

Validation of new mRNA semen specific markers for body fluid identification in forensics

Daniela Serra Borges

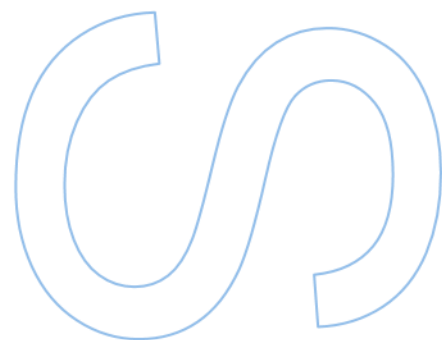
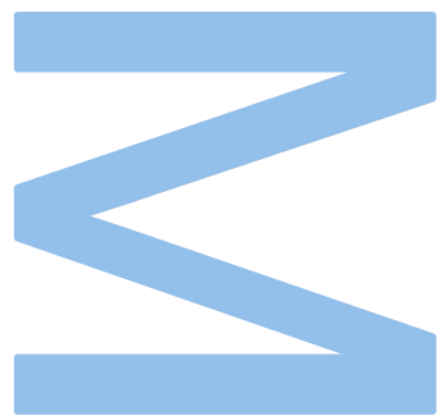
MSc in Forensic Genetics
Faculty of Sciences, University of Porto
2022/2023

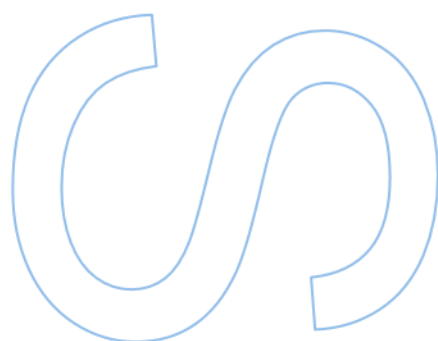
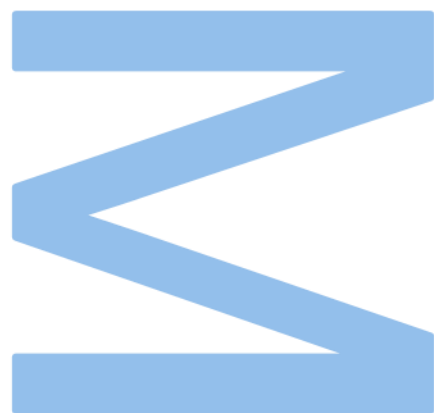
Supervisor

Iva Gomes, Associate Researcher, IPATIMUP/i3S

Co-supervisor

Laura Cainé, Teacher Assistant, FMUP





Acknowledgements

First, I would like to thank my supervisor, Dr. Iva Gomes, for accepting me on this project and for helping me during this time in the lab and with this dissertation. Thank you for your patience, your guidance, your understanding in all the difficulties during this challenging years and for not giving up on me.

To my co-supervisor, Dr. Laura Cainé, for always being available and for all the support that impacted this dissertation as well.

To the leader of Population Genetics & Evolution of i3S, Professor António Amorim, to all the group, including me in the group and for all the help in the lab, and to my colleague Bruno for helping me with doubts during this internship.

To all my family and friends for always believing in me and supporting me, no matter the decisions I make, especially my mom that always have the wisest words and find ways to motivate me, my friend Beatriz that always has my back helping me with technology, and Pedro for pressuring me, that was important too.

To all my part-time coworkers.

And, finally, to both my grandparents Celeste and Amadeu.

Thank you all for helping me during this dissertation, without all of you this would have been harder.

Resumo

O RNA mensageiro (mRNA) é o tipo de molécula de RNA mais utilizada para a análise forense na identificação de fluídos/ tecidos corporais a partir de amostras de manchas biológicas encontradas em cenas de crimes. Numa investigação da cena de um crime, esta identificação é crucial uma vez que pode oferecer detalhes sobre o crime, como por exemplo, a distinção entre sangue menstrual e sangue periférico que pode resultar de uma lesão traumática depois de agressão sexual. Por outro lado, o sêmen é um dos fluídos corporais mais relevantes no contexto da violência sexual pois numa situação de ejaculação, a investigação depende principalmente da deteção citológica e da visualização dos espermatozoides. No entanto, a ausência de espermatozoides não indica a ausência de sêmen e as técnicas existentes oferecem limitações.

Enquanto o perfil de DNA de um indivíduo fornece apenas informações sobre o dador de uma amostra biológica (confirmando ou excluindo o potencial doador da mancha), o RNA pode detetar genes expressos por tipos de células e garantir assim informações sobre os tipos celulares presentes e consequentemente identificar o tecido, ou seja, a origem da amostra. Como vantagem, um perfil de RNA pode ser analisado sem comprometer a análise do DNA devido à possibilidade da co-extração simultânea de RNA e DNA da mesma amostra.

O objetivo principal deste estudo foi desenvolver um sistema multiplex com (novos) marcadores de mRNA para a identificação de sêmen em contexto forense. O uso de um sistema de PCR em multiplex tem ajudado a aumentar a relevância da pesquisa de mRNA, uma vez que também estes tipos de marcadores, tal como marcadores de DNA, podem ser facilmente combinados em sistemas multiplex para uma única amplificação de PCR. Esta particularidade permite também a amplificação de vários marcadores para diferentes fluídos corporais ao mesmo tempo.

Deste trabalho resultou um sistema hexaplex para a amplificação, em uma só reação de PCR, dos seguintes marcadores PRM1, SEMG1, ODF1, SMCP, TcP11 e TNP1. A utilização destes marcadores em amostras com contexto forense também foi avaliada através de estudos de sensibilidade, especificidade e degradação. Os resultados obtidos, embora ainda preliminares, suportam a utilização destes marcadores, em particular, em casos de agressão sexual onde a identificação de sêmen é crucial como uma possível ferramenta molecular alternativa às técnicas tradicionais existentes.

É importante que no futuro mais marcadores sejam incluídos no multiplex desenvolvido de forma a completar esta ferramenta, não apenas marcadores específicos para o sémen, mas também outros marcadores relevantes para a identificação de fluídos e tecidos em contexto (agressão) sexual, como secreções vaginais, saliva e sangue (menstrual e periférico). O teste de especificidade também é importante ser melhorado, de modo a incluir um maior número de fluídos corporais e de um maior número de dados.

Palavras-chave: RNA, DNA, co-extração, marcadores de mRNA, identificação de fluídos corporais, multiplex PCR, sémen, agressão sexual, ciência forense, genética forense.

Abstract

Messenger RNA (mRNA) is the most used RNA molecule in the identification of body fluids/tissues from samples of biological stains found at crime scenes for forensic analyses. In crime scene investigation, this identification is crucial as it can provide details about the crime, such as the distinction between menstrual blood and peripheral blood that may result from a traumatic injury after sexual assault. On the other hand, semen is one of the most relevant body fluids in the context of sexual violence because in a situation of ejaculation, the investigation depends mainly on the cytological detection and visualization of sperm. However, the absence of sperm does not indicate the absence of semen and existing techniques have limitations.

