

OPEN

Distinct spectrum of microRNA expression in forensically relevant body fluids and probabilistic discriminant approach

Shuntaro Fujimoto¹, Sho Manabe¹, Chie Morimoto¹, Munetaka Ozeki¹, Yuya Hamano^{1,2}, Eriko Hirai¹, Hirokazu Kotani¹ & Keiji Tamaki¹

MicroRNA is attracting worldwide attention as a new marker for the identification of forensically relevant body fluids. A probabilistic discriminant model was constructed to identify venous blood, saliva, semen, and vaginal secretion, based on microRNA expression assessed via RT-qPCR. We quantified 15 candidate microRNAs in four types of body fluids by RT-qPCR and found that miR-144-3p, miR-451a-5p, miR-888-5p, miR-891a-5p, miR-203a-3p, miR-223-3p and miR-1260b were helpful to discriminate body fluids. Using the relative expression of seven candidate microRNAs in each body fluid, we implemented a partial least squares-discriminant analysis (PLS-DA) as a probabilistic discriminant model and distinguished four types of body fluids. Of 14 testing samples, 13 samples were correctly identified with >90% posterior probability. We also investigated the effects of microRNA expression in skin, semen infertility, and vaginal secretion during different menstrual phases. Semen infertility and menstrual phases did not affect our body fluid identification system. Therefore, the selected microRNAs were effective in identifying the four types of body fluids, indicating that probabilistic evaluation may be practical in forensic casework.

Body fluid identification is crucial in forensic casework for the reconstruction of the criminal activity. Body fluid identification may indicate how the sample was left at the crime scene^{1–4}. For example, saliva may have been deposited by chatting while semen may be attributable to sexual assault. Currently, morphological and enzymatic/immunological tests are performed on samples. However, these methods are difficult to apply successfully to degraded samples². In addition, the sample volumes are often very small and may not suffice for the multiple tests which may be required to distinguish body fluids².

Recently, RNA-based body fluid identification system has been developed. Certain messenger RNAs (mRNA) are expressed in a cell-specific manner^{5–24} and this property has been exploited to establish mRNA-based body fluid identification systems such as ‘Cell typer’^{14,16,17,25}. This system is applicable to various body fluids including the venous blood, saliva, nasal secretion, semen, vaginal secretion, menstrual blood and skin and requires no special instruments in the forensic laboratory. In forensic casework, mRNA may be remained in dried samples even if the samples are aged¹, but some studies have reported that the expression of mRNA is affected by various environmental factors^{26–28}.

In contrast, microRNAs are small molecules (17–22 bases) resistant to environmental factors because they are stored in exosomes²⁹, they have no splicing variants unlike mRNA and their expression also follows a cell-specific pattern^{30,31}. Recent studies have reported that several microRNAs are highly expressed in forensically relevant body fluids^{32–47} and can roughly distinguish semen, blood and epithelial cells. However, it is difficult to discriminate fluids with similar traits such as saliva and vaginal secretion. Sauer *et al.*⁴² proposed linear discriminant analyses for microRNA-based body fluid identification. This method is easy, systematic and comprehensible but their models could only distinguish between two body fluids at a time. Recently, Dørum *et al.*⁴⁷ reported probabilistic discrimination of body fluid based on microRNA sequencing and their method is highly effective for probabilistic

¹Department of Forensic Medicine, Kyoto University Graduate School of Medicine, Yoshida-Konoe-cho, Sakyo-ku, Kyoto, 606-8501, Japan. ²Forensic Science Laboratory, Kyoto Prefectural Police Headquarters, 85-3, 85-4, Yabunouchi-cho, Kamigyo-ku, Kyoto, 602-8550, Japan. Correspondence and requests for materials should be addressed to K.T. (email: ktamaki@fp.med.kyoto-u.ac.jp)

| Body fluid | microRNA | miRBase Accession ID | Amplification efficiency | References |
|-------------------|-------------|----------------------|--------------------------|------------|
| Venous blood | miR-16-5p | MIMAT0000069 | 1.86 | 32 |
| | miR-144-3p | MIMAT0000436 | 1.85 | 35 |
| | miR-451a-5p | MIMAT0001631 | 1.90 | 32 |
| Saliva | miR-200c-3p | MIMAT0000617 | — | 36,46 |
| | miR-203a-3p | MIMAT0000264 | 1.97 | 36 |
| | miR-205-5p | MIMAT0000266 | 1.50* | 32 |
| | miR-223-3p | MIMAT0000280 | 1.95 | 41 |
| | miR-658 | MIMAT0003336 | — | 32 |
| Semen | miR-10a-5p | MIMAT0000253 | 1.86 | 35 |
| | miR-888-5p | MIMAT0004916 | 1.87 | 37 |
| | miR-891a-5p | MIMAT0004902 | 1.90 | 37 |
| Vaginal secretion | miR-124-3p | MIMAT0000422 | — | 32 |
| | miR-155-5p | MIMAT0000646 | 1.96 | 42 |
| | miR-1260b | MIMAT0015041 | 1.98 | 37 |
| | miR-3685 | MIMAT0018113 | — | 42 |
| Reference RNA | 5S-rRNA | | 1.92 | 48 |
| | miR-484 | MIMAT0002174 | 1.92 | 48 |
| | miR-92a-3p | MIMAT0000092 | 1.85 | 48 |

Table 1. Candidate microRNAs and amplification efficiency in qPCR, analysed by LinRegPCR. *Low amplification efficiency; -: amplification failure.

discrimination with the disadvantages of cost and time required. It is preferable that irrelevant factors, such as skin contamination and donor traits, do not interfere with body fluid identification because they do not elucidate how the samples were left at the crime scene or who left them there². However, the actual effects of microRNA expression in skin and individual characteristics on microRNA expression have not yet been verified.

In the present study, microRNAs were quantified by RT-qPCR in forensically relevant body fluids including the venous blood, semen, saliva and vaginal secretion. The aim was to establish a microRNA-based body fluid identification system based on probabilistic discrimination. Furthermore, we investigated the effects of microRNA expression in skin on body fluid identification as well as specific donor traits such as seminal infertility and vaginal secretion changes with menstrual phase.

