Advancing forensic RNA profiling

Berge, M.W. van den

Advancing forensic RNA profiling

Margaretha Wilhelmina van den Berge

Colophon

Advancing forensic RNA profiling

Author:	Margreet van den Berge
ISBN/EAN:	978-90-827490-3-8
Printing:	Xerox (Nederland) B.V.
Cover design & Layout	Margreet van den Berge

Advancing forensic RNA profiling

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 28 juni 2018
klokke 12:30 uur

door

Margaretha Wilhelmina van den Berge

geboren te Willemstad

in 1990

Promotor	Prof. dr. P. de Knijff
Copromotor	Dr. ir. L.M.T. Sijen - *Nederlands Forensisch Instituut*
Leden promotiecommissie	Prof. dr. M. Tijsterman Prof. dr. M. Kayser - *Erasmus MC University Medical Center Rotterdam* Dr. S.A. Harbison - *Institute of Environmental Science and Research Limited (ESR), Auckland, New Zealand*

The research described in this thesis has been performed within the Research group of the Biological Traces division at the Netherlands Forensic Institute.
This work was financially supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 285487 (EUROFORGEN-NoE).
Publication of this thesis was financially supported by the Netherlands Forensic Institute.

Tell me and I forget. Teach me and I remember. Involve me and I learn.

Benjamin Franklin, 1706 – 1790

Table of contents

Outline

of this thesis

In forensic criminal investigations, DNA profiling is an important and frequently used tool, as it has the ability to reveal the identity of a donor of a trace with high evidentiary value. In forensic settings, this technique was first applied in 1985 and has undergone major technological advancements since then, now allowing for the analysis of traces in which only minute amounts of often degraded cellular material are present. Additionally, the predictive value of DNA regarding donor ethnicity, visible features and age is increasingly explored for the forensic context.

Next to the question of *who* contributed to the trace (DNA analysis), knowledge regarding the cellular origin of the evidentiary trace, thus *what* cell type contributed to the trace, is often key to facilitate inference of activities. Previously, cell type inference of body fluid and organ tissue traces commonly relied on techniques such as microscopic, histological or immunological tests, which due their limited sensitivity and specificity are not ideal in forensic settings. Alternative methods for cell type inference include RNA-based approaches which have shown their up come in forensic investigations since 1999. RNA profiling techniques rely on the fact that different cell types, such as blood, express a characteristic pattern of genes, such as haemoglobin beta (HBB) expressed in the red blood cells. RNA profiling is generally performed simultaneously and alongside DNA profiling from the same trace of evidence.

The chapters of this thesis describe the work performed through the years aiming to expand and advance forensic RNA profiling. Details per chapter are outlined below. The ultimate goal of forensic research is application in casework. Regarding RNA typing, the Netherlands Forensic Institute (NFI) is one of few labs in the world actually applying RNA profiling in forensic casework. Since the first application in an NFI case in 2010 and to the time of writing (December 2017), RNA profiling has been considered in over 200 cases. In approximately 70% of the RNA cases, identification of body fluids is requested. This body fluid system, referred to as the "Cell-typer", allows for the inference of blood, saliva, vaginal mucosa, menstrual secretion, semen and skin. Three years later, an organ tissue profiling system, allowing for the inference of brain, lung, liver, skeletal muscle, heart, kidney and skin, was introduced to RNA casework at the NFI and is requested in 30% of the RNA cases. Body fluid inferring RNA cases mainly involve sexual assault cases in which the presence of vaginal mucosa cells is disputed. Inference of organ tissues is mainly requested on objects involved in violent crimes. Up to date, ten NFI RNA cases have been presented in national court, while presented twice internationally.

As RNA profiling is in many ways very distinct from DNA profiling, expertise in analysing RNA profiles, clear guidelines for data interpretation and awareness regarding potential interpretation pitfalls are essential when RNA profiling is applied in casework. When proceeding to application in casework, NFI developed such accompanying interpretation guidelines. **Chapter 1** describes an RNA profiling exercise conducted as a collaboration between nine partners of the European Forensic Genetics Network of

Excellence (EUROFORGEN-NoE) in order to assess both the value of RNA analysis for cell type inference and the accompanying interpretation guidelines in forensic context.

As bodily secretions may be encountered at a crime scene whilst not targeted by the body fluid multiplex, six additional secretion types were analysed to investigate the possibility of obtaining false positive signals in these non-target cell types. **Chapter 2** describes the results of this specificity study and additionally describes studies performed aiming to further forensic RNA profiling, such as the use of alternative markers to replace RNA markers for the identification of vaginal mucosa, or the search for a relation in sensitivity of DNA versus RNA profiling in single donor body fluids.

Alike for DNA profiling, low level noise signals and dye blobs may be detected in RNA profiles. Increased noise can difficult interpretation of amplified signals and lead to false positive scoring of absent markers. **Chapter 3** describes the various causes and solutions that were considered to reduce the level of noise signals in RNA profiles.

Especially when minute evidentiary traces are analysed, background cell material unrelated to the crime may contribute to detectable levels in the genetic analyses. **Chapter 4** aims to increase our understanding regarding the prevalence of human cell material in background and activity scenarios. This chapter describes the results of analysing 549 samples comprising various public objects, private samples, transfer-related samples and laundered samples, which were analysed using DNA and RNA profiling. Several research questions were proposed for this study, for example "Do increased DNA yields lead to increased numbers of contributors or to the detections of other cell types than skin?" and "Is the owner of a private item always the major contributor to the DNA profile?".

Unlike in the study described in Chapter 4, forensic casework scenarios do not allow for the analysis of pre-activity background samples from the exact same location as post-activity samples. **Chapter 5**, a sequel of Chapter 4, describes therefore the analysis of post-activity background samples taken from an untouched area. Chapter 5 presents how these control region specimens may be useful when investigating activity-related scenarios.

Chapters 1 to 5 mainly focus on the inference of body fluids, while in some forensic cases, such as the event of a violent crime, knowledge regarding the tissue type may be of aid in the reconstruction of the event surrounding the crime. In **Chapter 6** the development of an organ tissue inferring RNA-based multiplex is described. This multiplex, referred to as the "Organtyper", allows for the inference of seven previously mentioned organ tissue types. After considering multiple candidate RNA markers, a selection of 17 markers was combined in a multiplex comprising at least two distinct markers per target organ tissue.

Primate- and target-specificity of the organ tissue typing system described in Chapter 6 is essential when applied in forensic casework. Initially, primate-specificity

of the system was confirmed based on *in silico* analysis using sequence alignment software. In **Chapter 7** specificity of the assay to human tissues was physically assessed by subjecting the assay to a set of non-human organ tissues. Additionally, cross-reactivity of the Organtyper markers to body fluids and of body fluid markers to organ tissues was assessed.

When postmortem intervals increase such as with longer burial times, human remains suffer increasingly from the taphonomic effects of decomposition processes such as autolysis and putrefaction. It is often assumed that RNA is less stable than DNA and may therefore be unsuitable for analysis of degraded samples. In **Chapter 8** the stability of DNA and RNA in long-buried human remains was assessed to examine for trends in nucleic acid degradation and the postmortem interval. This study gives insight in the remarkable stability of nucleic acids in severely degraded tissues and may lead to a change in sampling policies in identification of degrading cadavers.

In the last experimental chapter, a novel application of RNA typing was explored, namely for the inference of sex. While the presence of male cell material can be readily inferred from Y-chromosome specific signals in DNA quantitation and DNA profiling results, the presence of female cell material is inferred only indirectly, *i.e.* from absence or unbalanced response of the Y-chromosomal marker. On DNA level, no forensic marker exists to positively identify female cell material. **Chapter 9** describes the search for male- and female-specific RNA markers. This is the first assay enabling positive identification of female cellular material, and the first overlapping information in DNA and RNA profiles.

Chapter 10 reflects on the outcomes described in the earlier chapters of this thesis and appoints aspects of future and alternative cell type inferring techniques used in forensic casework. These include the future developments that lie in the application of massively parallel sequencing (MPS) both to increase understanding and as an alternative cell typing method.

Chapter 1

A collaborative European exercise on mRNA-based body fluid/skin typing and interpretation of DNA and RNA results

Forensic Science International: Genetics 10 (2014) 40-48

M. van den Berge
A. Carracedo
I. Gomes
E.A.M. Graham
C. Haas
B. Hjort
P. Hoff-Olsen
O. Maroñas
B. Mevåg
N. Morling
H. Niederstätter
W. Parson
P.M. Schneider
D. Syndercombe Court
A. Vidaki
T. Sijen

Abstract

The European Forensic Genetics Network of Excellence (EUROFORGEN-NoE) undertook a collaborative project on mRNA-based body fluid/skin typing and the interpretation of the resulting RNA and DNA data. Although both body fluids and skin are composed of a variety of cell types with different functions and gene expression profiles, we refer to the procedure as 'cell type inference'. Nine laboratories participated in the project and used a 20-marker multiplex to analyse samples that were centrally prepared and thoroughly tested prior to shipment. Specimens of increasing complexity were assessed that ranged from reference PCR products, cDNAs of indicated or unnamed cell type source(s), to challenging mock casework stains. From this specimen set, information on the overall sensitivity and specificity of the various markers was obtained. In addition, the reliability of a scoring system for inference of cell types was assessed. This scoring system builds on replicate RNA analyses and the ratio observed/possible peaks for each cell type [1]. The results of the exercise support the usefulness of this scoring system. When interpreting the data obtained from the analysis of the mock casework stains, the participating laboratories were asked to integrate the DNA and RNA results and associate donor and cell type where possible. A large variation for the integrated interpretations of the DNA and RNA data was obtained including correct interpretations. We infer that with expertise in analysing RNA profiles, clear guidelines for data interpretation and awareness regarding potential pitfalls in associating donors and cell types, mRNA-based cell type inference can be implemented for forensic casework.

Introduction

The potential of mRNA profiling to infer which body fluid or tissue resides in an evidentiary sample has been well demonstrated in the past decade. The identification of specific, sensitive and robust markers, poses a continuous search that is undertaken for a growing number of body fluids and organs [2–23]. Evaluation of the performance of these markers is well assisted by large collaborative exercises that have been performed for several body fluids [24–27]. Suitable markers have been combined in end-point reverse transcription PCR (RT-PCR) systems that assess the presence of multiple cell types simultaneously [28–31]. Most of these assays carry multiple markers per cell type, as expression of individual mRNAs varies with biological function and for individuals or physiological condition. As a result, RNA profiles appear much less balanced than DNA profiles and sometimes marker dropout occurs. Marker drop-in may arise from the presence of non-specific transcripts in a cell due to back ground level gene expression or as a result of spurious transcription that occurs whenever RNA polymerase binds to DNA. Variation in RNA profiling data may be further stimulated by the use of a rather high number of amplification cycles (33 in [30], 35 in [24–27], 30 in [29], 30 or 35 depending on the body fluid in [31]), and stochastic amplification effects [33] are seen when replicate RNA analyses are performed [32]. These issues have triggered the development of interpretation strategies for RNA profiles that allow for some marker absence and spurious signals [1,31]. The scoring methodologies have limitations with body fluids that express markers of other body fluids as well and cannot discriminate co-expressed cell types (*e.g.* blood co-expressed in menstrual secretion) from a mixture of the two cell types (*e.g.* peripheral blood and menstrual secretion). Mixed stains are challenging to interpret anyhow as one donor may give multiple celltypes or multiple donors the same cell type. Combined interpretation of RNA and DNA profiling results may be only possible when gender-specific body fluids and donors of different genders are involved, as peak height-based association of donor and cell type was found to be risky even in straightforward two donor–two cell type mixtures [32]. A collaborative exercise was organised among the partners of the European Forensic Genetics Network of Excellence (EUROFORGEN-NoE) in order to assess the value of mRNA analysis in a forensic context. An RNA analysis system was taken that is routinely used for casework at the Netherlands Forensic Institute (NFI), and each participant was provided with the same set of specimens that had increasing complexity. The most complex samples were mock casework stains for which laboratories not only scored the RNA results but also provided a forensic interpretation, building on the estimated number and genders of contributors, the cell types regarded present and, if possible, association of donor and cell types. As these stains were designed to be challenging and complex, potential pitfalls for integrated DNA and RNA data interpretation become apparent.

Materials and methods

Samples and materials provided

The mRNA profiling exercise was divided into two parts: part 1 included purified PCR products and cDNA specimens and part 2 comprised mock casework stains. Details are provided in Table 1.

Table 1. Overview of the samples and objectives assessed during the mRNA profiling exercise.

Part	Samples	Tasks and objectives
1	Purified[a] PCR products, cell type indicated 7 samples: each cell type[b] and blank	Adjust provided bin sets CE sensitivity differences for laboratories
	Single source cDNAs, cell type indicated 24 samples: 3 inputs[c] and minus RT for each cell type	One PCR per sample with 1 µL cDNA input Overall marker sensitivity and frequency of marker drop-in
	cDNAs with unspecified cell type 10 samples: communicated as single source, but including one mixture and one water sample	Two PCRs per sample: 0.5 µL and 2.0 µL input Familiarisation serial cDNA input approach Overall performance: marker drop-ins and marker dropouts
	Mixed cDNAs, unspecified cell types 4 samples: mixtures of two or three body fluids	Determine optimal cDNA input from 0.5 µL and 2.0 µL tests Generate four informative replicates for scoring RNA results Overall usefulness of scoring system
2A	Complex stains, unspecified cell types 4 samples: on a variety of substrates	Extraction, DNA/RNA profiling and questionnaire DNA: estimated minimum number of contributors and genders RNA: scoring peaks and cell types Interpretation: cell types present and association donor/cell type
2B	NFI dataset for the four complex stains 1 DNA profile and 4 replicate RNA profiles	Compare interpretations for same dataset

[a] Purification by MinElute columns using a low salt strength buffer to elute products. Prior to shipment it was tested that the fluorescently labelled PCR products are stable in this buffer, which is not the case if purified into water.
[b] Six cell types are included: blood, saliva, semen, vaginal mucosa, menstrual secretion and skin.
[c] cDNAs were derived from three amounts of RNA.

Specimens were prepared at the organising laboratory (NFI) with informed consent of the voluntary donors whose cell material was used. Saliva and semen were collected in tube, blood through a finger prick (Accu-chek, Softclix Pro, Roche Diagnostics GmbH, Germany), vaginal mucosa on cotton swab, menstrual secretion on Viba brushes (Rovers, Oss, the Netherlands) and skin by rubbing textiles over the face. Textiles used as substrates for stains were freed from contaminating DNA by irradiating each side with 254 nm UV light in a CL-1000 UV CrossLinker (UVP, Upland, USA) at 900 mJ/cm^2 for 30 min. For stain 1, equal amounts of saliva of two donors were mixed and 40 mL were added to cotton swabs. For stain 2, two Viba brushes with menstrual secretion (day 2 menstrual cycle) were gently shaken in 500 mL PBS (phosphate buffered saline) for 0.5 h after which 200 mL whole blood were added and 5 mL of the mixture were transferred to pieces of fleece. Since the menstrual secretion signals were surpassing those of blood, 50 mL neat blood were added on top of the stain. For stain 3, skin donor 1 rubbed one side and skin donor 2 the other side of a patch of linen over the face. Then, spots of 1 mL 100-fold diluted blood of one of these donors were placed on the linen and small areas of cloth

around these blood spots were excised. For stain 4, nail clippings of relatively short nails having contact with skin were collected and a fresh vaginal mucosa swab was rubbed over the nail clippings to transfer cell material. After drying, 1 mL seminal fluid of an azoospermic male was added. All specimens and reagents sent to participating laboratories were thoroughly tested prior to shipment. When shipping part 1 reagents, an amount of multiplex primer mix sufficient for both exercise parts was included so that all results were obtained using a single batch of primer mix. Part 1 reagents and specimens were sent on dry ice taking 1–7 days, when in part 2 only dried stains were sent normal mail was used. When additional reagents such as extraction chemistry were requested dry ice was used. These shipments took between 3 and 9 days.

DNA/RNA extraction, DNA quantification, reverse transcription

At the organising laboratory, samples were subjected to DNA/RNA co-extraction, quantification of DNA extracts, DNase treatment of RNA extracts and reverse transcription (both plus RT reaction and minus RT control) of 10 µL RNA aliquots as described in [30]. In each extraction, a negative and positive control (water or 5 µL blood spotted on FTA) was included. Negative controls did not show signals, positive controls showed correct peaks (housekeeping and blood markers only). For preparation of cDNA specimens, an appropriate number of cDNA batches (20 µL each) were pooled and redistributed into aliquots of 5 to 20 µL. All protocols were provided as example protocols, but participants could use chemistries and instrumentations of choice. An overview of the methodologies used during stain analysis is presented in Supplementary Table 1. Participating laboratories were asked to use the entire stain to prevent differences from using less material. For low level samples (indicated by a DNA yield below 2 ng), it was advised to concentrate the RNA extract down to 12 µL by applying an ethanol precipitation as described in [30] and use 10 µL in a single plus RT cDNA reaction and 2 µL in a minus RT cDNA control.

Endpoint PCR

The 19-plex described in [30] was supplemented with vaginal mucosa marker CYP2B7P1 [19] (amplicons size 146 bp; forward primer 5'-VIC-AGTCTACCAGGGATATGGCATG; reverse primer 5'-CTATCAGACACTGAGCCTCGTCC; final primer concentration 1.6 µM), menstrual secretion marker MMP10 [5,27] (amplicons size 107 bp; forward primer 5'-VICGCATCTTGCATTCCTTGTGCTGTTG; reverse primer 5'-GGTATTGCTGGGCAAGATCCTTGTT; final primer concentration 1.6 µM) and skin marker LCE1C [16,18] (amplicons size 99 bp; forward primer 5'-NED-TGTGACCCCGCTCCTGAATCCG; reverse primer 5'-CTTGGGAGGGCACTTGGGGGTG; final primer concentration 0.02 µM). To create the space in the multiplex to add these three markers, the general mucosa

markers KRT13 and SPRR2A were removed. In this way, a 20-plex was created. A large batch of 5x primer stock for this multiplex was prepared and aliquoted to provide all laboratories with the exact same primer mixture in all RT-PCRs.

The suggested cDNA inputs in the RT-PCR analyses were 1 µL for cDNA specimens with indicated cell type, 0.5 and 2 µL for cDNA samples with unspecified cell type(s), 0.1, 0.5, 1 and 4 µL for stain cDNAs and 3.5 µL for minus RT, negative and positive controls.

Capillary electrophoresis and analysis of DNA and RNA profiles

Amplified fragments were separated and detected on various types of standard genetic analysers using different separation matrices as indicated in Supplementary Table 1. As RT-PCR products are generated using a homemade multiplex, removal of dye-blobs prior to analysis is essential, and the various approaches used are indicated in Supplementary Table 1. For the analysis of RNA profiles, a 150 rfu detection threshold was suggested by the organising laboratory [30]. However, some labs used a lower threshold of 50 or 100 rfu as indicated in Supplementary Table 1. No intervention occurred because peaks were just below or above threshold. For the analysis of DNA profiles, participating laboratories used their own protocols.

DNA profiling

DNA profiles were generated using the AmpFℓSTR® NGM™ PCR Amplification Kit (NGM) (Life Technologies) using a maximum of 500 pg DNA. PCR products were separated according to standardized protocols [24] using a 3130XL Genetic Analyzer (Life Technologies) with POP-4 (Life Technologies) separation matrix using 3 kV, 15 s injection settings. Profile analysis was performed using Genemapper ID-X version 1.1.1 (Life Technologies) and a detection threshold of 50 rfus.

Scoring of RNA data

Following the procedure described in [1], four replicate RNA profiles were generated using an informative cDNA input, and the RNA data were evaluated by applying an '$x=n/2$' scoring system per cell type [1]. Here, 'x' reflects the number of observed and 'n' the number of theoretically possible peaks in all replicates. A cell type is scored as 'observed' if $x \geq n/2$, 'not observed' if $x=0$ and 'sporadically observed' if $0<x<n/2$. For co-expressed cell types, 'and fits' is added when (sporadically) observed. Cell types scored as 'sporadically observed' are generally regarded as 'not reliably observed' and tissues scored as 'and fits' as 'not present as such' [1].

Results

Analysis of cDNAs

Nine laboratories including the organising laboratory participated in the exercise. All laboratories had experience with RNA analyses; eight had participated (one as organiser) in EDNAP RNA exercises [24-27]. The exercise started with the analysis of reference RT-PCR, which enabled adjustment of the provided marker bin settings in the analysis software if necessary. Some differences in peak heights occurred; one participant had on average 2.5 times higher peaks than the organising laboratory while another had approximately 0.5 times lower peaks, which seems due to the use of lower injection settings (Supplementary Table 1).

The next set consisted of 24 cDNA specimens derived from three RNA inputs and a minus RT control for each of the six cell types. The results are presented in Table 2 and differences in marker sensitivity are evident. Although this finding may have been affected by the use of a single donation for each cell type, the trends befit observations in the organising laboratory. The difference is most extreme for the vaginal mucosa markers: whilst CYP2B7P1 responds almost fully for the three RNA inputs, no peaks are obtained for HBD1 and only some for MUC4. For HBD1 this probably derives from suboptimal performance in multiplex analysis [30]. For the housekeeping markers, GAPDH appears to be the least and 18S-rRNA the most robust marker. Non-specific signals are occasionally seen both in non-target cell types and in minus RT samples by all laboratories. Three cases appear more frequent: 1) blood and especially CD93 signals for vaginal mucosa, which is unprecedented and may have been the result of a trace of blood (or menstrual secretion for which CD93 appears the most prominent blood marker) in this specific donation; 2) skin marker signals (LOR and to lesser extent CDSN) for vaginal mucosa, indicating that LCE1C is the more specific skin marker and 3) MMP10 signals for various non-target cell types and minus RT blanks, which may be related to the relatively high signals for true MMP10 peaks, which are on average 5320 rfu, while for the other markers the average height ranges from 495 rfu (MUC4) to 3190 rfu (HBB). A lower primer concentration for MMP10 may be beneficial.

The next task involved the analysis of ten numbered cDNA specimens, indicated to be single source but purposefully comprising eight single source cDNAs (blood, saliva, two times skin, vaginal mucosa, menstrual secretion, semen fertile donor, semen azoospermic male), one mixture (vaginal mucosa with blood) and one water sample. To illustrate the effect of cDNA input on RNA profiling results, participants were asked to generate two RNA profiles with a four-fold difference in cDNA input (0.5 and 2.0 µL, respectively). In theory, the participants can detect a total of 208 cell type-specific peaks for the ten cDNA specimens with each cDNA input. Using the 0.5 µL input, 109 peaks were detected and with the higher input (2 µL except for one laboratory

that used 1 µL), 161 peaks were observed. Marker dropout was predominantly seen for vaginal mucosa marker HBD1 and blood marker AMICA1, which are both among the less sensitive markers when testing the RNA input series (Table 2).

Table 2. Percentages of detected markers when analysing cDNAs derived from three inputs of single source RNAs. For each cDNA specimen, one RNA profile was produced by each participant with an input of 1 µL cDNA (Table 1).

		Blood[a]			Saliva[a]			Semen[a]			Skin[a]			Menst secr[a]			Vag muc[a]			-RT[b]
RNA input (µL)		1.0	0.4	0.2	1.0	0.4	0.2	1.0	0.4	0.2	1.0	0.4	0.2	1.0	0.25	0.125	1.0	0.4	0.2	-
Cell type specific markers																				
Blood	HBB	100%[c]	88%	63%										13%	0%	0%				
	CD93	100%	63%	13%							13%			100%	50%	50%	50%	38%	13%	2%
	AMICA1	63%	25%	13%										38%	0%	0%	13%			
Saliva	STATH				88%	13%	13%													
	HTN3				75%	38%	38%													
Semen	SEMG1						25%	100%	100%	100%										
	PRM1							88%	88%	88%										
Skin	CDSN				13%						100%	100%	63%	13%	13%	13%	63%	25%	13%	
	LCE1C							13%			88%	38%	38%							
	LOR										100%	63%	0%	13%			88%	88%	63%	2%
Menstrual secretion	MMP10	13%				13%	13%	13%			13%		13%	100%	100%	88%	13%			10%
	MMP7				13%									100%	63%	25%				
	MMP11													88%	38%	25%				
Vaginal mucosa	HBD1										13%			0%	0%	0%	0%	0%	0%	
	MUC4										13%			100%	75%	38%	25%	0%	0%	
	CYP2B7P1													100%	100%	38%	100%	100%	88%	2%
General mucosa marker																				
	KRT4				88%	63%	38%			13%				100%	100%	75%	100%	88%	88%	
Housekeeping markers																				
	ACTB	100%	100%	88%	25%	0%	13%	38%	50%	13%	88%	38%	13%	100%	100%	100%	100%	100%	100%	
	18S-rRNA	100%	100%	88%	63%	38%	38%	75%	63%	50%	88%	88%	88%	100%	100%	100%	100%	88%	63%	
	GAPDH	100%	88%	38%	13%	0%	0%	25%	13%	0%	25%	13%	13%	100%	88%	63%	88%	63%	25%	

[a]Results are based on the data of eight laboratories as one laboratory used an unsuccessful method to purify PCR products.
[b]Six minus RT samples (one per cell type) were provided to each participant and percentages are based on 48 profiles.
[c]Correct marker signals are shaded green, false positive signals are shaded red. Darker shades indicate that a higher percentage of laboratories observe the signal.

Peaks detected using both the 0.5 µL and the 2 µL input (91 in total) are on average 3.9 times higher for the 2 µL, which complies with the four-fold increased input. With a higher amount of cDNA, more non-target peaks were seen as well: 21 for the lower and 42 for the higher input, which is a total of 63 marker drop-ins in 140 RNA profiles. Again, skin markers LOR and CDSN gave false positive signals in vaginal mucosa and menstrual secretion specimens (23 and 16 observations for LOR and CDSN respectively), and menstrual marker MMP10 showed occasional drop-in signals (13 times) among all specimens including the water sample (data not shown). The marker calling for the mixture and the water sample was not affected by labelling them as single source cDNAs.

The last set of cDNAs involved four mixtures having two or three cell types in balanced (1:1, based on the corresponding DNA profiles) or unbalanced (up to 1:10) ratios. Participants were asked to determine optimal cDNA input from the profiling results for a 0.5 and 2 µL input and generate four replicate RNA profiles for application of the '$x=n/2$' scoring system. The results are presented in Table 3.

Chapter 1

Table 3. Results of the analysis of four cDNA mixtures. Participants generated four informative RNA profiles and applied '*x*=*n*/2' counting system that scores results into five different categories [1].

Blood & Saliva (1:1)	Blood[a]	Saliva	Semen	Menstrual	Vaginal	Skin	Mucosa
Not observed			8/8[b]	6/8	6/8	7/8	
Observed	8/8	8/8					
Sporadically observed				2/8	2/8	1/8	
Observed & fits							8/8
Sporadically observed & fits							
Blood & Saliva & Skin (1:1:1)	**Blood**	**Saliva**	**Semen**	**Menstrual**	**Vaginal**	**Skin**	**Mucosa**
Not observed			8/8	3/8	5/8		
Observed	8/8	6/8				8/8	
Sporadically observed		2/8		5/8	3/8		
Observed & fits							8/8
Sporadically observed & fits							
Vaginal & Semen (fertile) (10:1)	**Blood**	**Saliva**	**Semen**	**Menstrual**	**Vaginal**	**Skin**	**Mucosa**
Not observed	6/8	8/8		2/8			
Observed			7/8		8/8	8/8	
Sporadically observed	2/8		1/8	6/8			
Observed & fits							8/8
Sporadically observed & fits							
Menstrual & Semen (sterile) & Skin (10:5:1)	**Blood**	**Saliva**	**Semen**	**Menstrual**	**Vaginal**	**Skin**	**Mucosa**
Not observed		8/8				2/8	
Observed	1/8		7/8	8/8	2/8	1/8	
Sporadically observed	1/8		1/8			5/8	
Observed & fits	3/8				6/8		8/8
Sporadically observed & fits	3/8						

[a]Colour coding for cell types is: black cell = present; white cell = not present; grey cell = co-expressed.
[b]Colour coding of table cells for the results is: green cell = correct i.e. not observed when not present or observed (& fits) when present; light green cell = sporadically observed (regarded not observed) when not present; red cell = incorrect i.e. observed but not present or not observed when present; light red cell = sporadically observed (regarded not observed) but present; grey cell = aberrant result for co-expressed cell types.

It becomes apparent that it is helpful to use the category 'sporadically observed' (which we generally regard as not observed) as this category was used 21 times for cell types not present in the mixtures and nine times for cell types present. In total, a cell type was missed eleven times, which in seven events concerned skin being the lowest component in mixture four (ratio 10:5:1) (Table 3). The other four missed inferences relate to three different components in three mixtures (Table 3). A false positive skin identification was obtained by all participants in the third mixture. This seems a consequence of the unintended responses of LOR and lesser extent CDSN in vaginal mucosa: in the 32 RNA profiles generated by the eight participants for this mixture, 29 LOR, 18 CDSN and no LCE1C signals were seen. It seems beneficial to remove LOR from the multiplex or exclude LOR results from interpretation by the scoring system.

Analysis of stains

Exercise part 1 on the cDNA sets familiarised participants with RNA profile analysis using the provided 20-plex and the application of the '*x*=*n*/2' counting system. The results were provided as feedback prior to the second exercise part that involved DNA and RNA analysis of challenging mock casework stains. The design of the four stains is given in Table 4.

Table 4. Design of the four stains that compose part 2 of the exercise.

Substrate	Preparation	Challenge
1: Cotton swab	Saliva ♀$_{D1}$ & saliva ♀$_{D2}$: daughter & mother; unequal DNA amounts	One body fluid by two donors; Parent-child relation between donors
2: Coloured fleece	Skin ♀$_{D1}$ & menstrual secr. $_{D1}$ ♀ & blood ♂$_{D2}$	Presence blood masked by menstrual secretion; Skin and menstrual secretion given by same donor
3: Green linen	Skin ♀$_{D1}$ & diluted blood ♀$_{D1}$ & skin ♀$_{D2}$	Low amount: EtOH precipitation may be needed; Donor 1 gives two cell types, skin given by two donors
4: Nail clipping	Nail/skin ♂$_{D1}$ & vaginal muc. ♀$_{D2}$ & semen azoospermic male ♂$_{D3}$	Only two of three donors will contribute DNA; Male DNA part represents skin and not seminal fluid

The participants were asked to estimate the number and genders of the donors, derive a result score for each cell type using to the '*x*=*n*/2' counting system and determine which cell types are present or not present (or 'no statement'). Both LOR and HBD1 were excluded during scoring because of non-specific signals with vaginal mucosa or insufficient multiplex amplification. Participants used various methodologies to extract nucleic acids and derive DNA and RNA profiles (Supplementary Table 1), and yields and profiling outcomes showed considerable variation (Table 5, Supplementary Fig. 1).

For stain 1 (two females giving saliva), all participants scored saliva, which is the only body fluid present, as observed. Four times a cell type not present in the stain was scored as observed: once blood, once menstrual secretion and twice skin. Two of these false positives (blood and once skin) relate to the use of a lower detection threshold than that of 150 rfu advised by the organising laboratory. The false menstrual secretion score relates to the analysis of slightly overloaded RNA profiles in which trailing signals occur about ten nucleotides before the parent peaks: the trailing signal of saliva marker HTN3 fits the bin of the MMP7 menstrual secretion marker (although 0.3 nt smaller than true MMP7 peaks), and together with MMP10 background signals (also seen in part 1, Table 2) menstrual secretion gets scored 'observed'. Actually all peaks including those for housekeeping markers show these trailing signals and with more expertise in analysing RNA profiles generated with the 20-marker multiplex, these false signals may have been recognised. The fourth false positive observation (skin) seems to relate to the analysis of slightly over-amplified profiles on a 3500 genetic analyser. The dynamic rfu range of the 3500 platform is approximately four-fold higher than that of the 3130XL platform that was used to generate the provided analytical thresholds. Concomitantly, true signals had up to 34000 rfu

Chapter 1

while the false skin signals were between 800 and 2400 rfu. When transferring RNA profiling to 3500 platforms, analytical thresholds need to be re-established.

For stains 2, 3 and 4, no false positive scorings occur (Table 5). However, for these three stains not all cell types present were scored as observed.

Table 5. Summarised results of the DNA and RNA data for the four stains (Table 4) that compose part two of the exercise.

Feature	Scoring[a]	Stain 1	Stain 2	Stain 3	Stain 4	All stains
Estimated minimum number and genders contributors	One	♀: 1/8	-	♀: 5/8	♀: 4/8	-
	Two	♀&♀: 7/8	♀&♂: 8/8	♀&♀: 3/8	♀&♂: 2/8 ♀& ? : 2/8	-
Average[b] % detected non-shared alleles	Donor 1	99%	100%	99%	16%	-
	Donor 2	84%	86%	28%	100%[c]	-
Cell types present	Observed[de]	8/8	19/24[f]	6/16	18/24	71% (51/72)
	Not observed	0/8	1/24	5/16	2/24	11% (8/72)
	Sporadically obs[d]	0/8	4/24	5/16	4/24	13% (13/72)
Cell types not present	Observed	4/40	0/16	0/32	0/24	4% (4/112)
	Not observed	25/40	15/16	27/32	17/24	75% (84/112)
	Sporadically obs	11/40	1/16	5/32	7/24	21% (24/112)

[a]Data of eight laboratories are included; for one laboratory the replicates seem contaminated by multiple cell types.
[b]No standard deviation is provided; results per participant are given in Supplementary figure 1.
[c]For donor 3 no specific alleles were observed.
[d]Both the categories 'observed' and 'observed & fits' or 'sporadically observed' and 'sporadically observed & fits'. For more details see Supplementary Table 1.
[e]Colour coding of table cells for the results is: green cell = correct *i.e.* observed when present or not observed or sporadically observed when not present; red cell = incorrect *i.e.* not observed or sporadically observed when present or observed but not present.
[f]Only the added cell types, not co-expressed cell types are regarded (so skin, blood and menstrual but not vaginal).

For stain 2 (female giving menstrual secretion and skin plus male giving blood), all participants observed menstrual secretion, while blood was scored as observed by six and skin by five of the eight laboratories. This is striking for blood as the blood donor corresponds to almost half of the total rfu weight in most of the DNA profiles. Apparently, blood gives relatively low RNA signals compared to the DNA signals or menstrual secretion gives relatively low DNA signals compared to the RNA signals indicating, as already shown in [32], that it is not appropriate to associate the strongest signals from RNA and DNA typing of co-extracted samples with each other. Stain 3 (skin and diluted blood from one female plus skin from another female) was challenging due to the low amounts of cell material present, which is reflected by a low number of observed scorings for present cell types: blood is scored only once and skin five times as observed (Table 5). Only two of the participants proceeded to ethanol precipitation of the RNA prior to cDNA synthesis, and the one positive scoring for blood was obtained by one of these participants. For stain 4 (nail clipping from a male with vaginal mucosa female and seminal fluid azoospermic male), skin and vaginal mucosa were scored as observed by six laboratories while semen was reported as observed four times. Since the semen contribution is of an azoospermic male, only the seminal fluid

marker SEMG1 will respond. To have semen scored as 'observed', SEMG1 needs to give a positive signal in all four replicates. Actually, this was only the case for three of the four laboratories: one participant had three (true) SEMG1 signals and one (false) PRM1 peak that was a pull-up from the LOR signal, which could have been recognised by peak shape and size as it was 0.9 nucleotide larger than a true PRM1 signal. When considering the scorings of all stains together, 4% were false positives scores that can be explained from suboptimal profile analysis and 11% were false negative scores (Table 5). The category 'sporadically observed (& fits)' was used for 13% of the scorings for cell types present and for 21% of the scorings for cell types not present, suggesting that this is a useful category as it seems to lower the number of false positive results.

Combined interpretation of DNA and RNA data for the stains

Using the inferences on donor numbers, their genders and the cell types observed, participants were requested to give a combined interpretation of DNA and RNA data. However, the underlying profiling data were so different that it was not constructive to compare these interpretations, and the dataset of the organising laboratory was sent out. This set consisted of the DNA quantification results (total and male-specific), one DNA profile and four RNA replicate profiles for each stain. Characteristics of this dataset are summarised in Table 6. The participants estimated the number and genders of the donors, derived a result score for each cell type and determined which cell types are present or not present, or receive 'no statement'. With stains 1, 3 and 4 participants provided different responses for the numbers and genders of donors (Table 6), indicating that the exact same DNA data are evaluated differently. The cell type scorings are mainly in accordance with the provided guidelines (Table 6). In all cases when cell types were scored as not observed, participants indicated these cell types as not present. Cell types scored as observed were always regarded present. Blood in stain 2 was classified four times as observed and three times as observed and fits because menstrual secretion signals were detected as well. In all instances when 'observed and fits' was chosen, 'no statement' was selected for presence. The sporadic signals seen for menstrual secretion in stain 3 and blood in stain 4 received different interpretations; six times the corresponding cell types were regarded as not present and six times 'no statement' was made.

The interpretations for the DNA and RNA data were combined into a verbal conclusion in which donors and cell types were associated, if possible. For stains 1, 2 and 3 the majority of the verbal conclusions are correct, have a correct interpretation among the multiple options given in the statement, or leave room for the correct alternative by using 'probably' (Table 6). For stain 2 twice an incorrect interpretation was given that appears to derive from not recognising that in case menstrual secretion is present blood signals may also originate from a peripheral blood contribution.

Table 6. Data (bold) and interpretations (not bold) for the NFI dataset on the four stains regarding DNA (A), RNA (B) and combined results (C).

A	Quant (ng/µL)[a]		AMEL (rfu)		Number alleles[b]		Donor 1[b]		Donor 2[b]		
	Total	Male	X	Y	Profile	Locus	Alleles	Rfu	Alleles	Rfu	Interpretation DNA results
Stain 1: ♀$_{D1\ child}$ & ♀$_{D2\ parent}$	1.02	0.0	1651	-	36	≤3	10/10	509	10/11	95	4x two ♀; 2x ♀ & unknown; 1x one ♀
Stain 2: ♀$_{D1}$ & ♂$_{D2}$	1.23	0.12	1719	345	52	≤4	26/26	620	26/26	576	7x ♀ & ♂
Stain 3: ♀$_{D1\ major}$ & ♀$_{D2\ minor}$	0.02	0.0	1348	-	35	≤4	17/17	523	6/19	75	3x two ♀; 2x ♀ & unknown; 2x one ♀
Stain 4: ♂$_{D1\ minor}$ & ♀$_{D2\ major}$ & ♂$_{D3\ sterile}$	0.54	0.01	2391	-	30	≤4	22/22	1621	3/23	113	3x ♀ & ♂; 1x ♀ & unknown; 3x one ♀

B	Blood	Saliva	Semen	Skin	Menstrual	Vaginal
Stain 1:	-	Saliva	-	-	-	-
Marker count	0/12	8/8	0/8	0/8	0/12	0/8
Result score[c]	7x not obs	7x obs	7x not obs	7x not obs	7x not obs	7x not obs
Presence score	7x not present	7x present	7x not present	7x not present	7x not present	7x not present
Stain 2:	Blood	-	-	Skin	Menstrual	-
Marker count	8/12	0/8	0/8	4/8	12/12	0/8
Result score	4x obs; 3x obs&fits	7x not obs	7x not obs	6x obs; 1x spor&fits	7x obs	7x not obs
Presence score	4x present; 1x not present; 2x no statement	7x not present	7x not present	6x present; 1x no statement[d]	7x present	7x not present
Stain 3:	Blood	-	-	Skin	-	-
Marker count	11/12	0/8	0/8	8/8	3/12	0/8
Result score	7x obs	7x not obs	7x not obs	7x obs	6x: spor; 1x not obs[e]	6x: not obs; 1x spor[e]
Presence score	7x present	7x not present	7x not present	7x present	2x not present; 5x no statement	7x not present
Stain 4:	-	-	Semen (sterile)	Skin (nail)	-	-
Marker count	2/12	0/8	4/8	8/8	0/12	6/8
Result score	6x spor; 1x spor&fits[f]	7x not obs	5x obs; 1x spor[g]; 1x spor&fits[h]	6x obs; 1x spor&fits[i]	7x not obs	6x obs; 1x not obs[j]
Presence score	2x no statement; 5x not present	7x not present	6x present (3x sterile); 1x no statement	7x present	7x not present	6x present; 1x not present

C	Overall interpretation	Correctness
Stain 1: ♀$_{D1\ child}^{saliva}$ & ♀$_{D2\ parent}^{saliva}$		
1x	♀$_{major}$ = saliva, ♀$_{minor}$ = saliva	Correct
3x	♀$_{major}$ = saliva, ♀$_{minor}$ = saliva or unknown cell type	Correct
1x	♀$_{major}$ = saliva, presence 2nd donor not confirmed	Correct although one donor missed
2x	Two ♀ donors and saliva, no association	Correct; some under-interpretation
Stain 2: ♀$_{D1}^{skin\ \&\ menstrual}$ & ♂$_{D2}^{blood}$		
1x	♀ = menstrual, ♂ = skin	Incorrect; presence blood missed
2x	♀ = menstrual, ♂ = skin or blood; ♀ = menstrual, ♂ = probably skin	Correct; little over-interpretation
2x	♀ = menstrual, ♂ = skin and/or blood; ♀ = menstrual, ♂ no association	Correct
1x	♀ = menstrual, skin no statement, blood either ♀ or ♂ donor	Correct; unclear formulation
1x	♀ & ♂ DNA donor, menstrual & skin (blood co-expressed), no association	Incorrect; presence blood missed
Stain 3: ♀$_{D1\ major}^{skin\ \&\ diluted\ blood}$ & ♀$_{D2\ minor}^{skin}$		
4x	Two ♀$_{DNA}$ donors, blood & skin, no association	Correct
1x	♀$_{major}$ = skin, ♀$_{minor}$ = blood or ♀$_{D1}$ = skin + blood, ♀$_{D2}$ = unknown	Correct; little over-interpretation
1x	Confident on one ♀ donor, blood & skin will be from her	Incomplete for DNA, risky on association
1x	♀ donor and skin present, possibly 2nd ♀ giving blood or sporadic menstrual with blood co-expressed, no association	Correct; but it is missed that 2nd donor could have given skin
Stain 4: ♂$_{D1\ minor}^{skin\ (nail)}$ & ♀$_{D2\ major}^{vaginal}$ & ♂$_{D3\ azoospermic}^{semen}$		
2x	♀$_{major}$ = vaginal + (probably) skin, ♂$_{minor}$ = semen	Incorrect; male donor did not give semen
1x	♀$_{major}$ = vaginal, ♂$_{minor}$ = semen, skin no statement	Incorrect; male donor did not give semen
2x	One ♀ donor = vaginal + skin, semen not reliable or only RNA	Incorrect; male donor not provide skin
1x	♀ and unknown, skin & vaginal & seminal fluid, no association	Under-interpretation; 2nd donor is ♂ (low male quant) so ♀ gave vaginal
1x	No report: discrepancy observing semen while no ♂ genotype	Incorrect; no PRM1/spermatozoa/DNA

[a]Sensitivity *Alu*-based DNA quantification system is >0.0005 ng/µL for total and >0.004 ng/µL for male human DNA concentration.
[b]Allele counts on 15 STRs, Amelogenin excluded. Donor-specific counts involve only non-shared alleles with homozygous alleles counted as two.
[c]Observed is abbreviated obs, sporadically observed as spor. Results of 7 participants: one did not return data and the organiser was aware of stain design.
[d]Laboratory 4: skin signals may be a false positive response (3 CDSN and 1 LCE1C peaks) with menstrual/vaginal.
[e]Laboratory 1: scoring for menstrual and vaginal swapped?
[f]Laboratory 7: the 2 blood peaks are sporadic and may fit vaginal.
[g]Laboratory 8: 4/8 semen peaks scored as 'sporadically observed', and interpreted as 'no statement'.
[h]Laboratory 1: 4/8 semen peaks scored as 'sporadically observed & fits', interpreted as 'present': mis-scoring?
[i]Laboratory 1: 8/8 skin peaks scored as 'sporadically observed & fits', interpreted as 'present': mis-scoring?
[j]Laboratory 1: 6/8 vaginal peaks scored as 'not observed', interpreted as 'not present': mis-scoring?

For stain 4, which was clearly the most complex stain that was sent out, no correct interpretation was provided. In this stain, the presence of the second male donor giving seminal fluid is fully masked at the DNA level by the female giving vaginal mucosa and the first male providing the nail clipping. Sterile seminal fluid does not carry spermatozoa, but low amounts of epithelial cells may be present and provide some DNA. Actually when a Y-STR profile was generated, all alleles of the nail donor and some low level signals for the seminal fluid donor were visible (results not shown). Most of the incorrect interpretations associated the male DNA component with the semen contribution without realising that this male may have provided the skin residing on the nail clipping and another male the seminal fluid. Other participants underestimated the number of contributors and consequently linked both vaginal mucosa and skin to the female donor, while she only provided vaginal mucosa. One participant reported a discrepancy for the lack of a male profile while a positive result for semen was obtained, which implies that the azoospermic status of the donor was not recognised. It is also noticed that some participants were very reluctant to make associations such as for stain 1 for which one participant did not link the major DNA contributor to saliva (Table 6). Interestingly, sometimes participants mentioned that DNA signals may derive from unknown cell types or cell types 'below detection', while they had indicated that the not observed cell types were all 'not present'. Apparently, 'not present' is rather used as 'seemingly not present, but there may be signals below detection' and terms like 'no indication for presence' or 'the presence can not be excluded' may be more appropriate.

Discussion and concluding remark

This collaborative EUROFORGEN-NoE exercise explored the usefulness of forensic cell type inference by mRNA profiling. The methodology would expedite from a good human RNA quantitation system. The exercise used a 20-marker multiplex in which most markers performed acceptably except skin marker LOR that showed cross-reactivity with vaginal mucosa and vaginal mucosa marker HBD1 that had low amplification success. These markers are best removed or replaced in an updated future multiplex. The addition of a second seminal fluid marker would increase the detection chance of semen from azoospermic males since now, a signal is needed in all four replicates to confirm presence. Furthermore, the primer concentrations for menstrual secretion marker MMP10 can be lowered to prevent background signals and over-amplification. Redesign of the multiplex is challenging because of multiplex spacing and marker balance. RNA amplicons are preferably sized between 70 and 150 bp to allow analysis of compromised samples. Consequently, limited space is available for markers and recurrent bleed-through and artefact signals (split peaks and trailing signals) that should not culminate in the bins of other markers. Primer concentrations need to be

Chapter 1

such that both good sensitivity and low noise levels are obtained. Furthermore, the use of a relatively high detection threshold appears beneficial to prevent false positive marker callings. Experience with the multiplex may aid profile analysis as artefact peaks are better recognised.

Even for an optimised multiplex, it may be inevitable that RNA profiles have signal imbalances, marker dropout and marker drop-in, as mRNA expression is influenced by various biological factors. To assess the validity of cell type signals, interpretation guidelines are applied. In this study, we used the '$x=n/2$' scoring system [1]. This system worked adequately as only few false positive scores were obtained (Table 5), which were predominantly due to signal calls on artefact peaks. Employing these guidelines may come at the cost of not inferring all cell types that are present, which is important to keep in mind during case interpretation. As an alternative approach, a numerical scoring method is described [31] in which values are assigned to each of the used mRNA markers (five per body fluid) based on correct and incorrect expression in samples of known origin. From these numerical values a body fluid score is calculated and positive body fluid identification is given when the combined marker value is higher than a pre-determined threshold value. This is clearly different from the '$x=n/2$' scoring system, in which all markers have the same weight, implying that all markers are evenly effective for cell type inference. This is not always the case as for instance seminal fluid and spermatozoa markers respond different when analysing semen of an azoospermic male (which could be compensated by adding a second seminal fluid marker or by scoring presence of seminal fluid upon regarding SEMG1 results only). On the other hand, the '$x=n/2$' method presents a general approach applicable to different mRNA profiling assays like cell and organ typing [21,30]. A comparative study including compromised samples would be informative to assess both interpretation strategies.

In a forensic case, DNA and RNA data need to be combined. Distribution of a DNA/RNA dataset derived for four truly challenging stains assessed this aspect. These stains covered the most prominent complications in DNA/RNA profiling such as same cell type given by multiple donors (stains 1 and 3), same donor giving multiple cell types (stains 2 and 3), masking of a cell type (blood, stain 2) by a co-expressing cell type (menstrual secretion), low level analysis (stain 3) and a cell type giving RNA but no (or hardly) DNA signals (seminal fluid azoospermic male, stain 4). No helpful context information such as 'the nails were clipped from person X' or reference profiles were given. In addition, not all participants were experienced with formulating forensic verbal statements. While for stains 1, 2 and 3 many correct interpretations were given, none of the interpretations for stain 4 were correct. This stain represents the forensic analysis of nail clippings taken from an assailant of digital penetration of a victim who had had previous intercourse with an azoospermic male. Although this is an unusual scenario, it may happen and serves to illustrate that awareness regarding such

interpretationpitfalls is important when proceeding to RNA analysis in forensic casework.

In conclusion, with expertise in analysing RNA profiles, clear guidelines for data interpretation and awareness regarding potential interpretation pitfalls mRNA-based cell type inference may be ready for implementation in forensic casework.

Acknowledgements

The work leading to these results was financially supported from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n8 285487 (EUROFORGENNoE). Daniela Niederwieser (Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria) and Gnanagowry Shanthan (Department of Forensic Genetics, Norwegian Institute of Public Health, Oslo, Norway) are acknowledged for their excellent technical assistance. We thank Corina Benschop (Netherlands Forensic Institute, The Hague, The Netherlands) for critically reading the manuscript and Petra Maaskant, Ingrid Blom and Ankie van Gorp (Netherlands Forensic Institute, The Hague, The Netherlands) for testing the NFI dataset for interpretation.

Supplementary material

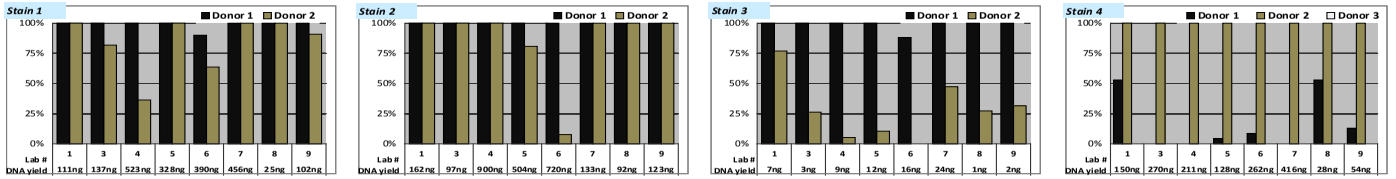
Supplementary Table 1. Methodologies used by the different laboratories for stain analysis.

Step	Methodologies used[a]
DNA/RNA extraction kit	5x **mirVana miRNA isolation kit & QIAamp mini** (Ambion/Qiagen) 4x Allprep RNA/DNA mini kit (Qiagen)
DNA elution	4x **100 µL** - 2x 80 µL - 3x 50 µL
DNA quantification	2x in house prepared system such as ***Alu* system** 2x Quantifiler Duo - 3x Quantifiler Human (Life Tech) 1x Qubit dsDNA HS assay (Invitrogen) 1x no quantification, standard 2.5 µL input
PCR kit	2x **NGM** - 2x NGM^{Select} - 1x Identifiler^{+} - 1x SEFiler^{+} - 1x SGM^{+} (Life Tech) 1x ESI17 - 1x ESI16 (both Promega)
Low template typing with stain 3	8x none - 1x 34 cycles duplicate
RNA elution	6x **60 µL H_2O** - 3x 30 µL H_2O
DNase digestion	8x **Turbo DNA-free kit** (Ambion) - 1x RNase-Free DNase (Qiagen)
RNA quantification	8x **none** - 1x Qubit RNA Assay kit (Invitrogen)
Reverse transcription kit	6x **RetroScript** (Ambion) - 1x iScript (Biorad) - 2x Superscript III (Invitrogen)
RNA input for reverse transcription	6x **10 µL** - 1x 12 µL - 1x 8 µL - 1x 20 ng
EtOH prec. RNA for low level stain 3	2x **yes** (full RNA extract) - 7x no
RT-PCR	9x **provided protocol and primer mix**
Informative cDNA input stain 1	4x 4 µL - 1x 1.5 µL - 2x 1 µL - 1x 0.25 µL[b] - 1x 0.1 µL
Informative cDNA input stain 2	2x 4 µL - 2x 1 µL - 2x 0.5 µL[b] - 1x 0.2 µL - 2x 0.1 µL
Informative cDNA input stain 3	1x 7.5 µL - 1x 5 µL - 4x 4 µL - 2x 3 µL[b] - 1x 1 µL
Informative cDNA input stain 4	2x 4 µL - 1x 3 µL[b] - 1x 1 µL - 2x 0.5 µL - 1x 0.25 µL - 2x 0.1 µL
Post PCR purification	8x **minElute** (Qiagen) - 1x Sephadex spin columns 96-well plate
Genetic analyser RNA profiles	7x **3130(XL)** - 1x 3730 - 1x 3500 (Life Tech)
Separation polymer RNA profiles	5x **POP-7** - 3x POP-4 - 1x POP-6 (Life Tech)
Injection setting	8x **3kV@10s**[c] - 1x 2kV@7s
Detection threshold RNA profiles	7x **150 rfu** - 1x 100 rfu - 1x 50 rfu

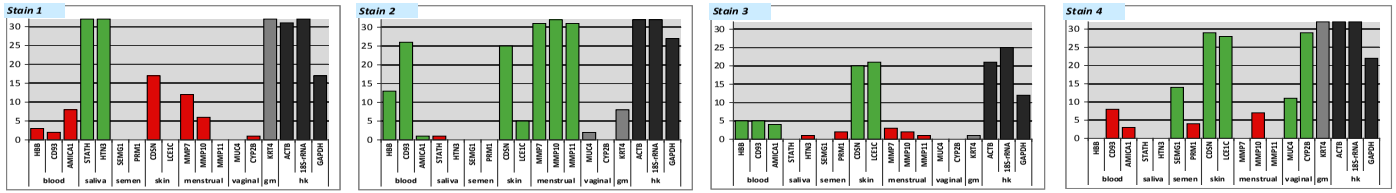
[a]In bold is the method for which an example protocol was provided, when the number is in grey, the method was by the participant whose data were excluded.
[b]Average input for lab using a range of inputs.
[c]One lab diluted the PCR products for stain 1 ten-fold and those for stain 2 five-fold.

