Advancing forensic RNA profiling

Berge, M.W. van den

Chapter 2

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 (2016) 119-129

Margreet van den Berge
Bryan Bhoelai
Joyce Harteveld
Anuska Matai
Titia Sijen

Abstract

The forensic identification of human body fluids and tissues by means of messenger RNA (mRNA) profiling is a long studied methodology that is increasingly applied to casework samples. Previously, we have described an mRNA multiplex system that targets blood, saliva, semen, menstrual secretion, vaginal mucosa and skin [1-2]. In this study we consider various topics to improve this mRNA profiling system or its use and adapt the method accordingly. Bodily secretions that may be encountered at a crime scene whilst not targeted by the multiplex –*id est* nasal mucosa, sweat, tears, faeces and urine- were examined for false positive signals. The results prompted us to identify a nasal mucosa marker that allows the discrimination of nasal mucosa from saliva or vaginal mucosa and nosebleed blood from peripheral blood. An updated version of the multiplex was prepared to which the nasal mucosa marker was added and in which markers for semen, vaginal mucosa and blood were replaced. *Lactobacillus* markers were regarded unsuitable as replacement for vaginal mucosa mRNA markers because of background signals on penile swabs that appeared devoid of female DNA. Furthermore, we provide approaches to deal with highly unbalanced mixtures. First, a differential extraction protocol was incorporated into a co-extraction protocol to allow DNA and RNA analysis of separated non-sperm and sperm fractions. In a second approach, besides the standard multiplex, a customized multiplex is used which excludes markers for prevailing cell types. This allows the use of lower cDNA inputs for the prevailing cell types and higher inputs for cell types that appear masked. Additionally, we assessed the relation between the percentage of alleles or markers detected in DNA or RNA profiles when decreasing sample amounts are analysed. While blood, saliva, semen and menstrual secretion show the trend that DNA profiling is more sensitive than RNA profiling, the reverse is seen for skin and variable results occur for vaginal and nasal mucosa. Lastly, we show that replicates are useful for interpretation of RNA data, as variations can be found even for true technical replicates. Increased numbers of replicates (over four) do, however, not cancel out the impact of this variation on data interpretation. Overall, the results of this study further forensic RNA profiling.

Introduction

Messenger RNA (mRNA) profiling for the purpose of human body fluid and organ tissue identification has been investigated intensively in the last decade [1-11]. The majority of mRNA profiling systems focus on the identification of body fluids most commonly encountered at a crime scene like blood, saliva, vaginal mucosa, menstrual secretion and semen. Other bodily secretions like nasal mucosa and faeces are generally not targeted by mRNA profiling systems as these cell types are less frequently relevant in forensics [10]. Awareness of possible cross-reactivity of mRNA markers to these secretions is, however, important to optimise data interpretation. In case of cross-reactivity, the addition of a marker specific to the cross-reacting cell type may be opportune. Studies describe, for example, false positive signals for vaginal mucosa and saliva mRNA markers in nasal mucosa [12-13]. Thus we assessed the level of cross-reactivity with our marker system and searched for a nasal mucosa-specific marker to aid in the distinction of nasal mucosa, saliva and vaginal mucosa. mRNA markers specific to vaginal mucosa are known to be particularly difficult to find as many of the target genes are expressed in a broader range of mucous membrane-enriched areas [14]. A suggested alternative approach for the identification of vaginal mucosa is by use of bacterial markers such as *Lactobacillus* species [15-19]. The suitability of bacterial markers for this purpose has been questioned as they are reported to respond also in non-vaginal samples [20]. Evidentiary samples frequently examined for the presence of vaginal mucosa are penile swabs and fingernail samples. We assessed the presence of *Lactobacillus* species on penile swabs lacking a detectable female DNA source. Furthermore we searched for ways to facilitate mRNA analysis of samples with highly deviating cell type ratios such as sexual assault samples comprising sperm and (female) epithelial cells. Separate analysis of DNA in the non-sperm (NF) and sperm fraction (SF) is commonly achieved by use of differential extraction. We explored the possibility of incorporating co-extraction of RNA into the differential extraction procedure so that next to DNA, RNA of the NF and SF can be analysed separately. In addition, we describe a method to perform differential analysis of RNA extracts post extraction. Over-amplification of markers for prevailing cell types can interfere with the detection of markers for underlying cell types. We assessed for the possibility of generation informative RNA profiles for underlying cell types by analysing markers for prevailing and underlying cell types in separate multiplexes. Furthermore, we examined the relation between the sensitivities of RNA and DNA profiling in single source samples of various donors, which is useful when interpreting both the DNA and the RNA results for a sample. Some of these samples were used to assess the effect of adding RNA profiling replicates in the RNA data interpretation system that uses a $x=n/2$ guideline [29].

Materials and methods

Sample collection

A set of control samples for body fluids and tissues commonly encountered in forensic cases was taken from a previous study [1]. This control set includes blood, saliva, vaginal mucosa, menstrual secretion, semen and skin samplings from four individuals. The four semen samples represent two samples from fertile and two samples from vasectomised donors.

Fresh nasal mucosa samples from 22 individuals were taken from both nostrils, using a single cotton swab per nostril. Seven of these individuals suffered from a cold. A total of 11 nosebleed samples from six donors were collected on tissue paper. Sweat, tear and urine samples from 10 donors were collected on cotton swabs while attempting to avoid skin contact. Each donor contributed all three body fluids. Swabs were air-dried and stored at room temperature in the dark until used. Faeces samplings were taken from specimens that had been stored at -20 °C. A total of 10 samples were collected and stored at -20 °C until further processing. Penile swabs were collected by 20 donors using 4N6FLOQSwabs™ with active drying system (Copan Diagnostics) [21]. Samples used for this study were collected with informed consent of the voluntary donors whose cell material was used.

DNA/RNA extraction, DNA quantification, ethanol precipitation, reverse transcription

Swabs were processed entirely, while sections of approximately 1 cm2 were excised from the nosebleed tissues and cut into small pieces before extraction. DNA/RNA co-extraction, DNase treatment, DNA quantification, ethanol precipitation and reverse transcription were performed as described in Lindenbergh et al. [1]. RNA extracts were ethanol-precipitated prior to reverse transcription when the total DNA yield of a sample was below 1 ng and processed as described in [2]. Differential co-isolation was performed using a customized mild lysis buffer for the separation of non-sperm fraction (NF) and sperm fraction (SF). This buffer is composed of phosphate buffered saline (PBS) buffer supplemented with 1.6 mg Proteinase K (20 mg/mL, QIAGEN) and 10 µM Ribonucleoside Vanadyl Complex (New England Biolabs), which is a potent inhibitor of various ribonucleases. Swabs are incubated in 504 µL of this mild lysis buffer for 20 min at 56 °C after which the lysate is separated from the carrier material using a QIAshredder Column (QIAGEN). The NF and SF are separated during 5 min centrifugation at 11.000 rpm. The pellet (SF) is washed twice using 400 µL PBS buffer with 10 µM Ribonucleoside Vanadyl Complex, centrifuging at 13.200 rpm for 5 min to retain the pellet. The sperm pellet is processed further into a DNA and RNA

extract according to standardized protocols [1], starting with a stringent lysis using the Lysis/Binding buffer provided in the mirVana miRNA Isolation Kit (Ambion). The supernatant (NF) is supplemented with Lysis/Binding buffer to a final volume of 400 µL and processed further according to standardized protocols [1] starting with the addition of the Homogenate Additive (mirVana miRNA Isolation Kit, Ambion).

RNA profiling

Several multiplexes were used in this study. The first experiments used the body fluid and skin typing multiplex described in [2], which is a 20-plex denoted Cell-typer V2 that has evolved from a 19-plex (denoted Cell-typer V1) described in [1]. As a result of the presented study, the multiplex was further updated as described in Table 1, and this 19-plex (denoted Cell-typer V3) was used in the later experiments. The changes comprise the following: 1) replacement of blood marker AMICA1, a leukocyte marker, for the more sensitive erythrocyte marker ALAS2; 2) addition of nasal mucosa marker BPIFA1 to aid in the distinction of vaginal mucosa, saliva and nasal mucosa (BPIFA1 was selected from five candidate nasal mucosa markers namely BPIFA1, BPIFB1, SCGB1A1, C6orf58, PPP1R9B [22-24]); 3) addition of KLK3 as a second seminal fluid marker to increase the chance of detecting semen from azoospermic males; 4) replacing vaginal mucosa marker HBD1, which is not functioning well in multiplex, for MYOZ1, which does function in multiplex; 5) removing less informative markers LOR (third skin marker, cross reacting with vaginal mucosa [2]), KRT4 (general mucosa marker, limited informative value) and GAPDH (third housekeeping marker, with large amplicon and thus less responsive in samples with degradation); 6) shifting the signal for menstrual secretion marker MMP7 by lengthening the amplicon with 1 nucleotide to distinguish the MMP7 signals from trailing signals of saliva marker HTN3 [2]; 7) lowering primer concentrations of menstrual secretion marker MMP10 and vaginal mucosa marker CYP2B7P1 to improve multiplex balance. Example electropherograms obtained when using Cell-typer V3 on target body fluids are shown in Supplementary Figure 1.

In addition a multiplex, denoted Lactoplex, was used which is an assembly of 16S-rRNA genes for four different *Lactobacillus* species, namely *L. crispatus*, *L. gasseri*, *L. jensenii* and *L. iners*. These markers are supplemented with mRNA markers for vaginal mucosa (HBD1, MUC4, CYP2B7P1), skin (CDSN, LOR, LCE1C) and the housekeeping markers ACTB and 18S-rRNA [1-2]. *Lactobacillus* primer sequences were adopted from literature [27-28] with minor adjustments to the forward primer of the *L. jensenii* marker to increase the annealing temperature (Supplementary Table 1).

