



## Genome-wide mRNA profiling and multiplex quantitative RT-PCR for forensic body fluid identification

Seong-Min Park<sup>a,b,1</sup>, Seong-Yeon Park<sup>c,1</sup>, Jeong-Hwan Kim<sup>a,1</sup>, Tae-Wook Kang<sup>a</sup>, Jong-Lyul Park<sup>a,b</sup>, Kwang-Man Woo<sup>c</sup>, Jong-Sik Kim<sup>c</sup>, Han-Chul Lee<sup>c</sup>, Seon-Young Kim<sup>a,b,\*</sup>, Seung-Hwan Lee<sup>c,\*\*</sup>

<sup>a</sup> Medical Genomics Research Center, KRIBB, Department of Functional Genomics, Republic of Korea

<sup>b</sup> University of Science of Technology, 111 Gwahang-no, Yuseong-gu, Daejeon, Republic of Korea

<sup>c</sup> DNA Forensic Division, Supreme Prosecutor's Office, 157 Banpodae-ro, Seocho-gu, Seoul, Republic of Korea

### ARTICLE INFO

#### Article history:

Received 5 March 2012

Received in revised form 16 August 2012

Accepted 6 September 2012

#### Keywords:

Body fluid identification

mRNA marker

Microarray

qRT-PCR

DNA/RNA co-extraction

### ABSTRACT

In forensic science, identifying a tissue where a forensic specimen was originated is one of the principal challenges. Messenger RNA (mRNA) profile clearly reveals tissue-specific gene expression patterns that many attempts have been made to use RNA for forensic tissue identification. To systematically investigate the body-fluid-specific expression of mRNAs and find novel mRNA markers for forensic body fluid identification, we performed DNA microarray experiment with 24 Korean body fluid samples. Shannon entropy and *Q*-values were calculated for each gene, and 137 body-fluid-specific candidate genes were selected. By applying more stringent criteria, we further selected 28 candidate genes and validated them by RT-PCR and qRT-PCR. As a result, we suggest a novel combination of four body-fluid-specific mRNA makers: PPBP for blood, FDCSP for saliva, MSMB for semen and MSLN for vaginal secretion. Multiplex qRT-PCR assay was designed using the four mRNA markers and DNA/RNA co-extraction method was tested for forensic use. This study will provide a thorough examination of body-fluid-specifically expressed mRNAs, which will enlarge the possibility of practical use of RNA for forensic purpose.

© 2012 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Identifying a tissue origin of a specimen is crucial in some crime cases like sexual assaults. Until now, immunologic detection of specific protein marker is mainly used for forensic body fluid identification: prostate specific antigen (PSA) for semen, and hemoglobin for blood identification [1–3]. However, immunologic tests occasionally cross-react to other species or tissues. And most immunologic tests rely on simple color-change which is hard to quantify, and sometimes hard to recognize especially there are very small amounts of forensic sample. Use of RNA for forensic body fluid identification has been discussed because of its reliable tissue specificity, and the available co-extraction methods to obtain RNA along with genomic DNA (gDNA) [4]. While gDNA is used for standard STR analysis, RNA can give tissue-specific information which DNA does not have. Furthermore, when used with multiplex PCR, numerous tissue-specific genes can be

simultaneously analyzed to identify multiple tissue types and thus save the limited amount of forensic samples. RNA analysis is expected to be adopted smoothly by using the reverse transcription and standard end-point PCR (RT-PCR) and reverse transcription and quantitative PCR (qRT-PCR) assays which are now widely used by forensic laboratories.

The major obstacle of forensic RNA analysis is the instability of RNA molecule which is easily degraded by ubiquitous ribonuclease (RNase) enzyme. But recent studies have shown that mRNA samples from crime scenes can be sufficiently stable for forensic use [5,6]. Using commercially provided RNA extraction kits and DNA/RNA co-extraction methods, RNA was successfully isolated from forensic samples [7–10]. These studies show that body fluid types can be distinguished using forensic RNA samples by measuring the expression levels of specific mRNA markers.

Several tissue-specific mRNA markers have been discovered for identification of frequently encountered forensic body fluids including blood, saliva, semen, vaginal secretion and menstrual blood [6,11–16]. In previous studies, many body-fluid-specific mRNA markers have been reported through single-gene-based assays. For example, SPTB, PBGD and HBA1 are reported as blood-specific mRNA markers; MMP7 and MMP11 as menstrual blood-specific markers; STATH and HTN3 as saliva-specific markers; PRM1, PRM2 and KLK3 as semen-specific markers; and HBD1 and

\* Corresponding author at: University of Science of Technology, 111 Gwahang-no, Yuseong-gu, Daejeon, Republic of Korea. Tel.: +82 042 879 8116; fax: +82 042 879 8119.

\*\* Corresponding author. Tel.: +82 02 3480 3730; fax: +82 02 3480 3606.  
E-mail addresses: [kimsy@kribb.re.kr](mailto:kimsy@kribb.re.kr) (S.-Y. Kim), [shlee@spo.go.kr](mailto:shlee@spo.go.kr) (S.-H. Lee).

<sup>1</sup> These authors contributed equally to this work.

MUC4 as vaginal-secretion-specific markers [4,15,17–20]. mRNA expression level was measured using RT-PCR or qRT-PCR [6,11–16,21,22]. Multiplex qRT-PCR was also developed to detect several mRNA markers at once [23]. These methods have advantages of sensitive analysis of mRNA expression, but only a few genes can be examined. To understand the gene expression signature of forensic body fluid samples and further investigate the possibility of novel mRNA markers, the systematic approach to analyze the whole transcriptome is needed.

DNA microarray is a useful tool to analyze the whole transcriptome and discover new biomarkers. Several researchers applied DNA microarray to their search for tissue-specific genes [24–26]. In forensic science, specific mRNA and microRNA markers for identification of blood and saliva had been found using DNA microarray [27,28]. However, investigating specific mRNA markers of four body fluids (blood, saliva, semen, and vaginal secretion) by genome-wide profiling has not been performed yet.

In this study, we performed genome-wide expression profiling with total RNAs isolated from four kinds of body fluid samples (blood, saliva, semen and vaginal secretion) obtained from Korean volunteers. Then, we calculated Shannon's entropy and  $Q$ -values for each gene to select 137 specific genes for each body fluid. The body-fluid-specific signature of the 137 genes was validated in an independent gene expression dataset prepared from public gene expression database. We also validated 18 genes by RT-PCR and qRT-PCR. As a result, we obtained novel specific mRNA markers for forensic body fluid identification.

## 2. Materials and methods

### 2.1. Sample collection and RNA preparation

Four kinds of body fluid samples, blood (18), saliva (18), semen (43) and vaginal secretion (157) were collected from healthy Korean volunteers, and treated anonymously following local ethical restrictions. To prepare intact RNA for microarray, aliquots of whole blood, saliva and semen were collected in microcentrifuge tubes and vaginal secretion samples were collected by immediate soaking of vaginal secretions in lysis buffer. Total RNA for microarray was isolated using Qiagen RNeasy Mini Kit according to the manufacturer's protocol. For saliva, several RNA samples were isolated using Qiagen RNeasy Protect Saliva Mini Kit. For DNA and RNA co-extraction, Qiagen AllPrep DNA/RNA Mini Kit was used to prepare genomic DNA and total RNA simultaneously. Extracted total RNA was analyzed with Experion<sup>TM</sup> RNA StdSens (BIO-RAD) to check its quality and quantity.

