function [2-7]. Although epigenetics, specifically through the study of differentially methylated CpG sites, is a well-established approach in gene expression studies, from a forensic standpoint the application of epigenetic markers and the analysis of DNA methylation patterns is a relatively new field.

Over the last few years there have been several studies published that attempt to utilise epigenetic markers in the field of forensic tissue identification [8-14]. It is evident why a DNA-based approach for identifying the cellular origin of a tissue stain can be beneficial over tissue-specific mRNA typing, which has been considered as confirmatory testing for forensic purposes. It has the potential to not only provide a direct link to the biological material used in personal identification, but also benefit from DNA's greater stability. It also has applicability in cold cases, where often minute stains or only DNA extracts may have been retained over time; however, we do acknowledge that this depends on sample availability and extraction method used. These studies of epigenetic markers have usually either validated previously reported tissue-specific markers or have attempted to discover new

CpG sites through genome-wide methylation analysis. Although some of these techniques seem to be quite sensitive [15] and therefore, suitable for use in forensic specimens, marker specificity issues have been reported. With the exception of semen identification, which has been proven relatively easy to differentiate due to its distinct cellular and biochemical nature [16-20], finding highly robust markers for more complex biological fluids such as blood, menstrual blood or vaginal fluid has been particularly challenging. It is important to appreciate that, whereas semen mainly consists of one cell type, meaning sperm cells (however, in certain cases it can occasionally contain other cell types such as white blood cells [21]), other body fluids like blood comprise many different types of cells, each able to demonstrate a distinct DNA methylation profile. DNA methylation analysis has already been proposed as a tool for cell typing in cell therapeutic approaches. For example, Baron *et al* identified panels of cell type-specific differentially methylated gene regions, which allowed for accurate identification and quantification of subpopulations in cell cultures [22].

Venous blood is of particular importance in forensic casework as it is considered the most common recovered biological material and there are cases where its origin can be vital to the outcome of a case. For example, current techniques would not normally distinguish venous from menstrual blood. While both would give positive reactions using presumptive chemical tests, even tissue-specific mRNA typing could still raise doubts about the origin, mainly due to the co-existence of blood peaks when menstrual blood is present [23, 24]. In DNA methylation-based studies in forensic identification for blood, potential markers such as cg06379435 [25, 26], cg08792630 [26], cg26285698 [12], cg04011671 and cg18454288 [14] have been identified through genome-wide methylation arrays, such as the Illumina Human Methylation 450K bead array. Others have focused more on a candidate gene approach in order to identify blood-specific differentially methylated regions, such as seen in the reported potentially blood-specific tDMR in the gene GAS2L1 [27]. However, it is evident that the discovery and identification of more forensically relevant tissue-specific CpG sites is necessary prior to the development of an applicable tissue-ID multiplex protocol.

In this study, and following a literature review of potential genes that demonstrate a blood-specific DNA methylation pattern, the embryonal Fyn-associated substrate (*EFS*) gene was selected as a likely candidate. Neumann *et al* analysed the methylation status of the *EFS* gene, and more specifically the region Chr14: 23,835,859-23,835,970 (GrCh37/hg19) including a total of 11 CpG sites, in various tissues including blood, buccal cells, sperm, brain and reported tissue-specific methylation patterns [28]. They found that *EFS* is highly methylated in blood, completely unmethylated in sperm and partially methylated in buccal cells. Therefore, this marker could be a good candidate for blood detection and the aim of this study was to investigate its potential and apply it to the analysis of other forensically relevant tissues.

2. Material and Methods

2.1. Sample Collection

Biological fluid samples used in this study were collected following full ethical approval by the appropriate Research Ethics Subcommittee at King's College London (BDM RESC 13/14-30). Full informed consent was obtained from all donors prior to sample collection. In total, 124 volunteers of both sexes and various ethnic backgrounds with ages ranging from 23 to 68 years old participated in this study. Individuals had the choice to donate one or more body fluids/tissues including whole blood, saliva, buccal cells, seminal fluid, vaginal fluid and menstrual secretion. Up to 20 ml of whole blood were collected by a trained phlebotomist in a clean clinical environment. All other body fluid samples were collected either using a cotton swab or a suitable receptacle by the participants in the privacy of their homes. Information such as subject's gender, geographic ancestry, and age were also recorded.

2.2. Validation samples

For validation purposes, a set of aged stains or DNA samples were used which had been previously collected for research purposes. The set included five blood samples stored for 9-18 years in the dark and at room temperature, and five blood DNA samples stored for 1-9

years at -20ºC. Using freshly collected samples, the following body fluid stains were also prepared: (a) mock casework blood samples including 1µl on a piece of fabric from a pair of denim jeans, 3µl on shirt which was then washed, 3µl on a towel and 5µl on tissue paper, (b) mixed stains including 1µl blood:1µl semen, and 2µl blood:1µl saliva on swabs, (c) artificially-degraded blood samples by exposing 1µl under UV for 0, 10, 30, 60, 90, 120 and 240 minutes, and lastly, (d) 5µl of blood on fabric stored at various temperatures (-20ºC, 4ºC, 25ºC, 37ºC, outdoors) for a week. All fresh stains were left at room temperature to dry overnight before analysis.

2.3. DNA extraction and quantification

DNA was isolated using the QIAamp DNA Investigator kit (QIAGEN, Hilden, Germany) using the appropriate protocol depending on the tissue and following the manufacturer's recommendation. To assess the quantity of the resulting solution, DNA samples were quantified using the Quantifiler® Human DNA investigation kit (ThermoFisher Scientific, Massachusetts, United States) according to the standard suggested protocol.

2.4. Bisulfite Conversion

DNA samples were treated with sodium bisulfite, which converts unmethylated cytosines into uracil, while the methylated ones remain unchanged. In this study, the EZ DNA methylation™ kit (ZymoResearch, Irvine, United States) was used. For the purpose of this study and depending on the experiment, 1-100ng or 10µl of each DNA sample were converted using the conversion conditions suggested by the manufacturer. Bisulfite treated DNA samples were eluted in 10-20µl of elution buffer and were stored at -20ºC for up to one month. Together with the samples, pre-defined DNA methylation controls (0-100%) (EpigenDx, Hopkinton, United States) were also used in order to assess both bisulfite conversion efficiency and the linearity of methylation quantification.