PCR amplification and product detection for all RNA analyses were performed according to standardized protocols [1]. A serial cDNA input (*e.g.* 0.2, 1 and 4 µL cDNA) PCR was used to determine the input providing an informative RNA profile. Subsequently, this input was used to generate a total of four PCR replicates per sample

(of which at least three are true technical replicates, depending on whether the selected input resided in the existing input series).

Table 1. Primer sequences for the different mRNA markers residing in the Cell-typer (V3) multiplex.

Marker name	Tissue	[primer] µM	Forward primer (5'-3') Reverse primer (5'-3')	Size (bp)	Dye	Reference
ALAS2	Blood	0.04	TTCTGCACCAGAAGGACTCAGCC [a] TAAATCTCGCACCCTGGCAGGATC	103	FAM™	[b]
CD93	Blood	0.25	ACCAGTACAGTCCGACAC TTGCTAAGATTCCAGTCCAG	151	NED™	[1]
HBB	Blood	0.035	GCACGTGGATCCTGAGAAC ATGGGCCAGCACACAGAC	61	FAM™	[1]
HTN3	Saliva	0.2	GCAAAGAGACATCATGGGTA GCCAGTCAAACCTCCATAATC	134	VIC®	[1]
STATH	Saliva/ nasal mucosa	0.3	TTTGCCTTCATCTTGGCTCT CCCATAACCGAATCTTCCAA	93	FAM™	[1]
BPIFA1	Nasal mucosa	0.2	CAAGTGAATACGCCCCTGGTCG GAATGGGTGCAGTCACCAAGGAC	131	PET™	[b]
KLK3	Seminal fluid	0.05	GACGTGGATTGGTGCTGCACC CTTCTCGCACTCCCAGCCTC	64	PET™	[b]
SEMG1	Seminal fluid	0.8	GGAAGATGACAGTGATCGT CAACTGACACCTTGATATTGG	121	FAM™	[1]
PRM1	Spermatozoa	0.3	AGACAAAGAAGTCGCAGAC TACATCGCGGTCTGTACC	91	NED™	[1]
CYP2B7P1	Vaginal mucosa	0.8[c]	AGTCTACCAGGGATATGGCATG CTATCAGACACTGAGCCTCGTCC	146	VIC®	[2]
MUC4	Vaginal mucosa	0.8	CTGCTACAATCAAGGCCA AAGGGAAGTTCTAGGTTGAC	141	FAM™	[1]
MYOZ1	Vaginal mucosa	0.8	GGGTTGGTGAGACAGGATCA TTTTCCCATGGGGAAATATAGGT	88	VIC®	[b]
MMP7	Menstrual secretion	0.8	GAACAGGCTCAGGACTATCTC TTAACATTCCAGTTATAGGTAGGCC	127	VIC®	[1]
MMP10	Menstrual secretion	0.1[c]	GCATCTTGCATTCCTTGTGCTGTTG GGTATTGCTGGGCAAGATCCTTGTT	107	VIC®	[2]
MMP11	Menstrual secretion	0.4	CAACCGACAGAAGAGGTTCG GAACCGAAGGATCCTGTAGG	76	NED™	[1]
CDSN	Skin	0.6	CTGGCTGGTCTCCTCCTG GGGTCCTTACAAGGGTCTGA	71	VIC®	[1]
LCE1C	Skin	0.02	TGTGACCCCGCTCCTGAATCCG CTTGGGAGGGCACTTGGGGGTG	99	NED™	[2]
ACTB	Housekeeping	0.2	TGACCCAGATCATGTTTGAG CGTACAGGGATAGCACAG	75	PET™	[1]
18S-rRNA	Housekeeping	0.025	CTCAACACGGGAAACCTCAC CGCTCCACCAACTAAGAACG	110	PET™	[1]

[a] Underlined nucleotides are 5' tails added to improve multiplex spacing
[b] Developed for this study using Ensembl and NCBI primer blast [24][25]
[c] Primer concentrations lowered compared to [2]

PCR products were purified [1] prior to detection using a 3130XL Genetic Analyzer (Life Technologies). Lactoplex amplification products were analysed using POP-7 (Life Technologies) separation matrix. All other RNA products were analysed using POP-4 (Life Technologies) as it became apparent that the average peak heights (relative fluorescence units, rfus) increased approximately 2-fold compared to POP-7, which can result in an increased number of detected markers (data not shown). Profile analysis was performed using Genemapper ID-X version 1.1.1 (Life Technologies) with a detection threshold of 150 relative fluorescence units.

RNA data interpretation was performed using the four PCR replicates and according to the $x=n/2$ rule as described in [29]. This method compares the number of observed (x) to the number of theoretically possible peaks (n) in all replicates. A cell type is scored "observed" when at least half of the possible peaks are observed ($x \geq n/2$), denoted "sporadically observed" when less than half of the possible peaks are observed ($0 < x < n/2$) and scored "not observed" when no peaks are detected ($x=0$). Cell types which are co-expressed with other detected cell types are scored "(sporadically) observed and fits" (*e.g.* vaginal mucosa when menstrual secretion is detected). "Sporadically observed" signals are generally regarded not reliable; "and fits" scorings are generally regarded not present as such.

DNA profiling

DNA profiles were generated using the AmpFℓSTR® NGM™ PCR Amplification Kit (Life Technologies) using a maximum of 500 pg DNA input based on quantification as described in [1]. PCR products were separated according to standardized protocols [1] using a 3130XL Genetic Analyzer (Life Technologies) with POP-4 (Life Technologies) separation matrix. Profile analysis was performed using Genemapper ID-X version 1.1.1 (Life Technologies) using a detection threshold of 50 rfus.

DNA/RNA sensitivity

The relation between the sensitivities of DNA and RNA profiling was examined using blood, saliva, vaginal mucosa, menstrual secretion, semen, skin and nasal mucosa samples as single cell types and four donors per sample type. DNA and RNA extracts were processed according to standardized protocols [1]. DNA extracts were quantified, diluted to 50 pg/µL after which 10 µL (500 pg), 2.5 µL (125 pg) and 1 µL (50 pg) inputs were used for STR profiling. The RNA extract of a sample was diluted in correspondence to the dilution applied for the DNA extracts and 10 µL was used for reverse transcription (it was previously established that this would result in full RNA profiles with 2 µL cDNA, which means that six times less RNA extract than DNA extract resides in the highest input of each analysis: RNA is extracted in 60 µL and DNA in 100 µL (RNA 1.67 times more concentrated) but only 2 of the 20 µL cDNA

preparation is used while the DNA dilution is straightforwardly used). RNA profiles were generated using 2 µL, 0.5 µL and 0.2 µL cDNA, which represent the same serial steps (1/1, 1/4, 1/10) as applied with DNA profiling. RNA analysis (Cell-typer V3) was performed using four technical PCRs replicates per input to allow profile interpretation as described in section "RNA profiling". Only cell type specific markers are regarded to determine the percentage of detected mRNA markers (for example in menstrual secretion samples, vaginal secretion and blood markers are not regarded).

Data interpretation using varying numbers of RNA profiling replicates

Three samples of the sample set described in section "DNA/RNA sensitivity" were additionally used to assess the effect of additional RNA replicates on RNA data interpretation using the $x=n/2$ guideline [29] as described in section "RNA profiling". The selected samples showed a range in the percentage of markers detected when three different cDNA inputs are used (semen donor A and C and menstrual secretion donor B in Supplementary Figure 2). Both from the existing cDNA batch (preparation described in section "DNA/RNA sensitivity") and from a newly prepared cDNA batch four replicates were generated with 0.2 and 0.5 µL cDNA inputs. Regarding the 2 µL cDNA input described in section "DNA/RNA sensitivity", four replicates were generated using the new cDNA batch but only one additional replicate could be prepared for the existing cDNA batch as insufficient cDNA remained. In total three data sets were compared for both body fluids, *i.e.* Set 1: Initial data (section "DNA/RNA sensitivity"), four replicates each input: 0.2, 0.5 and 2.0 µL cDNA; Set 2: Existing cDNA batch, four replicates with 0.2 and 0.5 µL inputs, 1 replicate with 2.0 µL input; Set 3: New cDNA batch prepared from the same RNA sample, four replicates each input: 0.2, 0.5 and 2.0 µL cDNA;. Thus, 12 RNA profiles were obtained for the 0.2 and 0.5 µL cDNA inputs and 9 profiles for the 2 µL input. The individual sets and combinations of sets were used to assess the effect of RNA profiling replicate number on the percentage of detected markers and the associated interpretation scores using the $x=n/2$ guideline.

Results and discussion

mRNA analysis of nasal mucosa samples and the identification of a nasal mucosa marker

The risk of false-positive signals caused by body fluids not targeted in our starting point mRNA assay (Cell-typer V2 [2]) was assessed on 22 nasal mucosa, 11 nosebleed specimens and 10 samples each for sweat, tear, faeces and urine donations. Analysis was performed using four PCR replicates and profile interpretation occurred accordingto

the '*x*=*n*/2' rule as previously described.

First the 22 nasal mucosa samples were analysed. Positive saliva signals were observed in all nasal mucosa samples: STATH in 84 and HTN3 in one of the 88 profiles, which summarised to an "observed" scoring for saliva in 86% of the samples (Figure 1A). These results are consistent with those described by Sakurada et al. [12] and Xu et al. [13], the latter regarding STATH a nasal secretion marker. General mucosa marker KRT4 was additionally detected in the majority of the nasal mucosa profiles. Regarding the three blood makers, signals were regularly detected for the two leukocyte markers (AMICA1 and CD93) and hardly for the erythrocyte marker (HBB, Figure 1B), which can be explained by the fact that nasal mucosa contains plasma leaking from blood vessels [12]. Vaginal mucosa markers were detected in many of the samples that originated from both female and male donors. Similar results have previously been reported by [13] and [14] for MUC4 and HBD1. We additionally analysed CYP2B7P1, which also responded in these nasal samples. Presence of leukocyte or vaginal marker signals did not show a relation to donors with (*n*=7) or without (*n*=15) a cold or donor gender (16 females, 6 males, data not shown). Skin was scored "observed" in one of the samples, which may be due to contact to the nose while sampling. None of the samples scored "observed" for menstrual secretion or semen (Figure 1).