### 2.2. DNA microarray experiment

Total RNA (1–5  $\mu$ g) was amplified with Illumina TotalPrep<sup>TM</sup> RNA Amplification Kit (Ambion). RNA amplification process is composed of cRNA (RNA derived from cDNA through standard RNA synthesis) synthesis, biotin labeling and *in vitro* transcription steps. Through these steps, RNA (500 ng) was converted into biotin-labeled cRNA, and then 750 ng of cRNA was hybridized with Illumina BeadChip according to manufacturer's protocol. After hybridization, BeadChip was scanned with Illumina BeadArray Reader<sup>TM</sup> laser scanner. Scanned images were converted into intensity values using Illumina BeadStudio<sup>TM</sup> 3.4 program.

### 2.3. Analysis of DNA microarray data

Intensity values from Illumina BeadStudio<sup>TM</sup> 3.4 program were globally normalized with Quantile method [29]. We used Shannon entropy ( $H$ ) to measure the overall specificity of a gene [30,31]. Given expression levels of a gene ( $g = 1, 2, \dots, G$ ) in  $N$  tissues ( $t = 1,$

$2, \dots, N$ ), the relative expression of a gene  $g$  in a tissue  $t$  was defined as

$$p_{t|g} = \frac{w_{g,t}}{\sum_{1 \leq t \leq N} w_{g,t}},$$

where  $w_{g,t}$  is the expression level of the gene in the tissue. The entropy of a gene's expression distribution was calculated as

$$H_g = \sum_{1 \leq t \leq N} -p_{t|g} \log_2(p_{t|g})$$

A low entropy value represents a specific gene expression pattern. However, while Shannon entropy provides a single metric for assessing the complete gene expression profile, it does not provide any information on the tissues in which a gene may be specifically expressed. A statistic,  $Q_{g|t}$ , was introduced to measure the degree of specificity of gene expression in each tissue type:

$$Q_{g|t} = H_g - \log_2(p_{t|g})$$

A low  $Q_{g|t}$  represents a specific expression pattern for that tissue [30,31].

Based on  $Q$ -value cut-off 1.8, candidate genes were selected. Co-expressed genes in more than two body fluids were removed manually. Unsupervised hierarchical clustering and visualization was performed with MEV 4.0 program (<http://www.tm4.org/>). Affymetrix U133 Plus 2 (GPL570) platform DNA microarray data about four body fluids were collected from Gene Expression Omnibus (GEO) database in NCBI (<http://www.ncbi.nlm.nih.gov/geo/>). GEO accession numbers of collected datasets were GSE6872, GSE7307, GSE7451, GSE8764, GSE11622, GSE12446, GSE13494, GSE14245, GSE14642, GSE17340, GSE20266, GSE22331 and GSE25518. Collected microarray data were globally normalized with MAS5 method by using affy package [32]. Two DNA microarray datasets about four body fluids have been visualized and provided in our own web server (<http://medical-genome.kribb.re.kr/forensic>). All primary data are deposited in the Gene Expression Omnibus (GEO) in NCBI under accession number GSE34844.

### 2.4. RT-PCR and qRT-PCR

Using total RNA (1–5  $\mu$ g) as a template, reverse transcription was carried out with iScript<sup>TM</sup> cDNA Synthesis Kits (BIO-RAD). Then, 1/150 of synthesized cDNA was used for RT-PCR and qRT-PCR. The RT-PCR primers were designed with Primer3 software. We designed primers and probes for RT-PCR and qRT-PCR based on sequences on exon–exon junctions to avoid possible false positives from genomic DNA contamination. RT-PCR was performed with Novelzyme<sup>TM</sup> Taq Plus Premix using the following parameters: initial denaturation at 98 °C for 5 min, followed by 35 cycles of denaturation at 98 °C for 30 s, annealing at 60 °C for 30 s, elongation at 72 °C for 30 s, and a final elongation at 72 °C for 5 min.  $\beta$ -Actin gene was used as a housekeeping control for normalization. PCR product was detected using agarose gel electrophoresis and ethidium bromide staining. The sequences of oligonucleotide primers are listed in [Supplementary Table 1](#).

Commercial dual labeled fluorogenic probes (Metabion) were used for qRT-PCR. The qRT-PCR reaction was carried out with THUNDERBIRD<sup>TM</sup> Probe qRT-PCR Mix on a CFX96<sup>TM</sup> Real-Time PCR machine (C1000<sup>TM</sup> Thermal Cycler, Bio-Rad) using the following parameters: initial denaturation at 94 °C for 1 min, followed by 45 cycles of denaturation at 94 °C for 15 s, and a final annealing/elongation at 56 °C for 1 min. GAPDH gene was used as a

housekeeping control for normalization. The sequences of oligonucleotide primers used in qRT-PCR are listed in [Supplementary Table 2](#). Expression level was quantified using delta  $C_t$  ( $\Delta C_t$ ) method.

For multiplex qRT-PCR, iQ<sup>TM</sup> Multiplex Powermix (Bio-Rad) was used. The qRT-PCR reaction was carried out on a CFX96<sup>TM</sup> Real-Time PCR machine (C1000<sup>TM</sup> Thermal Cycler, Bio-Rad) using the following parameters: initial denaturation at 95 °C for 1 min, followed by 45 cycles of denaturation at 95 °C for 15 s, and a final annealing/elongation at 56 °C for 1 min. GAPDH gene was used as a housekeeping control for normalization. Probes for multiplex qRT-PCR were labeled with five dyes: FAM, Yakima Yellow, IRD700, Cy5 and CAL Fluor Red 610. The intensities were detected using five filters: FAM, HEX, Quasar705, Cy5 and CAL Fluor Red 610. The sequences of oligonucleotide primers used in multiplex qRT-PCR are listed in [Supplementary Table 3](#). Probes and primers were added as the following concentrations for multiplex qRT-PCR: PPBP probe, 0.33  $\mu$ M; FDCSP probe, 0.67  $\mu$ M; MSMB probe, 0.33  $\mu$ M; MSLN probe, 0.33  $\mu$ M; GAPDH probe, 0.33  $\mu$ M; PPBP forward primer, 0.33  $\mu$ M; PPBP reverse primer, 0.33  $\mu$ M; FDCSP forward primer, 0.67  $\mu$ M; FDCSP reverse primer, 0.67  $\mu$ M; MSMB forward primer, 0.33  $\mu$ M; MSMB reverse primer, 0.33  $\mu$ M; MSLN forward primer, 0.67  $\mu$ M; MSLN reverse primer, 0.67  $\mu$ M; GAPDH forward primer, 0.33  $\mu$ M; GAPDH reverse primer, 0.33  $\mu$ M. We selected GAPDH as an endogenous control based on the previous reports [4,23]. Additionally, its expression was relatively invariable among samples and different tissues in our microarray experiment.

