Development of RNA profiling tools and the implementation in forensic casework

Lindenbergh, P.A.

Citation for published version (APA):
Lindenbergh, P. A. (2014). Development of RNA profiling tools and the implementation in forensic casework

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 6

General discussion
Determining cellular origin within forensic casework

Within forensic science, different biological investigations are performed in which the results can be used to formulate propositions on multiple levels of significance [1]. Results of DNA investigations on evidentiary items found at the scene of a crime are generally used to address propositions on the source-level, pursuing to answer the question “who is the source of the found biological material?” [1]. DNA profiles are compared to either reference profiles or DNA databases in order to assess whether a match can be identified. Despite that DNA profiles can have a crucial role within a case, DNA evidence alone does not present enough support for a conviction in The Netherlands. Additional proof can be found in the form of testimonies from witnesses, digital recordings or other tactical information. Furthermore, other relevant information may be gathered from the biological traces. This information may assist in the formulation of propositions on activity level-related to the events which led to deposition of the biological material [1]. To assess these kinds of propositions, information on the cellular origin of the biological sample might be used.

Initial indications of the presence of biological material, which can already be used at the scene of the crime, can be obtained by using alternative light sources (ALS) [2-5]. ALS are capable of detecting latent body fluid stains through fluorescence detection at specific excitation spectra. These non-invasive methods have the advantage that DNA profiling from analysed stains is not compromised. In the forensic laboratory, researchers not only search for but also analyse biological traces and have disposal over several methods to determine cellular origin. Mostly used are the chemical, catalytic and serological methods but the most straightforward method is microscopy which is most suitable for the detection (and determination) of sperm cells. Microscopy is less suited for the identification of other cell types unless histologically stained. The chemical assays used for body fluid identification rely on a visible colour changes upon reaction between the stain and reagents or by the process of chemoluminescence. An example of detection by chemoluminescence involves the use of luminol for latent blood stain detection. Upon oxidation of luminol by iron present in haemoglobin a green-blue colour is emitted which can be observed in a dark environment [6]. The method is sensitive and widely used [6] but is not human-specific and the illumination time of the stains is short. Another chemical used to indicate the presence of blood is tetrabase (TB). Here, tetrabase (tetra-methyl-diaminodifenylmethane) is converted to a blue coloured product in presence of peroxide and haemoglobin present in blood [7]. Catalytic tests are also based on colour forming reactions which are catalysed
by enzymes present in body fluid stains. Examples of catalytic methods are the acid phosphatase test, the amylase assay and the Kastle-Meyer test which presumptively identify respectively semen, saliva and blood. When brought in contact with a substrate (alpha-naphthyl) on a paper carrier, acid phosphatase, an enzyme present in seminal fluids which is secreted by the prostate [8], catalyses a purple colouring reaction on the carrier [2]. The amylase method for saliva works in a similar fashion and uses the catalytic activity of alpha-amylase [9]. In the Kastle-Meyer test an alkaline solution turns pink after oxidation of phenolphthalein by peroxide when blood is present [10].

Next to the chemical and catalytic tests, immunological assays have also become available for forensic purposes. The traditional enzyme-linked immunosorbent assay (ELISA) has been used from the early eighties for the identification of seminal fluids [11] and nowadays many more different assays are available for other body fluids including blood, saliva, semen and different organ tissues [10]. ELISAs are based on labelled antibody reactions, which can form coloured complexes in presence of body fluid specific antigens [12-14]. The method is sensitive but involves many different incubation and wash steps, which makes the method lengthy. Faster immunological-based methods were designed for saliva and blood and involved test strips and cards containing antibodies but were found to be inferior compared to catalytic tests and were only used to supplement other presumptive tests [10]. More recently the Rapid Stain Identification (RSID™) saliva/blood/semen/urine and the prostate specific antigen (PSA) test [15-19] were developed which are sensitive and easy to perform and take considerable less time compared to an ELISA.

With exception of semen identifications by means of microscopy, the above-described methods are presumptive and present the forensic scientist with only an indication on the presence of the tested tissue types. In addition to the described tests, many more presumptive methods are available for blood, saliva and semen [10]. For other forensically relevant tissues like vaginal mucosa, menstrual secretion and skin cells, presumptive tests are not available or of less specificity such as the lugol staining for vaginal cells [20,21]. It has been established that by the use of mRNA profiling, blood, saliva, semen, skin, vaginal mucosa and menstrual secretion can be assessed to a degree of specificity which cannot be attained by using presumptive tests [22-34]. Each body fluid is comprised of a complex mixture of different cell types that each contain a different set of expressed genes. By selecting and analysing those genes that appear to be predominantly expressed in the target tissue, it becomes possible to differentiate between tissues using the mRNA profile. RNA profiling provides the forensic community a tool that offers an unbiased, highly sensitive and human specific cell type determination [35]. The next section discusses some aspects that have to be considered when designing mRNA profiling assays for cell type determinations.
Considerations when setting up mRNA profiling assays for forensic casework