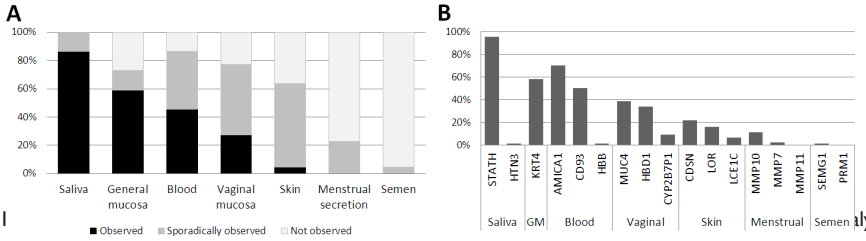


Figure 1 **lysis was performed in 4-fold to allow profile interpretation according to the *x*=*n*/2 guideline. (A) Scoring into categories for 22 samples. (B) Responsiveness of individual markers in 88 profiles (GM = General mucosa).**

STATH is thus both responding in saliva and nasal samples and only the absence of HTN3 would mark the difference between saliva and nasal mucosa. Using a negative marker is clearly an unfavourable situation, so we focussed on the identification of a marker giving a specific response for nasal mucosa. This decision was further driven by the observation that nasal samplings appear to contain high amounts of human cell material as high DNA yields (36 to >6400 ng) were obtained. Five candidate markers were selected from literature [22-24] (BPIFA1, BPIFB1, SCGB1A1, C6orf58, PPP1R9B, primer details in Supplementary Table 2) and their suitability to act as nasal mucosa-specific marker was assessed. Performance of the markers on target (thus nasal mucosa) samples was assessed using a range of primer concentrations (0.02 to 1.6 µM), for which markers were amplified together with housekeeping

marker 18S-rRNA in its standard primer concentration. This led to the exclusion of SCGB1A1 and C6orf58 from further analysis, as these markers performed poorly on target tissues (data not shown). Remaining markers were incorporated into the Cell-typer multiplex (Cell-typer V2)[2] to assess multiplex performance on non-target cell types (blood, saliva, vaginal mucosa, menstrual secretion, semen fertile/ sterile and skin, four donors each, except for semen sterile for which two donors were assessed). Analysis used four PCR replicates per sample. PPP1R9B and BPIFB1 were discarded because of cross-reactivity in blood (PPP1R9B), saliva (BPIFB1) and menstrual secretion (PPP1R9B and BPIFB1) was observed (data not shown). BPIFA1 was selected as a nasal mucosa marker as no cross-reactivity was observed in any of the assessed body fluids (Supplementary Table 3). The 22 nasal samples were re-analysed with a multiplex into which BPIFA1 was fitted (Cell-typer V3, Table 1). Guidelines for the interpretation of saliva/ nasal mucosa signals using shared marker STATH, saliva marker HTN3 and nasal mucosa marker BPIFA1 are described in Table 2.

Table 2. Guidelines for the interpretation of saliva/nasal mucosa signals using four RNA profiles and the *x*=*n*/2 rule. Scoring of saliva is performed using HTN3 (saliva marker) and STATH (shared saliva/nasal mucosa marker) signals. Nasal mucosa is scored using signals for BPIFA1 (nasal marker) and STATH (shared marker) signals. In case of four replicates, a maximum of eight signals can be detected for each of these body fluids.

Number of markers detected			Interpretation	
HTN3 (saliva)	STATH (shared)	BPIFA1 (nasal)	Saliva	Nasal mucosa
0	4	0	Saliva or nasal mucosa "observed"	
0	1-3	0	Saliva or nasal "sporadically observed"	
0	1-4	1-4	"(spor.) observed and fits" nasal*	"(spor.) observed"**
1-4	1-4	0	"(spor.) observed"	"(spor.) observed and fits saliva"
1-4	0-4	1-4	"(spor.) observed"	"(spor.) observed"

* 'Observed' when STATH + BPIFA1 signals together occur for at least half of the possible signals, otherwise 'sporadically observed'.
** 'Observed' when STATH + HTN3 signals together occur for at least half of the possible signals, otherwise 'sporadically observed'.

Nasal mucosa scored "observed" in 16 samples, while for the remaining six samples nasal mucosa could not be distinguished from saliva, as only shared marker STATH responded in each replicate (Figure 2A). For 14 of the 16 samples for which nasal mucosa was scored "observed", saliva was scored "observed and fits", as STATH was detected in all replicates next to BPIFA1 signals. Although the presence of saliva cannot be excluded, we tend to regard saliva not present as such because of the absence of HTN3 signals. Using the updated multiplex (Cell-typer V3), variable expression of vaginal mucosa markers is seen in the nasal samples, alike in the experiments using the previous multiplex (Cell-typer V2, Figure 1). For most samples, vaginal mucosa scored "sporadically observed" (Figure 2A) except for one that scored "observed". This is to be regarded "observed and fits", as we know from this study and from [13] that variable expression of vaginal mucosa markers in nasal mucosa samples occurs. The updated multiplex (Cell-typer V3) that included the nasal mucosa marker (Table 1) was used in following experiments.

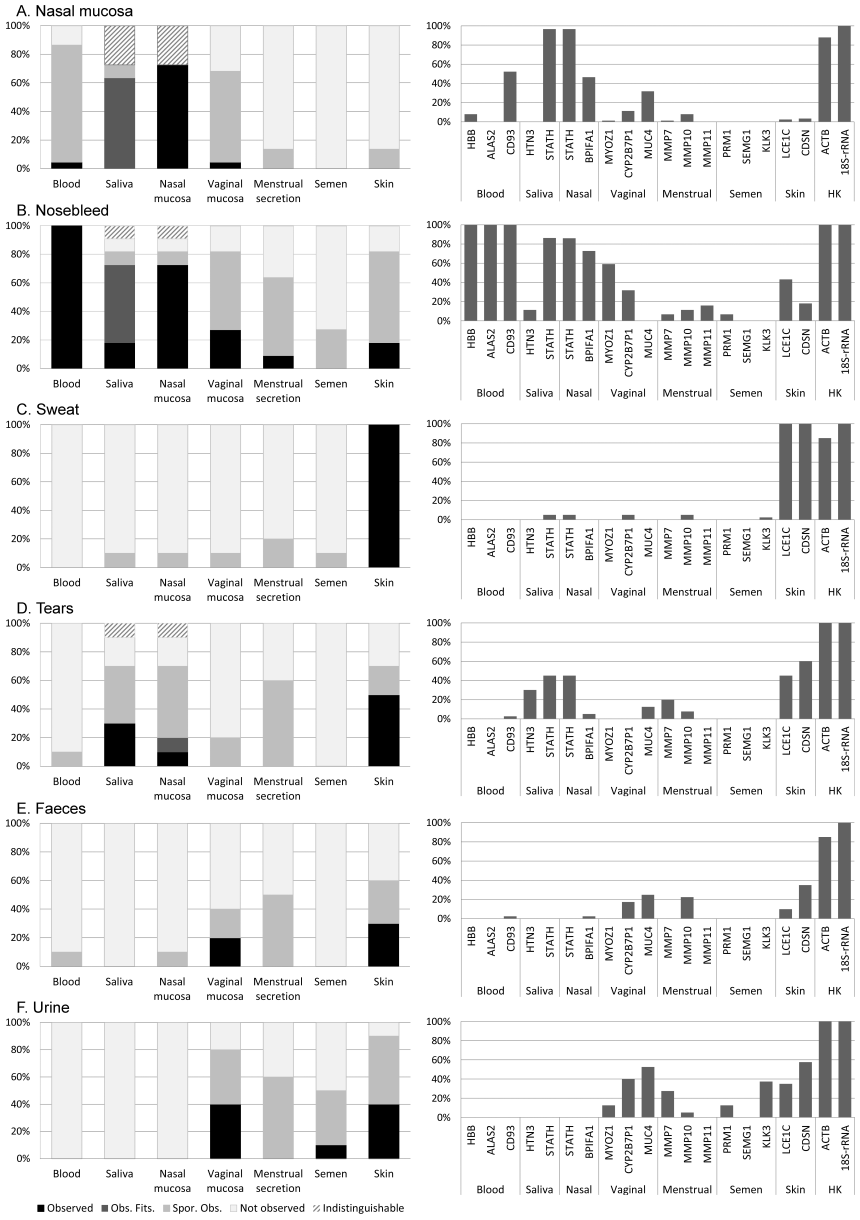


Figure 2. mRNA analysis of 22 nasal mucosa (A),11 nosebleed (B), and 10 sweat (C), tear (D), faeces (E) and urine (F) samples analysed with the updated Cell-typer (V3) multiplex Analysis was performed in 4-fold to allow profile interpretation based on the *x=n*/2 guideline (left). Saliva and nasal mucosa markers are marked "indistinguishable" when only shared marker STATH responded in each replicate, and no HTN3 (saliva) or BPIFA1 (nasal) signals were detected to distinguish the two cell types. Responsiveness of individual markers for each secretion are shown on the right. Housekeeping (HK) markers were scored "observed" in each sample.

mRNA analysis of nosebleed, sweat, tear, faeces and urine samples

As the nasal mucosa marker did not show cross-reactivity with peripheral blood samples, BPIFA1 could theoretically be used to discriminate peripheral from nosebleed blood. The distinction of nasal blood from peripheral blood has previously been described by Sakurada et al. [30], using the presence of STATH and the absence of HTN3, which we would now complement with nasal mucosa marker BPIFA1. A total of 11 nosebleed samples were analysed using the updated multiplex (Cell-typer V3) and blood was scored "observed" in all samples (Figure 2B). For eight of the samples, nasal mucosa was scored "observed" as well. For two of these eight samples, also saliva was scored "observed", while saliva was scored "observed and fits" for the other six samples. The vaginal mucosa and menstrual secretion markers responded to variable extent (menstrual secretion was scored "observed" once). This is probably due to the presence of nasal mucosa in the nosebleed samples as evident from the BPIFA1 response and therefore in accordance with the analysed nasal mucosa samples (Figure 1 and Figure 2A).