While an individual's DNA profile only provides information on the donor of a biological sample (allowing a confirmation or exclusion of the donor of the stain), RNA can detect genes expressed by different cell types and consequently identify the tissue, that is, the origin of the sample. As an advantage, an RNA profile can be analysed without compromising DNA analysis due to the possibility of simultaneous co-extraction of RNA and DNA from the same sample.

The main objective of this study was to develop a multiplex system with (new) mRNA markers for semen identification in forensic context. The use of a multiplex PCR system has helped to increase the relevance of mRNA research, as these types of markers, such as DNA markers, can easily be combined in multiplex systems for a single PCR amplification. This particularity also allows the amplification of several markers for different body fluids at the same time saving sample consumption.

This work resulted in the development of a hexaplex system for the amplification in a single PCR reaction of the following markers PRM1, SEMG1, ODF1, SMCP, TcP11 and TNP1. The potential application of these markers in samples from forensic backgrounds was also evaluated through sensitivity, specificity and degradation studies. The preliminary results obtained from these analyses support the use of mRNA markers in cases of sexual assault, where semen identification is crucial as a possible alternative molecular tool to the traditional techniques, due to the high sensitivity, high specificity and detection in older samples tested.

As future developments, it is important that more markers are included in the multiplex developed in order to complete this tool, not only markers specific to semen, but also other markers relevant for the identification of fluids and tissues obtained from sexual assaults, such as vaginal secretions, saliva and blood (menstrual and peripheral)

and include a greater number of body fluids from different donors to improve specificity interpretation.

Key words: RNA, DNA, co-extraction, mRNA markers, body fluid identification, PCR multiplex, semen, sexual assaults, forensic science, forensic genetics.

Table of contents

Acknowledgements.....	i
Resumo	ii
Abstract	iv
Table of contents.....	vi
Figure index	viii
Table index.....	x
List of abbreviations	xii
1. Introduction.....	1
1.1. Ribonucleic acid (RNA).....	1
1.1.1. Types of RNA	3
1.1.2. Messenger RNA (mRNA).....	8
1.2. RNA in Forensic Genetics.....	10
1.2.1. RNA versus DNA	10
1.2.2. Why (m)RNA?	11
1.2.3. Other mRNA applications	12
1.2.4. mRNA profiling in body fluid and tissue identification	13
1.2.5. Advantages and disadvantages of mRNA profiling	18
1.3. Forensic Genetics in sexual violence cases.....	20
1.3.1. Semen in body fluid identification.....	22
1.3.2. “In-use” semen markers.....	22
1.3.3. “New” semen markers	25
1.3.4. PCR multiplex development for semen identification	32
2. Objectives.....	34
3. Materials and methods	35
3.1. Collection of biological samples and RNA extraction.....	35
3.2. Coding and storage of biological samples.....	35
3.3. mRNA semen markers and new primer designs	35

3.4.	Reverse transcription reaction: cDNA synthesis.....	38
3.5.	PCR of mRNA markers (single amplifications)	38
3.6.	Polyacrylamide gel electrophoresis (PAGE).....	39
3.7.	Multiplex development and optimization	40
3.7.1.	Hexaplex mRNA system	40
3.7.2.	Multiplex optimization process	41
3.8.	Capillary Electrophoresis (CE).....	42
3.9.	Sensibility testing of mRNA semen-specific hexaplex	42
3.10.	Specificity testing of mRNA semen-specific hexaplex.....	43
3.11.	Degradation assessment of mRNA semen-specific hexaplex.....	44
3.11.1.	Samples coding and storage	44
3.11.2.	Co-extraction of DNA and RNA.....	44
3.11.3.	Degradation assessment of hexaplex of mRNA semen-specific markers	46
4.	Results and discussion	47
4.1.	Multiplex design: mRNA marker selection and final hexaplex mRNA semen specific assay	47
4.2.	Multiplex development and optimization	49
4.2.1.	Amplification performance of mRNA markers in singleplex	49
4.2.2.	Amplification performance of hexaplex mRNA system	51
4.2.3.	Multiplex optimization procedures.....	55
4.3.	Applicability of the mRNA semen-specific hexaplex in forensics	58
4.3.1.	Sensibility testing of the mRNA semen-specific hexaplex	59
4.3.2.	Specificity testing of mRNA semen-specific hexaplex	61
4.3.3.	Degradation assessment of mRNA semen-specific hexaplex	63
5.	Conclusions	65
6.	Future perspectives	67
7.	References	68

Figure index

Figure 1 - RNA's single-stranded structure and its nucleotides composed by a five carbon sugar (ribose), a phosphate group and one of four bases attached to each ribose. Figure adapted from https://www.genome.gov/genetics-glossary/RNA-Ribonucleic-Acid	2
Figure 2 - Classification of noncoding RNAs (ncRNAs) and its types. Figure adapted from https://www.bio-rad.com/en-pt/applications-technologies/coding-non-coding-rna?ID=Q1070M70KWE7	5
Figure 3 - Gene expression chart of PRM1 (ENSG00000175646.3). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: https://gtexportal.org/home/gene/PRM1	24
Figure 4 - Gene expression chart of SEMG1 (ENSG00000124233.11). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: https://gtexportal.org/home/gene/SEMG1	25
Figure 5 - Gene expression chart of ODF1 (ENSG00000155087.3). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: https://gtexportal.org/home/gene/ODF1	26
Figure 6 - Gene expression chart of ODF2(ENSG00000136811.16). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: https://gtexportal.org/home/gene/ODF2	27
Figure 7 - Gene expression chart of PHF7 (ENSG00000010318.20). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: https://gtexportal.org/home/gene/PHF7	29
Figure 8 - Gene expression chart of SMCP (ENSG00000163206.5). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: https://gtexportal.org/home/gene/SMCP	30
Figure 9 - Gene expression chart of TcP11 (ENSG00000124678.17). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: https://gtexportal.org/home/gene/TCP11	31
Figure 10 - Gene expression chart of TNP1 (ENSG00000118245.2) . Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: https://gtexportal.org/home/gene/TNP1	32
Figure 11 - Representation of the initially designed in silico hexaplex system with the eight selected markers. Numbers (80-180) represent the bps and peaks are located in the approximate expected amplicon sizes of primers.	48

Figure 12 - Representation of the final hexaplex system designed in silico with the six selected markers after elimination of ODF2 e PHF7. Numbers (80-180) represent the bps and peaks are located in the approximate expected amplicon sizes as designed. 49

Figure 13 - Polyacrylamide gel of singleplex assays of SMCP, TcP11, ODF1, ODF2, TNP1 and PHF7 semen-specific markers for samples S1 to S7. a) SMCP, TcP11 and ODF1 b) ODF2, TNP1 and PHF7..... 50

Figure 14 - Polyacrylamide gel of singleplex assays of SMCP, TcP11, ODF1, ODF2, TNP1 and PHF7 semen-specific markers for samples S8 to S14. a) SMCP, TcP11 and ODF1 b) ODF2, TNP1 and PHF7..... 51

Figure 15 - Polyacrylamide gel hexaplex assay for the first attempt of multiplex development. The PCR included the selected semen-specific markers amplified for all samples (1 to 14). Label: (*) SMCP, (**) TcP11, (***) ODF1, (*) ODF2, (**) TNP1, (***) PHF7, (green) nonspecific bands..... 52