2.5. Bisulfite PCR assay

This assay was designed to specifically amplify bisulfite-treated DNA using the BiSearch online tool (http://bisearch.enzim.hu/), which is able to address common issues seen in

bisulfite PCR such as low efficiency, mis-priming or non-specific amplification. The EFS assay includes a PCR primer set (the forward and biotin-labelled reverse); details regarding primer sequences and PCR product length are shown in Table 1. Due to location and assay design restrictions, only 10 out of the reported 11 CpG sites were investigated. Briefly, each PCR reaction consisted of 12.5μl of ZymoTaq PreMix (ZymoResearch), 1μl of 25mM $MgCl_2$ for a final concentration of 2.75mM (since the ZymoTaq™ Premix also contains 1.75mM $MgCl_2$), 1μl of each PCR primer (for a final concentration of 0.4μM), 1μl of bisulfite DNA template and 8.5μl of nuclease-free water, for a total reaction volume of 25μl. The thermocycling program used was: 95°C for 10 minutes, followed by 45 cycles of 94°C for 30 seconds, 50°C for 40 seconds, 72°C for 40 seconds, and a final extension step of 72°C for 15 minutes. Following amplification, the quality of PCR products was assessed on a 2% agarose gel.

Table 1. Designed bisulfite PCR assay

*The reverse primer is biotin-labelled at its 5'-end.

Assay	CpGs	Primer Sequence (5' → 3')		Length (bp)	%GC	Converted Cs	PCR Product (bp)
EFS	10	F	GGTTTTTTTTTTATTAGTTT	20	15.0	9	195
		R*	CTTCATATTATCACTAAAACC	21	28.6	5	

2.6. Pyrosequencing analysis

PCR products were converted to single-stranded via biotin-streptavidin selective binding: 15μl of biotin-labelled PCR products were mixed with 3μl of Streptavidin Sepharose High Performance Beads (GE Healthcare, Illinois, United States), 37μl of PyroMark Binding Buffer (QIAGEN) as well as 25μl of distilled water for a total volume of 80μl. The solutions were then vortexed at 1,000rpm for 30-40 minutes using a clear non-skirted 96-well plate to allow for efficient binding of the PCR products to the beads. Afterwards, the beads were isolated and captured utilising a Vacuum Prep Workstation (QIAGEN) according to the manufacturer's recommendations. Sequencing reactions were performed by mixing the

single-stranded templates with 11µl of PyroMark Annealing buffer (QIAGEN) and 1µl of the appropriate 10µM sequencing primer. In order to accommodate all 10 CpG sites, two primers were used: sequencing primer 1 - TTGTTTTTTTTATGGGAGGG (5'→3') analysing the chromosomal region Chr14: 23,835,857-91 (GrCh37/hg19) and sequencing primer 2 – GGTTTTTAGTAGAGTTTTTTTA (5'→3') analysing the adjacent region Chr14: 23,835,902-51. To ensure complete denaturation of DNA templates, these were heated at 80°C for 3 minutes. Reactions were left at room temperature for 5 minutes in order for the sequencing primer to bind before loading them on a PyroMark MD Pyrosequencer (QIAGEN). All four nucleotides together with the enzymes and substrates (PyroMark Gold Q96 reagents, QIAGEN) were placed into the instrument before analysis following the manufacturer's instructions.

2.7. Data analysis

Only pyrograms™ passing the instrument's quality control were used for analysis; this included the detection of desired peaks (no signals in dead injections) as well as detected bisulfite conversion rates of >95%. Bisulfite conversion rates were calculated using the peak heights of the C/T signals from built-in, non-CpG cytosines in the sequence. Similarly, for each CpG site the degree of methylation was measured as a frequency of C/T signals in the form of peak heights by the dedicated CpG methylation software, PyroMark CpG SW 1.0 (QIAGEN). DNA standards of known methylation levels (0%-100%) were used to assess the linearity of methylation quantification; the observed methylation ratio was plotted against the expected and the best-fitted regression line (linear) was chosen. Furthermore, to assess the reproducibility of quantification of the proposed pyrosequencing® assay, 20 blood samples were bisulfite-treated and amplified in triplicate using the EFS assay. For this experiment, only 10ng of DNA were used (instead of 100ng used in the specificity experiment) to mimic more the amounts obtained from forensic-type blood stains and to assess the method's reproducibility in sub-optimal conditions. The mean and standard deviation of the observed

methylation values were calculated for each sample and for each CpG site. Detected methylation data distribution was visualised by employing box-and-whisker plots. Student t-tests (assuming equal variance of DNA methylation amongst body fluids/tissues) were applied to assess whether tissue-specific DNA methylation patterns differ in a statistical significant way. This analysis was performed by comparing the levels of DNA methylation between body fluids/tissues for each CpG site. For data analysis, the IBM SPSS software (v.22, New York, United States) was employed.

3. Results

3.1. Accuracy and linearity of methylation quantification

Non-linear methylation quantification is common in DNA methylation studies, often due to potential different amplification efficiencies of the unmethylated and methylated allele during PCR [29, 30]. It has been previously suggested that careful primer design and extensive optimisation of the PCR step in terms of annealing temperature and $MgCl_2$ concentration might be sufficient to correct for this amplification bias [31, 32]. Here, following the optimisation of the EFS bisulfite pyrosequencing assay 100ng of each DNA methylation control (0-100%) were analysed in duplicate to assess its accuracy and linearity. For this purpose, the mean and standard deviation of every control and for each CpG site was calculated. The average standard deviation obtained from the duplicated samples was 0.05; however it was much higher in certain individual CpG sites, such as CpG 8, and in the 0.75 methylation control. Although a standard deviation of 0.05 could be considered as an acceptable degree of variability, a methylation difference of 0.26 for the same CpG site of the same sample was considered quite high and it is thought that this could be due to events during amplification. In general, large differences between CpG sites were sometimes observed, especially in the case of the highly methylated control, which resulted in a range of 0.43 - 0.92 detected methylation [Figure 1a]. According to the manufacturer, this particular control demonstrates a genome-wide >0.85 methylation; however, variations in individual CpG sites cannot be

excluded and can explain this wide range. Therefore to build the standard curve of methylation quantification, the expected methylation values were 'normalised' using the values obtained for the non-methylated and methylated controls [Figure 1b].