These vaginal mucosa and menstrual secretion signals are therefore regarded "(sporadically) observed and fits". The three samples that did not indicate the presence of nasal mucosa (once "sporadically observed", once "not observed", once nasal mucosa or saliva, Figure 2B) did not indicate the presence of any other cell type besides blood. The variability in the composition of nosebleed blood with more or less nasal mucosa is likely the cause of these results. Using BPIFA1, 73% of the samples could be identified as nosebleed samples. We could not include expirated blood samples but we expect similar RNA profiles as with nosebleed samples. Bloodstain pattern analysis may further assist in the discrimination of these sample types.

When analysing sweat samples (Figure 2C), skin was "observed" in each of the samples and only sporadic signals for the other body fluids were observed. These results are consistent with those described by Xu et al. and may be unavoidable as sweat is transported to the epidermal skin surface via sweat glands [13][31]. Studies describe the use of dermcidin (DCD) for the identification of sweat [13][31]. However, we reasoned that addition of a sweat marker is not imperative as this fluid has limited forensic relevance and no cross-reactivity with markers other than skin markers is observed. Tear samples (Figure 2D) resulted in positive scoring of saliva or nasal mucosa in 30% and 10% of the samples, respectively. Additionally, in one of the samples STATH was detected in each replicate and could thus not be distinguished from saliva. This pattern was not observed in any other analysed non-target body fluids (sweat, urine, faeces), which only resulted in sporadic marker signals for these cell types. Tear samples may however carry traces of nasal mucosa as tears are excreted through the nasolacrimal duct that also leads fluid to the nasal cavity [12]. These observations underlie our statement "saliva or nasal mucosa observed" when only STATH is observed in all replicates. Some cross-reactivity was observed for the vaginal mucosa markers

when analysing faeces samples (Figure 2E), which is probably due to the presence of mucous membranes in the gastrointestinal tract [32]. Vaginal mucosa markers, mainly MUC4, were also detected in urine (Figure 2F). As this gene is described to be one of the mucin genes predominantly expressed in human bladder [33] as well as urethra [32] this may explain its detection in samples from both female and male donors. Semen signals were occasionally detected in urine samples of male donors, which makes sense as both fluids pass through the urethra. Interestingly, MUC4 signals are not observed when analysing pure semen samples [1][5][7]. This may be explained from the activity of protease found in seminal fluid [32], in combination with the low extent to which MUC4 remains in the urethra. Lastly, sporadic menstrual secretion signals, mainly caused by MMP7, were observed in urine of males and females. This is consistent with results described in [13]. We assessed the possibility of using two known kidney markers, namely UMOD and FXYD2 [8], to act as urine-specific markers, as was done before for UMOD [13]. Signals were observed in a few urine samples and we inferred that the markers had limited sensitivity which did not improve with increased primer concentrations (up to 0.5 µM; optimized concentrations in [8] are 0.12 µM for UMOD and 0.06 µM for FXYD2). These markers are therefore regarded of too low sensitivity to identify urine, which is in contrast to findings described in [13].

Overall, cross-reactivity to a variable extent was observed for sweat, tears, faeces and urine samples. These samples are not expected to carry large numbers of human cells, and the occurrence of false positive signals seems unlikely when trace amounts of these secretions are present. The fact that the RNA extracts were ethanol-precipitated to generate informative RNA profiles for all these samples substantiates this. When we determined the average amount of human DNA (hDNA) per microgram secretion, we found that a vaginal swab contains approximately 12.400 or 22.500 times more human cell material than urine or faeces respectively (results not shown). While ethanol precipitation of the full RNA extract preceded the analysis of the urine and faeces samples, dilution of the cDNA was required to obtain informative RNA profiles for the vaginal samples. The addition of cell type specific markers for these fluids may therefore not be necessary.

Bacterial markers

The suitability of bacterial markers for the identification of vaginal mucosa was assessed using four *Lactobacillus* species. A multiplex (Lactoplex, Supplementary Table 1) was developed where 16S-rRNA transcripts for *L. jensenii*, *L. iners*, *L. crispatus* and *L. gasseri* were amplified alongside endogenous mRNA markers for vaginal mucosa, skin and housekeeping rRNA markers. As shown in Supplementary Table 1, *Lactobacillus* markers have relatively low primer concentrations compared to the mRNA markers. This is required to prevent over-amplification as the used *Lactobacillus* markers are ribosomal RNA markers, and ribosomal RNA is one of most predominant

types of RNA in cells. Specificity to vaginal mucosa was assessed on samples with vaginal mucosa (vaginal mucosa, menstrual secretion, four samples each) or without vaginal mucosa (skin from back, foot, hand, blood, saliva, semen, four samples each [1] and nasal mucosa, 22 samples). In addition, 20 penile swabs were analysed as penile swabs are regularly encountered in casework and preliminary studies had indicated the occurrence of *Lactobacillus* species on penile environments unrelated to direct vaginal contact [20]. Quantification and STR profiling of the penile swabs led to the exclusion of nine samples for RNA profiling. These samples either had DNA concentrations below the detection threshold of the quantification system (less than 0.5 pg/µL) or appeared to be mixtures due to which to presence of female vaginal cell material could not be excluded. The remaining 11 samples resulted in single-donor male DNA profiles. RNA analysis of these samples was performed in four-fold to allow profile interpretation using the $x=n/2$ guideline as previously described.

Regarding the ability to positively identify vaginal mucosa containing samples, the microbial markers functioned for 50% of the specimens, compared to 75% when using the mRNA markers (Figure 3). Next, the tendencies to give false positive signals were examined for the various skin specimens, body fluids and the penile swabs. *Lactobacillus* species were detected in 92% of the skin samples from the back, foot and hand, resulting in an "observed" scoring in 17% of the samples. None of the mRNA markers targeting vaginal mucosa responded in the skin samples. Sporadic signals were detected in blood and saliva for the mRNA markers, and in semen for both mRNA and *Lactobacillus* markers. With nasal mucosa samples the mRNA markers were detected in the majority of the samples, while the *Lactobacillus* markers did not respond. The absence of *Lactobacillus* in nasal mucosa has previously been demonstrated by Akutsu et al. [16], who also reported the cross-reactivity of *L. iners* in semen samples. Additionally, Haas et al. [18] reported the detection of *Lactobacillus* species in buccal and urine swabs.

The DNA profiles corresponding to the selected 11 penile swabs did not indicate any presence of female DNA (single source male profiles). Notwithstanding, based on the *Lactobacillus* species, vaginal material was scored "observed" in three of the 11 (27%) penile swabs. Only one of these samples also scored "observed" for vaginal mucosa based on mRNA markers. Since the sample appeared to have a single male donor origin, this sample plus the other ten penile samples were examined for the presence of secretions also produced by males that may cross react with the vaginal mRNA markers, such as nasal mucosa. Four PCR replicates of the Cell-typer V3 multiplex were generated using the same input as used for Lactoplex analysis. Results did not indicate the presence of nasal mucosa (data not shown). Overall, the majority of vaginal mucosa signals in all penile swabs were caused by mRNA marker MUC4 (46%), with an average peak height a 5-fold higher than the other vaginal mucosa markers. The detection of vaginal markers on penile swabs is probably caused by the presence of mucins, including MUC4, in foreskin secretion

[33-34]. Awareness of the possibility of non-specific product formation is important to prevent incorrect data interpretation, for example by disregarding MUC4 results from data interpretation when analysing penile swabs. Detection percentages and profile interpretation results of the two marker types are shown in Figure 3.

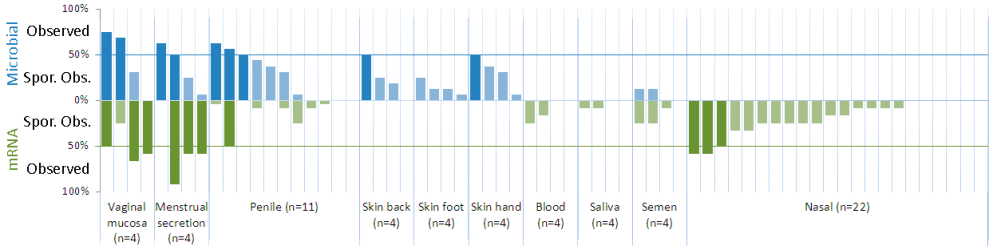


Figure 3. Marker detection percentages for vaginal mucosa containing, and vaginal mucosa lacking samples when analysed with microbial rRNA (blue) or endogenous mRNA (green) markers. Profile interpretation used the *x*=*n*/2 interpretation guideline and four RNA profiles. Detection percentages are determined based on a maximum of 16 markers for microbial markers (4 markers, 4 replicates), and 12 mRNA marker (3 markers, 4 replicates). Penile samplings were analysed both with the Lactoplex and the Cell-typer V3 multiplex, each containing 3 vaginal mucosa mRNA markers (thus the maximum number of detected markers is 24). Vaginal mucosa is scored "observed" if the marker detection percentage is ≥50% (solid bars), "sporadically observed" if >0% but <50% (transparent bars) and scored "not observed" if 0% of the markers are detected.