## 2.5. STR analysis

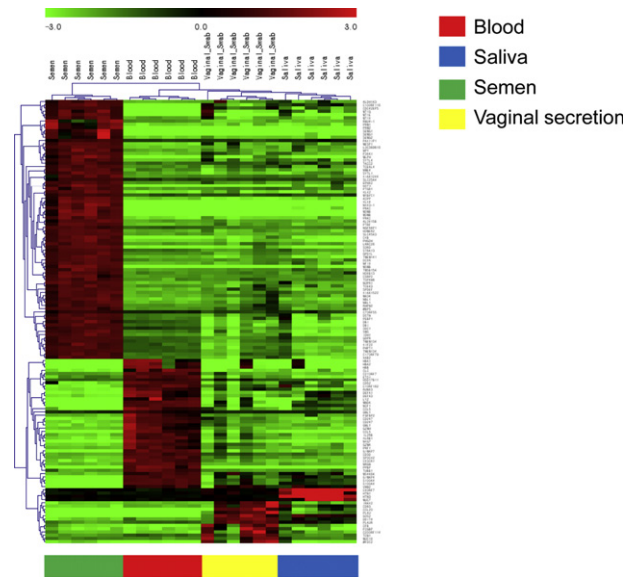
Four kinds of body fluid samples, blood (6), saliva (8), semen (7) and vaginal secretion (7) were collected from healthy Korean volunteers and used for DNA and RNA co-extraction. gDNA was prepared with Qiagen AllPrep DNA/RNA Mini Kit simultaneously with total RNA according to the manufacturer's protocol. gDNA was quantified with Quantifiler<sup>TM</sup> Human DNA Quantification Kit (Applied Biosystems) on 7500 Real-Time PCR System (Applied Biosystems) according to the manufacturer's protocol. Autosomal STR amplification was carried out using AmpFISTR<sup>®</sup> Identifier<sup>®</sup> PCR Amplification Kit (Applied Biosystems) according to the manufacturer's protocol. PCR product was resolved and detected with 3730 DNA Analyzer (Applied Biosystems) capillary electrophoresis system and STR profile was analyzed using GeneMapper ID software (Applied Biosystems).

## 3. Results

### 3.1. Gene expression profiling

We performed gene expression profiling experiments to find out specific mRNA markers for four types of body fluids including blood, saliva, semen, and vaginal secretion. Using commercial RNA preparation kits, we prepared total RNA from each body fluid samples and checked the quality of total RNA using Experion<sup>TM</sup> ([Supplementary Figure 1](#)). Not surprisingly, total RNA samples, especially from saliva and vaginal secretion, were in bad quality with extensive degradation possibly because of the abundant RNase in some of the body fluid samples ([Supplementary Figure 1](#)).

We then checked if DNA microarray experiments were possible for those RNA samples of bad quality. The first step of Illumina BeadChip Array<sup>TM</sup>, one of widely used DNA microarray platforms, is RNA amplification which performs cRNA synthesis, labeling with biotin and *in vitro* transcription. The success of Illumina BeadChip Array<sup>TM</sup> experiments is almost dependent on the success of RNA amplification. We selected 24 samples, 6 per



**Fig. 1.** Selected body-fluid-specific candidate genes. DNA microarray experiments were performed for 24 Korean body fluid samples, 6 samples per each body fluid, using Illumina BeadChip Array platform. Based on the  $Q$ -value ( $Q \leq 1.8$ ) of each body fluids, 137 body-fluid-specific candidate genes were selected. Unsupervised hierarchical clustering of 137 candidate genes was performed (red: blood, blue: saliva, green: semen, yellow: vaginal secretion). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

each body fluid, out of the 44 samples, based on the gel electrophoresis image rather than RQI value because we assumed that the RQI value might be less useful for highly degraded samples. With the selected samples, we performed RNA amplification and checked if cRNA was synthesized properly for subsequent steps ([Supplementary Table 4](#)). The qualities of some of the cRNA samples were not good, but most of them were amplified enough for hybridization to Illumina BeadChip. Thus, we hybridized them to Illumina BeadChip and scanned with laser scanner. During the scanning, we checked the overall intensities of 24 cRNA hybridization images ([Supplementary Figure 2](#)), and found that 6 out of 24 images had low overall intensity values. However, as we show later, even those samples were informative in selecting body-fluids-specific genes. These results show that Illumina BeadChip DNA microarray platform is robust enough to be applied to forensic body fluid investigation.

### 3.2. Gene expression data analysis

We used Shannon Entropy ( $H$ ) and  $Q$ -statistics ( $Q$ ) to select specific genes for each body fluid. Using normalized intensity values of 24 samples, we calculated  $H$  and  $Q$  values for all genes on the DNA microarray. Then, based on these  $Q$ -values, we selected candidate genes whose  $Q$ -values were lower than 1.8 in at least one body fluid. First, 158 genes were selected, but some of them were not specific to one body fluid. Especially, 21 genes were expressed in both saliva and vaginal secretion. After removing those 21 genes, we obtained 137 potential body-fluid-specific genes ([Supplementary Table 5](#)). These candidate genes consist of 40 blood-specific, 80 semen-specific, 4 saliva-specific and 13 vaginal-secretion-specific genes. When we performed unsupervised hierarchical clustering with the 137 candidate genes, we found that those genes distinguish the four body fluids ([Fig. 1](#)). This result indicates that the 137 candidate genes are successfully selected to show a body-fluid-specific signature from 24 Korean samples using Illumina BeadChip Array<sup>TM</sup> platform.

**Table 1**

41 candidate genes for the identification of 4 body fluids.