Biological aspects

Many characteristics unique to mRNA expression can pose challenges with respect to usage and interpretation of mRNA profiles and should be considered before implementing RNA-based identification assays in casework. Specific eukaryotic cells express specific sets of genes encoded in the DNA. Some genes are expressed in all cell types as they are involved in metabolic processes like maintenance, energy management, repair and structure and are often used as reference or housekeeping markers. Other genes are differentially expressed depending on the specific function of a cell. Actually, for such genes the mRNA expression is not restricted but rather enriched in these specific cells as in many other cell types very low background level expression occurs. This can clearly be observed when viewing the BIOGPS database and zooming in on the baseline [36]. Another possible explanation for the fact that mRNA expression is often not entirely specific for one tissue could be the multi-level of control in which gene expression is regulated. Because differences in gene expression are the fundamental basis of mRNA profiling, an understanding of the cause of expression variation may attribute to the use of mRNA in forensic context.

The regulation of gene expression in eukaryotic organisms is complex because of the many factors on which influence is exerted. On the DNA level, regulatory sites near the transcription start site of genes serve as a binding location for transcription factors, which can activate or repress RNA polymerase and hereby influence general transcriptional levels [37]. Variations in gene expression may be caused by the changes in e.g. concentrations of transcription factors which can affect expression of genes nearby the DNA binding location (cis) or genes on a distant stretch (trans), the so-called cis- and trans-regulatory effects [38]. Also DNA methylation is known to affect gene regulation by playing an important role in e.g. genomic imprinting and X-chromosome inactivation [39].

As a result of these molecular processes, small variations in gene expression occur between single cells of the same tissue-type of one individual [40]. Also between individuals large variations in gene expression patterns can be observed in tissues from the same origin. In a study conducted by Rogue et al., gene expression profiles were generated from hepatocytes originating from six donors. They observed very high variations between transcript levels of six hepatocyte populations for a number of genes, which were related to amino acid-, lipid-, carbohydrate- and drug metabolism [41]. Shown was also a difference in the number of expressed genes which varied from
8,335 to 9,501. In a larger study, gene expression variation of 18,000 genes was examined in peripheral blood lymphocytes from 75 healthy individuals. For 370 genes, a two-fold or greater variation from the mean was detected [42]. Inter-individual variations in gene expression in general are common and are caused by the complexity of gene regulation. Examples of factors which can be involved in the variation in expression between individuals (or the same individual at a different moment in time) are the intake of medication, changes in physiological state (e.g. a vasectomy resulting in dropout of *PRM1* in semen extracts [35]), nutrition (which it is known to affect methylation status [43,44]), stress and environmental differences influence gene expression [45]. Also on the population level genetic variation has been observed [46].

Identification and use of markers

One of the fundamental steps in successful mRNA profiling relies on the selection of suitable mRNA markers which ideally should have the ability to indisputably discriminate between a large panel of body fluids and tissues. Clearly, non-target marker expression would present a major issue since this could end as a false-positive result in a sample. When a marker would show too much cross-reactivity, single source samples could be wrongly identified as mixtures. For the identification of suitable mRNA marker candidates, three major strategies can be followed. 1) Many markers incorporated in mRNA multiplexes were found through searching literature and expression databases [23,29,30,47]. For instance saliva markers *STATH* and *HTN3* were obtained by searching the Cancer Genome Anatomy Project (CGAP) for genes, which were expressed in the salivary and parotid glands [29]. Another example is the BIOGPS database [36], in which all organ specific markers described in Chapter 5 were found. Some of the markers, which have been selected in this manner have proven very suitable for the use in forensic mRNA analyses. Some criticism has been expressed for using databases as some of these might contain biased gene information due to a non-random character of representation of clone libraries [33]. 2) A different approach was suggested by Zubakov et al., who performed whole-genome gene expression analysis on time-wise degraded saliva and blood samples, considering almost the entire human transcriptome [33]. 3) A third method which has recently been suggested involves whole transcriptome sequencing (RNA-Seq) of cDNA and presents a very sensitive mean of markers identification [28,48]. Using the described methods of marker detection, multiple research groups have published numerous assays for the identification of body fluids using a large panel of different mRNA markers. The most markers have been identified for blood. From the large pool of blood markers, *HBA* and *HBB* are considered the most sensitive.
For saliva the most reliable markers are \textit{STATH} and \textit{HTN3} \cite{22,29,49}. Zubakov \textit{et al.} have considered mRNA markers of which the respective proteins are highly proline and keratin-rich (e.g. \textit{KRT4}, \textit{KRT13} and \textit{SPRR2A}) \cite{33}. These markers were found to be highly expressed in other mucosal membranes like vaginal mucosa and menstrual secretion as well \cite{22}. Recently new saliva markers were identified (\textit{MUC7}, \textit{PRB4} and \textit{SMR3B}) \cite{50}, but more research on these markers will be necessary. For semen identification, only a few markers are generally used and can be divided into markers, which are expressed in the sperm cell (\textit{PRM1/2}) and in seminal fluid (\textit{SEMG1/2}, \textit{TGM4} and \textit{PSA}) \cite{49}. Expression of both \textit{PRM1} and \textit{PRM2} is dependent on the level of sperm cells which are being produced, which is highly variable between male individuals (absent in azoospemic males). Both markers are considered to be highly specific. The expression levels of \textit{SEMG1} and \textit{TGM4} are comparable \cite{51}. For menstrual secretion, inference is mostly based on the expression of matrix metalloproteinases e.g. \textit{MMP7, MMP10} and \textit{MMP11}. It has been reported that signals for \textit{MMP7} can drop-out when females are using hormonal contraceptives \cite{52}. The negative effects of possible drop-out of \textit{MMP7} on mRNA profiling of menstrual secretions in these cases can be overcome by also assessing the expression level of a minimum of two other menstrual secretion markers. Other markers, which have been used for identification of menstrual secretion are \textit{MSX1, SFRP4} and \textit{LEFTY2} but are less robust than the matrix metalloproteinases \cite{50}.