Figure 16 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 1..... 54

Figure 17 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 1 with Primer Mix 2. PRM1 has very weak and nearly “drops-out”. 55

Figure 18 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 1 with Primer Mix 3.. 56

Figure 19 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 3 with Primer Mix 4 (PCR 1 conditions)..... 56

Figure 20 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 3 with Primer Mix 4 (PCR 2 conditions)..... 57

Figure 21 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 2 with Primer Mix 4 (PCR 3 conditions)..... 57

Figure 22 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 1 with Primer Mix 1 (PCR 3 conditions)..... 58

Figure 23 - Specificity results of the VS, MB, and SAL samples. 63

Table index

Table 1 - List of RNA markers used during research for different purposes. Adapted from: Jawad, M., Noreen, S., Ashraf, MF., Asif, M., Asif, HM., Ali, S., Rakha, A. Application of ribonucleic acid (RNA) in Forensic sciences. Biomedica. 2020; 36 (3). 232-47.	15
Table 2 - Genes, tissue expression and genomic location (BUILD GRCh38/hg38) of the selected genes (data retrieved from www.genecards.org).....	36
Table 3 - mRNA markers with the respective primers sequences, melting temperature, 5' Fluorochrome, expected amplicon sizes and References: a) (Juusola & Ballantyne, 2005), b) (Haas et al., 2009).....	37
Table 4 - a) Quantity of H ₂ O added to markers without the label 5'; b) quantity of H ₂ O added to markers with the label 5'	37
Table 5 - Reagents and respective volumes used for one sample for the cDNA reaction.....	38
Table 6 - Reagents and respective volumes per sample for single amplification assays.	39
Table 7 - Thermocycling conditions each PCR reaction.....	39
Table 8 - Reagents and respective volume per sample for developed hexaplex...	41
Table 9 - Primer concentrations (µM) in the respective Primer Mix (1-4) during multiplex optimization. In bold are highlighted the major differences in primer concentrations among primer mixes optimization.....	41
Table 10 - Concentration of cDNA for respective dilutions prepared for sensibility testing.....	43
Table 11 - Coding samples and corresponding time of degradation among samples	44
Table 12 - Results comparison between PAGE and CE techniques in hexaplex assay. Label: (*) weak SEMG1 amplification, (**) weak TcP11 and ODF1 amplification, (***) weak TcP11 amplification, (****) weak ODF1 amplification.....	53
Table 13 - Expected and observed product sizes for respective markers for semen identification (sample 1).	54
Table 14 - Quantification of sample 6 with respective nucleic acid concentrations and absorbances.....	59
Table 15 - Representation of the height (RFU) of each amplified marker in the electropherogram results from the sensibility testing obtained in the software GeneMapper ID-X 1.4 Software 3500 CE system.	60
Table 16 - Quantification results of the different body fluids with respective nucleic acid concentrations and absorbances. Grey columns highlight the tested samples. ...	62

Table 17 - Representation of the height of each amplified marker in the electropherogram results from the degradation assay with samples from Instituto de Medicina Legal do Porto.	64
--	----

List of abbreviations

2D	Secondary
3D	Tertiary
A	Adenine
APAV	Associação Portuguesa de Apoio à Vítima
B	Blood
bp	Base pair
C	Cytosine
cDNA	Complementary DNA
CE	Capillary Electrophoresis
circRNA	Circular RNA
cRNA	Coding RNA
DNA	Deoxyribonucleic acid
EDNAP	European DNA Profiling Group
eRNA	Enhancer derived RNA
FME	Forensic Medical Examinations
FPP	Fertilization-Promoting Peptide
G	Guanine
KLK2 and 3	Kallikrein Related Peptidase 2 and 3
lincRNA	Long intergenic non-coding RNA
lncRNA	Long ncRNA
MB	Menstrual blood
miRNA	Micro RNA
ml	Millilitre

mRNA	Messenger RNA
MSMB	Microseminoprotein Beta
NAT	Natural Antisense Transcript
ncRNA	Non-coding RNA
ng	Nanogram
NKX3-1	Prostatic Tumour Suppressor Gene
nt	Nucleotide
ODF1 and 2	Outer Dense Fiber 1 and 2
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PHD	Plant Homeodomain
PHF7	PHD Finger Protein 7
PMI	Post-mortem interval
poRNA	Piwi-interacting RNA
pre-mRNA	Primary RNA
PRM1 and 2	Protamine 1 and 2
PSA	Prostate-Specific Antigen
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RT-PCR	Reverse Transcriptase- Polymerase Chain Reaction
S	Semen
SAL	Saliva
scaRNA	Cajal body specific RNA
SEMG1 and 2	Semenogelin 1 and 2
siRNA	Small interfering RNA

SMCP	Sperm Mitochondria Associated Cysteine Rich Protein
sncRNA	Short ncRNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein
TcP11	T-complex 11
TGM4	Transglutaminase 4
TNP1	Transition Protein 1
TNP1	Transition Protein 1
tRNA	Transfer RNA
TsD	Time since deposition
U	Uracil
VS	Vaginal secretions
WHO	World Health Organization
μl	Microliter
μM	Micromolar

1. Introduction

1.1. Ribonucleic acid (RNA)

According to the literature, it is justified that ribonucleic acid (RNA) has a dedicated field of biological research and specific definitions due to its roles in cellular processes and molecular mechanisms since it can provide catalytic, genetic, structural and regulatory functions in living cells (Copley et al., 2007; Sponer et al., 2018; Wu et al., 2014).

RNA is one of two classes of nucleic acid and it is present in all biological cells. The primary role of RNA is cellular protein synthesis, i.e., RNA converts genetic information from deoxyribonucleic acid (DNA) to proteins by coding amino acids (Gray et al., 2020). Even though DNA is identified as the molecular basis for genetic inheritance, RNA has a central role in information transfer from DNA to proteins (S. Li & Mason, 2014). RNA is not a simple copy of the genomic template DNA, since this molecule can sometimes be reordered or spliced in order to create alternative isoforms from the original transcript (S. Li & Mason, 2014). The nucleotide sequence of RNA is encoded in the DNA and can be modified after being transcribed (Gray et al., 2020; (S. Li & Mason, 2014). RNA is typically a linear polymer (single-stranded structure) and its basic structure includes ribose connected by phosphodiester bonds and nucleotides such as adenine (A), guanine (G), cytosine (C) and uracil (U) (Figure 1), similar to DNA (Chatterjee et al., 2022; Gray et al., 2020). Two characteristics that distinguish RNA from DNA and that have an important structural consequence are: 1) the substitution of -OH at 2' carbon of the ribose ring in RNA instead of the -H in the deoxyribose ring of DNA, and 2) the replacement of U in RNA for thymidine (T) in DNA (Sweeney et al., 2015). These differences make RNA more versatile than DNA regarding interactions that support complex structures (Sweeney et al., 2015).