Gene symbol	Full name	Entropy	Q <sub>blood</sub> <sup>a</sup>	Q <sub>saliva</sub>	Q <sub>semen</sub>	Q <sub>vaginal</sub>
<i>Blood</i>						
C21orf7	Chromosome 21 open reading frame 7	1.07	<b>1.41</b>	5.13	4.70	4.93
CCL5	Chemokine (C–C motif) ligand 5	0.69	<b>0.87</b>	5.74	5.78	4.85
CD247	CD247 molecule	1.11	<b>1.49</b>	5.23	5.43	4.16
CD3D	CD3d molecule, delta (CD3–TCR complex)	1.19	<b>1.60</b>	4.93	4.94	4.56
CD52	CD52 molecule	1.15	<b>1.55</b>	4.19	5.47	5.02
CLC	Charcot-Leyden crystal protein	0.77	<b>0.98</b>	5.15	5.37	5.33
CX3CR1	Chemokine (C–X3–C motif) receptor 1	1.01	<b>1.32</b>	5.06	5.16	4.73
ETS1	v-ets erythroblastosis virus E26 oncogene homolog 1 (avian)	1.08	<b>1.43</b>	5.62	4.49	4.74
FGFBP2	FGFBP2 fibroblast growth factor binding protein 2	1.16	<b>1.55</b>	4.90	4.94	4.68
GNLY	Granulysin	1.10	<b>1.46</b>	5.00	5.00	4.68
GZMH	Granzyme H (cathepsin G-like 2, protein h-CCPX)	1.06	<b>1.39</b>	5.03	4.92	4.83
HBA1	Hemoglobin, alpha 1	0.61	<b>0.76</b>	5.56	5.75	5.39
HBB	Hemoglobin, beta	0.34	<b>0.40</b>	6.67	6.81	5.85
KLRB1	Killer cell lectin-like receptor subfamily B, member 1	1.01	<b>1.32</b>	5.22	5.24	4.54
LYZ	Lysozyme	1.04	<b>1.41</b>	3.70	6.20	5.68
MNDA	Myeloid cell nuclear differentiation antigen	0.97	<b>1.28</b>	4.09	6.16	5.26
NKG7	Natural killer cell group 7 sequence	0.64	<b>0.80</b>	5.80	5.82	4.99
NRGN	Neurogranin (protein kinase C substrate, RC3)	1.01	<b>1.31</b>	4.95	5.00	5.00
PPBP	Pro-platelet basic protein (chemokine (C–X–C motif) ligand 7)	0.52	<b>0.63</b>	5.80	5.98	5.62
PRF1	Perforin 1 (pore forming protein)	1.13	<b>1.51</b>	4.90	5.09	4.60
RUNX3	Runt-related transcription factor 3	1.23	<b>1.72</b>	5.29	5.72	3.71
TUBB1	Tubulin, beta 1 class VI	1.18	<b>1.58</b>	4.86	4.84	4.76
<i>Saliva</i>						
FDCSP	Follicular dendritic cell secreted protein	0.77	5.47	<b>0.98</b>	5.23	5.15
HTN1	Histatin 1	0.83	5.23	<b>1.06</b>	5.37	4.99
HTN3	Histatin 3	0.37	6.29	<b>0.45</b>	6.20	6.24
MUC7	Mucin 7, secreted	1.26	4.74	<b>1.71</b>	4.79	4.73
<i>Semen</i>						
MSMB	Microseminoprotein, beta-	0.34	6.69	6.27	<b>0.41</b>	6.19
NEFH	Neurofilament, heavy polypeptide	0.69	5.73	5.01	<b>0.86</b>	5.58
NKX3-1	NK3 homeobox 1	0.38	6.44	5.76	<b>0.46</b>	6.54
PRM1	Protamine 1	0.97	5.10	5.05	<b>1.26</b>	4.91
PRM2	Protamine 2	0.66	5.55	5.39	<b>0.83</b>	5.46
SEMG1	Semenogelin I	0.48	5.87	5.98	<b>0.59</b>	5.82
SEMG2	Semenogelin II	1.21	4.89	4.70	<b>1.62</b>	4.80
KLK4	Kallikrein-related peptidase	0.22	7.06	6.85	<b>0.26</b>	7.16
<i>Vaginal secretion</i>						
PLAU	Plasminogen activator, urokinase	0.84	6.28	3.90	6.23	<b>1.10</b>
CFB	Complement factor B	0.86	5.55	4.71	5.28	<b>1.10</b>
IGFBP3	Insulin-like growth factor binding protein 3	1.35	4.68	4.44	4.96	<b>1.86</b>
SERPINB3	Serpin peptidase inhibitor, clade B (ovalbumin), member 3	1.38	4.84	4.39	4.78	<b>1.91</b>
S100A7	S100 calcium binding protein A7	1.46	4.74	4.32	4.78	<b>2.07</b>
MMP7	Matrix metalloproteinase 7	1.51	4.68	4.50	4.55	<b>2.16</b>
MSLN	Mesothelin	1.52	4.74	4.37	4.60	<b>2.18</b>

<sup>a</sup> Bold number means the lowest Q-value among the samples. The Q-value means the degree of specificity of gene expression in each tissue type.

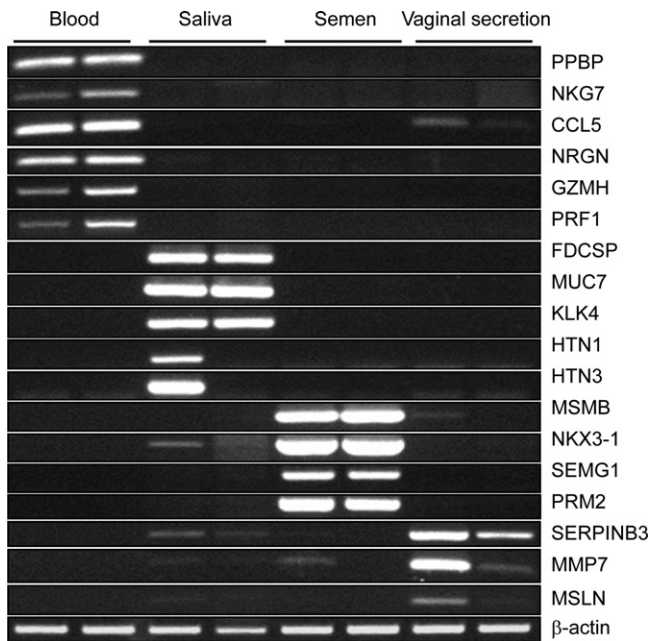
### 3.3. Analysis of public expression datasets

To validate the reliability of the 137 candidate genes in other dataset, we analyzed gene expression datasets publicly available from the Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>). We collected data from samples of the four body fluid types profiled using the Affymetrix U133 Plus2 platform. A total of 113 samples were collected, which contained 40 blood, 37 saliva, 13 semen and 23 vaginal samples. After global normalization of the intensity values summarized by using MAS5 algorithm, we extracted expression values of the 137 candidate genes and performed unsupervised hierarchical clustering. The 137 genes clearly separated the four body fluids in the 113 samples (Supplementary Figure 3). Therefore, it is concluded that the 137 body-fluid-specific genes can be applied to various datasets and microarray platforms as a general body-fluid-specific signature. We show the expression patterns of the well-known specific mRNA markers of the four body fluids: blood (HBA1 and PPBP), saliva (STATH and HTN3), semen (PRM1 and PRM2), and vaginal secretion (HBD1) in Supplementary Figure 4. A simple web interface is also prepared to help examine the gene expression

pattern of each gene in the two datasets (<http://medical-genome.kribb.re.kr/forensic>).

### 3.4. Further selection of 41 out of 137 candidate genes

For practical application, it is expensive and time-consuming to use all the 137 candidate genes for body fluid identification. We tried to select representative genes from the 137 candidates. Three criteria were used to select representative candidate genes: (1) the rank of Q-value in each body fluid, (2) the specificity in both datasets, (3) the absolute expression level. By these three criteria, we selected 22 blood-specific, 4 saliva-specific and 8 semen-specific representative candidate genes (Table 1 and Supplementary Figure 5). Unfortunately, for vaginal secretion samples, we could not obtain any proper specific genes fulfilling the three criteria. For vaginal secretion, we used higher Q-value cut-off and selected other candidate genes (Table 1). Many famous body-fluid-specific markers including PPBP (blood), HBA1 (blood), HTN1 (saliva), HTN3 (saliva), PRM1 (semen) and PRM2 (semen) were successfully selected from our dataset (Table 1), which suggests the effectiveness of our approach to identify specific markers for the four body fluids.