Vaginal mucosa presents one of the most forensically relevant body fluids for cases involving sexual assault. Identification of vaginal mucosa can be complex due to the histological and biochemical similarities with other mucous membranes \cite{33,53}. For this body fluid, \textit{HBD1} and \textit{MUC4} have been used but are considered controversial by numerous research groups because expression has earlier been observed in saliva \cite{33,47,54}. Although in our experiments we have not observed specificity issues with the two markers, \textit{HBD1} was found to perform sub-optimally in multiplex resulting in laborious additional monoplex analysis (Chapter 1, \cite{22}). An alternative strategy to indicate vaginal mucosa encompasses the use of microbial flora. In a study performed by Fleming \textit{et al.}, the 16S–23S intergenic spacer region of \textit{L. gasseri} and \textit{L. crispatus} were used to identify vaginal mucosa \cite{55}. The two bacterial species were detected in vaginal mucosa and menstrual secretions, but not in blood, saliva and semen (expression in skin was not addressed). In an extensive study performed by Benschop \textit{et al.}, who explored the vaginal microbiome \cite{56}, similar results were obtained for blood, saliva and semen but Lactobacillus species including \textit{L. gasseri} and \textit{L. crispatus}, were also found on skin samplings from the hand, female groin and penis. As in many sexual assault cases samples of the penis are investigated for presence of vaginal
mucosa, use of *L. gasseri* and *L. crispatus* on these samples could result in false positive identifications. In order to establish a firmer basis for the use of these bacteria, more penile samplings should be investigated as well as the effect of antibiotic treatments on the microbial flora. Using RNA-Seq, promising new markers were identified for vaginal mucosa. Especially *MYOZ1* and *CYP2B7P1* demonstrated a high level of sensitivity and specificity [48]. For these candidates the biological function is unknown making it unlikely that these would have been identified through literature searches. Recent studies have also identified *ESR1* [57] and *MSLN* [58] as vaginal markers but both markers have not yet been used in validation studies.

Also using RNA-Seq, a marker was found for skin cells namely *LCE1C* [28]. Together with *CDSN* and *LOR* [32] these markers form the panel of markers for skin determinations in the assay described in this thesis. Hanson *et al.* did find expression of *CDSN* and *LOR* in vaginal samples which was also observed in our study (Chapter 1, [22]). As *CDSN* and *LOR* are predominantly expressed in keratinised epithelium, which is biochemically similar in vaginal, oral and skin epithelium, these results can be explained [32] but has to be considered when interpreting sexual assault cases. As *LCE1C* is more specific for skin cells it is recommended to also include this marker in a multiplex additional to either *CDSN* and/or *LOR*.

Overall, it appears favourable to have at least two markers per tissue to accommodate variation in expression between individuals or with physiological condition.