Results

Selection of candidate RNA based on amplification success and efficiency. Before comparing microRNA expression levels in the body fluids, the RT-qPCR amplification efficiency of candidate microRNAs and three reference RNAs that were selected in our previous study⁴⁸ was determined for all samples. Of the 15 candidate RNAs (Table 1), four candidates were not amplified at all, namely, miR-200c-3p and miR-658 (saliva), and miR-124-3p and miR-3685 (vaginal secretion) (Table 1). In addition, miR-205 (saliva) showed a low amplification efficiency (1.50; Table 1). Therefore, we excluded these microRNAs because they were not suitable for our detection system and used the ten remaining candidates (miR-16-5p, miR-144-3p and miR-451a-5p for venous blood; miR-203a-3p and miR-223-3p for saliva; miR-10a-5p, miR-888-5p and miR-891a-5p for semen; miR-155-5p and miR-1260b for vaginal secretion) in the subsequent analyses.

Comparison of microRNA expression among five body fluids. Figure 1 shows the expression levels of ten microRNAs in each type of body fluid. The microRNAs that were relevant for venous blood, miR-16-5p, miR-144-3p and miR-451a-5p, were highly expressed in venous blood (Figure 1a). The miR-16-5p was highly expressed in all target body fluids whereas miR-144-3p was expressed only in the venous blood and miR-451a-5p showed higher expression in venous blood than that in other body fluids. When the venous blood discrimination threshold was set to a certain value such as $\Delta Cq = 10$ for miR-144-3p and 0 for miR-451a-5p in our dataset, the fluid type was easily distinguished.

The candidate microRNAs for saliva, miR-203a-3p and miR-223-3p, were highly expressed in both the saliva and vaginal secretion (Figure 1b). Hence, it was difficult to qualitatively discriminate between saliva and vaginal secretion using these microRNAs. Nevertheless, this technique could effectively discriminate saliva from venous blood and semen.

The microRNA candidates for semen, miR-10a-5p, miR-888-5p and miR-891a-5p, were highly expressed in semen. The miR-10a-5p was detected in venous blood and vaginal secretion at the same expression level as in semen. The latter two were detected at low expression levels or not detected in other body fluids (Figure 1c). Therefore, miR-888-5p and miR-891a-5p were effective for semen identification.

The microRNAs miR-155-5p and miR-1260b were detected in vaginal secretion (Figure 1d). However, the former was expressed at the same level in all body fluids and was not specific to vaginal secretion whereas the latter was highly expressed in vaginal secretion. On the other hand, it was also highly expressed in semen and saliva. Consequently, it is difficult to identify vaginal secretion using miR-1260b expression alone, but combining the

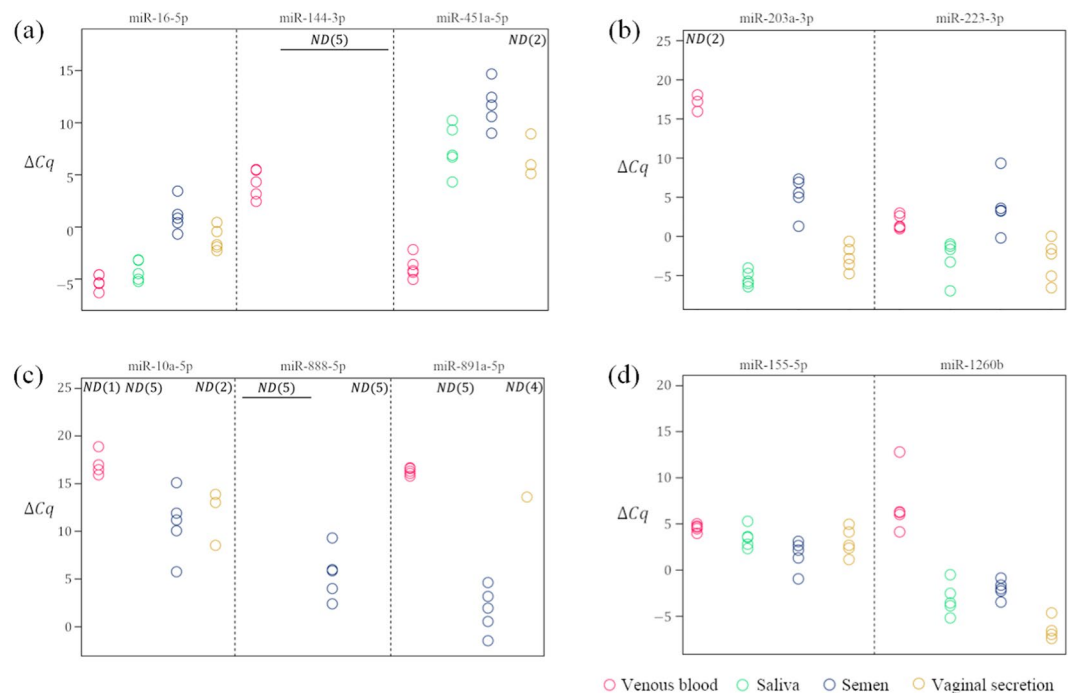


Figure 1. Candidate microRNAs expression in four types of body fluids. (a) Venous blood-relevant microRNAs miR-16-5p, miR-144-3p, and miR-451a-5p; (b) saliva-relevant microRNAs miR-203a-3p, and miR-223-3p; (c) semen-relevant microRNAs miR-10a-5p, miR-888-5p, and miR-891a-5p; (d) vaginal secretion-relevant microRNAs miR-155-5p, and miR-1260b. Red circles: venous blood; green circles: saliva; blue circles: semen; yellow circles: vaginal secretion (all $n = 5$). ‘Not detected’ is abbreviated ND. Number within brackets represents the number of ND samples.

other target microRNAs, such as miR-203a-3p and miR-888-5p, would be effective for identification of vaginal secretion.

We identified seven effective microRNAs (miR-144-3p and miR-451a-5p for venous blood, miR-203a-3p and miR-223-3p for saliva, miR-888-5p and miR-891a-5p for semen and miR-1260b for vaginal secretion) and used them in subsequent examinations.

Effects of microRNA expression in skin, semen infertility and vaginal secretion during different menstrual phases. In forensic practice, skin is sometimes present in crime scene samples and we hoped that expression of our target microRNAs in skin would not disturb the identification of the four types of body fluids. We investigated the effects of microRNA expression in skin by comparing the expression levels of each candidate microRNA in skin samples with those in each body fluid. Several candidate microRNAs were expressed at low levels or not at all in skin (Figure 2a). However, miR-203a-3p, miR-223-3p and miR-1260b (saliva or vaginal secretion) were detected in four out of five skin samples. In one skin sample, miR-451a-5p (venous blood) was highly expressed. Therefore, the expression levels of miR-451a-5p, miR-203a-3p, miR-223-3p and miR-1260b could be affected by microRNA expression in skin if a biological fluid was swabbed from skin, but none of the microRNAs in skin were not expressed at similar levels to those of the target body fluids.