**Fig. 2.** Validation of selected body-fluid-specific mRNA markers using RT-PCR. RT-PCR assay was carried out with two total RNA samples per each body fluid. Body-fluid-specific expression patterns of 18 representative candidate genes were validated. PCR product was detected using agarose gel electrophoresis and ethidium bromide staining.

### 3.5. Validation of body-fluid-specific markers using RT-PCR and qRT-PCR

We initially tested each of the primer pairs of the 41 candidate genes, but found that only 28 primer pairs worked in our RT-PCR condition (data not shown), therefore those 28 genes were selected for validation by RT-PCR (Supplementary Table 1). First, we performed RT-PCR assay with two samples for each body fluid (Fig. 2 and Supplementary Figure 6). Eighteen out of 28 genes were successfully validated as body-fluid-specific genes (Fig. 2), while 10 genes were not specific (Supplementary Figure 6). Successfully validated mRNA markers included six blood-specific (PPBP, NKG7, CCL5, NRGN, GZMH, PRF1), five saliva-specific (FDCSP, MUC7, KLK4, HTN1, HTN3), four semen-specific (MSMB, NKX3-1, SEMG1, PRM2) and three vaginal-secretion-specific (SERPINB3, MMP7, MSLN) genes. Several candidate genes including HBA1 (blood) and PRM1 (semen) are already known as body-fluid-specific mRNA markers, but in our RT-PCR data, their expression patterns were not body-fluid-specific in Korean samples (Supplementary Figure 6).

We tested qRT-PCR probes of the 18 genes, but only ten probes worked in a preliminary qRT-PCR test (data not shown). Therefore, 10 mRNA markers were validated again by qRT-PCR using TaqMan<sup>®</sup> probes (Supplementary Table 2). We first performed qRT-PCR assay with two samples per each body fluid (Supplementary Figure 7). Most mRNA markers, except KLK4, showed the same body-fluid-specific expression patterns as RT-PCR results. For multiplex qRT-PCR, we selected four markers, only one mRNA marker per each body fluid (PPBP for blood, FDCSP for saliva, MSMB for semen and MSLN for vaginal secretion), and their expression patterns were further validated in a larger number of samples using qRT-PCR. Fig. 3 shows the body-fluid-specific expressions of the four representative genes, PPBP, FDCSP, MSMB and MSLN.

For the practical application in forensic body fluid identification, we designed multiplex qRT-PCR probes of the four mRNA markers (Supplementary Table 3). Multiplex qRT-PCR was carried

out using those probes labeled with 5 different dyes (Fig. 4). Four body fluids were successfully distinguished in 5-dye multiplex qRT-PCR assay.

We then tried to evaluate the four mRNA markers. Two datasets were used for the evaluation: one was the microarray data of the 113 samples from public microarray database and the other was our additional qRT-PCR data of 18 samples per each body fluid. First, we performed receiver operating characteristic (ROC) analysis and calculated area under curve (AUC) value (Table 2, Supplementary Figure 8 and 9). Second, we calculated sensitivity and specificity of the four mRNA markers (Table 2). The selected markers showed good sensitivity, specificity and AUC value in the microarray and the qRT-PCR datasets except MSLN. These results suggest that our mRNAs were efficient body-fluid-specific markers though the vaginal secretion marker, MSLN, was less optimal in its sensitivity as we previously expected.

### 3.6. DNA/RNA co-extraction and simultaneous body fluid identification and STR analysis

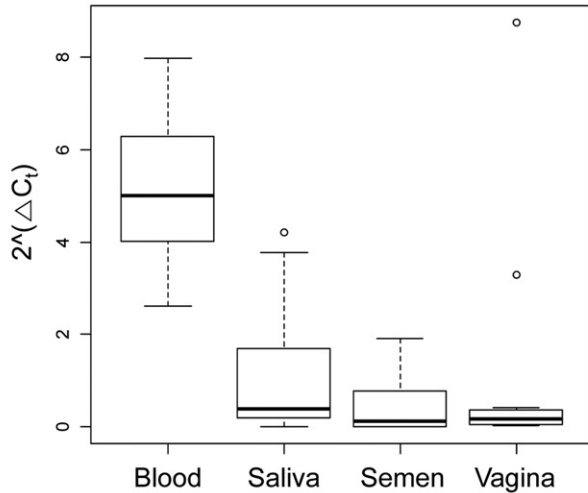
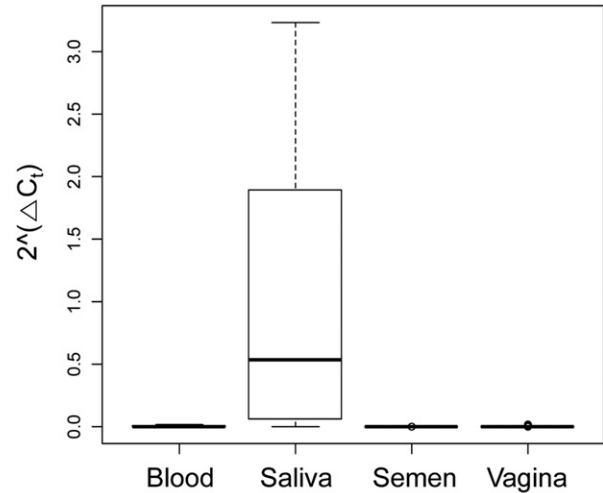
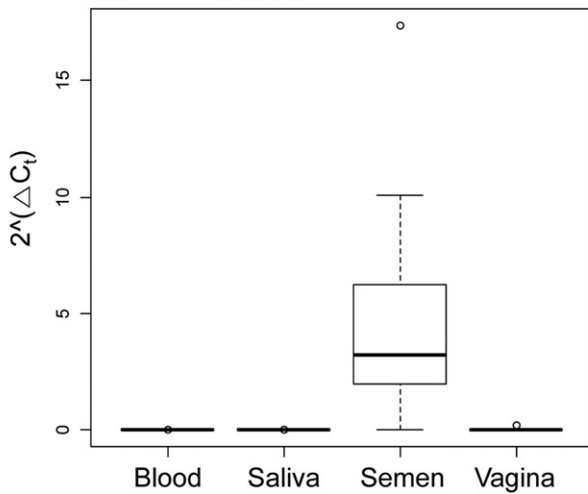
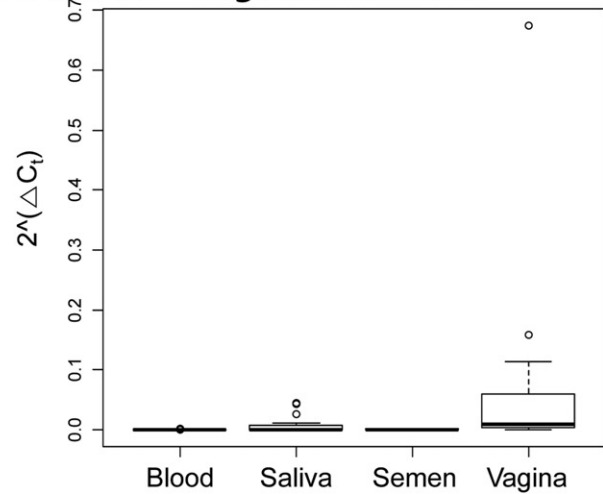
In most forensic samples, it is indispensable to conduct human identification by purifying gDNA followed by STR amplification and fragment analysis. Simultaneous extraction of DNA and RNA in the same spot of sample can save the sample consumption, and it allows gDNA extraction even though RNA is badly degraded. Commercially provided DNA/RNA co-extraction kit (AllPrep DNA/RNA Mini Kit, Qiagen) was tested for independent forensic body fluid samples: blood, semen, saliva and vaginal secretion. Multiplex qRT-PCR was followed using the RNA fraction, while gDNA was used for autosomal STR amplification according to the standard protocol.