**Practical considerations**

**RNA isolation**

For combined mRNA and DNA profiling from the same forensic sample, an effective and robust co-isolation method is required which does not compromise the yield of either type of nucleic acids. For these purposes, numerous co-isolation protocols have been developed [22,26,31,59-61]. In the co-isolation procedure we developed, the standard method of DNA isolation (with exception of sample lysis) at the Netherlands Forensic Institute was combined with an RNA isolation kit capable of also isolating small RNA species which offers the opportunity to detect both mRNA and miRNA markers [22]. Although this study is focused on mRNA profiling, in the developmental stage of the project the possibility to analyse miRNAs was considered an alternative. In general, miRNAs are considered less prone to degradation
compared to the mRNA molecules which might be advantageous when analysing (partially) degraded samples. More on analysis of miRNAs will be discussed later on in this chapter. In the standard casework procedure, the DNA/RNA containing lysate is first transferred to a DNA column to separate the DNA and the RNA. Instead of directly continuing with both the DNA and RNA isolation, the RNA containing fraction is stored at -80 °C awaiting the results of DNA profiling. This freezing step does not affect the quality of the mRNA profiles (personal communication Bhoelai and Sijen). When no results are obtained from DNA profiling, the RNA fraction can be discarded, as putative mRNA profiling results are not informative in the case report. The choice to perform a co-isolation has to be made in advance, as mRNA profiling is not possible for a stored DNA extract obtained by a standard DNA isolation. An organic extraction is used to isolate the RNA from the stored fraction after which RNA is immobilised using a glass-fiber containing solid phase in presence of ethanol. The organic extraction of RNA is robust, yields pure RNA and is commonly used in other published mRNA profiling studies [59,60,62]. The co-isolation procedure described in this thesis has proven to be effective as also with low amounts of cellular material full RNA profiles could be generated. Additional improvements to the method could be made with respect to (1) automation and (2) combination with differential lysis. Automation of sample lysis is explored by many laboratories in order to reduce handling-errors and gain productivity. For reference material, automated robot lines are already used to isolate DNA [63-65] and with increasing experience and technological advances it is likely that also casework samples will follow soon. Unfortunately, the organic extraction step is not very amendable for automation because separation of the water and organic phase after phenol extraction might prove problematic. Instead, magnetic bead based co-isolations are more suitable for automated processes as has been shown by Bowden et al., which combined the Promega DNA IQ™ system with the Zymo Research Mini RNA Isolation Kit™ II and were able to isolate very low concentrations of RNA (<20 pg/µL) [61].

Differential lysis is used for samples from sexual-assault cases in order to separate sperm and non-sperm cells [66]. Especially these cases are of interest for mRNA profiling as the presence of for instance vaginal mucosa can hold incriminating information while no presumptive tests exist for this tissue. When a relatively small part of the cell material comprises of sperm cells, the ratio sperm to non-sperm can be unfavourable and signals from the high template component may mask the signals of low template components in both the DNA and RNA profiles. For this reason, it would be advantageous if the differential lysis protocol could be combined with mRNA profiling. At present, efforts are undertaken at the NFI to combine the two protocols.
RNA quantification

In forensic DNA analyses, the amount of human DNA present in a DNA extract is measured to achieve a balanced STR profile. For optimal RNA analysis, it may be beneficial to quantify the amount of human RNA as well. Although several methods for RNA quantification exist, such as RiboGreen, BioAnalyzer, Nanodrop, none of these methodologies is human-specific. There are several options to obtain some kind of input regulation in RNA analyses. First, one could determine the amount of total RNA as an indicator for the amount of human RNA and use this value to drive the input in the cDNA synthesis reaction. However, this approach may go amiss when lots of microbial RNA is present as can be the case in vaginal specimens. Second, one could take advantage of the fact that cDNA synthesis reactions can generally take micrograms of RNA input [22] and will not be easily overloaded when using a fixed volume of RNA extract. Then, either the cDNA is subjected to a human-specific quantification method to obtain an indicative value (one now assumes all RNA is transcribed equally efficiently) or an exploratory profiling approach is taken by use of various cDNA inputs in different RNA assays (serial input approach). Third, one could regard the human-specific DNA quantification result as an indicator of the amount of human RNA present. However, this relation may be very different for the different body fluids. Also, it is very likely that DNA and RNA are differently affected by degradation-inducing conditions. Overall, no ideal method is available yet. Some laboratories subject RNA extracts to DNA quantification. This approach has two goals: 1) verify that no DNA is detected in the RNA extract and 2) test for the presence of inhibitors by assessing the performance of the IPC (internal positive control) that is generally present in commercial human DNA quantification kits.