Supplementary Figure 1. Detailed results of the DNA and RNA data generated by eight participants for the four mock case stains. The results of one laboratory are not included as replicate RNA analyses appeared contaminated by various cell types (whereas RNA profiles from the serial input analysis were not contaminated). Skin marker LOR and vaginal marker HBD1 are not used for scoring results. In the plots illustrating the number of detected marker peaks, the cell types present are coloured green, the cell types not present red, co-expressed cell types (such as general mucosa, gm) grey and housekeeping (hk) markers dark grey. For the scoring results black cells indicate cell types present, white cells mark cell types not present and grey cells shown co-expressed cell types. Scorings are coloured green if correct (light green in case of sporadic scoring, which is regarded not observed) and red if incorrect (light red in case of sporadic scoring, which is regarded not observed). Aberrant scorings regarding co-expressed markers or when falsely the category '& fits' is chosen are marked in grey.

DNA: Percentage detected alleles for each donor (based on non-shared alleles only)



RNA: Numbers of detected marker peaks for 32 profiles (8 participants, 4 replicates each)



RNA: Cell type scorings for 8 participants

Stain 1	blood	saliva	semen	skin	menstrual	vaginal	mucosa
observed		7		2	1		
obs. & fits	1	1					8
spor. obs.	1			3	4		
spor. & fits	1			1		1	
not obs.	5		8	2	3	7	

Stain 2	blood	saliva	semen	skin	menstrual	vaginal	mucosa
observed	2			5	7		
obs. & fits	4				1		2
spor. obs.	1	1		2			1
spor. & fits	1					2	2
not obs.		7	8	1		6	3

Stain 3	blood	saliva	semen	skin	menstrual	vaginal	mucosa
observed	1			5			
obs. & fits							0
spor. obs.	3	1	1	1	3		1
spor. & fits				1			
not obs.	4	7	7	1	5	8	7

Stain 4	blood	saliva	semen	skin	menstrual	vaginal	mucosa
observed			4	6		6	2
obs. & fits				1		1	6
spor. obs.	2		2	1	2	1	
spor. & fits	2				1		
not obs.	4	8	2		5		

References

1. A. Lindenbergh, P. Maaskant, T. Sijen, Implementation of RNA profiling in forensic casework, Forensic Sci. Int. Genet. 7 (2012) 159-166.
2. M. Bauer, A. Kraus, D. Patzelt, Detection of epithelial cells in dried blood stains by reverse transcriptase-polymerase chain reaction, J Forensic Sci 44 (1999) 1232-1236.
3. M. Bauer, D. Patzelt, Protamine mRNA as molecular marker for spermatozoa in semen stains, Int J Legal Med 117 (2003) 175-179.
4. J. Juusola, J. Ballantyne, Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification, Forensic Sci Int 135 (2003) 85-96.
5. J. Juusola, J. Ballantyne, mRNA profiling for body fluid identification by multiplex quantitative RTPCR, J Forensic Sci 52 (2007) 1252-1262.
6. M. Bauer, D. Patzelt, Identification of menstrual blood by real time RT-PCR: technical improvements and the practical value of negative test results, Forensic Sci Int 174 (2008) 55-59.
7. M. Setzer, J. Juusola, J. Ballantyne, Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains, J Forensic Sci 53 (2008) 296-305.
8. D. Zubakov, E. Hanekamp, M. Kokshoorn, W. van Ijcken, M. Kayser, Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of timewise degraded samples, Int J Legal Med 122 (2008) 135-142.
9. 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.
10. C. Haas, C. Muheim, A. Kratzer, W. Bar, C. Maake, mRNA profiling for the identification of sperm and seminal plasma, Forensic Sci Int Genetics Supplement Series 2 (2009) 534-535.
11. K. Sakurada, H. Ikegaya, H. Fukushima, T. Akutsu, K. Watanabe, M. Yoshino, Evaluation of mRNAbased approach for identification of saliva and semen, Leg Med (Tokyo) 11 (2009) 125- 128.
12. R.I. Fleming, S. Harbison, The use of bacteria for the identification of vaginal secretions, Forensic Sci Int Genet 4 (2010) 311-315.
13. K. Sakurada, T. Akutsu, H. Fukushima, K. Watanabe, M. Yoshino, Detection of dermcidin for sweat identification by real-time RT-PCR and ELISA, Forensic Sci Int 194 (2010) 80-84.
14. C. Haas, E. Hanson, A. Kratzer, W. Bar, J. Ballantyne, Selection of highly specific and sensitive mRNA biomarkers for the identification of blood, Forensic Sci Int Genet 5 (2011) 449- 458.
15. J. Jakubowska, A. Maciejewska, R. Pawlowski, mRNA profiling in identification of biological fluids in forensic genetics. Problems of Forensic Science 87 (2011) 204-215.
16. E. Hanson, C. Haas, R. Jucker and J. Ballantyne, Identification of skin in touch/contact forensic samples by messenger RNA profiling. Forensic Science International: Genetics Supplement Series 3 (2012) 305-306.
17. S.M. Park, S.Y. Park, J.H. Kim, T.W. Kang, J.L. Park, K.M. Woo, J.S. Kim, H.C. Lee, S.Y. Kim, S.H. Lee, Genome-wide mRNA profiling and multiplex quantitative RT-PCR for forensic body fluid identification, Forensic Sci Int Genet 7 (2013) 143-150.
18. E. Hanson, C. Haas, R. Jucker and J. Ballantyne, Specific and sensitive mRNA biomarkers for the identification of skin in 'touch DNA' evidence. Forensic Sci. Int. Genet. 6 (2012) 548-558.
19. 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.
20. J. Jakubowska, A. Maciejewska, R. Pawlowski, K.P. Bielawski, mRNA profiling for vaginal fluid and menstrual blood identification, Forensic Sci Int Genet 7 (2013) 272-278.
21. 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.
22. D. van der Meer, M.L. Uchimoto, G. Williams, Simultaneous Analysis of Micro-RNA and DNA for Determining the Body Fluid Origin of DNA Profiles. J Forensic Sci 58 (2013) 967–971.
23. B. Bhoelai, B. de Jong, T. Sijen, Can mRNA markers distinguish traces generated by different types of contact? Forensic Sci Int Genet Suppl Series 4 (2013) e3-e4.
24. C. Haas, E. Hanson, W. Bar, R. Banemann, A.M. Bento, A. Berti, E. Borges, C. Bouakaze, A. Carracedo, M. Carvalho, A. Choma, M. Dotsch, M. Duriancikova, P. Hoff-Olsen, C. Hohoff, P. Johansen, P.A. Lindenbergh, B. Loddenkotter, B. Ludes, O. Maronas, N. Morling, H. Niederstatter, W. Parson, G. Patel, C. Popielarz, E. Salata, P.M. Schneider, T. Sijen, B. Sviezena, L. Zatkalikova, J. Ballantyne, mRNA profiling for the identification of blood--results of a collaborative EDNAP exercise, Forensic Sci Int Genet 5 (2011) 21-26.
25. C. Haas, E. Hanson, M.J. Anjos, W. Bar, R. Banemann, A. Berti, E. Borges, C. Bouakaze, A. Carracedo, M. Carvalho, V. Castella, A. Choma, G. De Cock, M. Dotsch, P. Hoff-Olsen, P. Johansen, F. Kohlmeier, P.A. Lindenbergh, B. Ludes, O. Maronas, D. Moore, M.L. Morerod, N. Morling, H. Niederstatter, F. Noel, W. Parson, G. Patel, C. Popielarz, E. Salata, P.M. Schneider, T. Sijen, B. Sviezena, M. Turanska, L. Zatkalikova, J. Ballantyne, RNA/DNA co-analysis from blood stains--results of a second collaborative EDNAP exercise, Forensic Sci Int Genet 6 (2012) 70-80.
26. C. Haas, E. Hanson, M.J. Anjos, R. Banemann, A. Berti, E. Borges, A. Carracedo, M. Carvalho, C. Courts, G. De Cock, M. Dotsch, S. Flynn, I. Gomes, C. Hollard, B. Hjort, P. Hoff- Olsen, K. Hribikova, A. Lindenbergh, B. Ludes, O. Maronas, N. McCallum, D. Moore, N. Morling, H. Niederstatter, F. Noel, W. Parson, C. Popielarz, C. Rapone, A.D. Roeder, Y. Ruiz, E. Sauer, P.M. Schneider, T. Sijen, D.S. Court, B. Sviezena, M. Turanska, A. Vidaki, L. Zatkalikova, J. Ballantyne, RNA/DNA co-analysis from human saliva and semen stains--results of a third collaborative EDNAP exercise, Forensic Sci Int Genet 7 (2013) 230-239.
27. C. Haas, E. Hanson, M.J. Anjos, K. Ballantyne, R. Banemann, B. Bhoelai, E. Borges, M. Carvalho, C. Courts, G. De Cock, K. Drobnic, M. Dotscr, R. Fleming, C. Franchix I. Gomes, G. Hadzic, S.A Harbison, J. Harteveld, B. Hjort, C. Hollard, P. Hoff-Olsen, C. Huls, C. Keyser, O. Maronas, N. McCallum, D. Moore, N. Morling, H. Niederstatter, F. Noel, W. Parson, C. Phillips, C. Popielarz, A.D. Roeder, L. Salvaderi, E. Sauer, P.M. Schneider, G. Shantan, D. Syndercombe Court, M. Turanska, R.A.H.van Oorschot, M. Vennemann, A. Vidaki, L. Zatkalikova, J. Ballantyne, 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.
28. J. Juusola, J. Ballantyne, Multiplex mRNA profiling for the identification of body fluids, Forensic Sci Int 152 (2005) 1-12.
29. R.I. Fleming, S. Harbison, The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids, Forensic Sci Int Genet 4 (2010) 244-256.
30. 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.
31. 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 (2012) 707-721.
32. 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.
33. 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.

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).

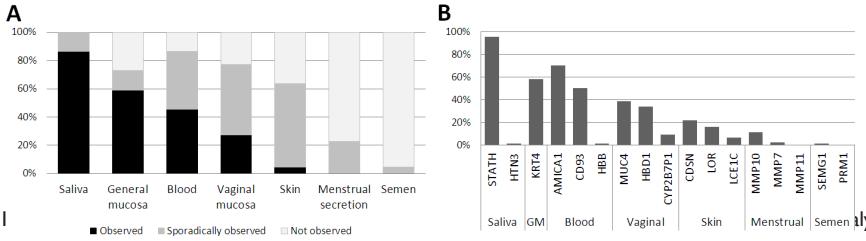


Figure 1 **lysis was performed in 4-fold to allow profile interpretation according to the *x*=*n*/2 guideline. (A) Scoring into categories for 22 samples. (B) Responsiveness of individual markers in 88 profiles (GM = General mucosa).**

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.

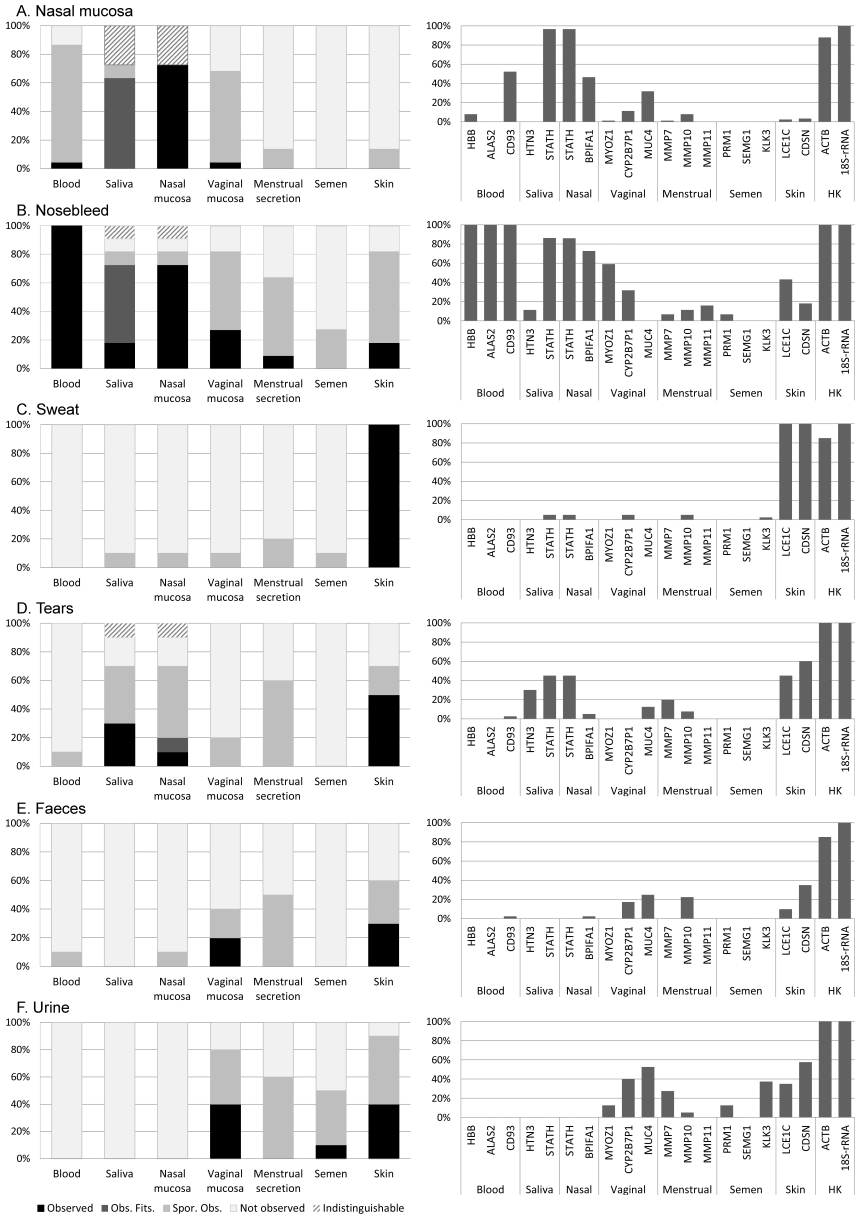


Figure 2. mRNA analysis of 22 nasal mucosa (A),11 nosebleed (B), and 10 sweat (C), tear (D), faeces (E) and urine (F) samples analysed with the updated Cell-typer (V3) multiplex Analysis was performed in 4-fold to allow profile interpretation based on the *x=n*/2 guideline (left). Saliva and nasal mucosa 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.

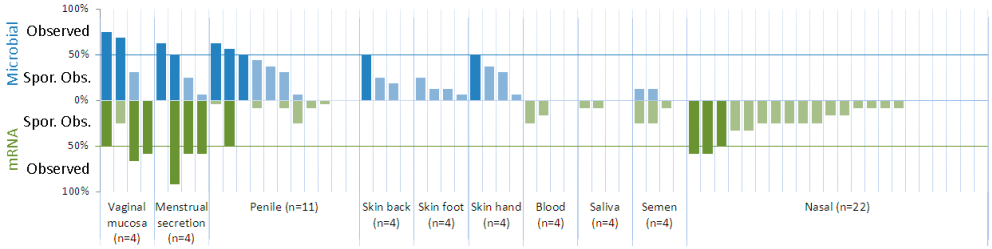


Figure 3. Marker detection percentages for vaginal mucosa containing, and vaginal mucosa lacking samples when analysed with microbial rRNA (blue) or endogenous mRNA (green) markers. Profile interpretation used the *x*=*n*/2 interpretation guideline and four RNA profiles. Detection percentages are determined 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.

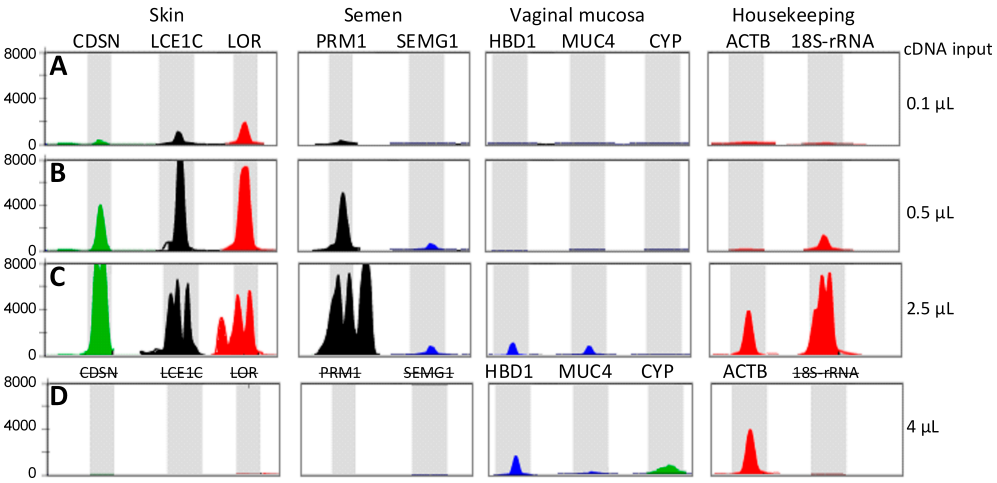


Figure 4. Differential analysis of an imbalanced vaginal mucosa/semen/skin mixture. A serial input (0.1, 0.5, 2.5 µL cDNA, A, B, C) was used to determine the optimal cDNA input. Vaginal mucosa signals are visible using a higher input (C) but this results in overloaded skin and semen signals. Replicate 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

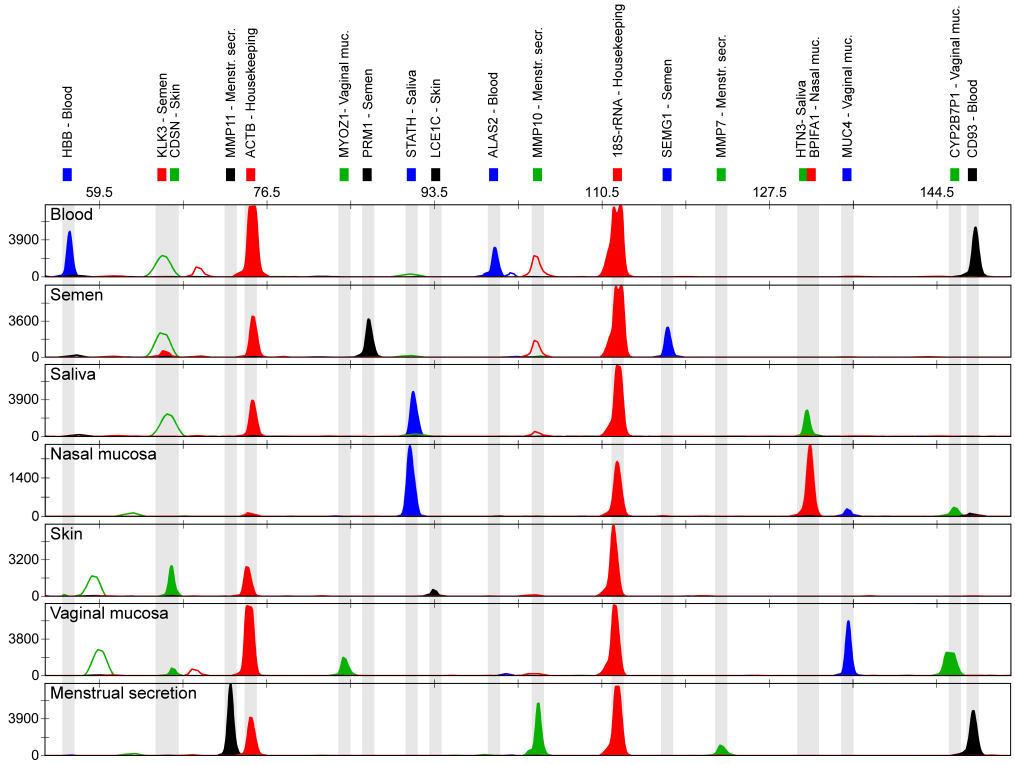
Concluding remarks

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.

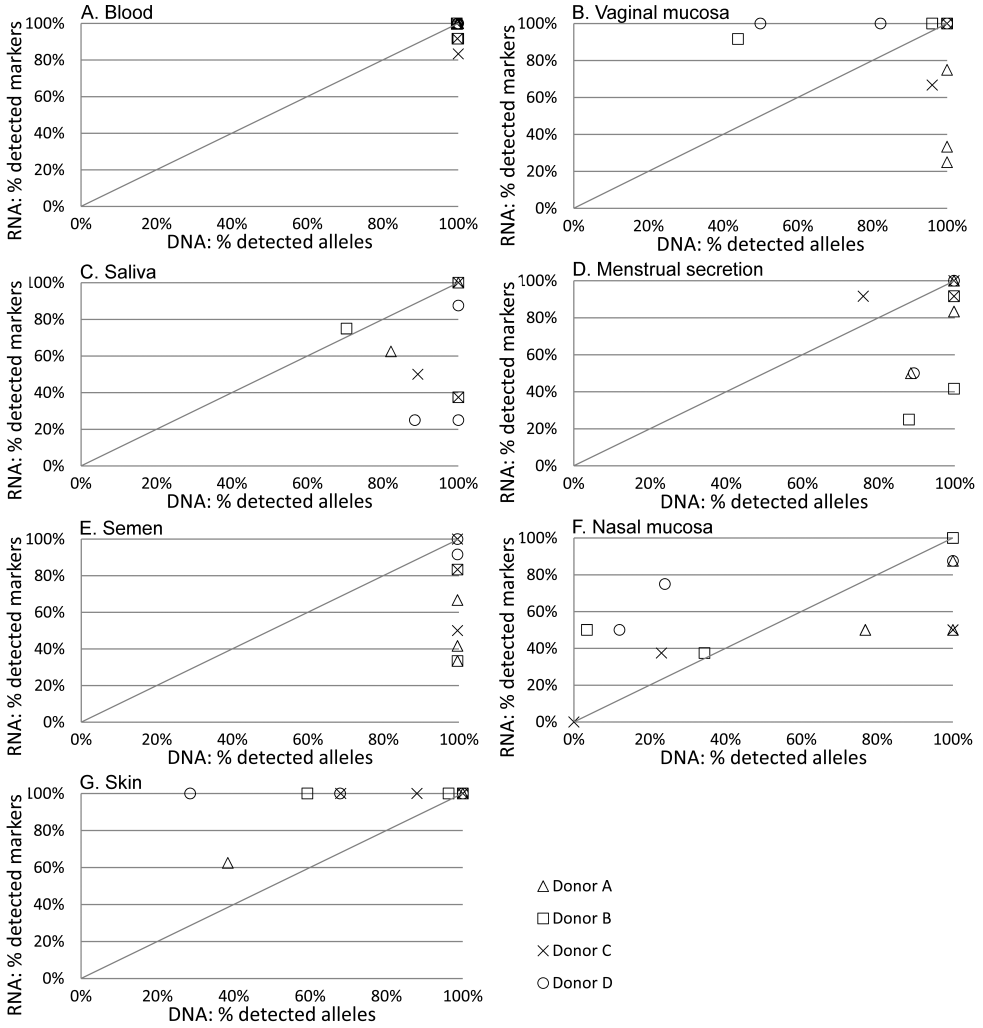
Acknowledgements

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).

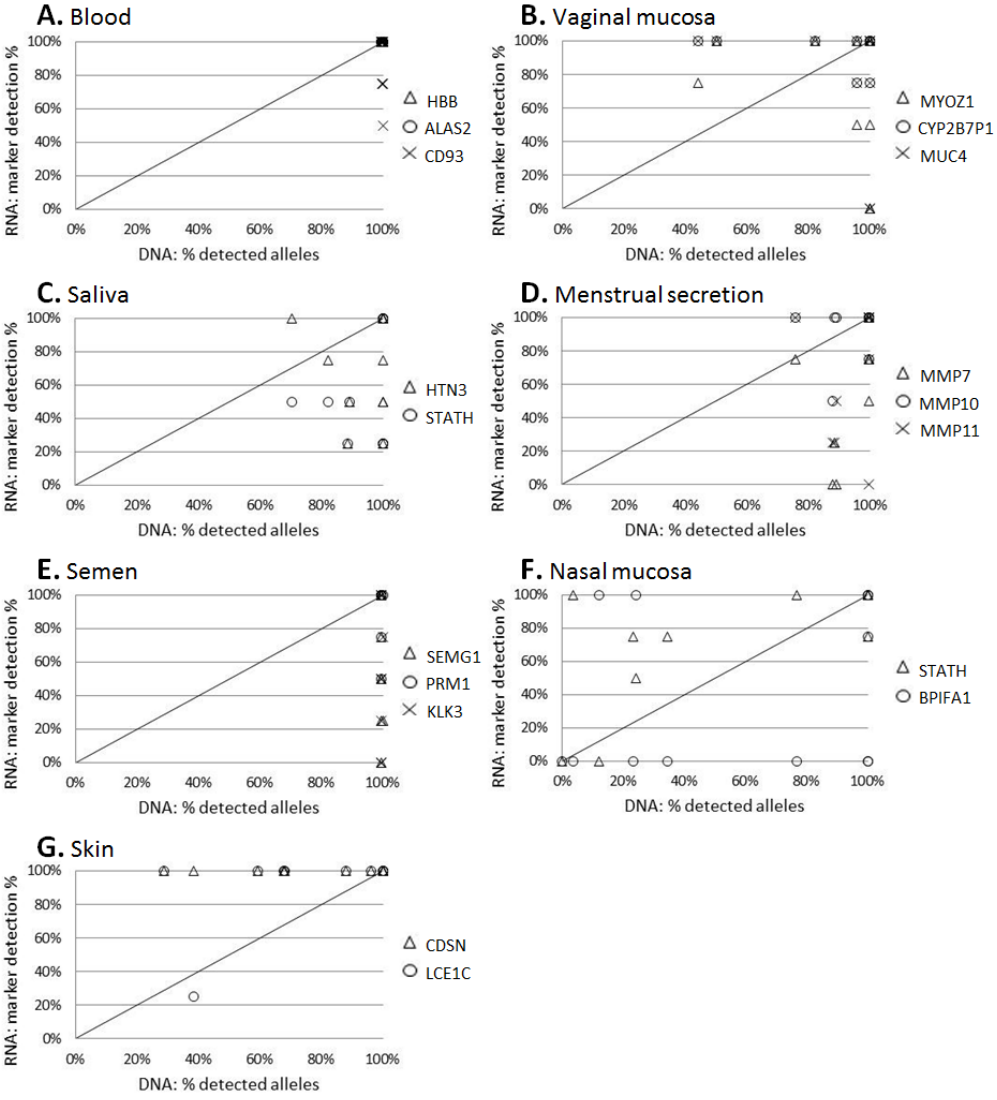
Supplementary material



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.

Chapter 3

Advancing forensic RNA profiling: Preventing noise signals in RNA profiling by adding the multiplex buffer last

Forensic Science International: Genetics Supplement Series 5 (2015) e518-e519

Margreet van den Berge
Titia Sijen

Abstract

We noticed variation in the level of background noise when using custom-made fluorescently labelled primer mixes during PCR amplification. Various aspects that may affect the level of noise generated during PCR set-up were assessed. Best results are obtained when the multiplex buffer is added as a last component.

Introduction

Custom-made fluorescently labelled primer mixes, such as RNA typing multiplexes, are increasingly used in forensic analyses. These multiplexes are regularly used in combination with commercially available multiplex amplification chemicals, for example the QIAGEN multiplex kit. We experienced variation in the occurrence of background noise, and noticed higher noise levels when more samples were analysed. This background noise also observed in blank PCR controls suggested that it is caused during PCR set-up. Various causes and solutions were considered to reduce the level of background signals.

Materials and methods

PCR amplification and product detection was performed according to standardized protocols [1], which uses a master mix of 12.5 µL 2x QIAGEN multiplex reaction mix supplemented with 5 µL primer mix for 19 amplicons based on [2], followed by the addition of the appropriate amount of cDNA sample and/or water that together make a volume of 7.5 µL. PCR products were purified [1] prior to detection using a 3130XL Genetic Analyzer (Life Technologies), POP-4 (Life Technologies) separation matrix and 3 kV, 10 s injection settings. Various aspects were assessed using NTCs (no template controls or blank consisting of nuclease-free water) for PCR amplification. We evaluated the use of QIAGEN multiplex *Plus* buffer, Q-solution (2.5 µL/reaction), increased annealing temperature (from 60 °C to 64 °C) and using a multiplex with reduced complexity (from 19 to 11 labelled primers).

Results and discussion

An overlay electropherogram of a skin sample amplified using the standard amplification protocol is shown in Figure 1A. Visible is the background noise, which severely interferes with the interpretation of amplified peaks (*id est* skin and housekeeping markers) and may lead to false positive scoring of absent markers. We noticed that more noise occurred with larger sample sizes, that it also occurred for the blank control sample that was pipetted last and that it occurred more frequent with less experienced laboratory personal. Thus we hypothesized that the time that buffer and primer mix remain mixed in concentrated form may have a role in the formation of noise signals (note that the four components of the reaction – buffer, primer mix, water and cDNA – are not pipetted per sample but each component

is pipetted for the full sample series). A time series was performed varying the time that multiplex buffer and primer mix stand combined before adding water. Results for the 60,12 and 0 min incubation are shown in Figure 1B. A decrease in background signals is observed with reduced time intervals. Similar time series experiments were repeated to assess the effect when the buffer and primer mix stand combined in less concentrated forms, that is after adding water to the reaction volume of 25 µL. Results are shown in Figure 1C and a clear reduction of background noise was observed compared to the 60 min profile shown in Figure 1B. The use of QIAGEN multiplex *Plus* buffer was compared to the use of the regular QIAGEN multiplex kit (both use a HotStartTaq DNA Polymerase to prevent polymerase activity at room temperature; the *Plus* buffer has a reduced activation time), but this did not reduce the level of background noise (data not shown). Addition-ally, Q-solution was assessed as this is described to improve amplification of difficult template by modifying the melting behaviour of the template. However, the addition of Q-solution led to drop-out of various markers while background signals remained visible (data not shown). Increased annealing temperatures and a reduced complexity of the primer mix also failed to reduce background noise (data not shown). Most promising results are obtained when adding the multiplex buffer as a last component, visible in Figure 1D.

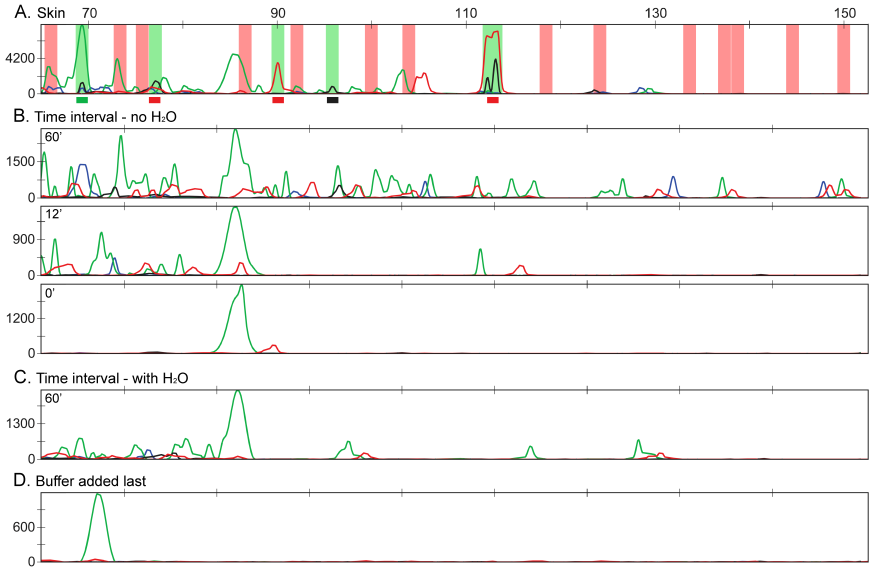


Figure 1. Overlay electropherograms. A: A skin sample amplified showing background noise. Red markers bins indicate locations for which no signals are expected; light green marker bins those of skin and housekeeping markers. The bars below the green bins indicate the fluorophore colour of the expected signal. B: Blank controls amplified after primer mix and multiplex buffer stood combined in concentrated (no water added) for 60, 12 or 0 minutes. C: Blank control amplified after primer mix and multiplex buffer were combined in diluted form (water added) for 60 minutes. D: Blank control amplified when the multiplex buffer was added last, and the PCR was started immediately.

Concluding remarks

We infer that, when using custom-made fluorescently labelled primer mixes, background noise increases when primer mix and multiplex buffer stand combined in concentrated form. By adding the QIAGEN multiplex reaction mix as a last component during PCR set-up of RNA profiling multiplexes, the occurrence of background noise is reduced. Since noise occurs without the presence of template DNA, we assume the noise is caused by destabilisation or aggregation of the fluorescently labelled primers. However, as the composition of the commercial buffer is unknown, we cannot speculate what components may cause this instability. In our laboratory several assays use this multiplex buffer but since these assays do not use fluorescently labelled primers, this background noise is not observed.

Acknowledgements

Sanne Zijlstra, Anuska Matai, Vincent Warnaar and Kristiaan van der Gaag are thanked for assistance. TS and MvdB received financial support from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 285487 (EUROFORGEN-NoE).

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, 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.

Chapter 4

Prevalence of human cell material: DNA and RNA profiling of public and private objects and after activity scenarios

Forensic Science International: Genetics 21 (2016) 81-89

Margreet van den Berge
Gosku Ozcanhan
Sanne Zijlstra
Alexander Lindenbergh
Titia Sijen

Abstract

Especially when minute evidentiary traces are analysed, background cell material unrelated to the crime may contribute to detectable levels in the genetic analyses. To gain understanding on the composition of human cell material residing on surfaces contributing to background traces, we performed DNA and mRNA profiling on samplings of various items. Samples were selected by considering events contributing to cell material deposits in exemplary activities (*e.g.* dragging a person by the trouser ankles), and can be grouped as public objects, private samples, transfer-related samples and washing machine experiments. Results show that high DNA yields do not necessarily relate to an increased number of contributors or to the detection of other cell types than skin. Background cellular material may be found on any type of public or private item. When a major contributor can be deduced in DNA profiles from private items, this can be a different person than the owner of the item. Also when a specific activity is performed and the areas of physical contact are analysed, the "perpetrator" does not necessarily represent the major contributor in the STR profile. Washing machine experiments show that transfer and persistence during laundry is limited for DNA and cell type dependent for RNA. Skin conditions such as the presence of sebum or sweat can promote DNA transfer. Results of this study, which encompasses 549 samples, increase our understanding regarding the prevalence of human cell material in background and activity scenarios.

Introduction

In recent years, the sensitivity of DNA profiling methodologies has increased tremendously and now allows for analysis of minute evidentiary samples previously regarded unsuitable for STR profiling. This increased sensitivity, however, has also led to an increased possibility of profiling DNA that is not related to a crime, but residing in the background. Collecting data regarding the prevalence, transfer, persistence and recovery of human cell material has increasingly gained attention, for instance because of the importance in addressing activity-level questions [1-3]. To name a few studies: human DNA has been analysed in dust, neck and fingernail samplings [1][4-9]; the DNA on hands or touched items was found to often originate from DNA-rich areas like the mouth or nose [10] and the presence of sebaceous glands in skin surfaces [11] or cell-free nucleic acids in sweat was shown to significantly affect transfer of DNA [12]. Others describe the direct and indirect transfer of DNA traces in specific situations or in the washing machine [13-16]. Furthermore, factors influencing the transfer and recovery of DNA have been investigated, like the moisture level of the stain, the type of biological substance, the type of surface, the application of friction and individual shedding state [16-23]. All these studies, however, focussed on inventorying background DNA. Additional knowledge may reside in examining what cell types are present. mRNA profiling can be used for the inference of body fluids as well as organ tissues [1][24-32]. Presumably, skin is the cell type most abundantly present in contact traces, as an individual sheds approximately $5x10^8$ cells each day [33]. However, skin cells do not contain much DNA [34], thus minute amounts of body fluids that have higher DNA contents may affect the genetic analyses.

In this study, DNA and mRNA profiling is performed on samplings of various items, which may contribute to or be a vector for the deposition of background cell material. These items are derived from considering the events contributing to human cell material deposits in an exemplary activity, namely dragging a person at the trouser ankles as depicted in Figure 1. The items can be grouped as everyday public objects and private samples. Besides, transfer and persistence of cellular material in the washing machine is investigated to determine possibilities of cell residues on fabrics after washing. Additionally, the effect of sebum and sweat on the transfer of cellular material is assessed.

Materials and methods

Experimental setup and sample collection

A total of 549 samples were analysed for this background study. Samples are divided

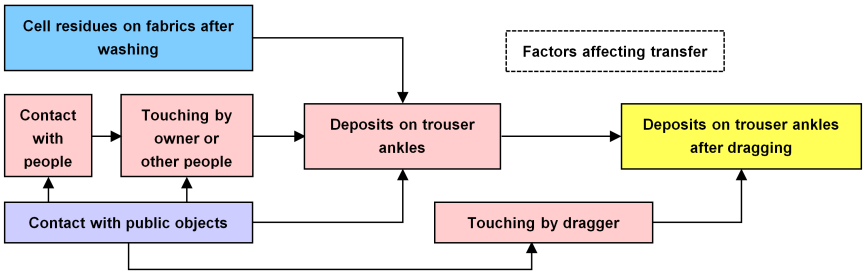


Figure 1. Overview of various factors, which may contribute to background traces found on items prior and post an activity-related scenario (in this example dragging by trouser ankles). In this study, we focussed on analysing public items (purple box), private samples (red), activity-related samples (yellow), factors affecting transfer (white) and washing machine samples (blue).

over four categories, namely public, private, transfer-related and washing machine samples. An overview of the samples and sampling locations is shown in Table 1. All samples except the fingernail samples were collected using tape lifts [35] and areas were stubbed until tapes were saturated. Fingernail samples were collected using a moistened (Nuclease-free water, Ambion) mini-tip swab (MWSCS, MWE Medical Wire). Patches used for the sweat, (non-) sebaceous skin and washing machine experiments were cleared from contaminating nucleic acids by irradiating each side in a CL-1000 UV-CrossLinker at 900 mJ/cm^2 for 60 minutes. The front and back of the cloths/patches were sampled with an individual tape lift.

The first sample set includes 105 public samples as presented in Table 1, set 1. Public items were sampled at areas most accessible to physical contact. Banknotes and coins were collected from wallets of 13 individuals, thus the last person handling the money is known.

The private sample set consists of 164 samples subdivided in samples collected directly from the volunteers' skin (neck, hands, fingernails; 80 samples in total) or clothes (winter gloves, trouser leg, armpits of shirts; 84 samples in total). Sampling locations for the private samples are described in Table 1, set 2.

The third set consists of 168 transfer-related samples: 94 activity-related samples (grabbing and dragging (Table 1, set 3.1)) and 74 samples to assess the effect of skin condition on transfer (Table 1, set 3.2). In the grabbing scenario, a grabber wore his/her personal winter glove on his/her dominant hand and firmly grabbed the bare arm of a "victim" during 5 seconds during which the "victims" were asked to resist the activity. Both the external surface of the grabber's winter glove and the victim's arm were sampled immediately after the activity. For the dragging scenarios, a "victim" who was sitting in a chair was dragged along a corridor at the trouser ankles or armpits of shirts during one minute. Draggers used their bare hands for the activity without

any instructions regarding hand washing. Immediately after the activity, samples were taken from exterior surfaces of the ankle area of the trousers and the armpit area of the shirt. Couples performing these activity-related scenarios were selected to have a low number of shared alleles (three to eight when considering NGM loci) to maximise the information per donor. The same winter gloves, trousers and shirts had also been sampled prior to the activity and represent the samples described in set 2.2 (Table 1). To allow for new "background" to be deposited, a considerable amount of time was left between the prior-activity sampling and the activity namely at least six hours for the shirts and trousers during which they were worn all the time and at least 5 days for the winter gloves during which they were worn approximately half an hour each day without washing. The skin conditions examined for effect on the transfer of DNA are sebaceous and sweaty skin (Table 1, set 3.2). For five volunteers, the forehead and palm of the hand was stubbed to collect sebaceous and non-sebaceous skin, respectively. Volunteers were asked to wash the sampling locations with water and soap, dry with paper tissues and not touch/use these skin areas for 30 minutes. Non-sweaty and sweaty skin samples were prepared by pulling UV-irradiated cloths prior to and after running activities, respectively. Two samples were collected from UV-irradiated cloths prior to pulling to assess for possible remainders of contaminating. For 16 volunteers, post and prior to running the right and left hand were sampled, resulting in 64 samples.

The last sample set comprises 112 samples generated from washing machine experiments (Table 1, set 4). Sampling areas of 5x5 cm were marked with a permanent marker on denim cloths (15 x 21 cm) prior to UV-irradiation. Ten samples were collected from five UV-treated denims prior to washing to assess for possible remainders of contaminating DNA or RNA. For part one of the washing machine experiments, two volunteers combined four patches of denim with their regular laundry; patch 1 and 2 had two marked "blank" areas. Patch 3 and 4 both had four marked areas; a dried bloodstain (500 µL), a dried saliva stain (1000 µL) and two marked "blank" areas. Patch 1 and 2 each resided in a separate washing bag to assess indirect transfer, while patch 3 and 4 were tumbled within the laundry without a washing bag (to study direct transfer). Patch 1 and 3 were air-dried; patch 2 and 4 tumble-dried. Each volunteer performed the experiment twice, resulting in a total of 96 samples. Part two of the washing experiment involved the washing of two "blank" denim patches (alike those used in part one) in individual washing bags, without any laundry. One patch was air-dried; the other tumble-dried. Each volunteer performed the experiment twice, with no other washes in between, giving a total of 16 samples. Both volunteers used a 40 °C wash program and included fabric softener.

All samples used for this study were collected with informed consent of the voluntary donors whose cell material was used.

Table 1. Overview of the samples collected for this study. The number of DNA profiles generated equals the number of samples analysed. RNA analysis was performed in four-fold to allow profile interpretation according to "*x*=*n*/2" guidelines [36].

Set	Item	# for DNA profiling	# for RNA profiling
1	Public objects	105	420
	Escalator rails train station, rubber surface, *n*=6	6	24
	Stair rails train station, painted metal surface, *n*=6	6	24
	Public toilet door handle, smooth metal surface, *n*=6	6	24
	Public toilet flush button, plastic surface, *n*=6	6	24
	Shopping cart handle, plastic surface, *n*=6	6	24
	Shopping basket handle, plastic surface, *n*=6	6	24
	Library books, laminated carbon, *n*=6	6	24
	Coin money, *n*=6	6	24
	Handle bar indoor, smooth metal surface, *n*=6	6	24
	Banknotes, *n*=51	51	204
2	Private samples	164	656
2.1	Skin	80	320
	Neck samples, front and back, *n*=10 each	20	80
	Hands, right and left, *n*=20 each	40	160
	Fingernails (index, middle, ring finger combined), right and left hand, *n*=10 each	20	80
2.2	Clothing	84	336
	Winter gloves dominant hand, fingers and thumb area, *n*=10 each (5 polyester, 3 leather, 2 wool)	20	80
	Trouser, right and left leg, *n*=24 each (jeans only)	48	192
	Armpits of shirt, right and left, *n*=8 each (regular shirts)	16	64
3	Transfer-related samples	168	0
3.1	Activity-related scenarios	94	0
	Winter gloves after grabbing, finger and thumb area, *n*=10 each	20	0
	Arm victim after grabbing by glove, *n*=10	10	0
	Trouser after dragging, right and left, *n*=24 each	48	0
	Armpits of shirt after dragging, right and left, *n*=8 each	16	0
3.2	Factors affecting DNA transfer	74	0
	Sebaceous vs non-sebaceous skin, forehead and palm of hand, *n*=5 each	10	0
	Sweaty vs non sweaty hands, right and left hand, before and after running, *n*=16 each	64	0
4	Washing machine: transfer and persistence	112	448
4.1	Cloths combined with regular washes, air-dried and tumble-dried, *n*=48 each	96	384
	Areas with blood or saliva stain, *n*=16 each	32	128
	Blank areas on spotted or blank cloths, *n*=32 each	64	256
4.2	Blank cloths, blank washes, air-dried and tumble-dried, *n*=8 each	16	64

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

DNA/RNA co-extraction, DNase treatment, DNA quantification, ethanol precipitation and reverse transcription were performed as described in Lindenbergh et al. [24]. Tape lifts were processed entirely and when two tape lifts were derived for a sample, these were extracted together. 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 Ref. [25]. DNA extracts were ethanol-precipitated [24] when the DNA concentration of a sample was below 50 pg/µL. An amount of DNA extract was taken so that approximately 500 pg DNA resided in the ethanol precipitate. For all transfer-related samples (set 3.1 and 3.2, Table 1), only the DNA fraction was processed after co-isolation.

RNA profiling

Cell type inference was performed using in-house developed multiplexes [24-25] [37]. The multiplexes described in Ref. [24] and Ref. [25] were used for the public objects, neck, winter gloves, armpits of shirt and some trouser ankle samples (Table 1) and allow for the inference of blood, saliva, vaginal mucosa, menstrual secretion, semen and skin; the multiplex described in Ref. 37 was used for hands, fingernails, most trouser and washing machine samples (Table 1) and targets besides the aforementioned body fluids nasal mucosa. Amplification and product detection was performed according to standardized protocols [24]. A serial cDNA input (*e.g.* 0.2, 1, 4 µL cDNA) PCR was used to determine the input providing an informative RNA profile. Supplementary PCRs were performed to obtain four informative PCR replicates per sample. When RNA extracts were ethanol-precipitated prior to reverse transcription, the 20 µL cDNA was divided over four replicate PCRs with 5 µL input. PCR products were purified [24] prior to detection using a 3130XL Genetic Analyzer (Life Technologies). Amplification products were analysed using POP-7 or POP-4 (Life Technologies) separation matrix using 3 kV, 10 s injection settings. Profile analysis was performed using Genemapper ID-X version 1.1.1 (Life Technologies) with a detection threshold of 150 relative fluorescence units (rfu).

DNA profiling

DNA profiles were generated using the AmpFℓSTR® NGM™ PCR Amplification Kit (NGM) (Life Technologies) using a maximum of 500 pg DNA. PCR products were separated according to standardized protocols [24] using a 3130XL Genetic Analyzer (Life Technologies) with POP-4 (Life Technologies) separation matrix using 3 kV, 15 s injection settings. Profile analysis was performed using Genemapper ID-X version 1.1.1 (Life Technologies) and a detection threshold of 50 rfus.

Data analyses

Donors with known genotypes were used for all samples (except the public items), so that the percentage of total rfus in the STR profile for each donor or for unknown contributors could be determined. In case of multiple known donors (*e.g.* activity-related samples), shared alleles were considered as a separate class. Profile

interpretation was performed using the 15 STR loci in the NGM kit (Amelogenin was not included). Maximum allele count [38] was used to determine the minimum number of contributors in a profile. All samples collected after dragging or grabbing experiments were analysed using the LoCIM-tool [39] to deduce the alleles of the most prominent component in the STR profile and determine whether the "perpetrator" or the "victim" represents the major contributor in the DNA profile.

RNA data interpretation was performed according to the "$x=n/2$" rule as described in Ref. [36]. 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 detected ($x \geq n/2$), scored "sporadically observed" when less than half of the possible peaks are detected ($0 < x < n/2$) and scored "not observed" when no peaks are detected ($x=0$). Cell types that 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.

Results

Public objects

Public items such as shopping basket handles, stair rails in train stations or money are regularly touched by or localize near multiple people. To gain understanding of the composition of human cell material residing on such surfaces, we investigated a set 105 public items (Table 1).

Details on DNA and RNA profiling results of the analysed public items are shown in Figure 2. DNA yields range from 0.0 to 41.1 ng, the minimum number of contributors lies between 0 to 6 donors and the number of observed cell types between 0 and 3. High DNA yields do not necessarily relate to increased numbers of cell types or contributors. Reliable quantification results (above 0.5 pg/μL) were obtained for 91% of the samples. In 17% of the samples a major contributor could be deduced in the STR profile. For the banknotes (n=51) and coin money (n=6) the last user was known, but this last user was not necessarily the major contributor in the DNA profile: for the 9 samples for which a major was deduced, 5 times this represented the last user. Skin was scored "observed" in 96% of the samples; no cell type was scored "observed" in the remaining 4% of samples. Skin was occasionally (78%) scored "observed" for samples with DNA quantification results below the detection threshold of the quantification system (0.5 pg/μL). "Observed" scoring of a cell type other than skin occurred in 35% of the samples and mainly involved saliva and occasionally vaginal mucosa, menstrual secretion or semen (results not shown).

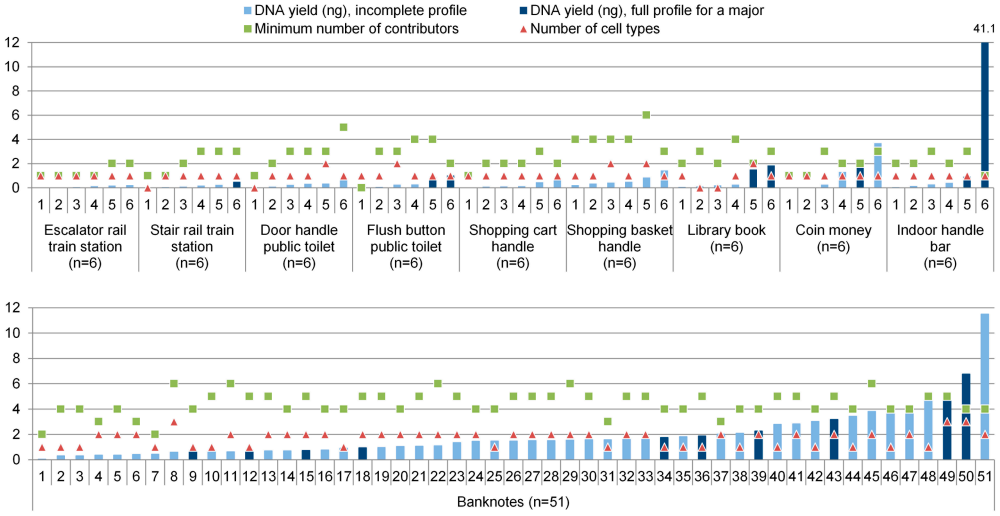


Figure 2. DNA profiling and mRNA cell typing results of 105 public items. Blue bars indicate the DNA yield (ng) for each sample. Light blue bars indicate the samples that produced incomplete STR profiles; dark blue bars represent samples for which a single source full DNA profile or a clear major contributor was observed. Additionally, the minimum number of contributors (green dot) and the number of observed cell types (red dot) are shown. When the DNA yield exceeds 12 ng (the maximum value at the y-axis), the yield is indicated above the bar.