(b)

<InlineShape1>

(a)

<InlineShape2>

(a) Box plot showing the minimum, first quartile, median, third quartile and maximum methylation levels detected for each methylation standard taking into account all ten CpG sites, (b) Standard curve showing the observed *vs*. expected methylation ratio of the average methylation per standard after normalisation. Error bars correspond to standard deviation.

3.2. Tissue-specific *EFS* methylation

As mentioned above, Neumann *et al* [28] detected differential methylation patterns of the *EFS* gene when analysing various human tissues including the forensically relevant tissues: blood ("full" methylation), sperm ("no" methylation) and buccal cells ("partial" methylation). In order to apply *EFS* differential methylation patterns for the detection of whole blood in a forensics, various forensically relevant tissues also needed to be investigated; therefore, a total of 77 body fluid samples (20 blood, 10 semen, 16 saliva, 10 buccal, 10 vaginal fluid and 11 menstrual blood samples) were analysed. DNA samples were extracted and quantified; the mean (standard deviation) DNA yield per µl of starting material in liquid samples was as follows: 17 (7) ng for blood, 62 (56) ng for semen and 14 (31) ng for saliva. In relation to material collected on swabs, a buccal swab yielded an average of 2,802 (1,615) ng while a

vaginal or menstrual secretion swab generated 2,247 (1,542) ng and 2,657 (2,487) ng of DNA respectively.

For each sample, 100ng of DNA was bisulfite-treated and amplified using the proposed EFS PCR assay. All pyrograms™ passed the quality control of the instrument (expected sequence pattern, peak height >70 relative light units, rlu) and the average bisulfite conversion rate of the eight built-in conversion controls (non-CpG cytosines included in the sequence) was 94 ± 5%. In general, it was noticed that the obtained individual conversion rates decreased only slightly as the sequencing reaction progressed, possibly due to technical reasons (previously unincorporated cytosines); however, overall the measured bisulfite conversion rates were satisfactory and in agreement with what the manufacturer suggests.

In this analysis and for each CpG site included in the EFS assay we considered two aspects: (1) the distribution of DNA methylation levels within a tissue, and (2) the difference of DNA methylation levels between different tissues and for each CpG site. While we consider that the DNA methylation distribution should be as small as possible to indicate reproducible and robust DNA methylation profiles, the difference between tissues should be as large as possible to allow for accurate tissue identification. Firstly, to assess these aspects box-and-whisker plots were drawn for each CpG site and body fluid/tissue [Figure 2]. As shown in Figure 2, the average methylation proportion in blood was the highest amongst the body fluids examined (0.67 ± 0.16), while semen was found to be almost completely unmethylated (0.06 ± 0.04) in all samples, with the exception of two samples, where the mean methylation was 0.67 and 0.54 respectively (shown as outliers – outside the range ‘mean±standard deviation’) [Figure 2b]. It is believed that this could be due to either natural inter-variability in methylation levels or possible presence of blood in semen. The remaining tested tissues demonstrated partial methylation levels (saliva - 0.43 ± 0.12, buccal cells - 0.26 ± 0.1,

vaginal fluid - 0.22 ± 0.07 and menstrual blood - 0.22 ± 0.05) [Figure 2c-f]. Interestingly, there were also two samples of menstrual blood that showed a slightly higher methylation level (5-15%) for a subset of the investigated CpG sites, which we believe could be due to natural inter-individual variation or the effect of the day of the menstrual cycle that the samples were collected. Furthermore, the presence of circulatory blood in menstrual blood as a result of trauma of some kind cannot be excluded. Overall, there was an inter-individual methylation range of 0.33 ± 0.1 when taking into account all samples at each CpG site, with saliva showing the highest mean range of 0.5. This range corresponds to the observed inter-individual variability; therefore the lower it is the more robust the proposed CpG site is considered.

Boxes represent the first and third quartile while the horizontal line represents the median value. Error bars correspond to the minimum and maximum detected methylation value. Outliers are shown as red ‘x’ dots.

Moreover, we aimed to 'translate' these differences in DNA methylation levels into statistically significant differences that can be used for tissue identification. For this reason, we performed independent two-sample t-tests (considering unequal sample sizes but equal DNA methylation variance) by comparing the results between all body fluids and tissues [Supplementary Table S1]. Despite the large inter-individual variability in certain CpGs/tissues or the small DNA methylation differences of certain CpGs between some tissues, the results of this statistical analysis are promising. When comparing the DNA methylation results of blood with all other tissues, a highly significant difference is observed for all CpG sites ($p<0.001$) [Table S1]. Also, the same is observed when comparing the very low methylation profiles of semen samples with the other tissues (despite the two outliers). However, the differences are not so clear and significant when tissues of epithelial origin are compared. For example, while saliva samples show significantly different DNA methylation profiles in comparison to buccal, vaginal and menstrual blood cells, the same cannot be concluded when comparing the latter with each other [Table S1]. We believe that these results are in agreement with these tissues' cell type composition, and can also be seen in Figure 2 as most samples demonstrate low to intermediate DNA methylation levels for all tested CpGs (<0.5).

Considering each CpG site separately, it was observed that CpG 4 seemed to be more promising as a blood-specific marker. This is due to the fact that this CpG site not only shows the greatest methylation difference between blood (>0.8) and the rest of the tissues (<0.8) (with the exception of the two semen outliers), but also the strongest average p-value when taking into account all blood to non-blood tissue comparisons ($p=3.88E-11$). Considering the limitations of existing methods including low sensitivity and specificity in some instances, a DNA methylation-based approach for the identification of blood using the proposed marker (together with others from the literature) can be very beneficial either as a complementary method or as a stand-alone assay (and particularly in cases where only

DNA is available). However, caution is needed when analysing saliva or other body fluids that can contain leucocytes, as this could result in obtaining 'altered' DNA methylation values [33].