To determine the effect of semen infertility, we collected oligospermic semen (three out of five were azoospermic) and asthenospermic semen and evaluated the expression of each candidate microRNA (Figure 2b). Semen-relevant microRNAs (miR-888-5p and miR-891a-5p) were detected at the same expression levels in these samples as fertile semen ($P > 0.05$; Welch’s *t*-test). The other fluid-relevant microRNAs were also expressed in fertile semen at the same levels as oligospermic and asthenospermic semen ($P > 0.05$; Welch’s *t*-test).

The expression of seven candidate microRNAs was compared between the follicular and luteal phases of vaginal secretion (Figure 2c). There were no significant differences in the expression level of any of the candidate microRNAs between the follicular and luteal phases of vaginal secretion ($P > 0.05$; Welch’s *t*-test).

Probabilistic discrimination of four types of body fluids by PLS-DA. We performed probabilistic discriminant analysis by combining the ΔCq -values for seven microRNAs to identify four types of body fluids. We assigned a partial least squares-discriminant analysis (PLS-DA) to the discriminant model (Supplementary Figure 1). PLS-DA has the advantage in data analysis that the interaction between the variables can be considered by searching for latent variables (PLS components) and the PLS components are adapted to discriminate the training dataset. The model was constructed with 32 training samples and tested using 14 testing samples as shown in Materials and Methods. Figure 3a shows the optimal number of PLS components and indicates that the classification error rate decreased with increasing number of PLS components in all three distance measures (*i.e.*,

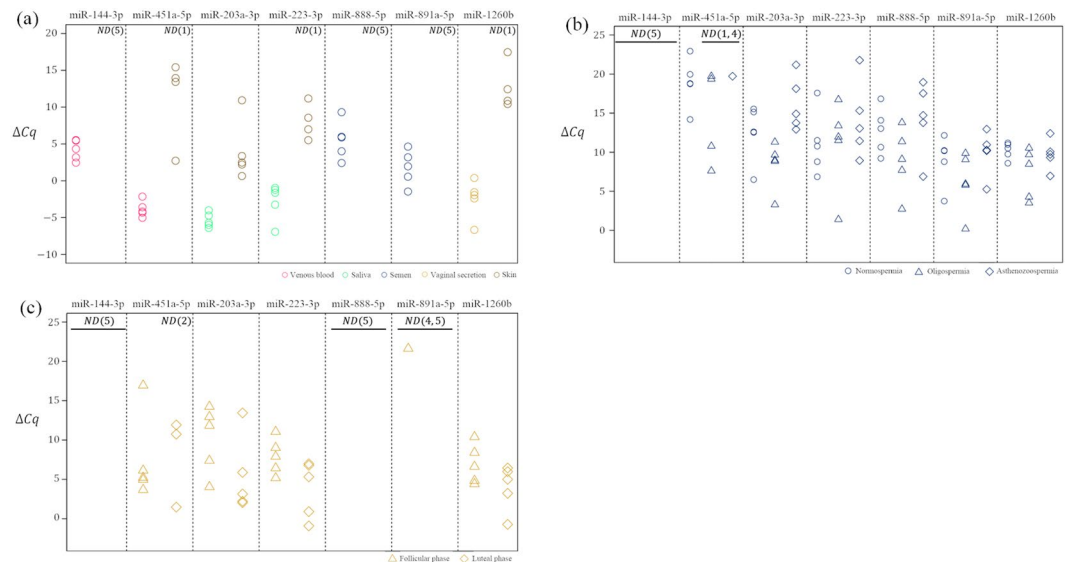


Figure 2. Effects of microRNA expression in skin, semen infertility and different menstrual phases on selected microRNA expression. **(a)** Effects of microRNA expression in skin on body fluid identification. Brown circles: skin; red circles: venous blood; green circles: saliva; blue circles: semen; yellow circles: vaginal secretion. **(b)** Effects of semen infertility compared among normospermic (circles), oligospermic (triangles) and asthenospermic (rectangles). **(c)** Relative microRNA expression in vaginal secretion between follicular (triangles) and luteal (rectangles) phases. $N = 5$ for all samples. 'Not detected' is abbreviated ND. Number within brackets represents the number of ND samples.

maximal prediction value, centroid distance and Mahalanobis distance.) However, the use of >3 PLS components did not markedly change the classification error rate. In general, the model could overfit our training data if many components were included. Therefore, we selected three PLS components to construct the probabilistic discriminant model. The correlation loading plot (Figure 3b) shows the correlation between the explanatory variables and the PLS components. For the model based on three PLS components, the correlation loading plots disclosed that miR-144-3p and miR-451a-5p, miR-203a-3p and miR-223-3p, and miR-888-5p and miR-891a-5p all had same loading value trend. Moreover, PLS component 1 was positively affected by venous blood-specific microRNAs, PLS component 2 was negatively affected by semen-specific microRNAs and PLS component 3 was positively affected by saliva-relevant microRNAs. The saliva markers were more positively correlated with PLS component 2 than with PLS component 3. Venous blood and vaginal secretion makers also positively correlated with PLS component 2, but they had negative effects in PLS component 3. The vaginal secretion-relevant microRNA (miR-1260b) was correlated to miR-203a-3p and miR-223-3p in PLS components 1 and 2, but for PLS component 3 the saliva- and vaginal secretion-relevant microRNAs were negatively correlated. For this reason, PLS component 3 was effective at discriminating between saliva and vaginal secretion. Figure 3c reveals that the four types of body fluids were separately distributed into each category. Venous blood and semen were well separated from the other body fluids whereas saliva and vaginal secretion partially overlapped. Almost all testing samples were correctly plotted in each category, but one vaginal secretion was apparently grouped with the saliva samples.

We also assigned a posterior probability for each body fluid according to Bayes theorem by fitting the PLS components to multivariate normal distribution (Supplementary Figure 1 and Table 2). All testing samples were correctly identified by maximal probability. Venous blood and semen were correctly predicted at approximate 100% posterior probability. Except for one vaginal fluid sample, saliva and vaginal secretion were predicted with $>90\%$ posterior probability. The outlying vaginal fluid showed the posterior probability 77.9% and corresponded to the sample that was apparently plotted in the saliva group (Figure 3c). The posterior odds of vaginal secretion to saliva in the sample was 3.52. Moreover, five samples of azoospermic semen were examined for proof of no effects of semen infertility in probabilistic discrimination (Supplementary Table 1). All five azoospermic semen samples were correctly identified to be semen with over 98% posterior probability.