Private samples

With private samples, both cell material of the owner and other people may be present due to everyday activities. We investigated the composition of 164 private samples (Table 1). The first subset (set 2.1, Table 1) comprised samples taken directly from various skin areas for different reasons, namely 1) neck samples for information on background in cases of manual strangulation; 2) hand samples as hands make contact with public object, private items and skin areas (own and non-own) and 3) fingernail dirt samples to gain knowledge on background in *e.g.* sexual assault or murder cases. Additionally, private objects such as clothing can be involved in criminal acts: gloves can be worn during manual strangulations and a victims' clothing can be touched when a perpetrator relocates the body and such items reside in subset 2.2 (Table 1).

An overview of DNA and RNA profiling results of the analysed private items is shown in Table 2 and details per individual sample are provided in Figures 3A-F. DNA yields vary more between individuals and less for samples from the same person (*e.g.* right and left hand for one individual tend to give similar yields except for few instances). No obvious differences for donors of different gender or varying fabric types were seen (data not shown). The estimated minimum number of contributors ranged up to five; most often at least two (55% of skin and 42% of clothing samples)

Table 2. Summary of DNA and RNA profiling results after analysing 164 private samples.

		Skin			Clothing		
		Neck	Hands	Fingernails	Winter gloves	Trouser ankles	Armpits of shirts
		Front + Back	R+L[a]	3 fingers both hands	Fingers + thumb dominant hand	R+L	R+L
		(*n*=20)	(*n*=40)	(*n*=20)	(*n*=20),	(*n*=48)	(*n*=16)
Results		Fig. 3A	Fig. 3B	Fig. 3C	Fig. 3D	Fig. 3E	Fig. 3F
DNA	Avg. DNA yield (ng)	0.4	5.9	6.6	0.7	3.2	0.2
	Range DNA yield (ng)	0.0-5.0	0.0-67.5	0.3-58.6	0.0-2.5	0.1-29.0	0.0-0.6
	Number of contributors	1-3	1-4	1-3	1-4	2-5	1-3
	Full profile for sampled donor	67%	88%	25%	80%	83%	56%
	Sampled donor is major contributor	100%	93%	100%	100%	60%	100%
	Samples with background alleles	83%	88%	80%	90%	100%	100%
	Avg. % non-owner rfus	10	9	6	13	20	12
	Range % non-owner rfus	0-40	0-66	0-20	0-26	0-62	0-27
	Avg. PH[a] ratio donor/background	7.6	8.8	6.0	4.2	4.4	5.3
	Range PH ratio donor/background	1.2-19.1	0.5-21.8	1.7-16.2	1.9-8.3	0.8-19.2	1.2-21.5
RNA	Skin "observed"	100%	100%	100%	95%	100%	100%
	No cell type "observed"	0%	0%	0%	5%	0%	0%
	Extra cell type "observed" next to skin	17%	18%	5%	15%	8%	19%
	DNA yield (ng) samples extra cell type	0.2-5.0	0.2-67.5	58.6	0.4-2.0	2.3-12.9	0.1-0.3

[a] R, right; L, left; PH, peak height,

or three (42% of skin and 40% of clothing samples) contributors appear present. Non-owner alleles are observed in the majority of samples, but generally the donor alleles are substantially higher than non-owner alleles. Hence, for most samples a major contributor could be deduced that corresponded to the owner except for some hand and trouser samples for which no major contributor was deduced (Figure 3B+E). For some of these samples the average peak height of donor alleles is higher than non-donor alleles (peak height ratio above 1, Figure 3B+E), but the presence of few rather high background alleles prevents deducing the donor as major contributor in the STR profile. RNA profiling resulted in an "observed" scoring for skin in all but one of the samples. For this sample no cell types were scored "observed". Other cell types than skin are occasionally (14% of the samplings) scored "observed" and these mainly involve blood, saliva or vaginal mucosa. Nasal mucosa is scored "observed" in few samples, which are the samples resulting in a relatively high DNA yield (at least 58.6 ng) [37]. "Sporadic" scorings, which we regard unreliable as described in Ref. [36] were detected in 70% of the samples.

Transfer-related samples

Activity-related scenarios

In crimes with physical contact, DNA can be transferred between the contact areas of "perpetrator" and "victim". We investigated the composition of 94 samples (Table 1) obtained from three different activities that exemplify activities in criminal acts, namely 1) grabbing during which the "perpetrator" wore a winter glove and grabbed the bare

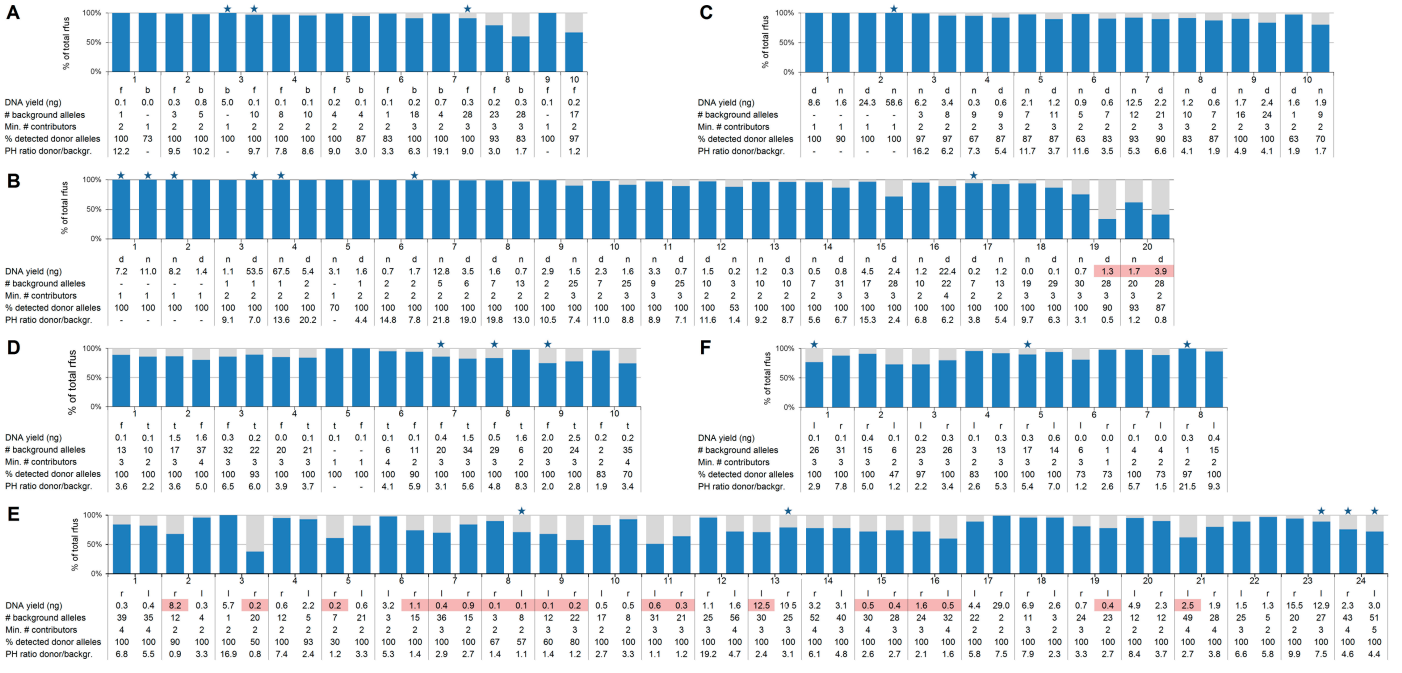
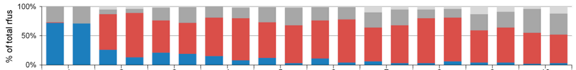


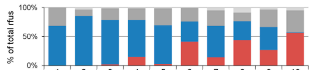
Figure 3. Detailed information for the 164 analysed private samples: A) 20 samples from the front (f) and the back (b) of the neck of ten volunteers. For two samples (9 and 10) results were obtained only for samples collected from the front side of the neck; B) 40 samples from the dominant (d) and non-dominant (n) hand of 20 volunteers; C) 20 fingernail samples from the dominant (d) and non-dominant (n) hand of ten volunteers; D) 20 samples from the finger (f) and thumb (t) area of the winter glove worn on the dominant hand of ten volunteers; E) 48 samples from the right (r) and left (l) trouser leg ankles of 24 volunteers; F) 16 samples from the right (r) and left (l) armpits of shirts of eight volunteers. Indicated per sample are the percentages of total rfus in the STR profile belonging to the donor (blue bars) and background (light grey bars). A blue star above the bar indicates samples for which next to skin additional cell types were scored "observed". Red markings in the DNA yield section indicate samples for which the major contributor deduced from in the STR profile was not the owner.

A



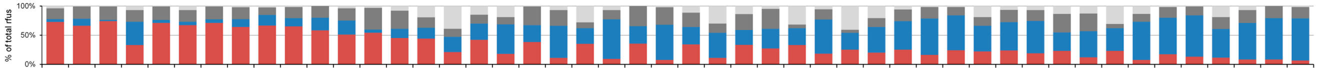
	1 f	1 t	2 t	2 f	3 f	3 t	4 f	4 t	5 t	5 f	6 t	6 f	7 f	7 t	8 f	8 t	9 f	9 t	10 f	10 t
DNA yield (ng)	5.4	125.3	2.8	9.5	1.5	1.2	1.3	1.3	0.5	0.6	1.1	3.1	3.8	6.7	2.2	1.5	6.1	7.7	1.8	0.8
# background alleles	1	-	18	16	5	1	-	13	2	6	4	3	25	16	7	14	10	10	10	10
Min. # contributors	2	1	3	3	3	2	2	3	3	2	3	3	4	3	2	3	2	2	3	2
% detected alleles victim	100	100	100	100	100	57	96	88	77	32	77	58	71	38	35	60	22	17	29	19
% detected alleles grabber	35	0	100	100	100	95	100	100	100	100	100	100	100	100	100	100	100	100	100	95
PH ratio victim/backgr.	6.9	-	3.8	2.0	1.9	1.3	-	1.9	2.2	1.4	1.9	1.2	1.0	1.4	0.4	1.4	0.7	1.2	0.9	0.7
PH ratio grabber/backgr.	0.2	-	8.9	11.1	5.2	2.2	-	16.3	3.3	9.3	8.4	12.0	6.8	10.1	4.2	11.9	2.0	3.3	7.4	2.2

B



	1	2	3	4	5	6	7	8	9	10
DNA yield (ng)	13.4	-	0.0	-	0.6	-	1.1	0.3	-	-
# background alleles	-	2	7	6	2	4	15	16	2	4
Min. # contributors	2	2	2	3	2	3	3	3	2	2
% detected alleles victim	100	100	100	100	100	100	100	100	78	0.05
% detected alleles grabber	5	0	50	100	19	100	100	100	71	95
PH ratio victim/backgr.	-	1.9	19.3	13.0	4.0	7.7	8.0	2.3	1.3	0.8
PH ratio grabber/backgr.	-	-	1.0	3.1	0.9	9.0	2.1	3.5	1.1	2.5

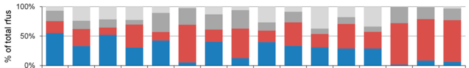
C



	1 r	1 l	2 r	2 l	3 l	3 r	4 r	4 l	5 r	5 l	6 l	6 r	7 l	7 r	8 r	8 l	9 l	9 r	10 l	10 r	11 l	11 r	12 r	12 l
DNA yield (ng)	1.8	3.4	8.7	0.9	0.5	1.1	9.6	5.4	0.4	0.9	3.5	1.9	2.8	1.4	0.3	0.2	0.4	0.1	2.1	0.6	1.3	2.0	17.0	3.2
# background alleles	12	7	4	25	3	15	2	5	3	3	-	7	5	13	19	18	22	11	2	20	15	16	1	8
Min. # contributors	3	3	3	3	2	3	2	2	2	2	2	3	2	3	3	3	3	3	3	3	2	3	2	4
% detected alleles victim	83	96	50	100	53	63	70	93	63	63	100	100	60	87	67	60	97	83	100	97	67	100	100	100
% detected alleles grabber	100	100	100	100	100	100	100	100	93	100	100	100	100	100	100	63	97	47	97	83	80	80	100	84
PH ratio victim/backgr.	0.7	2.7	0.7	6.2	1.3	1.1	0.7	1.3	1.5	1.2	-	1.7	0.8	1.4	1.4	0.7	1.6	1.5	3.1	7.3	1.0	6.2	4.6	10.2
PH ratio grabber/backgr.	9.6	14.7	12.5	5.3	8.1	6.7	5.8	6.3	3.3	3.2	-	3.5	3.7	3.2	1.9	0.8	2.6	1.3	4.0	1.7	1.0	1.0	6.4	1.7

	13 l	13 r	14 r	14 l	15 l	15 r	16 r	16 l	17 l	17 r	18 r	18 l	19 r	19 l	20 l	20 r	21 l	21 r	22 l	22 r	23 r	23 l	24 r	24 l
DNA yield (ng)	10.0	6.2	5.7	3.6	0.8	1.1	6.6	3.6	2.0	2.5	2.0	2.0	0.6	0.6	0.7	0.7	1.0	3.1	3.9	5.6	3.2	5.5	19.8	3.1
# background alleles	20	35	21	14	26	22	43	44	8	2	5	21	10	10	19	21	18	33	9	2	25	15	2	10
Min. # contributors	3	3	3	3	4	3	4	5	3	2	3	3	3	3	3	3	4	4	3	2	4	3	2	3
% detected alleles victim	100	100	100	100	100	100	100	100	100	93	100	97	100	100	97	100	92	100	100	100	100	100	100	100
% detected alleles grabber	100	87	100	100	96	100	86	90	93	63	97	90	93	93	97	83	88	67	97	97	83	65	92	75
PH ratio victim/backgr.	2.3	2.2	2.0	5.0	1.0	9.6	1.3	4.0	2.6	2.9	5.3	2.0	4.1	3.2	2.1	3.4	1.0	6.7	10.2	10.8	3.1	6.6	29.8	16.4
PH ratio grabber/backgr.	2.7	0.7	2.3	3.5	1.2	2.9	1.5	2.1	1.4	1.2	2.2	1.1	1.8	1.6	1.6	1.1	0.6	1.1	2.9	2.0	0.8	1.1	3.7	2.0

D



	1 l	1 r	2 r	2 l	3 l	3 r	4 l	4 r	5 r	5 l	6 l	6 r	7 r	7 l	8 r	8 l
DNA yield (ng)	0.8	1.8	0.3	0.8	2.2	1.5	0.2	1.4	0.2	0.2	1.4	0.3	0.5	2.9	1.4	3.2
# background alleles	22	14	31	18	31	11	25	24	28	16	43	24	40	2	3	14
Min. # contributors	4	3	4	4	5	3	3	4	4	3	5	4	4	2	3	3
% detected alleles victim	100	100	100	88	100	95	100	95	100	100	96	100	91	22	83	67
% detected alleles grabber	100	92	75	88	100	100	100	100	91	100	92	100	100	100	100	100
PH ratio victim/backgr.	7.2	0.8	3.1	1.2	6.4	1.4	3.5	2.5	2.0	3.0	1.6	1.6	1.6	1.4	1.9	1.7
PH ratio grabber/backgr.	2.6	0.8	1.0	1.5	1.9	13.6	2.1	10.9	1.0	3.3	1.2	2.3	1.6	13.6	12.9	13.1

Figure 4. Detailed information for the of 94 analysed activity-related samples: A) 20 samples from the finger (f) and thumb (t) areas of winter gloves worn on the dominant hand of ten "perpetrators" who grabbed the bare arm of a "victim"; B) samples from grabbed arms of ten "victims" after the grabbing activity as described in A; C) 48 samples from the right (r) and left (l) trouser leg ankles of 24 volunteers collected after a dragging activity; D) 16 samples from the right (r) and left (l) armpits of shirts of eight volunteers collected after a dragging activity. Indicated per sample are the percentages of total rfus in the STR profile belonging to the "victim" (blue bars), "perpetrator" (red bars), shared alleles (dark grey bars) and background (light grey bars). Markings in the DNA yield section indicate samples for which the "victim" (blue) or "perpetrator" (red) was deduced as major contributor in the STR profile, absence of colour indicates samples for which no major could be deduced.

arm of the "victim"; 2) dragging of a "victim" by a "perpetrator" at the trousers ankles; 3) dragging of a "victim" by a "perpetrator" at armpits of shirts. The winter gloves, trouser ankles and armpits of shirts that were sampled after these activities represent the same items analysed without an activity in section "Private samples" (this sampling had occurred at least six hours earlier).

An overview of DNA and RNA profiling results is shown in Table 3 and details per individual sample are provided in Figures 4A-D. Alike the public items (section "Private samples"), large variations in DNA yields and number of contributors are observed. For most samples, an increase in DNA yield is observed when comparing yields prior (section "Private samples") and post activity (Figure 3D-F *versus* Figure 4A,C-D).

Alleles corresponding to both the "grabber" and the "victim" were detected in the majority of the samples. The "perpetrator" was not necessarily the major contributor in the STR profile (Figure 4, Table 3). When no major contributors could be deduced (34% of all samples), this was mainly because of approximately equal peak heights for alleles corresponding to the "victim" and the "perpetrator" and not because of high non-owner signals. For most samples the donor alleles could be distinguished from non-owner alleles as these had generally relatively low peak heights (Table 3).

Table 3. DNA profiling results of 94 activity-related samples.

	Results	Winter gloves Fingers + thumb dominant hand (*n*=20) Fig. 4A	Grabbed arm (*n*=10) Fig. 4B	Trousers ankle R+L[a] (*n*=48) Fig. 4C	Armpits of shirts R+L (*n*=16) Fig. 4D
DNA	Avg. DNA yield (ng)	9.2	3.1	3.4	1.2
	Range DNA yield (ng)	0.5 - 125.3	0.0 - 13.4	0.1 - 19.8	0.2 - 3.2
	Higher DNA yield after activity (compared to Table 2)	95%	na[a]	71%	81%
	Number of contributors	1 - 4	2 - 3	2 - 5	2 - 5
	Samples with background signals	90%	90%	98%	100%
	Average % non-owner rfus	4	3	10	15
	Range % non-owner rfus	0-13	0-9	0-41	0-37
	Avg. % detected alleles grabber	91	64	91	96
	Range % detected alleles grabber	0-100	0-100	47-100	75-100
	Avg. % detected alleles victim	64	88	90	90
	Range % detected alleles victim	17-100	0-100	50-100	22-100
	Full grabber profiles	80%	40%	40%	69%
	Full victim profiles	25%	80%	56%	50%
	Grabber is major contributor in STR profile	90%	10%	21%	31%
	Victim is major contributor in STR profile	10%	60%	25%	19%
	Samples lacking a major contributor	0%	30%	54%	50%
	Avg. PH[a] ratio grabber/background	25.4	2.9	3.2	5.2
	Range PH ratio grabber/background	0.2-240.4	0.9-9	0.6-14.7	0.8-13.6
	Avg. PH ratio victim/background	5.6	6.5	4.2	2.6
	Range PH ratio victim/background	0.4-60.2	0.8-19.3	0.7-29.8	0.8-7.2

[a] R, right; L, left; na, not applicable; PH, peak height

Factors affecting DNA transfer

Sebaceous fluid secreted by the sebaceous glands is described to contribute to the amount of DNA available for transfer [11]. To confirm and extend these findings, we assessed the effect of the presence of sebum or sweat on skin surfaces on DNA transfer. In total 74 samples were analysed (Table 1). In addition, two samples were taken from UV-irradiated cloths, which confirmed removal of DNA.

An overview of DNA results is shown in Figure 5A-B. Large variations in DNA yields are observed for these various types of skin samples: yields ranged between 0.1 to 2.4 ng for non-sebaceous samples and between 0.7 to 19.9 ng for sebaceous skin samples. Yields for non-sweaty skin samples ranged between 0.1 and 10.8 ng and those for sweaty skin samples between 0.1 and 191.8 ng. When comparing the yields for the two skin types collected per donor, the sebaceous skin samples always resulted in higher DNA yields (2.6 to 186 times higher) than the non-sebaceous skin samples. The sweaty hands had higher DNA yields in 72% of the comparisons (up to 153.6 times higher).

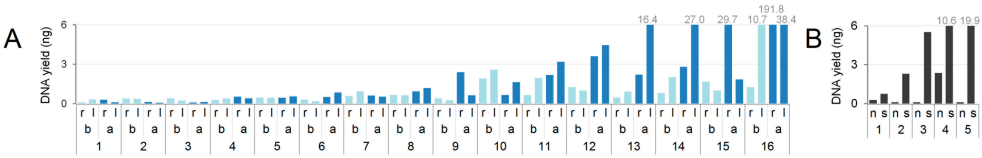


Figure 5. DNA yields (ng) for the various skin samples analysed to assess the effect of the presence of sweat or sebum on DNA transfer: A) 64 sweat and non-sweat samples of the right (r) and left (l) hands of 16 individuals before (b) and after (a) running activities; B) non-sebaceous (n) and sebaceous (s) skin samples of five individuals. Light blue bars indicate DNA yields for samples taken before the activity, dark blue the samples taken after the activity. For samples with a DNA yield above 6 ng (the maximum of the y-axis), the yield is indicated in grey above the bar.

Washing machine: transfer and persistence

Washing machine experiments were performed to assess the transfer and persistence of DNA and RNA during laundry. The denim patches were derived from worn jeans and treated with UV irradiation. Although no DNA resided on these patches, skin mRNA markers were detected in some of the patches (Table 4, analysis of ten samples). It was therefore decided to regard the skin marker results with caution. The experiment had two parts. In part one "blank" patches and patches with controlled inputs of two body fluids were combined with regular laundry to assess indirect (to the "blank" patches) and direct (within a patch on which body fluid stains are deposited) transfer of DNA and RNA, and persistence at the stain location. An overview of the DNA and RNA profiling results is shown in Table 4.

Table 4. Details on DNA and RNA profiling results after analysing 112 washing machine samples. Average yields were determined including samples with concentrations below the detection threshold of the quantification system.

					Cell types "observed"		
Part			Average recovered DNA yield (ng)	Persistence rate (%)[a]	Blood	Saliva	Skin
0	UV-treated patches	Blank areas (*n*=10)	0.01 (0.00-0.01)	na[b]	0%	0%	60%
1		Blood stain (*n*=16)	0.11 (0.03-0.42)	0.001 (0.000-0.003)[c]	0%	0%	94%
	Spotted patches	Saliva stain (*n*=16)	0.04 (0.02-0.09)	0.001 (0.000-0.002)[d]	0%	0%	81%
		Blank areas (*n*=32)	0.05 (0.02-0.17)	na	3%	0%	97%
	Blank patches	Blank areas (*n*=32)	0.07 (0.03-0.22)	na	6%	0%	91%
2	Wash 1	Blank areas (*n*=8)	0.00 (0.00-0.00)	na	0%	0%	63%
	Wash 2	Blank areas (*n*=8)	0.00 (0.00-0.00)	na	0%	0%	38%

[a]Determined by comparing the DNA yield recovered at the stain area after washing to the amount of DNA spotted on the patch

[b]Not applicable

[c]A blood stain is inferred to contain 13,339 ng DNA

[d]A saliva stain is inferred to contain 5,550 ng DNA

Persistence rates of blood and saliva on the stained denims were determined by comparing the amount of DNA spotted on the patch (derived from determining the DNA yield in an aliquot of liquid body fluid *i.e.* 10 µL blood or 20 µL saliva) to the DNA yield recovered at the stain area after washing. Average persistence rates of less than 0.001% were determined for both body fluids (Table 4). Since DNA from the patches is recovered by tape lifting, these persistence rates will be an underrepresentation. DNA yields were low for all samplings and ethanol precipitation of the full DNA extract was required to generate profiles with information. The DNA profiles derived for the body fluid stain areas contain on average 19% of the non-shared alleles of the blood donor and 3% of the saliva donor at the respective spot locations. Since in addition some alleles not corresponding to the blood or saliva donor were detected, we cannot be fully sure that all matching alleles derive from the spotted body fluids. Direct transfer of blood and saliva was assessed by sampling "blank" areas on the same cloth. Incomplete profiles with a maximum of 9 alleles were obtained and the origin of the alleles (blood, saliva, washing volunteer or other) was unclear. Indirect transfer, referring to transfer to patches that had not been in contact with the body fluid stains as these patches resided inside a washing bag, resulted in profiles with only very few alleles of unclear origin.

Both the locations where body fluids were spotted and the "blank" areas, were assessed using mRNA profiling. Blood nor saliva was scored "observed" on the stain locations. Blood was however scored "observed" in three of the analysed blank areas that resided either on a spotted or a blank cloth (Table 4). Saliva was not scored "observed" in any of the sampled areas. Skin mRNA was observed in the majority of samples (Table 4), but the origin of these skin signals is unclear, as skin mRNA was not fully removed upon UV irradiation of the patches. Possibly skin signals are added from co-washed laundry or remnants in washing machine, as skin is scored "observed" in

a higher percentage of the patches (Table 4). In the second part of the experiment we assessed whether skin mRNA is removed upon washing by performing multiple and subsequent "blank patch washes", during which blank patches were washed without the presence of any other laundry. Wash 1 followed a normal laundry load and wash 2 immediately followed this wash 1. Results are shown in Table 4 and again very low DNA yields and only empty or STR profiles with few alleles were obtained. Skin remained detectable using mRNA profiling, although the percentage of positive patches is lower than when regular laundry is present as well (Table 4). For both parts of the experiment the effect of tumble- or air-drying was additionally studied, however, no clear differences were observed when comparing results of air- or tumble-dried samples (data not shown).

Discussion and conclusion

In this study, we assessed the amount and composition of human cell material prevailing on various items. To maximise profiling information, we concentrated DNA and RNA extracts prior to profiling when DNA yields were low. By this approach we observed background cellular material that predominantly gives skin cell typing signals on any type of public or private item although large variations for the DNA yield and number of contributors were seen. Various aspects can have a role such as the function of an item that determines how often, how long and with what intensity an item is touched, the type of substrate from which an item is made, the type of biological substance that was deposited by users, whether or not friction is applied and the individual shedder status of persons touching the items [11-12][16-23]. Results show that high DNA yields do not necessarily relate to an increased number of contributors, or the detection of other cell types than skin. The observation that the detection of skin-specific mRNAs is relatively sensitive when comparing to other cell types or skin-derived DNA [37] and our observation that skin mRNA signals are robust and can even be detected after UV-irradiation, may have a role here as it is possible that a body fluid was deposited of which the DNA is detected and the RNA not. Some body fluids such as nasal mucosa [37] contain more nucleated cells than skin [34]. Results of this study support this, as mainly high DNA yields are obtained for samples with nasal mucosa scored "observed". The study involved 549 samples that were collected and analysed during a substantial time period. Therefore, three versions of our cell typing multiplex were applied in which the two earliest versions did not carry a marker specific for nasal mucosa. With these multiplexes we occasionally detect vaginal mucosa and menstrual secretion mRNA markers that may originate from the presence of nasal mucosa, as the markers for these body fluids can be co-expressed in nasal mucosa to variable extent [37].

STR profiles were generated to obtain information on whether a major contributor is present and whether this matches the owner in case of a private item. With public objects, 17% of the samples showed a clear major contributor but no relation could be made with type of item or DNA yield. With private objects, 88% of the samples showed a major contributor; 99% of these match the owner (consistent to transfer studied described in Ref. [10]), 1% not (these were all hand samples) and we did not attempt to identify this person and determine the relation to the owner of the private item. Irrespective of the presence of a major contributor, we observed non-owner alleles in 90% of the private samples, confirming the concept that private items carry non-self cell material. Consistent with results described by van Oorschot et al. [40] we found that the DNA yields of samplings vary mainly between individuals and less within samples from the same person. Regarding cell type composition, only few samples contained an additional cell type besides skin and no trends can be derived whether this observation occurs more within or between persons, or for a specific gender.

For the majority of samples collected after activity-related scenarios at least partial DNA profiles for both the victim and perpetrator could be retrieved. In only a portion of the samples a major contributor is deduced and approximately evenly frequent this is the "victim" or the "perpetrator" (29% and 38% of the samples, respectively). Often, no clear major could be deduced (34% of the samples). Since the activities were performed in a controlled setting and because we generally see a similar result for the right and the left sample, we infer that specific aspects of the couples engaged in the activity have a role, such as the type of fabric, number of days that clothing had been worn, shedding features of individuals or personal habits [10][19][21][23][41]. We confirmed [11] that the presence of sebum promotes DNA transfer and habits like regularly touching face or scalp can give an individual a higher DNA transfer rate. We assessed whether the presence of sweat could have a similar effect as sebum, but noticed a less strong effect. As sweat itself contains only limited amounts of DNA [37], moisture may provoke increased DNA yields by releasing more skin cells.

We also studied transfer and persistence during laundry. Generally, low persistence rates are obtained for the saliva and bloodstains that were used, which is consistent with earlier studies describing difficulties in producing full STR profiles of saliva and to a lesser extend blood stains after transfer (not necessarily during laundry) [13][17][23]. The moisture level of a stain may influence the transferability [18][42] and the rehydration of the stains during laundry may increase transfer although one would assume that the washing detergent deteriorates the cell material. No conclusions can be drawn from the detection of skin mRNA after washing, as these markers remained detectable after UV-irradiation of the patches, which indicates a remarkable stability of the skin mRNA markers [37]. Even though only limited cell types and sample volumes were assessed, the overall view is that the transfer and persistence of DNA in a washing machine is very limited, while the persistence of RNA appears cell type dependent.

Even though many other aspects regarding the transfer, persistence and recovery of human cell material remain to be evaluated, results of this study increase insight in the prevalence of human cell material on various sample types and in activity-related scenarios. This study may aid reporting officers in evaluating the strength of evidence at the activity level for instance when using a Bayesian network approach [43-46]. In such an approach, the probabilistic relationships between variables are assessed through a graphical model. These probabilities can be based on expert opinion or – preferably- derived from experimental data such as those presented in this study. For instance when considering a dragging scenario, a probability of 0.21 may be given to the event that the perpetrator occurs as a major in the DNA profile and one of 0.25 that the victim occurs as major when considering the data presented in Figure 4C (in which in ten of the 48 DNA profiles the perpetrator occurs as major and 12 times the victim). Inevitably, experimental designs will not be the exact same as the hypothesized crime scene conditions but information on the occurrences of a certain event and the extent of variation that is obtained, will assist in assigning probabilities. Clearly, this is a challenging task as a large variation is already observed for experiments with straightforward conditions as in this study.

Acknowledgements

The authors are grateful to all volunteers who donated for this study. We thank Amy Keetels and Lindy Clarisse for technical assistance and 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).

References

1. T. Sijen, Molecular approaches for forensic cell type identification: On mRNA, miRNA, DNA methylation and microbial markers, Forensic Sci. Int.: Genet. (2015).
2. L. McKenna, Understanding DNA results within the case context: importance of the alternative proposition. Frontiers in Genetics 4 (2013) 242.
3. P. Gunn, S. Walsh and C. Roux, The nucleic acid revolution continues, will forensic biology become forensic molecular biology? Frontiers in Genetics 5 (2014) 44.
4. van Oorschot RA, Ballantyne KN, Mitchell RJ (2010) Forensic trace DNA: a review. Investig Genet 1:14.
5. Toothman MH, Kester KM, Champagne J, Cruz TD, Street WS, Brown BL: Characterisation of human DNA in environmental samples. Forensic Sci Int 2008, 178:7-15.
6. E.A.M. Graham, G.N. Rutty, Investigation into "normal" background DNA on adult necks:

Implications for DNA profiling of manual strangulation victims, J. Forensic Sci. 53 (5) (2008) 1074–1082.
7. Cook O, Dixon L: The prevalence of mixed DNA profiles in fingernail samples taken from individuals in the general population. Forensic Sci Int Genet 2007, 1:62-68.
8. Dowlman EA, Martin NC, Foy MJ, Lochner T, Neocleous T: The prevalence of mixed DNA profiles on fingernail swabs. Sci Justice 2010, 50:64-71.
9. Wiegand P, Bajanowski T, Brinkmann B: DNA typing of debris from fingernails. Int J Leg Med 1993, 106:81-83.
10. R.A. Wickenheiser, Trace DNA: a review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact, Journal of Forensic Sciences 47.3 (2002): 442-450.
11. Zoppis S, Muciaccia B, D'Alessio A, Ziparo E, Vecchiotti C, Filippini A, DNA fingerprinting secondary transfer from different skin areas: morphological and genetic studies. Forensic Sci Int Genet 11 (2014) 137–143
12. I. Quinones, B. Daniel, Cell free DNA as a component of forensic evidence recovered from touched surfaces, Forensic Sci. Int. Genet. 6 (2012) 26–30.
13. Kamphausen T, Fandel SB, Gutmann JS, Bajonowski T, Poetsch M: Everything clean? Transfer of DNA traces between textiles in the washtub. Int J Legal Med 2015, 1-6.
14. H. Brayley-Morris, A. Sorrell, A.P. Revoir, G.E. Meakin, D. Syndercombe-Court, R.M. Morgan, Persistence of DNA from laundered semen stains: Implications for child sex trafficking cases, Forensic Sci. Int.: Genet. 19 (2015) 165-171.
15. S.F. Petricevic, J.A. Bright, S.L. Cockerton, DNA profiling of trace DNA recovered from bedding, Forensic Sci Int 159.1 (2006): 21-26.
16. V.J. Lehmann, R.J. Mitchell, K.N. Ballantyne, R.A.H. van Oorschot, Following the transfer of DNA: How does the presence of background DNA affect the transfer and detection of a target source of DNA?, Forensic Sci. Int.: Genet. 19 (2015): 68-75.
17. Wiegand P, Heimbold C, Klein R, Immel U, Stiller D, Klintschar M: Transfer of biological stains from different surfaces. Int J Legal Med 2012, 125:727-731
18. R.A.H. van Oorschot, M. Goray, E. Eken, R.J. Mitchell, Impact of relevant variables on the transfer of biological substances, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 547–548.
19. T.J. Verdon, R.J. Mitchell, R.A.H. van Oorschot, The influence of substrate on DNA transfer and extraction efficiency, Forensic Sci. Int.: Genet. 7 (1) (2013) 167–175.
20. M. Phipps, S. Petricevic, The tendency of individuals to transfer DNA to handled items, Forensic Sci. Int. 168 (2007) 162–168.
21. R.K. Farmen, R. Jaghø, P. Cortez, E.S. Frøyland, Assessment of individual shedder status and implication for secondary DNA transfer, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 415–417.
22. A. Lowe, C. Murray, J. Whitaker, G. Tully, P. Gill, The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces, Forensic Sci. Int. 129 (2002) 25–34.
23. M. Goray, E. Eken, R.J. Mitchell, R.A. van Oorschot, Secondary DNA transfer of biological substances under varying test conditions, Forensic Sci. Int. Genet. 4 (2010) 62–67.
24. 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.
25. 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.
26. R.I. Fleming, S. Harbison, The use of bacteria for the identification of vaginal secretions, Forensic Sci. Int.: Genet. 4 (2010) 311–315.
27. 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.
28. 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.
29. 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.
30. J. Juusola, J. Ballantyne, Multiplex mRNA profiling for the identification of body fluids, Forensic Sci. Int. 152 (2005) 1-12.
31. 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.
32. 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.
33. L.M. Milstone, Epidermal desquamation, Journal of dermatological science 36.3 (2004): 131-140.
34. D.H. Warshauer, P. Marshall, S. Kelley, J. King, B. Budowle, An evaluation of the transfer of saliva-derived DNA, Int. J. Legal Med, 126 (2012), 851-861.
35. B. Bhoelai, F. Beemster, T. Sijen, Revision of the tape used in a tape-lift protocol for DNA recovery, Forensic Sci. Int. Genet. Suppl. Ser. 4.1 (2013): e270-e271.
36. A. Lindenbergh, P. Maaskant, T. Sijen, Implementation of RNA profiling in forensic casework, Forensic Sci. Int.: Genet. 7 (2013) 159–166.
37. M. van den Berge, B. Bhoelai, J. Harteveld, A. Matai, T. Sijen, 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 Sci. Int.: Genet. 20 (2016) 119-129.
38. T.M. Clayton, J.P. Whitaker, R. Sparkes, P. Gill, Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Sci. Int. 91 (1998) 55-70.
39. C.C.G. Benschop, T. Sijen, LoCIM-tool: an expert's assistant for inferring the major contributor's alleles in mixed consensus DNA profiles, Forensic Sci. Int.: Genet. 11 (2014): 154-165.
40. R.A.H. Van Oorschot, et al. Are you collecting all the available DNA from touched objects? International Congress Series. Vol. 1239 (2003) 803-807.
41. R.A.H. Van Oorschot, et al. Persistence of DNA deposited by the original user on objects after subsequent use by a second person, Forensic Sci. Int.: Genet. 8.1 (2014): 219-225.
42. R.A.H. van Oorschot, et al. DNA transfer: The role of temperature and drying time. Legal Medicine, 16.3 (2014): 161-163.
43. I.W. Evett, P.D. Gill, G. Jackson, J. Whitaker, C. Champod, Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks, Journal of Forensic Sciences, 47.3 (2002), 520-530.
44. D. Taylor, D. Abarno, T. Hicks, C. Champod, Evaluating forensic biology results given source level propositions, Forensic Sci. Int.: Genet. http://dx.doi.org/10.1016/j.fsigen.2015.11.009
45. A. Biedermann, F. Taroni, Bayesian networks for evaluating forensic DNA profiling evidence: A review and guide to literature. Forensic Sci. Int.: Genet. 6 (2012), 147-57.
46. F. Taroni, C. Aitken, P. Garbolino, A. Biedermann, Bayesian Networks and Probabilistic Inference in Forensic Science. Chichester: John Wiley & Sons, Ltd.; 2006.

Chapter 5

DNA transfer and cell type inference to assist activity level reporting: Post-activity background samples as a control in dragging scenario

Forensic Science International: Genetics Supplement Series 6 (2017) e591-e592

Margreet van den Berge
Linda van de Merwe
Titia Sijen

Abstract

In an earlier study we examined the composition of human cell material prevailing on trousers in an activity-related scenario (controlled dragging-by-ankles experiment) van den Berge et al. (2016). This included DNA and RNA profiling of samples taken from the location of interest prior to the activity, and DNA profiling of similar samples after the activity. Here, we present results of a third sample set that was simultaneously collected post-dragging, which involves an area of the trousers that was not touched by the dragger. Results of this study increase our understanding regarding the importance of including control region samples when investigating activity-related scenarios.

Introduction

The prevalence, transfer, persistence and recovery of human cell material has increasingly gained attention, for instance because of the importance in addressing activity-level questions [2]. In a previous study we described the analysis of samples taken from the exterior surface of the ankle area of the trousers prior (background/control samples) and post (activity samples) a controlled dragging-by-ankle experiment [1]. Results showed that samples taken prior to an activity resulted in STR profiles where a major contributor could be deduced that corresponded to the owner of the trousers, despite the occurrence of often many non-donor alleles. The majority of samples collected after the dragging scenario resulted in at least partial DNA profiles where both the victim and perpetrator could be retrieved with the perpetrator representing a deducible major in approximately a quarter of the samples. Unlike this study [1], forensic casework scenarios do not allow for the analysis of pre-activity background samples from the exact same location as post-activity samples. We now describe the analysis of post-activity samples taken in the proximity of the grabbed area, that was however not touched by the dragger. Analysis of these samples provide insight in the suitability of including background control regions.

Materials and methods

Samples were collected by tape lifting 26 knee areas of trousers after performing dragging experiments as described in Ref. [1]. The tape lifts were processed in its entire and DNA was extracted using the QIAamp DNA Mini Blood kit (QIAGEN) according to manufacturer's instructions. DNA quantification and ethanol precipitation was performed as described in Ref. [1]. DNA profiles were generated using the AmpFlSTR® NGM™ PCR Amplification Kit (Life Technologies). Amplification, separation and interpretation was performed as described in Ref. [1].

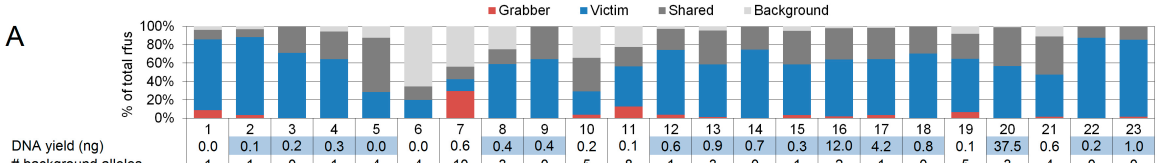
Donors with known genotypes (and informed consent) were used, so that the percentage of total rfus in the STR profile for each donor or for unknown contributors could be determined. Shared alleles were considered as a separate class. Profile interpretation was performed using the 15 STR loci in the NGM kit (Amelogenin was not included). Maximum allele count was used to determine the minimum number of contributors in a profile. Alike in Ref. [1], the LoCIM-tool was used to deduce the alleles of the most prominent component in the STR profile and to determine whether the "victim" or the "grabber" represents the major contributor in the DNA profile. STR profiles with ≤7 alleles were not subjected to further analysis.

Results and discussion

Three of the 26 samples were not subjected to further analysis, as these samples showed ≤7 alleles. An overview of DNA profiling results of the remaining 23 knee-area-of-trouser samples is shown in Figure 1A. Shown in Figure 1B and C are the results that had previously been obtained for the same trouser leg pre-, and post-dragging, respectively [1]. Alike for the ankle pre-dragging samples (Figure 1B), the majority of post-activity knee samples (Figure 1A) tend to show lower DNA yields compared to the ankle samples taken after the dragging activity (Figure 1C). Alleles corresponding to the victim (owner) were detected in the 100% of the knee samples (Figure 1A). Non-owner alleles are observed in 71% of the knee samples, but generally the owner (victim) alleles are substantially higher than non-owner alleles (peak height ratio victim/background). Hence, for 70% of the samples a major contributor could be deduced that corresponded to the owner of the trousers, while for the remaining samples no major contributor was deduced. Difficulty in deducing a major for these samples is most likely caused by the presence of some relatively high background alleles, preventing the possibility to deduce a major contributor in the STR profile. The estimated minimum number of contributors for the knee samples ranged up to two, which occurred in 61% of the samples. Grabber alleles are detected in 57% of the samples, with a maximum percentage of detected grabber alleles of 23% (Figure 1A, sample 7). An increase in detected grabber alleles can be observed especially in samples that show low numbers of owner alleles and high numbers of background alleles. These grabber alleles cannot be distinguished from background alleles, as can be seen by the low grabber/background peak height ratios. These results present a similar view as the results presented in Figure 1B and are in contrast to the post-activity samples (Figure 1C) in which the grabber stands out more clearly from the background alleles with an increased peak height ratio between grabber to background alleles.

Concluding remarks

With the increased interest in activity-related scenarios, the use of appropriate control regions for STR-profiling of this type of samples is highly relevant. We showed that in post-activity samples, a mixture of the owner of a private object and a perpetrator can be distinguished from background signals. In both pre- and post-activity control region samples, on the contrary, alleles matching a perpetrator may appear that are indistinguishable from background signals.



A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
DNA yield (ng)	0.0	0.1	0.2	0.3	0.0	0.0	0.6	0.4	0.4	0.2	0.1	0.6	0.9	0.7	0.3	12.0	4.2	0.8	0.1	37.5	0.6	0.2	1.0
# background alleles	1	1	0	1	4	4	10	3	0	5	8	1	3	0	1	2	1	0	5	3	4	0	0
Min. # contributors	1	2	1	2	2	1	2	1	1	2	2	2	2	1	2	2	2	1	2	2	2	1	1
% detected alleles victim	30	55	23	53	16	5	9	32	33	9	59	68	68	45	36	74	53	53	53	100	65	30	75
% detected alleles grabber	11	5	0	0	0	0	23	0	0	5	18	6	6	0	5	5	5	0	15	0	5	0	5
PH ratio victim/backgr.	2.5	1.7	-	0.9	3.0	1.2	1.4	0.9	-	1.2	1.1	1.6	2.3	-	1.2	3.4	3.0	-	2.7	7.7	1.2	-	-
PH ratio grabber/backgr.	0.0	0.5	-	-	-	-	0.2	-	-	0.6	0.4	0.7	0.0	-	0.0	2.0	2.0	-	0.3	-	0.3	-	-

B	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
DNA yield (ng)	8.2	0.2	0.6	0.4	0.9	0.1	0.1	0.2	0.5	0.6	0.3	12.5	10.5	4.4	29.0	6.9	2.6	0.7	0.4	4.9	2.3	1.5	1.3
# background alleles	9	3	13	25	10	1	8	14	5	20	14	24	19	17	2	6	2	19	14	10	10	17	4
Min. # contributors	2	2	2	3	2	2	3	3	2	3	3	3	3	2	2	3	2	3	2	2	2	3	2
% detected alleles victim	85	27	100	100	100	59	55	82	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
% detected alleles grabber	16	20	40	65	29	10	18	36	16	50	32	33	33	26	0	23	5	25	45	11	11	41	9
PH ratio victim/backgr.	0.9	1.2	4.5	3.0	1.9	1.8	1.5	1.4	3.4	1.0	1.6	2.4	3.0	5.9	7.5	10.4	2.7	3.7	2.8	7.4	3.5	6.4	7.8
PH ratio grabber/backgr.	0.0	0.2	0.2	0.3	0.0	0.0	0.4	0.2	0.9	0.4	0.8	0.3	0.4	0.3	-	0.4	0.0	0.1	0.3	0.4	0.9	0.2	0

C	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
DNA yield (ng)	8.7	0.4	0.9	2.8	1.4	0.3	0.4	0.1	0.6	1.3	2.0	10.0	6.2	2.0	2.5	2.0	2.0	0.6	0.6	0.7	0.7	3.9	5.6
# background alleles	4	3	3	5	13	19	22	11	20	15	16	20	35	8	2	5	21	10	10	19	21	9	2
Min. # contributors	3	2	2	2	3	3	3	3	3	2	3	3	3	3	2	3	3	3	3	3	3	3	2
% detected alleles victim	50	63	63	60	87	67	97	83	97	67	100	100	100	100	93	100	97	100	100	97	100	100	100
% detected alleles grabber	100	93	100	100	100	100	97	47	83	80	80	100	87	93	63	97	90	93	93	97	83	97	97
PH ratio victim/backgr.	0.7	1.5	1.2	0.8	1.4	1.4	1.6	1.5	7.3	1.0	6.2	2.3	2.2	2.6	2.9	5.3	2.0	4.1	3.2	2.1	3.4	10.2	10.8
PH ratio grabber/backgr.	12.5	3.3	3.2	3.7	3.2	1.9	2.6	1.3	1.7	1.0	1.0	2.7	0.7	1.4	1.2	2.2	1.1	1.8	1.6	1.6	1.1	2.9	2.0

Figure 1. Detailed information for A) samples from the knee area of the trousers that were collected after a dragging activity (this area was not touched by the grabber during the activity and are therefore regarded as background/control samples); B) corresponding background samples taken from the trouser leg ankle (dragging location) prior to performing the dragging scenario; C) corresponding samples from the trouser leg ankles collected after a dragging activity. Indicated per sample are the percentages of total rfus in the STR profile belonging to the grabber (red bars), victim (blue bars), shared alleles (dark grey bars) and background (light grey bars). Markings in the DNA yield section indicate samples for which the grabber (red) or victim (blue) was deduced as major contributor in the STR profile. Absence of colour indicates samples for which the deduced major does not match the grabber or victim.

Acknowledgements

Sanne Zijlstra is thanked for technical assistance, Jerry Hoogenboom is thanked for critically reading the manuscript.

References

1. M. van den Berge, G. Ozcanhan, S. Zijlstra, A. Lindenbergh, T. Sijen, Prevalence of human cell material: DNA and RNA profiling of public and private objects and after activity scenarios. Forensic Sci. Int.: Genet. 21 (2016) 81-89.
2. T. Sijen, Molecular approaches for forensic cell type identification: on mRNA, miRNA, DNA methylation and microbial markers, Forensic Sci. Int. Genet. 18 (2015): 21-32.

Chapter 6

Development of a mRNA profiling multiplex for the inference of organ tissues

International Journal of Legal Medicine (2013) 127:891-900

Alexander Lindenbergh*
Margreet van den Berge*
Roelof-Jan Oostra
Cindy Cleypool
Annette Bruggink
Ate Kloosterman
Titia Sijen

* *These authors contributed equally*

Abstract

Forensic characterisation of organ tissue generally occurs through histological and immunological assays of limited sensitivity. Here, we explore an alternative approach and examine a total of 41 candidate mRNA markers for their ability to differentiate between brain, lung, liver, skeletal muscle, heart, kidney and skin. Various selection rounds are applied involving 85 organ tissues (36 excised autopsy specimens and 49 frozen tissue sections, with at least ten specimens for each organ type), 20 commercially available RNAs from different human tissues and at least two specimens of blood, saliva, semen, vaginal mucosa, menstrual secretion or touch samples. Finally, 14 markers are regarded tissue-specific and included in an endpoint RT-PCR multiplex together with one general muscle, one blood and one housekeeping marker. This 17-plex is successfully used to analyse a blind test set of 20 specimens including mixtures, and samples derived from stabbing of organ tissues. With the blind test set samples, it is shown that an earlier described interpretation strategy for RNA cell typing results [1] is also effective for tissue inference. As organ-typing is embedded in a procedure of combined DNA/RNA extraction and analysis, both donor and organ type information is derived from the same sample. Some autopsy specimens presented DNA profiles characteristic for degraded DNA. Nevertheless, the organ-typing multiplex could generate full RNA profiles, which is probably due to small sizes of the amplicons. This assay provides a novel tool for analysis of samples from violent crimes.

Introduction

Evidentiary items collected at the scenes of violent crimes can include trauma-causing objects such as guns, bullets, knives or impact-inflicting implements. These objects can contain cell material from the victim and/or the perpetrator, which can be examined by DNA profiling [2, 3]. The biological material on these trauma-causing objects often represents blood and/or skin cells but in addition tissue fragments from organs may be present due to the perforation of the body [4]. These organ remnants may hold forensically relevant information. For instance, the presence of organ tissue matching the DNA profile of the victim on a knife provides a link to the crime and argues against an accidental finger cut. With shootings, inferring what type of cell material resides on which bullet is useful for crime scene reconstruction. For complex cases involving multiple shooters, specific bullets with particular organ tissues may be linked to the individual firearms thereby indicating from which gun the lethal shot was fired. Furthermore in practical cases of run-over traffic accidents, identification of human brain tissue was found to be important [5].

For the forensic identification of human organ tissues, enzyme-linked immunosorbent assays and histological methods are mostly employed [5–8]. These methods have limited sensitivity, and when little cell material is present, organ-typing analyses are complicated [9]. We reasoned that an alternative approach may lie in the use of mRNA markers specific for various organ tissues. In recent years, mRNA profiling has emerged as a sensitive, human-specific method for the simultaneous inference of several body fluids and skin cells [10–16]. RNA profiling is typically accompanied by a DNA/RNA co-isolation enabling DNA profiling on the exact same sample, which is of criminalistic importance. Whilst RNA molecules are generally considered to be less stable than DNA molecules, successful mRNA profiling on aged and degraded body fluid samples has been achieved upon selection of stable markers and the use of small amplicons [16–18]. It is not known whether forensic RNA analysis for tissues is possible, as the ex vivo stability of their RNA is unclear. Tissues with a low turnover rate and low concentrations of digestive enzymes such as the central nervous system may have the best quality of RNA [19]. Lung and liver, on the other hand, are described to be most prone to degradation resulting in low-quality RNA [20].

Here, we investigated a total of 41 candidate mRNA markers for the ability to identify brain, lung, liver, skeletal muscle, heart, kidney or skin. We focused on these tissues, as these are most likely found on trauma-causing objects. Specific and sensitive markers were selected and combined into an endpoint RT-PCR (reverse transcriptase PCR) multiplex system that allows analysis by capillary electrophoresis (CE) following routines standard in forensic laboratories. Such an assay will be of value in forensic casework.

Materials and methods

Organ samples

This study aimed at mRNA-based inference of human organ types. Organ specimens were collected at medical institutions and the necessary approvals were obtained from the required ethical examination commissions by these medical institutions. Seven different organ tissues (brain, lung, liver, skeletal muscle, heart, kidney and skin) were studied for which 85 specimens were collected. Thirty-six specimens represented autopsy samples and were excised from four individuals who voluntarily donated their body for research to the Academic Medical Centre (AMC, Amsterdam, The Netherlands). Cadavers were fresh frozen upon arrival and thawed (for other purposes) several days before sample excision. This resulted in 12 brain (four frontal lobe, four occipital lobe, four cerebellum), three lung, three liver, four skeletal muscle, three heart, three kidney and eight skin (four finger, four palm) samples. Excised tissues were transferred to a container with RNA*later*® (Ambion® by Life Technologies, Carlsbad, USA), transported to the forensic institute and stored at −80 °C until use. For each excision, new scalpel blades were used to minimize the risk of contamination. For each DNA/RNA co-isolation, tissue cubes of around 5×5×5 mm were used. Forty-nine specimens involved sections prepared from frozen tissues (University Medical Centre Utrecht (UMC) Central Biobank, Utrecht, The Netherlands). This set comprised seven samples of each of the seven organ types. These samples mainly represented tissues remaining after surgery or diagnostic research that were frozen directly after collection or diagnosis. Sections were stored at −80 °C until use. For extraction of nucleic acids, tissue squares of around 2×2 mm were excised from the sections.