3.3. Inter-individual variability of *EFS* methylation

To further test potential inter-individual variability of *EFS* methylation in blood, an independent cohort of blood samples was analysed. To account for possible age-, ethnicity- and gender-associated effects, this set included a total of 47 blood samples from female and male individuals aged 20-70 years from different ethnic backgrounds [Supplementary Figure S1a-b]. The same experimental conditions employed in the specificity experiment were used (100ng DNA) and only the first four CpG sites were analysed using the sequencing primer 1. The previously obtained levels in blood [Figure 2] were confirmed in this experiment, with no significant methylation differences and no outliers in blood. As shown in Figure S1c, CpG 4 was again confirmed to be the most useful marker since it not only demonstrates the highest level of methylation in blood (>0.8) but also the lowest variability; the median (horizontal line) is close to 1 (0.98). In contrast, CpG 1 demonstrated a large variation amongst the blood samples (0.2 - 0.9), which does not support its use as a blood-specific marker. Lastly, there was no significant correlation between methylation and age, ethnicity or gender ($p>0.05$).

3.4. Applicability to forensic casework

It is essential that the observed methylation status of the *EFS* gene when analysing freshly collected body fluids is also detected in samples that are of low quality and/or quantity. To assess the stability of *EFS* methylation as well as its applicability as a blood-specific marker in forensic casework, a set of mock casework samples were prepared and analysed. One of the most significant advantages of applying DNA methylation profiling for the identification of tissues is its potential applicability in cold cases. In most of these cases, forensic tissue

typing using mRNA profiling cannot be employed either because in stored stains mRNA molecules significantly degrade over time or because minute samples or only DNA samples are kept (although we understand that this depends on the case circumstances and extraction method used). To assess the stability of *EFS* methylation, a total of five blood stains stored at room temperature protected from light for 9-18 years as well as five blood DNA samples stored at -20^{o}C for 1-9 years were analysed. In all cases, the expected methylation ratios for the investigated CpG sites were obtained (CpG 1 – 0.72 ± 0.05, CpG 2 – 0.63 ± 0.05, CpG 3 – 0.72 ± 0.05 and CpG 4 – 0.94 ± 0.08), highlighting the stability of DNA methylation at these sites.

Furthermore, depending on case circumstances, forensic specimens could be mixed or degraded due to exposure to sunlight, high temperature and adverse weather conditions. In an attempt to recreate 'forensic-like' scenarios, various stains were prepared as described in the methods section and their methylation level in the first four EFS CpG sites was quantified. The whole swab or stain was used for DNA extraction and eluted in 10µl; then, the entire solution of extracted DNA was used for bisulfite treatment and converted DNA was also eluted in 10µl. Briefly, all stains stored at various temperatures for a week yielded the 'expected' blood methylation pattern indicating that under these conditions the methylation of *EFS* gene is stable (for example, we obtained >0.90 methylation for all samples stored at -20^{o}C, 4^{o}C, 25^{o}C, 37^{o}C and outdoors). This is very important as crime stains are often exposed to sunlight, so it is encouraging for DNA methylation-based applications to be able to prove that methyl groups are stable under these temperatures. On the other hand, considering the artificially UV-degraded samples, the 'expected' methylation pattern was observed for all bloodstains incubated under UV for up to 90 minutes (no statistically significant difference compared to the non-UV-treated sample, $p>0.05$). However, the stain that was UV-degraded for 120 minutes resulted in a completely non-methylated profile. This could indicate that the methylation levels could have been altered due to the UV light.

Potential stochastic effects cannot also be excluded, as it is known to be common in highly degraded stains. In fact, a likely explanation is that methylated cytosine were deaminated following the UV exposure, which were then converted to thymines, hence showing as unmethylated sites. This effect of specific intensities and wavelengths of UV light on DNA methylation patterns could be further explored in the future. Finally, the stain that was incubated for 4 hours under UV light was too degraded to produce an amplicon ('unsuccessful' STR profiling of this sample also confirmed the extent of its degradation, data not shown).

Additionally, although bloodstains on the towel and tissue paper gave successful pyrograms™, the methylation obtained from the washed bloodstain was very low (0.08). The stained shirt was washed in a public washing machine; therefore potential 'contamination' with other body fluids/tissues or the effect of chemical reagents could explain this result. These factors should be taken into account when interpreting methylation profiles from crime scene stains. Also, a reduced methylation was obtained from the stain on the denim jeans (average of 0.33), which, in this case, could be due to the effect of dyes on the analysis. As all these samples were generally of low-quantity/quality DNA, we cannot exclude that a lower DNA template input could partially explain the occasionally observed lower DNA methylation levels. Nevertheless, all pyrograms™ passed the built-in quality controls in terms of bisulfite conversion and peak quality, however, the peak heights were generally just above the threshold of 70rlu.

Lastly, blood, semen and saliva were used to generate two mixed stains; the obtained methylation values of both single-source and mixed stains using CpG 4 are presented in Figure 3. The potential difficulty of mixture analysis is considered one of the main drawbacks of DNA methylation-based tissue identification due its quantitative (value) rather than

qualitative (peaks) nature. Hence, if applicable in the future, it should be used in combination with presumptive testing (if sufficient casework stain is available), mRNA profiling (if mRNA is available) and also, DNA profiling, to initially assess the number of potential contributors. It was, therefore, important to assess how this assay performed when analysing mixed stains. As illustrated, using the observed methylation of single-source tissues, the 'expected' methylation value was calculated. Although in both cases the observed methylation ratio was not significantly different from the 'expected' value, correction via the equation of the linear regression line in the standard curve for CpG 4 [Figure 3b] improved the accuracy of the results. These results are very encouraging; however, the analysis of more mixed stains generated in various ratios is necessary before making conclusions about its reliability in body-fluid mixtures.

(b)

<InlineShape3>

(a)

<InlineShape4>

Samples		**CpG 4**
Blood	Observed	0.97
Semen	Observed	0.00
Saliva	Observed	0.51
Blood:Semen (1:1)	Expected	0.49
	Observed	0.61
	Corrected	0.53
Blood:Saliva (2:1)	Expected	0.82
	Observed	0.91
	Corrected	0.84

(a) Observed methylation for all individual and mixed samples. Individual methylation values were used to calculate the expected values for the mixed stains. (b) Standard curve for EFS

CpG 4 showing the observed *vs*. expected methylation ratios of pre-defined DNA methylation controls. The equation of the fitted linear regression line was used to 'correct' the observed methylation for mixed stains.