In all body fluid samples tested for DNA/RNA co-extraction, the 15 Identifiler autosomal STR loci, along with Amelogenin locus, were successfully analyzed. The representative STR analysis results of each body fluid are shown in Supplementary Figure 10. RNA prepared simultaneously with gDNA was tested for body fluid identification with multiplex qRT-PCR method. Our multiplex qRT-PCR system using newly selected four mRNA markers succeeded in distinguishing the body fluid types by showing the specific expression of its own body fluid marker (Supplementary Figure 11). These results indicate that gDNA extracted by DNA/RNA co-extraction method is of sufficiently good quality for STR analysis, and also RNA prepared together is suitable for RT-PCR assay. This would open up the possibility of forensic use of mRNA markers for body fluid identification.

## 4. Discussion

We collected 24 body fluid samples from Korean individuals, and selected 137 specific genes from microarray gene expression profiling data. Those candidate genes were validated in another dataset prepared from public gene expression datasets. We then further validated 18 genes by RT-PCR and qRT-PCR. We also developed multiplex qRT-PCR system using the newly selected four representative mRNA markers and its usability was demonstrated with the DNA/RNA co-extraction method for practical forensic application.

One challenge for gene expression profiling experiment with body fluid samples is the quality and quantity of RNA. Not surprisingly, we found that many RNA samples from body fluids were degraded and of poor quality. However, even with the degradation, we could obtain enough cRNA samples after RNA amplification and hybridize them to Illumina BeadChip. These results suggest that Illumina BeadChip Array platform is sensitive enough to be applicable to highly degraded RNA samples from crime scenes.

**A. PPBP - Blood****B. FDSCP - Saliva****C. MSMB - Semen****D. MSLN - Vaginal secretion**

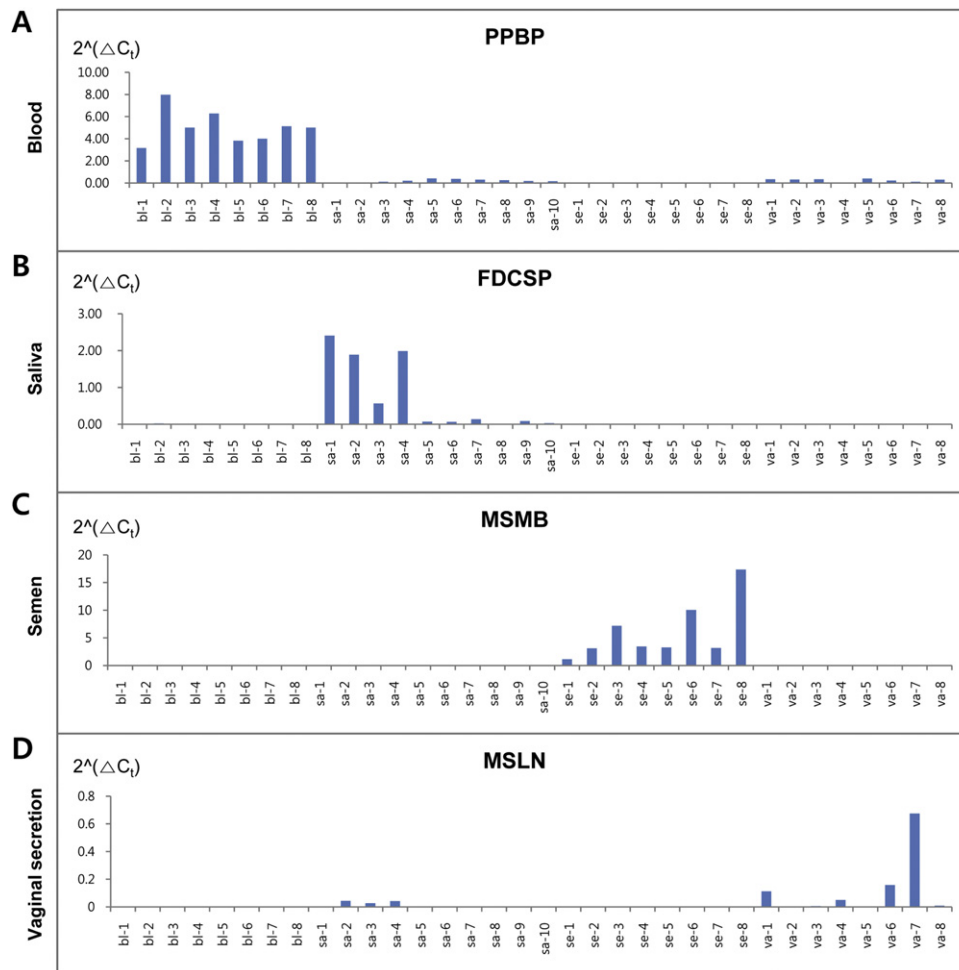
**Fig. 3.** Validation of selected body-fluid-specific mRNA markers using qRT-PCR. qRT-PCR assay was carried out with 18 total RNA samples per each body fluid. Expression patterns of four mRNA markers including PPBP (blood), FDSCP (saliva), MSLN (Semen) and MSMB (vaginal secretion) were measured by qRT-PCR and represented as a box plot. (A) Blood, (B) saliva, (C) semen, (D) vaginal secretion.

The aim of this study was to investigate new body-fluid-specific mRNA markers using DNA microarray technology to distinguish the four body fluids frequently obtained from crime scenes. We performed gene expression profiling experiment for 24 body fluid samples from Korean individuals, and starting from 29,377 genes, selected 137 body-fluid-specific genes by applying Shannon's entropy and  $Q$ -statistics. We then validated the 137 genes in another dataset of 113 non-Korean individuals profiled by using Affymetrix platform. As a result, we showed that the 137 candidate genes display a body-fluid-specific gene expression signature not restricted to Koreans or to a single microarray platform. Previous researches tried to find body-fluid-specific markers using DNA microarray, but only two kinds of body fluids, blood and saliva, were studied so far [27,28]. As our study systematically investigated four body fluids frequently encountered in crime scenes, it is expected to contribute to discovering more efficient combination of mRNA markers for body fluid identification. We provide a simple web interface to examine our data for other forensic scientists at <http://medical-genome.krabb.re.kr/forensic>.