Differential mRNA analysis

Differential DNA/RNA co-extraction

Differential extraction (DE) allows for the separate analysis of DNA from sperm and non-sperm cells. In forensics this technique is most commonly applied to sexual assault samples that are expected to have a surplus of (vaginal) epithelial cells and few sperm cells (intimate swabs), or reversely few vaginal epithelial cells and a surplus of sperm cells. The basic concept of DE is to use a mild lysis for the extraction of nucleic acids of the epithelial cells, followed by pelleting the sperm cells, removing non-sperm cell DNA, and lysing the sperm cells using a more stringent protocol. Currently, DE protocols do not allow for the co-extraction of RNA. We investigated ways to incorporate DE into a co-extraction protocol to separate RNA and DNA of both the non-sperm fraction (NF) and sperm fraction (SF). Mixtures of saliva (female origin) and semen (40:1 ratio) were used to first assess the suitability of our standard co-extraction buffer to function as a mild lysis buffer [1]. This buffer, however, appeared too stringent as a portion of the spermatozoa was lysed during mild lysis resulting in PRM1 (spermatozoa marker) signals in both NF and SF (saliva and seminal fluid markers occurred mainly in the NF, data not shown). Next, the RSID semen buffer (Galantos Genetics GmbH), a PBS-like buffer, was tested as mild lysis buffer. Even though proteinase K was added that can break down RNases, this buffer did not enable the isolation of intact RNA in both

the NF and SF. DNA, however, was isolated. Finally mild lysis was achieved using PBS buffer (alike RSID semen buffer) with the addition of ribonucleoside-vanadyl complex, a known strong inhibitor of various ribonucleases [35-37]. On DNA level, the efficiency of this differential lysis buffer was assessed by determining the distribution of DNA belonging to the saliva and semen donor in the NF and the SF. Ratios were determined based on peak heights for non-shared alleles in STR profiles. Since the majority of DNA belonging to the saliva donor resided in the NF, and the majority of DNA belonging to the semen donor in the SF (Table 3) epithelial and sperm cells seemed separated. A small portion of the semen donors' DNA was however also detected in the NF, which may originate from low amounts of male epithelial cells or some lysed spermatozoa in the NF. In a previous study we describe the detection of alleles of a sterile seminal fluid donor, which probably originate from low amounts of epithelial cells or white blood cells present in semen [2]. The addition of the ribonucleoside-vanadyl complex assisted in separation of the RNA molecules of the different cells as only saliva and seminal fluid markers are detected in the NF, while spermatozoa markers are only detected in the SF (Table 3). The additional detection of saliva mRNA markers in the SF (Table 3) can be explained by the surplus of these cells in the sample, which affects separation efficiency. Overall, the use of a PBS buffer with the ribonucleoside-vanadyl complex is useable for the differential co-extraction of DNA and RNA over SF and NF.

Table 3. DNA and RNA results of a differentially co-extracted saliva-semen mixture using PBS buffer with Ribonucleoside-vanadyl complex as mild lysis buffer. Percentages of DNA belonging to the saliva and semen donor are determined based on peak height ratios in STR profiles (data not shown). RNA results (based on 3 PCR amplifications) separately show the detection percentages of semen markers for seminal fluid (KLK3 and SEMG1) and spermatozoa (PRM1), as spermatozoa markers are expected only in the SF.

	Total DNA		mRNA marker detection percentages		
	Saliva donor	Semen donor	Saliva	Seminal fluid	Spermatozoa
NF	90%	10%	100%	17%	0%
SF	4%	96%	50%	0%	100%

Differential analysis

Usually in our laboratory, the optimal cDNA input for RNA profiling is determined using a serial input. However, in samples with highly deviating cell type ratios, one cell type can show overloaded signals while others show signals around the detection level. Differential analysis, *i.e.* analysis using an adjusted multiplex from which overloaded markers are excluded, could simplify data interpretation. This approach could allow the identification of underlying cell types as it allows for using different cDNA inputs for the full multiplex (Cell-typer V3) and an adjusted multiplex. To test this approach, a mixed sample consisting of three cell types (semen, vaginal mucosa, skin) was analysed using a serial cDNA input. The mixture ratio based on peak heights of the

DNA profile is 1:0.7, RNA results are shown in Figure 4A-C. Whilst skin, semen and housekeeping are well detected with a low input (0.1 or 0.5 µL, Figure 4AB), vaginal mucosa markers are detected only in profiles with high input (2.5 µL, Figure 4C). Analysing overloaded profiles is unfavourable as bleed-through signals and trailing products may impede correct data interpretation. Whilst the presence of skin and semen is evident from amplifications using inputs as presented in Figure 4B, a higher cDNA input is appropriate to examine the presence of vaginal mucosa. Therefore, a multiplex is assembled from which all skin, all semen and the highest housekeeping marker (18S-rRNA) are excluded. Primer concentrations of the remaining primers and the amplification conditions are equal to those used for Cell-typer V3. Four PCR replicates were generated using this adjusted multiplex with a higher cDNA input (Figure 4D). Profile interpretation of these replicates resulted in an "observed" scoring for vaginal mucosa. No peaks were detected for saliva, blood, nasal mucosa and menstrual secretion. This illustrates the proof of concept that underlying cell types may be detected by using a multiplex from which markers for major cell types are excluded. This approach should only be applied when there are indications for presence of an underlying cell type as apparent from Figure 4C to prevent biased analysis.

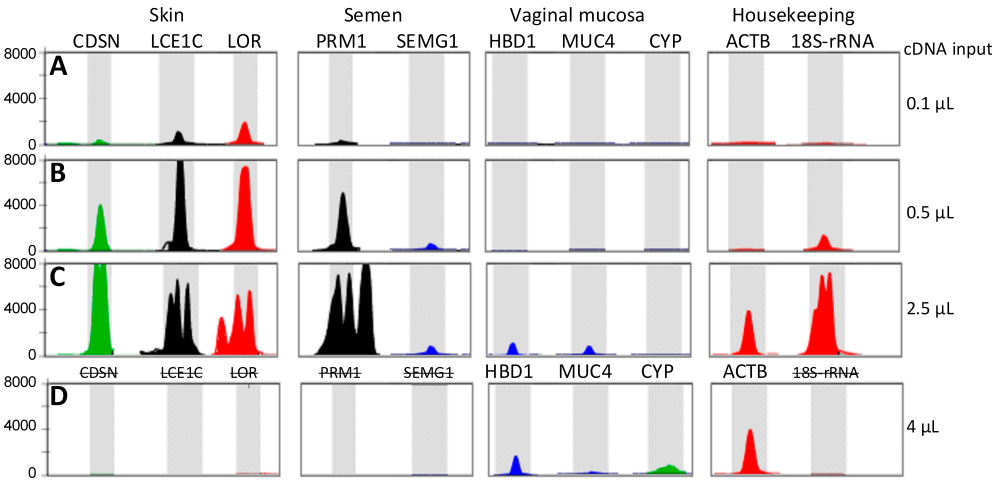


Figure 4. Differential analysis of an imbalanced vaginal mucosa/semen/skin mixture. A serial input (0.1, 0.5, 2.5 µL cDNA, A, B, C) was used to determine the optimal cDNA input. Vaginal mucosa signals are visible using a higher input (C) but this results in overloaded skin and semen signals. Replicate PCRs were performed using an adjusted multiplex from which overloaded markers are excluded (18S-rRNA, skin and semen marker: strikethrough) to allow the use of a higher input (4 µL cDNA, D). Menstrual secretion, blood, saliva and nasal mucosa markers are not shown, as no signals were observed.

DNA/RNA relation

Whilst the amount of genomic DNA is identical for diploid nucleated cells, the

amount of the individual mRNAs will vary in the various cell types. Consequently, there may be differences between the sensitivity of DNA and RNA profiling for different cell types. We examined DNA and RNA profiling (Cell-typer V3) sensitivity for single cell type samples using the same dilution steps for the DNA and cDNA samples. The percentage detected STR alleles and the percentage detected mRNA markers were determined and plotted as shown in Supplementary Figure 2 to study the relation between the sensitivity of DNA and RNA profiling. Results show that for some samples, like blood and semen, DNA profiling appears to be more sensitive than RNA profiling, as the percentage of detected alleles remains 100% whilst mRNA marker detection percentages decrease (Supplementary Figure 2AE). Skin, on the other hand, shows opposite results, as mRNA markers remain detectable whilst DNA allele detection percentages decrease (Supplementary Figure 2G). Especially nasal mucosa and vaginal mucosa data show a large variation (Supplementary Figure 2BF). This variation appears to be largely donor dependent, as for example vaginal mucosa donor A shows a decrease in percentage detected mRNA markers whilst, for DNA, always all alleles were observed. Donor D, in contrast, shows the opposite results (Supplementary Figure 2B). Additionally, we found that some mRNA markers, like CD93 and BPIFA1, are more prone to drop-out than others (Supplementary Figure 3). Overall, these results are consistent with the study analysing two component mixtures, where the straightforward association of DNA and RNA results based on peak heights is discouraged [11]. Results of this study give insight in the sensitivity of DNA and RNA profiling of single source samples, now including nasal mucosa.

Effect of the number of replicates when using the $x=n/2$ interpretation guideline

Our standardized RNA data interpretation is performed using four PCR replicates and scoring according to the $x=n/2$ guideline. We assessed the effect of using a higher number of PCR replicates for RNA data interpretation given this guideline. For three samples (previously used in the sensitivity study in section "DNA/RNA relation", semen donor A and C and menstrual secretion donor B, Supplementary Figure 2), additional profiles were generated to obtain three separate datasets for each of the three cDNA inputs with ideally four replicates per input. Thus, for each cDNA input of a sample up to 12 RNA profiling replicates were generated. The set descriptions and results are presented in Table 4.