4. Discussion

Neumann *et al* [28] detected differential methylation patterns of the *EFS* gene when analysing various human tissues including blood, sperm and buccal cells. It is known that *EFS* encodes for a protein playing a role in coordinating cell adhesion via tyrosine-kinase-based signalling. EFS has been recognised to be a member of the CRISPR-associated (CAS) protein family but little is known regarding its function. It has been shown to contain a Src homology 3 (SH3) domain and has already been shown to interact with Src-family kinases in mice [34]. Research has also suggested that it is an adapter protein that is regulated via phosphorylation but has no enzymatic activity [35]. Additionally, its methylation mechanism indicates EFS involvement in the differentiation of the hematopoietic cell lineage (T-lymphocyte regulation). Although it has not been linked to cancer to date, other members of the CAS family are known oncogenes acting as prognostic markers of metastasis [36].

The reported blood-specific region of the EFS gene consisting of ten CpG sites was further investigated using various forensically relevant tissues and its potential for blood identification was confirmed through the detection of the previously reported high methylation levels in this tissue. While buccal samples gave a lower methylation range than saliva, the latter produced a wider range among individuals, which could be potentially be supported by the presence of leukocytes in saliva due to either immunological conditions or to gum bleeding. Additionally, it should be noted that while in semen the investigated sites were completely unmethylated (<0.2), in 20% of the analysed samples we obtained much higher levels for all CpG sites at the investigated *EFS* locus (0.3-0.8 more methylation depending on the CpG site) [Figure 2b]. This could once again be supported by the

presence of leukocytes in semen, which is not uncommon and it could be attributable to various reasons such as sexually transmitted infections or indicate an unknown pathological condition [21, 37]. Interestingly, these two semen samples in particular, have also previously resulted in 'odd' DNA methylation results for other tested genomic loci [38], perhaps indicating a wider, 'abnormal' DNA methylation profile. To further assess the frequency of these outliers, it is necessary that a greater number of semen and other samples need to be explored in future studies. These observations and the possibility of observing natural inter-individual methylation variation among individuals should be taken into account during interpretation via a suitable statistical approach.

Nevertheless, one of the analysed CpG sites (CpG 4) was found to be the most useful marker for the identification of blood as it demonstrated the highest average methylation difference (at least >0.3) between blood and other body fluids. Ideally, in order to be able to report with very high certainty that a particular tissue in present, the desired difference of a tissue compared to others should display an 'on-off' methylation; however this is not always possible because of the complexity of tissue constituents. Given that the mean difference in methylation between blood and menstrual blood in some previously proposed tissue-specific sites is rather low (for example in markers cg03363565 – 0.19 and cg09696411 – 0.33 [12], cg06379435 – 0.33 and cg01543184 – 0.30 [25]), the mean detected methylation difference for CpG4 in this assay between blood (0.94) and menstrual blood (0.25) should not be overlooked as a potential discriminating marker.

Additionally, the assay successfully passed an initial validation phase showing not only that methylation levels among a total of 65 blood samples were consistent but also that such an assay can be applied in aged stains (up to 18 years at room temperature and protected from light), stains stored at a wide range of different temperatures (ranging from -20^{o}C to 37^{o}C for

up to a week) and degraded samples (up to 90 minutes under UV). However, further research is needed to establish the effect of clothing dyes, washing detergents and substrates. Moreover, although pyrosequencing® has been proposed as a method that produces accurate results, the increased number of PCR cycles (45 cycles) as well as the sequencing reaction itself could result in stochastic events and therefore, jeopardise its reproducibility. In our case, the mean standard deviation for CpGs 1, 2 and 3 was 0.08, while an average standard deviation of 0.05 was obtained for CpG 4 [Figure S1d]. We also believe that replicate analysis can be very useful and improve the robustness of observed DNA methylation levels, since, in some cases, the observed standard deviation per sample was as high as 0.26 (CpG 3).

In an attempt to re-create common environmental conditions often faced in forensic casework, such as washed clothes, stains exposed to heat and UV or aged samples, we analysed various stains, which most gave the expected *EFS* methylation profile. This indicates that the proposed marker is not only highly blood-specific and robust, but also sensitive and stable over time, characteristics that are highly important in such analysis. Also, as demonstrated here, it is important not to underestimate the potential of epigenetic markers in mixed stains, however more tissue-specific CpG sites should be included for each tissue for conclusive results. In complex tissue identification cases, it would perhaps be advantageous for one to combine all available approaches, such as tissue-specific mRNA profiling. Especially in the case of differentiating blood and menstrual blood, mRNA typing has been useful for the inclusion of menstrual blood through the presence of the corresponding menstrual blood-specific peaks, but there is always a possibility that the presence of non-menstrual blood masks these. DNA methylation profiling could then act as a very useful alternative to exclude the presence of non-menstrual blood in situations where a non-blood methylation profile is obtained (methylation levels of <0.80 in the case of EFS CpG4 suggested in this study). Differences in DNA methylation levels between whole blood

and menstrual blood are, to some extent, expected to occur, since apart from white blood cells menstrual blood swabs usually also contain other cell types, such as vaginal cells and skin.

Lastly, future work could include not only an extensive validation of the blood-specific marker found in this study, including sensitivity analysis and assessment of its performance combined with short tandem repeat (STR) profiling, but also the investigation of the entire *EFS* locus or its promoter, which could reveal more potentially useful CpG sites. Even though the proposed CpG 4 showed great robustness in its methylation levels after analysing 65 blood samples, to account for potential, common inter-individual differences in DNA methylation and consequently, gene expression, a larger dataset of body fluids and tissues should be analysed. It would be interesting to also include nasal blood in the analysis, which has not been investigated in any similar study before. As nasal blood is often recovered in physical assault cases, it would be beneficial to see if the EFS assay would still be able to differentiate venous and capillary blood. Additionally, since investigators' questions are often more complex and involve whether it is possible to determine what body fluids and tissues are the source of complex mixed DNA profiles obtained in a forensic case, such as samples resulted from a head shot or sexual assault samples, it is important to further investigate the proposed markers' usefulness by testing more types of tissues and mock casework scenarios.

Furthermore, given the suspected involvement of EFS gene in disease as a member of the CAS protein family, analysing diseased tissues of both blood (cancer) and uterus would explore the impact on DNA methylation levels. Using a similar approach, investigating other genes that demonstrate tissue-specific gene expression (the epigenetic profile), such as the haemoglobin gene, could reveal more markers of interest that could then be selected during

the discovery phase for assay development. A more gene-oriented targeted approach, looking at methylation sites potentially involved in gene regulation, could result in markers showing ‘stronger’ differences.