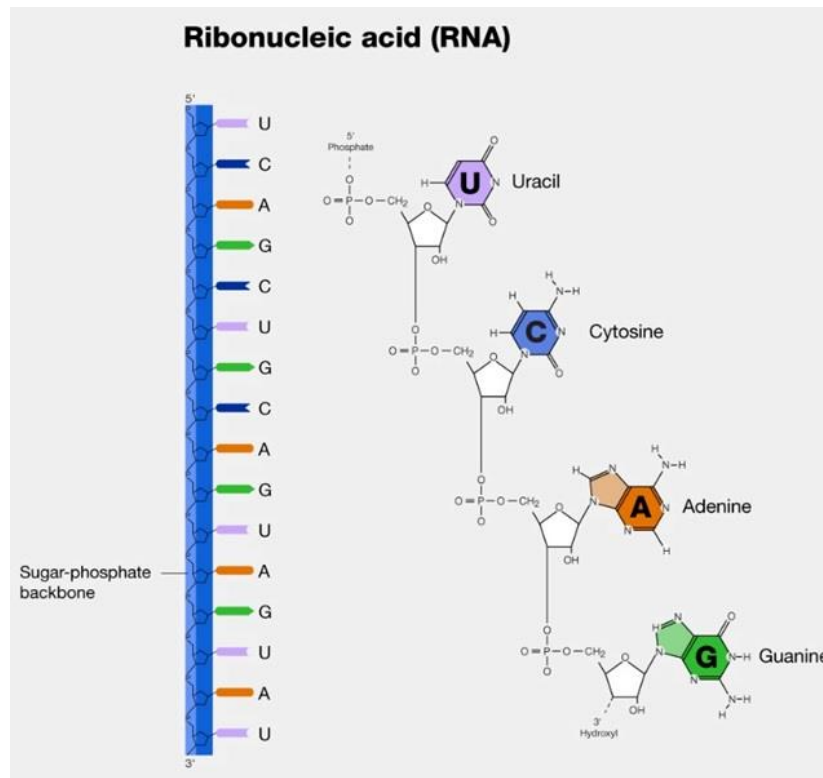


Figure 1 - RNA's single-stranded structure and its nucleotides composed by a five carbon sugar (ribose), a phosphate group and one of four bases attached to each ribose. Figure adapted from <https://www.genome.gov/genetics-glossary/RNA-Ribonucleic-Acid>

The existence of sequences that are self-complementary in the RNA strand leads to an interchain base-pairing and folding of the ribonucleotide chain into large domains such as duplexes, tetraloops, junctions, bulges, and pseudoknots, and it can be rigid or flexible (Runnels et al., 2018). Therefore, it is possible to define RNA molecules in a hierarchy manner by analysing its structures: primary, secondary (2D) and tertiary (3D) (Sponer et al., 2018). Primary structure is defined by an RNA sequence relatively to its constituent nucleotides, and it is possible that they suffer post-transcriptional chemical modifications (Sponer et al., 2018). 2D structure is defined by involving canonical base-pairs (A=U and C=G) and its interactions are stable and usually stronger than interactions in 3D structure (Mathews et al., 1999; Sponer et al., 2018). However, RNA molecules are generally single-stranded, and most mixed RNA sequences create series of antiparallel canonical double helices since they can fold back upon themselves (Sponer et al., 2018). Finally, tertiary structure's RNA motifs (3D), folds and architectures are stabilized by a range of interactions between individual nucleotides (Sweeney et al., 2015). It is known that base-pairing is the most specific of interactions since it allows RNA base pairing to be

generalized as a network of sequence-specific internucleotide interactions (Sponer et al., 2018); whereas base-stacking gives much of the stabilization energy for RNA folding (Sweeney et al., 2015). According to (Uhlenbeck, 1995) and (Firdaus-Raih et al., 2014) most RNA molecules fold into a well-defined three-dimensional structures being critical for its stability and biological function (to bind proteins, participate in catalysis or to serve as a template). Modifications in ribose and nitrogenous bases allows the formation of chemical bonds between distant regions in the RNA strand, leading to contortions in the RNA chain that will stabilize the RNA structure (Chatterjee et al., 2022; (Bhansali, 2021).

1.1.1. Types of RNA

There are various types of RNA that perform in protein synthesis; and the three most well-known are messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA). While both tRNA and rRNA are involved in part of the protein-synthesizing machinery, mRNA intermediates information in protein synthesis (Hornos & Hornos, 1993). mRNA is defined as a type of single-stranded RNA that originated from a DNA template during transcription. It plays a crucial role in the regulation of protein synthesis since it carries the information from the DNA in the nucleus to the ribosome that will then be read and each three-based codon will be translated into its corresponding amino acid in a growing protein chain (Sen, 2023). The genetic information embedded in mRNAs into polypeptides are then decoded by ribosomes (Wingo et al., 2019). So, the ribosome is required for the translation of all cellular proteins and is composed of two subunits (Sloan et al., 2016). The small 40S subunit of the eukaryotic ribosome includes one rRNA chain (18S) and 33 proteins whereas the large 60S subunit involves three rRNA molecules (25S, 5.8S and 5S) and 46 proteins and both subunits exist separately in the cytoplasm but the two join on the mRNA molecule (Ben-Shem et al., 2011; Khatter et al., 2015). rRNA is directly connected to the catalytic process of the protein synthesis, which means that these molecules play an important role regarding the attachment of each new amino acid to the growing chain, acting as a ribozyme (Pierce, 2012). The ribosomal subunits also contain binding sites specialized for RNA molecules, such as tRNA (Pierce, 2012). tRNA is a small non-coding RNA that delivers amino acids to the ribosome allowing the genetic code to be correctly interpreted for protein synthesis (Guy et al., 2014; Torres, 2019). Once tRNA is charged with the associated amino acid, they can read the nucleotide triplets on mRNA (codons) via base-pairing with the nucleotide triplets on tRNA (anticodon) meaning that it's an adaptor molecule (Torres, 2019). So, it

is important that the transmission of information from a nucleotide to an amino acid sequence is ensured taking into the account the diversity of tRNA isoacceptors or isotypes (tRNA with different anticodon that charge the same amino acid), isodecoders (molecules with the same anticodon but different structure) and its modified states (Ehrlich et al., 2021). In short, within the ribosome, mRNA and tRNA complexes are held together facilitating base-pairing while rRNA catalyzes the attachment of each new amino acids to the growing chain (Clancy & Brown, 2008).

In addition, RNAs can also be classified into two different categories: coding (cRNA) and non-coding RNAs (ncRNA). The primary transcript of RNA (pre-messenger RNA – pre-mRNA) includes both exons and introns and before it moves out of the nucleus to the cytoplasm, RNA goes through a splicing process that removes the introns, splicing the exons, forming the mature mRNA (Molecular Biology Review, n.d.). Consequently, the coding region is the portion of mRNA that is translated into protein (Kapranov et al., 2019; Molecular Biology Review, n.d.). Also, according to (Palazzo et al., 2015), the genomes of large multicellular eukaryotes are mostly comprised of non-protein coding DNA, meaning that most transcribed RNA is not translated into proteins, unlike mRNA. In fact, the ncRNAs act through complementary base pairing with target RNAs instead of being translated into proteins (Weick & Miska, 2014). ncRNAs are composed of housekeeping ncRNAs and regulatory ncRNAs; regulatory ncRNAs are further classified according to their size: long ncRNAs (being longer than 200 nucleotides) and short ncRNAs (smaller than 200 nucleotides) (Figure 2) (Coding and Non-Coding RNA, 2023; Geisler & Coller, 2013).