When looking at the individual RNA profiles, variation in the percentage of detected markers becomes apparent, even for true technical replicates. For example, marker detection percentages within a set could range between 0% and 67% (*e.g.* Menstrual secretion donor B, 0.5 µL input, set 2, "individual" profiles, Table 4). This supports the use of a replicate-based system for data interpretation [29], as a cell type is scored "not

observed" when a replicate with 0% markers detected is regarded, and "observed" when a replicate with 67% markers detected is considered. Combining replicates provides a more reliable interpretation scoring, to our view, as it resembles the consensus approach for low template DNA profiling [38] by which variation in individual profiles is balanced by looking at the overall result. The sample set we chose for this experiment includes suboptimal (too low) cDNA inputs due to which the body fluid that is present was not scored "observed" in the original set of replicates (sets 1, Table 4, semen sample A, 0.2 and 0.5 µL input, and menstrual secretion sample B, 0.2 and 0.5 µL input). Preparing a different replicate set (either from the same cDNA batch *i.e.* sets 2, or from a fresh cDNA batch prepared from the same RNA sample *i.e.* sets 3, Table 4) or increasing the number of replicates to eight or 12 profiles did not invariably lead to an "observed" scoring for the body fluid as the variation for the individual profiles within a set of replicates underlies the scoring result (Table 4). Actually, when considering all available replicates (12 for the 0.2 and 0.5 µL inputs and nine for the 2.0 µL input) the same scoring result was obtained as for the original replicate set (set 1, Table 4). Notably, no false positive scorings for non-target cell types occurred in any of the sets.

Lastly we applied the $x=n/2$ rule per individual replicate and in Table 4 the number of profiles for which this criterion is met is indicated. This again corroborates the use of multiple replicates. For instance when regarding the results for the menstrual secretion sample with 0.2 µL input that seems a too low input to detect the menstrual secretion markers, for all replicate sets the score is "sporadically observed" and the overall percentage detected markers is 28%. Notwithstanding for two of the 12 replicates more than 50% of the markers are detected. Reversely, for semen sample A with 2 µL input, the overall percentage of detected markers is 67% and an "observed" score is obtained for all replicate sets. However, for only seven of the nine individual profiles 50% or more of the semen markers are detected. Furthermore, there is no strict relation between the number of individual profiles for which 50% or more of the markers are detected and the overall percentage of detected markers considering all available replicates: semen sample A 0.5 µL input and menstrual sample B 0.5 µL input both have 47% of the markers detected considering all 12 replicates while the number of profiles having ≥50% markers detected is four and seven respectively (Table 4).

Overall, results indicate the necessity of replicates for interpretation of RNA data. However, high numbers of replicates (over four) do not cancel out the impact of variation in individual profiles on data interpretation.

Table 4. mRNA marker detection percentages for up to 12 replicates using various cDNA inputs for semen and menstrual secretion samples. Three different replicate sets (each set representing true technical replicates) were compared for each cDNA input for each sample. Set 1 and 2 are prepared from the same cDNA batch, set 3 from a distinct cDNA batch. Within each set, the range of detection percentages for the individual profiles is given. When 4, 8 or 12 replicates are considered the overall percentage is given. Additionally, the number of profiles for which at least half of the markers is detected is shown. Grey colour coding indicates the cell type would be scored "observed" when applying the $x=n/2$ guidelines.

		Semen sample A						Semen sample C						Menstrual secretion sample B					
		Semen markers (3)			HK[a] markers (2)			Semen markers			HK markers			Menstrual markers (3)			HK markers		
# Replicates	Profiles	0.2 µL	0.5 µL	2.0 µL	0.2 µL	0.5 µL	2.0 µL	0.2 µL	0.5 µL	2.0 µL	0.2 µL	0.5 µL	2.0 µL	0.2 µL	0.5 µL	2.0 µL	0.2 µL	0.5 µL	2.0 µL
Individual	Set 1[b]	33%	33-67%	33-100%	50-100%	100%	100%	33-67%	67-100%	100%	50-100%	100%	100%	0-67%	0-67%	67-100%	100%	100%	100%
Individual	Set 2[c]	33-67%	33-100%	100%[e]	50-100%	50-100%	100%[e]	0-67%	67-100%	100%[e]	50-100%	100%	100%[e]	0-33%	0-67%	67%[e]	100%	100%	100%[e]
Individual	Set 3[d]	33-67%	33%	33-67%	50%	50%	50-100%	33-67%	33%	100%	50-100%	50-100%	100%	33-67%	33-67%	67-100%	100%	100%	100%
4	Set 1	33%	42%	67%	63%	100%	100%	50%	83%	100%	88%	100%	100%	25%	42%	92%	100%	100%	100%
4	Set 2	50%	67%	-	88%	88%	-	50%	83%	-	63%	100%	-	17%	42%	-	100%	100%	-
4	Set 3	42%	33%	58%	50%	50%	63%	58%	33%	100%	88%	75%	100%	42%	58%	92%	100%	100%	100%
8	Set 1 & 2	42%	54%	-	75%	94%	-	50%	83%	-	75%	100%	-	21%	42%	-	100%	100%	-
8	Set 1 & 3	38%	38%	63%	56%	75%	81%	54%	58%	100%	88%	88%	100%	33%	50%	92%	100%	100%	100%
8	Set 2 & 3	46%	50%	-	69%	69%	-	54%	58%	-	75%	88%	-	29%	50%	-	100%	100%	-
12	Set 1, 2 & 3	42%	47%	67%[f]	67%	79%	83%[f]	53%	67%	100%[f]	79%	92%	100%[f]	28%	47%	89%[f]	100%	100%	100%[f]
#profiles ≥50% markers detected		3/12	4/12	7/9	12/12	12/12	9/9	8/12	8/12	9/9	12/12	12/12	9/9	2/12	7/12	9/9	12/12	12/12	9/9

[a]HK, Housekeeping
[b]Set 1: Initial data (section "DNA/RNA sensitivity"), four replicates each input: 0.2, 0.5 and 2.0 µL cDNA.
[c]Set 2: Remnant of cDNA batch used for set 1, four replicates using 0.2 and 0.5 µL input, 1 replicate using 2.0 µL input.
[d]Set 3: New cDNA batch (same RNA as used for set 1 and 2), four replicates each input: 0.2, 0.5 and 2.0 µL cDNA.
[e]The 2 µL input of set 2 used 1 replicate.
[f]Since the 2 µL input of set 2 comprised of 1 replicate, 9 replicates are considered

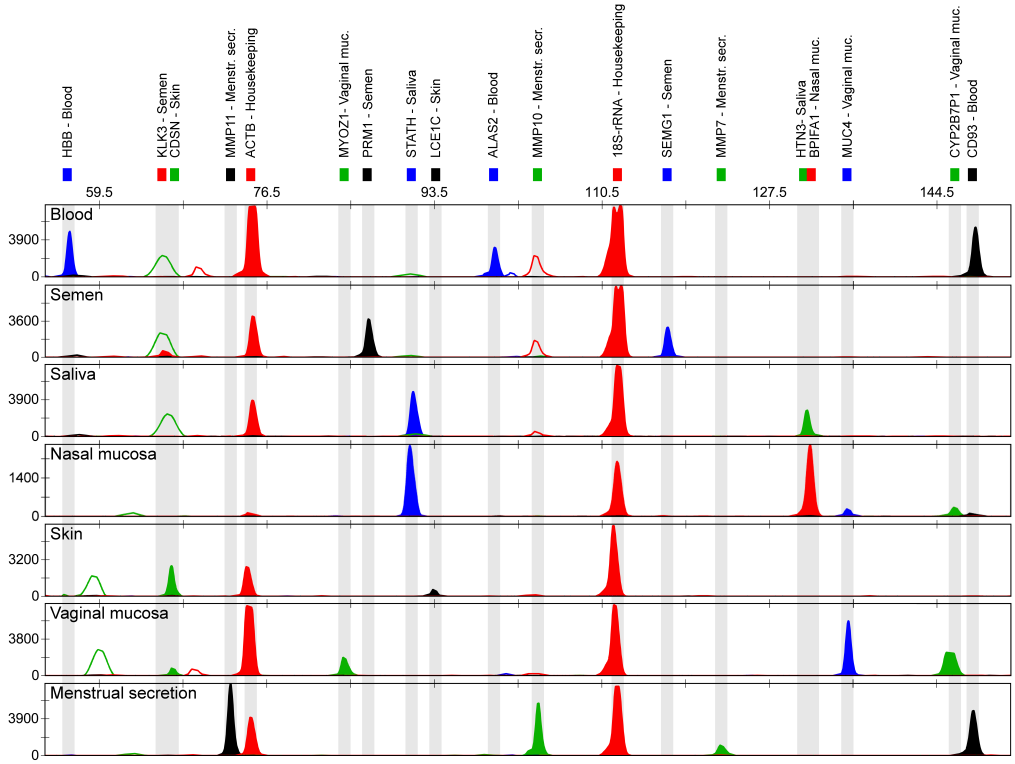
Concluding remarks

In this study, we assessed various ways of advancing mRNA-based body fluid typing. Firstly, the methodology is improved by adding a nasal mucosa marker and developing a differential extraction strategy that generates DNA and RNA fractions for both the SF and NF. Secondly, insight is increased by determining the possibility of obtaining non-specific signals for non-target cell types and studying the relation between the sensitivity of DNA and RNA results. We derive that false positive signals can be obtained when mRNA markers show cross-reactivity with body fluids or secretions not yet targeted by a multiplex. The addition of specific mRNA markers for these cross-reacting body fluids seems especially important when a body fluid or secretion carries a large number of human cells. In addition we noticed that for some sample types an adjusted interpretation approach is appropriate. For example, vaginal marker MUC4 is best not considered when penile swabs are analysed. Furthermore we discourage the association of DNA and RNA results as the relative sensitivity of the two profiling systems varies for different cell types, donors and markers. Lastly, the variation seen for individual RNA profiles, even when true technical replicates, substantiates the use multiple profiles for data interpretation, but there appears no real added value to increase the number over four replicates.