5. Conclusion

In summary, in this study we proposed a novel blood-specific differentially methylated gene locus, namely embryonal Fyn-associated substrate (*EFS*) gene that could act as a good candidate marker to confirm the presence of blood as part of a future DNA methylation-based multiplex tool, which also demonstrated its potential applicability in casework samples of questioned quantity and/or quality. In particular, CpG 4 in the EFS assay showed great robustness with low inter-individual variation and has potential as an excellent marker to assist in the differentiation between whole and menstrual blood. Further research into validating such an assay, but also into including additional adjacent CpGs, would shed more light into the usefulness of this locus.

6. Conflict of Interest

The authors declare no conflicts of interest.

7. Acknowledgements

The authors gratefully acknowledge Dr David Ballard, King’s College London, for his support and advice throughout this project and all volunteers who kindly provided samples for this study. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. AV was supported by doctorate funding awarded by the Papadaki Foundation, National and Kapodistrian University of Athens.

8. References

[1] M. Szyf, DNA methylation and demethylation probed by small molecules, Biochim Biophys Acta 1799(10-12) (2010) 750-9.
[2] R. Illingworth, A. Kerr, D. DeSousa, H. Jorgensen, P. Ellis, J. Stalker, D. Jackson, C. Clee, R. Plumb, J. Rogers, S. Humphray, T. Cox, C. Langford, A. Bird, A novel CpG island set identifies tissue-specific methylation at developmental gene loci, PloS Biology 6(1) (2008) 37-51.
[3] C. De Bustos, E. Ramos, J.M. Young, R.K. Tran, U. Menzel, C.F. Langford, E.E. Eichler, L. Hsu, S. Henikoff, J.P. Dumanski, B.J. Trask, Tissue-specific variation in DNA methylation levels along human chromosome 1, Epigenetics & Chromatin 2(1) (2009) 7.
[4] K.E. Varley, J. Gertz, K.M. Bowling, S.L. Parker, T.E. Reddy, F. Pauli-Behn, M.K. Cross, B.A. Williams, J.A. Stamatoyannopoulos, G.E. Crawford, D.M. Absher, B.J. Wold, R.M. Myers, Dynamic DNA methylation across diverse human cell lines and tissues, Genome Res 23 (2013) 555-67.
[5] V.K. Rakyan, T.A. Down, N.P. Thorne, P. Flicek, E. Kulesha, S. Graf, E.M. Tomazou, L. Backdahl, N. Johnson, M. Herberth, K.L. Howe, D.K. Jackson, M.M. Miretti, H. Fiegler, J.C. Marioni, E. Birney, T.J. Hubbard, N.P. Carter, S. Tavare, S. Beck, An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs), Genome Res 18(9) (2008) 1518-29.
[6] S. Sato, S. Yagi, Y. Arai, K. Hirabayashi, N. Hattori, M. Iwatani, K. Okita, J. Ohgane, S. Tanaka, T. Wakayama, S. Yamanaka, K. Shiota, Genome-wide DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) residing in mouse pluripotent stem cells, Genes Cells 15(6) (2010) 607-18.
[7] T.M. Thompson, D. Sharfi, M. Lee, C.M. Yrigollen, O.Y. Naumova, E.L. Grigorenko, Comparison of whole-genome DNA methylation patterns in whole blood, saliva, and lymphoblastoid cell lines, Behav Genet 43(2) (2013) 168-76.
[8] D. Frumkin, A. Wasserstrom, B. Budowle, A. Davidson, DNA methylation-based forensic tissue identification, Forensic Sci Int Genet 5(5) (2011) 517-24.
[9] T. Madi, K. Balamurugan, R. Bombardi, G. Duncan, B. McCord, The determination of tissue-specific DNA methylation patterns in forensic biofluids using bisulfite modification and pyrosequencing, Electrophoresis 33(12) (2012) 1736-45.
[10] C.M. Bartling, M.E. Hester, J. Bartz, E. Heizer, Jr., S.A. Faith, Next-generation sequencing approach to epigenetic-based tissue source attribution, Electrophoresis 35(21-22) (2014) 3096-101.
[11] T. Muangsub, J. Samsuwan, P. Tongyoo, N. Kitkumthorn, A. Mutirangura, Analysis of methylation microarray for tissue specific detection, Gene 553 (2014) 31-41.
[12] S. Forat, B. Huettel, R. Reinhardt, R. Fimmers, G. Haidl, D. Denschlag, K. Olek, Methylation markers for the identification of body fluids and tissues from forensic trace evidence, PLoS ONE 11(2) (2016) e0147973.
[13] S. Moran, C. Arribas, M. Esteller, Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhance sequences, Epigenomics 8(3) (2016) 389-99.
[14] H. Heyn, S. Moran, I. Hernando-Herraez, S. Sayols, A. Gomez, J. Sandoval, D. Monk, K. Hata, T. Marques-Bonet, L. Wang, M. Esteller, DNA methylation contributes to natural human variation, Genome Res 23 (2013) 1363-72.
[15] D.S.B.S. Silva, J. Antunes, K. Balamurugan, G. Duncan, C.S. Alho, B. McCord, Developmental validation studies of epigenetic DNA methylation markers for the detection of blood, semen and saliva samples, Forensic Science International: Genetics 23 (2016) 55-63.