Methods of detection

The assays which have been designed for mRNA-based tissue inference are either based on multiplex endpoint PCR in combination with CE detection or use quantitative PCR (q-PCR). Endpoint PCR with CE detection is popular within forensic genetics because of the sensitivity and the ability to simultaneously detect multiple PCR products [22,25,26,30]. Alternatively, quantitative PCR (q-PCR) may be used which is a sensitive technique capable of detecting marginal changes in template concentrations [26,31-33]. In order to correctly interpret q-PCR data, normalisation strategies have to be applied using reference genes. Just as other genes, reference genes may exhibit differences in expression level within cell types and choice of reference is crucial [67,68,69].
Housekeeping genes commonly used include B2M, UCE, UBC, GAPDH, 18S-rRNA and ACTB. GAPDH is considered highly variable and most prone to drop-out, whereas 18S-rRNA and ACTB perform best in these assays [70]. Using modern q-PCR technology, it is possible include to a maximum number of five amplifications into a multiplex [71] of which two represent an internal reference and a reference marker. As it is desired to analyse a minimum number of two markers per tissue [22,35] q-PCR analysis would require a significant amount of sample and time. Efforts to improve this limitation are on-going [72].

Taken together, at present endpoint PCR may be better suited to forensic applications because 1) the methodology is the same as used with common DNA profiling and 2) much more transcripts can be measured in one single run. Designing an end-point multiplex presents the next challenge. Since RNA molecules are more sensitive to degradation than DNA molecules, small amplicon sizes are to be preferred (between 70 and 150/200 bp). Also, markers need to be well spaced. With CE, bleed-through signals can occur in other fluorescent dye channels when a marker is highly amplified. Thus, marker spacing needs to consider all dye channels. When over-amplification occurs, artefact signals occur like split peaks or trailing peaks some 10 nucleotides larger and preferably no other markers have a bin at these positions. Unfortunately, the amplicon signals do not always locate at the designed position but can shift some nucleotides up or down, and this will necessitate lengthening or shortening of primers. Besides the spacing of markers, attention needs to go to balancing of the profile regarding the signal strength for the different markers for the same cell type. As mRNA expression will vary for individuals and with physiological condition and because distinct mRNAs may be differently affected by degradation-inducing conditions, full balance is not realistic. A general balance can be obtained by adjusting the primer concentrations for each amplicon or selecting a stronger or less strong fluorescent dye. Much optimisation effort will precede a final RNA multiplex assay.

Beyond the RNA profile

Implementation of RNA profiling at a forensic laboratory

In spite of the possibilities and advantages compared to conventional approaches, mRNA profiling is only actively used in criminal casework by a small number of forensic laboratories. The techniques involved in mRNA profiling befit the expertise and infrastructure present in forensic laboratories, and information on the
performance of the RNA assays (such as repeatability, reproducibility, detection limit, sensitivity and robustness) is available from the published developmental validation studies. It seems thus likely that the reason for the reserved attitude towards mRNA profiling lies beyond the technical aspects. The concept of using sets of specific mRNA markers to detect different tissues is generally recognised, but presumably forensic scientists experience uncertainty about the validity of applying RNA assays to crime scene samples: Do markers for a cell type present robust signals? Can spurious signals for non-target tissues give false positive cell type identifications? Is RNA profiling suited for degraded samples? Is the presence of multiple cell types effectively recognised? These issues invoke that implementation of RNA profiling needs to be accompanied by interpretation guidelines.

**Interpretation of RNA profiles**

The performance of the markers in a multiplex can be heavily influenced by biological and environmental factors. Thus, signal imbalances are inevitable in RNA profiles, to the level of complete marker absence. As a solution, a minimum of two markers for one specific body fluid or tissue is included in most RNA assays. One hopes to have all markers for a cell type responding, but how should results be interpreted when only one marker presents a signal? Especially since cross-reactive signals may occur for one or even multiple markers, so how can true and false signals be discriminated? In Chapter 2 and [35], we propose an interpretation strategy that builds on two components: 1) replicate analysis and 2) a \( x=n/2 \) scoring system that places the RNA results in various categories (‘\( x \)’ is the number of observed and ‘\( n \)’ the number of theoretically possible peaks) for a cell type considering all replicates. A tissue is scored ‘observed’ if \( x \geq 2 \), ‘not observed’ if \( x=0 \) and ‘sporadically observed’ if \( 0 < x < 2 \). For co-expressed cell types, ‘and fits’ is added when (sporadically) observed. During case interpretation, tissues scored as ‘sporadically observed’ are effectively regarded as ‘not observed’ and tissues scored as ‘and fits’ as ‘not present as such’, which is effectively ‘not observed’. All markers have the same weight in this approach and ideally all markers are as effective regarding cell type identification. When one marker can have a lower efficiency with specific samples (like degraded samples, or in specific biological situations), it may be advisable to have three instead of two markers for this tissue. For instance, semen identification in case of an azoospermic male needs the presence of SEMG1 peaks in all replicates, as no signals for PRM1 will occur. Thus, it will be opportune to add a third semen marker to the multiplex that is specific to seminal fluid (not spermatozoa). Likewise, adding more effective vaginal mucosa markers will be helpful. Recently an alternative scoring
method has been published by Roeder et al. who assigned numerical values to each of the used mRNA markers based on correct and incorrect expression in samples of known origin [50]. When assessing samples, a body fluid score was calculated by adding the individual assigned numerical values of the markers with a signal higher than the set detection threshold. Body fluids were positively scored when the combined marker value was found higher than a pre-determined threshold value. The scoring thresholds were set stringently to minimise the chance of false positive results. For both this method and the one described in this thesis, it is possible to positively identify tissues for which incomplete profiles were generated. Our method provides a general method, applicable to any RNA multiplex, while the Roeder method needs re-evaluation with each alternation like changing a marker and maybe even using a new batch of primer stock. The Roeder methodology uses five markers for five body fluids residing in four different multiplexes. For these four multiplexes, a single amplification is used so basically, both methodologies usurp a similar amount of cDNA: either 4 replicates for one multiplex or 1 replicate for 4 different multiplexes. A true comparative study including compromised samples would be needed to assess which approach is more effective: a method having 2 to 3 markers per cell type that are analysed 4 times, or a method with 5 markers per cell type analysed once.