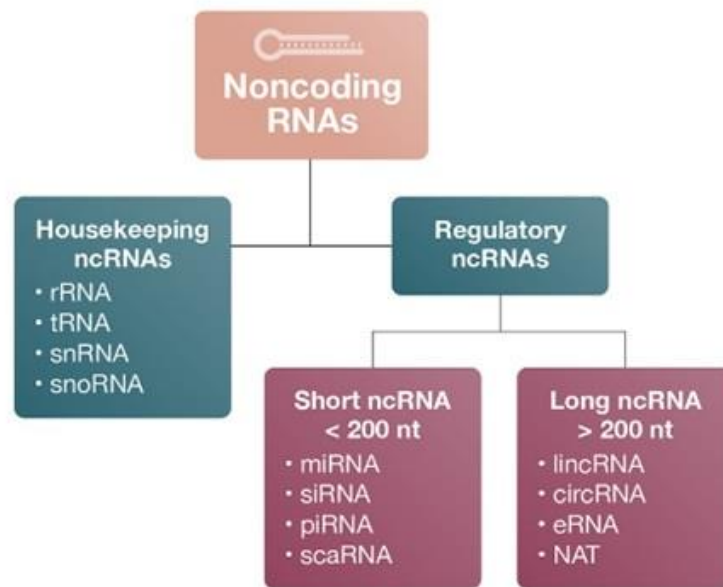


Figure 2 - Classification of noncoding RNAs (ncRNAs) and its types. Figure adapted from <https://www.biorad.com/en-pt/applications-technologies/coding-non-coding-rna?ID=Q1070M70KWE7>

Regarding the housekeeping ncRNAs (Figure 2), small nuclear RNA (snRNA) is a critical component of the spliceosome once it catalyzes the splicing of pre-mRNA (J. Meng et al., 2021). snRNA is an abundant molecule present in the nucleus that participates in the pre-mRNA splicing reaction and RNA processing (Coding and Non-Coding RNA, 2023; Karijolic & Yu, 2010). snRNA works as a small nuclear ribonucleoprotein (snRNP) complex that involves five snRNA (U1, U2, U4, U5 and U6) and several protein components, forming the spliceosome (Karijolic & Yu, 2010).

On the other hand, small nucleolar RNAs (snoRNAs) appear typically in the nucleolus where they play a role for both modification and processing of rRNA (Allmang et al., 1999; Holley & Topkara, 2011). rRNA modifications are directed by two large families of snoRNA, the box C/D snoRNAs and the box H/ACA snoRNAs, that differ relatively to their sequence and structural elements, binding partners and nature of the chemical modification catalysed (Bachellerie et al., 2002; Scott & Ono, 2011). According to Bachellerie et al., 2002, the two families of snoRNAs specify the sites to be modified in rRNA through the formation of a specific duplex in the rRNA modification site while the catalytic function is provided by a common protein enzyme (methylase or pseudouridine synthase) associated with the snoRNA. To allow the proper function of the spliceosome, modifications of snRNA are required, conducted by snoRNAs granting complementary

with rRNA that determines which nucleosides to be modified (Brosius & Raabe, 2016; Holley & Topkara, 2011).

Short ncRNAs, also known as sncRNAs, (< 200 nt) can be grouped into three main categories based on their length, mode of biogenesis and effector proteins: micro RNAs (miRNAs), small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs) (Patil et al., 2014) (Figure 2)

miRNAs are the most comprehensively studied types of RNA molecules and the interest of miRNA-mediated regulation of gene expression has been growing, with 21-23 nt in length and over 38.589 sequences deposited in the miRBase database (MiRBase, n.d.; Schöniger & Arenz, 2013). miRNA is known to play an important role to regulate the post-transcription of most of the genes and consequently protein production (Garcia et al., 2021; Vos et al., 2019). A common characteristic that defines all miRNA molecules is the fact they undergo a mechanism from a long double-stranded primary transcript (pre-miRNA) to a mature short double-stranded miRNA with ~22 nt (Vos et al., 2019). miRNA and siRNA use similar machinery relatively to their biogenesis, to the assembly into RNA-protein complexes and regarding the ability to regulate gene transcripts negatively in diverse eukaryotes (Tang, 2005). While miRNA is processed from a long double-stranded RNA, siRNA is derived from a shorter RNA hairpin-loop structure (Coding and Non-Coding RNA, 2023). siRNAs are usually exogenous; they are established into the cell by invading viruses and the viral double-stranded RNA is cleaved into short fragments of ~21-23 nt (Jackson & Linsley, 2010; Ramaswamy & Slack, 2002). Unlike miRNA, siRNA share a full sequence complementary to their targets, triggering an enzymatic cleavage of the matched mRNA at the opposite nt position of the molecule guiding strand (Jackson & Linsley, 2010). piRNAs play an important role leading associated chromatin-silencing machinery to transposon-encoding DNA sequences (that can replicate and insert into the genome potentially detrimental DNA damage) in order to maintain genome integrity in animal germ lines (Chalbatani et al., 2019; Nishimasu et al., 2012). These molecules involve endonucleolytic cleavage of the target sequence after complementary base-pair recognition preventing the accumulation of changes in the genome of the next generation (Weick & Miska, 2014). There are more than 30.000 piRNAs that have been found in humans, with usually ~26-31 nt and it is known that a subset of piRNAs may be capable to regulate protein-coding genes via DNA methylation (Chalbatani et al., 2019). Finally, Cajal body specific RNAs (scaRNAs), are required for catalysing post-transcriptional

modifications, acting as a “guide RNA” that direct modification of the snRNA (Enwerem et al., 2014; Porro et al., 2015).

Long ncRNAs (lncRNAs) are another class of noncoding RNAs (Figure 2) that have also been determined to be involved in transcriptional regulation and translation of target genes by modulating the functions of other ncRNAs such as miRNA, snoRNA, etc. (Dahariya et al., 2019; Shibata et al., 2015). LncRNA, with more than 200 nt, cannot be translated into proteins as well (Dahariya et al., 2019). The classification of lncRNA that has been proposed is based on their length, transcript properties, location (particularly to genomic annotations), regulatory elements and function (Jarroux et al., 2017). Long intergenic non-coding RNAs (lincRNAs) are usually defined as autonomously transcribed non-coding RNAs that do not overlap annotated coding genes (Ransohoff et al., 2017). LincRNAs share characteristics with other transcripts from the lncRNA family and they represent more than half of lncRNA transcripts in humans, being the largest class of lncRNA which are not located near protein-coding genes (Ransohoff et al., 2017). The functions of these molecules are mainly involved in biological processes such as control of gene expression, scaffold formation, epigenetic control and they also share some characteristics with sncRNAs regarding the stabilization of specific RNA-RNA and RNA-DNA interactions (Deniz & Erman, 2017).