The four body fluid types were clearly distinguished probabilistically and the calculated probability could prove useful in the reconstruction of the criminal activity by body fluid identification.

Discussion

The present study demonstrated that microRNAs are effective at identifying forensically relevant body fluids. Venous blood is readily identified with miR-144-3p and miR-451a-5p and semen is easily detected using miR-888-5p and miR-891a-5p. Nevertheless, saliva and vaginal secretion were not qualitatively distinguishable. Saliva and vaginal secretion have similar properties in that they contain epithelial cells, indigenous bacteria and mucus. Therefore, they are expected to express the same microRNAs at similar levels. However, miR-203a-3p and miR-223-3p in saliva and the miR-1260b in vaginal secretion were expressed more highly than they were in the other body fluids. Probabilistic discrimination, then, should effectively identify saliva and vaginal secretion.

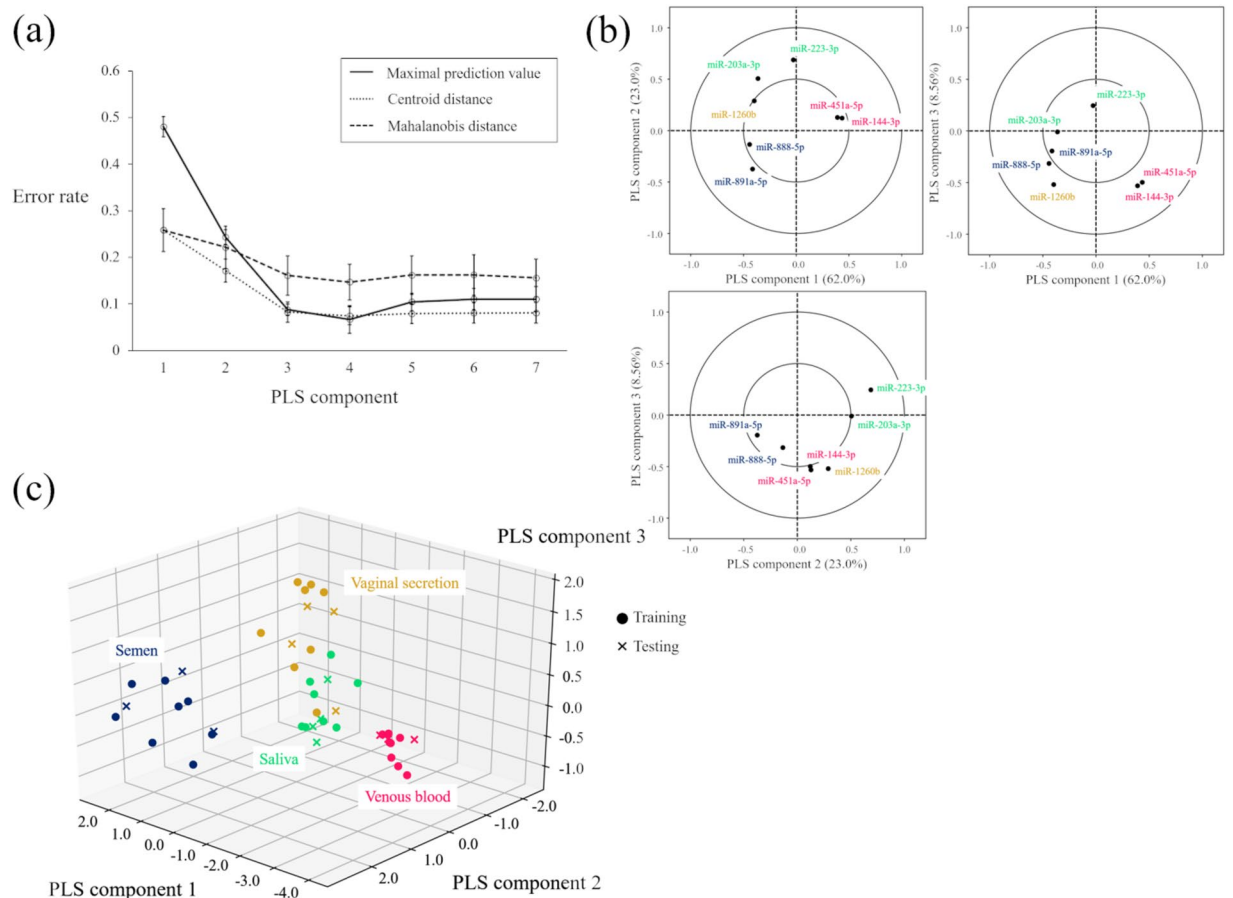


Figure 3. Probabilistic discrimination of four types of body fluids by seven microRNA expressions. (a) Optimal number of PLS components validated by five-fold cross-validation. Horizontal axis is the number of PLS components used. Vertical axis is the prediction error rate based on the training datasets. Solid lines represent the maximal prediction values. Dotted line represents results predicted by centroid distance. Dashed line represents results predicted by Mahalanobis distance. (b) Correlation loading plots between microRNAs and PLS components (1–3). All combinations of the three PLS components are shown. The axes of each plot describe PLS component and the percentage of explained variance. (c) Plots of four types of body fluid samples are shown in PLS component 1–3. Red circles: venous blood ($n = 11$); green circles: saliva ($n = 12$); blue circles: semen ($n = 11$); yellow circles: vaginal secretion ($n = 12$). Training samples indicated in 'o'. Testing samples shown in 'x'.

Certain microRNAs were expressed at levels different from those reported in previous studies^{32,36,42,46}. In addition, four microRNAs were not detected in all body fluids. One possible reason is the difference in the detection methods used. We used a Universal cDNA Synthesis Kit II (Exiqon, Veboek, Denmark) for the RT reaction. It adds a poly (A) tail to the 3'-terminal regions of the targets and transforms the RNAs in all samples to cDNAs. The advantage of this method is that it mitigates the influence of variability among target RNAs in the RT reaction. Certain previous studies used a stem-loop primer that reacts with each target microRNA by short-length hybridisation. This method has a higher specificity for target microRNA than ours, but the efficacy of RT reaction would be various in each target microRNA. Furthermore, it is impractical because the extracted RNA is lost by quantifying the target microRNAs every time. For quantification, the SYBR Green or TaqMan[®] Assays (Applied Biosystems[™], Weiterstadt, Germany) was generally used. TaqMan[®] Assays quantify target microRNAs more specifically than SYBR Green methods. In the present study, however, the use of locked nucleic acid (LNA) technology with amplification primers increased the specificity to the same level as that for the TaqMan[®] Assays. To our knowledge, no studies have explored this discrepancy in microRNA detection between methods; however, certain reports including ours, have supported this finding^{37,42}.