To assess specificity on a larger set of human tissues than those targeted by the selected markers, the FirstChoice® Human Total RNA Survey Panel (Ambion® by Life Technologies, Carlsbad, USA) was used which contains pools of total RNA (of at least three donors) for 20 different human tissues in a concentration of 1 µg/µL. These RNAs were diluted to 10 ng/µL to facilitate analysis.

Body fluid samples

To assess the performance of the organ-typing multiplex when body fluids are involved, RNA extracts of donations of blood, saliva, semen, menstrual secretion, vaginal mucosa and skin (touched cotton patches) were used [16]. When testing the final multiplex, two samples provided by different individuals were used for each of these cell types except for semen for which two donations from fertile donors and two specimens from sterile donors were used. Volunteers had given informed consent.

Mock casework samples

For the preparation of a blind test set, small portions of the autopsy samples were cut, transferred to a petri dish and rubbed with dry swabs (Deltalabs, Barcelona, Spain) to collect some of the cell material. Mixed samples were prepared by sampling multiple excised portions with the same swab (using different sides of the swab if possible) and each tissue portion was discarded straight after a sampling. The aim was to have approximately equal amounts of each organ type on the swab. The blind test consisted of 20 swabs representing three single organ samples, one neat blood specimen and 16 mixed samples: eight two-organ mixtures, five three-organ mixtures, two four-organ mixtures and one seven-organ mixture. Each organ was represented four to ten times and neat blood was present once.

To mimic violent crime scene samples, autopsy samples were stabbed using scalpel knives. Scalpels were inserted once and pulled out immediately. To mimic an abdominal stabbing, a stacking of liver, skeletal muscle and skin was prepared. In addition, liver covered with skin, skeletal muscle covered with skin and skin alone were stabbed. Liver and skeletal muscle samplings had a thickness of 6 to 10 mm and skin specimens that were excised from finger, a thickness of 2 to 3 mm (width and length were around 1.5 to 2 cm). Immediately after stabbing, cell material was collected from the scalpels using swabs.

Retrieval of candidate markers

Marker candidates for brain, lung, liver, skeletal muscle, heart or kidney were selected from the gene annotation portal BIOGPS [21]. Candidates were included because of high expression values (in the first selection round above 1,600, later above 900) and expected specificity, which refers to much higher expression values in the target tissue than in other tissues, especially those assessed in this study.

Primer design

For the mRNA markers, cDNA-specific primer sets were designed that either spanned an intron or had one primer covering an exon–exon junction using Ensemble and NCBI Primer-BLAST [22, 23]. For 18S-rRNA and HBB, primers were adopted from literature [11]; for LOR [16] and LCE1C [13], primers were redesigned. In silico BLAST analyses affirmed human specificity of the amplifications. Details for the amplicons and primers of the markers included in the final multiplex are given in Table 1. In the process of developing this final multiplex, other multiplexes were used for which markers and amplicon numbers are depicted in Table 2. The labelled primers underwent HPLC-purification (Applied Biosystems™, Warrington, United Kingdom).

Table 1. Characteristics of markers included in organ-typing multiplex. For each amplicon one of the primers covers an exon-exon junction unless indicated otherwise.

Tissue	Gene	[Primer] µM	Forward primer (5'-3') Reverse primer (5'-3')	Dye	Size (bp)	Biological function
Brain	SNAP25	0.02	TCGGCGGCTCCACCACAGTT TTGGCTCTGGACCTGGGCTTCTC	6FAM	78	Synaptic vesicles: maintenance & signalling Reported expression in spinal cord [24]
	RTN1	0.08	AGGCCACCCTTTCAAGGCCTACTT AGAACTGCAGGCAGTCCGTGT	NED	86	Neuroendocrine cells: maintenance & signalling Reported expression in spinal cord [24]
Lung	SFTPB	0.06	CCATGATTCCCAAGGGTGCGCTA CGCCACCAGAGGTACCACGC	NED	68	Stability of pulmonary tissues by reducing surface tension
	SFTPD[a]	0.10	CCTGCCTGGTCGCGATGGAC CCAGGCATCCCTGCTTGCCC	6FAM	99	Stability of pulmonary tissues by defence against inhaled micro-organisms
Liver	AMBP	0.03	TTGGCTGACCGAGGTGAATGTGTCC[b] ACCAGTTGCCCACCCCCTGAT	VIC	119	Protease inhibitor & lipocalin transport protein secreted in plasma
	VTN	0.05	GTGCAAGCCCCAAGTGACTCG CATAGACCGTGTACTCATCCTCCG	VIC	65	Secreted by hepatocytes dampens early MAC formation
Skeletal muscle	TNNI2	0.13	ATGTCTGAAGTGCAGGAGCTCTGC GTCGTACTTCTCCTCTTCAGCCGC	PET	72	Fast-twitch skeletal muscle protein. Reported expression in thyroid and tongue [24]
	MYH1	0.8	TTCGCATCTCTACGCCAGGGTCCTTA[b] AGGAAAGGAGCAGCCTCCCCAAA	NED	104	Heavy chain myosin, muscle contraction Reported expression in thyroid [24]
Heart	MYL7	0.08	AAGCTCAATGGGACAGACCCCG CACCACCCCTTTGCCGCTGG	VIC	82	Light chain regulatory myosin motor protein
	MYBPC3	0.4	TCATAGAGGCAGAGAAGGCAGAGC CCCTTTGGGACTTGGGGCACT	VIC	130	Myosin-associated protein Cardiac isoform
General muscle	NMRK2	0.4	GGAGGAGAAGTACCCGCAACTACA ACACCGTTGGCCTCCATCTCCT	6FAM	109	Muscle β1 integrin binding protein regulation of terminal myogenesis
Kidney	UMOD	0.12	ATCACACGGAAAGGTGTCCAGGC TTTTTTGGAGCACAGGGCTTTCCGC[b]	6FAM	145	Formation of a water permeable barrier inhibition calcium crystallization in renal fluids
	FXYD2	0.06	CTCCATCCAGGCCCCAGGCA CGGTCTCATAGTCATAGTAGAACGGG	PET	143	Modulates the activity of the Na,K-ATPase Reported expression in pancreatic islet [24]
Skin	LCE1C	0.2	TGTGACCCCGCTCCTGAATCCG CTTGGGAGGGCACTTGGGGGTG	NED	99	Component of the cornified envelope Mouse homolog expressed in stomach [24]
	LOR	0.6	CTTTGGGCTCTCCTTCCT TCTTCACGCAGTCCACTG	PET	84	Component of the cornified envelope In terminally differentiated epidermal cells
House-keeping	18S-rRNA[c]	0.015	CTCAACACGGGAAACCTCAC CGCTCCACCAACTAAGAACG	PET	110	Small ribosomal subunit protein biosynthesis in all cells
Blood	HBB	0.1	GCACGTGGATCCTGAGAAC ATGGGCCAGCACACAGAC	6FAM	61	Component hemoglobin oxygen binding in red blood cells

[a] Primers span an intron.

[b] Underlined nucleotides are 5' tails added to improve multiplex spacing.

[c] Primer pair does not cover an exon-exon junction nor spans an intron.

RNA profiling

DNA/RNA co-isolation was performed as described by Lindenbergh et al. [16]. Tissue debris that had not fully lysed after 2 h of incubation at 56 °C was discarded using a QIA shredder column. RNA extracts were DNase-treated [16], and copy DNA (cDNA) was synthesised [16] using 10 µL RNA extract. Variable amounts (up to 5 µL)

Table 2 Performance of all candidate markers in various assays and selection rounds.

		Selection round[a]					
Tissue	Marker	1[b]	2[c]	3[d]	4[e]	5[f]	Remark
Brain	MBP						
	SNAP25						
	OLFM1						
	RTN1						
	NEFL						
	PLP1						
Lung	SFTPC						
	SFTPB						Among two most sensitive markers in round 1
	BPIFB1						Specific but less sensitive than SFTPB & SFTPD, so not taken to round 2
	SCGB1A1						Specific but less sensitive than SFTPB & SFTPD, so not taken to round 2
	SFTPD						Among two most sensitive markers in round 1
	ALOX5						
Liver	APOA1						
	ALB						
	APOC3						
	AMBP amp1[g]						1 extra T residue added to rv primer in round 2 for multiplex spacing
	AMBP amp2						Amplicon 2 dismissed after round 3 because superfluous
	CYP2E1						
	AHSG						
	VTN						Included in round 3 to have 2nd liver marker
Skeletal muscle	TNNC2						
	MYLPF						
	MYL1						
	TNNI2						
	MYH2						
	ATP2A1						
	NEB						
	MYH1						2 extra Ts to rv primer in round 2 for multiplex spacing; thyroid expr.
	TNNT1						
Heart	ACTC1						
	TNNI3						
	MYL7 amp1[g]						
	MYL7 amp2						Amplicon 2 dismissed after round 3 because of low sensitivity
	TNNT2						
	MYBPC3						Included in round 5 to have 2nd heart marker
General muscle	NMRK2						Initially included in round 3 for heart, found to mark general muscle
Kidney	UMOD						5 extra T residues to rv primer in round 4 for multiplex spacing
	ABP1						
	SLC13A3						
	PDZK1IP1						Not specific when assayed via CE in round 2
	FXYD2						Included in round 3 to have 2nd kidney marker
Skin	LOR						
	LCE1C						
Housekeeping	18S-rRNA						Final primer concentration lower than in [16] to prevent saturation
Blood	HBB						Co-expressed in several tissues
Number of markers		37	14	16	16	17	
Number of amplicons		37	16	18	16	17	

[a]Color coding: good performance | dismissed | not analysed

[b]1: Tissue-oriented small multiplexes & detection on agarose gel; 85 single source samples - see Supplementary Table 1

[c]2: 16-plex & detection by CE; selected set of single source samples; results not shown

[d]3: 18-plex & detection by CE; 85 single source samples (each two cDNA inputs); 20 human tissues (various RNA amounts); six common forensic samples (body fluids and touch DNA, four different donors each, one cDNA input); 20 blind samples including mixtures (each two cDNA inputs); for all analyses results not shown

[e]4: 16-plex & detection by CE; selected sample set; results not shown

[f]5: Final 17-plex & detection by CE; 48 single source samples (one cDNA input) - see Supplementary Table 2; 20 human tissues (various RNA amounts) - see Suppl. Table 3; six common forensic samples (body fluids and touch DNA, two different donors each, one cDNA input) - see Supplementary Table 4; 20 blind samples including mixtures (each four replicates) - see Table 4

[g]After round 1 only one tissue candidate remained. Two amplicons for different exons were assessed to increase amplification chance

of cDNA were used as inputs for the various RT-PCRs. To analyse picogram amounts of RNAs of the FirstChoice® Human Total RNA Survey Panel, cDNAs were diluted accordingly. For the various candidate markers that were assessed by agarose gel electrophoresis RT-PCR primer concentrations were 0.2 µM. Primer concentrations of the optimized multiplex are reported in Table 1. The RT-PCRs employed amplification settings as described in [16].

RNA profiling

DNA/RNA co-isolation was performed as described by Lindenbergh et al. [16]. Tissue debris that had not fully lysed after 2 h of incubation at 56 °C was discarded using a QIA shredder column. RNA extracts were DNase-treated [16], and copy DNA (cDNA) was synthesised [16] using 10 µL RNA extract. Variable amounts (up to 5 µL) of cDNA were used as inputs for the various RT-PCRs. To analyse picogram amounts of RNAs of the FirstChoice® Human Total RNA Survey Panel, cDNAs were diluted accordingly. For the various candidate markers that were assessed by agarose gel electrophoresis RT-PCR primer concentrations were 0.2 µM. Primer concentrations of the optimized multiplex are reported in Table 1. The RT-PCRs employed amplification settings as described in [16].

When a novel candidate marker was examined, first the length of the amplified cDNA fragment was confirmed using an appropriate RNA extract. In all cases the observed fragment lengths approximated the designed sizes. Minus RT reactions were included in all analyses (*i.e.* both when assessing results on agarose gels and when performing CE analysis) and did not show signals.

When RT-PCR products were assessed by agarose gel electrophoresis, 10 µL PCR product were supplemented with 2 µL BlueJuice Loading Buffer (Life Technologies, Bleiswijk, The Netherlands) and separated on 4 % agarose gels (Multi Purpose, Roche, Almere, The Netherlands) in 1× TBE buffer (10×, Gibco, Invitrogen, Bleiswijk, The Netherlands) containing 0.5 µg/ml Ethidium Bromide (Invitrogen, Bleiswijk, The Netherlands). Fragments were separated for approximately 60 min at 80 V (PowerPac Basic™, BioRad, Ede, The Netherlands), visualised using UV-light (Chemi Genius2, Syngene, Cambridge, United Kingdom) and analysed using GeneSnap Software (Syngene, Cambridge, United Kingdom). When the RT-PCR products were analysed by CE, purified PCR products [16] were separated using POP-7 as separation matrix. In case of overloaded profiles, RT-PCR products were diluted 100- or 20-fold in water and re-injected for CE. GeneMapper ID-X version 1.1.1 (AB) was used to analyse the RNA profiles with 150 rfu as detection threshold.

RNA profile interpretation

RNA extracted from neat blood was used to show functional methodology for cDNA synthesis and PCR amplification when multiplex amplification and CE analysis is performed, as one blood marker resides in the multiplex. The 18S-rRNA marker in the multiplex has the role to indicate the presence of RNA and absence of factors inhibiting cDNA synthesis or PCR, which is especially informative when none of the other markers give a signal. Since 18S-rRNA signals can vary in strength for several reasons (*e.g.* differences in the proportion mRNA/rRNA isolated from a sample, relatively low or high sensitivity of organ-specific markers compared to the 18S-rRNA marker and stacking of 18S-rRNA signals in case of mixed samples), presence of the 18S-rRNA signal in a profile is not absolute, and thus, profiles lacking the 18S-rRNA signal can be considered for data interpretation.

DNA profiling

DNA was quantified using the highly sensitive *Alu* repeat system that provides accurate measurements down to 0.5 pg/µL, according to the protocol of Nicklas et al. [24] with minor adjustments to the primer concentrations as 200 nM *Alu* forward primer and DYZ5 reverse primer were used. DNA profiling was performed using the AmpFℓSTR® NGM™ PCR Amplification Kit (Applied Biosystems (AB), Foster City, Texas, USA). Where possible, 500 pg of genomic DNA (gDNA) was used as input. Amplification products were separated using CE on a 3130XL genetic analyser (AB) with POP-4 (AB) as separation matrix as earlier described [16]. Analysis of DNA profiles was performed with GeneMapper ID-X version 1.1.1 (AB) using 50 relative fluorescence units (rfu) as detection threshold [16].

Results and discussion

Selection of candidate markers

Candidate markers for organ typing were retrieved by a BIOGPS [21] search and included six candidates for brain, six for lung, six for liver, nine for skeletal muscle, four for heart and four for kidney (genes are depicted in Table 2, column first selection round). Additionally, two reported skin markers (LOR [16, 25] and LCE1C [13]) were included in the candidate set, resulting in a total of 37 genes to test.

Test material consisted of 36 excised autopsy tissues and 49 frozen tissue sections. These were subjected to RNA/DNA co-isolation [16]. DNA profiling learned that the quality of the autopsy samples was less than that of the frozen sections as 13 out of 36 autopsy samples resulted in partial DNA profiles while all of the 49 sections resulted in

full profiles (Table 3). The partial DNA profiles had a "skislope" appearance typical for degraded DNA and occurred mainly for lung, liver, heart and kidney samples.

For the mRNA marker candidates, tissue-orientated multiplexes were prepared containing all markers for each of the seven tissue types. The markers within each of these multiplexes had uniquely sized amplicons (between 58 to 175 nucleotides), which allowed assessment of their individual performance. Specificity of the candidate markers was determined by testing all 85 tissues with all seven tissue orientated multiplexes, which resulted in 595 amplifications. When visualising the PCR products on agarose gel, 14 markers appeared specific. These represent one to four markers for each of the organ tissues (Table 2, details in Supplementary Table 1).

To befit the standard forensic procedures, we aimed for an end-point multiplex RT-PCR assay and analysis of the fluorescently labelled products by CE. The process of developing such a multiplex included various selection rounds as summarised in Table 2. Prior to round 3, new marker candidates were retrieved from BIOGPS for the three organ types (heart, liver and kidney) for which only one marker remained after selection round 2. Finally, two apparently specific markers were obtained for each of the seven tissues (Table 2). These 14 tissue-specific markers were supplemented with one general muscle marker recognising both skeletal and heart muscle, one blood marker (HBB [16]) and one housekeeping marker (18S-rRNA [16]). This created a 17-plex that underwent optimisation for primer concentrations and marker spacing.

Table 3 Results of DNA analyses of the excised autopsy samples and frozen tissue sections.

Type of sampling	*n*	Full DNA profile	Partial DNA profile	Average DNA concentration (ng/µL)	Range of DNA concentrations (ng/µL)
Autopsy					
Frontal lobe	4	4	0	40.99	2.82 – 66.54
Occipital lobe	4	4	0	25.71	1.79 – 44.87
Cerebellum	4	3	1	38.14	2.46 – 76.50
Lung	3	0	3	2.85	1.41 – 4.03
Liver	3	0	3	0.77	0.54 – 1.03
Skeletal muscle	4	4	0	6.61	0.10 – 13.58
Heart	3	0	3	2.19	0.66 – 3.20
Kidney	3	0	3	2.08	0.39 – 5.00
Skin, hand palm	4	4	0	29.33	10.08 – 58.92
Skin, fingertip	4	4	0	33.14	27.38 – 41.76
Frozen sections					
Brain	7	7	0	15.23	5.57 – 31.42
Lung	7	7	0	67.18	6.71 – 269.09
Liver	7	7	0	3.82	2.52 – 5.15
Skeletal muscle	7	7	0	3.62	1.96 – 7.07
Heart	7	7	0	5.41	2.48 – 9.55
Kidney	7	7	0	13.77	5.20 – 30.65
Skin	7	7	0	18.47	1.81 – 49.53

Specificity and sensitivity of the organ-typing 17-plex regarding human tissues

The organ-typing 17-plex was subjected to performance testing. First, 48 frozen tissue specimens were analysed. Signals for the correct tissue-specific markers were obtained for all 48 samples (Supplementary Table 2). In some samples only one of the tissue-specific markers responded. These samples involve tissues for which generally one of the markers gives lower signals (Figure 1A). Non-specific signals of relatively low heights were sporadically observed (Supplementary Table 2). Co-expression of blood was seen for brain, liver, skeletal muscle and heart. Typical electropherograms (epgs) are shown in Figure 1A. In these epgs, peak heights vary not only for the tissue-specific markers, but also for the 18S-rRNA marker, which does not necessarily reflect a low expression of 18S-rRNA for certain tissues; some tissue-specific markers give relatively high peaks, due to which lower inputs were used for the profiles presented in Figure 1A. In those cases, the 18S-rRNA signal is detected in profiles with higher inputs. Thus, we do not regard 18S-rRNA a strict positive control required to give a signal in each analysed profile. Nevertheless, 18S-rRNA is very useful to infer whether RNA/cDNA is present, which is especially important when no other markers give peaks.

Next, RNAs included in the FirstChoice® Human Total RNA Survey Panel were analysed with the 17-plex. This panel contains 20 different human tissues including six of the tissues (brain, lung, liver, skeletal muscle, heart and kidney) targeted by the multiplex. Since these RNAs are of excellent quality and have a fixed RNA concentration, they were used to determine the sensitivity of the multiplex. For informative profiles, picogram amounts of RNA in the range of 50, 10 and 2 pg sufficed; nanogram amounts gave strong over-amplification and artefacts. The brain, liver and kidney markers required 2 pg in the multiplex PCR to present full RNA profiles, while the lung, skeletal muscle and heart markers needed 10 pg. The housekeeping marker required 50 pg RNA for most tissues (Supplementary Table 3). Blood is detected for only some of the tissues. Cross-reactions are evident for three markers: lung marker SFTPB and skeletal muscle markers TNNI2 and MYH1 all show signals for the thyroid (Supplementary Table 3). Thyroid expression of these latter two markers is reported in BIOGPS [21] albeit to lower level than in skeletal muscle.

Performance of the organ-typing 17-plex on body fluids

The performance of the organ-typing 17-plex was also assessed for cell types generally encountered in forensic context: blood, saliva, semen (from fertile and sterile donors), menstrual secretion, vaginal mucosa and skin (in the form of touched cotton patches). For each cell type, two independent samples were analysed. The organ-typing 17-plex contains one blood marker (HBB) that should respond in blood and menstrual secretion samples and two skin markers (LCE1C and LOR) expected to give signals in touch samplings. Strong signals for HBB and LCE1C were observed for the expected

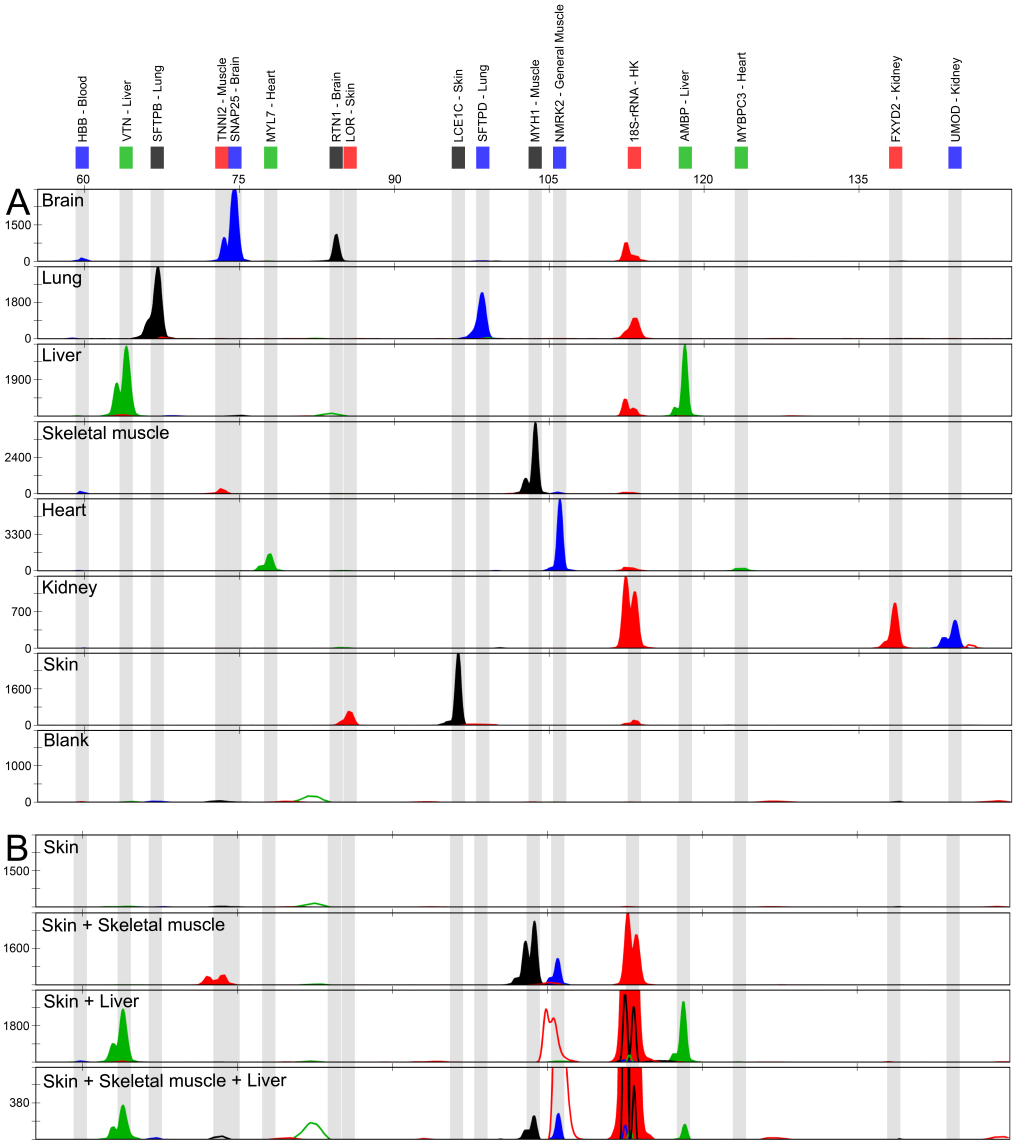


Figure 1. Typical overlay electropherograms as produced by the optimised 17-plex. A: Examples of the single source samples brain, lung, liver, skeletal muscle, heart, kidney and skin. Some tissue-specific markers generally give lower signals than their counterparts: AMBP in liver, TNNI2 in skeletal muscle, MYBPC3 in heart and LOR in skin. B: Results of mock stabbing experiments. A,B: Not filled in peaks correspond to either dye blobs which could not be removed by purification (*e.g.* at 85 nt), or artefact peaks due to over-amplification. In blanks and minus RT reactions no marker expression was observed.

sample types (Supplementary Table 4). LOR amplification failed, which may be due to the low cell content in touch specimens [16]. Additional non-specific signals (Supplementary Table 4) were sporadic and of relatively low height and we expect that these signals will not be reproduced upon replicate RNA analyses [1]. Muscle marker TNNI2 is reported to give expression in tongue [21], but apparently this does not produce TNNI2 signals in saliva.

Performance of the organ-typing 17-plex on a blind test set

A blind test set of 20 samples prepared from autopsy material was examined. Since multiple tissues may be hit upon violent actions, mixtures were included in this set. In [1] we introduced an mRNA interpretation strategy based on two components: (1) replicate RNA analyses (preferably four) and (2) interpretation rules that categorise the results as 'observed', 'sporadically observed' or 'not observed'. Scoring is based on a 'x=n/2' rule in which 'x' is the number of found and 'n' the number of theoretically possible peaks for a tissue. A tissue is scored 'observed' if $x=n/2$, 'not observed' if $x=0$ and 'sporadically observed' if $0<x<n/2$. During case interpretation, tissues scored as 'sporadically observed' are effectively regarded as 'not observed'; the separate category is used to translate all signals to the scorings. When next to other observed tissues, a score 'observed' is obtained for blood; the presence of blood is not regarded as a contribution of peripheral blood since blood may coincide in the organ tissues. This interpretation strategy was applied to the blind test set RNA data.

The scores derived from four replicate RNA profiles are presented in Table 4. In the 20 samples a total number of 49 tissues are present; 39 of these are scored 'observed', six 'sporadically observed' and four 'not observed'. For 12 samples all input tissues are considered 'observed' (Table 4). The lung, liver, heart and kidney material that was used to create this blind test set, had presented partial, degraded DNA profiles due to the condition of the cadavers from which they were excised (Table 3). Nevertheless, full RNA profiles and scorings 'observed' were obtained most times these tissues resided in the blind test set (results not shown and Table 4). Skin, heart and skeletal muscle appear most prone to marker drop out (Table 4). The skin markers are less effectively detected in the excised skin specimens than in touch specimens (Supplementary Table 4), especially when considering the DNA yields that are much higher for the excised skin specimens (Table 3) [16]. Presumably, LCE1C and LOR are predominantly expressed in upper epidermal layers that represent only a fraction of the excised skin specimens, as these comprise the epidermal, dermal and hypodermal layers plus adipose tissue. For both heart and skeletal muscle, marker drop out occurs especially for one of the two markers (MYBPC3 and TNNI2 respectively), which is in agreement with the observation that these markers required 10 pg and not 2 pg RNA to give signals in the sensitivity test (Supplementary Table 3).

Table 4 Results for the blind test set which includes mixed samples. Four RNA profiles were generated for each sample. Indicated are the number and type of tissues present in a sample ('present', marked by black cell;'-' means not present) and the scoring results using the interpretation guidelines described in [1]. The color coding is used to present the scoring results as indicated below the table.

Sample	# Tissues present[b]	Neat blood		Brain		Lung		Liver		Skeletal muscle[a]		Heart[a]		Kidney		Skin	
		present	result[c]	present	result	present	result	present	result	present	result	present	result	present	result	present	result
1	1	-		-				-		-		-		-		-	
2	1	-		-		-		-		-		-		-			
3	1	-		-		-				-		-		-		-	
4	1			-		-		-		-		-		-		-	
5	2	-		-		-				-				-		-	
6	2	-				-		-		-		-		-			
7	2	-				-		-				-		-		-	
8	2	-		-		-		-				-		-			
9	2	-		-				-				-		-		-	
10	2	-		-		-				-		-				-	
11	2	-		-				-		-		-				-	
12	2	-		-		-		-						-		-	
13	2	-		-				-		-				-		-	
14	3	-		-				-		-						-	
15	3	-		-				-				-		-			
16	3	-				-				-				-		-	
17	3	-		-						-		-				-	
18	4	-		-				-						-			
19	4	-		-						-						-	
20	7	-															

[a] General muscle marker NMRK2 not included during interpretation.

[b] Samples are ordered by number of tissues present; during analysis the samples were randomized.

[c] Blood may coincide in organ tissues; positive scores are colored green.

[d] No absent tissues were scored as observed.

	Present	Absent
Observed		none[d]
Sporadically observed		
Not observed		

Importantly, no false positive identifications, *id est* scores 'observed' for tissues not present in a sample, have occurred for this blind test set; the peaks seen for tissues not present in the samples all result in a score 'sporadically observed' which is effectively regarded 'not observed'. Blood was scored 'observed' both in the sample containing neat blood and for six samples carrying mixtures of organ tissues. The 18S-rRNA marker gave a signal in all the 80 RNA profiles. Thus, the multiplex allows assessment of mixed samples, and organ-typing results can be interpreted using the approach described in [1].

Performance of the organ-typing 17-plex on mock stabbings

To assess performance of the organ-typing multiplex in a mock violent case, an abdominal stabbing incident was mimicked. When analysing cell material present on the scalpel knife used to stab a stacking of skin, skeletal muscle and liver, signals for both liver and skeletal muscle were retrieved (Figure 1B). This means that also the lowest tissue (liver) is detected. The signal for 18S-rRNA is overloaded due to the additive effect from the multiple tissues. This hampers the use of a higher cDNA input in the 17-plex that may have facilitated improved detection of tissue-specific markers that generally give lower peaks (Figure 1A) such as TNNI2 that shows drop out for the triple layer stabbing. Skin was not detected which is probably a consequence of the vertical cut due to which hardly cells from the upper epidermal layer (where the skin markers are thought to be expressed) reside on the knife (Figure 1B). The control samples (stabbings of skin only, liver covered with skin and skeletal muscle covered with skin) neither show signals for skin, while skeletal muscle and liver were detected. This experiment shows that cells from the lower tissues remain present on the scalpel after withdrawal through the upper layers including skin.

Concluding remarks

Here, we created an end-point multiplex RT-PCR assay to infer the presence of brain, lung, liver, skeletal muscle, heart, kidney and skin. Performance tests using this organ-typing 17-plex showed high specificity and sensitivity, and the ability to identify different sources in case of mixed samples. Non-specific signals were only evident for the thyroid (from both skeletal muscle markers and one lung marker). However, not all bodily tissues were tested. Nevertheless, we do not anticipate many more non-specific signals as only very few non-tested tissues are reported in BIOGPS to give expression for the selected markers (Table 1). For degraded samples that presented partial DNA profiles, full RNA profiles were obtained which may be due to the small sizes of the amplicons. The RNA profiles generated by the 17-plex can show peak

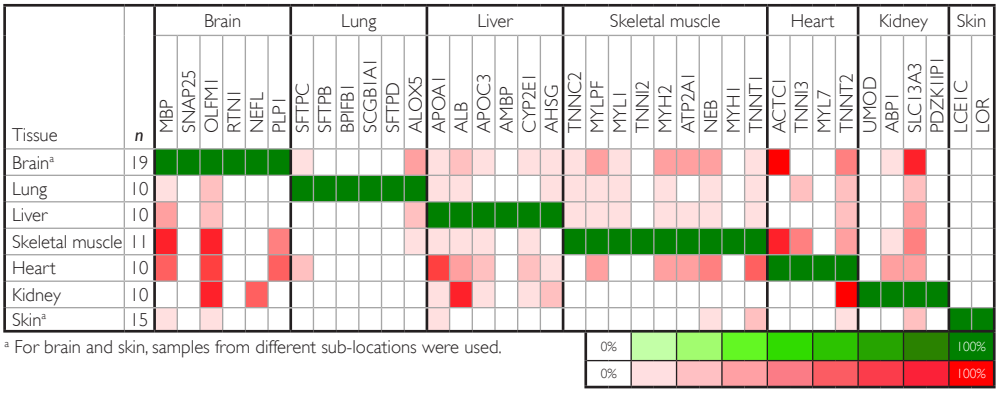
imbalance for the different markers (Figure 1). Also, RNA profiles in general show peak height variation due to differences in gene expression for distinct donors and with changes in physiological condition. Furthermore, variation between RNA replicates occurs [26]. This variation was also observed with the organ-typing multiplex described here. Therefore, we have recommended to perform multiple RNA analyses and use interpretation guidelines as described in [1]. This strategy was successfully applied to a blind test set of 20 samples including mixtures. In conclusion, a novel tool for forensic analysis of samples from violent crimes is presented.

Acknowledgements

The authors are very grateful to all donors from whom tissues have been used in this study. We thank Kees Krijgsman and Eppo van Houten for excising the autopsy samples. Role of funding: this study was supported by a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN). The work leading to these results also received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 285487 (EUROFORGEN-NoE). We thank Lindy Clarisse for technical assistance and Antoinette Westen for assistance with the manuscript.

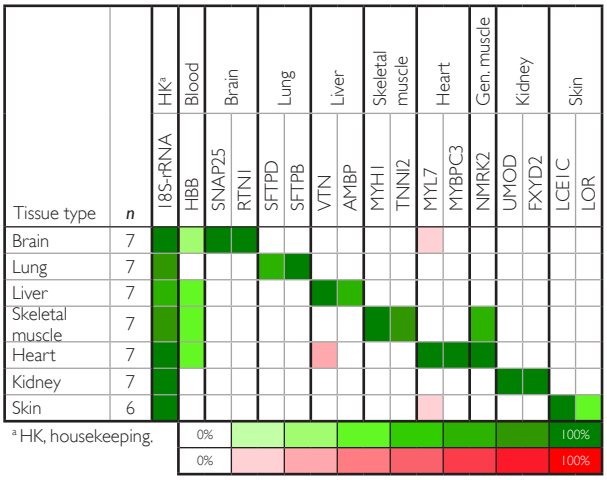
Supplementary material

Supplementary Table 1 Results of specificity tests for 37 putative organ-specific markers using 85 single sources samples. Tissue-oriented multiplexes are used and results are analysed by agarose gel electrophoresis. Darker color tones correspond to increasing percentages of positive signals. Green cells indicate signals corresponding to the target tissue, red cells signals for non-target tissues. A minimum number (*n*) of 10 donors was analysed for each tissue type.



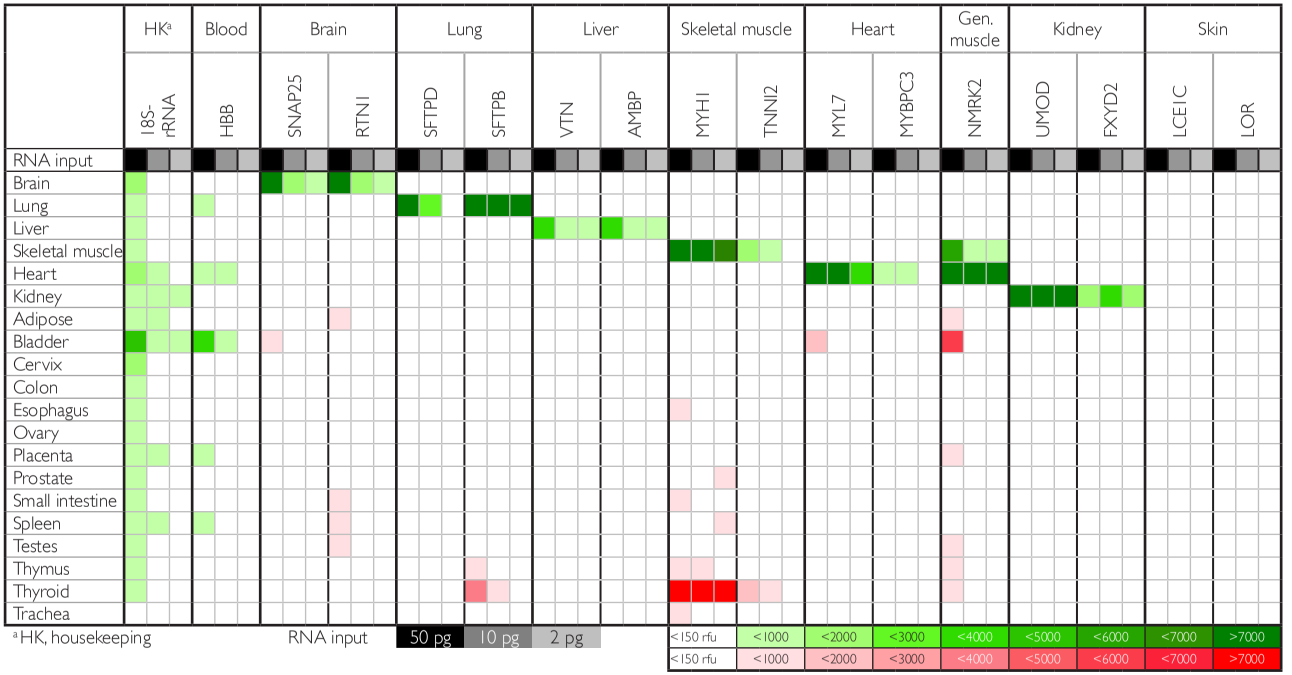
[a] For brain and skin, samples from different sub-locations were used.

Supplementary Table 2 Results of specificity tests for the markers in the organ-typing 17-plex using 48 frozen tissue sections. The number (*n*) of samples per tissue is indicated. The percentage of samples in which marker expression was observed is indicated as a color scale in which darker tones represent a higher percentage. Green cells indicate signals corresponding to target tissue; red cells signals for non-target tissues.

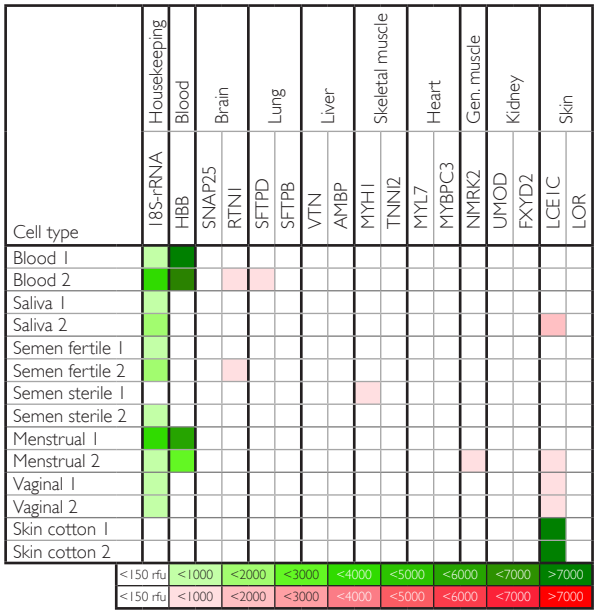


[a] HK, housekeeping.

Supplementary Table 3 Specificity and sensitivity results of the organ typing 17-plex tested on 20 different tissues of the FirstChoice® Human Total RNA Survey Panel. RNA inputs in the RT-PCR were 50, 10 or 2 pg (indicated by grey scale). Color marks indicate tissues for which signals for a marker are observed. Darker color tones indicate higher peak heights (rfu; see color scale below table). Green cells indicate signals corresponding to the target tissue, red cells signals for non-target tissues. Skin samples are not represented in the panel.



Supplementary Table 4 Results for the specificity of the markers in the organ typing 17-plex when analysing blood, saliva, semen (fertile/sterile), menstrual secretion, vaginal mucosa and touched cotton patches (two specimens of different donors each). Darker color tones indicate higher peak heights (rfu). Green cells indicate specific signals, red cells mark non-specific signals.



References

1. Lindenbergh A, Maaskant P, Sijen T (2013) Implementation of RNA profiling in forensic casework. Forensic Sci Int Genet 7:159–166
2. Karger B, Meyer E, DuChesne A (1997) STR analysis on perforating FMJ bullets and a new VWA variant allele. Int J Legal Med 110:101–103
3. Dieltjes P, Mieremet R, Zuniga S, Kraaijenbrink T, Pijpe J, de Knijff P (2011) A sensitive method to extract DNA from biological traces present on ammunition for the purpose of genetic profiling. Int J Legal Med 125:597–602
4. Visser R (1998) Bullet cytology; results of a pilot study. Sixth Cross Channel Conference on Forensic Medicine p11
5. Takata T, Miyaishi S, Kitao T, Ishizu H (2004) Identification of human brain from a tissue fragment by detection of neurofilament proteins. Forensic Sci Int 144:1–6
6. Kimura A, Ikeda H, Yasuda S, Yamaguchi K, Tsuji T (1995) Brain tissue identification based on myosin heavy chain isoforms. Int J Legal Med 107:193–196
7. Seo Y, Kakizaki E, Takahama K (1997) A sandwich enzyme immunoassay for brain S-100 protein and its forensic application. Forensic Sci Int 87:145–154
8. Takahama K (1996) Forensic application of organ-specific antigens. Forensic Sci Int 80:63–69
9. Nichols CA, Sens MA (1990) Recovery and evaluation by cytologic techniques of trace material

retained on bullets. Am J Forensic Med Pathol 11:17–34
10. Fleming RI, Harbison S (2010) The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. Forensic Sci Int Genet 4:244–256
11. Haas C, Klesser B, Maake C, Bar W, Kratzer A (2009) mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. Forensic Sci Int Genet 3:80– 88
12. Hanson E, Haas C, Jucker J, Ballantyne J (2011) Identification of skin in touch/contact forensic samples by messenger RNA profiling. Forensic Sci Int Genetics Supplement Series 3:305–306
13. Hanson E, Haas C, Jucker R, Ballantyne J (2012) Specific and sensitive mRNA biomarkers for the identification of skin in 'touch DNA' evidence. Forensic Sci Int Genet 6:548–558
14. Hanson EK, Lubenow H, Ballantyne J (2009) Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. Anal Biochem 387:303–314
15. Juusola J, Ballantyne J (2007) mRNA profiling for body fluid identification by multiplex quantitative RT-PCR. J Forensic Sci 52:1252–1262
16. Lindenbergh A, de Pagter M, Ramdayal G, Visser M, Zubakov D, Kayser M, Sijen T (2012) A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. Forensic Sci Int Genet 6:565–577
17. Zubakov D, Hanekamp E, Kokshoorn M, van IJcken W, Kayser M (2008) Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of timewise degraded samples. Int J Legal Med 122:135–142
18. Zubakov D, Kokshoorn M, Kloosterman A, Kayser M (2009) New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains. Int J Legal Med 123:71–74
19. Lee HY, Park MJ, Choi A, An JH, Yang WI, Shin KJ (2011) Potential forensic application of DNA methylation profiling to body fluid identification. Int J Legal Med 126:55–62
20. Inoue H, Kimura A, Tuji T (2002) Degradation profile of mRNA in a dead rat body: basic semi-quantification study. Forensic Sci Int 130:127–132
21. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodge CL, Haase J, Janes J, Huss JW III, Su AI (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol 10:R130
22. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden T (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinforma 13:134
23. Flicek P et al (2012) Ensembl 2012. Nucl Acid Res 40:84–90
24. Nicklas JA, Buel E (2006) Simultaneous determination of total human and male DNA using a duplex real-time PCR assay. J Forensic Sci 51:1005–1015
25. Visser M, Zubakov D, Ballantyne KN, Kayser M (2011) mRNAbased skin identification for forensic applications. Int J Legal Med 125:253–263
26. Harteveld J, Lindenbergh A, Sijen T (2013) RNAcell typing andDNA profiling of mixed samples: Can cell types and donors be associated? Sci. Justice, http://dx.doi.org/10.1016/j.scijus.2013.02.001

Chapter 7

Extended specificity studies of mRNA assays used to infer human organ tissues and body fluids

Electrophoresis 2017, 38, 3155–3160

Margreet van den Berge
Titia Sijen

Abstract

Messenger RNA (mRNA) profiling is a technique increasingly applied for the forensic identification of body fluids and skin. More recently, an mRNA-based organ typing assay was developed which allows for the inference of brain, lung, liver, skeletal muscle, heart, kidney, and skin tissue. When applying this organ typing system in forensic casework for the presence of animal, rather than human, tissue is an alternative scenario to be proposed, for instance that bullets carry cell material from a hunting event. Even though mRNA profiling systems are commonly in silico designed to be primate specific, physical testing against other animal species is generally limited. In this study, human specificity of the organ tissue inferring system was assessed against organ tissue RNAs of various animals. Results confirm human specificity of the system, especially when utilizing interpretation rules considering multiple markers per cell type. Besides, we cross-tested our organ and body fluid mRNA assays against the target types covered by the other assay. Marker expression in the nontarget organ tissues and body fluids was observed to a limited extent, which emphasizes the importance of involving the case-specific context of the forensic samples in deciding which mRNA profiling assay to use and when for interpreting results.

Introduction

The use of messenger RNA (mRNA) profiling for the identification of body fluids and skin has been increasingly investigated in the past decades [1-11]. More recently, an mRNA assay was developed which allows for the inference of seven organ tissue types [12-14]. Previously, enzyme-linked immunosorbent assays or histological methods were used to infer the type of cell [15-18]. These methods are, however, often not of sufficient sensitivity to assess the limited amount of cells present on forensic evidentiary traces [19]. The organ typing mRNA assay described in [12], in contrast, allows analysis of low level and degraded samples and can be used to determine multiple tissue types simultaneously [13]. As DNA is simultaneously coextracted alongside RNA, DNA profiles can be generated from the same evidentiary stain to identify the donor of a trace. In forensic settings, this organ typing mRNA assay (referred to as the Organtyper) can be used to aid in the identification of organ tissue remnants present on, for example, evidentiary items of a violent crime and therewith provide information on the course of events. An alternative scenario to be proposed in case of a violent crime is the presence of animal rather than human cells, for instance as the result of a hunting event, a car collision with an animal, or even from food preparation or spilling. Specificity to human tissues can, therefore, be very informative when RNA profiling is applied in forensic context.

In 1977, the DNA genome of bacteriophage PhiX174 [20] was fully sequenced and since then, thousands of organisms (including over 130 mammalian species) have followed [21], enabling comparative genome analyses to in silico assess the potential formation of unintended amplification products from nontarget genes and/or organisms. The amount and the positioning of mismatches at oligonucleotide binding sites are the key in amplification efficiency [22-26]. However, experimental confirmation of human specificity of RNA profiling multiplexes adds to addressing hypotheses in a forensic context, and we subjected organ typing RNA assays to a selection of animal and nontarget organ tissue RNAs. Additionally, we cross-tested our organ and body fluid mRNA assays against the target types covered by the other assay; an experiment inspired by a forensic case involving facial violence (head, nose, and throat area) that led to a request to subject the evidentiary samplings from knives to both organ and body fluid typing. The rationale was that (besides blood and skin) nasal mucosa, saliva, skeletal muscle, and brain could be present. This request made it necessary to cross-test the Organtyper and Cell-typer on body fluids and organs, including organs not targeted by the multiplex.

Materials and methods

Sample collection

A set of 19 animal target organ tissues was used to assess human specificity of organ-typing markers in nonhuman samples. This sample set included commercially available total RNAs (Zyagen) of whole brain for cow, chicken, dog, guinea pig, pig, and rabbit; cerebellum for cat and sheep; lung, liver, skeletal muscle, heart, and kidney for cow and dog; and universal RNA (encompassing brain, heart, kidney, liver, lung, skeletal muscle, pancreas, spleen, stomach, intestine, colon, testis, ovary, and uterus) for chicken. Besides, organ specimen RNAs as described in [12] and the FirstChoice® Human Total RNA Survey Panel (Applied Biosystems) were used; the latter containing pools of total RNA for 20 different human tissues including six of the organs targeted by the Organtyper multiplex [12]. Skin was added to this organ specimen set [12]. Lastly, six human body fluids (*i.e.*, blood, saliva, nasal mucosa, vaginal mucosa, menstrual secretion, and semen) as described in [1] were included (different donors were used throughout the experiments). All necessary approvals from ethical examination commissions at the medical institutions where organs were collected and informed consent of the voluntary donors contributing body fluids were obtained.

RNA profiling

Reverse transcription of the commercial RNAs (animal origin or Human Total RNA Survey Panel) was performed as described in [1]. As the RNA concentrations of these samples were provided, defined RNA inputs to an end concentration of 20 pg/µL were used during reverse transcription. A cDNA input representing 10 pg RNA was used in duplicate PCR amplifications for each animal sample (which is twice the amount of human RNA used to optimize the Organtyper system [12], as we aimed to challenge the multiplex for the animal specimens). A third amplification carried not only animal cDNA (~10 pg RNA), but also human kidney cDNA (~5 pg RNA) that served as a positive amplification control. In all these amplifications, the kidney markers responded. Amplification of the Human Total RNA Survey Panel samples used a cDNA input representing 5 pg RNA.

PCR amplification was performed using Brain-plex, Organtyper [12], and Cell-typer [3]. The Brain-plex is a small multiplex combining four markers specifically aiming for the identification of central nervous system tissue. These include astrocyte (astrocytes constitute nearly half of the human brain [27]) markers GFAP, S100B, ACSBG1, and oligodendrocyte (oligodendrocytes constitute the majority of myelinating cells in the brain [28]) marker OPALIN together with housekeeping marker 18S-rRNA. Brain-plex primers were designed using Ensembl and NCBI primer blast [21, 29], labeled

with a 6-FAM label to obtain the highest sensitivity (Brain-plex primer sequences are provided in Supplementary Table 1).

Minor adjustments were made to the Organtyper [12] and Cell-typer [3] multiplexes. In the Organtyper multiplex brain marker RTN1 was replaced by the more sensitive markers GFAP and OPALIN, which also extended the targeted central nervous system cell types. The primers for heart marker MYBPC3, housekeeping marker 18S-rRNA, skin marker LOR, and blood marker HBB were adjusted to increase annealing temperature and reduce primer–dimer formations. Also, the primer concentrations of lung marker SFTPB and kidney marker UMOD were reduced to improve signal balance. In the Cell-typer multiplex [3], the primers for 11 markers were adjusted to enable an increased annealing temperature and optimize marker spacing. Primer details for these updated Organtyper and Cell-typer multiplexes can be found in Supplementary Table 2 and Supplementary Table 3, respectively.

PCR amplifications involved the PCR protocol as described in [1] with the annealing temperature increased to 64°C. PCR products were purified [1] prior to detection using a 3130XL Genetic Analyzer (Life Technologies). Amplification products were analyzed using POP-4 (Life Technologies) separation matrix and 3 kV, 10 s injection settings. 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 according to the "$x=n/2$" guidelines as described in [30] based on multiple PCR replicates (three) per sample. 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 ($0 < x < n/2$), meaning that no reliable statement is possible. A cell type is scored "not observed" when no peaks are detectedt ($x=0$).

Appropriate controls (*i.e.*, positive controls including samples of known origins, and/or no template controls) were included during each step of the protocol.