[16] H.Y. Lee, M.J. Park, A. Choi, J.H. An, W.I. Yang, K.J. Shin, Potential forensic application of DNA methylation profiling to body fluid identification, Int J Legal Med 126(1) (2012) 55-62.
[17] A. Wasserstrom, D. Frumkin, A. Davidson, M. Shpitzen, Y. Herman, R. Gafny, Demonstration of DSI-semen--A novel DNA methylation-based forensic semen identification assay, Forensic Sci Int Genet 7(1) (2013) 136-42.
[18] J. Antunes, D.S.B.S. Silva, K. Balamurugan, G. Duncan, C.S. Alho, B. McCord, High-resolution melt analysis of DNA methylation to discriminate semen in biological stains, Anal Biochem 494 (2016) 40-45.
[19] L. Bai, P. Yan, C. Ximei, L. Jia, C. Zhang, D. Guo, Methylation-sensitive restriction enzyme nested real time PCR, a potential approach for sperm DNA identification, J Forensic Leg Med 34 (2015) 34-9.
[20] J. Mill, B.T. Heijmans, From promises to practical strategies in epigenetic epidemiology, Nature Reviews Genetics 14 (2013) 585-94.
[21] P. Kumar, S. Kapoor, V. Nargund, Haematospermia - a systematic review, Ann R Coll Surg Engl 88(4) (2006) 339-42.
[22] U. Baron, I. Turbachova, A. Hellwag, F. Eckhardt, K. Berlin, U. Hoffmuller, P. Gardina, S. Olek, DNA methylation analysis as a tool for cell typing, Epigenetics 1(1) (2006) 55-60.
[23] A. Lindenbergh, M. de Pagter, G. Ramdayal, M. Visser, D. Zubakov, M. Kayser, T. Sijen, A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces, Forensic Sci Int Genet 6(5) (2012) 565-77.
[24] M. van den Berge, A. Carracedo, I. Gomes, E.A. Graham, C. Haas, B. Hjort, P. Hoff-Olsen, O. Maronas, B. Mevag, N. Morling, H. Niederstatter, W. Parson, P.M. Schneider, D.S. Court, A. Vidaki, T. Sijen, A collaborative European exercise on mRNA-based body fluid/skin typing and interpretation of DNA and RNA results, Forensic Sci Int Genet 10 (2014) 40-8.
[25] H.Y. Lee, J.H. An, S.E. Jung, Y.N. Oh, E.Y. Lee, A. Choi, W.I. Yang, K.J. Shin, Genome-wide methylation profiling and a multiplex construction for the identification of body fluids using epigenetic markers, Forensic Sci Int Genet 17 (2015) 17-24.
[26] J.L. Park, O.H. Kwon, J.H. Kim, H.S. Yoo, H.C. Lee, K.M. Woo, S.Y. Kim, S.H. Lee, Y.S. Kim, Identification of body fluid-specific DNA methylation markers for use in forensic science, Forensic Sci Int Genet 13C (2014) 147-153.
[27] X.D. Fu, J. Wu, J. Wang, Y. Huang, Y.P. Hou, J. Yan, Identification of body fluid using tissue-specific DNA methylation markers, Forensic Science International: Genetics Supplementary Series 5 (2015) e151-3.
[28] L.C. Neumann, A. Weinhausel, S. Thomas, B. Horsthemke, D.R. Lohmann, M. Zeschnigk, EFS shows biallelic methylation in uveal melanoma with poor prognosis as well as tissue-specific methylation, BMC Cancer 11 (2011) 380.
[29] P.M. Warnecke, C. Stirzaker, J.R. Melki, D.S. Millar, C.L. Paul, S.J. Clark, Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA, Nucleic Acids Research 25(21) (1997) 442-4426.
[30] E.A. Moskalev, M.G. Zavgorodnij, S.P. Majorova, I.A. Vorobjev, P. Jandaghi, I.V. Bure, J.D. Hoheisel, Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression, Nucleic Acids Research 39(11) (2011) e77.
[31] L. Shen, Y. Guo, X. Chen, S. Ahmed, J.-P. Issa, Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis, BioTechniques 42(1) (2007) 48-58.

[32] T.K. Wojdacz, T. Borgbo, L.L. Hansen, Primer design versus PCR bias in methylation independent PCR amplifications, Epigenetics 4(4) (2009) 231-234.
[33] C. Thiede, G. Prange-Krex, J. Freiberg-richter, M. Bornhauser, G. Ehninger, Buccal swabs but not mouthwash samples can be used to obtain pretransplant DNA fingerprints from recipients of allogenic bone marrow transplants, Bone Marrow Transplantation 25 (2000) 575-7.
[34] M. Ishino, T. Ohba, J. Inazawa, H. Sasaki, Y. Ariyama, T. Sasaki, Identification of an Efs isoform that lacks the SH3 domain and chromosomal mapping of human Efs, Oncogene 15(14) (1997) 1741-1745.
[35] N. Tikhmyanova, A.V. Tulin, F. Roagiers, E.A. Golemis, Dcas supports cell polarization and cell-cell adhesion complexes in development, PLoS ONE 5(8) (2010) e12369.
[36] L.T. Donlin, N. Danzl, C. Wanjalla, K. Alexandropoulos, Deficiency in expression of the signaling protein Sin/Efs Leads to T lymphocyte activation and mucosal inflammation, Molecular Cell Biology 25 (2005) 11035-11046.
[37] K.B. Stefanovic, P.C. Gregg, M. Soung, Evaluation and treatment of hematospermia, American Family Physician 80(12) (2009) 1421-7.
[38] A. Vidaki, F. Giangasparo, D. Syndercombe Court, Discovery of potential DNA methylation markers for forensic tissue identification using bisulphite pyrosequencing, Electrophoresis 37(21) (2016) 2767-79.