We examined the effects of microRNA expression in skin. Our results showed that miR-203a-3p, miR-223-3p and miR-1260b (saliva and vaginal secretion) were detected in four out of five skin samples. Skin contains epithelial cells, indigenous bacteria and sebum. Some of these may also appear in saliva and vaginal secretion. In the present study, body fluid samples were carefully collected to avoid contaminating them with skin. In forensic practice, however, such contamination is inevitable under certain circumstances, especially when the fluids must be sampled directly from victim body surfaces. In the identification of saliva and vaginal secretion, then, the effect of skin contamination must be considered. In addition, miR-451a-5p was detected at a high expression level in

| Testing sample | Prediction | | | |
|---------------------|------------------------|------------------------|------------------------|------------------------|
| | Blood | Saliva | Semen | Vaginal secretion |
| Blood 1 | ≈1* | ≈0** | ≈0** | 5.12×10^{-17} |
| Blood 2 | ≈1* | ≈0** | ≈0** | 9.36×10^{-15} |
| Blood 3 | ≈1* | ≈0** | ≈0** | 7.29×10^{-19} |
| Saliva 1 | 6.48×10^{-23} | 0.965 | 4.85×10^{-11} | 0.0346 |
| Saliva 2 | 3.66×10^{-48} | 0.982 | 1.53×10^{-5} | 0.0180 |
| Saliva 3 | 2.54×10^{-52} | 0.989 | 4.95×10^{-6} | 0.0106 |
| Saliva 4 | 7.46×10^{-38} | 0.990 | 5.41×10^{-7} | 0.0104 |
| Semen 1 | ≈0** | 1.12×10^{-14} | 0.9999999958 | 4.22×10^{-9} |
| Semen 2 | ≈0** | 1.74×10^{-7} | 0.999955 | 4.50×10^{-5} |
| Semen 3 | ≈0** | 2.96×10^{-17} | 0.99999999941 | 5.91×10^{-10} |
| Vaginal secretion 1 | 7.74×10^{-22} | 0.0633 | 4.31×10^{-21} | 0.937 |
| Vaginal secretion 2 | 2.92×10^{-19} | 0.221 | 6.45×10^{-8} | 0.779 |
| Vaginal secretion 3 | 1.00×10^{-43} | 0.0426 | 5.38×10^{-7} | 0.957 |
| Vaginal secretion 4 | 2.48×10^{-28} | 5.38×10^{-13} | 1.03×10^{-19} | ≈1* |

Table 2. Posterior probability of testing samples assigned by PLS-DA. *Posterior probability >0.9999999999 ; **Posterior probability $<1.0 \times 10^{-100}$.

one skin sample. Four out of five skin samples expressed miR-451a-5p but its expression level in three of these samples was much lower than that of venous blood. The RNA yield was low for the skin sample highly expressing miR-451a-5p (Supplementary Table 2). Moreover, 5S-rRNA of a skin sample was expressed at comparably lower levels (Cq of the sample: 30.56 and mean of Cq of the others: 18.95). The miR-451a-5p upregulation in the skin sample may have been the result of the unstable amplification of a small amount of extracted RNA. Therefore, the stability of the RT reaction and amplification in low template samples must be verified prior to their application in forensic practice.

Moreover, we examined the effects of semen infertility on semen identification. Expression of the candidate microRNAs was not affected by semen infertility. A previous study used miR-888-5p and miR-891a-5p and reported that infertile semen affected semen identification⁴⁹. This finding may be due to the difference in sample size ($n = 144$) compared with the present study ($n = 10$). Nevertheless, this discrepancy can also be explained by the difference in the detection systems used in these studies. The small nucleolar RNAs (SNORD24 and SNORD38B) used as reference RNAs in the previous study might have been inappropriate for semen identification. In a previous study, we found that certain semen samples did not express small nucleolar RNAs⁴⁸. Using our detection systems, we found for our data that miR-888-5p and miR-891a-5p could be used in semen identification regardless of semen infertility. Our data showed high variation of microRNA expression in each category. The variation was mainly due to individual variances because the degree of infertility was varied and another reason could be the use of reference RNA as only 5S-rRNA for normalisation because some samples showed extremely low expression of miR-484 which was used as reference RNA in probabilistic identification of four types of body fluids.

We constructed a probabilistic discrimination model for four types of body fluids based on seven selected microRNAs. This method might have applications in the forensic casework addressed in previous reports on mRNA-based body fluid identification^{24,50}. In a trial, a jury could easily understand the result of the analysis of a stain and evaluate the evidence used in crime reconstruction. Our results showed that venous blood, saliva, semen and vaginal secretion were correctly identified and predicted at high posterior probability ($>90\%$). In addition, azoospermic semen samples were identified to be semen. Therefore, probabilistic evaluation by PLS-DA was effective for discriminating four types of body fluids despite the small number of training samples and regardless of semen infertility. However, the posterior probability for one vaginal fluid sample was substantially lower than that for the other fluids (77.9%). Moreover, the probability that it was misidentified as saliva was 22.1%. These results suggest that saliva and vaginal secretion have similar traits. At 77.9% probability, it could be difficult to conclude that a crime scene sample was vaginal fluid because the posterior odds between vaginal secretion and saliva was only 3.52. In forensic practice, a probabilistic statement may help the jury render a verdict even when the probability is low. The judgment is executed on the basis of a comprehensive estimation of criminal activity by use of body fluids, criminal investigation information and eyewitness testimony^{3,4}. A previous study reported that criminal activity can be probabilistically estimated by a Bayesian network⁵¹. We concur with the probabilistic reconstruction of criminal activity and believe that our method can assign probabilities for body fluids via microRNA expression. Additional training samples and markers such as mRNA and DNA methylation could enhance the discriminant power of our constructed model. To apply it to actual cases, we plan to validate our probabilistic model using aged and/or degraded samples and mixed body fluids with additional training samples in future studies.

In conclusion, we identified seven microRNAs to discriminate four types of forensically relevant body fluids. We assessed forensically practical issues including microRNA expression in skin and donor characteristics such as semen infertility and menstrual cycles. We then proposed a probabilistic approach for distinguishing the body

fluid samples. Our findings support the application of a microRNA-based body fluid identification system to forensic practice and we expect the system to be useful solving forensic challenges.