In silico assessment of human specificity

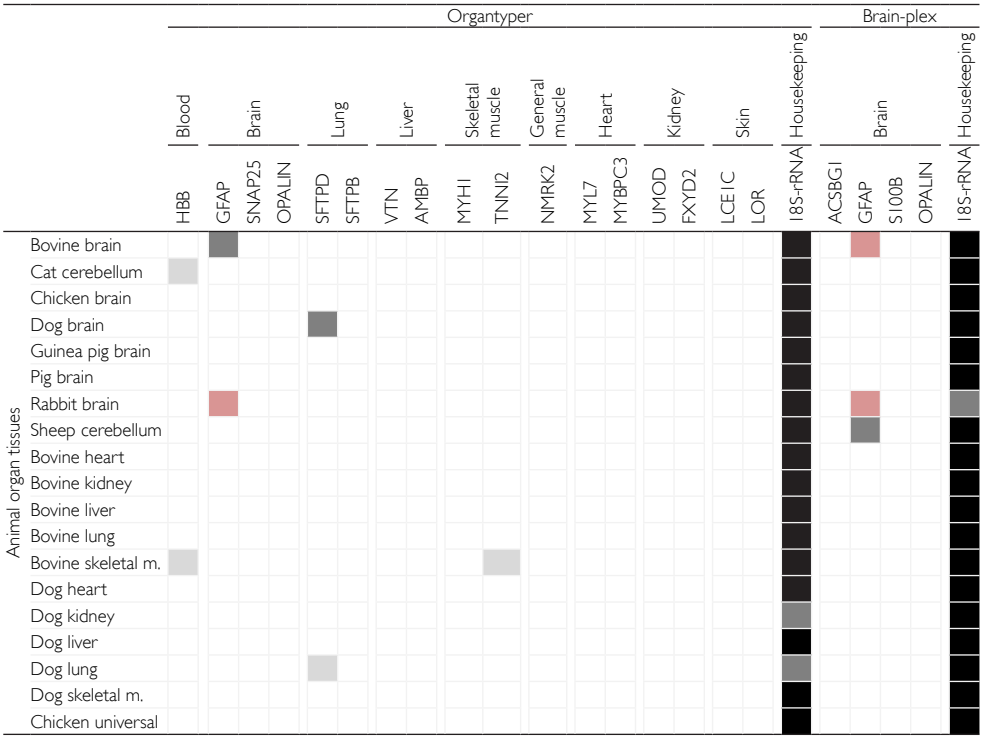
Primer sequences of the Brain-plex, Organtyper, and Cell-typer multiplex (Supplementary Tables 1–3) were subjected to in silico BLAST analyses [21] using primer specificity stringency settings in which a maximum of six mismatches were allowed on any target. RNA sequences for ribosomal RNA marker 18S-rRNA were obtained from the SILVA rRNA database project [31]. A summary of the possible products on animal templates considering cow, cat, chicken, dog, guinea pig, pig, rabbit, and sheep using these databases is shown in Supplementary Table 4.

Results and discussion

Animal organ tissues

Specificity of organ-typing markers (*i.e.*, markers residing in the Organtyper and Brain-plex) for human tissues was tested by using various animal RNAs. Animals were selected for their presence in an average household and included pets (cat, dog, guinea pig, and rabbit) and foods (cow, chicken, pig, and sheep). Results are presented in Table 1. Only few false positive signals (*i.e.*, cross-specific amplifications) were obtained, mainly for astrocyte marker GFAP (specifically with cow, rabbit, and sheep brain RNA, see Table 1). As a positive scoring for an organ tissue requires the detection of at least 50% of all markers for that tissue (section "RNA profiling", "RNA data interpretation"), this cross-reactivity will not result in a false "observed" scoring.

Table 1 Results of specificity tests for the markers residing in the Organtyper and Brain-plex (that are designed to be human specific) using animal organ tissue samples. Different colored fillings indicate the number of times a marker is detected out of three replicate profiles: white when no signals are detected, light grey when detected once, and dark grey when detected twice. Markers detected in all of the three replicates are colored black or red depending on whether marker signals were expected or not (18S-rRNA is not a human-specific marker so signals are to occur).



Housekeeping marker 18S-rRNA was detected in most animal amplifications and is, thus, not a human specific housekeeping marker, which is consistent with the results described in [32]. The main function of this housekeeping marker in these end-point RT-PCR assays is to confirm successful RNA analysis by indicating the presence of RNA, and the absence of factors inhibiting cDNA synthesis or PCR, which is especially informative when none of the other markers give a signal [12]. Few signals were observed for blood marker HBB, lung marker SFTPD, and skeletal muscle marker TNNI2, but these signals were sporadic. Overall, the detection of these marker signals can be explained by the positioning and number of mismatches in the primers when aligned to the animal templates (see Supplementary Table 4 for detailed BLAST results). For example in silico BLAST analyses for astrocyte marker GFAP to cow, rabbit, or sheep show one or two mismatches per primer that reside at least four nucleotides upstream of the 3' end of the primer (Supplementary Table 4A). Species that show more (3' end) mismatches (cat and guinea pig, Supplementary Table 4A) did not result in false positive GFAP signals (Table 1). The Brain-plex results more GFAP false positive responses than the Organtyper (Table 1), which is consistent with the higher sensitivity of a smaller multiplex.

No nonhuman body fluids (other than blood residing in organs) were analyzed as sample collection of vaginal mucosa, nasal mucosa, skin, and menstrual secretion was too complex or impossible and no full picture would be obtained.

Human tissues and case example

Specificity of the Organtyper and Cell-typer multiplex was cross-tested on nontarget human body fluids and organ tissues, including organs not targeted by the multiplex. For the Organtyper, similar tests have previously been performed and described in [12] and these findings were confirmed (Table 2A). Most of the detected Organtyper signals were sporadic marker amplifications seen in only one of the replicates. Skeletal muscle markers were however detected in esophagus, thymus, thyroid, and trachea (all residing in the upper chest/ throat area). Even though the biological function of the skeletal muscle markers in these tissues is unknown, expression of these markers in thyroid tissue has been reported in BioGPS [33]. Additionally, one of the lung markers is detected in thyroid tissue, and both skin markers in vaginal mucosa samples. These results are consistent with findings previously described in [2].

Brain-plex markers GFAP and OPALIN did not result in any unexpected signals when analyzing nontarget organ tissues or body fluids (Table 2). Reproducible ACSBG1 signals are, however, detected in trachea tissue. Additionally, S100B signals are reproducibly detected in adipose, colon, esophagus, small intestine and trachea tissue. S100B expression was reported predominantly in astrocytes [34]. Various studies, however, describe the detection of S100B in non-nervous tissues,

Table 2 Results of specificity tests for the markers residing in the Organtyper, Brain-plex, and Cell-typer multiplex (that are designed to be target-specific) using human organ tissue (A); and human body fluid (B) samples. Different colored fillings indicate the number of times a marker is detected out of three replicate profiles: white when no signals are detected, light grey when detected once, and dark grey when detected twice. Markers detected in all of the three replicates are colored green or red depending on whether markers are detected on target organs, or a result of cross-reacting markers, respectively. Dark red colorings indicate samples for which an "observed" scoring would occur using the "x = n/2" rule, light red when samples would be scored "sporadically observed" which is generally regarded not reliable.

such as adipocytes [35]. Additionally, deregulated expression of S100 family genes in human diseases, especially in cancer studies has been described [36]. Notwithstanding the evident role of S100B in astrocyte functioning, all functions of S100B are not fully clear [37], and explanations regarding the detection of this marker in these nontarget tissues cannot readily be given based on gene expression databases such as BioGPS or GTEx Portal [33, 38]. These findings support the choice of supplementing the original Organtyper multiplex with GFAP and OPALIN (section "RNA profiling"), for which no signals were detected in non-target tissues.

When analyzing the same set of human organ tissues and body fluids with the Cell-typer multiplex, marker signals for vaginal mucosa marker MYOZ1 were additionally detected in skeletal muscle samples. Even though the exact function of MYOZ1 in vaginal mucosa is unknown [7], various studies do describe that the protein encoded by the MYOZ1 (Myozenin-1) gene is predominantly expressed in skeletal muscle [39, 40] and this is also evident in the BioGPS and GTEx Portal gene expression databases [33, 38]. Menstrual secretion marker MMP7 is observed in kidney tissue and seminal fluid marker SEMG1 is detected in bladder tissue, which is consistent with results described in [3, 32, 41]. The detection of menstrual secretion marker MMP11 in placenta, seminal fluid marker SEMG1 in prostate, and spermatozoa marker PRM1 in testes tissue are implied by the association of these tissue types with the body fluids targeted by these markers and have previously been reported [42-44]. The markers indicative for the presence of nasal mucosa (*i.e.* STATH and BPIFA1) are also detected in trachea tissue, and thus appear expressed in multiple tissues (BPIFA1 in nasal, oropharyngeal, and lung epithelia [3, 45], and STATH in salivary glands and trachea [33]). The detection of vaginal mucosa marker MUC4 in trachea tissue, the cross-reactivity of skin markers in various body fluids, and expression of vaginal mucosa markers and white blood cell marker CD93 in nasal mucosa samples has previously been reported (Table 2) [2, 3, 46]. Although skin marker LCE1C is common in the Organtyper and Cell-typer, results for this marker on human body fluid samples were not identical (Table 2B). This is likely cause by the use of different body fluid donors, and emphasized the variability that may be seen between donors especially for cross-reactive signals.

When considering the false positive responses not per marker but for all markers indicating a body fluid or organ jointly by the use of the "$x=n/2$" interpretation guideline, trachea seems prone to false positive scorings while especially skeletal muscle markers may provoke false positive results (Table 2). In the context of the above-described case in which Organtyper and Cell-typer were requested for facial injuries, we discuss the following points of attention: (1) the MYOZ1 signals in the Cell-typer multiplex should be regarded both in context with the two vaginal mucosa markers and the two skeletal muscle markers in the Organtyper multiplex; in case skeletal muscle markers MYH1 and TNNI2 respond and vaginal mucosa markers CYP2B7P1 and MUC4 not, MYOZ1 signals may derive in the presence of skeletal muscle and MYOZ1 is best excluded from

the inference for vaginal mucosa; (2) when trachea may be injured, the presence of nasal mucosa and skeletal muscle is unsure; and (3) when multiple organs in the upper chest/ facial area are violated, one should be extremely careful to infer presence of skeletal muscle and nasal and vaginal mucosa.

Concluding remarks

In this study, specificity analysis of a set of mRNA markers used for the inference of body fluids and organ tissues was performed to gain insight and awareness regarding the possibilities and limitations of forensic mRNA profiling. Results stress the importance of using multiple markers per cell type, as single markers may be detected in nontarget tissues and the use of increased number of markers per cell type will allow inference with more certainty. Additionally, awareness of unintended marker detection in nontarget tissues is crucial when applying mRNA profiling in casework, as this may lead to incorrect associations. Results of this study emphasize the importance of furthering the knowledge and understanding of mRNA markers when applied to forensic casework.

Acknowledgements

The authors are grateful to all donors from whom samples have been used in this study. 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 Table 1 Primer sequences for the markers residing in the Brain-plex.

Marker name	Tissue	[primer] µM	Forward primer (5'-3') Reverse primer (5'-3')	Size (bp)	Dye
GFAP	Astrocyte	0.04	TGGAGAGGAAGATTGAGTCGCTGG CGAACCTCCTCCTCGTGGATCTTC	70	6FAM™
OPALIN	Oligodendrocyte	0.8	GCCATGGAGGAAAGTGACAGACC CTCATGTGTGGGTGATCTCCTAGG	93	6FAM™
ACSBG1	Astrocyte	0.2	CTACACTTCCGGCACCACTGG GTCCACGTGATATTGTCTTGACTCAG	66	6FAM™
S100B	Astrocyte	0.3	GCAGCAAGGAGACCAGGAAGG AACGTCGATGAGGGCCACCATG	81	6FAM™
18S-rRNA	Housekeeping	0.013	GACTCAACACGGGAAACCTCACC CTCCACCAACTAAGAACGGCCATG	110	PET™

Supplementary Table 2 Primer sequences for the markers residing in the Organtyper multiplex.

Gene	Tissue	[Primer] µM	Forward primer (5'-3') Reverse primer (5'-3')	Size (bp)	Dye
SNAP25	Brain	0.02	TCGGCGGCTCCACCACAGTT TTGGCTCTGGACCTGGGCTTCTC	78	6FAM™
GFAP	Brain	0.04	TGGAGAGGAAGATTGAGTCGCTGG CGAACCTCCTCCTCGTGGATCTTC	70	6FAM™
OPALIN	Brain	0.8	GCCATGGAGGAAAGTGACAGACC CTCATGTGTGGGTGATCTCCTAGG	93	6FAM™
STFTB	Lung	0.04	CCATGATTCCCAAGGGTGCGCTA CGCCACCAGAGGTACCACGC	68	NED™
SFTPD	Lung	0.10	CCTGCCTGGTCGCGATGGAC CCAGGCATCCCTGCTTGCCC	99	6FAM™
AMBP	Liver	0.03	TTGGCTGACCGAGGTGAATGTGTCC[a] ACCAGTTGCCCACCCCCTGAT	119	VIC®
VTN	Liver	0.05	GTGCAAGCCCCAAGTGACTCG CATAGACCGTGTACTCATCCTCCG	65	VIC®
TNNI2	Skeletal muscle	0.13	ATGTCTGAAGTGCAGGAGCTCTGC GTCGTACTTCTCCTCTTCAGCCGC	72	PET™
MYH1	Skeletal muscle	0.8	TTCGCATCTCTACGCCAGGGTCCTTA[a] AGGAAAGGAGCAGCCTCCCCAAA	104	NED™
MYL7	Heart	0.08	AAGCTCAATGGGACAGACCCCG CACCACCCCTTTGCCGCTGG	78	VIC®
MYBPC3	Heart	0.4	AGGCAGAGAAGGCAGAGCCCAT AGCTTGACCCTTTGGGACTTGGG	129	VIC®
NMRK2	General muscle	0.4	GGAGGAGAAGTACCCGCAACTACA ACACCGTTGGCCTCCATCTCCT	109	6FAM™
UMOD	Kidney	0.06	ATCACACGGAAAGGTGTCCAGGC TTTTTTGGAGCACAGGGCTTTCCGC[a]	145	6FAM™
FXYD2	Kidney	0.06	CTCCATCCAGGCCCCAGGCA CTCTCCAAAAAGCAGAGACAGCAGG CGGTCTCATAGTCATAGTAGAACGGG	143	PET™
LCE1C	Skin	0.2	TGTGACCCCGCTCCTGAATCCG CTTGGGAGGGCACTTGGGGGTG	99	NED™
LOR	Skin	0.6	TGCTTTGGGCTCTCCTTCCTTCTC GGTCTTCACGCAGTCCACTGG	88	PET™
18S-rRNA	Housekeeping	0.013	GACTCAACACGGGAAACCTCACC CTCCACCAACTAAGAACGGCCATG	110	PET™
HBB	Blood	0.1	GCACGTGGATCCTGAGAACTTCAG ATGGGCCAGCACACAGACCAG	61	6FAM™

[a] Underlined nucleotides are 5' tails added to improve multiplex spacing.

Supplementary Table 3 Primer sequences for the markers residing in the Cell-typer multiplex.

Gene	Tissue	[Primer] µM	Forward primer (5'-3') Reverse primer (5'-3')	Size (bp)	Dye
HBB[a]	Blood	0.035	GCACGTGGATCCTGAGAACTTCAG ATGGGCCAGCACACAGACCAG	61	6FAM™
CD93[a]	Blood	0.075	GCTCTGGGGCTACTGGTCTATC TCCCAGGTGTCGGACTGTACTG	151	NED™
ALAS2	Blood	0.05	TTCTGCACCAGAAGGACTCAGCC[b] TAAATCTCGCACCCTGGCAGGATC	103	6FAM™
HTN3[a]	Saliva	0.2	CTTCACTTCAGCTTCACTGACTTCTG CTTTGCATGTGAATCAGCTCCAGTC	132	VIC®
STATH[a]	Saliva/ nasal mucosa	0.3	TTCATCTTGGCTCTCATGGTTTCCATG GCCATACCCATAACCGAATCTTCCA	93	6FAM™
BPIFA1	Nasal mucosa	0.15	CAAGTGAATACGCCCCTGGTCG GAATGGGTGCAGTCACCAAGGAC	131	PET™
SEMG1	Seminal fluid	0.6	GGAAGATGACAGTGATCGT CAACTGACACCTTGATATTGG	121	6FAM™
PRM1[a]	Spermatozoa	0.05	GAGAGCCATGAGGTGCTGCC AGGCAGGAGTTTGGTGGATGTGC	90	NED™
KLK3	Seminal fluid	0.05	GACGTGGATTGGTGCTGCACC CTTCTCGCACTCCCAGCCTC	64	PET™
MUC4[a]	Vaginal mucosa	0.8	CTGCTACAATCAAGGCCACTGCTAC AAGGGAAGTTCTAGGTTGACAGTTGG	141	6FAM™
CYP2B7P1[a]	Vaginal mucosa	0.05	CCTCATGTCGCAGAGAGAGTCTAC CCCATGGGGAGAAGGTCAGCA	146	VIC®
MYOZ1[a]	Vaginal mucosa	0.8	CGTGTTCTCCGGTCACAGCAG TGGATTCAGCCGGCTGCTCG	88	VIC®
MMP7	Menstrual secretion	0.4	GAACAGGCTCAGGACTATCTC TTAACATTCCAGTTATAGGTAGGCC[b]	127	VIC®
MMP10	Menstrual secretion	0.07	GCATCTTGCATTCCTTGTGCTGTTG GGTATTGCTGGGCAAGATCCTTGTT	107	VIC®
MMP11	Menstrual secretion	0.15	CAACCGACAGAAGAGGTTCG GAACCGAAGGATCCTGTAGG	76	NED™
CDSN[a]	Skin	0.05	TCCTCCTGCCAGGGACCTTG GATACGCGTGGGGTCCTTACAG	71	VIC®
LCE1C	Skin	0.03	TGTGACCCCGCTCCTGAATCCG CTTGGGAGGGCACTTGGGGGTG	99	NED™
18S-rRNA[a]	Housekeeping	0.013	GACTCAACACGGGAAACCTCACC CTCCACCAACTAAGAACGGCCATG	110	PET™
ACTB[a]	Housekeeping	0.045	CAGAGCCTCGCCTTTGCCGAT CGCGGCGATATCATCATCCATGGT	75	PET™

[a] Primer concentrations lowered compared to [3].
[b] Underlined nucleotides are 5' tails added to improve multiplex spacing.

Supplementary Table 4 BLAST results for the primers included in the Organtyper and Brain-plex (A), and the Cell-typer multiplex (B) when considering various animals as target. Shown are the locations at which mismatches occur, numbered in 5'-3' direction. To infer the distance from the 3' end of the primer, the primer length is indicated as well. Housekeeping marker 18S-rRNA resulted in full primer-template matches for all animals considered (not shown). Markers and animals without comments did not result in hits when considering the primer BLAST specificity stringency settings described in section "In silico assessment of human specificity". Colour coding for Organtyper and Brain-plex (A) correspond to results presented in Table 1, which indicates the number of times a marker is detected out of three replicate profiles: white when no signals are detected, light grey when detected once and dark grey when detected twice. Markers detected in all of the three replicates are coloured red. Indicated in the dotted square are samples with possible non-specific product formation, which were not assessed in this study as RNA samples for these organs were not included.

	Tissue	Gene	Primer length (bp)	Bos taurus Bovine	Canis lupus familiaris Dog	Gallus gallus Chicken	Felis catus Cat	Cavia porcellus Guinea pig	Sus scrofa Pig	Oryctolagus cuniculus Rabbit	Ovis aries Sheep
A	Blood	HBB	F 24			10T>C 13T>C*	13T>C				
			R 21			9G>C 13A>G 15A>T*	1A>G 16G>C				
	Brain	GFAP	F 24	20G>T			17G>A 20G>T	3G>A 8G>A 11G>A 20G>T		20G>C	20G>T
			R 24	3A>C 15G>A			1C>T 3A>C 15G>A	1C>T 3A>C 15G>A		3A>C 20T>C	3A>C 15G>A
		ACSBG1	F 21	7T>G 10C>G 19T>C	4C>T 7T>C		7T>C 10C>G 19T>C		7T>A 10C>G	7T>C 10C>G	
			R 26	18T>C	6C>T 18T>C		12A>G 18T>C		6C>T	12A>G 21A>G	
	Lung	SFTPD	F 20						13C>T 19A>T		
			R 20						12T>C 16T>C		
Organtyper	Skeletal muscle	TNNI2	F 24		6T>C		24C>T		6T>G 9A>G 12G>C	4T>G 6T>C 9A>G 21C>G	
			R 24	4G>A 16T>C 19A>G	4G>A 16T>C 19A>G 22C>T		4G>A 16T>C 19A>G	16T>C 19A>G	16T>C 19A>C	4G>A 16T>C 19A>G	4G>A 16T>C 19A>G
	General muscle	NMRK2	F 24	13C>A 14C>G 17C>A 19A>G 24A>G		8A>G 14C>A 19A>G					13C>G 14C>G 17C>A 19A>G 24A>G
			R 22	22T>C		3A>C 7T>C 10G>T					22T>C
	Heart	MYL7	F 22		15A>G 21C>G				9T>C		
			R 20		7C>A 10T>C				7C>G 10T>C		7C>G 16G>A
		MYBPC3	F 22							3G>T 13T>C 15G>A	
			R 23								
	Kidney	FXYD2	F 25	1C>G 7A>C 8A>C 18A>G 23A>G					7A>C 8A>C 9A>G 10A>C 18A>G		
			R 26	8A>G 23C>T					2G>T 14A>G 20G>A 23C>T		
B	Blood	HBB	F 24			10T>C 13T>C*	13T>C				
			R 21			9G>C 13A>G 15A>T*	1A>G 16G>C				
		CD93	F 22				6G>T 9G>A 12A>T			15G>T 21T>C	
			R 22				5A>T 8T>C 11C>T			8T>C 11C>G 20C>T	
		ALAS2	F 21				9G>A 12G>A 16T>G	12G>A 15C>T	1C>T 15C>T	15C>T	
			R 24	9G>A 20G>A			7T>G 8C>T	6C>T 9C>A	9G>A	12C>A	
	Nasal mucosa	BPIFA1	F 22				1C>A 3A>T			1C>A 2A>G 9T>C	
			R 23				3A>G 6G>A 21G>A			2A>G 3A>G 20G>T	
Cell-typer	Vaginal mucosa	MYOZ1	F 21		3T>A			2G>A 3T>G			3T>G 10C>T 11G>A
			R 20		1T>C 2G>A 9G>A 20G>A			1T>A 2G>T 4A>G 20G>A			6T>G 7C>G 20G>A
	Menstrual secretion	MMP7	F 21						14A>G		
			R 25						4A>G 20T>C 25C>G	6A>C 20T>C	
		MMP10	F 25							1G>C 9C>T 10A>C 11T>C 23T>C	
			R 25							12G>A 17A>T 21T>A	
		MMP11	F 20	14A>C						14A>C	
			R 20	7A>G						7A>C	7A>G
	Skin	CDSN	F 20	10C>A					10C>A	14G>A 19T>G	10C>A
			R 22	1G>A 7C>T 19A>G				4A>T 19A>G	1G>A 19A>G	4A>C 7C>T	7C>T 19A>G

*The product length of amplification using HBB markers to Gallus gallus template deviates from the expected product length (130 vs 61 bp).

References

1. Lindenbergh A, de Pagter M, Ramdayal G, Visser M, Zubakov D, Kayser M, Sijen T, 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. van den Berge M, Carracedo A, Gomes I, Graham EAM, Haas C, Hjort B, Hoff-Olsen P, Maronas O, Mevag B, Morling N, Niederstätter H, Parson W, Schneider PM, Syndercombe Court D, Vidaki A and Sijen T. 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. 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 Sci. Int.: Genet. 20 (2016) 119-129.
4. Fleming RI, Harbison S, The use of bacteria for the identification of vaginal secretions, Forensic Sci. Int.: Genet. 4 (2010) 311–315.
5. Gomes I, Kohlmeier F, Schneider PM, Genetic markers for body fluid and tissue identification in forensics, Forensic Sci. Int.: Genet. Suppl. Ser. 3 (2011) e469– e470.
6. Haas C, Klesser B, Maake C, Bar W, Kratzer A, mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR, Forensic Sci. Int. Genet. 3 (2009) 80–88.
7. Hanson EK, Ballantyne J, Highly specific mRNA biomarkers for the identification of vaginal secretions in sexual assault investigations, Sci. Justice 53 (2013) 14–22.
8. Juusola J, Ballantyne J, Multiplex mRNA profiling for the identification of body fluids, Forensic Sci. Int. 152 (2005) 1-12.
9. Roeder AD, Haas C, 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. Sijen T, Molecular approaches for forensic cell type identification: On mRNA, miRNA, DNA methylation and microbial markers, Forensic Sci. Int.: Genet. (2015).
11. Harteveld J, Lindenbergh A, Sijen T, RNA cell typing and DNA profiling of mixed samples: can cell types and donors be associated, Sci. Justice 53 (2013) 261–269.
12. Lindenbergh A, van den Berge M, Oostra RJ, Cleypool C, Bruggink A, Kloosterman A, Sijen T, Development of a mRNA profiling multiplex for the inference of organ tissues, Int. J. Legal Med. 127 (2013) 891–900.
13. van den Berge M, Wiskerke D, Gerretsen R, Tabak J, Sijen T, DNA and RNA profiling of excavated human remains with varying postmortem intervals. Int. J. Legal Med. 130 (2016) 1471-1480.
14. van den Berge M, Sijen T, A male and female RNA marker to infer sex in forensic analysis. Forensic Sci. Int.: Genet. 26 (2017) 70-76.
15. Takata T, Miyaishi S, Kitao T, Ishizu H, Identification of human brain from a tissue fragment by detection of eurofilament proteins. Forensic Sci Int 144 (2004) 1–6.
16. Kimura A, Ikeda H, Yasuda S, Yamaguchi K, Tsuji T. Brain tissue identification based on myosin heavy chain isoforms. Int J Legal Med 107 (1995) 193–196.
17. Seo Y, Kakizaki E, Takahama K, A sandwich enzyme immunoassay for brain S-100 protein and its forensic application. Forensic Sci Int 87 (1997) 145–154.
18. Takahama K, Forensic application of organ-specific antigens. Forensic Sci Int 80 (1996) 63–69.
19. Nichols CA, Sens MA, Recovery and evaluation by cytologic techniques of trace material retained on bullets. Am J Forensic Med Pathol 11 (1990) 17–34.
20. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, Hutchison CA, Slocombes PM, Smith M, Nucleotide sequence of bacteriophage phi X174 DNA, Nature 265 (1977) 687 – 695.
21. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL, Primer-BLAST: a tool to design

target-specific primers for polymerase chain reaction, BMC Bioinf. 13 (2012) 134.
22. Whiley DM, Sloots TP, Sequence variation in primer targets affects the accuracy of viral quantitative PCR. J Clin Virol. 2005, 34 (2)
23. Sipos R, Szekely AJ, Palatinszky M, Revesz S, Marialigeti K, Nikolausz M, Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. FEMS Microbiol Ecol. 2007, 60 (2)
24. Waterfall CM, Eisenthal R, Cobb BD, Kinetic characterisation of primer mismatches in allele-specific PCR: a quantitative assessment. Biochem Biophys Res Commun. 2002, 299 (5)
25. Ghedira R, Papazova N, Vuylsteke M, Ruttink T, Taverniers I, De Loose M, Assessment of primer/template mismatch effects on real-time PCR amplification of target taxa for GMO quantification. J Agric Food Chem. 2009, 57 (20)
26. Stadhouders R, Pas SD, Anber J, Voermans J, Mes TH, Schutten M, The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay. The Journal of Molecular Diagnostics, 12 (2010) 109-117.
27. Cahoy, John D., et al. "A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function." The Journal of Neuroscience 28.1 (2008): 264-278.
28. Sallis ESV, Mazzanti CM, Mazzanti A, Pereira LAV, Arroteia KF, Fustigatto R, Pelizzaril C, Rodriguesl A, Graça DL, OSP-Immunofluorescent remyelinating oligodendrocytes in the brainstem of toxically-demyelinated Wistar rats. Arquivos de neuro-psiquiatria, 64 (2006) 240-244.
29. Cunningham F, et al., Ensembl 2015, Nucleic Acids Res. 43 (2015) D662-D669
30. Lindenbergh A, Maaskant P, Sijen T, Implementation of RNA profiling in forensic casework, Forensic Sci. Int. Genet. 7 (2013) 159–166.
31. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Acids Res. 41 (2013): D590-D596.
32. Xu Y, Xie J, Cao Y, Zhou H, Ping Y, Chen L, Gu L, Hu W, Bi G, Ge J, Development of highly sensitive and specific mRNA multiplex system (XCYR1) for forensic human body fluids and tissues identification, PLoS One 9 (2014) e100123.
33. Wu C, Jin X, Tsueng G, Afrasiabi C, Su AI, BioGPS: building your own mash-up of gene annotations and expression profiles. Nucl. Acids Res. 44 (2016) D313-D316.
34. Gonçalves CA, Leite MC, Guerra MC, Adipocytes as an important source of serum S100B and possible roles of this protein in adipose tissue. Cardiovasc Psychiatry Neurol. 2010, 2010: 790431
35. Michetti F, Dell'Anna E, Tiberio G, Cocchia D, Immunochemical and immunocytochemical study of S-100 protein in rat adipocytes. Brain Research, 262 (1983) 352–356.
36. Ji J, Zhao L, Wang X, Zhou C, Ding F, Su L, Zhang C, Mao X, Wu M, Liu Z, Differential expression of S100 gene family in human esophageal squamous cell carcinoma. Journal of cancer research and clinical oncology, 130 (2004) 480-486.
37. Teratani T, Watanabe T, Kuwahara F, Kumagai H, Kobayashi S, Aoki U, Ishikawa A, Arai K, Nozawa R, Induced transcriptional expression of calcium-binding protein S100A1 and S100A10 genes in human renal cell carcinoma. Cancer letters, 175 (2002) 71-77.
38. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, The genotype-tissue expression (GTEx) project, Nat. Genet. 45 (2013) 580–585.
39. Braun T, Gautel M, Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. Nat Rev Mol Cell Biol, 12 (2011) 349-361.
40. Takada F, Vander Woude DL, Tong HQ, Thompson TG, Watkins SC, Kunkel LM, Beggs AH, Myozenin: an alpha-actinin-and gamma-filamin-binding protein of skeletal muscle Z lines. Proc. Natl. Acad. Sci.

98 (2001) 1595-1600.
41. Fang R, Manohar CF, Shulse C, Brevnov M, Wong A, Petrauskene OV, Brzoska P, Furtado MR, Real-time PCR assays for the detection of tissue and body fluid specific mRNAs. In International Congress Series 1288 (2006) 685-687.
42. Basset P, Bellocq JP, Wolf C, Stoll I, Hutin P, Limacher JM, Podhajcer OL, Chenard MP, Rio MC, Chambon P, A novel metalloproteinase gene specifically expressed in stromal cells of breast carcinomas. Nature 348 (1990) 699 – 704.
43. Lundwall Å, Bjartell A, Olsson AY, Malm J, Semenogelin I and II, the predominant human seminal plasma proteins, are also expressed in non-genital tissues. Molecular human reproduction, 8 (2002) 805-810.
44. Wykes SM, Nelson JE, Visscher DW, Djakiew D, Krawetz SA, Coordinate expression of the PRM1, PRM2, and TNP2 multigene locus in human testis. DNA and cell biology, 14 (1995) 155-161.
45. Liu Y, Bartlett JA, Di ME, Bomberger JM, Chan YR, Gakhar L, Mallampalli RK, McCray PB Jr., Di YD, SPLUNC1/BPIFA1 contributes to pulmonary host defense against Klebsiella pneumoniae respiratory infection, Am J Pathol, 182 (2013), pp. 1519–1531
46. Cossu C, Germann U, Kratzer U, Baer W, Haas C, How specific are the vaginal secretion mRNA-markers HBD1 and MUC4, Forensic Sci. Int.: Genet. Suppl. Ser. 2 (2009) 536–537.

Chapter 8

DNA and RNA profiling of excavated human remains with varying postmortem intervals

International Journal of Legal Medicine 130 (2016) 1471-1480

Margreet van den Berge
Demi Wiskerke
Reza Gerretsen
Jonathan Tabak
Titia Sijen

Abstract

When postmortem intervals (PMIs) increase such as with longer burial times, human remains suffer increasingly from the taphonomic effects of decomposition processes such as autolysis and putrefaction. In this study, various DNA analysis techniques and a messenger RNA (mRNA) profiling method were applied to examine for trends in nucleic acid degradation and the postmortem interval. The DNA analysis techniques include highly sensitive DNA quantitation (with and without degradation index), standard and low template STR profiling, insertion and null alleles (INNUL) of retrotransposable elements typing and mitochondrial DNA profiling. The used mRNA profiling system targets genes with tissue specific expression for seven human organs as reported by Lindenbergh et al. (Int J Legal Med 127:891-900, [27]) and has been applied to forensic evidentiary traces but not to excavated tissues. The techniques were applied to a total of 81 brain, lung, liver, skeletal muscle, heart, kidney and skin samples obtained from 19 excavated graves with burial times ranging from 4 to 42 years. Results show that brain and heart are the organs in which both DNA and RNA remain remarkably stable, notwithstanding long PMIs. The other organ tissues either show poor overall profiling results or vary for DNA and RNA profiling success, with sometimes DNA and other times RNA profiling being more successful. No straightforward relations were observed between nucleic acid profiling results and the PMI. This study shows that not only DNA but also RNA molecules can be remarkably stable and used for profiling of long-buried human remains, which corroborate forensic applications. The insight that the brain and heart tissues tend to provide the best profiling results may change sampling policies in identification cases of degrading cadavers.

Introduction

The determination of a postmortem interval (PMI; also known as time since death (TSD)) is an important aspect in forensic death investigations. Traditionally, the PMI is estimated based on various physical and biochemical changes occurring shortly after death, such as rigor, algor and livor mortis [1, 2]. These factors can, however, only be used for estimating relatively short PMIs with a wide window of estimation [2, 3]. More recently, studies have been performed to investigate postmortem stability of nucleic acids and search for trends relating nucleic acid degradation and PMI [4–15]. The analysis of hard tissue such as the bone and dental tissues are described to be more favourable for DNA analysis due to the relatively higher stability of nucleic acids in these tissue types [11]. Processing of these types of samples, however, is time consuming and labour intensive, which makes the analysis of soft tissue favourable [11]. Especially soft postmortem tissues are affected by decomposition processes such as autolysis and putrefaction [2, 15]. The rate of decomposition is known to be influenced by external variables such as type of clothing, presence of a body bag and/or coffin, soil type, burial depth, water context, ambient temperature, weather conditions, air circulation, accessibility to insects, body mass index (BMI) of the deceased, agonal state and microbiome composition [5, 16–18]. Moreover, it is described that different tissue types may be affected differently by these degradation processes, as tissues are differentially shielded from the external factors or microbial sources and because enzymes tend to be more active in tissues such as the kidney and liver, resulting in putrefaction, thus early DNA degradation in these tissue cells [3].

DNA profiling of postmortem tissues mainly aims for genetic identification through short tandem repeat (STR) profiling [3–5, 10, 11, 14]. Besides, messenger RNA (mRNA) profiling has become increasingly utilized in forensics in the past few decades [19]. mRNA profiling uses markers targeting mRNA transcripts of gene combinations predominantly expressed in, and thus characteristic for, specific cells or conditions. This technique is commonly applied for the identification of body fluids [20–26], and more recently years also for the inference of organ tissue type [27]. Few studies have been performed to investigate the applicability of RNA profiling in postmortem human tissues to investigate RNA degradation as a possible indicator of the postmortem interval [18, 28–31] or to determine the cause of death [32]. In this current study, we apply various DNA analysis techniques (*i.e.* DNA quantification [33–35], STR profiling, InnoTyper profiling [35, 36], mitochondrial DNA (mtDNA) SNaPshot [37]) and mRNA profiling to search for trends relating nucleic acid degradation and PMI in 81 exhumed human organ tissues (*i.e.* brain, lung, liver, skeletal muscle, heart, kidney and skin samples) with burial times/PMIs ranging from 4 to over 42 years.

Materials and methods

Sample collection

The organ specimens used in this study were collected during exhumations at two cemeteries in The Netherlands. The reason for evacuation of the graves was redevelopment of the cemetery. In The Netherlands, a space for a grave is made available for 15 or 20 years. Exhumations were as in forensic setting, performed by a forensic anthropologist MD and a forensic archaeologist with necessary approvals from the required ethical examination commissions and the municipality. In total, 81 organ tissue samples comprising brain, lung, liver, skeletal muscle, heart, kidney and skin were collected from 19 exhumed human remains. Organ tissues were collected in plastic storage containers and stored at −80 °C for 7.5 years until processed. Due to the varying states of decomposition of the bodies, it was not always possible to find and thus collect all organ types from all bodies. Burial times of the donors, ranged between 4 and over 42 years. Additional information regarding the different donors, such as burial times, sex and age, are provided in Table 1.

For three of the donors, the burial times are unknown, and it is expected that these samples range between 23 and 42 years, as this is the trend for the remaining donors buried at this cemetery. Table 1 additionally includes details regarding the state of decomposition according to the decomposition Staging Scale [38] and the condition of the human remains. The decomposition scale uses scores ranging from 1 to 10, in which 1 represents a complete fresh body and 10 represents a completely skeletonized body. The condition of the preservation of the skeleton is rated from 1 to 4, in which 1 represents whole and visually undamaged skeletal elements and 4 represents skeletal elements that are reduced to a powdery substance.

DNA/RNA extraction, DNA quantification

In preparation for DNA/RNA co-extraction, small tissue sections (approximate size 3 mm3) were excised from the centre of a tissue sample. These sections were finely cut and transferred to 2-mL tubes to improve cell lysis. The excised tissue sections were weighed prior to isolation in order to determine the DNA concentration per gram of crude tissue for each sample. After excision, the tissues immediately underwent cell lysis according to the protocol described in Ref. [20] by using 600 µL lysis binding buffer (Ambion), 40 µL proteinase K (20 mg/mL, QIAGEN) and incubating at 56 °C for 2 h. After 30 min of incubation, an additional 40 µL proteinase K was added to samples that seemed not completely lysed. After cell lysis, lysates were stored at −80 °C until further processing according to the DNA/RNA extraction as described in Ref. [20], in which residuals of tissue debris are removed by the use of QIAshredder columns.

Table 1 Information regarding burial and exhumation conditions for the 19 donors from whom samples were collected for this study.

										Tissues included						
Donor	Burial ground	Soil type	# years buried	Gender	Age at death	Stage of decomposition[a]	Condition skeleton	Depth interment (cm)	Additional information	Brain	Heart	Kidney	Lung	Skeletal muscle	Liver	Skin
1	A	Sand	4y6m[b]	M	82y3m	7	1	80	Water in coffin	✓	✓	✓	✓	✓	✓	
2			4y6m	M	83y10m	7	1	80	Water in coffin	✓	✓	✓	✓	✓	✓	
3			4y10m	M	84y3m	7	1	80	Water in coffin	✓	✓	✓	✓	✓	✓	
4			5y2m	M	83y10m	7	1	80	Water in coffin and mould growth on waterline	✓	✓	✓	✓	✓	✓	
5			5y4m	M	88y2m	7	1	80	Water in coffin	✓	✓	✓	✓	✓	✓	
6			5y6m	M	80y4m	7	1	80	Water in coffin	✓	✓	✓	✓	✓	✓	
7			6y11m	M	79y5m	7	1	80	Water in coffin	✓	✓	✓	✓	✓		
8	B	Sandy loams with gravel inclusions	23y9m	M	-	8	2	110						✓		✓
9			23y9m	F	-	9	2.5	220								✓
10			26y6m	F	-	8	1	120		✓				✓		
11			30y4m	F	-	8	2	120	Underwent autopsy prior to burial	✓	✓		✓	✓		✓
12			36y6m	M	-	8	1	100		✓	✓	✓	✓	✓	✓	✓
13			37y3m	M	-	8	1	180		✓			✓	✓	✓	
14			39y	-	-	8/9	2	180						✓		
15			40y4m	F	-	7/8	1/2	120						✓		✓
16			42y4m	F	-	8	3	120								✓
17			-	M	-	7/8	1	180		✓	✓	✓	✓	✓	✓	✓
18			-	F	-	7/8	1	80						✓		✓
19			-	F	-	7	1	80		✓	✓	✓		✓	✓	✓
									Number of samples per tissue type	13	11	10	11	17	10	9

[a]According to the Decomposition Staging Scale described in Ref. [38].
[b]'y' representing the number of years, 'm' representing the number of months.

Small adjustments were made to the protocol described in Ref. [20] regarding elution volumes, which were reduced to 50 µL for DNA and 40 µL for RNA extracts (compared to 100 and 60 µL, respectively). DNase treatment and DNA quantification (*Alu* assay [33]) were performed, according to protocols described in Ref. [20]. Besides the *Alu* assay quantification system, an additional quantification was performed using the InnoQuant™ Kit (InnoGenomics). The InnoQuant system is one of the more novel quantification systems which, in addition to a total human DNA quantification, provides degradation and inhibition analysis [34, 35, 39]. The system can accurately measure the DNA quantity in a sample down to 1 pg/µL (compared to 0.5 pg/µL for the total DNA marker and 4 pg/µL for the male DNA indicator in the *Alu* assay, as determined in house), and DNA concentrations are determined by amplification of two separate high copy number nuclear DNA targets, *i.e.* a short (80 bp) *Alu* target and a long (207

bp) SVA target. A degradation index (DI) indicating the degree of DNA degradation in a sample is determined by calculating the ratio between the quantitative values of the short target and the long target (DI = [short/long]. A DI value of one implies no degradation in a sample; increased DI values imply degradation. A synthetic target is simultaneously amplified as internal positive control (IPC) to assess for the presence of PCR inhibitors. The IPC is an indicative of a PCR inhibitor when cycle threshold (Ct) values of the IPC in a sample exceed the mean IPC Ct value for the dilution standards by more than two units [34, 35]. InnoQuant quantification was performed according to manufacturer's instructions using 2 µL DNA extract per reaction during a 32-cycle PCR on a 7500 Real-Time PCR system (Applied Biosystems). Data analysis was performed using SDS v2.3 with a Ct set as described by manufacturer's instructions [34]. Maximal PCR efficiency (100 % efficiency equals a slope of −3.3) was acquired by applying empirically established baseline settings (3–11 for the short fragment, 3–12 for the long fragment and 3–15 for the IPC). When an indication for inhibition was obtained (Ct values above 34 for the *Alu* assay or based on IPC for the InnoQuant), tenfold dilutions were prepared and submitted to quantification.

STR profiling

STR amplification was performed on all 81 DNA extracts using the AmpFℓSTR® NGM™ PCR Amplification Kit (Life Technologies) during a 29-cycle PCR and a maximum of 500 pg DNA input (or 10 µL) based on quantification results as described in [20]. For samples with low DNA concentrations (*i.e.* below 0.05 ng/µL), the maximum input of 10 µL DNA extract per reaction was used. PCR products were separated according to standardized protocols [20] using a 3130XL Genetic Analyser (Life Technologies) with POP-4 (Life Technologies) separation matrix and 3 kV, 15 s injection settings. Profile analysis was performed using Genemapper ID-X version 1.1.1 (Life Technologies), a detection threshold of 50 relative fluorescence units (rfu) and locus-specific stutter ratio thresholds as described in Ref. [40]. Samples resulting in incomplete profiles with less than eight detected alleles were subjected to +5-cycle NGM PCR amplification [40]. Amplification products were separated and analysed as described above using 1.5× stutter filter thresholds [40]. Samples resulting in profiles with less than eight detected alleles in the 29 + 5-cycle STR profiles were subjected to mtDNA analysis using a SNaPshot assay [37]. The percentage of detected alleles in the STR profiles was determined considering the 15 STR loci in the NGM kit (Amelogenin excluded). When no reference profile could be deduced from the profile set for a donor (for instance when none of the tissues presented a full profile), single allele calls were regarded as heterozygote alleles accompanied with a drop-out. Degradation rates of the STR profiles were calculated based on ski slope measurements by comparing the average peak height of long (>225 bp) versus short (<225 bp) alleles.

InnoTyper profiling

InnoTyper profiling was performed for all 81 DNA extracts using the InnoTyper® 21 Human DNA Analysis Kit (InnoGenomics) [35, 36]. A maximum of 500 pg or 16 µL DNA extract was used for amplification during a 31-cycle PCR according to manufacturer's instructions. PCR products were separated according to manufacturer's instructions using a 3130XL Genetic Analyser (Life Technologies) with POP-4 (Life Technologies) separation matrix and 3 kV, 15 s injection settings. Profile analysis was performed using Genemapper ID-X version 1.1.1 (Life Technologies) and a detection threshold of 50 rfu. The percentage of detected alleles was determined considering the 20 bi-allelic loci (Amelogenin excluded).

mtDNA SNaPshot

Mitochondrial DNA SNaPshot was performed on 14 of the DNA extracts for which less than eight alleles were detected using enhanced (29 + 5-cycles) NGM profiling. Reference samples (DNA extract of different tissue type for the same donor for which STR profiling was successfully applied) were included in mtDNA SNaPshot analyses.

Amplification of mtDNA fragments was performed as described in Ref. [37] using the mini-mtDNA method [41, 42]. This method aims to amplify ten overlapping mini-amplicons covering the entire control region in two multiplex PCR assays (set 1 and set 2). Due to limited amounts for the DNA extracts, only multiplex set 1 was used for which 5 µL extract was amplified in a reaction volume of 50 µL [37]. Minor adjustments involved the increase of the number of PCR cycles by 1 and adjustments in primer concentrations. mtDNA product formation was checked using QIAxcel capillary electrophoresis [37]. Prior to performing the single-base extension (SBE) PCR, the mtDNA PCR products were purified by using ExoSAP-IT® (Affymetrix) according to manufacturer's protocol.

The SNaPshot PCR was performed as described in Ref. [37] by using SBE-primers in combination with fluorescently labelled ddNTPs to extend the SBE-primers at the single nucleotide polymorphisms (SNP) position. These 18 SBE-primers are equally divided over two assays (SBE set 1 and 2) that match the two mini-mtDNA PCR sets. However, as the two mini-mtDNA PCR multiplexes have overlapping amplicons, quite some SBE-primers selected for set 2 also function with set 1 PCR products and *vice versa*. Thus, when both SBE-primer sets are (separately) applied to mini-mtDNA set 1 PCR products, 15 of the 18 SNPs can be analysed (nine set one SNPs: 16270, 16278, 16519, 195, 16362, 185, 16294, 182 and 16311 and six set 2 SNPs: 16223, 16129, 16126, 150, 146 and 152). Unincorporated ddNTPs were removed by using Shrimp Alkaline Phosphatase (2U, Affymetrix) according to manufacturer's protocol. Fragments were separated and detected as described in Ref. [37] on a 3130XL Genetic

Analyser (Life Technologies) using POP-4 (Life Technologies) separation matrix and 1.2 kV, 15 s injection settings. Profile analysis was performed using Genemapper® v4.0 (Life Technologies) with a detection threshold of 50 rfu and an allele balance cut-off value of 0.3. The percentage of detected SNPs was determined considering a maximum of 15 SNPs.

RNA profiling

Organ type inference was performed on all 81 RNA extracts using an updated version for an in-house developed multiplex (van den Berge and Sijen, manuscript in preparation), which allows for the inference of brain, lung, liver, skeletal muscle, heart, kidney and skin tissues. All RNA extracts, regardless of DNA quantification results [33], were subjected to ethanol precipitation prior to reverse transcription to maximize profiling results (and using the knowledge that the reverse transcription reaction can take up to 2 µg RNA, which will not be reached with the small tissues sections used for extraction). Ethanol precipitation, reverse transcription, PCR amplification and product detection were performed according to standardized protocols [20]. After reverse transcription, a serial input of 0.5 and 5 µL cDNA was used in the PCR to determine the input providing an informative RNA profile. Supplementary PCRs were performed to obtain three informative PCR replicates per sample. PCR products were purified [20] prior to detection using a 3130XL Genetic Analyser (Life Technologies). Amplification products were analysed using POP-4 (Life Technologies) separation matrix and 3 kV, 10 s injection settings. Profile analysis was performed using Genemapper ID-X version 1.1.1 (Life Technologies) with a detection threshold of 150 rfu.

RNA data interpretation was performed according to the "$x=n/2$" rule as described in [43]. 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 ($x=0$).

Results

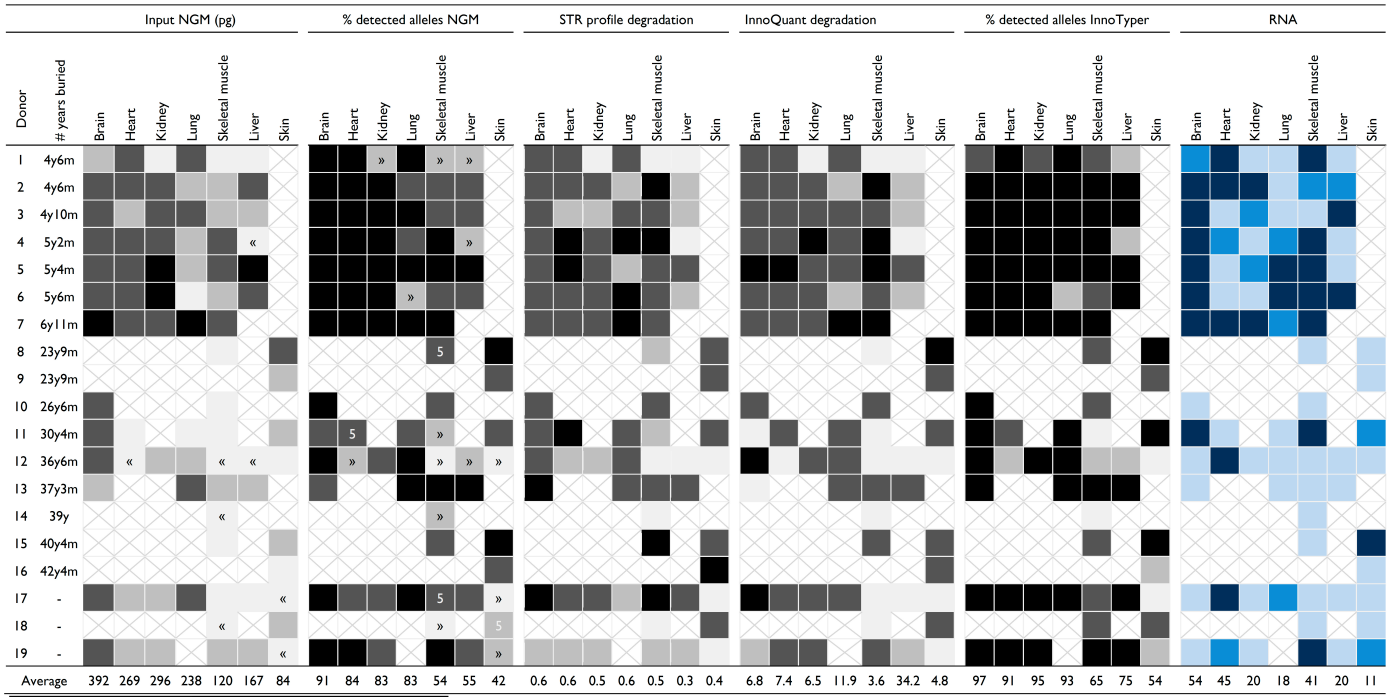
DNA quantification

For this study, a total of 81 decomposed organ tissues were analysed comprising 13 brain, 11 lung, 10 liver, 17 skeletal muscle, 11 heart, 10 kidney and 9 skin samples. Two highly sensitive DNA quantitation methods were used: firstly, the *Alu* assay described in Ref. [33] that provides accurate quantification results down to 0.5 pg/µL total human

DNA (Table 2). For eight samples, quantification results below this detection threshold were obtained. These samples represent three skeletal muscle, two liver, two skin and one heart sample (indicated "«" in Table 2). Although these samples most likely do not carry human DNA, they were processed like the samples that have detectable human DNA concentrations. As each excised sample was weighed prior to isolation, DNA yields could be normalized and presented in nanograms DNA per gram of isolated tissue for each sample (Table 2). The first six donors presented in Table 2 all provided each organ type except skin tissue and had been buried for approximately similar times (4 years and6months–5 years and6months). When comparing DNA yields per gram crude tissue for the different organ tissues within one donor, none of the organs stands out to produce the highest yield, as large variations in yields are observed for all tissue types. When regarding DNA yields for samples with longer burial times, less samples with high yields are seen although various samples have good yields; thus, indicating that DNA yields do not necessarily decrease with increased burial times (Table 2). The *Alu* assay also carries a male DNA indicator, which is less sensitive than the total indicator as it is based on much less repetitive units [33]. Nevertheless, male gender was correctly indicated for all samples with a concentration above the detection threshold (4 pg/µL).