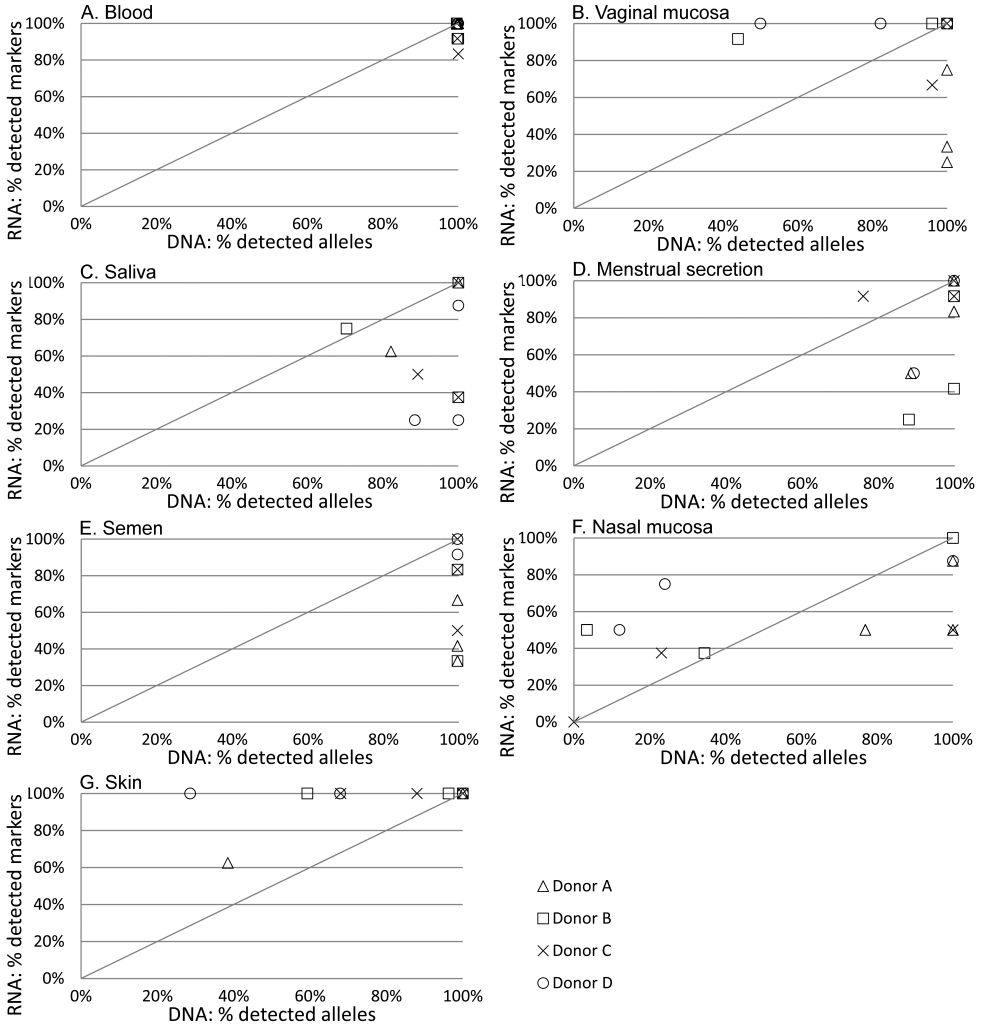
Acknowledgements

The authors are grateful to all volunteers who donated samples for this study. We thank Corina Benschop for critically reading the manuscript. TS and MvdB received financial support from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 285487 (EUROFORGEN-NoE).

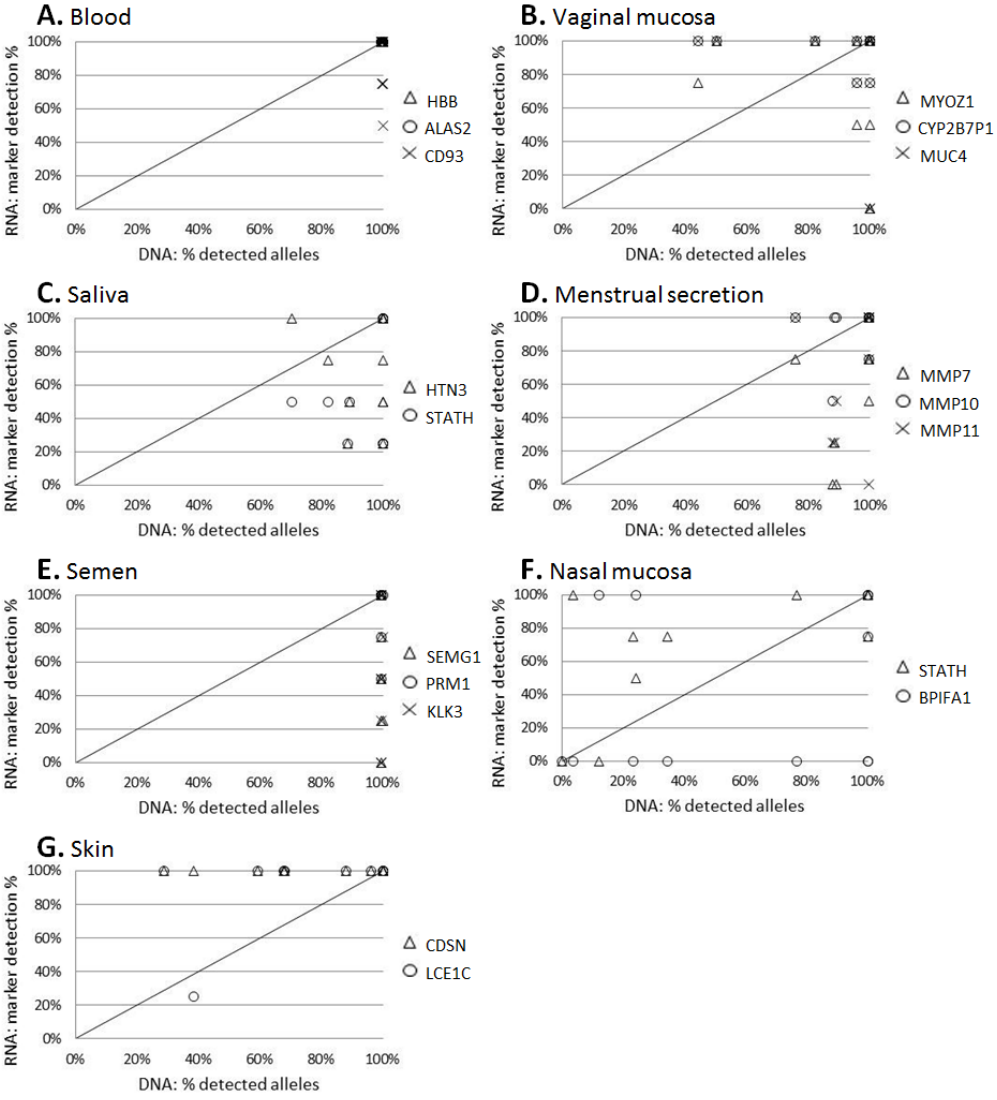
Supplementary material



Supplementary Figure 1. Examples of overlay electropherograms obtained when applying Cell-typer V3 on single source target body fluids.



Supplementary Figure 2. Percentage detected STR alleles and mRNA markers for different cell types using equivalent dilutions. mRNA marker detection percentages are determined for each cell type based solely on target mRNA marker detection of four PCR replicates. When DNA and RNA profiling would be equally sensitive, a relative indicated by the diagonal line would arise.



Supplementary Figure 3. Percentage detected STR alleles and mRNA markers for different cell types using equivalent dilutions. mRNA marker detection percentages are shown for the individual markers for each cell type based solely on target mRNA marker detection of four PCR replicates. When DNA and RNA profiling would be equally sensitive, a relative indicated by the diagonal line would arise.

Supplementary Table 1. Primer sequences for the different bacterial and mRNA markers residing in the Lactoplex.

Marker name	Tissue	[primer]	Forward primer (5'-3')	Size	Dye	Reference
		µM	Reverse primer (5'-3')	(bp)		
L. crispatus	Vaginal mucosa	0.1	GATTTACTTCGGTAATGACGTTAGGA	137	6FAM™	[26]
			AGCTGATCATGCGATCTGCTTTC			
L. gasseri	Vaginal mucosa	0.1	AGCGAGCTTGCCTAGATGAATTTG	171	6FAM™	[27]
			TCTTTTAAACTCTAGACATGCGTC			
L. jensenii	Vaginal mucosa	0.02	ACCTGCCCTTAAGTCTGGGA [a]	91	6FAM™	[27]
			ACGCCGCCTTTTAAACTTCTT			
L. iners	Vaginal mucosa	0.02	TTGAAGATCGGAGTGCTTGC	97	6FAM™	[26]
			TTATCCCGATCTCTTGGGCA			
HBD1	Vaginal mucosa	0.8	GAAATCCTGGGTGTTGCC	101	FAM™	[1]
			AAAGTTACCACCTGAGGCC			
LOR	Skin	0.6	CTTTGGGCTCTCCTTCCT	89	PET™	[1]
			AGAGGTCTTCACGCAGTC			
MUC4	Vaginal mucosa					
CYP2B7P1	Vaginal mucosa					
CDSN	Skin		As described in Table 1,			
LCE1C	Skin		except for primer concentration CYP2B7P1: 1.6 µM			
ACTB	Housekeeping					
18S-rRNA	Housekeeping					

[a] Developed for this study using Ensembl and NCBI primer blast [24][25]

Supplementary Table 2. Primer sequences for the different tested candidate nasal mucosa markers.