Materials and Methods

Ethical statement. All volunteers provided written informed consent. This study was approved by the ethics committee of the Graduate School of Medicine of Kyoto University (Approval No. G1053-2) and was conducted according to the principles of the Declaration of Helsinki.

Sample preparation. Five forensically relevant body fluids (venous blood ($n = 11$); saliva ($n = 12$); semen ($n = 16$ including infertile semen); vaginal secretion ($n = 12$, divided into follicular and luteal phases); and skin ($n = 5$)) were collected from volunteers. Sample collection was performed by the same procedure as that used in our previous study⁴⁸, except for the skin, which was collected by swabbing the foreheads of volunteers within 2 h after washing their faces.

Total RNA extraction and RNA yield and integrity measurements were performed by the same protocols used in our previous study⁴⁸. RNA yield and integrity data are listed in the additional information (Supplementary Table 2).

RT-qPCR. The 15 candidate microRNAs were selected from previous studies in which they were highly expressed in body fluids (Table 1)^{32,35–37,41,42,46}. For quantitative analysis, we assigned 5S-rRNA, miR-484 and miR-92a-3p as reference RNAs and evaluated their expression in all body fluids. These were validated in our previous study⁴⁸. To investigate microRNA expression in skin, semen infertility and different menstrual phases, 5S-rRNA was used as a reference RNA because certain samples did not express miR-484 or miR-92a-3p in our preliminary experiment (data not shown).

RT-qPCR was performed with the miRCURY LNATM Universal RT miRNA PCR System (Exiqon) in triplicates for each sample according to the manufacturer's protocol. Template total RNA 10 ng in 10- μ L reaction volume in reverse transcription reaction and 4- μ L of 80-fold-diluted template cDNA was applied in qPCR. RT-qPCR quality was validated against a negative H₂O control, a negative RT-enzyme control and a positive synthetic RNA spike-ins control (Exiqon). Melting analyses and polyacrylamide gel electrophoresis verified that no by-products were generated in the PCR assays.

Cq values were assigned with the LinRegPCR program (January 2016 version)^{52,53}, which set the noise baseline and computed the Cq values and amplification efficiencies. The median of the triplicate Cq values was selected for the subsequent analyses. The exclusion criterion was set to Cq > 35. When any of the triplicate Cq values in a microRNA was >35, we determined that the microRNA was not detected in a sample (N.D.).

Δ Cq were calculated as follows: Δ Cq = Cq (target RNA) – Cq (reference RNA). The Cq (reference RNA) were assigned as the geometric mean of expression of three reference RNAs expression⁵⁴ except for when examining skin samples, semen infertility and the different menstrual phases. In such cases, 5S-rRNA expression was assigned. A negative value of Δ Cq for a microRNA means that it is highly expressed in a sample.

All experiments fully complied with the essential information in the MIQE guidelines⁵⁵ and allowed reliable and unequivocal interpretation of the qPCR results.

Assessing the effects of semen infertility and vaginal secretion at different phases. The semen and vaginal secretion samples were randomly selected so the number of samples was the same in each category ($n = 5$). Semen samples included normospermic semen, oligospermic semen (of which three were azoospermic) and asthenospermic semen and the vaginal secretion samples were taken from the follicular and luteal phases.

Welch's *t*-tests were used to compare the expression levels of the selected microRNAs between normospermic and oligospermic, normospermic and asthenospermic and follicular and luteal vaginal secretions. A *P*-value < 0.05 was considered statistically significant. Statistical tests were performed in R (v.3.5.1)⁵⁶.

Probabilistic discrimination of four types of body fluids. We performed probabilistic discrimination at two steps, PLS-DA and Bayes theorem (Supplementary Figure 1). PLS-DA has various advantages such as a smaller sample size than explanatory variables can be applied, the interaction between the variables can be considered by searching for latent variables (PLS components) and the PLS components are adapted to discriminate the training dataset. The analysis of PLS-DA was executed by the 'mixOmics' package in R^{57,58}. Of the 46 body fluid samples, 32 (eight samples in each body fluid) were used in model building and 14 (three in venous blood and semen and four in saliva and vaginal secretion) were used in probabilistic prediction. In addition, five azoospermic semen samples were tested probabilistically for the examination of no-sperm effects. When no microRNA was detected, a conservative value of Cq = 35 was assigned to the microRNA expression in the sample. To determine the optimal number of PLS components, a five-fold cross-validation was conducted using the 'mixOmics' package. This procedure was repeated for every number of PLS components 1,000 times and the prediction error was calculated by maximal prediction value, centroid distance and Mahalanobis distance.

Probabilistic evaluation was performed on the basis of Bayes theorem. Based on a typical forensic approach using DNA-based personal identification, hypotheses were tested for the four body fluid types (H_i : H_{blood} , H_{saliva} , H_{semen} , $H_{vaginal}$). The results of the microRNA expression were represented as *E*. A posterior probability for the blood sample is shown below:

$$Pr(H_{blood}|E) = \frac{Pr(H_{blood})Pr(E|H_{blood})}{\sum_i Pr(H_i)Pr(E|H_i)}$$

where $Pr(H_{blood}|E)$ is the posterior probability that the crime scene sample is blood given the microRNA expression in the sample and $Pr(H_{blood})$ is the prior probability. The prior probabilities are set equally ($Pr(H_{blood}) = 0.25$). $Pr(E|H_{blood})$ is the likelihood that a blood sample have such a microRNA expression profile. The likelihood was assigned by fitting the PLS components of the training dataset to a multivariate Gaussian distribution for each body fluid type.