A next step to take during interpretation is the decision whether cell types are “present” or “not present”. Straightforward interpretation would be to regard ‘observed’ (our method) or above threshold (Roeder method) scorings as ‘present’ and ‘not observed’, ‘sporadically observed’, ‘and fits’ or below threshold scorings as ‘not present’. However, ‘not present’ may be a far too strong statement and ‘no indication for presence’ or ‘presence cannot be excluded’ may be more appropriate. The co-expression of cell types like blood with menstrual secretion makes it very challenging to recognise true mixtures of such two body fluids, whatever scoring procedure used. Context information, results of other evidentiary traces in the case and the corresponding DNA results of RNA-analysed sample may assist a forensic scientist. The study described in Chapter 3 clearly illustrates the risk of ‘association fallacy’ when one would associate the major cell type in an RNA profile to the major donor in a DNA profile. Thus, careful interpretation seems a wise route together with appoint specialized forensic scientists that build expertise for RNA analyses in a forensic context. And foremost, more studies on markers, cell type specificity, relation DNA and RNA results and interpretation approaches will be informative.
Beyond RNA profiling

Alternative marker types for cell type determinations

**microRNAs**

Over the past five years, alternative bio-molecular strategies have been developed, which can be used to assess the tissue type of forensic samples. Focus has been given to another type of RNA namely the microRNA (miRNA). MiRNAs, typically 19–24 nucleotides in length, regulate protein levels post-transcriptionally by hybridizing to complementary sequences at the 3’-UTR of specific mRNA molecules causing gene silencing by either degradation or translational repression [73,74]. Thereby, these non-coding small RNAs play an important role in a diversity of cellular processes like regeneration, replication and differentiation [73,74]. MiRNAs appear less susceptible for degradation by factors, which are known to degrade mRNA like high humidity and temperatures, moist and UV-light [2] as miRNAs form intimate complexes with Argonaute proteins in order to do their function [75]. This may increase protection against these environmental induced factors of degradation than mRNA. This property is beneficial as many forensic samples are compromised. In a number of studies, several potential tissue specific miRNAs were identified and subsequently assays were designed which could be used for the detection of several body fluids using as little as 50 pg of input material [76-80]. At present, strong miRNA candidates are available for blood and saliva but few for vaginal mucosa, menstrual secretions and semen. None of the investigated miRNAs were found to be absolutely tissue specific and showed also expression in non-target tissues. However, body fluids could be discriminated by using relative expression ratios of the miRNAs through q-PCR analysis [77]. The analysis of mixed samples, especially unbalanced or high order mixtures, may be quite complicated with q-PCR. It also requires the use of statistical models, which have to be validated before implementation in forensic casework. As the method currently relies on q-PCR, the multiplexing capability is limited. As the number of found miRNAs is continuously increasing, it is plausible that more specific candidates will be identified in the future. However, it is likely that the most abundant miRNAs have already been discovered. Many miRNAs have highly homologous family members of which some may differ only a single basepair from each other and may easily be co-amplified. However, TaqMan miRNA assays are available which can distinguish between these family members and through which analysis of the correct miRNA can be affirmed [81]. Also Next Generation Sequencing (NGS) might be used for analysis of miRNAs [82]. The use of NGS within the field of cell type determination will be discussed later on.
DNA methylation