Marker name[a]	Forward primer (5'-3')	Size
	Reverse primer (5'-3')	(bp)
BPIFA1	CAAGTGAATACGCCCCTGGTCG	131
	GAATGGGTGCAGTCACCAAGGAC	
BPIFB1	CCGCTGCTCAGTGCCATGC	103
	TGATGACCTTCAGCCAGATGATGTGC	
SCGB1A1	AACCAGAGACGGGCCAGAGCAT	123
	ACGCTGAAAGCTCGGGCAGATC	
C6orf58	GGACAGGCAGATTAGCTGATCCAAC	145
	CCAGAATCAACCGCAGCAAGAAAGG	
PPP1R9B	GATGACGAGGAGACGGGAGAG	120
	GGACAGTGCATCCTCGTTCTCC	

[a] All primer sequences are developed for this study using Ensembl and NCBI primer blast [24][25]

Supplementary Table 3. Specificity results for nasal mucosa marker BPIFA1. Additionally shown are results for saliva marker HTN3 and shared saliva/ nasal mucosa marker STATH, as these markers are used for the interpretation of saliva/nasal mucosa samples. Marker detection percentages are determined based on four replicate PCRs.

	HTN3	STATH	BPIFA1
Blood (*n*=4)	6%	-	-
Saliva (*n*=4)	94%	100%	-
Nasal mucosa (*n*=4)	-	88%	75%
Vaginal mucosa (*n*=4)	-	-	-
Menstrual secretion (*n*=4)	-	-	-
Semen sterile (*n*=4)	-	-	-
Semen fertile (*n*=2)	-	-	-
Skin (*n*=4)	-	-	-

References

1. A. Lindenbergh, M. de Pagter, G. Ramdayal, M. Visser, D. Zubakov, M. Kayser, T. Sijen, A multiplex (m) RNA-profiling system for the forensic identification of body fluids and contact traces, Forensic Sci. Int.: Genet. 6 (2012) 565–577.
2. M. van den Berge, A. Carracedo, I. Gomes, E.A.M. Graham, C. Haas, B. Hjort, P. Hoff-Olsen, O. Maronas, B. Mevag, N. Morling, H. Niederstätter, W. Parson, P.M. Schneider, D. Syndercombe Court, A. Vidaki and T. Sijen. A collaborative European exercise on mRNA-based body fluid/skin typing and interpretation of DNA and RNA results. Forensic Sci. Int.: Genet. 10 (2014) 40-48.
3. R.I. Fleming, S. Harbison, The use of bacteria for the identification of vaginal secretions, Forensic Sci. Int.: Genet. 4 (2010) 311–315.
4. I. Gomes, F. Kohlmeier, P.M. Schneider, Genetic markers for body fluid and tissue identification in forensics, Forensic Sci. Int.: Genet. Suppl. Ser. 3 (2011) e469– e470.
5. C. Haas, B. Klesser, 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.
6. E.K. Hanson, J. Ballantyne, Highly specific mRNA biomarkers for the identification of vaginal secretions in sexual assault investigations, Sci. Justice 53 (2013) 14–22.
7. J. Juusola, J. Ballantyne, Multiplex mRNA profiling for the identification of body fluids, Forensic Sci. Int. 152 (2005) 1-12.
8. A. Lindenbergh, M. van den Berge, R.J. Oostra, C. Cleypool, A. Bruggink, A. Kloosterman, T. Sijen, Development of a mRNA profiling multiplex for the inference of organ tissues, Int. J. Legal Med. 127 (2013) 891–900.
9. A.D. Roeder, C. Haas, mRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification, Int. J. Legal Med. 127 (2013) 707–721.
10. T. Sijen, Molecular approaches for forensic cell type identification: On mRNA, miRNA, DNA methylation and microbial markers, Forensic Sci. Int.: Genet. (2015).
11. J. Harteveld, A. Lindenbergh, T. Sijen, RNA cell typing and DNA profiling of mixed samples: can cell types and donors be associated, Sci. Justice 53 (2013) 261–269.
12. K. Sakurada, T. Akutsu, K. Watanabe, Y. Fujinami, M. Yoshino, Expression of statherin mRNA and protein in nasal and vaginal secretions, Legal Med. 13 (2011) 309–313.
13. Y. Xu, J. Xie, Y. Cao, H. Zhou, Y. Ping, L. Chen, L. Gu, W. Hu, G. Bi, J. Ge, Development of highly

sensitive and specific mRNA multiplex system (XCYR1) for forensic human body fluids and tissues identification, PLoS One 9 (2014) e100123.
14. C. Cossu, U. Germann, A. Kratzer, W. Baer, C. Haas, How specific are the vaginal secretion mRNA-markers HBD1 and MUC4, Forensic Sci. Int.: Genet. Suppl. Ser. 2 (2009) 536–537.
15. R.I. Fleming, S. Harbison, The use of bacteria for the identification of vaginal secretions, Forensic Sci. Int.: Genet. 4 (2010) 311–315.
16. T. Akutsu, H. Motani, K. Watanabe, H. Iwase, K. Sakurada, Detection of bacterial 16S ribosomal RNA genes for forensic identification of vaginal fluid, Legal Med. 14 (2012) 160-162.
17. S. Giampaoli, A. Berti, F. Valeriani, G. Gianfranceschi, A. Piccolella, L. Buggiotti, C. Rapone, A. Valentini, L. Ripani, V. Romano Spica, Molecular identification of vaginal fluid by microbial signature, Forensic Sci. Int.: Genet. 6 (2012) 559–564.
18. C. Haas, E. Hanson, M.J. Anjos, K.N. Ballantyne, R. Banemann, B. Bhoelai, E. Borges, M. Carvalho, C. Courts and G. De Cock, RNA/DNA co-analysis from human menstrual blood and vaginal secretion stains: Results of a fourth and fifth collaborative EDNAP exercise. Forensic Sci. Int.: Genet. 8 (2014) 203-212.
19. B. Ma, L.J. Forney, J. Ravel, The vaginal microbiome: rethinking health and diseases, Annu. Rev. Microbiol. 66 (2012) 371.
20. C.C. Benschop, F.C. Quaak, M.E. Boon, T. Sijen, I. Kuiper, Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context, Int. J. Legal Med. 126 (2012) 303–310.
21. C.C. Benschop, D.C. Wiebosch, A.D. Kloosterman, T. Sijen, Post-coital vaginal sampling with nylon flocked swabs improves DNA typing, Forensic Sci. Int.: Genet. 4 (2010) 115-121.
22. K.M. Stankovic, H. Goldsztein, D.D. Reh, M.P. Platt, R. Metson, Gene Expression Profiling of Nasal Polyps Associated With Chronic Sinusitis and Aspirin-Sensitive Asthma, The Laryngoscope 118 (2008) 881-889.
23. S.H. Song, H.U. Jang, J.W. Oh, J.S. Kim, Gene Expression Analysis in Nasal Polyp Using Microarray, Otorhinolaryngology-Head Neck Surg. 54 (2011) 55-61.
24. L. Bingle, S.S. Cross, A.S. High, W.A. Wallace, D.A. Devine, S. Havard, M.A. Campos, C.D. Bingle, SPLUNC1 (PLUNC) is expressed in glandular tissues of the respiratory tract and in lung tumours with a glandular phenotype, J. Path. 205 (2005) 491-497.
25. P. Flicek, et al., Ensembl 2014, Nucleic Acids Res. 42 (2014) D749-D755.
26. J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T.L. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction, BMC bioinformatics 13 (2012) 134.
27. M. Zozaya-Hinchliffe, R. Lillis, D.H. Martin, M.J. Ferris, Quantitative PCR assessments of bacterial species in women with and without bacterial vaginosis, J. Clin. Microbiol. 48 (2010) 1812–1819.
28. D.H. Yan, Z. Lü, J.R. Su, Comparison of main lactobacillus species between healthy women and women with bacterial vaginosis, Chin. Med. J. 122 (2009) 2748-2751.
29. A. Lindenbergh, P. Maaskant, T. Sijen, Implementation of RNA profiling in forensic casework, Forensic Sci. Int.: Genet. 7 (2013) 159–166.
30. K. Sakurada, T. Akutsu, K. Watanabe and M. Yoshino, Identification of nasal blood by real-time RT-PCR. Legal Medicine 14 (2012) 201-204.
31. K. Sakurada, T. Akutsu, H. Fukushima, K. Watanabe and M. Yoshino, Detection of dermcidin for sweat identification by real-time RT-PCR and ELISA. Forensic Science International 194 (2010) 80-84.
32. J. N'Dow, J. Pearson, D. Neal, Mucin gene expression in human urothelium and in intestinal segments transposed into the urinary tract, J. Urol. 164 (2000) 1398-1404.
33. C.L. Russo, S. Spurr-Michaud, A. Tisdale, J. Pudney, D. Anderson, I.K. Gipson, Mucin gene expression in human male urogenital tract epithelia, Hum. Reprod. 21 (2006) 2783-2793.

34. G. Hadžić, A. Lukan, K. Drobnič, Practical value of the marker MUC4 for identification of vaginal secretion in penile swabs, Forensic Sci. Int.: Genet. Suppl. Ser. 3 (2011): e222-e223.
35. S.L. Berger, C.S. Birkenmeier, Inhibition of intractable, nucleases with ribonucleoside-vanadyl complexes: isolation of messenger ribonucleic acid from resting lymphocytes, Biochemistry 18 (1979) 5143-5149.
36. J.C. Gray, The inhibition of ribonuclease activity and the isolation of polysomes from leaves of the French bean, Phaseolus vulgaris. Arch. Biochem. Biophys. 163 (1974) 343-348.
37. E. Egberts, P.B. Hackett, P. Traub, Hoppe-Zeyler's Z, Physiol. Chem. 358 (1971): 475-490
38. C.C. Benschop, C.P. van der Beek, HC, Meiland, A.G. van Gorp, A.A. Westen, T. Sijen, Low template STR typing: effect of replicate number and consensus method on genotyping reliability and DNA database search results, Forensic Sci Int Genet 5 (2011) 316-328.