Using three stringent criteria, we selected 41 candidate genes out of 137 genes for further validation (Table 1). 18 of them were

successfully validated by RT-PCR and qRT-PCR. Well-known body-fluid-specific mRNA markers which were reported in previous studies were included in the 41 selected candidate genes: HBA1, MMP7, HTN3, PRM1 and PRM2 [4,15,17–20]. However, some genes showed discrepancies in expression patterns. For example, HBA1 is known as a blood-specific mRNA marker and PRM1 is known as a semen-specific mRNA marker. However, in our RT-PCR data, HBA1 and PRM1 did not show the body-fluid-specific expression in Korean samples. It is also examined that HBA1 is expressed in some of vagina samples in non-Korean GEO data set as well (Supplementary Figure 5). Meanwhile, our selected representative blood-specific mRNA marker, PPBP, showed specific expression patterns in both our own and GEO datasets (Supplementary Figure 5). This blood-specific expression of PPBP was validated using RT-PCR and qRT-PCR assay (Figs. 2 and 3). These results suggest that we may not apply some of previously reported body-fluid-specific mRNA markers to Korean samples without thorough observation of their expression patterns in wide datasets.

Among the four body fluid types, vaginal secretion samples showed the most heterogeneous gene expression patterns, so we had to apply higher  $Q$ -value cut-off to select candidate genes for



**Fig. 4.** Validation of multiplex qRT-PCR probes. Multiplex qRT-PCR was performed with probes labeled with five different dyes. GAPDH gene expression was measured together and used as endogenous control (bl: blood, sa: saliva, se: semen, va: vaginal secretion). (A) Blood, (B) saliva, (C) semen, (D) vaginal secretion.

vaginal secretion. Higher  $Q$ -value cut-off means lower fidelity of vaginal secretion candidate genes, and consequently lower accuracy in validation. RT-PCR assay revealed that SERPINB, MMP7 and MSLN were good mRNA markers for vaginal secretion (Fig. 2), however, large sample to sample variations were shown as expected (Figs. 2 and 3). These large variations would cause false negative result in vaginal secretion sample and may require more specific markers for identification of vaginal secretion. Meanwhile, when discussing the inconsistent expression patterns of MMP7 in vaginal secretion samples, it needs to be considered that MMP7 is also reported as a marker for menstrual blood [4]. Because we had no information about the menstrual cycles of sample donors, variations of MMP7 might be a result of hormonal effects.

In some cases, the inconsistency in results of RT-PCR, qRT-PCR, Illumina and Affymetrix microarray is examined, possibly caused by the differences in sequences used as primers and probes. For example, KLK4 was selected as a semen-specific mRNA marker in Illumina microarray platform, but in Affymetrix microarray platform, it showed higher expression in saliva samples (data not shown). Moreover, KLK4 was considered as a saliva-specific marker in RT-PCR result (Fig. 2), but its expression was much higher in semen samples in qRT-PCR result (Supplementary Figure 7). Another example, NKX3-1 was selected as both saliva and semen-specific markers in Affymetrix microarray platform because there were two different probes for the gene on Affymetrix microarray platform (data not shown). Therefore, it is considered

**Table 2**  
Evaluation of four body-fluid-specific mRNA markers.

Gene	Body fluid type	Microarray			qRT-PCR		
		AUC <sup>a</sup>	Sensitivity	Specificity	AUC	Sensitivity	Specificity
PPBP	Blood	1.000	1.000	1.000	0.968	1.000	0.911
FDSCP	Saliva	0.998	0.946	1.000	0.957	0.944	0.982
MSMB <sup>b</sup>	Semen	0.981	0.846	0.980	0.932	0.900	1.000
MSMB <sup>c</sup>		0.997	1.000	0.940			
MSLN	Vaginal secretion	0.857	0.459	0.934	0.857	0.778	0.911

<sup>a</sup> AUC (area under curve) means the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.

<sup>b</sup> 207430\_s\_at.

<sup>c</sup> 210297\_s\_at.

to be important to examine various gene expression studies using different assay platforms and various regions where the probes or primers bind to select genuine representative body-fluid-specific markers.

DNA/RNA co-extraction method tested in this study is especially useful for forensic scientists, because it gives gDNA and RNA simultaneously from the same spot of a sample. gDNA is used for standard human identification based on STR analysis, while RNA gives tissue-specific information which could not be gained from DNA (Supplementary Figure 8 and 9). In spite of the great tissue specificity and all the efforts to use RNA for forensic purpose, it is hard to obtain intact RNA without degradation in body fluid samples (Supplementary Figure 1). RNase is ubiquitous and very stable enzyme that it is crucial to keep the sample in the environment which reduces the RNase activity. However, forensic samples are not under our control before they arrive at the laboratories. Considering the massive degradation and limited amount of forensic samples, continuous experimental attempts will be needed to practically use RNA in crime cases. Multiplex qRT-PCR designed in this study is to conduct body fluid identification with limited amount of forensic samples (Fig. 4). Previous studies explored the possibility of microRNAs for forensic use because of its short length [28,33,34]. Next generation sequencing would be another technological alternative which provides more sensitive detection of much larger number of genes all together [35].

## 5. Conclusion

We successfully obtained four body-fluid-specific mRNA markers: PPBP for blood, FDCSP for saliva, MSMB for semen and MSLN for vaginal secretion. Additionally, many putative mRNA markers were investigated through systematic mRNA profiling. In this study, we validated expression patterns of several previously known mRNA markers, like PPBP, and also found novel mRNA markers for forensic body fluid identification. Finally, multiplex qRT-PCR was designed using the selected mRNA markers and tested in RNA samples simultaneously extracted with gDNA for forensic use. It is expected that our extensive gene expression study and novel body-fluid-specific mRNA markers will advance the use of RNA in forensic field.

## Acknowledgement

This work was supported by the research project for practical use and advancement of forensic DNA analysis (2012) of Supreme Prosecutors' Office, Republic of Korea.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.fsigen.2012.09.001>.