Secondly, a novel quantitation system named the InnoQuant™ kit by InnoGenomics [34, 35] was used. The InnoQuant kit provides quantitative analysis and additionally determines a degradation index (DI) for quality and an internal positive control (IPC) for integrity, which is highly beneficial for forensic samples that are often degraded [34, 35]. The *Alu* assay and InnoQuant results follow the same trend (with InnoQuant 10 samples remain under the detection level including the eight samples for which the *Alu* assay did not return a quantification result), although differences are seen, which can be explained by the use of different repetitive elements in both systems. Most added value of the InnoQuant lies in providing a DI and assessing for PCR inhibitors. The DI is determined based on the ratio of quantitative values for the long and short amplicons and are presented in Table 2. For only a few samples, no degradation is detected (14 %). For the majority of samples, the DI indicates moderate (DI 2.5–20; 56 %) to severe (DI >20; 9 %) degradation, while for the remaining 22 %, no DI could be determined due to drop-out of the long fragment (which indicates very severe degradation). The skeletal muscle appears to be the organ type resulting in the most degraded DNA, as for 41 % of the skeletal muscle samples, no DI could be determined, followed by liver, skin, brain, kidney and heart tissues. Based on IPC results, two samples showed indication of PCR inhibitors (brain donor 11 and 13, data not shown). For one of these samples, inhibition was also observed using the *Alu* assay (undetermined Ct value).

Table 2 Overview of the DNA and RNA profiling results obtained after analysing 81 postmortem organ tissues collected after varying burial times.



Input NGM (pg)	≥500	499-250	249-50	<50
% detected alleles NGM	≥90	89-30	29-5	<5
STR profile degradation	>0.7	0.7-0.5	<0.5	n.d.
InnoQuant degradation	<2.5	2.5-20	>20	n.d.
% detected alleles InnoTyper	≥90	89-30	29-5	<5
RNA	Obs.	Sporadically Obs.		Not Obs.

« Samples with quantification results below the threshold of the *Alu* assay quantification system (0.5 pg/µL).

5 Samples subjected to five additional NGM PCR cycles.

» Samples subjected to five additional NGM PCR cycles and mtDNA SNaPshot.

n.d. Not determined ratio e.g. due to absence of longer fragments.

STR profiling

NGM profiling was performed for all 81 tissue samples. Not all DNA yields allowed for a full 500 pg DNA input, and Table 2 shows the range in pg DNA that was submitted to NGM profiling (either 500 pg DNA or 10 µL DNA extract). For 78 % of the samples, informative profiles with eight alleles or more were obtained after performing the standard 29- cycle PCR. PCR products of the 18 samples that resulted in profiles with seven or less detected alleles underwent an additional 5-cycle amplification (indicated "5" in Table 2). These include the eight samples with quantification results below the detection threshold of the *Alu* assay (section "DNA quantification"). As a result of the additional cycles, the percentage of samples with informative profiles (eight alleles or more) increased to 83 %. Besides the final percentage of detected alleles for each sample, Table 2 shows the average percentage of detected alleles for each organ type, from which we infer that the brain is the most promising organ tissue to be used for DNA profiling followed by heart, lung and kidney, liver, skeletal muscle and lastly skin tissue. STR profiles were additionally used to determine degradation rates by comparing the average peak height of short (75 to 225 bp) versus long (225 to 375 bp) alleles. For 16 % of the samples, this ratio was above 0.75, which indicates no or little degradation. The majority of samples showed moderate (ratio 0.1–0.75, 70 %) or severe ("n.d.", no long fragments detected, 14 %) degradation in the profiles. On average, brain, heart and lung samples show the least degraded profiles, followed by skeletal muscle and kidney, skin and lastly, liver samples.

InnoTyper profiling

For all 81 samples, additional autosomal profiling was applied through the InnoTyper™ 21 kit [35,36]. This kit targets genomic sites known to carry variation for the presence or absence of retrotransposable elements (RE) and amplifies either the insertion or the null (INNUL) allele. Primer design achieved small amplicons (60–124 bp) for both allelic states [35,36]. The InnoTyper kit is described to be highly sensitive, tolerant to degradation and inhibition and applicable to extremely degraded or low template samples. The kit accommodates 16 µL extract as input, due to which numerous samples received a higher DNA input than with NGM profiling that accommodates 10 µL input. The percentage of detected alleles was determined based on the 20 bi-allelic INNUL markers. An overview of the results can be found in Table 2. Full InnoTyper profiles were obtained for 65 % of the samples. The lowest input used to obtain a full profile was 47 pg. However, with this sample set that comprises many severely degraded samples, a certain DNA input will not guarantee a full profile, as for example, an allele detection percentage of 50 % was obtained for an input of 9 pg DNA and 53 % detected alleles for an input of 336 pg. Overall, higher average allele detection percentages were obtained compared to those obtained using NGM profiling (Table 2.

mtDNA SNaPshot

Furthermore, mtDNA analysis was performed. Due to the much higher copy number for mitochondrial DNA molecules than for genomic DNA (gDNA), mtDNA profiling is commonly applied to samples that lack or have severely degraded gDNA [44]. We used an mtDNA SNaPshot system described in Ref. [37] that uses SBE primers to target 18 SNPs with a high discriminatory power in a European Dutch population. The mtDNA SNaPshot was applied to 14 samples for which seven or less alleles were detected using enhanced NGM profiling (indicated "»" in Table 2). These samples are mainly skeletal muscle, liver and skin samples, including the eight samples with gDNA quantification results below the detection threshold of the quantification system (section "DNA quantification"). The percentage of detected SNPs was determined by comparing SNaPshot results of the 14 samples to corresponding reference samples. For all 14 samples, 100% of the expected SNPs could be typed (data not shown), and we infer that full mtDNA SNaPshot profiles would be achieved for all samples as these 14 samples represent the samples performing worst in autosomal profiling.

RNA profiling

mRNA profiling for the inference of seven organ tissue types was performed on all 81 RNA extracts, and the results are presented in Table 2. The brain appears the most successful organ to be sampled for mRNA tissue profiling, as samples were scored "observed" in 54 % of the brain tissue samples, followed by heart, skeletal muscle, kidney and liver, lung and lastly skin tissues. There is no clear relation between the burial time and the mRNA profiling results, as can be seen for example when regarding the heart tissue results. Overall, the skin appears to be the least successful tissue type to use for mRNA profiling, although only relatively old (23 years and up) samples were collected for this tissue type. Notwithstanding, a skin sample was scored observed which had been buried for over 40 years.

Discussion and concluding remarks

This study aims to search for trends in nucleic acid degradation of exhumed organ tissue samples with increased PMIs. For this purpose, DNA/RNA co-extractions were performed on 81 exhumed organ tissues with PMIs ranging from 4 to over 42 years. A summary of the outcome of various DNA and mRNA profiling techniques can be found in Table 3. Although kidney tissue presents the highest DNA yield on average, the organ does not always present the highest DNA yield when the various tissues for the same donor are regarded, which is consistent with results described in Ref. [5].

Table 3 DNA concentrations in ng per gram crude tissue.

Burial ground	Donor	# years buried	Brain	Heart	Kidney	Lung	Skeletal muscle	Liver	Skin
A	1	4y6m	3.74	267.51	0.36	6.24	0.69	0.11	
	2	4y6m	22.79	110.70	323.73	10.08	1.65	33.64	
	3	4y10m	139.83	7.47	24.71	145.98	0.82	2.06	
	4	5y2m	148.37	5.80	7.15	1.75	45.98	0.00	
	5	5y4m	22.99	36.40	1544.66	11.72	80.24	19.06	
	6	5y6m	71.95	134.39	746.36	0.29	6.11	9.84	
	7	6y11m	170.45	82.41	16.40	1354.27	15.11		
B	8	23y9m					0.21		32.15
	9	23y9m							4.83
	10	26y6m	6.80				0.97		
	11	30y4m	44.29	0.72		1.33	0.25		2.83
	12	36y6m	79.27	0.00	3.71	4.62	0.00	0.00	0.25
	13	37y3m	4.34			25.96	4.98	3.06	
	14	39y					0.00		
	15	40y4m					0.71		6.82
	16	42y4m							0.54
	17	-	345.71	8.61	5.48	177.09	0.68	0.89	0.00
	18	-					0.00		1.64
	19	-	66.57	7.17	5.98		12.16	6.50	0.00

Additionally, DNA yields did not always decrease with increased burial times, implying there is no straightforward relation between DNA concentration and the PMI. The degradation index provided by the InnoQuant kit and the peak heights at long and short amplicons in the STR profiles were used to determine the DNA degradation level in a sample.

As shown in Tables 2 and 4, varying degrees of degradation are observed for the different tissue types; such as a tendency for more degradation in liver and kidney and less degradation in brain tissue. This variation can be explained by various factors such as the high activity of hydrolysis enzymes in liver tissue [14], the high degree of autolysis in kidney tissue [15], a low tissue turnover rate and poor source of digestive enzymes in brain tissue plus a well-protected anatomical location of the brain in the body [9, 14, 45]. This also explains why certain tissue types with high DNA inputs show degradation, while other samples with lower inputs show less degradation (*e.g.* Table 2 input and profiling results for liver and kidney versus brain tissue in donor 5). The InnoQuant kit additionally provides an IPC value to indicate the presence of PCR inhibitors. Inhibition was observed only in a few brain samples, which may be caused by increased levels of proteins and inhibiting substances present in cerebrospinal fluid [46, 47]. The inhibiting effect could be overcome by reanalysing these samples with less input.

The INNUL profiling system appears of added value for analysing severely degraded samples, as the kit slightly outperforms the NGM STR profiling kit based on the percentage of detected alleles. This seems to be due to the small amplicon sizes and the larger DNA extract volume that can be added as PCR input. However, the InnoTyper kit performed less well than the mtDNA SNaPshot assay that resulted in full profiles for all samples for which gDNA profiling was unsuccessful. There is a much higher copy number for mtDNA than gDNA [44] although the mtDNA/gDNA ratio

Table 4 Ranking of the different tissue types for the different analysis methods. Tissues are ranked from 1 (best) to 7 (worst).

	DNA concentration / g tissue	InnoQuant degradation	% detected alleles NGM	STR profile degradation	% detected alleles InnoTyper	RNA
Brain	3	1-3	1	1-3	1	1
Heart	2	1-3	2	1-3	4	2
Kidney	1	1-3	3-4	4-5	2	4-5
Lung	4	4	3-4	1-3	3	6-7
Skeletal muscle	5	7	6	4-5	6	3
Liver	6	6	5	7	5	4-5
Skin	7	5	7	6	7	6-7

can vary between tissues and is described to be specifically high in tissues with higher ATP requirements (*e.g.* skeletal and heart muscle) [44, 48] and in liver tissue, where hepatocytes are described to have high number of mitochondrial genomes [11].

It is stated that RNA is less stable than DNA because of the hydroxyl group at the 2' position of the ribose sugar, which makes RNA more prone to hydrolysis than DNA [49, 50]. This hydroxyl group and the fact that G-U base pairing occurs within RNA molecules allow RNA to form secondary and tertiary structures [49–51], which may explain the remarkable stability in the analysed postmortem tissues. Stability can be transcript-dependent due to the structure at the 3' region [52, 53]. Furthermore, the presence of ribonucleases may vary in tissues, which affects overall RNA abundance. Correspondingly, the lowest RNA stability was found in lung and skin samples, which are known to be ribonuclease-rich organs [18]. Studies have previously described a correlation in transcript analysis in various tissue types and the PMI [18, 54] using (compared to this study) relatively short PMIs up to 11 h [18] and various factors are described to significantly affect mRNA expression levels, such as the agonal state [55], sex, or age at death [56], but these relations were not straightforwardly observed in this study.

The results of this study encourage the use of brain and heart tissue for postmortem DNA and RNA profiling. The association of DNA and/or RNA degradation and the postmortem interval is discouraged, as no trends were observed for the aspects regarded. Evidently, many factors both pre- and postmortem have a role in degradation and relations may >occur for factors not assessed in this study, such as the effect of taphonomic differences between the cemeteries and individuals, which may strongly influence determination of the PMI [16]. Furthermore, no relations were observed between DNA and RNA profiling success, which is in concordance with previous studies [57] and important knowledge for the interpretation of combined RNA and DNA profiling data. .

Acknowledgements

The authors are grateful to all the donors from whom tissues have been used

in this study. We thank Natalie Weiler (Netherlands Forensic Institute), Gina Pineda (InnoGenomics) and Sudhir Sinha (InnoGenomics) for technical assistance. Frank van de Goot and W.J. Mike Groen are thanked for sample collection and providing information on the exhumed bodies. Corina Benschop is thanked 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).

References

1. Goff ML (2009) Early post-mortem changes and stages of decomposition in exposed cadavers. Exp Appl Acarol 49(1–2):21–36
2. Hau TC, Hamzah NH, Lian HH, Hamzah SPAA (2014) Decomposition process and post mortem changes: review. Sains Malaysiana 43(12):1873–1882
3. Johnson LA, Ferris JA (2002) Analysis of postmortem DNA degradation by single-cell gel electrophoresis. Forensic Sci Int 126(1): 43–47
4. Calacal GC, Apaga DLT, Salvador JM, Jimenez JAD, Lagat LJ, Villacorta RPF, De Ungria MCA (2015) Comparing different post-mortem human samples as DNA sources for downstream genotyping and identification. Forensic Sci Int Genet 19:212–220
5. Courts C, Sauer E, Hofmann Y, Madea B, Schyma C (2015) Assessment of STR typing success rate in soft tissues from putrefied bodies based on a quantitative grading system for putrefaction. J Forensic Sci 60(4):1016–1021
6. Hansen J, Lesnikova I, Funder AMD, Banner J (2014) DNA and RNA analysis of blood and muscle from bodies with variable postmortem intervals. Forensic Sci Med Pathol 10(3):322–328
7. Hoff-Olsen P, Mevåg B, Staalstrøm E, Hovde B, Egeland T, Olaisen B (1999) Extraction of DNA from decomposed human tissue: an evaluation of five extraction methods for short tandem repeat typing. Forensic Sci Int 105(3):171–183
8. Ogata M, Mattern R, Schneider PM, Schacker U, Kaufmann T, Rittner C (1990) Quantitative and qualitative analysis of DNA extracted from postmortem muscle tissues. Zeitschrift für Rechtsmedizin 103(6):397–406
9. Itani M, Yamamoto Y, Doi Y, Miyaishi S (2011) Quantitative analysis of DNA degradation in the dead body. Acta Med Okayama 65(5):299–306
10. Pooniya S, Lalwani S, Raina A, Millo T, Dogra TD (2014) Quality and quantity of extracted deoxyribonucleic acid (DNA) from preserved soft tissues of putrefied unidentifiable human corpse. J Lab Phys 6(1):31
11. Schwark T, Heinrich A, von Wurmb-Schwark N (2011) Genetic identification of highly putrefied bodies using DNA from soft tissues. Int J Legal Med 125(6):891–894
12. Kaiser C, Bachmeier B, Conrad C, Nerlich A, Bratzke H, EisenmengerW, Peschel O (2008) Molecular study of time dependent changes in DNA stability in soil buried skeletal residues. Forensic Sci Int 177(1):32–36
13. Ebuehi OA, Amode M, Balogun A, Fowora A (2015) Postmortem time affects brain, liver, kidney and heart DNA in male rat. Am J Biochem 5(1):1–5
14. Ludes B, Pfitzinger H, Mangin P (1993) DNA fingerprinting from tissues after variable postmortem periods. J Forensic Sci 38(3):686–690

15. Bär W, Kratzer A, Mächler M, Schmid W (1988) Postmortem stability of DNA. Forensic Sci Int 39(1):59–70
16. de Leeuwe R, Groen WJM, A taphonomic study based on observations of 196 exhumations and 23 clandestine burials, Taphonomy of human remains: Forensic analysis of the dead and the depositional environment. E. Schotmans, N. Márquez-Grant, S. Forbes, Wiley-Blackwell, in press.
17. Koppelkamm A, Vennemann B, Lutz-Bonengel S, Fracasso T, Vennemann M (2011) RNA integrity in post-mortem samples: influencing parameters and implications on RT-qPCR assays. Int J Legal Med 125(4):573–580
18. Sampaio-Silva F, Magalhães T, Carvalho F, Dinis-Oliveira RJ, Silvestre R (2013) Profiling of RNA degradation for estimation of post morterm interval. PLoS One 8(2):e56507
19. Sijen T (2015) Molecular approaches for forensic cell type identification: on mRNA, miRNA, DNA methylation and microbial markers. Forensic Sci Int Genet 18:21–32
20. Lindenbergh A, de Pagter M, Ramdayal G, Visser M, Zubakov D, Kayser M, Sijen T (2012) A multiplex (m) RNA-profiling system for the forensic identification of body fluids and contact traces. Forensic Sci Int Genet 6(5):565–577
21. van den BergeMet al (2014) A collaborative European exercise on mRNA-based body fluid/skin typing and interpretation ofDNA and RNA results. Forensic Sci Int Genet 10:40–48
22. Haas C, Klesser B, Maake C, Bär W, Kratzer A (2009) mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. Forensic Sci Int Genet 3(2):80–88
23. Hanson EK, Ballantyne J (2013) Highly specific mRNA biomarkers for the identification of vaginal secretions in sexual assault investigations. Sci Justice 53(1):14–22
24. Juusola J, Ballantyne J (2005) Multiplex mRNA profiling for the identification of body fluids. Forensic Sci Int 152(1):1–12
25. Roeder AD, Haas C (2013) 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(4):707–721
26. Harteveld J, Lindenbergh A, Sijen T (2013) RNA cell typing and DNA profiling of mixed samples: can cell types and donors be associated? Sci Justice 53(3):261–269
27. Lindenbergh A, van den Berge M, Oostra RJ, Cleypool C, Bruggink A, Kloosterman A, Sijen T (2013) Development of a mRNA profiling multiplex for the inference of organ tissues. Int J Legal Med 127(5):891–900
28. Bauer M, Gramlich I, Polzin S, Patzelt D (2003) Quantification of mRNAdegradation as possible indicator of postmorteminterval—a pilot study. Legal Med 5(4):220–227
29. Heinrich M, Matt K, Lutz-Bonengel S, Schmidt U (2007) Successful RNA extraction from various human postmortem tissues. Int J Legal Med 121(2):136–142
30. Koppelkamm A, Vennemann B, Fracasso T, Lutz-Bonengel S, Schmidt U, Heinrich M (2010) Validation of adequate endogenous reference genes for the normalisation of qPCR gene expression data in human post mortem tissue. Int J Legal Med 124(5):371–380
31. Poór VS, Lukács D, Nagy T, Rácz E, Sipos K (2016) The rate of RNA degradation in human dental pulp reveals post-mortem interval. Int J Legal Med 130(3):615–619
32. Maeda H, Zhu BL, Ishikawa T, Michiue T (2010) Forensic molecular pathology of violent deaths. Forensic Sci Int 203(1):83–92
33. Nicklas JA, Buel E (2006) Simultaneous determination of total human and male DNA using a duplex real-time PCR assay. J Forensic Sci 51(5):1005–1015
34. Pineda GM, Montgomery AH, Thompson R, Indest B, Carroll M, Sinha SK (2014) Development and validation of innoQuant™, a sensitive human DNA quantitation and degradation assessment method for forensic samples using high copy number mobile elements Alu and SVA. Forensic Sci Int Genet 13:224–235

35. Sinha S, Murphy G, Brown H, Montgomery A, Carrol M, Tabak J (2015) Retrotransposable elements: novel and sensitive DNA markers and their application in human identity. Forensic Sci Int Genet Suppl Ser 5:e627–e629
36. LaRue BL, Sinha SK, Montgomery AH, Thompson R, Klaskala L, Ge J, King M, Turnbough M, Budowle B (2012) INNULs: a novel design amplification strategy for retrotransposable elements for studying population variation. Hum Hered 74(1):27–35
37. Weiler NEC, de Vries G, Sijen T (2016) Development of a control region-based mtDNA SNaPshot™ selection tool, integrated into a mini amplicon sequencing method. Sci Justice 56(2):96–103
38. Clark MA, Worrell MB, Pless JE (1997) Postmortem changes in soft tissues, Forensic Taphonomy; W.D. Haglund, M.H. Sorg, CRC Press 161
39. Ewing MM, Thompson JM, McLaren RS, Purpero VM, Thomas KJ, Dobrowski PA, DeGroot GA, Romsos EL, Storts DR (2016) Human DNA quantification and sample quality assessment: developmental validation of the PowerQuant® system. For Forensic Sci Int Genet 23:166–177
40. Westen AA, Grol LJ, Harteveld J, Matai AS, de Knijff P, Sijen T (2012) Assessment of the stochastic threshold, back-and forward stutter filters and low template techniques for NGM. Forensic Sci Int Genet 6(6):708–715
41. Eichmann C, Parson W (2008) 'Mitominis': multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples. Int J Legal Med 122(5):385–388
42. Berger C, Parson W (2009) Mini-midi-mito: adapting the amplification and sequencing strategy of mtDNA to the degradation state of crime scene samples. Forensic Sci Int Genet 3(3):149–153
43. Lindenbergh A, Maaskant P, Sijen T (2013) Implementation of RNA profiling in forensic casework. Forensic Sci Int Genet 7(1):159–166
44. Foran DR (2006) Relative degradation of nuclear and mitochondrial DNA: an experimental approach. J Forensic Sci 51(4):766–770
45. Lee J, Hever A, Willhite D, Zlotnik A, Hevezi P (2005) Effects of RNA degradation on gene expression analysis of human postmortem tissues. FASEB J 19(10):1356–1358
46. Mitchell PS, Espy MJ, Smith TF, Toal DR, Rys PN, Berbari EF, Osmon DR, Persing DH (1997) Laboratory diagnosis of central nervous system infections with herpes simplex virus by PCR performed with cerebrospinal fluid specimens. J Clin Microbiol 35(11):2873–2877
47. Ratnamohan VM, Cunningham AL, Rawlinson WD (1998) Removal of inhibitors of CSF-PCR to improve diagnosis of herpesviral encephalitis. J Virol Methods 72(1):59–65
48. D'Erchia AM, Atlante A, Gadaleta G, Pavesi G, Chiara M, De Virgilio C, Gissi C (2015) Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity. Mitochondrion 20:13–21
49. Fordyce SL, Kampmann ML, Van Doorn NL, Gilbert MTP (2013) Long-term RNA persistence in postmortem contexts. Investig Genet 4(1):1
50. Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362(6422):709–715
51. Varani G, McClain WH (2000) The G·U wobble base pair. EMBO Rep 1(1):18–23
52. Suay L, SalvadorML, Abesha E, KleinU(2005) Specific roles of 5' RNA secondary structures in stabilizing transcripts in chloroplasts. Nucleic Acids Res 33(15):4754–4761
53. Mitchell P, Tollervey D (2001) mRNA turnover. Curr Opin Cell Biol 13(3):320–325
54. Birdsill AC, Walker DG, Lue L, Sue LI, Beach TG (2011) Postmortem interval effect on RNA and gene expression in human brain tissue. Cell Tissue Bank 12(4):311–318
55. Cummings TJ, Strum JC, Yoon LW, Szymanski MH, Hulette CM (2001) Recovery and expression of

messenger RNAfrom postmortem human brain tissue. Mod Pathol 14(11):1157–1161
56. Preece P, Cairns NJ (2003) Quantifying mRNA in postmortem human brain: influence of gender, age at death, postmortem interval, brain pH, agonal state and inter-lobe mRNAvariance. Mol Brain Res 118(1):60–71
57. van den Berge M, Bhoelai B, Harteveld J, Matai A, Sijen T (2016) 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 Sci Int Genet 20:119–129

Chapter 9

A male and female RNA marker to infer sex in forensic analysis

Forensic Science International: Genetics 26 (2017) 70-76

Margreet van den Berge
Titia Sijen

Abstract

In forensics, DNA profiling is used for the identification of the donor of a trace, while messenger RNA (mRNA) profiling can be applied to identify the cellular origin such as body fluids or organ tissues. The presence of male cell material can be readily assessed by the incorporation of Y-chromosomal markers in quantitation or STR profiling systems. However, no forensic marker exists to positively identify female cell material; merely the presence of female DNA is deduced from the absence of a Y peak, or unbalanced X-Y signals at the Amelogenin locus or unbalanced response of the total and Y-specific quantifier. The presence of two X-chromosomes in female cells invokes dosage compensation, which is achieved through inactivation of one of the X-chromosomes in females. Since this process involves specific RNA molecules, identification of female cellular material may be possible through RNA profiling. Additionally, male material may be identified through RNAs expressed from the Y-chromosome. RNAs preferentially expressed in either sex were assessed for their potential to act as sex markers in forensic RNA assays. To confirm sex-specificity, body fluids and organ tissues of multiple donors of either sex were tested. Additionally, sensitivity of the markers and the suitability of positively identifying male-female mixtures were assessed and degraded samples were used to assess performance of the markers in forensic settings. The addition of sex-specific markers is of added informative value in any RNA profiling system and both markers were incorporated into existing RNA assays that either target body fluids or organs. These are the first forensic assays that enable positive identification of female cellular material.

Introduction

In forensics, DNA profiling is generally used for the identification of the donor of a trace, in which the sex (referring to the genetic origin of sex) of the donor is generally deduced from the response of the single copy gene Amelogenin (AMEL) gene that is located on the X and the Y sex chromosomes [1,2]. AMEL-based sex determination methods rely on the simultaneous amplification of differently sized X- and Y-chromosome targets and use the presence or absence of the Y-signals to determine sex [1–3]. The AMEL-Y can drop-out because of low-level DNA input or primer binding site mutations which can lead to misinterpretation of males as females [1–4]. The most recently developed STR (Short Tandem Repeats) kits circumvent this issue as they carry additional Y-chromosomal STRs markers [3,5–7]. There are, however, currently no forensic markers available to positively identify female cell material. Basically the absence or relatively low peak height of the AMEL-Y or Y-STR signals is used to determine female sex [1–3,5], which is more difficult to apply in male-female mixtures. On the DNA level the X-chromosome can therefore not be used to distinguish between female and male DNA. Female-specific markers may, however, be found on RNA or protein level as certain cellular processes are specific to females. Since forensic analyses may involve all kinds of body fluids or organ tissues, we aim for a process occurring in all diploid female cells such as the process leading to X-chromosome inactivation. In mammals, the inactivation of one X-chromosome provides functional X-chromosomal dosage equivalence between XY males and XX females. Identification of female cellular material may therefore be possible via targeting cellular products involved in X-chromosome inactivation. Since X-chromosome inactivation is an RNA-induced process (see the description of the process below), RNA markers seem more promising than protein-based markers.

Messenger RNA (mRNA) profiling has been increasingly applied in the last decade for the inference of body fluids [8–15] and organ tissues [16,17] residing in an evidentiary trace. RNA profiling is generally performed alongside DNA profiling via DNA/RNA co-extraction of a stain. mRNA profiling, however, currently lacks the possibility of providing information regarding the donor of a trace. The incorporation of sex-biased genes into presently employed mRNA profiling systems could provide additional information concerning the donors' sex. Female- and male-specific RNA markers have previously been described for use mainly in evolutionary biology, medical- and clinical genetics [18,19] such as to examine sex differences in cancer genetics [20,21], neuropsychiatric disorders [22] or human diseases [20,23]. In forensics, no applications have been described.

For this study, two markers with global sex-biased expression [24] were assessed, namely XIST and RPS4Y1. XIST (X Inactive Specific Transcript), the candidate female RNA marker, is a long non-coding RNA (lncRNA) involved in X-chromosome

inactivation (XCI) [25]. This process leads to the random silencing of one of the two X-chromosomes in females, resulting in an active and an inactive X (Xa and Xi, respectively), in order to provide X-chromosome dosage equivalence in males and females [26–28]. XCI is controlled by the X-chromosome inactivation centre (Xic) which is involved in counting the number of X-chromosomes to be inactivated, selecting and initiating the XCI and the in cis spread of the inactivation on the entire Xi [29,30]. The Xic contains four non-translated RNA genes, namely JPX, FTX, XIST and TSIX [27,28]. Both JPX and FTX escape XCI and show a female-specific role in the up-regulation of the expression of XIST [31,32]. Expression of these two genes, however, appears not fully female-specific [28,33]. XIST expression is initiated in early embryogenesis and is expressed exclusively from the Xic of the Xi. XIST RNA is not translated, although it is spliced, capped and polyadenylated. The largest human XIST transcript is 19 kb and remains in the nucleus where it "coats" the Xi to inactivate the chromosome. Males do not have dosage compensation, which makes XIST expression female-specific [34]. XIST transcription in male cells has only been described in the germ cells prior to meiosis [30,35]. The other marker candidate, TSIX, is transcribed antisense to XIST and is a negative regulator of XIST. TSIX is therefore expressed prior to inactivation and from both X-chromosomes; its expression is repressed once inactivation is established [29,30]. Consequently, XIST was the selected female marker candidate in this study.

The majority of male-specific genes are located on the Y-chromosome, such as DBY, SMCY, UTY, RPS4Y, and USP9Y [22,24,36]. Y-chromosome encoding ribosomal protein RPS4Y1 (Ribosomal protein S4, Y-linked 1) [36,37] was selected, as this gene is described to have the highest average expression difference between females and males [24] and is ubiquitously expressed in human tissues [36,37].

In this study we assessed the functionality of incorporating these sex-specific RNA markers in currently employed mRNA multiplexes [15,17], which can provide new possibilities in forensic analyses.

Materials and methods

Sample collection

Primer concentration optimization of the sex markers was performed using a set of control samples originating from the FirstChoice® Human Total RNA Survey Panel (Ambion® by Life Technologies, Carlsbad, USA). This panel contains total RNA pools of at least three donors for various human tissues and details on donor sex are provided by the manufacturer. The samples selected for use in this study contained material of solely male or solely female donors, *i.e.* colon, prostate, trachea, testes and spleen of

male donors and cervix, skeletal muscle, placenta and ovary of female donors. RNA pools come in a concentration of 1 µg/µL and they were processed so that after reverse transcription a cDNA input corresponding to 5 pg RNA was used in the PCR (which is also the amount of human RNA used to optimise the Organtyper [16]).

Single donor samples from various body fluids and organ tissues previously described in Refs. [15–17] were used to confirm sex-specificity of the candidate markers. These samples include 30 single donor body fluids from ten male and 20 female donors comprising four samples each for blood, saliva, vaginal mucosa, menstrual secretion, fertile semen, skin, nasal mucosa, plus two sterile semen samples obtained from vasectomised donors. The organ tissue sample set includes a total of 21 single donor autopsy organ tissues comprising brain, lung, liver, skeletal muscle, heart, kidney and skin tissues of one male and two female donors.

Mixed samples were analysed to confirm detection of both sex markers in case of male-female mixtures and to assess marker drop-out of the two sex-specific markers when the male-female ratios are unbalanced. This set includes seven blood:blood mixtures with varying inputs of male and female contribution, *i.e.* male-female ratios of 1:0, 0:1, 1:1, 1:½ , ½:1, 1:⅕ and ⅕ :1. The mixtures were prepared by combining cDNAs of single donor male and female blood extracts, the described ratios do therefore not relate to DNA ratios.

The applicability of the sex markers to (severely) degraded samples was assessed by analysing a set of ten exhumed organ samples. These samples have previously been used in a study which aimed to search for trends in nucleic acid degradation and the post mortem interval [17]. This exhumed organ sample set comprises ten organ tissue samples originating from four females and six males which had been exhumed after burial times ranging from four to over 40 years.

Lastly, the added value of the sex-specific RNA markers in forensic settings is presented through the analysis of a male nosebleed sample. This sample has been used in a previous study [15] and could then, based on mRNA profiling results, not be specified as such due to co-expression of various body fluid markers [15].

The autopsy and exhumed organ specimens have been used in previous studies and carry the necessary approvals [16,17]. The body fluid samples used for this study were collected with informed consent of the voluntary donors whose cell material was used.

Marker selection and primer design

Sex-specific markers were selected from RNA sequencing data generated by the Genotype-Tissue Expression (GTEx) project as described in Ref. [24]. Two markers, namely between the two sexes XIST (female) and RPS4Y1 (male), showing highest average expression differences in females versus males were selected from a list of

global sex-biased expressed genes [24].

XIST and RPS4Y1 primer sets were developed using Ensembl and NCBI primer blast [38,39], with at least one of the primers covering an exon–exon junction. Primers were designed in such a way that amplification products would fit in two existing mRNA profiling multiplexes, namely the Cell-typer [15] and Organtyper [16,17] multiplexes. Details on primer sequences of the two sex markers are shown in Table 1.

Preliminary tests were performed to assess marker perfor- mance, fragment sizing and optimise primer concentrations. Initially, both sex markers and housekeeping marker 18S-rRNA [16,17] were combined in a triplex and applied to samples containing material of solely male or female donors (FirstChoice® Survey Panel, see "Sample collection" section). Primer concentrations of the sex markers were optimized by assessing a range of concentrations (0.2, 0.4 and 0.8 µM) while 18S-rRNA remained in its standard concentration (0.015 µM [9,17]). The optimal primer concentration for both XIST and RPS4Y1 was set on 0.2 µM, after which both sex markers were combined into the Cell-typer [15] and the Organtyper [17] multiplex.

Table 1 Primer sequences for the two sex-specific markers.

Marker name	Tissue	[primer] µM	Forward primer (5'-3') / Reverse primer (5'-3')	Size (bp)	Dye
XIST	Female	0.2	ATTTTAACTGATCCCATTGAAGATACCACGC [a]	83	
			TCAGAATGTCCAAGAGGAGCCTAAGG		PET™ [b]
RPS4Y1	Male	0.2	TGGAAGAGGCAAAGTACAAGTTGTGC	63	NED™
			GGATTCCCTTCACTCCCACAGTAAT		

[a] Underlined nucleotides are 5' tails added to obtain the desired product size.
[b] Reverse primer labelled.

RNA profiling

DNA/RNA co-extraction, DNase treatment, DNA quantification, ethanol precipitation and reverse transcription were performed as described in Ref. [8]. 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 Ref. [9].

The reverse transcription protocol was adjusted when it became apparent that sex marker XIST was poorly amplified in the Cell-typer [15] and Organtyper [17] multiplexes. To this aid, an unlabelled version of the XIST reverse primer presented in Table 1, referred to as XIST RT-primer, was added during the first step of the reverse transcription protocol, in which Random Decamers (5 µM per reaction) are combined with RNA extracts and denatured. Random Decamers and XIST RT-primer concentrations were optimized to a final concentration of 4.7 µM and 0.625 µM per reaction, respectively (after considering a range of 2.5–4.8 µM Random Decamers and 5–0.3 µM XIST RT-primer). The input of RNA extract (maximum 10 µL) remained unadjusted. No negative effects were observed by this change in protocol, while the

average peak height of XIST signals was boosted an approximate fivefold (data not shown).

PCR amplification and product detection for all RNA analyses were performed according to standardized protocols [8]. PCR products were purified [8] prior to detection via a 3130XL Genetic Analyzer (Life Technologies). Purified amplification products were analysed using POP-4 separation matrix (Life Technologies) and 3 kV, 10 s injection settings. Profile analysis was performed using Genemapper ID-X version 1.1.1 (Life Technologies) with a detection threshold of 150 relative fluorescence units (rfu).

RNA data interpretation was performed according to the "$x=n/2$" guidelines and four PCR replicates per sample as described in [40]. 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 ($x = 0$).

DNA profiling

DNA profiles were generated using the AmpFℓSTR® NGM™ Amplification Kit (Life Technologies) using a maximum of 500 pg DNA input based on quantification as described in [8]. PCR products were separated according to standardized protocols [8] 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.

Results and discussion

Specificity

After selection of one candidate sex-specific RNA marker for each sex, primer sets were designed and tested on a set of control samples (sections "Sample collection", "Marker selection and primer design") with positive results. The assay conditions were optimized (sections "Marker selection and primer design", "RNA profiling") after which both markers were incorporated into the Cell-typer and Organtyper multiplexes (now referred to as Sex-Cell-typer and Sex-Organ- typer). Sex-specificity of the markers was assessed using single donor body fluids (n=30, four technical replicates each) and organ tissues (n=21, four technical replicates each). Single donor contribution was confirmed for all samples via NGM profiling, except the semen donations from vasectomised males, that lack spermatozoa and DNA (data not shown).

An overview of the specificity results using the single donor body fluids and organ tissues is shown in Table 2. Female sex was correctly scored "observed" in 97% of

the female samples (one sample scored "sporadically observed") and male sex was correctly scored "observed" in 94% of the male samples (one sample resulted in a "not observed" scoring). This includes the detection of the male sex marker in semen samples from vasectomised donors, for which no DNA-profile could be obtained using standard NGM profiling (data not shown). The detection of RPS4Y1 in azoospermic semen can be explained as RPS4Y1 is a ribosomal RNA marker, expression is therefore not restricted to the spermatozoa, and is thus also expressed in seminal fluid (alike the KLK3 and SEMG1 marker included in the Cell-Typer [15]). Furthermore, no incorrect "observed" scorings of female marker XIST in male samples or male marker RPS4Y1 in female samples occurred, indicating the markers are sex-specific with no false positive scoring.

Table 2 **XIST and RPS4Y1 RNA profiling results after analysing single donor body fluids (*n*=30) and organ tissues (*n*=21). Interpretation of the sex markers is performed using the "*x*=*n*/2" guidelines and four PCR replicates per sample. Shown are the number of female and male samples analysed for the specified body fluids and organ tissues. (Light) green marking indicate the correct "(sporadically) observed" scoring of markers, missed marker signals are indicated by a grey marking.**

Single donor	Female donor (*n*)	Male donor (*n*)	XIST correctly "observed"	XIST "sporadically observed"	RPS4Y1 incorrectly "observed"	RPS4Y1 correctly "observed"	RPS4Y1 "sporadically observed"	XIST incorrectly observed
Body fluids			Female samples			Male samples		
Blood	3	1	100%			100%		
Saliva	3	1	100%			0%		
Vaginal mucosa	4	-	100%			-		
Menstrual secretion	4	-	100%			-		
Semen*	-	6	-			100%		
Skin	3	1	100%			100%		
Nasal mucosa	3	1	100%			100%		
# samples	20	10						
Organ tissues			Female samples			Male samples		
Brain	2	1	100%			100%		
Lung	2	1	100%			100%		
Liver	2	1	100%			100%		
Skeletal muscle	2	1	100%			100%		
Heart	2	1	100%			100%		
Kidney	2	1	50%	50%		100%		
Skin	2	1	100%			100%		
# samples	14	7						

*Four fertile semen and two sterile semen samples.

Example overlay RNA profiling electropherograms after analysing single donor body fluids and organ tissues are presented in Figure 1. Shown is the positioning of the sex markers in the Sex-Cell- typer and Sex-Organtyper multiplexes and the correct detection of the sex markers in male and female samples. It has been described in medical research, such as cancer studies, that expression of XIST is correlated with the development of some cancers [20,21,35], leading to the detection of female marker XIST in male cancer cells. In forensic scenarios this may be relevant, for example when bullets are analysed that potentially perforated cancerous organ tissues.

To investigate this cross-reactivity a set of 28 biopsy samples, most likely holding cancer cells as they are taken from patients in hospitals, was analysed using the Sex-Organtyper multiplex. These biobank samples include 14 male and 14 female donors divided over four samples each for brain, lung, liver, skeletal muscle, heart, kidney and skin tissue. Male marker RPS4Y1 was detected, as expected, only in samples obtained

from male donors. Female marker XIST was detected in 50% of the male samples, which, based on four technical replicates per sample, resulted in an "observed" scoring for female cells in 43% of the analysed male donor samples and the "sporadically observed" scoring in one male sample (data not shown). This may be explained by the dysregulation of XIST in cancer cells, for which the expression of long, abundantly expressed non-coding transcripts is described to be altered [21]. Unfortunately, we cannot retrieve the detailed information on which sample was positively described to be cancerous.

Mixture analysis

Next, a set of blood:blood mixtures (n=7, four technical replicates each) with varying contributions of male and female material were analysed using the Sex-Cell-typer multiplex to confirm detection of both sex markers in case of male-female mixtures and to assess marker drop-out of the two sex-specific markers when the male-female ratios are unbalanced. Illustrative electropherograms are presented in Figure 2 and show the detection of male marker RPS4Y1 in each of the mixtures containing male RNA, and female marker XIST detection in samples containing female RNA. As can be seen from the varying signal heights of the male and female marker, sample input does not go hand in hand with peak height. This befits the use of 33 amplification cycles and has previously been described in Ref. [41].

Degraded samples

The performance of the sex markers on degraded samples was assessed on a set of exhumed organ tissues [17]. Using the Sex- Organtyper multiplex, sex was correctly scored "observed" for 80% of the samples; the remaining two samples resulted in a "sporadically observed" scoring. This includes the positive infer- ence of sex for a skin sample which had been buried up to 40 years, indicating that the sex markers can successfully be applied to degraded samples. Again, no false positive scorings occurred (Table 3).

Application in casework

The added value of the sex-specific RNA markers in forensic casework becomes clear when applied to a sample in which, without sex markers, the combination of detected mRNA markers could lead to incorrect assumptions. Recently, the detection of vaginal mucosa markers, normally linked to the presence of female cell material, has been described in expirated blood, nosebleed blood and nasal mucosa samples, regardless of the sex of the donor [15]. For example, when analysing a male donor

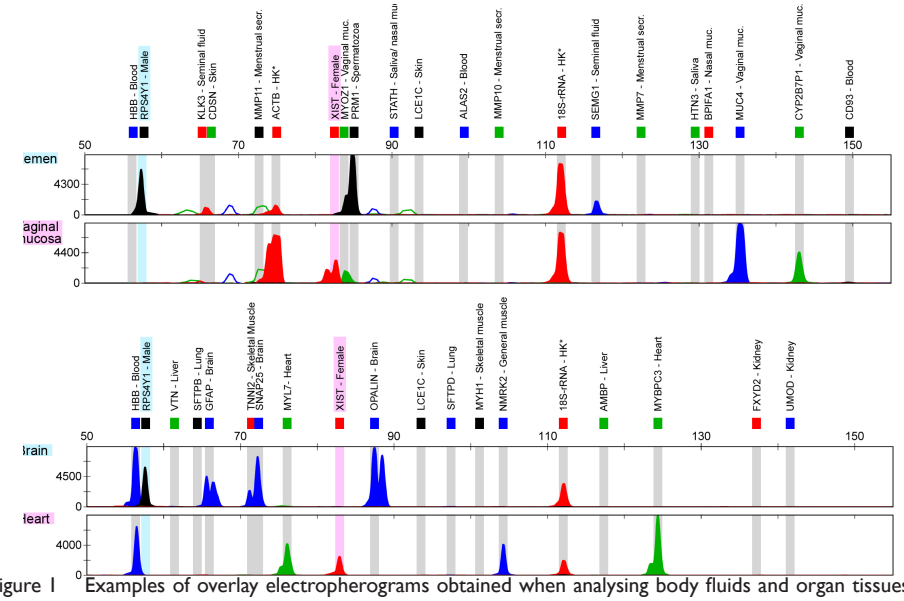


Figure 1 Examples of overlay electropherograms obtained when analysing body fluids and organ tissues from a single donor of either sex using the Cell-typer and Organtyper multiplexes supplemented with a male and a female sex marker. Highlighted in light-blue are the male sex marker RPS4Y1 and the male samples. The female marker and female samples are indicated by a pink highlight. *HK= housekeeping.

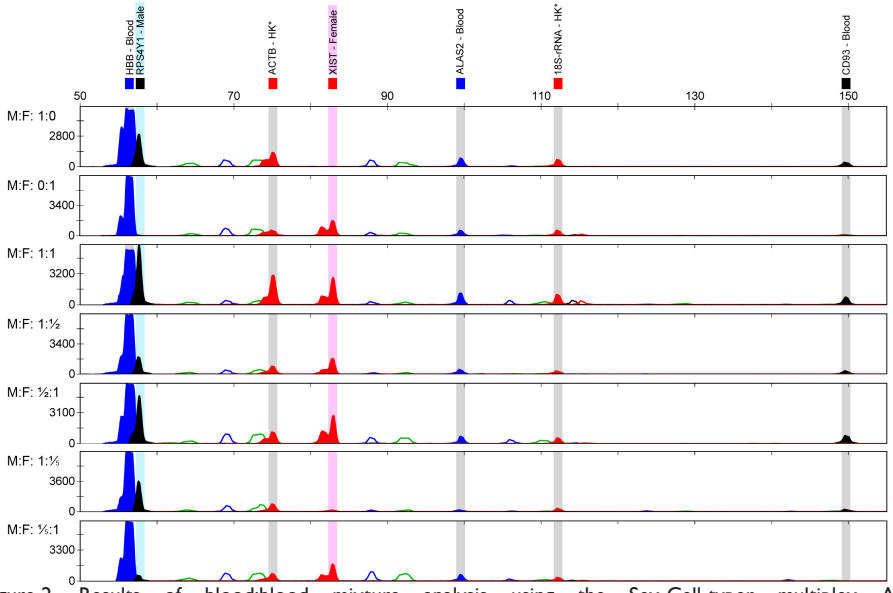


Figure 2 Results of blood:blood mixture analysis using the Sex-Cell-typer multiplex. A series of samples with varying ratios of male:female (M:F) contributions were analysed to assess the markers' suitability to detect both markers simultaneously. Highlighted in light-blue are the male sex marker RPS4Y1 and the male samples. The female marker XIST and female samples are indicated by a pink highlight. *HK= housekeeping. Only the bins for the responding RNA markers are shown.

Table 3 XIST and RPS4Y1 profiling results after analysing exhumed human organ tissues (*n*=10), which had been buried between 4 and up to 40 years. For three of the samples the exact burial times were unknown (# years buried indicated by "-"). Interpretation of the sex markers is performed using the "*x*=*n*/2" guidelines and four PCR replicates per sample. Green marking indicate the correct scoring of the target marker, light green markings indicate the "sporadic" detection of the target marker. No non-target markers were detected in non-target tissues (indicated in green).

Degraded samples	# years buried	Sex	Correctly "observed"	"Sporadically observed"	Incorrectly "observed"
Exhumed tissues			Target marker		Non-target
Kidney	4y6m	M	Y		N
Liver	4y10m	M	Y		N
Brain	5y2m	M	Y		N
Liver	5y6m	M		Y	N
Kidney	6y11m	M	Y		N
Brain	30y4m	F	Y		N
Skin	40y4m	F	Y		N
Heart	-	M	Y		N
Skin	-	F		Y	N
Skeletal muscle	-	F	Y		N

*Y= yes, N=no

nosebleed sample, blood, saliva, vaginal mucosa and menstrual secretion markers are detected, as shown in Figure 3A. A nasal mucosa marker has previously been developed and incorporated in an updated version of the Cell-typer multiplex [15] to aid in the distinction of a nosebleed samples from a mixture of blood, saliva and vaginal mucosa (Figure 3B). The addition of sex markers now confirms the male sex of the donor (Figure 3C). Thus, the addition of sex-specific RNA markers aids the inference that the sample involves a nosebleed sample from a male donor rather than a mixture of cells from a male and a female donor.

Concluding remarks

This study describes the incorporation of sex-specific RNA markers in currently employed forensic mRNA profiling multiplexes to simultaneously identify the sex of the donor and the cell type of the material in an evidentiary trace. The use of sex-specific markers has been described by various studies, mostly in medical context, but is new to forensics [18,22,23]. The Sex-Cell-typer and Sex-Organtyper multiplexes were successfully applied to single donor samples, mixtures and degraded samples. No false positive results were obtained, except for the XIST marker in biopsy samples that are likely tumorous. Thus, when cell material derives from an individual carrying cancerous cells in the targeted tissue, the data regarding the female marker should be interpreted carefully as long non-coding RNAs, such as XIST, are known to have dysregulated gene expression in tumour cells. We realise, however, that the health status of individuals involved in a crime will often be unknown. Additionally, certain syndromes such as Klinefelter (*e.g.* 47,XXY) and Triple X syndrome (47,XXX) will lead

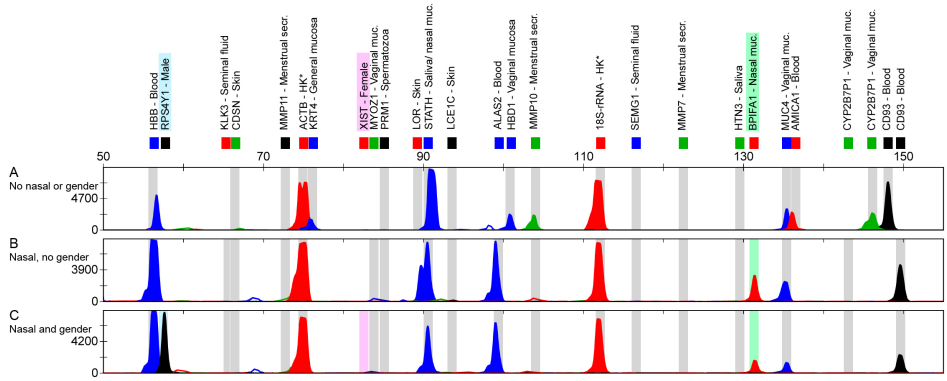


Figure 3 The added value of the Sex-Cell-typer multiplex shown by the analysis of a male nosebleed sample using three versions of the Cell-typer multiplex. A: Analysis using the Cell-typer multiplex as described in Ref. [9]. Results indicate a blood, saliva, vaginal mucosa, menstrual secretion mixture, which cannot be classified as a male nosebleed sample due to the absence of nasal mucosa and sex markers. B: The addition of a nasal mucosa marker (Ref. [15]) allows for the classification as a nosebleed sample. C: The addition of sex-specific RNA markers allows for inference of a nosebleed sample originating from a male donor. (Note that some markers have been replaced or spaced differently during the development of the multiplexes. Presence of bins under marker names indicate the presence of the respective marker per analysis.)

to the inactivation of multiple X-chromosomes and therefore increased XIST expression, while females with Tuner syndrome (45,X) will show no XIST expression as no X-chromosome inactivation occurs [42,43]. Syndromes due to skewed X-chromosome inactivation (non-random choice regarding which X is inactivated) are not expected to affect XIST marker detection [44]. The incorporation of sex-specific RNA markers provides the first overlapping information in DNA and RNA profiles.