Another novel approach for tissue identification uses the DNA extract and is based on differential methylation levels [83-85]. DNA methylations are epigenetic modifications which play a role in gene silencing (by affecting the structure of chromatin [83]) or gene activation in important processes like cellular differentiation and energy homeostasis [86,87]. DNA methylation occurs at CpG-dinucleotides at the C5 position of cytosine. In the human genome between 60 % and 90 % of all CpGs are methylated [83,88]. Through epigenome sequencing it became apparent that also DNA methylation patterns can be tissue specific [2,89]. Forensically relevant methylated segments are denoted tissue-specific differentially methylated regions (tDMRs) and have allowed the identification of blood, saliva, semen, vaginal mucosa and skin [84]. Analysis of tDMRs is preceded by a restriction of the DNA using a methylation sensitive restriction enzyme (MSRE), which cleaves a CpG containing recognition site only when the cytosine residue is unmethylated. Two loci are subsequently amplified using PCR (MSRE-PCR); methylated loci will amplify efficiently resulting in a strong signal whereas unmethylated loci will not amplify as the template has been cut. The ratio of the signals obtained by capillary electrophoresis differs between the two loci to such a degree that allows distinction between the body fluids [83]. For full profiles, the input of DNA in the restriction can be as little as 31 pg, which is much lower compared to other methods of methylation detection like e.g. pyrosequencing for which much higher amounts of DNA are needed [90]. Using differential DNA methylations, a kit (DSI-semen) has recently been developed and validated with the aim of fully replacing microscopic examinations of semen [91]. It uses differential methylations patterns at five genomic positions. Because the restriction sites of the enzyme used (HhalI) are not present within 65 basepairs of the repetitive sequences of the commonly used UK, European, and CODIS STR marker sets [91], it was possible to combine the DNA and the DSI-semen markers in one reaction. This would present benefits for sexual assault samples as analysis of the semen does not consume additional sample or requires additional chemicals. It has been reported that during development and aging, DNA methylations can be susceptible to change which could present challenge as results for forensic purposes should not be age dependent. Also environmental factors, nutrition and lifestyle can influence DNA methylations. A study by An et al. showed that for three tDMRs on loci used for semen identifications (USP49, PRMT2 and DACT1) methylation levels were not significantly different in semen from young and elderly men [85]. Together this adds support that methylation-based methods have the potential to become a valuable tissue identification tool although validation studies have yet to show if mixtures can be analysed.
Proteins

Over the past decades, the development of new technologies has not only revolutionized the field of genetics but also that of proteomics. Advances in specifically mass spectrometry (MS) instrumentation have allowed the very sensitive and accurate determination of a large range of molecules including proteins. With MS, the relative abundance and mass-to-charge ratios of molecular fragments from ionized sample/cell extracts can be determined and compared to databases containing the ratios of known peptides. Usually MS is preceded by a separation technique like Gas Chromatography (GC) or High Performance Liquid Chromatography (HPLC) in order to obtain more pure analysable digested protein fractions. The combined methodology of GC/HPLC with MS has proved a powerful diagnostic tool in various fields of research like endocrinology, clinical toxicology and pharmacokinetics. In forensic context MS is currently used in environmental analyses, fire and explosives investigations and drug detection [92]. Recently, MS applications for the detection of blood, saliva, semen, vaginal mucosa and menstrual secretion have also been developed which rely on the detection of functional proteins enriched in or specific to these cell types [93,94]. In a study performed by Yang et al., the most abundant proteins in blood, saliva and semen were identified. MS analysis of multiple of these proteins per body fluid resulted in MS profiles unique for each body fluid, which makes this assay confirmatory [95]. Compared to mRNA profiling, MS was able to detect smaller volumes (50 fL, 10 nL and 1 nL for respectively blood, saliva and semen) and has the advantage that results are automatically compared to a database, which can simplify profile interpretation.

Next to qualitative measurements, MS also provides quantitative data on the presence of proteins, which may assist in the determination of the relative contributions of cell types in the event of a mixture. MS has the potential to assist forensic scientists with highly sensitive information on presence of cell types. Before implementation of MS in casework, profiling interpretation models are to be developed. These models have been described for mRNA profiling and are essential for implementation. More work is also needed on the isolation method to allow uncompromised extraction of both good quality protein and DNA. Preferably this method is amendable to automation. A mass spectrometer might not be part of the standard instrumentation available on a forensic laboratory and requires specific technical expertise. Therefore MS is not as easily implementable as mRNA profiling. Future research will reveal more information on the divers proteomes of different cell types, which will benefit the MS application for forensic purposes.
Additional forensic RNA applications

Apart from cell type determinations, different research groups have used mRNA in forensic context previously in efforts to determine the age of biological stains and post-mortem intervals through quantifying its degradation state [96,97]. Initial work was performed on dried bloodstains by Anderson et al. who showed that age estimation was possible by comparing the relative ratios of two different RNAs: 18S r-RNA and ACTB [98,99]. Useful estimates could be made under experimental conditions to a degradation time of 150 days. As in mixtures multiple cell types are present in which the same housekeeping genes are expressed, the rRNA and mRNA ratio cannot be used to assess the age of the single components. Alternatively, reflectance spectrometry can be used to determine hemoglobin derivatives (oxyhemoglobin, methemoglobin and hemichrome), which are formed during degradation [100,101]. By using this hemoglobin reaction kinetics, the age of the bloodstains could be determined to a maximum of 60 days with an uncertainty for 14 days. Further work in this field could benefit the forensic process significantly.