References

- Li, R. *Forensic Biology, Second Ed.* (CRC Press, 2015).
- Sijen, T. Molecular approaches for forensic cell type identification: On mRNA, miRNA, DNA methylation and microbial markers. *Forensic science international. Genetics* **18**, 21–32, <https://doi.org/10.1016/j.fsigen.2014.11.015> (2015).
- Taylor, D., Kokshoorn, B. & Biedermann, A. Evaluation of forensic genetics findings given activity level propositions: A review. *Forensic science international. Genetics* **36**, 34–49, <https://doi.org/10.1016/j.fsigen.2018.06.001> (2018).
- Gill, P. *et al.* DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence - Guidelines highlighting the importance of propositions: Part I: evaluation of DNA profiling comparisons given (sub-) source propositions. *Forensic science international. Genetics* **36**, 189–202, <https://doi.org/10.1016/j.fsigen.2018.07.003> (2018).
- Juusola, J. & Ballantyne, J. Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification. *Forensic science international* **135**, 85–96 (2003).
- Juusola, J. & Ballantyne, J. Multiplex mRNA profiling for the identification of body fluids. *Forensic science international* **152**, 1–12, <https://doi.org/10.1016/j.forsciint.2005.02.020> (2005).
- Juusola, J. & Ballantyne, J. mRNA profiling for body fluid identification by multiplex quantitative RT-PCR. *Journal of forensic sciences* **52**, 1252–1262, <https://doi.org/10.1111/j.1556-4029.2007.00550.x> (2007).
- Zubakov, D., Kokshoorn, M., Kloosterman, A. & Kayser, M. New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains. *International journal of legal medicine* **123**, 71–74, <https://doi.org/10.1007/s00414-008-0249-z> (2009).
- Haas, C. *et al.* mRNA profiling for the identification of blood—results of a collaborative EDNAP exercise. *Forensic science international. Genetics* **5**, 21–26, <https://doi.org/10.1016/j.fsigen.2010.01.003> (2011).
- Visser, M., Zubakov, D., Ballantyne, K. N. & Kayser, M. mRNA-based skin identification for forensic applications. *International journal of legal medicine* **125**, 253–263, <https://doi.org/10.1007/s00414-010-0545-2> (2011).
- Lindenbergh, A. *et al.* A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. *Forensic science international. Genetics* **6**, 565–577, <https://doi.org/10.1016/j.fsigen.2012.01.009> (2012).
- Park, S. M. *et al.* Genome-wide mRNA profiling and multiplex quantitative RT-PCR for forensic body fluid identification. *Forensic science international. Genetics* **7**, 143–150, <https://doi.org/10.1016/j.fsigen.2012.09.001> (2013).
- Roeder, A. D. & Haas, C. mRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification. *International journal of legal medicine* **127**, 707–721, <https://doi.org/10.1007/s00414-012-0794-3> (2013).
- van den Berge, M. *et al.* A collaborative European exercise on mRNA-based body fluid/skin typing and interpretation of DNA and RNA results. *Forensic science international. Genetics* **10**, 40–48, <https://doi.org/10.1016/j.fsigen.2014.01.006> (2014).
- Song, F., Luo, H. & Hou, Y. Developed and evaluated a multiplex mRNA profiling system for body fluid identification in Chinese Han population. *Journal of forensic and legal medicine* **35**, 73–80, <https://doi.org/10.1016/j.jflm.2015.08.006> (2015).
- van den Berge, M. & Sijen, T. Extended specificity studies of mRNA assays used to infer human organ tissues and body fluids. *Electrophoresis*, <https://doi.org/10.1002/elps.201700241> (2017).
- van den Berge, M. & Sijen, T. A male and female RNA marker to infer sex in forensic analysis. *Forensic science international. Genetics* **26**, 70–76, <https://doi.org/10.1016/j.fsigen.2016.10.018> (2017).
- Akutsu, T., Watanabe, K., Takamura, A. & Sakurada, K. Evaluation of skin- or sweat-characteristic mRNAs for inferring the human origin of touched contact traces. *Legal medicine (Tokyo)* **33**, 36–41, <https://doi.org/10.1016/j.legalmed.2018.05.003> (2018).
- Albani, P. P. & Fleming, R. Novel messenger RNAs for body fluid identification. *Science & justice* **58**, 145–152, <https://doi.org/10.1016/j.scijus.2017.09.002> (2018).
- Hanson, E., Ingold, S., Haas, C. & Ballantyne, J. Messenger RNA biomarker signatures for forensic body fluid identification revealed by targeted RNA sequencing. *Forensic science international. Genetics* **34**, 206–221, <https://doi.org/10.1016/j.fsigen.2018.02.020> (2018).
- Ingold, S. *et al.* Body fluid identification using a targeted mRNA massively parallel sequencing approach—results of a EUROFORGEN/EDNAP collaborative exercise. *Forensic science international. Genetics* **34**, 105–115, <https://doi.org/10.1016/j.fsigen.2018.01.002> (2018).
- Liang, Q. *et al.* Development of New mRNA Markers for the Identification of Menstrual Blood. *Annals of clinical and laboratory science* **48**, 55–62 (2018).
- Zhang, Y. *et al.* Evaluation of the inclusion of circular RNAs in mRNA profiling in forensic body fluid identification. *International journal of legal medicine* **132**, 43–52, <https://doi.org/10.1007/s00414-017-1690-7> (2018).
- Dorum, G. *et al.* Predicting the origin of stains from next generation sequencing mRNA data. *Forensic science international. Genetics* **34**, 37–48, <https://doi.org/10.1016/j.fsigen.2018.01.001> (2018).
- van den Berge, M., Bhoelai, B., Harteveld, J., Matai, A. & Sijen, T. Advancing forensic RNA typing: On non-target secretions, a nasal mucosa marker, a differential co-extraction protocol and the sensitivity of DNA and RNA profiling. *Forensic science international. Genetics* **20**, 119–129, <https://doi.org/10.1016/j.fsigen.2015.10.011> (2016).
- Humphreys-Beher, M. G., King, F. K., Bunnell, B. & Brody, B. Isolation of biologically active RNA from human autopsy for the study of cystic fibrosis. *Biotechnology and applied biochemistry* **8**, 392–403 (1986).
- Johnson, S. A., Morgan, D. G. & Finch, C. E. Extensive postmortem stability of RNA from rat and human brain. *Journal of neuroscience research* **16**, 267–280, <https://doi.org/10.1002/jnr.490160123> (1986).
- Lin, M. H., Jones, D. F. & Fleming, R. Transcriptomic analysis of degraded forensic body fluids. *Forensic science international. Genetics* **17**, 35–42, <https://doi.org/10.1016/j.fsigen.2015.03.005> (2015).
- Valadi, H. *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature cell biology* **9**, 654–659, <https://doi.org/10.1038/ncb1596> (2007).
- Sood, P., Krek, A., Zavolan, M., Macino, G. & Rajewsky, N. Cell-type-specific signatures of microRNAs on target mRNA expression. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 2746–2751, <https://doi.org/10.1073/pnas.0511045103> (2006).
- Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414, <https://doi.org/10.1016/j.cell.2007.04.040> (2007).
- Hanson, E. K., Lubenow, H. & Ballantyne, J. Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. *Analytical biochemistry* **387**, 303–314, <https://doi.org/10.1016/j.ab.2009.01.037> (2009).
- Vaz, C. *et al.* Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood. *BMC genomics* **11**, 288, <https://doi.org/10.1186/1471-2164-11-288> (2010).