Acknowledgements

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

References

1. A. Mannucci, K.M. Sullivan, P.L. Ivanov, P. Gill, Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene Amelogenin, Int J Legal Med, 106 (1994), pp. 190–193
2. N. von Wurmb-Schwark, H. Bosinski, S. Ritz-Timme, What do the X and Y chromosomes tell us about sex and gender in forensic case analysis?. Journal of forensic and legal medicine, 14(1), (2007) 27-30.
3. M.A. Jobling, I.C.C. Lo, D.J. Turner, G.R. Bowden, A.C. Lee, Y. Xue, T. Kraaijenbrink, J. Henke, G. Guanti, B. McKeown, R.A.H. van Oorschot, R.J. Mitchell, P. de Knijff, C. Tyler-Smith, E.J. Parkin, Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing Amelogenin Y. Human molecular genetics, 16(3), (2007) 307-316.
4. B. Shadrach, M. Commane, C. Hren, I. Warshawsky, A Rare Mutation in the Primer Binding Region of the Amelogenin Gene Can Interfere with Gender Identification, The Journal of molecular diagnostics, 6(4), 401-405.
5. J. Henke, L. Henke, P. Chatthopadhyay, M. Kayser, M. Dülmer, S. Cleef, et al. Application of Y-chromosomal STR haplotypes to forensic genetics, Croatian Med J, 43 (2001), pp. 292–297
6. D.Y. Wang, S. Gopinath, R.E. Lagacé, W. Norona, L.K. Hennessy, M.L. Short, J.J. Mulero, Developmental validation of the GlobalFiler® Express PCR Amplification Kit: a 6-dye multiplex assay for the direct amplification of reference samples. Forensic Science International: Genetics, 19, (2015) 148-155.
7. M.G. Ensenberger, K.A. Lenz, L.K. Matthies, G.M. Hadinoto, J.E. Schienman, A.J. Przech, M.W. Morganti, D.T. Renstrom, V.M. Baker, K.M. Gawrys, M. Hoogendoorn, C.R. Steffen, P.Martín, A. Alonso, H.R. Olson, C.J. Sprecher, D.R. Storts, Developmental validation of the PowerPlex® Fusion 6C System. Forensic Science International: Genetics, 21, (2016) 134-144.
8. 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.
9. M. van den Berge, A. Carracedo, I. Gomes, E.A.M. Graham, C. Haas, B. Hjort, P. Hoff-Olsen, O. Maronas, B. Mevag and N. Morling, 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.
10. 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.
11. 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.
12. J. Juusola, J. Ballantyne, Multiplex mRNA profiling for the identification of body fluids, Forensic Sci. Int. 152 (2005) 1-12.
13. 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.
14. T. Sijen, Molecular approaches for forensic cell type identification: On mRNA, miRNA, DNA methylation and microbial markers, Forensic Sci. Int.: Genet. (2014).
15. M. van den Berge, B. Bhoelai, J. Harteveld, A. Matai, T. Sijen, 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 Sci. Int.: Genet. 20 (2016) 119-129.
16. 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.
17. M. van den Berge, D. Wiskerke, R. Gerretsen, J. Tabak, T. Sijen, DNA and RNA profiling of excavated

human remains with varying postmortem intervals, In press
18. F. Staedtler, N. Hartmann, M. Letzkus, S. Bongiovanni, A. Scherer, P. Marc, K.J. Johnson, M.M. Schumacher, Robust and tissue-independent gender-specific transcript biomarkers. Biomarkers, 18 (2013) 436-445.
19. H. Ellegren, J. Parsch, The evolution of sex-biased genes and sex-biased gene expression. Nature Reviews Genetics, 8(9), (2007) 689-698.
20. X. Shi, M. Sun, H. Liu, Y. Yao, Y. Song, Long non-coding RNAs: a new frontier in the study of human diseases. Cancer letters, 339(2), (2013) 159-166.
21. D.S. Perez, T.R. Hoage, J.R. Pritchett, A.L. Ducharme-Smith, M.L. Halling, S.C. Ganapathiraju, P.S. Streng, D.I. Smith, Long, abundantly expressed non-coding transcripts are altered in cancer. Human molecular genetics, 17(5), (2008) 642-655.
22. M.P. Vawter, S. Evans, P. Choudary, H. Tomita, J. Meador-Woodruff, M. Molnar, J. Li, J.F. Lopez, R. Myers, D. Cox, S.J. Watson, H. Akil, E.G. Jones, W.E. Bunney, Gender-specific gene expression in post-mortem human brain: localization to sex chromosomes. Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology, 29(2), (2004) 373.
23. H. Si, R.S. Banga, P. Kapitsinou, M. Ramaiah, J. Lawrence, G. Kambhampati, A. Gruenwald, E. Bottinger, D. Glicklich, V. Tellis, S. Greenstein, D.B. Thomas, J. Pullman, M. Fazzari, K. Susztak, Human and murine kidneys show gender-and species-specific gene expression differences in response to injury. PLoS One, 4(3), (2009) e4802.
24. M. Melé, P.G. Ferreira, F. Reverter, D.S. DeLuca, J. Monlong, M. Sammeth, T.R. Young, J.M. Goldmann, D.D. Pervouchine, T.J. Sullivan, R. Johnson, A.V. Segrè, S. Djebali, A. Niarchou, The GTEx Consortium, F.A. Wright, R. Lappalainen, M. Calvo, G. Getz, E.T. Dermitzakis, K.G. Ardlie, R. Guigó, The human transcriptome across tissues and individuals. Science, 348(6235), (2015) 660-665.
25. K. Plath, S. Mlynarczyk-Evans, D.A. Nusinow, B. Panning, Xist RNA and the mechanism of X chromosome inactivation. Annual review of genetics, 36(1), (2002) 233-278.
26. K. Ng, D. Pullirsch, M. Leeb, A. Wutz, Xist and the order of silencing. Embo reports, 8(1), (2007) 34-39.
27. E. Maclary, M. Hinten, C. Harris, S. Kalantry, Long nonoding RNAs in the X-inactivation center. Chromosome research, 21(6-7), (2013) 601-614.
28. T. Nagano, P. Fraser, No-nonsense functions for long noncoding RNAs. Cell, 145(2), (2011) 178-181.
29. J. Lee, L.S. Davidow, D. Warshawsky, Tsix, a gene antisense to Xist at the X-inactivation centre. Nature genetics, 21(4), (1999) 400-404.
30. G.D. Penny, G.F. Kay, S.A. Sheardown, S. Rastan, N. Brockdorff, Requirement for Xist in X chromosome inactivation. Nature, 379(6561), (1996) 131-137.
31. J.G. van Bemmel, H. Mira-Bontenbal, J. Gribnau, Cis-and trans-regulation in X inactivation. Chromosoma, 125(1), (2016) 41-50.
32. C. Li, T. Hong, C.H. Webb, H. Karner, S. Sun, Q. Nie, A self-enhanced transport mechanism through long noncoding RNAs for X chromosome inactivation. Scientific Reports, 6 (2016)
33. J.B. Berletch, F. Yang, J. Xu, L. Carrel, C.M. Disteche, Genes that escape from X inactivation. Human genetics, 130(2), (2011) 237-245.
34. J.C. Chow, Z. Yen, S.M. Ziesche, C.J. Brown, Silencing of the mammalian X chromosome. Annu. Rev. Genomics Hum. Genet., 6, (2005) 69-92.
35. S.M. Weakley, H. Wang, Q. Yao, C. Chen, Expression and function of a large non-coding RNA gene XIST in human cancer. World journal of surgery, 35(8), (2011) 1751-1756.
36. H. Skaletsky, T. Kuroda-Kawaguchi, P.J. Minx, H.S. Cordum, L. Hillier, L.G. Brown, S. Repping, T. Pyntikova, J. Ali, Tamberlyn Bieri, Asif Chinwalla, Andrew Delehaunty, Kim Delehaunty, Hui Du, Ginger Fewell,

Lucinda Fulton, Robert Fulton, T. Graves, S. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfing, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S. Yang, R.H. Waterston, R.K. Wilson, S. Rozen, D.C. Page. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature, 423(6942), (2003) 825-837.

37. D. Planchard, Y. Loriot, A. Goubar, F. Commo, J.C. Soria, Differential expression of biomarkers in men and women. In Seminars in oncology, 36 (6), (2009) 553-565.
38. P. Flicek, et al., Ensembl 2014, Nucleic Acids Res. 42 (2014) D749–D755.
39. J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T.L. Madden, PrimerBLAST: a tool to design target-specific primers for polymerase chain reaction, BMC Bioinf. 13 (2012) 134.
40. A. Lindenbergh, P. Maaskant, T. Sijen, Implementation of RNA profiling in forensic casework, Forensic Sci. Int.: Genet. 7 (2013) 159–166.
41. Harteveld, J., Lindenbergh, A., Sijen, T. (2013). RNA cell typing and DNA profiling of mixed samples: can cell types and donors be associated?. Science & Justice, 53(3), 261-269.
42. B. Ji, K.K. Higa, J.R. Kelsoe, X. Zhou, Over-expression of XIST, the master gene for X chromosome inactivation, in females with major affective disorders. EBioMedicine, 2(8), (2015) 909-918.
43. C. Turner, N.R. Dennis, D.H. Skuse, P.A. Jacobs, Seven ring (X) chromosomes lacking the XIST locus, six with an unexpectedly mild phenotype. Human genetics, 106(1), (2000) 93-100.
44. R.M. Plenge, B.D. Hendrich, C. Schwartz, J.F. Arena, A. Naumova, C. Sapienza, R.M. Winter, H.F. Willard, A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. Nature Genetics, 17(3), (1997) 353-356.

Chapter 10

General discussion

Challenges and opportunities in forensic RNA cell type inference

Introduction to cell type inference

In forensic criminal investigations, DNA profiling is an important and frequently used tool as it has the ability to reveal the identity of the donor of a trace, which is important for both investigative and evaluative purposes. In the hierarchical evaluation structure that goes from sub-source, source, activity to offence level, donor inference reflects sub-source level reporting [1-3]. In current days it is increasingly questioned how evidentiary traces got deposited (rather than by who) resulting in activity level reporting by forensic scientists. One of the aspects that can provide activity level information are the cell type(s) contributing to an evidentiary trace, although this in itself is source level information (a trace represents a certain cell type from a certain person). Conventional methods for cell type inference include chemical, enzymatic and histological assays that tend to be of limited specificity or sensitivity, which can be complex when analysing forensic samples that often comprise little cell material [4-9]. Additionally, these methods are often presumptive in nature and only suitable for identification of one cell type at a time [4][10]. More recently, genetic- and protein-based cell type inferring methods have been proposed that include the use of micro RNA (miRNA), DNA methylation, microbial species, proteins and messenger RNA (mRNA). Our choice to exploit mRNA markers for forensic cell type inference is mainly derived from the limitations regarding specific markers for the other methods. A summary of the advantages and limitations of the methods other than mRNA profiling can be found in **Box 1**.

mRNA plays a key role in cell function as it operates as a messenger between a DNA gene and the final protein product The amount and composition of mRNA molecules in a cell reflects the function of that cell. The human body is composed of trillions of cells, encompassing at least 400 different cell types [11-12], such as adipocytes, erythrocytes and monocytes, that have different morphologies and functions. Even though the genome information (DNA) is the same for practically all cells (except haploid gametes and anucleated cells such as erythrocytes or thrombocytes), different cells express only subsets of the ~21,000 coding genes that constitute the full human genome [10][13]. A gene is a region of the DNA that encodes function. It is transcribed from DNA into RNA, which can be mRNA (or other types of RNAs that have a role in protein synthesis, post-transcriptional modification, DNA replication or gene regulation). In a process called translation, mRNAs serve as a chemical blueprint for the synthesis of proteins. Cell types are defined during early embryonic development by their specific pattern of regulated gene expression [10][14]. Expression of genes may be restricted to a specific cell type, as for example genes that code for proteins involved in oxygen transport need to be transcribed in red blood cells and not in spermatozoa, while other genes may be expressed ubiquitously. While the DNA in one's blood is unique for that individual, the mRNA expressed in blood cells is similar between all individuals.

Box 1 – Alternative genetic- and protein-based methods for cell type inference

Both messenger RNA (mRNA) and **micro RNA** (miRNA) contribute to cell type differentiation [10] [70]. miRNAs are small non-coding regulatory RNA molecules that function in gene silencing and gene expression regulation [71]. Silencing is induced by base pairing with (partially) complementary sequences within the 3'UTR (untranslated region) of the targeted mRNA [72]. With a predominant length of 22 nucleotides, miRNAs are much shorter that mRNAs. Since miRNAs are *in vivo* embedded in protein complexes and protected against RNases (RNA-degrading enzyme) they have a higher resistance to degradation compared to mRNA, which could make miRNA markers advantageous in forensic analysis.

However, miRNAs may lack specificity as only partial base binding of miRNA-mRNA is required, which allows individual micro RNAs to have up to hundreds of targets. Additionally, individual mRNAs can contain binding sites for multiple and different miRNAs [73].

Cell type information can also be obtained from DNA-based methods, although not through the actual genetic code but from changes occurring in the epigenome. The epigenome comprises the chemical changes to DNA and histones as a way to regulate gene expression [74]. One of these epigenetic modifications is called **DNA methylation**, which is the modification of DNA by the addition of a methyl group onto cytosine that can result in modification of gene function [75]. Various studies have revealed tissue-specific differentially methylated regions (tDMRs) that could provide an alternative to (m)RNA profiling [76-77]. DNA methylation-based assays employ either of two methods: the most applied method uses bisulphite conversion in which unmethylated cytosine nucleotides are converts into uracil, while the methylated cytosines remain unchanged. Data is analysed based on differential methylation levels, in which the ratio between methylated and unmethylated cytosines for each methylation site is determined (although SNaPshot CE-based visualisation results in "on-off" information rather than methylation rates, as the signal strength of different dyes differ [78]). A drawback of bisulphite conversion methods is that the conversion requires large quantities of DNA and goes hand-in-hand with DNA degradation [79]. The second method uses methylation-sensitive restriction enzymes (MSRE) [80-81] in a PCR targeting tDMRs. An advantage of this method is that it does not require bisulphite conversion, and therefore requires less DNA. This technique starts with digestion of genomic DNA using a methylation-sensitive restriction enzyme that digests the recognition site only if the cytosine is not methylated; when the cytosine is methylated the enzyme cannot cut. After digestion, either a PCR amplification is performed using primers that surround the digestion sites (only undigested *i.e.* methylated DNA fragments allow amplification) or a single base extension is performed (the extension primer will only be extended if the restriction site was not cut, *i.e.* the DNA was methylated). Suitable tDMRs with high methylation level differences have been discovered mainly for semen (most of which relate to hypermethylation). So far, DNA methylation is less successful for the inference of the remaining body fluids as markers for these body fluids show less distinct methylation levels. Additionally, methylation levels show inter-person variation and can be influenced by factors such as age, nutrition and environmental factors (*e.g.* stress, smoking, sun exposure) [82]. Furthermore, results can be affected by incomplete conversion or restriction, inhibition or degradation. One of the major advantages of DNA methylation markers compared to RNA markers is that it can be applied on DNA extracts. Therefore, it can readily be applied in virtually all forensic casework wherein DNA extracts are available, while RNA assays require additional procedures including the co-extraction of RNA. A recent study proposed the idea of implementing DNA methylation markers into currently existing STR kits [81]. This could be of aid for example for the semen fraction after differential lyses in which the presence of semen can be confirmed by detection of semen-specific methylation markers. Moreover, multiple independent studies have identified useful methylation markers for the purpose of age determination with improved predicting accuracy and practical applicability compared to alternative age predicting methods (such as traditional morphological-, chemical-, or genetic-based methods like mitochondrial DNA deletions and telomere shortening).

Another alternative marker for cell type inference lies in **microbial species**. Microorganisms can be found virtually everywhere on earth, including the trillions of microbial cells that can be found in the human body. They function both inside and outside the human body to secure human health and support food digestion. Different locations house distinct microbial communities and even though the makeup of these communities may vary depending on factors such as age or health issues, these communities show sufficient similarities between individuals to differentiate between certain body fluids or tissues [83-84]. For forensic cell type inference purposes especially the vaginal micro flora has received a lot of attention, mainly because of the difficulty in finding true vaginal mucosa specific (m)RNA markers [84-85]. The use

of bacterial markers has been proposed and especially the *Lactobacillus* species were found suitable for the identification of samples containing vaginal mucosa. There are however some drawbacks in the use of microbial markers, such as the variations that may be observed between persons, the fact that micro flora may change with disease (*e.g.* vaginosis), microbes may not be human-specific and the microbiome may transfer upon contact [86-87]. Additionally, multiple studies describe the detection of various microbial markers in samples that do not contain vaginal mucosa, such as other areas of the skin (including samples devoid of female DNA) [17][85][88]. Use of these markers for forensic body fluid identification settings may therefore not be desirable.

Rather than using mRNA markers that target uniquely expressed mRNAs, **proteomics** is based on the detection of specific protein biomarkers, such as Semenogelin for semen. Proteomes can be mapped using two-dimensional high-performance liquid chromatography (HPLC) [89-91]. After identifying potentially useful biomarkers, mass-spectrometry-based assays can be used to detect even low-abundance proteins in high resolution against a background of non-target molecules. This way, the proteomes of various body fluids such as menstrual blood, blood, semen, and saliva have readily been investigated and distinguishable proteomic profiles have been identified. The major challenge in proteomic analysis is, however, that it requires a separate extraction of proteins that has so far not been incorporable in DNA and/or DNA/RNA co-extraction protocols. Furthermore, the abundancy of certain proteins, such as keratins in skin, can impede the successful application of proteomic markers in forensic science.

RNA profiling at the NFI

The forensic RNA-based cell type inferring assays at the NFI target genes that are predominantly expressed in forensically relevant body fluids (blood, saliva, vaginal mucosa, menstrual secretion, semen and nasal mucosa are targeted by the Cell-typer multiplex) and organ tissues (brain, lung, liver, skeletal muscle, heart, kidney and skin are targeted by the Organtyper multiplex). These assays allow for forensic assessments in for instance sexual assault cases where the presence of vaginal mucosa cells is questioned, or can be of aid in for example the reconstruction of a violent crime.

At the NFI, mRNA profiling for body fluid identification was first applied in casework in 2010. Three years later, mRNA-based organ typing was introduced in casework. In total, mRNA profiling has been considered in over 200 cases. Upon request of RNA profiling, the co-extracted DNA is first subjected to STR profiling to assess suitability and usefulness for RNA profiling. Depending on the case this may for example involve finding female DNA in a sampling from male skin areas or clothing or *vice versa*. In approximately 60% of the cases considered for RNA profiling, DNA profiling results led to the subsequent request of subjecting a sample to RNA profiling. About half of the DNA results of the samples that are sent through for Cell-typer requests are single-donor profiles, with the other half being mixed profiles. For the Organtyper requests, the majority of cases sent trough for RNA profiling are single-donor based on DNA profiling results (90%). For the Cell-typer multiplex the majority of samples selected for RNA profiling relate to sexual assault cases and include the analysis of samples obtained from *e.g.* the penis, underpants or fingers. For the Organtyper multiplex, mainly clothing or trauma-causing objects such as bullets or knives are examined. The Cell-typer multiplex is requested in approximately 70% of the RNA cases, compared to the Organtyper multiplex in approximately 30% of the cases. In two of the 200 cases, both the Cell-typer and the Organtyper multiplexes were requested for the same RNA extract. RNA profiling itself is generally performed unbiased – with no case context – after which results are handed to the reporting officers.

Over the years, the multiplexes have almost continuously been under optimisation to adapt to minor or major challenges that were encountered in casework or research [9][15-20]. For example, nasal mucosa was generally not considered a forensically relevant body fluid, and markers for the cell

type were therefore initially not included in the assay. The cross-reactivity of various Cell-typer markers in nasal mucosa and the observation that nasal samplings contain high amounts of human cell material, however, has led to the addition of a nasal mucosa marker in the Cell-typer multiplex [17]. This allows for alternative scenarios such as secondary transfer of nasal mucosa, blood from nose bleedings or expirated blood to be assessed. Overall, out of the 19 markers included in the Cell-typer multiplex that was first used in casework in 2010, only 11 remain in the currently used multiplex (although for all 11 markers the primers were redesigned to improve performance). Six markers were removed because they had limited informative value (*e.g.* general mucosa markers that are co-expressed in various cell types and skin markers that were found to provide limited added value in casework even when observed), two markers were replaced (a vaginal and a blood marker), and six markers were added. These new markers included three novel cell types: a female and male gender marker, and the previously-mentioned nasal mucosa marker that aids in distinguishing nasal mucosa from samples containing saliva and/or vaginal mucosa. The RNA-specific gender markers allow, for the first time, to establish a link between RNA and DNA profiles [19]. Also the Organtyper multiplex has had its improvements as it now targets a broader range of cell types that reside in the central nervous system (CNS), *i.e.* neuronal cells, astrocytes and oligodendrocytes and it carries the gender markers.

Challenges in forensic RNA profiling

There are various differences between DNA and RNA that challenge the workflow and data interpretation for RNA profiling. These differences can raise concerns or beliefs regarding the use of RNA in forensic settings that will be addressed below.

"RNA is less stable than DNA and therefore not suitable for use in forensic casework"

It is often disputed that RNA is less stable than DNA due to the hydroxyl group at the 2' position of the ribose sugar, which makes RNA more prone to hydrolysis than DNA. Some argue that, due to this lower stability, RNA is not suited for forensic casework. Yes, the hydroxyl group at the 2' position of the ribose sugar makes RNA more prone to hydrolysis than DNA [21-22]. This hydroxyl group, however, and the fact that G-U base pairing occurs within RNA molecules, allows RNA to form secondary and tertiary structures which probably adds to a remarkable RNA stability when fluids or tissues are in dried state [21-23]. *In vivo*, stability is known to be transcript-dependent based on the structure at the 3' region [24-25]. Also, RNA stability is provided from forming ribonucleoprotein complexes [26]. This stability has also been shown in Chapter 8 of this study, where RNA profiling was successfully applied on excavated samples that had been buried for over 40 years [18].

To enable the analysis of degraded RNA, small amplicon sizes are used in our multiplexes. To prevent preferential amplification for a cell type, the targeted tissues have amplicons in the same size range. The main risk of low stability is that no RNA remains available, thus that no results are obtained and false negative interpretations may occur. Keeping in mind that absence of detection does not mean that a body fluid is absent(note that the stability of RNA may vary for body fluids, possibly related to the microbes present in a fluid [10][27]), RNA profiling is well suitable for use even in the challenging samples that can be encountered in forensic settings.

Other studies describe how the over-time stability changes of RNA can actually be used for other forensic purposes. For example, the degradation pattern of mRNA could be used to determine the time since deposition of a sample, or signals for circadian rhythm expressed RNA molecules could be used to determine at what time during the 24 hour day/night cycle a sample was deposited at a crime scene. More detailed information regarding alternative uses of mRNA is presented in **Box 2**.

Box 2 - Alternative applications using mRNA

There are various studies that use mRNA-based techniques for other purposes than cell type inference. Many of these studies involve the ex-vivo degradation of RNA. As the ex-vivo stability of RNA changes over time, the degradation pattern of mRNA can potentially be used to establish a system to determine the time a sample, such as a blood stain, was placed. Regarding **time of placement**, there are two distinct research focuses that are currently being explored. The first aims to determine the time since deposition (TSD), giving an indication of the time in hours/days/weeks since a trace was deposited. The ability to determine the TSD can provide information by establishing an approximation of the time a criminal offense took place, which could also be used to select/ensure crime-related stains or to assess a suspect's alibi [92-93]. More recently, a novel trace timing aspect was examined aiming to determine **at what time during the 24 hour day/night cycle** a sample was deposited at a crime scene [94-96]. This study builds on the intrinsic rhythms of core clock-controlled genes involved in the circadian rhythm, such as the hormones melatonin and cortisol. This method can be promising, however, more clock-controlled genes with different peak times in the 24-hour day/night cycle need to be identified to increase accuracy and reliability of this method. Furthermore, the robustness of the method in varying conditions (*e.g.* humidity, temperature, different substrates) and the effect of factors such as medication need to be assessed.

The determination of a **postmortem interval** (PMI) is an important aspect in forensic death investigations. Traditionally, the PMI is estimated based on various physical and biochemical changes occurring shortly after death, such as rigor, algor and livor mortis [97-98]. These factors can, however, only be used for estimating relatively short PMIs with a wide window of estimation [98-99]. More recently, studies have described the use of nucleic acid degradation patterns to estimate the PMI. The rationale behind this is that when postmortem intervals increase such as with longer burial times, human remains suffer increasingly from the taphonomic effects of decomposition processes such as autolysis and putrefaction. Varying results are described in different studies [100-104] using either of two methods. An end-point PCR method can be used in which tissue specific mRNA markers (and STR markers) are used to examine for trends in degradation. Alternatively, a quantitative PCR approach is used in which RNA markers are selected based on their abundancy in various tissues that is either stable (as a reference) or becomes altered when the PMI elongates (as a target biomarker). The overall trend of these studies it that even though promising results can be obtained for mock-cases, PMI estimation may be much more complicated in real cases as the effects of many pre- and postmortem factors with a role in degradation are unknown. All of these studies do however show that RNA, which is often assumed to be much less stable than DNA, can be remarkably stable. While the previously described studies aim to investigate the postmortem decrease of mRNA expression levels, a more recent study examined the transcripts generated in a window up to 96 hours after death with varying peak abundances for different genes and the striking expression of genes that were not expressed before death. One of the explanations of the increase of these transcripts is that when the nucleosome structure unwinds, remnant transcription machinery can transcribe genes that were previously silenced by their chromatin structure. These so-called "zombie genes" appear to "wake up" after death and monitoring of their expression could potentially be of aid in determining the time of death more accurately [105-106].

Alternatively, mRNA analysis could be of aid in determining the **cause of death** by examining alterations in mRNA expression patterns as a result of the traumatic effects surrounding death. For example, the amount of EPO (Erythropoietin) synthesized by the kidneys depends on the amount of oxygen present in blood. In case of for example drowning, strangulation or hanging, the expression level of EPO decreases. Similarly, various biomarkers are known to respond differently depending on the type of trauma that occurs. Although this field of study is relatively new in forensics, it may in the future be possible to distinguish traumas leading to death, such as hypoxia, injury (*e.g.* a different respond to brain, lung, or

cardiac damage), intoxication and drug adverse effects, or fire fatality based on changes in biomarker expression [107-108].

Wound age estimation is another issue in forensic science, that is routinely determined in forensic autopsy based macroscopic (*e.g.* skin colour changes) and immunohistochemical examinations [109-111]. Determination of the early stage (days) of wound age is especially important in case of suspected non-accidental injuries. Information regarding the age of the wound (dermal or intestinal) can give insight in the timing of the injury, the order of infliction, the survival time after injury and the relation between the time a trauma was inflicted and the wounds detected on a body. From a genomic point of view, the process of healing involves many biological changes that can be divided in three phases, namely inflammation, proliferation and maturation [110][112-113]. The timely changes in gene expression that occur during these three phases are caused by various biological mediators such as cytokines, growth factors and other bioactive molecules [110][114]. The key to mRNA-based wound age estimation is to identify a combination of genes that is indicative for specific time points or phases in the wound healing process. For this, mRNA expression of genes is monitored during the wound healing process to assess their suitability to be used to estimate wound age. Studies describe how especially mRNAs encoding cytokines and growth factors are useful markers for the determination of wound age [110][114]. Other studies have been performed aiming to determine the **age of bruises**, which is especially important in child abuse cases [115]. Alike for traditional wound age determination, morphological characteristics such as the transition of colour are important in determining the age of a bruise. There are however some factors that affect that affect the duration and appearance of a bruise, such as the location, size and skin colour [116]. Genes that play a role in bruise healing can aid in the determination of the age of a bruise. Various markers have been proposed that have the potential to be used to more accurately estimate the age of bruises and/or wounds [109-123]. However, aspects such as the effect of treatment like antiseptics, varying gene expression and the fact that most of the current bruise/wound age estimation methods are developed targeting animal models and are thus not fully representative for humans, difficult use of these assays in forensic casework [109][117][119][120].

"Lack of an RNA quantification system, determining input is not objective"

With DNA profiling, a quantitative real-time PCR approach targeting a human-specific sequence (nowadays a multicopy or even repetitive element to increase sensitivity [28]) is used. A similar method for RNA is complicated for two main reasons: 1) qPCR requires RNA to be copied into cDNA for which the input in the reverse transcription (RT) reaction needs to be inferred first (and one then assumes that this RT reaction is equally efficient for each sample) and 2) even when a ubiquitously expressed target sequence is used, this RNA is subject to expression differences for cell types, donors and in response to external factors [29-31]. At the NFI, we therefore utilize a serial input PCR approach which is a pragmatic solution to overcome these quantification difficulties [9]. For the RT reaction, we use a fixed amount of RNA following the rationale that the RT reaction is not readily overloaded with forensic samples (according to the manufacturer, up to 5 micrograms of total RNA can be reverse transcribed). Then, we use three cDNA inputs (*e.g.* 0.2, 1 and 4 µL) in the PCR and based on results of this serial input, an informative PCR input is selected which is subsequently used to generate replicate PCRs (generally four) [32]. Due to for example stochastic amplification effects (when template is limited), background expression (mRNA expression is enriched, not restricted in cell types) or spurious transcription (transcription is a relatively sloppy process with many cellular processes in place to destabilise aberrant transcripts) these PCR replicates may not be identical [10][32]. To overcome these issues, an approach alike the low-template DNA profile consensus approach is utilized [33], meaning the use of multiple PCR replicates per sample and scoring all markers for a cell type in all replicates jointly according to the "$x=n/2$" interpretation guideline. This method compares

the number of observed (*x*) to the number of theoretically possible peaks (*n*) in all replicates [32].

"Outcomes depend on interpretation approach"

The "*x*=*n*/2" guideline anticipates for the detection of background signals as cell types are only scored "observed" when ≥50% of the possible signals are detected [32]. In this approach, each marker has the same "weight" in data interpretation. This is theoretically not optimal as marker expression is known to vary between donors depending on various biological, physiological or environmental factors [32][34]. The exact composition of an evidentiary sample may vary as body fluids and organs are composed of various cell types. Thus, a specimen may contain fewer or more cells of the cell type in which a target marker is expressed. For example, at the NFI, brain is examined by a neuronal, astrocyte and oligodendrocyte mRNA marker. The use of multiple genes simultaneously in combination with the ≥50% value averages out these variations. This guideline has been shown to prevent false positive scorings [16]. This may however come at the cost of not inferring all cell types that are present, which is important to keep in mind during case interpretation (not seen ≠ absent). Results for cell types that are scored "observed" are reported as "fits with the presence of … cells". When no signals are seen the cell type is scored "not observed" which is reported as "no indication for the presence of … cells". When signals are seen for <50% of the possible signals, the scoring is "sporadically observed" which is reported as "no statement can be made regarding the presence of … cells" [32]. Thus, the ≥50% rule represents a fall of the cliff approach in which one observation more or less can make a categorical difference between "fits with the presence of" and "no statement can be made". Although challenging to achieve due to the many variables, reporting of RNA data could significantly benefit from probabilistic reporting, *i.e.* to assign weight of evidence to the results. An initial study shows both the outline and the challenges that come with this approach [35-36].

"Housekeeping markers are not used for normalization"

Clearly, the cell type specific markers are the most informative and important markers in RNA multiplexes, but there is debate regarding the importance of appropriate control/housekeeping markers. Housekeeping markers may be used to mark that RNA was present in the sample, but this requires that a housekeeping marker is always detected when RNA is present. This is not always the case with forensic samples (the expression of the housekeeping marker may vary or the housekeeping marker may not be detected due to a lower primer concentration that is selected to have a balanced end-point RT-PCR profile). Also, housekeeping markers can be used for normalisation when a quantitative RT-PCR approach is followed [29][36], but for the studies described in this thesis we apply end-point RT-PCR. The main purpose of the housekeeping markers in our analyses is to indicate absence of substances inhibiting cDNA synthesis and/or cDNA amplification for a sample [15][20]. This is especially informative when none of the other markers gives a signal. Thus, we use a housekeeping marker to confirm successful PCR amplification. Housekeeping markers in end-point PCRs should therefore be expressed ubiquitously, so that there is no issue on dependence for a cell type, which is the case in the assays described in this thesis. Various other controls are included throughout the RNA workflow, such as blank controls to confirm that no contamination has occurred, positive controls to confirm that all the used chemicals function, minus-RT controls (RT control lacking the reverse transcriptase enzyme) to assess for residuals of genomic DNA present in RNA extracts after DNase treatment.

"DNA and RNA results cannot be linked"

The use of multiple genes for different cell types in a multiplex allows for the analysis of cell type

mixtures. While DNA peaks are generally in line with the DNA contribution when contributions are not low template, RNA peaks may be unbalanced for a single contribution due to varying levels of expression per marker that can vary per individual depending on various biological, physiological or environmental factors [32][34]. Moreover, one contributor can donate more than one cell type and a single cell type can be given by multiple contributors. Up to now, profiling results are reported as DNA and RNA results, as currently DNA and RNA profiling outcomes cannot readily be linked. Only with certain body fluids, such as semen and vaginal mucosa, association of cell type and donor is possible based on gender (when the sample is inferred to comprise one female and one male donor) [19]. An alternative way of associating DNA and RNA based on peak heights has been assessed [37], but this was found to be impossible as the major in the DNA profile was found not to necessarily correspond to the highest RNA signals. The underlying reason may be that DNA and RNA profiling can have different sensitivities for cell types and donors.

Recently, the sequence of the RNA molecule has been proposed to provide information for association of DNA and RNA results [38]. Single nucleotide polymorphisms (SNPs) in transcribed regions are carried over to the RNA molecules. Clearly, within the open reading frame (ORF) there is selection against frameshifts, early stop codons and amino acid changes with deleterious effects on protein function and as a result SNPs may be less frequent in ORFs than in untranslated regions [39] (UTRs, note that 5'UTR changes may influence expression level and 3'UTR changes may affect stability [40-41]). Transcribed regions of marker genes can be examined for SNP locations in various databases such as the Ensembl database [42]. "Reference" SNP profiles can be obtained using DNA extracts; RNA extracts will confirm the cell type specific expression. To effectively associate donors (DNA) and cell types (RNA), multiple SNPs are required. This can be a collection of SNPs with varying minor allele frequencies (MAFs), in which the SNPs with a higher MAF are more common and thus more likely to be observed, while SNPs with a lower MAF are less common and thus more discriminative. Initially, RNA SNP profiles could be used to determine whether a suspect can have contributed to a trace. In the future, it may be possible to assign a weight of evidence to the RNA SNP profile, which enables moving towards distinguishing between individuals. Previous studies have reported that SNP mixtures can be statistically evaluated in a similar way as STR mixtures [43-44]. In contrary to STRs, SNPs do not stutter, which simplifies statistical analysis. For performing such statistical analyses, frequency data of the individual SNPs are required.

"RNA profiling is time consuming and labour intensive with value in limited numbers of cases"

Yes, RNA profiling described in this thesis represents a labour intensive approach and should be applied in cases only where the labour is likely to merit the information for the case. Notwithstanding, there may be strategies to use a simplified form of RNA typing in many more forensic cases when the method is adapted to befit pre-screening of crime stains (as also outlined in the ParaDNA Body Fluid ID Test [45-46]). Then, the method could replace and extend conventional presumptive and/or confirmatory tests (the first giving an indication regarding the identity of a sample, the latter being more specific). In forensic settings these tests, such as the RSID (Rapid Stain Identification) or prostate specific antigen (PSA) test, are applied mainly for the indicative identification of seminal fluid, blood and saliva [4]. The major drawback of these tests is that they can only be used to identify a limited number of cell types and false positive and negative scorings can occur [4][10]. With RNA profiling, the number of cell types available for pre-screening would significantly increase and encompass for example vaginal mucosa and gender (note that the XIST RNA marker is the only true female marker described in forensic literature [19]). A direct RT-qPCR approach is proposed to this aim. The approach is highly similar to that used in the DNA6hours procedure, which is currently used for rapid STR profiling [47-48]. A lot of time is gained by omitting steps such as cell lysis, co-extraction and DNase treatment. From an evidentiary sample a small amount

is collected by tape-lifting using a mini-stub [47-48]; the remainder of the sample stays dry and intact. The cell material on the mini-tape is submitted to a short lysis, and the lysate is used as input for a one-step RT-qPCR which, in one reaction, combines two enzymes for reverse transcription of RNA to cDNA and the cDNA amplification of genes of interest. These genes can be the same as the ones used in currently-used mRNA-based assays, thus using the knowledge obtained in all earlier RNA work. Alike the commonly used RNA assays, interpretation of the direct RT-qPCR assay will provide information on the presence of a body fluid. Furthermore, the obtained lysates could additionally be used for direct STR-profiling assays. This rapid assay, that is more sensitive, human-specific and additionally allows to simultaneously target more body fluids, could potentially replace currently-used presumptive and confirmatory tests. In forensic casework, this rapid tool can serve as a selection tool, in which knowledge regarding the cell type including the gender of the donor contributing to a trace can aid in selecting the traces for which to proceed with regular DNA and/or RNA profiling. This tool can for example be of aid in identifying a male blood stain in a collection of mainly female blood spatters. This rapid RNA assay is more investigative in nature while current RNA profiling is foremost evaluative. Thus, the assays will not replace each other.

"MPS is the future of forensics, including RNA profiling"

In the past few years, massively parallel sequencing (MPS) has made its rise in forensic science. MPS is a high-throughput sequencing technique that allows for the simultaneous determination of individual sequences of large numbers of molecules [49]. Conventional methods for DNA and/or RNA analysis mostly focus on profiling via PCR amplification and capillary electrophoresis (CE). Although the CE-approach is highly useful for analysis of samples that vary in lengths, MPS additionally provides information regarding the actual sequence of the amplified fragment, which greatly increases discriminating power [49].

MPS for forensic RNA applications is relatively new and so far, only few approaches have been used. Of course, whole transcriptome sequencing (RNA-Seq) can provide massive amounts of information regarding the full transcriptome of organisms and tissue types [50]. RNA-Seq data can be used in many ways, for example to identify expressed transcripts, examine the transcriptional structure of genes (such as start sites, 5' and 3' ends, (alternative) splicing patterns and RNA editing) and to examine changes in expression levels [51]. For forensic cell type inferring purposes, whole transcriptome sequencing is mainly used to identify novel markers. This can be complicated when a forensically relevant sample contains a low percentages of human material (for example saliva and vaginal samples comprise mainly microbes; only 5-10% of sequences are reported to map the human genome [27]). Previously, microarray studies were performed to examine gene expression by using a selection of sequences with known identities. MPS allows analysis of the complete transcriptome including both known and unknown genes. Once genes of interest are identified, targeted sequencing can be applied. In the future, this could be used to solve the earlier described drawback regarding the deconvolution of RNA profiles to associate cell types with individual donors [52]. Although MPS may indeed be the future of forensics on DNA level, its added value on RNA level will depend on the feasibility to associate DNA and RNA results based on RNA SNPs. Should RNA SNPs be of insufficient or limited added value, RNA profiling may well be performed using capillary electrophoresis (CE), as then MPS provides similar information compared to CE-based methods. If RNA marker analysis is to be incorporated in MPS multiplexes targeting DNA (which can be various types of markers such as autosomal-, Y- and X-STRs, or SNPs for identification, ancestry or externally visible characteristics), the issue of reverse transcription needs to be accommodated in the protocol.

Alternatively, MPS could be considered to establish a more quantitative analysis to provide insight in the gene expression variations that can be observed for different markers. This could additionally provide insight when analysing mixtures of multiple cell types. This could be achieved by using a two-step PCR, targeted sequencing approach [53]. In the first PCR, random tag primers are incorporated to the target

that consist of 1) a gene-specific primer (thus targeted); 2) a unique tag of *e.g.* 10 random nucleotides in length (in this example yielding 410 or 1,048,576 unique tags); 3) a common barcode, that is the target of the second PCR. During the first PCR, these random tag primers are merely annealed and extended (no exponential amplification). A second enrichment PCR is performed that uses primers targeting the common barcodes (3). This way, each original target of interest is tagged by a unique tag (2). After sequencing, the number of unique tags per target represents the number of original molecules, providing quantitative information for each target. Up to now, this approach has mainly been tested and promoted by commercial companies. One of the major drawbacks of this method up to date is the low sensitivity of the assay due to which it is not yet applicable in forensic settings.

Concluding remarks

Since the first publication in 1999 [54], RNA profiling for the purpose of body fluid identification has been investigated almost continuously [55- 69]. While RNA profiling can be applied also for alternative aims (such as determining the time since deposition or estimating the postmortem interval), in forensic casework, its main application at the moment is inference of cell types. Up to date, there are at least two forensic laboratories that routinely apply RNA profiling for the purpose of cell type inference in casework. These are ESR (the institute of Environmental Science and Research) in New Zealand, that applies RNA profiling for the inference of body fluids since 2010, and the NFI in The Netherlands, that uses both body fluid and organ tissue inferring assays since 2010 and 2012, respectively. Organ typing has not been implemented at ESR because of a low rate of shootings (SallyAnn Harbison, personal communication). Besides, some laboratories use RNA profiling occasionally, often as a follow-up of collaborative exercises in which NFI primer mixes were shared. Other laboratories seem reluctant in implementing RNA profiling in forensic casework, which may be due to the labour intensity of both the implementation procedure and the RNA profiling itself and the low frequency of cases in which RNA profiling is requested in smaller laboratories. Although RNA profiling is not subject to DNA law, absence of acceptance in court in a country may withhold laboratories to proceed to RNA profiling especially in countries that run by the adversarial system. The fact that RNA profiling is routinely used for forensic casework in both a country in which the adversarial is used and a country run by the inquisitorial legal system may be helpful. In our experience, a steady inflow of cases is required to retain the essential awareness and expertise in analysing RNA data. The use of a uniform set of markers and interpretation guidelines may be necessary in making RNA profiling more approachable. Also, when RNA markers would be incorporated into commercial MPS kits with clear protocols and interpretation software, the implementation hurdle might be lessened for many laboratories. Major future advances could lie with rapid RNA-based screening, which could serve not only as a replacement or extension for currently-used presumptive tests, but also as a selection tool for identifying traces of interest. Furthermore, when incorporated in MPS platforms and thereby extending the analysis from amplicon length-based to sequence-based, RNA may in the future provide information regarding the cell type as well as information to associate cell type to a donor. This could provide the missing link between RNA and DNA profiling.

References

1. Jackson G. (2009). Understanding forensic science opinions, in Handbook of Forensic Science, eds Fraser J., Williams R., editors. (Cullompton: Willan Publishing;), 419–455
2. Cook, R., Evett, I.W., Jackson, G., Jones, P.J., Lambert, J.A., A hierarchy of propositions: deciding which level to address in casework, Sci. Justice 38 (1998) 231–239.

3. Evett, I.W., Gill, P.D., Jackson, G., Whitaker, J., Champod, C., Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks, J. Forensic Sci. 47 (2002) 520–530.
4. Virkler, K., Lednev, I. K. Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene. Forensic Sci. Int. (2009) 188(1), 1-17.
5. Takata T, Miyaishi S, Kitao T, Ishizu H., Identification of human brain from a tissue fragment by detection of neurofilament proteins. Forensic Sci Int (2004) 144:1–6
6. Kimura A, Ikeda H, Yasuda S, Yamaguchi K, Tsuji T., Brain tissue identification based on myosin heavy chain isoforms. Int J Legal Med (1995) 107:193–196
7. Seo Y, Kakizaki E, Takahama K., A sandwich enzyme immunoassay for brain S-100 protein and its forensic application. Forensic Sci Int (1997) 87:145–154
8. Nichols CA, Sens MA., Recovery and evaluation by cytologic techniques of trace material retained on bullets. Am J Forensic Med Pathol (1990) 11:17–34
9. Lindenbergh, A., de Pagter, M., Ramdayal, G., Visser, M., Zubakov, D., Kayser, M., Sijen, T., A multiplex (m) RNA-profiling system for the forensic identification of body fluids and contact traces, Forensic Sci. Int.: Genet. 6 (2012) 565–577.
10. Sijen,T., Molecular approaches for forensic cell type identification: on mRNA miRNA,DNA methylation and microbial markers, Forensic Sci. Int. Genet 18 (2015) 21–32.
11. Bianconi E, Piovesan A, Facchin F, et al.: An estimation of the number of cells in the human body. Ann Hum Biol 2013; 40:463-471.
12. Fantom Consortium. A promoter-level mammalian expression atlas. Nature, (2014) 507(7493), 462-470.
13. Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M. F., Kellis, M., Lindblad-Toh, K., Lander, E. S. Distinguishing protein-coding and noncoding genes in the human genome. Proceedings of the National Academy of Sciences, (2007) 104(49), 19428-19433.
14. Shapiro, J. A. Revisiting the central dogma in the 21st century. Annals of the New York Academy of Sciences, (2009) 1178(1), 6-28.
15. Lindenbergh, A., van den Berge, M., Oostra, R.J., Cleypool, C., Bruggink, A., Kloosterman, A., Sijen, T., Development of a mRNA profiling multiplex for the inference of organ tissues, Int. J. Legal Med. 127 (2013) 891–900.
16. van den Berge, M., Carracedo, A., Gomes, I., Graham, E.A.M., Haas, C. Hjort, B., Hoff-Olsen, P., Maronas, O., Mevag, B., Morling, N., A collaborative European exerciseon mRNA-based body fluid/skin typing and interpretation of DNA and RNAresults, Forensic Sci. Int.: Genet. 10 (2014) 40–48.
17. 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 Sci. Int.: Genet. 20 (2016) 119-129.
18. van den Berge, M., Wiskerke, D., Gerretsen, R.R.R., Tabak, J., Sijen, T., DNA and DNA and RNA profiling of excavated human remains with varying postmortem intervals, Int. J. Legal Med. 130 (2016) 1471-1480.
19. van den Berge, M., Sijen, T., A male and female RNA marker to infer sex in forensic analysis. Forensic Sci. Int.: Genet. 26 (2017) 70-76.
20. van den Berge, M., Sijen, T., Extended specificity studies of mRNA assays used to infer human organ tissues and body fluids. Electrophoresis (2017)
21. Fordyce SL, Kampmann ML,Van Doorn NL, Gilbert MTP., Long-term RNA persistence in postmortem contexts. Investig. Genet (2013) 4(1):1
22. Lindahl T Instability and decay of the primary structure of DNA. Nature (1993) 362(6422):709–715
23. Varani G, McClain WH The G·U wobble base pair. EMBORep (2000) 1(1):18–23
24. Suay L, Salvador ML, Abesha E, Klein U., Specific roles of 5'RNA secondary structures in stabilizing

transcripts in chloroplasts.Nucleic Acids Res (2005) 33(15):4754–4761
25. Mitchell P,Tollervey D mRNA turnover. CurrOpin CellBiol (2001) 13(3):320–325
26. Pérez-Ortín, J. E., Alepuz, P., Chávez, S., & Choder, M. Eukaryotic mRNA decay: methodologies, pathways, and links to other stages of gene expression. Journal of molecular biology, (2013) 425(20), 3750-3775.
27. Weinbrecht K.D., Fu, J., Payton M.E., Allen, R.W.,Time-dependent loss of mRNA transcripts fromforensic stains, Research and Reports in Forensic Medical Science 2017:7 1–12
28. Nicklas, J. A., Noreault-Conti, T., Buel, E. Development of a Real-Time Method to Detect DNA Degradation in Forensic Samples. Journal of forensic sciences, (2012) 57(2), 466-471.
29. Moreno, L. I., Tate, C. M., Knott, E. L., McDaniel, J. E., Rogers, S. S., Koons, B. W., Kavlick, M.F., Craig, R.L., Robertson, J. M. Determination of an effective housekeeping gene for the quantification of mRNA for forensic applications. Journal of forensic sciences, (2012) 57(4), 1051-1058.
30. Suzuki, T., Higgins, P. J., Crawford, D. R. Control selection for RNA quantitation. Biotechniques (2000) 29(2), 332-337.
31. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., &Speleman, F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology, (2002) 3(7), research0034-1.
32. Lindenbergh A, Maaskant P, Sijen T., Implementation of RNAprofiling in forensic casework. Forensic Scilnt Genet (2013) 7(1):159–166
33. Benschop, C.C.G., van der Beek, C.P., Meiland, H.C., van Gorp, A.G., Westen, A.A., Sijen, T., Low template STR typing: effect of replicate number and consensus method ongenotyping reliability and DNA database search results, Forensic Sci. Int. Genet. 5(2010) 316–328.
34. 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, (2013) 127(4), 707-721.
35. de Zoete, J., Curran, J.,Sjerps, M. A probabilistic approach for the interpretation of RNA profiles as cell type evidence. Forensic Science International: Genetics, (2016) 20, 30-44.
36. Haas, C., Klesser, B., Maake, C., Bär, W., Kratzer, A. mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. Forensic Science International: Genetics, (2009) 3(2), 80-88.
37. Harteveld, J., Lindenbergh, A., Sijen, T. RNA cell typing and DNA profiling of mixed samples: can cell types and donors be associated?. Science & Justice, (2013) 53(3), 261-269.
38. Omedei, M., Gino, S., Inturri, S., Pasino, S.,Robino, C. Individual assignment of body fluids in mixed stains by synonymous SNP analysis. Forensic Science International: Genetics Supplement Series, (2013) 4(1), e19-e20.
39. Kumar, P., Henikoff, S., Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols, (2009) 4(7), 1073-1081.
40. Hinnebusch, A. G., Ivanov, I. P.,Sonenberg, N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. Science, (2016) 352(6292), 1413-1416.
41. Andreassi, C., Riccio, A. To localize or not to localize: mRNA fate is in 3' UTR ends. Trends in cell biology, (2009) 19(9), 465-474.
42. Flicek, P., et al., Ensembl 2014, Nucleic Acids Res. 42 (2014) D749–D755.
43. Bleka, Ø., Eduardoff, M., Santos, C., Phillips, C., Parson, W., Gill, P. Open source software EuroForMix can be used to analyse complex SNP mixtures. Forensic Science International: Genetics, (2017) 31, 105-110.
44. Gill, P., Haned, H., Eduardoff, M., Santos, C., Phillips, C., Parson, W. The open-source software LRmix can be used to analyse SNP mixtures. Forensic Science International: Genetics Supplement Series, (2015) 5, e50-e51.

45. Blackman, S., Stafford-Allen, B., Hanson, E., Panasiuk, M., Brooker, A., Rendell, P., Ballantyne, J., Wells, S. Developmental Validation of the ParaDNA® Body Fluid ID System. Forensic Science International: Genetics Supplement Series (2017).
46. Doole, S. An investigation into the use of the ParaDNA® body fluid identification system in forensic examinations. Forensic Science International: Genetics Supplement Series (2017).
47. Verheij, S., Harteveld, J., & Sijen, T., A protocol for direct and rapid multiplex PCR amplification on forensically relevant samples. Forensic Science International: Genetics, (2012) 6(2), 167-175.
48. Verheij, S., van Gorp, A. G., Benschop, C. C., Harteveld, J., Matai, A. S., Sijen, T., Implementation and first case results of a rapid DNA profiling service. Forensic Science International: Genetics Supplement Series, (2011) 3(1), e427-e428.
49. Parson, W., Ballard, D., Budowle, B., Butler, J. M., Gettings, K. B., Gill, P., Gusmão, L., Hares, D.R., Irwin, J.A., King, J.L., de Knijff, P., Morling, N., Prinz, M., Schneider, P.M., Van Neste, C., Willuweit, S., Phillips, C., Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. Forensic Science International: Genetics, (2016) 22, 54-63.
50. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, A., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data, Nat Biotechnol. 2011 July; 29(7): 644–652.
51. Wang, Z., Gerstein, M., Snyder, M., RNA-Seq: a revolutionary tool for transcriptomics, Nat Rev Genet. 2009 Jan; 10(1): 57–63.
52. Ingold, S., Haas, C., Dørum, G., Hanson, E., & Ballantyne, J. Association of a body fluid with a DNA profile by targeted RNA/DNA deep sequencing. Forensic Science International: Genetics Supplement Series. (2017)
53. Peng, Q., Satya, R.V., Lewis, M., Randad, P. and Wang, Y., Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. BMC genomics, 2015, 16(1), p.589.
54. Bauer, M., Kraus, A. and Patzelt, D., . Detection of epithelial cells in dried blood stains by reverse transcriptase-polymerase chain reaction. Journal of Forensic Science 1999, 44(6), pp.1232-1236.
55. Haas C,, Klesser, B., Maake, C., Bar, W., Kratzer, A., mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR, Forensic Sci. Int. Genet. 3 (2009) 80–88.
56. Hanson, E.K. Ballantyne, J., Highly specific mRNA biomarkers for the identification of vaginal secretions in sexual assault investigations, Sci. Justice 53 (2013) 14–22.
57. J. Juusola, J. Ballantyne, Multiplex mRNA profiling for the identification of body fluids, Forensic Sci. Int. 152 (2005) 1-12.
58. Nussbaumer, C., Gharehbaghi-Schnell, E. and Korschineck, I., . Messenger RNA profiling: a novel method for body fluid identification by real-time PCR. Forensic science international 2006, 157(2), pp.181-186.
59. Xu, Y., Xie, J., Cao, Y., Zhou, H., Ping, Y., Chen, L., Gu, L., Hu, W., Bi, G., Ge, J. and Chen, X., Development of highly sensitive and specific mRNA multiplex system (XCYR1) for forensic human body fluids and tissues identification. PloS one, 2014, 9(7), p.e100123.
60. Gomes, I., Kohlmeier, F. and Schneider, P.M., Genetic markers for body fluid and tissue identification in forensics. Forensic Science International: Genetics Supplement Series, 2011, 3(1), pp.e469-e470.
61. Park, S.M., Park, S.Y., Kim, J.H. , Kang, T.W., Park, J.L., Woo, K.M., Kim, J.S., Lee, H.C., Kim, S.Y., Lee S.H., Genome-wide mRNA profiling and multiplex quantitative RT-PCR for forensic body fluid identification, Forensic Sci. Int.: Genet. 7 (2013) 143–150.
62. Cossu, C., Germann, U., Kratzer, A., Baer, W., Haas, C., How specific are the vaginal secretion mRNA-markers HBD1 and MUC4, Forensic Sci. Int.: Genet. Suppl. Ser. 2 (2009) 536–537.