In Chapter 5 we have shown that mRNA profiling can be used for organ typing as an alternative to histological assays. Another application of mRNA within forensic pathology is called ‘advanced molecular autopsy’, which investigates the genetic basis of death [102]. For example, transcripts of EPO, HIF1A and VEGF can be used to assess the cause of a myocardial injury [103]. The use of this application would require the establishment of post-mortem mRNA expression databases.

Future technologies for cell type determinations

Current forensic DNA and mRNA typing technologies predominantly rely on PCR of (c)DNA molecules followed by size separation, single nucleotide extension (for instance SNaPshot assay) or nucleotide sequencing (Sanger) for which capillary electrophoresis is employed. Although CE offers scientists a robust and sensitive analysis platform, advances in Next Generation Sequencing (NGS) technologies have been noticed by the forensic community. DNA analyses may improve from NGS because 1) more loci than the number currently used in STR amplification kits may be analysed, 2) more information from degraded DNA could be obtained and 3) a higher resolution of sequence reads will be obtained, which could aid mixture analysis of both genomic and mitochondrial DNA [104]. Also, NGS approaches using bisulphite converted DNA have been developed for the assessment of genome wide methylation levels [105,106]. This can forensically be relevant not only for tissue identification but
also for cases involving monozygotic twins as it has been shown that monozygotic twins can display differences in DNA methylation [107]. A comparison between genome-wide methylation levels from DNA found at a crime scene and DNA from twins might give leads towards involvement of one twin individual, whereas the current DNA technologies are not able to discriminate between twins. Deep sequencing approaches are not limited to genomic research but can also be extended towards transcriptomics, using whole transcript sequencing or RNA-Seq. The method can be used as a replacement tool for the traditional micro-arrays experiments to investigate gene expression profiling under various conditions. Recently the method has also been used in forensic context to identify candidates for vaginal mucosa and skin for use in mRNA profiling assays [108,109]. As RNA-Seq is able to quantitatively assess the amount of cDNA sequences present in a specific sample, the combined sequence data of thousands of cDNAs might possibly be used to infer cell type. The amount of RNA needed to performed a successful RNA-Seq run will likely decrease due to technological innovations. However, the success is also depending on the quality of the recovered RNA and more research is needed in order to assess whether degraded samples are able to return good quality sequence reads. In addition, analysis, validation and bioinformatics interpretation guidelines need to be developed before possible incorporation into forensic casework [110].

In Conclusion

The introduction of DNA profiling has revolutionised the forensic community and law enforcement by providing the opportunity to match individuals to biological material found at a crime scene. Through mRNA profiling, an extra level of information has become available, which can be used to ascertain the biological origin of forensic traces. RNA profiling can reliably be used for not only body fluids but also for tissues. The instability of the mRNA molecule and its susceptibility to degradation by ubiquitously present RNases is mentioned often as one of the major drawbacks of mRNA profiling for forensic purposes. This thesis as well as other published work, reports that even from (very) old stains, mRNA profiling is robust and capable of positively identifying minute and very old stains and that degradation issues appear to be a less critical factor. Also, notwithstanding the variable nature of mRNA expression within and between individuals and cell types, robust inference is very well possible through the use of interpretation models. Nevertheless, mRNA profiling has not been developed to fully replace presumptive testing. The method should only be performed
when confirmation about cellular origin is desired or when presence of cell types is questioned for which no presumptive assays are available. For each forensic case being investigated the relevance of mRNA profiling should thus be assessed. The increasing sensitivity of future molecular biological techniques like NGS or MS will make it possible to extract and analyse mRNA, DNA and protein from a minute number of cells. Although this can be advantageous for forensic samples, it also brings extra attention to the criminalistic value of results. Regardless of its detection method, the expression of mRNA will play a pivotal role in future cell type determinations. Continuing efforts into the identification of suitable markers, validation of developed multiplexes and collaborative RNA exercises (e.g. from The European DNA Profiling Group (EDNAP) [49,111,112]) are likely to further establish the role of mRNA in the determination of the biological origin of forensic traces.


56. Benschop CC, Quaak FC, Boon ME, Sijen T, Kuiper I. Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context? Int J Legal Med 2012; 126: 303-310.