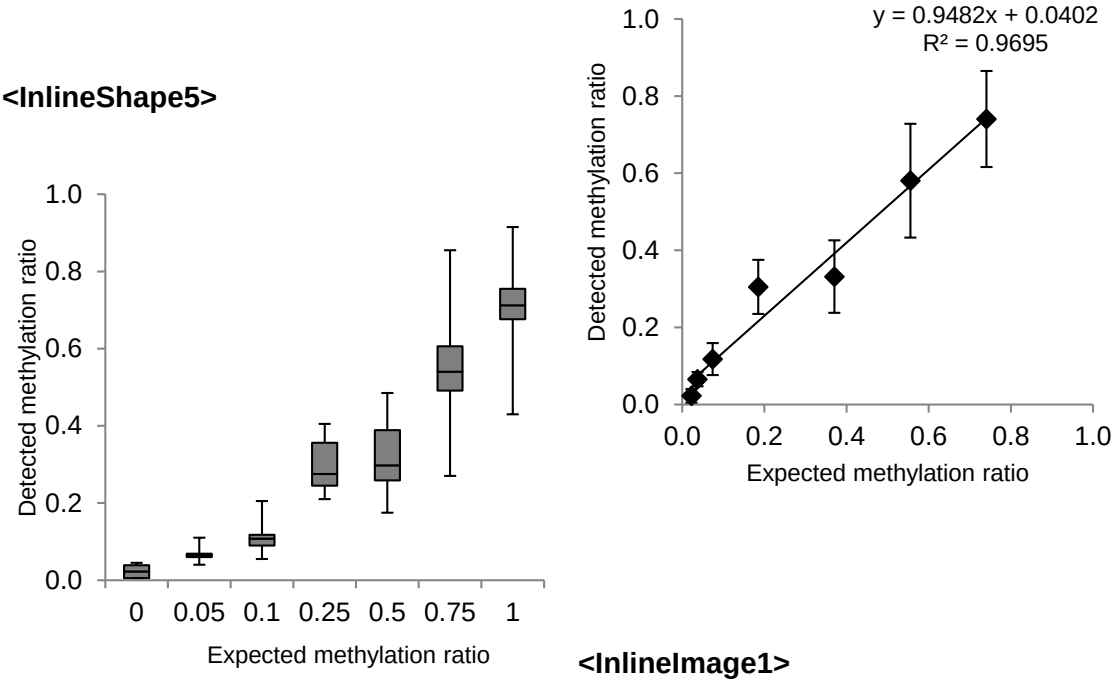
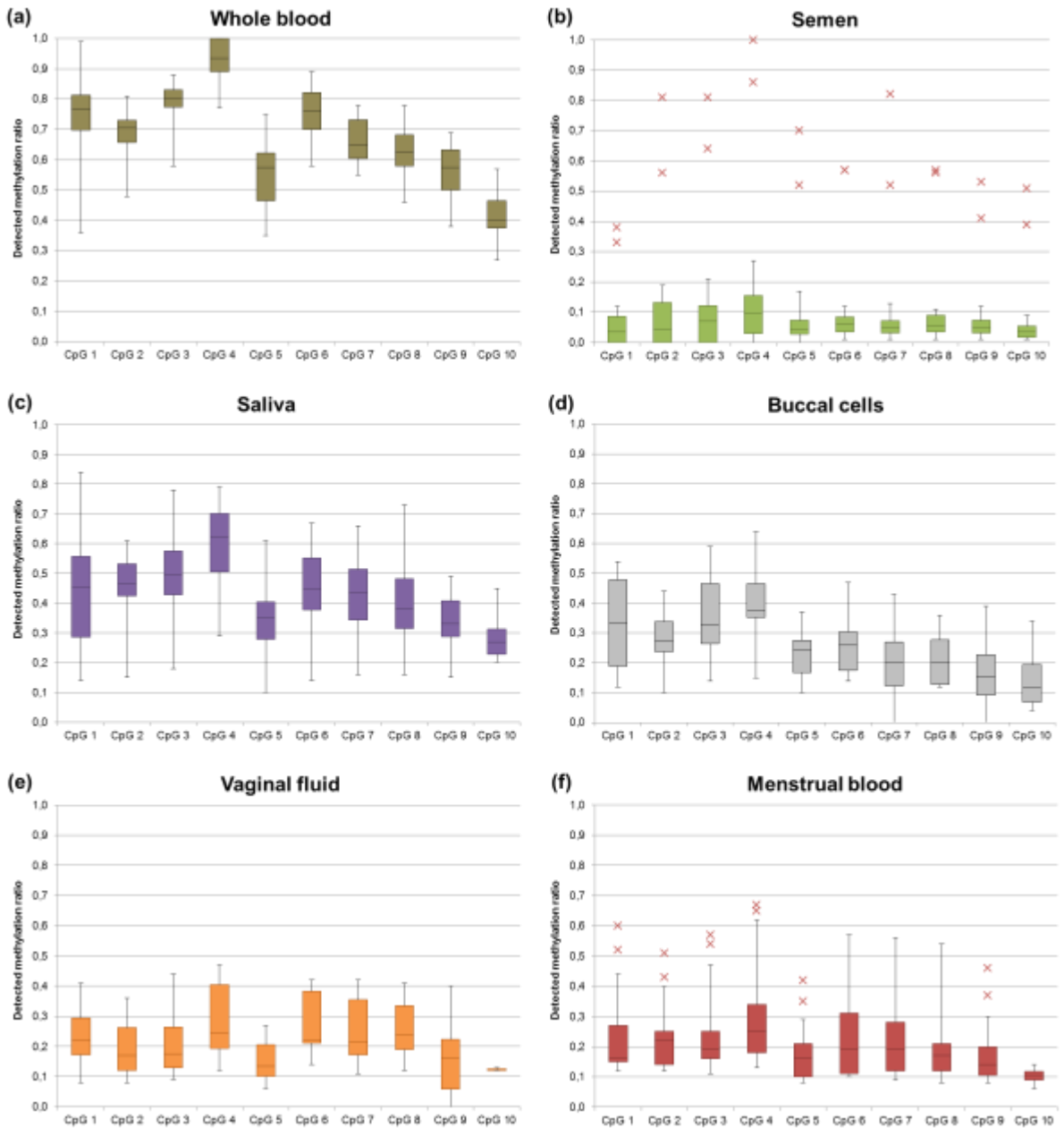
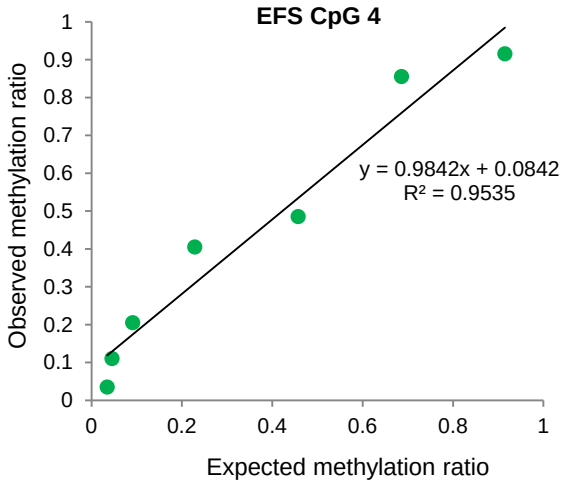


Figure 1. Linearity of methylation quantification in the EFS assay



<InlineImage2>

Figure 2. Box-and-whisker plots showing the detected methylation levels in (a) blood (n=20), (b) semen (n=10), (c) saliva (n=16), (d) buccal cells (n=10), (e) vaginal fluid (n=10), (f) menstrual blood (n=11) for all ten CpG sites included in the EFS assay (total n=77)



<InlineImage3>

Figure 3. Methylation analysis of EFS CpG 4 in mixed stains