63. M.Visser, D. Zubakov, K.N. Ballantyne, M. Kayser, mRNA-based skin identification for forensic applications, Int. J. Legal Med. 125 (2011) 253–263.
64. R. Fang, C.F. Manohar, C. Shulse, M. Brevnov, A. Wong, O.V. Petrauskene, P. Brzoska, M.R. Furtado, Real-time PCR assays for the detection of tissue and body fluid specific mRNAs, Int. Congr. Ser. 1288 (2006) 685–687.
65. R.I. Fleming, S. Harbison, The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids, Forensic Sci. Int.: Genet. 4 (2010) 244–256.
66. M. van den Berge, T. Sijen, Advancing forensic RNA profiling: Preventing noise signals in RNA profiling by adding the multiplex buffer last. Forensic Sci. Int.: Genet. Suppl. Ser. 5 (2015) e518-e519.
67. M. van den Berge, G. Ozcanhan, S. Zijlstra, A. Lindenbergh, T. Sijen, Prevalence of human cell material: DNA and RNA profiling of public and private objects and after activity scenarios. Forensic Sci. Int.: Genet. 21 (2016) 81-89.
68. M. van den Berge, L. van de Merwe, T. Sijen, DNA transfer and cell type inference to assist activity level reporting: Post- activity background samples as a control in dragging scenario. Forensic Sci. Int.: Genet. Suppl. Ser. 5 (2017) e518-e519.
69. Hanson, E., & Ballantyne, J. (2017). Human Organ Tissue Identification by Targeted RNA Deep Sequencing to Aid the Investigation of Traumatic Injury. Genes, 8(11), 319.
70. A.Q. Gomes, S. Nolasco, H. Soares, Non-coding RNAs multi-tasking molecules in the cell, Int. J. Mol. Sci. 14 (2013) 16010–16039.
71. He, L., & Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nature Reviews Genetics, 5(7), 522-531.
72. Plasterk, R. H. (2006). Micro RNAs in animal development. Cell, 124(5), 877-881.
73. B.P. Lewis, C.B. Burge, D.P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell 120 (2005) 15–20.
74. Bird, A., & Jaenisch, R. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature genetics, 33, 245.
75. Vidaki, A., Daniel, B., & Court, D. S. (2013). Forensic DNA methylation profiling—Potential opportunities and challenges. Forensic Science International: Genetics, 7(5), 499-507.
76. Rakyan, V. K., Down, T. A., Thorne, N. P., Flicek, P., Kulesha, E., Graf, S., Tomazou, E. M., Backdahl, L., Johnson, N., Herberth, M., Howe, K. L., Jackson, D. K., Miretti, M. M., Fiegler, H., Marioni, J. C., Birney, E., Hubbard, T. J., Carter, N. P., Tavare, S., Beck, S., Genome Res. 2008, 18, 1518–1529.
77. Thompson, T. M., Sharfi, D., Lee, M., Yrigollen, C. M., Naumova, O. Y., Grigorenko, E. L., Behav. Genet. 2013, 43, 168–176
78. Lee, H. Y., An, J. H., Jung, S. E., Oh, Y. N., Lee, E. Y., Choi, A., Yang, W.I., Shin, K. J. (2015). Genome-wide methylation profiling and a multiplex construction for the identification of body fluids using epigenetic markers. Forensic Science International: Genetics, 17, 17-24.
79. C. Grunau, S.J. Clark, A. Rosenthal, Bisulfite genomic sequencing: systematic investigation of critical experimental parameters, Nucleic Acids Res. 29 (2001) E65.
80. Frumkin, D., Wasserstrom, A., Budowle, B., Davidson, A., Forensic Sci. Int. Genet. 2011, 5, 517–524
81. Y.C. Lin, L.C. Tsai, J.C.I. Lee, C.W. Su, J.T.C. Tzen, A. Linacre, H.M. Hsieh, Novel identification of biofluids using a multiplex methylation sensitive restriction enzyme-PCR system. Forensic Science International: Genetics (2016), 25, 157-165.
82. Lam, L. L., Emberly, E., Fraser, H. B., Neumann, S. M., Chen, E., Miller, G. E., & Kobor, M. S. (2012). Factors underlying variable DNA methylation in a human community cohort. Proceedings of the National Academy of Sciences, 109(Supplement 2), 17253-17260.
83. J. Kuczynski, E.K. Costello, D.R. Nemergut, J. Zaneveld, C.L. Lauber, D. Knights, O. Koren, N. Fierer, S.T. Kelley, R.E. Ley, Direct sequencing of the human microbiome readily reveals community differences,

Genome Biol. 11 (2010) 210.
84. Fleming, R. I., & Harbison, S. (2010). The use of bacteria for the identification of vaginal secretions. Forensic Science International: Genetics, 4(5), 311-315.
85. Benschop, C. C., Quaak, F. C., Boon, M. E., Sijen, T., & Kuiper, I. (2012). Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context?. International journal of legal medicine, 126(2), 303-310.
86. P. Jorth, K.H. Turner, P. Gumus, N. Nizam, N. Buduneli, M. Whiteley, Metatranscriptomics of the human oral microbiome during health and disease, mBio 5 (2014) e01012–e01014.
87. B.A. White, D.J. Creedon, K.E. Nelson, B.A. Wilson, The vaginal microbiome in health and disease, Trends Endocrinol. Metab. 22 (2011) 389–393.
88. C. Haas, E. Hanson, M.J. Anjos, K.N. Ballantyne, R. Banemann, B. Bhoelai, E. Borges, M. Carvalho, C. Courts, 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.
89. Harbison, S.A., Fleming, R., Forensic body fluid identification: state of the art, Research and Reports in Forensic Medical Science 2016:6
90. Dammeier, S., Nahnsen, S., Veit, J., Wehner, F., Ueffing, M., & Kohlbacher, O. (2015). Mass-Spectrometry-Based Proteomics Reveals Organ-Specific Expression Patterns To Be Used as Forensic Evidence. Journal of proteome research, 15(1), 182-192.
91. Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Al-Khalili Szigyarto, C., Odeberg, J., Djureinovic, D., Ottosson Takanen, J., Hober, S., Alm, T., Edqvist, P., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., Pontén, (2015). Tissue-based map of the human proteome. Science, 347(6220), 1260419.
92. Anderson, S., Howard, B., Hobbs, G. R. & Bishop, C. P. A method for determining the age of a bloodstain. Forensic Sci. Int. 148, 37–45 (2005).
93. Anderson, S. E., Hobbs, G. R. & Bishop, C. P. Multivariate analysis for estimating the age of a bloodstain. J. Forensic Sci. 56, 186–193 (2010)
94. Lech, K., Liu, F., Ackermann, K., Revell, V. L., Lao, O., Skene, D. J., & Kayser, M. (2016). Evaluation of mRNA markers for estimating blood deposition time: Towards alibi testing from human forensic stains with rhythmic biomarkers. Forensic Science International: Genetics, 21, 119-125.
95. Lech, K., Ackermann, K., Revell, V. L., Lao, O., Skene, D. J., & Kayser, M. (2016). Dissecting daily and circadian expression rhythms of clock-controlled genes in human blood. Journal of biological rhythms, 31(1), 68-81.
96. Sibbens, L., Van de Voorde, W., Decorte, R., & Bekaert, B. (2017). The development of a forensic clock to determine time of death. Forensic Science International: Genetics Supplement Series.
97. Goff ML (2009) Early post-mortem changes and stages of decomposition in exposed cadavers. ExpApplAcarol 49(1–2):21–36
98. Hau TC, Hamzah NH, Lian HH, Hamzah SPAA (2014) Decomposition process and post mortem changes: review. SainsMalaysiana 43(12):1873–1882
99. Johnson LA, Ferris JA (2002) Analysis of postmortem DNA degradation by single-cell gel electrophoresis. Forensic SciInt 126(1): 43-47
100. Calacal GC, Apaga DLT, Salvador JM, Jimenez JAD, Lagat LJ, Villacorta RPF, De Ungria MCA (2015) Comparing different post-mortem human samples as DNA sources for downstream genotyping and identification. Forensic SciInt Genet 19:212–220
101. Courts C, Sauer E, Hofmann Y, Madea B, Schyma C (2015) Assessment of STR typing success rate in

soft tissues from putrefied bodies based on a quantitative grading system for putrefaction. J Forensic Sci 60(4):1016–1021
102. Hansen J, Lesnikova I, Funder AMD, Banner J (2014) DNA and RNA analysis of blood and muscle from bodies with variable postmortem intervals. Forensic Sci Med Pathol 10(3):322–328
103. Itani M, Yamamoto Y, Doi Y, Miyaishi S (2011) Quantitative analysis of DNA degradation in the dead body. Acta Med kayama 65(5):299–306
104. Ebuehi OA, Amode M, Balogun A, Fowora A (2015) Postmortem time affects brain, liver, kidney and heart DNA in male rat. Am J Biochem 5(1):1–5
105. Pozhitkov, A. E., & Noble, P. A. (2017). Gene expression in the twilight of death. Bioessays, 39(9).
106. Pozhitkov, A. E., Neme, R., Domazet-Lošo, T., Leroux, B. G., Soni, S., Tautz, D., & Noble, P. A. (2017). Tracing the dynamics of gene transcripts after organismal death. Open biology, 7(1), 160267.
107. Zhao, D., Ishikawa, T., Quan, L., Michiue, T., Zhu, B. L., & Maeda, H. (2009). Postmortem quantitative mRNA analyses of death investigation in forensic pathology: an overview and prospects. Legal Medicine, 11, S43-S45.
108. Maeda, H., Ishikawa, T., & Michiue, T. (2011). Forensic biochemistry for functional investigation of death: concept and practical application. Legal Medicine, 13(2), 55-67.
109. J. Sun, X. Zhu, T. Dong, X. Zhang, Q. Liu, S. Li, Q. Du, An "up, no change, or down" system: Time-dependent expression of mRNAs in contused skeletal muscle of rats used for wound age estimation, Forensic Science International 272 (2017) 104–110
110. T. Kondo, Y. Ishida, Molecular pathology of wound healing, Forensic Science International 203 (2010) 93–98
111. T. Kondo, T. Ohshima, W. Eisenmenger, Immunohistochemical and morphometrical study on the temporal expression of interleukin-1alpha (IL-1alpha) in human skin wounds for forensic wound age determination, Int J Legal Med (1999) 112 :249–252
112. S.F. Ibrahim, M. Issak, A.A. Bayoumy, D.S. Abd El-Fatah, Cutaneous (tPA) and Skeletal (TnI) mRNA as Markers of Aging in Contused Wound, J Forensic Sci, July 2016, Vol. 61, No. 4
113. Q. Du, J. Sun, L. Zhang, X. Liang, X. Guo, C. Gao, Y. Wang, Time-dependent expression of SNAT2 mRNA in the contused skeletal muscle of rats: a possible marker for wound age estimation, Forensic Sci Med Pathol (2013) 9:528–533
114. Y. Sato, T. Ohshima, The expression of mRNA of proinflammatory cytokines during skin wound healing in mice: a preliminary study for forensic wound age estimation (II), Int J Legal Med (2000) 113 :140–145
115. M. Takamiya, K. Saigusa, R. Kumagai, N. Nakayashiki, Y. Aoki, Studies on mRNA expression of tissue-type plasminogen activator in bruises for wound age estimation, Int J Legal Med (2005) 119: 16–21
116. P. Vanezis, Interpreting bruises at necropsy. J ClinPathol (2001) 54:348–355
117. Y. Wang, Y. Yamamoto, Y. Kuninaka, T. Kondo, F. Furukawa, Forensic Potential of MMPs and CC Chemokines for Wound Age Determination, J Forensic Sci, November 2015, Vol. 60, No. 6
118. S. Palagummi, SA Harbison, R. Fleming, A time-course analysis of mRNA expression during injury healing in human dermal injuries, Int J Legal Med (2014) 128:403–414
119. J. Sun, L. Nan, C. Gao, Y. Wang, Validation of reference genes for estimating wound age in contused rat skeletal muscle by quantitative real-time PCR, Int J Legal Med (2012) 126:113–120
120. J. Sun, Y. Wang, L. Zhang, C. Gao, L. Zhang, Z. Guo, Time-dependent expression of skeletal muscle troponin I mRNA in the contused skeletal muscle of rats: a possible marker for wound age estimation, Int J Legal Med (2010) 124:27–33
121. Y. Ishida, A. Kimura, M. Nosaka, Y. Kuninaka, E. Shimada, H. Yamamoto, K. Nishiyama, S. Inaka, T. Takayasu, W. Eisenmenger, T. Kondo, Detection of endothelial progenitor cells in human skin wounds and its application for wound age determination, Int J Legal Med (2015) 129:1049–1054

122. T. Yu, Z. Li, R. Zhao, Y. Zhang, Z. Zhang, D. Guan, Time-dependent Expression of MMP-2 and TIMP-2 after Rats Skeletal Muscle Contusion and Their Application to Determine Wound Age, J Forensic Sci, March 2016, Vol. 61, No. 2
123. X. Zhu, Q. Du, S. Li, J. Sun, Comparison of the homogeneity of mRNAs encoding SFRP5, FZD4, and Fosl1 in post-injury intervals: Subcellular localization of markers may influence wound age estimation, Journal of Forensic and Legal Medicine 43 (2016) 90e96

Chapter 11

Summary
Nederlandse samenvatting
List of publications
Dankwoord
Curriculum vitae

Summary

This thesis describes studies that assisted the expanding and advancement of forensic RNA profiling assays used for the inference of body fluids and organ tissues.

Even though DNA and RNA profiling are generally performed alongside and simultaneous to each other, both processes present their own challenges and require distinct expertise and interpretation strategies. **Chapter 1** describes a collaborative exercise undertaken between partners of the European Forensic Genetics Network of Excellence (EUROFORGEN-NoE) on mRNA-based body fluid typing and the interpretation of RNA and DNA data. Nine laboratories participated in the project for which an RNA assay and samples of increasing complexity were centrally prepared and thoroughly tested prior to distribution. The study aimed to give information on the overall sensitivity and specificity of the RNA assay and the reliability of the scoring system. Results of the study support the usefulness of this scoring system, which builds on replicate RNA analysis and the ratio between the number of observed and possible signals for each cell type. This study raises awareness regarding various potential interpretation pitfalls of RNA profiling, such as the realization that the same cell type can be donated by multiple donors, the same donor can donate multiple cell types, cell types can be masked by co-expressed cell-types (*e.g.* blood in a menstrual secretion sample) and cell types can provide RNA but no DNA (*e.g.* seminal fluid of an azoospermic male).

Chapter 2 describes various topics assessed to improve RNA profiling. Amongst others, bodily secretions that may be encountered at a crime scene whilst not targeted by the RNA assay were examined for false positive signals. This prompted us to identify a nasal mucosa marker, as 1) nasal mucosa carries a large number of human cells and represents therefore a likely alternative scenario and 2) markers for vaginal mucosa and menstrual secretion can be co-expressed in nasal mucosa. In addition, bacterial markers (*Lactobacillus*) are described to be unsuitable as an alternative for vaginal mucosa RNA markers, as background signals were detected in penile swabs that appeared devoid of female DNA. Furthermore, it is described that vaginal mucosa RNA marker MUC4 is best not considered when penile swabs are analysed. Also, a differential extraction protocol is proposed to facilitate RNA and DNA analysis of both the sperm and non-sperm fraction in sexual assault samples with highly deviating cell type ratios, such as in case of a surplus of (vaginal) epithelial cells and few sperm cells. Next, the relation between DNA and RNA profiling sensitivity was examined by comparing the percentage of alleles or markers on DNA and RNA level when analysing decreasing sample amounts. The associating of DNA and RNA results is discouraged, as the relative sensitivity of the two profiling systems varies for different cell types, donors and markers. Lastly, we show the necessity of using replicate analysis for interpretation of RNA data, as variations can be found even for true technical

replicates.

Alike for DNA profiling, increased noise signals can arise in RNA profiles that difficult interpretation of amplified signals and lead to false positive scoring of absent markers. In **Chapter 3**, various causes and solutions were considered to reduce low level noise signals detected in RNA profiles. The RNA typing multiplexes described in this thesis use custom-made fluorescently labelled primers in combination with commercially available multiplex amplification chemicals. As an increase in noise signals was observed also without the presence of template DNA, it is assumed that noise is caused by destabilisation or aggregation of the fluorescently labelled primers. However, as the composition of the commercial buffer is unknown, we cannot speculate what components may cause this instability.

In recent years, the sensitivity of profiling methodologies has increased tremendously and now allows for analysis of minute evidentiary samples that were previously regarded unsuitable for profiling. Especially when minute evidentiary traces are analysed, background cell material unrelated to the crime may contribute to detectable levels in the genetic analyses. **Chapter 4** describes a study that increased the understanding of the composition of human cell material residing on surfaces contributing to background in evidentiary traces. Samples were selected by considering events contributing to cell material deposits in exemplary activities (*e.g.* dragging a person by the trouser ankles), and can be grouped as public objects, private samples, transfer-related samples and washing machine experiments. To maximize profiling information, DNA and RNA extracts were concentrated prior to profiling when DNA yields were low. This study shows that high DNA yields do not necessarily relate to an increased number of contributors or to the detection of other cell types than skin, which is the cell type predominantly observed on any type of public or private item. Background cellular material may be found on any type of public or private item. When a major contributor can be deduced in DNA profiles from private items, this can be a different person than the owner of the item. Also, when a "perpetrator" performs a specific activity and the areas of physical contact are analysed, the "perpetrator" does not necessarily represent the major contributor in the DNA profile. While it was confirmed that the presence of sebum can promote DNA transfer, a less strong effect was seen for the presence of sweat, which may be because sweat itself contains only limited amounts of DNA. Washing machine experiments for saliva and bloodstains showed that transfer and persistence during laundry is limited for DNA, and cell type dependent for RNA.

Unlike in the activity-related experimental design described in Chapter 4, forensic casework does not allow for the analysis of pre-activity background samples from the exact same location as post-activity samples. **Chapter 5**, a sequel of Chapter 4, describes therefore the analysis of post-activity background samples taken from an untouched area, that were simultaneously collected post-dragging. This chapter highlights the relevance of including appropriate control region specimens for STR-

profiling in activity-related scenarios. While in post-activity samples alleles matching a perpetrator can be distinguished from background signals, both pre- and post-activity control region samples may yield alleles matching a perpetrator, that are however indistinguishable from background alleles.

The first five chapters of this thesis mainly focus on the RNA-based identification of body fluids, while in forensic cases, such as the event of a violent crime, knowledge regarding the tissue type may be of aid in the reconstruction of the event surrounding the crime. **Chapter 6** describes the development of a novel RNA-based tool for the inference of organ tissues. After applying various selection rounds including target and non-target tissues, markers regarded tissue-specific were combined in a multiplex assay. This assay allows for the inference of seven organ tissue types that are most likely to be found on trauma-causing objects, namely brain, lung, liver, skeletal muscle, heart, kidney and skin tissue. This tool can be of aid for example in shootings, where inferring what type of cell material resides on which bullet is useful for crime scene reconstruction. For complex cases involving multiple shooters, specific bullets with particular organ tissues may be linked to the individual firearms thereby indicating from which gun the lethal shot was fired.

For the applicability of the organ tissue typing assay in forensic casework, primate and target specificity of the system are essential as the presence of animal, rather than human tissue, is an alternative scenario (such as bullets carrying cell material from a hunting event) to be proposed. In **Chapter 7**, primate specificity of the Organtyper assay was confirmed by assessing organ tissue RNAs of various animals. Cross-testing results of both the Organtyper and the Cell-typer assay against target types covered by the other assay stress the importance of using multiple markers per cell type, as single markers may be detected in non-target tissues and the use of increased number of markers per cell type will allow inference with more certainty. Additionally, this chapter emphasizes the importance of awareness regarding unintended marker detection in non-target tissues that is crucial when applying mRNA profiling in casework, as this may lead to incorrect associations.

When postmortem intervals (PMIs) increase such as with longer burial times, human remains suffer increasingly from the taphonomic effects of decomposition processes such as autolysis and putrefaction. It is often assumed that RNA is less stable than DNA and may therefore be unsuitable for forensic analysis. In **Chapter 8**, an RNA-based organ typing assay and a range of DNA profiling techniques were applied to organ tissue samples obtained from excavated graves with burial times ranging from 4 to 42 years. The severely degraded organ tissues either showed poor overall profiling results or varying results for DNA and RNA profiling success, with sometimes DNA and other times RNA profiling being more successful. The association of DNA and/or RNA degradation and the postmortem interval is discouraged, as no relations were observed between nucleic acid profiling results and the PMI. Additionally, no relations were

observed between DNA and RNA profiling success, which is important knowledge for the interpretation of combined RNA and DNA profiling data. Furthermore, results show that not only DNA but also RNA molecules can be remarkably stable (successful for a tissue buried for over 40 years), which corroborate forensic applications. Lastly, insight that the brain and heart tissues tend to provide the best profiling results may change sampling policies in identification cases of degrading cadavers.

In forensics, DNA profiling is used for the identification of the donor of a trace, while mRNA profiling can be applied to identify the cellular origin such as body fluids or organ tissues. In **Chapter 9**, a novel application of RNA typing was explored, namely for the inference of sex. While the presence of male cell material can readily be inferred on DNA level from Y-chromosomal markers in quantitation or DNA profiling, no forensic markers exist for the positive identification of female cell material. The presence of female cell material is inferred only indirectly, *i.e.* from absence or unbalanced response of the Y-chromosomal marker. Female-specific markers may, however, be found on RNA level as certain cellular processes are specific to females. In mammals, the inactivation of one X-chromosome provides X-chromosomal dosage equivalence between XY males and XX females. Since this process involves specific RNA molecules, it was examined if such a marker could be used to identify female cellular material. Additionally, RNAs expressed from the Y-chromosome were examined for the identification of male material. After assessing for sex-specificity, sensitivity, suitability of identifying male-female mixtures and performance on degraded samples, the two sex-specific markers were incorporated in existing RNA assays. This is the first forensic assay enabling positive identification of female cellular material, and the first overlapping information in DNA and RNA profiles.

Chapter 10 is the general discussion of this thesis, which describes amongst others various general aspects regarding cell type inference, such as the use of alternative markers other than mRNA or the use of mRNA for alternative purposes. An overview is given regarding the development of RNA profiling assays at the NFI and their application to casework. Additionally, various concerns or beliefs that may arise concerning RNA profiling in forensic settings, and future developments regarding RNA-based cell type inference are discussed.

Nederlandse samenvatting

In forensisch strafrechtelijk onderzoek is DNA-onderzoek een belangrijk en vaak gebruikt hulpmiddel voor het identificeren van de donor van een spoor. In hedendaags forensisch onderzoek is de vraag echter niet alleen van wie een spoor afkomstig kan zijn, maar steeds vaker ook hoe het spoor daar is gekomen. Een aspect dat informatief kan zijn voor het bepalen van de ontstaanswijze van een spoor is het type cellen dat in het spoor aanwezig is. Het menselijk lichaam bestaat uit triljoenen cellen en tenminste 400 verschillende celtypen, zoals vetcellen en rode- of witte bloedcellen, die ieder hun eigen morfologie en functie hebben. Dit proefschrift beschrijft onderzoek naar het identificeren van celtypen met behulp van messenger RNA (mRNA). mRNA speelt een belangrijke rol in het bepalen van de functie van een cel, omdat het fungeert als boodschapper tussen de informatie vanuit genen in DNA en het uiteindelijke eiwitproduct dat het echte werk in de cel verzet. mRNA fungeert in een cel als de architect die de bouwtekeningen in het DNA leest en de opbouw van de eiwitten dirigeert. Het is niet zo dat in alle cellen van ieder gen mRNA en eiwit wordt gemaakt. De expressie van genen kan gelimiteerd zijn tot een specifiek celtype; genen die coderen voor eiwitten die betrokken zijn in zuurstoftransport komen bijvoorbeeld wel tot expressie in rode bloedcellen, maar niet in vetcellen. Andere genen kunnen in ieder celtype aanwezig zijn, zoals zogenaamde "huishoudgenen". Hoewel de genomische informatie (DNA) van een persoon voor alle cellen gelijk is (met uitzondering van de geslachtscellen die elk de helft van het genoom bevatten), komt in elke cel slechts een subset van de genen als mRNA tot expressie. Dit aspect is bruikbaar in forensisch RNA-onderzoek om af te leiden met welk(e) celtype(n) we te maken hebben in een spoor.

De RNA-gebaseerde celtype-identificatiemethoden (assays) van het Nederlands Forensisch Instituut (NFI) richten zich op genen die hoofdzakelijk tot expressie komen in forensisch relevante lichaamsvloeistoffen (met name bloed, speeksel, vaginale mucosa, menstruele secretie, sperma en nasale mucosa (beter bekend als nasaal vocht) die onderscheiden kunnen worden met de "Cell-typer" assay) en orgaan weefsels (met name hersen-, long-, lever-, spier-, hart- en nierweefsel die onderscheiden kunnen worden met de "Organtyper" assay). Deze assays kunnen bijvoorbeeld worden toegepast in zedendelicten waarin de aanwezigheid van vaginale cellen of sperma ter discussie staat, of kunnen bijvoorbeeld van hulp zijn in de reconstructie van een geweldsmisdrijf door het eventueel aanwezige orgaanweefsel op een mes, kogel of ander betrokken voorwerp te onderzoeken.

Op het NFI is mRNA-profilering voor het onderscheiden van lichaamsvloeistoffen voor het eerst toegepast in forensisch zaakonderzoek in 2010. Drie jaar later werd de mRNA-gebaseerde orgaantypering-assay geïntroduceerd in zaakonderzoek. Sinds de

eerste toepassing in een NFI-zaak in 2010 en de tijd van schrijven (december 2017) zijn onderdelen van RNA-profilering in meer dan 200 zaken toegepast. Tot op heden zijn de RNA-resultaten in tien van deze zaken meegenomen in uitspraken in een Nederlandse rechtbank en twee keer in een buitenlandse rechtbank.

De verschillende hoofdstukken van dit proefschrift hebben het onderzoek naar het uitbreiden en verbeteren van forensische RNA-profilering voor het identificeren van lichaamsvloeistoffen en orgaanweefsels als leidraad. Dit onderzoek richtte zich niet alleen op de technische aspecten van de assays maar ook op het toetsen van interpretatierichtlijnen en het uitbreiden van de achtergrondkennis. Hierbij moet gedacht worden aan het aanpassen van de markers in zowel de lichaamsvloeistof- als de orgaantypering-assay wegens uiteenlopende redenen, een Europees samenwerkingsonderzoek naar onder andere de effectiviteit van interpretatierichtlijnen bij RNA-profilering, onderzoek naar de aanwezigheid van achtergrondcelmateriaal en naar de gevoeligheid, specificiteit en robuustheid van de assays bij monsters (samples) van allerlei origine (zoals dierlijk of uit opgegraven lichamen). Onderstaande alinea's geven een meer gedetailleerde samenvatting per hoofdstuk.

Hoewel DNA- en RNA-profilering in het algemeen vanuit eenzelfde bemonstering en op vergelijkbare wijze worden uitgevoerd, kennen beide processen hun eigen uitdagingen en vereisen ze specifieke expertise en interpretatiestrategieën. **Hoofdstuk 1** van dit proefschrift beschrijft een samenwerkingsonderzoek uitgevoerd tussen partners van het European Forensic Genetics Network of Excellence (EUROFORGEN-NoE) met betrekking tot RNA-gebaseerde lichaamsvloeistoftypering en de interpretatie van RNA- en DNA-data. Negen laboratoria deden mee in het project waarvoor een RNA-assay en samples van toenemende complexiteit centraal werden voorbereid en grondig werden getest voordat ze over de verschillende laboratoria werden verspreid. De studie had als doel om informatie over de algehele gevoeligheid en specificiteit van de RNA-assay en de betrouwbaarheid van het scoringssysteem te verkrijgen. Resultaten van dit onderzoek ondersteunen de bruikbaarheid van dit scoringssysteem, dat bouwt op herhaalde RNA-analyse en het percentage waargenomen ten opzichte van het aantal mogelijke signalen voor ieder celtype. Deze studie verhoogt het bewustzijn met betrekking tot verschillende mogelijke interpretatievalkuilen van RNA-profilering, zoals het besef dat hetzelfde celtype door meerdere donoren kan worden gedoneerd (bijvoorbeeld speeksel van twee verschillende personen), dezelfde donor meerdere celtypen kan bijdragen (bijvoorbeeld bloed en speeksel van dezelfde persoon), celtypen kunnen worden gemaskeerd door celtypen met gedeelde expressie (bijvoorbeeld bloed in het geval van menstruele secretie) en er celtypen zijn die wel RNA, maar geen DNA bijdragen (bijvoorbeeld spermavloeistof van een azoöspermische man (zoals in geval van sterilisatie)).

Hoofdstuk 2 beschrijft verschillende onderwerpen die zijn onderzocht om RNA-

profilering te verbeteren. Onder andere zijn lichaamsvloeistoffen die wel op een plaats delict kunnen worden aangetroffen maar niet onderscheiden kunnen worden met de RNA-assays onderzocht op de mogelijkheid van het geven van vals-positieve signalen (denk hierbij aan urine, feces, tranen en zweet). Dit heeft geleid tot de identificatie van een nasale mucosa-marker, omdat nasale mucosa een groot aantal menselijke cellen bevat en niezen een gangbare gebeurtenis is. Zodoende is de aanwezigheid van nasale mucosa een realistische bron van celmateriaal van een verdachte (in forensische termen: een realistisch alternatief scenario). De nasale mucosa-marker helpt ook om bloedneusbloed of uitgeademd bloed te kunnen identificeren. Daarnaast kunnen de markers voor vaginale mucosa en menstruele secretie soms ook tot expressie komen in nasale mucosa en met de aanwezigheid van een nasale mucosa-marker in de assay kan nu herkend worden dat het dan niet om vaginale mucosa of menstruele secretie gaat maar om nasale mucosa. Daarnaast werd getoetst of markers voor de bacteriële soort Lactobacillus, die in de literatuur waren beschreven als vaginale markers, hiervoor inderdaad geschikt waren. Echter omdat achtergrondsignalen werden gedetecteerd in penisafstrijkjes waar geen vrouwelijk DNA in aangetoond werd, zijn deze bacteriële markers verworpen. Het onderzoek aan deze penisafstrijkjes liet bovendien zien dat de vaginale mucosa RNA-marker MUC4 beter buiten beschouwing gelaten kan worden wanneer dergelijke bemonsteringen worden geanalyseerd omdat er vals-positieve signalen kunnen ontstaan. In zedenzaken kan bij inwendige bemonstering van vrouwen een speciaal protocol worden toegepast waardoor het DNA uit de grote overmaat van vaginale cellen en het DNA uit de weinige eventueel aanwezige spermacellen in andere fracties terecht komt (de sperma- en de niet-spermafractie). Dit heet differentiële extractie en er werd een protocol voorgesteld om naast DNA ook het RNA in deze twee fracties te extraheren. Daarnaast is de relatie tussen DNA- en RNA-profileergevoeligheid onderzocht door het percentage gedetecteerde kenmerken of markers op DNA- en RNA-niveau te vergelijken bij het analyseren van samples met afnemende hoeveelheid materiaal. Omdat de gevoeligheid op DNA- en op RNA-niveau niet gelijk op gaat en kan verschillen tussen de verschillende celtypen, donoren en de verschillende markers binnen hetzelfde celtype wordt het associëren van DNA- en RNA-resultaten afgeraden. Ten slotte wordt de noodzaak van het uitvoeren van herhaalde RNA-analyses voor de interpretatie van RNA-gegevens benadrukt, omdat zelfs in technische herhalingen variaties kunnen worden gevonden.

Net als bij DNA-profilering kunnen verhoogde ruissignalen ontstaan in RNA-profielen, welke de interpretatie kunnen bemoeilijken en in het ergste geval zouden kunnen leiden tot het vals-positief scoren van afwezige markers. In **hoofdstuk 3** zijn verschillende oorzaken en oplossingen voor het verminderen van deze ruissignalen in RNA-profielen getoetst. De RNA-assays beschreven in dit proefschrift maken gebruik van zelfontworpen primerparen (deze primerparen bepalen welk deel in een sequentie wordt geamplificeerd) met een fluorescent label aan één van de primers

en deze primers worden gebruikt in combinatie met commercieel verkrijgbare amplificatiechemicaliën. Omdat een toename van ruissignalen ook werd waargenomen zonder de aanwezigheid van celmateriaal, wordt aangenomen dat de ruis wordt veroorzaakt door destabilisatie of aggregatie van de fluorescent gelabelde primers. Omdat de samenstelling van de commercieel verkrijgbare buffer onbekend is, kunnen we echter niet achterhalen welke componenten deze instabiliteit kunnen veroorzaken; wel blijken de ruissignalen sterk te verminderen wanneer primers en buffer zo kort mogelijk hoog-geconcentreerd gemengd zijn.

In de afgelopen jaren is de gevoeligheid van de profileringsmethoden enorm toegenomen, wat het analyseren van minimale hoeveelheden celmateriaal, voorheen ongeschikt geacht voor profilering, mogelijk maakt. Vooral wanneer minimale hoeveelheden celmateriaal worden geanalyseerd kan achtergrondcelmateriaal dat geen verband houdt met het misdrijf gedetecteerd worden in de genetische analyse. **Hoofdstuk 4** beschrijft een onderzoek dat de kennis en het begrip vergroot aangaande de aanwezigheid van humaan celmateriaal op oppervlakken die bijdragen aan de "achtergrond" materie in bewijsmateriaal. Bemonsteringen werden geselecteerd door het beschouwen van gebeurtenissen die bijdragen aan het achterlaten van celmateriaal in activiteitgerelateerde scenario's (bijvoorbeeld het slepen van een persoon aan de broekspijpen). Onderzoek is gedaan naar openbare voorwerpen (denk aan deurklinken), privé-objecten (bijvoorbeeld kleding maar ook handen), bemonsteringen na een activiteit (het slepen van een vrijwilliger aan de enkels van de broekspijpen) en bemonsteringen na wasmachine-experimenten (is een broek echt helemaal vrij van humaan celmateriaal na wassen?). Deze studie leverde een aantal inzichten zoals: 1) hoge DNA-opbrengsten hebben niet noodzakelijkerwijs betrekking op een groot aantal donoren of de detectie van andere celtypen dan huid (het celtype dat voornamelijk wordt gedetecteerd op een openbaar of privé-voorwerp); 2) de donor waarvan afgeleid kan worden dat deze het meest heeft bijgedragen in de bemonstering (de zogenaamde DNA-hoofddonor) kan een ander persoon zijn dan de eigenaar van het item; 3) ook de persoon die de "daderrol" had bij een specifieke activiteit is maar in een beperkt percentage van de bemonsteringen de hoofddonor in het DNA-profiel. Deze inzichten helpen het interpreteren van resultaten in de forensische praktijk, met name wanneer niet zozeer betwist wordt of het celmateriaal van een verdachte op een stuk van overtuiging aanwezig is, maar vooral hoe het daar gekomen is (interpretatie op activiteitenniveau). Ook werd gekeken of de conditie van de huid een effect had op het bevorderen van DNA-overdracht; een sterk positief effect werd gevonden voor talg (de huid van de handpalmen bevat geen talgklieren, die op het gezicht en hoofdhuid wel; gewoontes als aanraken van gezicht of hoofdhuid bevorderen dus DNA-overdracht). Een beperkt effect werd gezien voor zweet, mogelijk omdat zweet zelf maar een beperkte hoeveelheid DNA bevat. De wasmachine-experimenten met speeksel- en bloedvlekken toonden aan dat overdracht en persistentie tijdens het wassen beperkt

zijn voor DNA en celtype-afhankelijk zijn voor RNA (speeksel en bloed wordt vrijwel helemaal weggewassen, terwijl huid nog enigszins detecteerbaar blijft).

In tegenstelling tot de activiteitgerelateerde experimentele opzet beschreven in Hoofdstuk 4, is het in forensisch zaakwerk niet mogelijk om achtergrondbemonsteringen af te nemen van exact dezelfde locatie vóór als na een bepaalde activiteit. **Hoofdstuk 5**, een vervolg op Hoofdstuk 4, beschrijft daarom de analyse van bemonsteringen van de kniegebieden van de broekspijpen wanneer vrijwilligers werden versleept door vastpakken bij de enkel. Voor een deel van de enkelbemonsteringen genomen na het slepen, worden duidelijk de DNA-kenmerken overeenkomend met de "dader" gedetecteerd; in de enkelbemonsteringen vóór slepen en de kniebemonsteringen na slepen komen de kenmerken die zouden kunnen overeenkomen met de "dader" niet boven de achtergrond uit. Daarmee benadrukt dit hoofdstuk de relevantie van het meenemen van bemonsteringen van geschikte controlegebieden bij activiteitgerelateerde scenario's.

De eerste vijf hoofdstukken van dit proefschrift richten zich voornamelijk op de RNA-gebaseerde identificatie van lichaamsvloeistoffen, terwijl in forensisch zaakonderzoek, zoals in het geval van een gewelddadig misdrijf, kennis over het type orgaanweefsel van nut kan zijn in het reconstrueren van de gebeurtenissen rondom het misdrijf. **Hoofdstuk 6** beschrijft de ontwikkeling van een RNA-gebaseerde assay voor het onderscheiden van orgaanweefsels. Na het toepassen van verschillende selectieronden, waarvoor onder andere niet-beoogde (non-target) weefsels geanalyseerd zijn om te onderzoeken of kruisreacties konden optreden in weefseltypen waarvoor de assay niet is ontwikkeld, zijn markers die weefselspecifiek werden geacht gecombineerd in één assay. Met deze assay kan onderscheid gemaakt worden tussen zeven orgaantypen die het meest waarschijnlijk zijn om aan te treffen op geweldsvoorwerpen, namelijk hersen-, long-, lever-, spier-, hart-, nier- en huidweefsel. Deze assay kan bijvoorbeeld worden gebruikt in het geval van een schietpartij, waar het afleiden van het type celmateriaal aanwezig op een kogel kan worden gebruikt voor de reconstructie van een misdrijf. In complexe zaken waarbij meerdere schutters betrokken zijn, kunnen specifieke kogels met een bepaald orgaanweefsel gelinkt worden aan individuele vuurwapens, waarmee kan worden achterhaald met welk vuurwapen het dodelijke schot is afgevuurd.

Voor de toepasbaarheid van de orgaantypering-assay in forensisch zaakwerk is humaan- en weefselspecificiteit van de assay essentieel, omdat de aanwezigheid van dierlijk in plaats van humaan materiaal als alternatief scenario kan worden geopperd (zoals een kogel met celmateriaal van een dier na jagen). In **hoofdstuk 7** is bevestigd dat de Organtyper-assay specifiek is voor mensen (strikt genomen primaten) door RNAs uit orgaanweefsel van verschillende dieren te analyseren. Toen de celtypen die doelwit (target) zijn van de Organtyper getest werden met de Cell-typer-assay en vice versa, bleken enkele individuele markers een respons te geven in non-target celtypen.

Dit benadrukt het belang van het gebruik van meerdere markers per celtype, zodat de zekerheid waarmee conclusies getrokken kunnen worden vergroot wordt.

Algemeen wordt aangenomen dat RNA minder stabiel is dan DNA en daarom wordt RNA vaak beperkt geschikt beschouwd voor forensische analyse. Zonlicht, vocht en tijd worden gezien als factoren die nucleïnezuren destabiliseren maar ook microben of de tafonomische processen (postmortale processen als autolyse en putrefactie) die plaatsvinden na overlijden kunnen een negatieve werking hebben. In **hoofdstuk 8** is de op RNA gebaseerde orgaantypering-assay en een scala aan DNA-profileringstechnieken toegepast op orgaanweefselsamples verkregen van lichamen variërend in postmortaal interval doordat deze lichamen 4 tot 42 jaar begraven hadden gelegen. Voor een aanzienlijk deel van deze organen konden de nucleïnezuren nog geanalyseerd worden ondanks dat de organen sterk aangetast waren door de tafonomische processen; soms was hierbij DNA- en soms RNA-profilering meer succesvol. Het wordt afgeraden om de mate van DNA- en/of RNA-degradatie te gebruiken om het postmortaal interval af te leiden, omdat er geen relaties zijn gevonden tussen nucleïnezuur resultaten en het postmortaal interval. Waarschijnlijk hebben vele andere factoren zoals het vochtgehalte van de bodem, type kleding en kist waarin overledene begraven is, sekse en doodsoorzaak een rol. Het feit dat met succes een RNA-profiel werd verkregen uit een weefsel dat meer dan 40 jaar begraven was ondersteunt sterk de forensische toepassing van RNA-profilering. Ten slotte kan het inzicht dat hersen- en hartweefsel de beste profileringsresultaten opleveren het bemonsteringsbeleid veranderen in geval van identificatiezaken aan ontbindende lichamen.

In forensisch onderzoek wordt DNA-profilering gebruikt voor de identificatie van de donor van een spoor, terwijl RNA-profilering kan worden gebruikt voor het identificeren van de cellulaire origine, zoals lichaamsvloeistoffen of orgaanweefsels. In **hoofdstuk 9** is een nieuwe toepassing van RNA-typering onderzocht, namelijk voor het onderscheiden van sekse. Terwijl de aanwezigheid van mannelijk celmateriaal kan worden afgeleid op DNA-niveau met behulp van Y-chromosomale markers in kwantificatie of DNA-profilering, bestaan er geen forensische markers die direct de aanwezigheid van vrouwelijk celmateriaal aantonen. De aanwezigheid van vrouwelijk materiaal wordt slechts indirect afgeleid, namelijk door de afwezigheid van de Y-chromosomale marker of een ongebalanceerd resultaat bij de bepaling van het X/Y-chromosomale kenmerk bij een man-vrouw mengsel. Vrouwspecifieke markers zouden echter op RNA-niveau gevonden kunnen worden, gezien bepaalde cellulaire processen specifiek voor de vrouw zijn. Zo komt bijvoorbeeld het proces dat zorgt voor de inactivatie van één van de twee vrouwelijke X-chromosomen ten behoeve van doseringsequivalentie tussen XY-mannen en XX-vrouwen alleen bij vrouwen voor. Omdat bij dit proces specifieke RNA-moleculen betrokken zijn, is onderzocht of deze RNAs als marker voor vrouwelijk celmateriaal gebruikt kunnen worden. Daarnaast zijn RNAs die op het Y-chromosoom tot expressie komen onderzocht voor het aantonen

van mannelijk celmateriaal tijdens RNA-profilering. Na het onderzoeken van de sekse-specificiteit, de gevoeligheid, de geschiktheid voor het identificeren van man-vrouw mengsels en de prestaties van de markers op gedegradeerde samples, zijn de twee sekse-specifieke markers opgenomen in de bestaande RNA-assays. Dit is de eerste forensische assay die het mogelijk maakt om vrouwelijk celmateriaal aan te tonen en de eerste overlappende informatie in DNA- en RNA-profielen.

Hoofdstuk 10, de algemene discussie van dit proefschrift, beschrijft onder andere verschillende algemene aspecten met betrekking tot het onderscheiden van celtypen, zoals het gebruik van alternatieve markers (anders dan mRNA), of het gebruik van mRNA voor alternatieve doeleinden. Een overzicht wordt gegeven van de ontwikkelingen van RNA-assays die gebruikt worden bij het NFI en hun toepassing in zaakwerk. Daarnaast worden verschillende bekommernissen of overtuigingen behandeld die kunnen leven ten aanzien van RNA-profilering in forensische context. Ook worden toekomstige ontwikkelingen met betrekking tot RNA-gebaseerde celtypeonderscheiding besproken.

List of publications

Publications on which this thesis was based

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.
**Both authors contributed equally*

M. van den Berge, A. Carracedo, I. Gomes, E.A.M. Graham, C. Haas, B. Hjort, P. Hoff-Olsen, O. Maronas, B. Mevag and N. Morling, H. Niederstätter, W. Parson, P.M. Schneider, D. Syndercombe Court, A. Vidaki, T. Sijen, A collaborative European exercise on mRNA-based body fluid/skin typing and interpretation of DNA and RNA results, Forensic Sci. Int.: Genet. 10 (2014) 40-48.

M. van den Berge, T. Sijen, Advancing forensic RNA profiling: Preventing noise signals in RNA profiling by adding the multiplex buffer last. Forensic Sci. Int.: Genet. Suppl. Ser. 5 (2015) e518-e519.

M. van den Berge, G. Ozcanhan, S. Zijlstra, A. Lindenbergh, T. Sijen, Prevalence of human cell material: DNA and RNA profiling of public and private objects and after activity scenarios. Forensic Sci. Int.: Genet. 21 (2016) 81-89.

M. van den Berge, B. Bhoelai, J. Harteveld, A. Matai, T. Sijen, 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 Sci. Int.: Genet. 20 (2016) 119-129.

M. van den Berge, D. Wiskerke, R.R.R. Gerretsen, J. Tabak, T. Sijen, DNA and DNA and RNA profiling of excavated human remains with varying postmortem intervals, Int. J. Legal Med. 130 (2016) 1471-1480.

M. van den Berge, T. Sijen, A male and female RNA marker to infer sex in forensic analysis. Forensic Sci. Int.: Genet. 26 (2017) 70-76.

M. van den Berge, T. Sijen, Extended specificity studies of mRNA assays used to infer human organ tissues and body fluids. Electrophoresis (2017) 38, 3155–3160

M. van den Berge, L. van de Merwe, T. Sijen, DNA transfer and cell type inference to assist activity level reporting: Post- activity background samples as a control in dragging scenario. Forensic Sci. Int.: Genet. Suppl. Ser. 6 (2017) e591-e592.

Publications besides this thesis

S. Verheij, L. Clarisse, **M. van den Berge**, T. Sijen, RapidHIT™ 200, a promising system for rapid DNA analysis, Forensic Sci. Int.: Genet. Suppl. Ser. 4 (2013) e254–e255

C. Haas, E. Hanson, R. Banemann, A.M. Bento, A. Berti, Á. Carracedo, C. Courts, G. De Cock, K. Drobnic, R. Fleming, C. Franchi, I. Gomes, G. Hadzic, S.A. Harbison, B. Hjort, C. Hollard, P. Hoff-Olsen, C. Keyser, A. Kondili, O. Maroñas, N. McCallum, P. Miniati, N. Morling, H. Niederstätter, F. Noël, W. Parson, M.J. Porto, A.D. Roeder, E. Sauer, P.M. Schneider, G. Shanthan, T. Sijen, D. Syndercombe Court, M. Turanská, **M. van den Berge**, M. Vennemann, A. Vidaki, L. Zatkalíková, J. Ballantyne, RNA/DNA co-analysis from human skin and contact traces–results of a sixth collaborative EDNAP exercise, Forensic Sci. Int.: Genet. 16 (2015) 139-147.

M. Marcińska, E. Pośpiech, J. Dyrberg Andersen, **M. van den Berge**, Á. Carracedo, M. Eduardoff, A. Marczakiewicz-Lustig, N. Morling, T. Sijen, M. Skowron, J. Söchtig, D. Syndercombe-Court, N. Weiler, The EUROFORGEN-NoE Consortium, D. Ballard, C. Børsting, W. Parson, C. Phillips, W. Branicki, Evaluation of DNA variants associated with androgenetic alopecia and their potential to predict male pattern baldness, PLoS ONE (2015); 10(5): e0127852.

E. Pośpiech, J. Karłowska-Pik, M. Marcińska, S. Abidi, J.D. Andersen, J. D., **M. van den Berge**, A. Carracedo, M. Eduardoffi A. Freire-Aradas, N. Morling, T. Sijen, M. Skowron, J. Söchtig, D. Syndercombe-Court, N. Weiler, P.M. Schneider, D. Ballard, C. Børsting, W. Parson, C. Phillips, W. Branicki, Evaluation of the predictive capacity of DNA variants associated with straight hair in Europeans. Forensic Sci. Int.: Genet. 19 (2015) 280-288.

E. Carnevali, D. Lacerenza, S. Severini, F. Alessandrini, C. Bini, C. Di Nunzio, M. Fabbri, P. Fattorini, A. Piccinini, E. Ponzano, G. Portera, C. Previderè, A. Renieri, F. Scarnicci, A. Verzeletti, S. Pelotti, **M. van den Berge**, T. Sijen, C. Robino. A GEFI collaborative exercise on DNA/RNA co-analysis and mRNA profiling interpretation. Forensic Sci. Int.: Genet. Suppl. Ser. 6 (2017) e18-e20.

S. Ingold, G. Dørum, E. Hanson, A. Berti, W. Branicki, P. Brito, P. Elsmore, K.B. Gettings, F. Giangasparo, T.E. Gross, S. Hansen, E.N. Hanssen, M.-L. Kampmann, M. Kayser, F.-X. Laurent, N. Morling, A. Mosquera-Miguel, W. Parson, C. Phillips, M.J. Porto, E. Pośpiech, A.D. Roeder, P.M. Schneider, K. Schulze Johann, C.R. Steffen, D. Syndercombe-Court, M. Trautmann, **M. van den Berge**, K.J. van der Gaag, J. Vannier, V. Verdoliva, A. Vidaki, C. Xavier, J. Ballantyne, C. Haas (2018). Body fluid identification using a targeted mRNA massively parallel sequencing approach–results of a EUROFORGEN/EDNAP collaborative exercise. Forensic Sci. Int.: Genet. 34 (2018) 105-115.

Dankwoord

Het voelt als de dag van gisteren dat ik als "broekie" begon aan mijn avontuur bij het NFI. De tijd is voorbij gevlogen, "het boekje" is af en het einde van mijn promotietijd in zicht. Natuurlijk schrijf je zo'n proefschrift niet alleen maar heb ik het geluk gehad dat ik gesteund ben door een heleboel mensen, waarvan ik er een aantal specifiek wil noemen.

In eerste plaatst uiteraard Titia, zonder wiens vertrouwen, vrijheid en steun ik deze mijlpaal nooit had kunnen halen. Vol passie, enthousiasme en humor kon je sparren over experimenten, resultaten en manuscripten. Ik waardeer enorm de tijd en energie die je door de jaren heen in mij hebt gestoken en waarmee we deze uitdaging tot een succesvol einde zullen brengen.

Geen promotie zonder promotor. Peter, grote waardering heb ik voor het vertrouwen en de kans die je mij hebt gegeven om bij jou te promoveren.

Wim, oud-afdelingshoofd van HBS, met wiens steun ik dit promotietraject kon starten. Alle (ex-)collega's en (ex-)stagiaires van het R&D-team, waarbij ik er een aantal in het bijzonder wil noemen. Alex, vol enthousiasme was jij mijn stagebegeleider toen ik voor het eerst bij het NFI begon. Ik denk nog vaak terug aan deze gave tijd: het opzetten van de orgaantypering-assay en hoe deze nu toegepast wordt in forensisch zaakonderzoek! Corina, bij wie ik altijd terecht kan voor vragen, input of "gewoon" een luisterend oor. Daarnaast uiteraard de stagiaires die ik heb mogen begeleiden gedurende de jaren, waardoor ikzelf ook ben gegroeid. Mijn (ex-)kamergenoten die het door de jaren heen met mij hebben volgehouden en zorgden voor gezelligheid, een minder serieuze noot en uiteraard de nodige flauwe grap.

Collega's binnen het NFI met wie ik de afgelopen jaren heb mogen samenwerken, met name van de divisies BiS, BDE en CFS en in het bijzonder de RNA-deskundigen op het gebied van RNA-zaakonderzoek.

Alle coauteurs van de wetenschappelijke artikelen die uit dit promotietraject zijn voortgekomen. Daarbij horen uiteraard ook de verschillende donoren die bereid waren monsters af te staan voor onderzoek. Daarnaast het EUROFORGEN netwerk, voor de financiële ondersteuning en de kansen die verbonden zijn aan een internationale samenwerking.

Dan kom ik nu bij de twee mensen die naast mij zullen staan tijdens de verdediging. Mijn broer Kees en collega Natalie. Fijn dat jullie mij als paranimf willen ondersteunen.

Last but not least: familie en vrienden. In het bijzonder uiteraard het thuisfront. Mijn ouders Krijn en Sjouk en de rest van het nest: Kees-Micha, Jouke-Sanne, Marjolein-Itay, Kristiaan-Sharon, Jasper-Paul en zeker ook Sam-Job-Max-Ben, Julius-Isabelle-Babette, Seth-Emma. Met zoveel steun, levendigheid en enthousiasme komt het altijd goed.

Curriculum vitae

Margreet van den Berge was born in Willemstad, the Netherlands, on April 19th, 1990. She attended high school at the CSG Willem van Oranje in Oud-Beijerland and graduated HAVO in 2008. The same year, she started the Bachelor degree programme Biology and Applied Medical Laboratory Technology at the Avans University of Applied Sciences in Breda. This Bachelor degree included two internships; one at Flinders University in Adelaide, Australia under the supervision of Prof. A. Linacre; and one at the Research and Development (R&D) group of the department of Human Biological Traces at the Netherlands Forensic Institute (NFI) under the supervision of Dr. P.A. Lindenbergh and Dr. L.M.T. Sijen. After her graduation in 2012, she was offered a research position in the R&D group at the NFI, which led to the start of a PhD under the supervision of Dr. L.M.T. Sijen and Prof.dr. P. de Knijff (Leiden University Medical Centre, LUMC). During her time at the NFI, Margreet developed, validated and challenged RNA profiling tools for the identification of body fluids and organ tissues, aiming to expand and advance forensic RNA profiling. After completing her PhD, she will continue to work in the research group at the NFI as a postdoc.