On the other hand, circular RNAs (circRNAs) (Figure 2) are a class of endogenous non-coding RNA that are covalently closed loop structures without a 5' cap or a 3' Poly A tail (S. Meng et al., 2017). The advance in computational analysis and RNA sequencing techniques made it possible to recognize circRNA as a single stranded RNA, differing from the known linear RNA, since it is continuously closed in on itself (Lu, 2020). Even though the mechanisms of circRNA generation and functions are still not fully clear, recent studies have shown that these molecules act as post-transcriptional regulators and that also have functions in cell proliferation, motility, and metastasis, as well as, in cell cycle progression, angiogenesis, and apoptosis (Verduci et al., 2021). It is also described that the best-described mechanism that explains the biological function of these molecules is the ability to function as miRNA sponges by regulating protein function or by being translated themselves (Kristensen et al., 2019; S. Meng et al., 2017; Verduci et al., 2021).

Relatively to enhancer derived RNAs (eRNAs) it is known that they share characteristics with other lncRNAs, especially in the activation of mRNA transcription, challenging the initial thoughts that enhancers are sites for transcription factor assembly

(Kim et al., 2015). The interactions between enhancers and promoters can be bidirectional and promoters are needed to activate enhancer transcription (Kim et al., 2015). eRNAs are originated from enhancers regions and these molecules' main function is to facilitate enhancer-promoter interaction or activate promoter-driven transcription (Kim et al., 2015).

Finally, natural antisense transcripts (NATs) (Figure 2) are RNA molecules that are transcribed from the opposite DNA strand to other transcripts overlapping, in part, with sense RNA (both sense and antisense RNA can encode proteins or be non-protein-coding transcripts) (Faghihi & Wahlestedt, 2009). NATs interfere with translation and have equally inhibitory (blocking hairpin structures in mRNA that are essential for ribosome binding) and stimulatory effects (neutralizing inhibitory structures and enabling efficient translation initiation or elongation) (Wight & Werner, 2013). So, these molecules have regulatory functions regarding the transcriptional level and regulate gene expression levels post-transcriptionally (Rosikiewicz & Makałowska, 2016).

Despite the diversity of RNA types, according to the scientific forensic community, miRNA and mRNA are the commonly used RNA molecules in forensic science analysis, specially to identify the cellular origin such as body fluids or organ tissues from biological stains obtained from crime scenes or evidence items (Bai et al., 2013; Glynn, 2020; Haas, Muheim, et al., 2009; van den Berge & Sijen, 2017).

1.1.2. Messenger RNA (mRNA)

As previously mentioned, mRNA is described as a single-stranded RNA sequence originated from a DNA strand template and has an important role in protein synthesis regulation since it will carry DNA information to a ribosome to assemble amino acids in their correct order. While these molecules (mRNA) are the least abundant type of the total of RNA, they are also the most heterogeneous, since an mRNA is used momentarily for the synthesis of a large variety of polypeptides (Pelley, 2012).

When DNA was first identified as the source of genetic information, it was also discovered that it could not directly encode proteins since this molecule occupies the nucleus and protein synthesis occurs in the cytoplasm (Pierce, 2012). So, there is an additional molecule that carries genetic information for protein structure to the ribosome

and messenger RNA name was based on that carrier (Pierce, 2012). An important step in gene expression is the mediated transfer of mRNA of genetic information from the nucleus to cytoplasm so it can then be translated into proteins (Fasken & Corbett, 2005). Variation of gene expression in mRNA results from a combination of interactions between transcription factors and the state of chromatin or promoter/silencing DNA regions (Sijen, 2015). Initially, transcription of genes occurs in the DNA sequences that serves as a template and an enzyme called RNA polymerase II catalyses the formation of a primary mRNA (pre-mRNA) molecule also known as primary transcript (Clancy & Brown, 2008). Pre-mRNA is composed of coding sequences (exons) and non-coding sequences (introns) and this immature molecule goes through a set of post transcriptional modifications to create the functional and mature mRNA (Green, 1986). These modifications occur in the nucleus involving 3 stages: the precise removal of introns (RNA splicing), the addition of a 5' cap (7-methylguanosine RNA) at the 5' and a 3' cleavage end with the addition of a poly(A) tail (50 to 250 or more adenine nucleotides) (Aregger & Cowling, 2016; Green, 1986; Pelley, 2012). After all these resulting modifications, only a small proportion of mRNA (~2%) is translated into a protein molecule (Sijen, 2015). Researchers also believe that not all transcribed RNA will be a functional RNA since it can result in noncoding intergenic RNA or a molecule with an imperfect termination, meaning that gene expression is controlled by regulatory systems to guarantee the quality of mRNA and defected RNA is then degraded (Sijen, 2015).

1.2. RNA in Forensic Genetics

1.2.1. RNA versus DNA

In a forensic scenario it is common to find biological stains derived from blood, semen, saliva, or mixed body fluids, so it is important to know the source of the sample and the body fluid or tissue that originated the stain. This collection of information is vital for casework conduction because it may offer crucial details about case circumstances (Albani & Fleming, 2018). For example, if the presence of circulatory blood in a sample stain is positive, it may result from a traumatic injury, whilst the presence of menstrual blood reveals sexual activity; or the presence of skin cells suggests social contact, while vaginal secretions support a sexual interaction (Albani & Fleming, 2018; E. Hanson et al., 2018).

Expression of specific mRNAs differ depending on the cell type, so with mRNA-based assay it is possible to identify the type of biological fluids/cells, whereas with DNA analysis a genetic profile of an individual is obtained from a biological sample recovered from the crime scene (Juusola & Ballantyne, 2003; Roeder & Haas, 2013). Consequently, there is a comparison between both DNA's (from the crime scene and from the individual in question), so the person is either included or excluded as a potential donor of the stain (Juusola & Ballantyne, 2003; Moreno et al., 2012). DNA profiling does not reveal the biological material from the crime scene under which it was transferred but provides information on the donor. Thus, it is not possible to conclusively link a DNA profile to a sample stain derived from a particular body fluid (E. Hanson et al., 2018; Sijen, 2015). DNA profiling context and relevance depends on the identification of the body fluid from which it has originated since a forensic stain could test positive for a certain body fluid and the acquired DNA profile from another body tissue (skin cells); or in case of being present more than one body fluid (mixture of fluids), the body fluid that DNA profile originated cannot be determined (Fleming & Harbison, 2010).

When the type and abundance of specific targeted mRNAs allows the detection of body fluids/ tissues sources of forensic samples, there is the possibility of simultaneously analysing RNA and DNA from the same sample so that no genetic material is lost (Bauer & Patzelt, 2003; E. Hanson et al., 2018). mRNA profiling systems include the ability to co-extract RNA and DNA and consequently, use mRNAs markers for body fluid detection without compromising the yield of either what is required (Bowden et al., 2011; Roeder & Haas, 2013). This process of co-extraction has the capacity of analysing multiple

markers and tissue types, at the same time, in a single polymerase chain reaction (PCR) assay and the potential for automation and increased specificity, preserving samples and only using a small portion of the extract that is discarded when analysing DNA alone (Juusola & Ballantyne, 2007; Roeder & Haas, 2013). So, mRNA is a valuable tool since it provides high sensitivity and compatibility with DNA analysis, and in turn DNA quantification from recovered biological material ensures the quality of mRNA profiling process (Albani & Fleming, 2018; Juusola & Ballantyne, 2003; Moreno et al., 2012).