## References

- [1] N. Khaldi, A. Miras, K. Botti, L. Benali, S. Gromb, Evaluation of three rapid detection methods for the forensic identification of seminal fluid in rape cases, *J. Forensic Sci.* 49 (2004) 749–753.
- [2] L.I. Stowell, L.E. Sharman, K. Hamel, An enzyme-linked immunosorbent assay (ELISA) for prostate-specific antigen, *Forensic Sci. Int.* 50 (1991) 125–138.
- [3] Y. Itoh, S. Matsuzawa, Detection of human hemoglobin A (HbA) and human hemoglobin F (HbF) in biological stains by microtiter latex agglutination–inhibition test, *Forensic Sci. Int.* 47 (1990) 79–89.
- [4] M. Bauer, RNA in forensic science, *Forensic Sci. Int. Genet.* 1 (2007) 69–74.
- [5] M. Setzer, J. Juusola, J. Ballantyne, Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains, *J. Forensic Sci.* 53 (2008) 296–305.
- [6] D. Zubakov, M. Kokshoorn, A. Kloosterman, M. Kayser, New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains, *Int. J. Legal Med.* 123 (2009) 71–74.
- [7] C. Haas, E. Hanson, M.J. Anjos, W. Bar, R. Banemann, A. Berti, E. Borges, C. Bouakaze, A. Carracedo, M. Carvalho, V. Castella, A. Choma, G. De Cock, M. Dotsch, P. Hoff-Olsen, P. Johansen, F. Kohlmeier, P.A. Lindenbergh, B. Ludes, O. Maronas, D. Moore, M.L. Morerod, N. Morling, H. Niederstatter, F. Noel, W. Parson, G. Patel, C. Popielarz, E. Salata, P.M. Schneider, T. Sijen, B. Sviezena, M. Turanska, L. Zatkalikova, J. Ballantyne, RNA/DNA co-analysis from blood stains—results of a second collaborative EDNAP exercise, *Forensic Sci. Int. Genet.* (2011).
- [8] A. Bowden, R. Fleming, S. Harbison, A method for DNA and RNA co-extraction for use on forensic samples using the Promega DNA IQ system, *Forensic Sci. Int. Genet.* 5 (2011) 64–68.
- [9] M. Alvarez, J. Juusola, J. Ballantyne, An mRNA and DNA co-isolation method for forensic casework samples, *Anal. Biochem.* 335 (2004) 289–298.
- [10] M. Bauer, D. Patzelt, A method for simultaneous RNA and DNA isolation from dried blood and semen stains, *Forensic Sci. Int.* 136 (2003) 76–78.
- [11] K. Sakurada, T. Akutsu, K. Watanabe, Y. Fujinami, M. Yoshino, Expression of statherin mRNA and protein in nasal and vaginal secretions, *Leg. Med. (Tokyo)* 13 (2011) 309–313.
- [12] C. Haas, E. Hanson, A. Kratzer, W. Bar, J. Ballantyne, Selection of highly specific and sensitive mRNA biomarkers for the identification of blood, *Forensic Sci. Int. Genet.* 5 (2011) 449–458.
- [13] M.L. Richard, K.A. Harper, R.L. Craig, A.J. Onorato, J.M. Robertson, J. Donfack, Evaluation of mRNA marker specificity for the identification of five human body fluids by capillary electrophoresis, *Forensic Sci. Int. Genet.* (2011).
- [14] C. Haas, B. Klessner, C. Maake, W. Bar, A. Kratzer, mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR, *Forensic Sci. Int. Genet.* 3 (2009) 80–88.
- [15] C. Nussbaumer, E. Gharehbaghi-Schnell, I. Korschneck, Messenger RNA profiling: a novel method for body fluid identification by real-time PCR, *Forensic Sci. Int.* 157 (2006) 181–186.
- [16] J. Juusola, J. Ballantyne, Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification, *Forensic Sci. Int.* 135 (2003) 85–96.
- [17] M. Bauer, D. Patzelt, Evaluation of mRNA markers for the identification of menstrual blood, *J. Forensic Sci.* 47 (2002) 1278–1282.
- [18] J. Juusola, J. Ballantyne, Multiplex mRNA profiling for the identification of body fluids, *Forensic Sci. Int.* 152 (2005) 1–12.
- [19] G. Ferri, C. Bini, S. Ceccardi, S. Pelotti, Successful identification of two years old menstrual bloodstain by using MMP-11 shorter amplicons, *J. Forensic Sci.* 49 (2004) 1387.
- [20] M. Bauer, D. Patzelt, Protamine mRNA as molecular marker for spermatozoa in semen stains, *Int. J. Legal Med.* 117 (2003) 175–179.
- [21] M. Visser, D. Zubakov, K.N. Ballantyne, M. Kayser, mRNA-based skin identification for forensic applications, *Int. J. Legal Med.* 125 (2011) 253–263.
- [22] J. Gauvin, D. Zubakov, J. van Rhee-Binkhorst, A. Kloosterman, E. Steegers, M. Kayser, Forensic pregnancy diagnostics with placental mRNA markers, *Int. J. Legal Med.* 124 (2010) 13–17.
- [23] J. Juusola, J. Ballantyne, mRNA profiling for body fluid identification by multiplex quantitative RT-PCR, *J. Forensic Sci.* 52 (2007) 1252–1262.
- [24] R. Shyamsundar, Y.H. Kim, J.P. Higgins, K. Montgomery, M. Jordan, A. Sethuraman, M. van de Rijn, D. Botstein, P.O. Brown, J.R. Pollack, A DNA microarray survey of gene expression in normal human tissues, *Genome Biol.* 6 (2005) R22.
- [25] A. Saito-Hisaminato, T. Katagiri, S. Kakiuchi, T. Nakamura, T. Tsunoda, Y. Nakamura, Genome-wide profiling of gene expression in 29 normal human tissues with a cDNA microarray, *DNA Res.* 9 (2002) 35–45.
- [26] A.I. Su, M.P. Cooke, K.A. Ching, Y. Hakak, J.R. Walker, T. Wiltshire, A.P. Orth, R.G. Vega, L.M. Sapinoso, A. Moqrich, A. Patapoutian, G.M. Hampton, P.G. Schultz, J.B. Hogenesch, Large-scale analysis of the human and mouse transcriptomes, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 4465–4470.
- [27] D. Zubakov, E. Hanekamp, M. Kokshoorn, W. van Ijcken, M. Kayser, Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples, *Int. J. Legal Med.* 122 (2008) 135–142.
- [28] C. Courts, B. Madea, Specific Micro-RNA signatures for the detection of saliva and blood in forensic body-fluid identification, *J. Forensic Sci.* 56 (2011) 1464–1470.
- [29] B.M. Bolstad, R.A. Irizarry, M. Astrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, *Bioinformatics* 19 (2003) 185–193.
- [30] E.W. Klee, Data mining for biomarker development: a review of tissue specificity analysis, *Clin. Lab. Med.* 28 (2008) 127–143, viii.
- [31] J. Schug, W.P. Schuller, C. Kappen, J.M. Salbaum, M. Bucan, C.J. Stoeckert Jr., Promoter features related to tissue specificity as measured by Shannon entropy, *Genome Biol.* 6 (2005) R33.
- [32] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A.J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J.Y. Yang, J. Zhang, Bioconductor: open software development for computational biology and bioinformatics, *Genome Biol.* 5 (2004) R80.
- [33] D. Zubakov, A.W. Boersma, Y. Choi, P.F. van Kuijk, E.A. Wiemer, M. Kayser, MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation, *Int. J. Legal Med.* 124 (2010) 217–226.
- [34] E.K. Hanson, H. Lubenow, J. Ballantyne, Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs, *Anal. Biochem.* 387 (2009) 303–314.
- [35] M.L. Metzker, Sequencing technologies—the next generation, *Nat. Rev. Genet.* 11 (2010) 31–46.