34. Weber, J. A. *et al.* The microRNA spectrum in 12 body fluids. *Clinical chemistry* **56**, 1733–1741, <https://doi.org/10.1373/clinchem.2010.147405> (2010).
35. Zubakov, D. *et al.* MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *International journal of legal medicine* **124**, 217–226, <https://doi.org/10.1007/s00414-009-0402-3> (2010).
36. Courts, C. & Madea, B. Specific micro-RNA signatures for the detection of saliva and blood in forensic body-fluid identification. *Journal of forensic sciences* **56**, 1464–1470, <https://doi.org/10.1111/j.1556-4029.2011.01894.x> (2011).
37. Wang, Z. *et al.* Screening and confirmation of microRNA markers for forensic body fluid identification. *Forensic science international. Genetics* **7**, 116–123, <https://doi.org/10.1016/j.fsigen.2012.07.006> (2013).
38. Cheng, L., Sun, X., Scicluna, B. J., Coleman, B. M. & Hill, A. F. Characterization and deep sequencing analysis of exosomal and non-exosomal miRNA in human urine. *Kidney international* **86**, 433–444, <https://doi.org/10.1038/ki.2013.502> (2014).
39. Hanson, E. K., Mirza, M., Rekab, K. & Ballantyne, J. The identification of menstrual blood in forensic samples by logistic regression modeling of miRNA expression. *Electrophoresis* **35**, 3087–3095, <https://doi.org/10.1002/elps.201400171> (2014).
40. Hu, L., Wu, C., Guo, C., Li, H. & Xiong, C. Identification of microRNAs predominately derived from testis and epididymis in human seminal plasma. *Clinical biochemistry* **47**, 967–972, <https://doi.org/10.1016/j.clinbiochem.2013.11.009> (2014).
41. Park, J. L. *et al.* Microarray screening and qRT-PCR evaluation of microRNA markers for forensic body fluid identification. *Electrophoresis* **35**, 3062–3068, <https://doi.org/10.1002/elps.201400075> (2014).
42. Sauer, E., Reinke, A. K. & Courts, C. Differentiation of five body fluids from forensic samples by expression analysis of four microRNAs using quantitative PCR. *Forensic science international. Genetics* **22**, 89–99, <https://doi.org/10.1016/j.fsigen.2016.01.018> (2016).
43. Seashols-Williams, S. *et al.* High-throughput miRNA sequencing and identification of biomarkers for forensically relevant biological fluids. *Electrophoresis* **37**, 2780–2788, <https://doi.org/10.1002/elps.201600258> (2016).
44. Sirker, M., Fimmers, R., Schneider, P. M. & Gomes, I. Evaluating the forensic application of 19 target microRNAs as biomarkers in body fluid and tissue identification. *Forensic science international. Genetics* **27**, 41–49, <https://doi.org/10.1016/j.fsigen.2016.11.012> (2016).
45. Mayes, C., Seashols-Williams, S. & Hughes-Stamm, S. A capillary electrophoresis method for identifying forensically relevant body fluids using miRNAs. *Legal medicine (Tokyo)* **30**, 1–4, <https://doi.org/10.1016/j.legalmed.2017.10.013> (2017).
46. Wang, Z. *et al.* Identification of Saliva Using MicroRNA Biomarkers for Forensic Purpose. *Journal of forensic sciences* **60**, 702–706, <https://doi.org/10.1111/1556-4029.12730> (2015).
47. Dørum, G. *et al.* Predicting the origin of stains from whole miRNome massively parallel sequencing data. *Forensic science international. Genetics* **40**, 131–139, <https://doi.org/10.1016/j.fsigen.2019.02.015> (2019).
48. Fujimoto, S. *et al.* Optimal small-molecular reference RNA for RT-qPCR-based body fluid identification. *Forensic science international. Genetics* **37**, 135–142, <https://doi.org/10.1016/j.fsigen.2018.08.010> (2018).
49. Tian, H. *et al.* Semen-specific miRNAs: Suitable for the distinction of infertile semen in the body fluid identification? *Forensic science international. Genetics* **33**, 161–167, <https://doi.org/10.1016/j.fsigen.2017.12.010> (2018).
50. de Zoete, J., Curran, J. & Sjerps, M. A probabilistic approach for the interpretation of RNA profiles as cell type evidence. *Forensic science international. Genetics* **20**, 30–44, <https://doi.org/10.1016/j.fsigen.2015.09.007> (2015).
51. Biedermann, A. & Taroni, F. Bayesian networks for evaluating forensic DNA profiling evidence: a review and guide to literature. *Forensic science international. Genetics* **6**, 147–157, <https://doi.org/10.1016/j.fsigen.2011.06.009> (2012).
52. Ruijter, J. M. *et al.* Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic acids research* **37**, e45, <https://doi.org/10.1093/nar/gkp045> (2009).
53. Ruijter, J. M. *et al.* LinRegPCR (2014.x) Analysis of quantitative RT-PCR data, Amsterdam, the Netherlands. <http://LinRegPCR.nl/>.
54. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, Research0034.1, <https://doi.org/10.1186/gb-2002-3-7-research0034> (2002).
55. Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry* **55**, 611–622, <https://doi.org/10.1373/clinchem.2008.112797> (2009).
56. R Core team. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
57. Lê Cao, K. A., Boitard, S. & Besse, P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC bioinformatics* **12**, 253, <https://doi.org/10.1186/1471-2105-12-253> (2011).
58. González, I., Lê Cao, K. A., Davis, M. J. & Déjean, S. Visualising associations between paired ‘omics’ data sets. *BioData mining* **5**, 19, <https://doi.org/10.1186/1756-0381-5-19> (2012).

Acknowledgements

This work was supported by a Grant-in-Aid for Challenging Exploratory Research (JSPS KAKENHI Grant No. JP16K15401). We thank Dr. Kentaro Ichioka (Ichioka Urological Clinic) for the sample collection and the diagnosis of azoospermia. We also thank Prof. Noriomi Matsumura (Department of Gynecology and Obstetrics, Kinki University Faculty of Medicine) for the sample collection.

Author Contributions

Designed the study: K.T. Carried out the experiment, analysed the data and wrote the manuscript: S.F. Data interpretation: S.M., C.M. and Y.H. Collected the samples: S.F., C.M. and E.H. Critically reviewed the manuscript: S.M., M.O. and H.K.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-50796-8>.

Competing Interests: The authors declare no competing interests.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019