1.2.2. Why (m)RNA?

Although, in forensic science, most investigators have focused on DNA research, the potential technologies for RNA analyses are increasing. RNA transcripts in body fluids can differ according to a specific tissue and can be detected by using molecular technologies facilitating the forensic analyses of stains (Fu & Allen, 2019). For example, body fluids such as menstrual blood, semen, saliva, urine and sweat can be identified in dried stains by molecular assays that target tissue specific transcripts accommodated in the fluid cells (Fu & Allen, 2019). So, the forensic genetic community has been using two main RNA molecules, mRNA and miRNA, because both are cell types specific granting high specificity for the identification of human body fluids (Alshehhi & Haddrill, 2019; E. K. Hanson et al., 2009; Lindenbergh et al., 2012). miRNA has also been proposed for the study of body fluids identification, since it they have short length (resulting in more resistance for degradation), high tissue specificity and binding to proteins making it a stable marker (Bamberg et al., 2022; O'Leary & Glynn, 2018). However, studies have shown inconsistent results, for example, while a study confirms miR-205 as a good biomarker for saliva identification, another concludes that it is possible to be epithelium specific but not saliva (Silva et al., 2015). Also, miRNA is not an ideal marker for semen identification due to its overlapping expression and co-expression and the absence of homogeneity and the non-reproducibility of results from different labs difficult the miRNA markers validation (Silva et al., 2015; Tian et al., 2018).

There are reports of unexpectedly high stability of RNA markers under certain conditions, with tissue-specific expression patterns, confirming its efficiency, so the instability should not be the reason to prevent implementation of RNA as an additional tool in forensic genetics (Bauer, 2007; Setzer et al., 2008; Zubakov et al., 2008). The European DNA Profiling Group (EDNAP) collaborative exercises on mRNA profiling for

body fluid identification support the use of an mRNA-based system for body fluid identification as an alternative tool to traditional serological tests or along conventional DNA profiling, so this technique could be used in forensic casework laboratories (Haas et al., 2012; Haas, Hanson, Bär, et al., 2011; Haas, Hanson, Morling, et al., 2011).

To overcome the limitations of conventional body fluid identification methods, the use of mRNA profiling has been proposed. mRNA profiling systems have become a relevant field for the scientific community over the past decade since it allows the identification of a variety of forensically relevant human cell types derived from body fluids/ tissues in small amounts of samples depending only on mRNA expression (Lindenbergh, Maaskant, et al., 2013; Richard et al., 2012). The analysis of body fluids and tissues is crucial to identify the nature and source of genetic material, so it's possible to reveal and reconstruct events of a crime scene, and consequently to contribute solving crimes (Blackman et al., 2018; Zubakov et al., 2008). Therefore, identification of cellular/ body fluid source of biological evidence can determine enough number of mRNAs that are expressed entirely in cells that comprise that cell type/ body fluid (Juusola & Ballantyne, 2003). So, according to the literature, with RNA profiling there is the possibility to detect rare genes expressed by specific cell types providing insights into which cells or fluids are present in a certain stain (Moreno et al., 2012).

1.2.3. Other mRNA applications

Beyond body fluid/ tissue identification, mRNA analysis has other main forensic genetic applications regarding diseases and mechanisms leading to death, determination of the cause of death due to, for example, post-mortem RNA stability (Bauer, 2007). Regarding post-mortem human tissue, mRNA might expose information about gene expression at the point of death or near to its occurrence (Koppelkamm et al., 2011). After death, RNA degradation is facilitated by ribonucleases already present in the cell or introduced through bacteria or other environmental contamination, as well as physical and chemical factors (Bauer, 2007; Finger et al., 1987; Yasojima et al., 2001). According to the literature, agonal factors like coma and hypoxia can affect the amount of some RNAs as well as its integrity in post-mortem brain studies, which lead to the question that analysing mRNA could serve as a diagnostic tool for evaluating the type and duration of coma and agony, plus the level of oxygen deprivation (Barton et al., 1993;

Bauer, 2007; Tomita et al., 2004). So, this would offer valuable insights into cause and progression of death.

mRNA markers applications can also include wound age determination, post-mortem interval (PMI), the age of stains and so on (Bauer, 2007; Sijen, 2015). Wound examination it is the most important procedure in forensic pathology for an exact diagnosis of the cause and manner of death, it is crucial to describe a detailed and objective result of the wounds as much as possible (Ohshima, 2000). Wound age determination using RNA markers can help answer questions like the survival time after the fatal injury (Ohshima, 2000). Additionally, PMI is defined as the time that passed since death and the examination of the corpse and studies have shown RNA deterioration and/or the loss of RNA transcripts is more prone than DNA and proteins concerning rapidity and temporal correlation (Poór et al., 2016; Scrivano et al., 2019). So, RNA could potentially be used for PMI estimation (Poór et al., 2016). On the other hand, establishing the age of a stain or time since deposition (TsD) can indicate the time when a crime happened and/ or whether the biological evidence was placed before or after the crime (Alshehhi & Hadrill, 2019). Once forensic stains are often exposed to a large variety of environmental conditions for long periods of time before being collected and transferred to the laboratory, it is important to address nucleic acid stability under these conditions (Bauer et al., 2003). For example, studies have demonstrated that in old dried stains (10 to 20 years old) it is possible to detect some levels of several RNA transcripts, providing evidences in favour of the application of RNA markers for forensic cases involving stains of unknown age (at least under certain environmental conditions) (Fu & Allen, 2019; Zubakov et al., 2009).

1.2.4. mRNA profiling in body fluid and tissue identification

mRNA profiling is currently the most established method to identify body fluids serving as a confirmatory test, proving to be a viable alternative to conventional methods (Blackman et al., 2018; Cortellini et al., 2020). This approach relies on the distinct expression patterns of mRNAs allowing the identification and characterization of specific biological fluids in stain samples (Cortellini et al., 2020). Several mRNA assays have already been developed to detect all relevant body fluids commonly found at crime scenes investigations (Blackman et al., 2018; Richard et al., 2012).

mRNA profiling is usually performed alongside a DNA/RNA co-isolation procedure allowing DNA profiling to be performed on the exact same sample with no sample loss, holding the fact that mRNA markers can be analysed in a multiplex reaction with significant forensic value (Kohlmeier & Schneider, 2012; Lindenberg, van den Berge, et al., 2013). So, this method involves co-extracting RNA from a fraction of the DNA extract (which is typically discarded when in standard extraction DNA procedures) and numerous RNA and RNA/DNA coextraction kits are already commercially available for forensic stain investigations (Roeder & Haas, 2016). Through the usage of mRNA profiling, studies have confirmed the successful identification of blood, menstrual blood, and vaginal secretions in stain samples in laboratories lacking prior experience with RNA and RNA/DNA co-extraction analysis (Richard et al., 2012).

Numerous studies using mRNA markers for recognition of the source of biological evidence in forensics have been proposed and have already identified promising genes that exhibit tissue-specific expression achieving this aim (Gomes et al., 2013; Haas et al., 2013; E. Hanson et al., 2018; Juusola & Ballantyne, 2003). In a scenario where, for example, a biological sample such as blood is recovered from the suspect's clothing, a dispute may arise between the suspect and the prosecution regarding its origin: suspect argues that the blood originated from the victim's nose, while the prosecution claims that it resulted from a sexual encounter (Jawad et al., 2020). In this case, it is crucial to determine the true source of the blood stain in order to validate the respective claim. A similar case can happen with any type of body fluid being highly relevant to identifying the origin of a biological sample evidence.

A list of RNA markers (miRNA and mRNA) has been adapted and withdrawn from a Review Article from Jawad et al., 2020 (Table 1) that contains many different types of tissues and body fluids and associated RNA based markers.

Table 1 - List of RNA markers used during research for different purposes. Adapted from: Jawad, M., Noreen, S., Ashraf, MF., Asif, M., Asif, HM., Ali, S., Rakha, A. Application of ribonucleic acid (RNA) in Forensic sciences. Biomedica. 2020; 36 (3). 232-47.

Body Fluid	RNA Markers
Circulatory Blood/ Blood Stains	Delta-aminolevulinate synthase (ALAS2)
	Haemoglobin alpha (HBA)
	Haemoglobin beta (HBB)
	Beta-spectrin (SPTB)
	Porphobilinogen deaminase (PBGD)
	CD3 γ molecule (CD3G)
	Ankyrin 1 (ANK1)
	Glycophorin A (GLY)
	CASP2, C1QR1, ALOX5AP, AQP9, C5R1, NCF2, MNDA, ARHGAP26
	AM1CA1
	GYPA
	Pro-platelet basic protein (PPBP)
	ALAS2, ANK1, HBB
	SPTB, PBGD
	HBA, ALAS2, GlycoA
	miR-190a-5p, miR-454-3p
	HBD, SLC4A1
	miR451
	hsa-miR-144-3p
	CASP1, AMICA1, C1QR1, ALOX5AP, AQP9, C5R1, NCF2, MNDA, ARHGAP26
Menstrual Blood	Matrix metalloproteinase 7 (MMP7)
	Matrix metalloproteinase 10 (MMP10)
	Matrix metalloproteinase 11 (MMP11)
	Glycophorin A (GlycoA)
	Keratin 19 (CK19), Progesterone Receptor (PR)
	NCF2, MNDA, CASP1, C1QR1, ARHGAP26, AMICA1, ALOX5AP, AQP9, C5R1
	LEFTY2, MMP10
	HBA, ALAS2, GlycoA, HBD1, MUC4, Lcris, Lgas, MMP7, MMP10, MP11
	miR-141-3p, miR-497-5p
	MMP3, STC1
Saliva/ Saliva Stains	miR16
	miR-412-3p
	Statherin (STATH)
	Histatin 3 (HTN3)
	PRB4
	SPRR3, SPRR1A, KRT6A
	KRT4, KRT13
	Follicular dendritic cell secreted protein (FDSCP)
	HTN3, MUC7, STATH
	STATH, HTN3
	miR-935, miR-3168
	SPRR3, SPRR1A, KRT4, KRT6A, and KRT13

	KRT4, KRT6A, KRT13, SPRR1A, SPRR2A
	STATH, HTN3, MUC7
Nasal Secretions	Statherin (STATH)
	Protamine 1 (PRM1)
	Protamine 2 (PRM2)
	KLK/prostate-specific antigen (PSA)
	Transglutaminase 4 (TGM4)
	Semenogelin 1 (SEMG1)
	SEMG1
Semen	MSP
	Microseminoprotein (MSMB)
	PRM2, SEMG1, TGM4
	PRM1, PRM2
	PRM2, SEMG1, TGM4
	miR-888-5p, miR-135a-5p
	TNP1, KLK2
	miR10b, miR374
	miR-10b-5p
	Human β -defensin 1 (HBD1)
	Mucin 4 (MUC4)
	16S rRNA
Vaginal Secretions	Human β -defensin 1 (HBD1)
	ESR1
	SFTA2, FUT6, DKK4, IL19, MYOZ1, CYP2B7P1
	Mesothelin (MSLN)
	CYP2A7, CYP2B7P1, DKK4, FUT6, IL19, MYOZ1, NOXO1
	MUC4
	HBD1, MUC4, Lcris, Lgas
	miR-3134, miR-31-3p
	miR1280, miR4286, miR124, miR203, miR205
	miR-124a-3p
	Corneodesmosin (CDSN)
	Loricrin (LOR)
	Cytokeratin 9 (KRT9)
Skin/ Touch-DNA	LCE1D, LCE2D, CCL27, IL1F7, LCE1C
	CST6
	CDSN, LOR, KRT9
	CDSN, CST6, DSC1
	CCL27, IL1F7, KRT9, LCE1C, LCE2D
	miR-4761-5p, miR-137, miR-4473, miR-585-3
	miR3169
	LCE1C, LCE1D, LCE2D, CCL27, IL1F7
Oral Mucosa	Keratin 4 (KRT4)
Urine	Uromodulin (UMOD)
Sweat	DCD

After the occurrence of a crime, biological samples play crucial roles during a forensic investigation and provides valuable evidence since it is often transference between the victim, the offender, and the crime scene (Alvarez et al., 2004; An et al., 2012). In a sexual assault scenario, the presence of certain body fluids offers important information about the sequence of events occurred, allowing the reconstruction of the crime circumstances, as well as probative value of the evidence to be enhanced in a court through the confirmation of the biological evidence (An et al., 2012; Fujimoto et al., 2017; Loureiro et al., 2019; Takamura et al., 2017). It is necessary to detect and identify the nature of biological stains, because for instance, blood stains might show a physical struggle, an assault or even murder, and on the contrary, a semen or vaginal stain suggest the involvement of a sexual encounter assault (An et al., 2012; Loureiro et al., 2019). Body fluids such as blood, semen, saliva, vaginal secretions, urine, and sweat are the most common found at a crime scene and have a valuable together with the contribution of DNA evidence (An et al., 2012; Virkler & Lednev, 2009). Each of these body fluids have screening tests, however, as said before, such tests have limitations when it comes to identifying all fluids that are significant in forensic analysis and show cross-reactivity issues, leading to no reliable testing (Albani & Fleming, 2018; Blackman et al., 2018).

More recently, studies have shown that mRNA analysis is the most established approach regarding body fluid identification, providing a confirmatory test determining the presence of a body fluid and several specific and sensitive mRNA assays have already been developed (Table 1) (Blackman et al., 2018). Each tissue or body fluid include individual cell types with a distinct gene expression pattern (transcriptome), which is defined by the presence and abundance of specific mRNAs (Setzer et al., 2008). Genes are either ubiquitously expressed by housekeeping genes that are responsible for cell maintenance or are tissue-specific genes that are expressed exclusively in certain tissues (Juusola & Ballantyne, 2005). So, mRNA could be determined in a stain/ tissue recovered from the crime scene to allow identification of the body fluid or tissue in question (Juusola & Ballantyne, 2005; Setzer et al., 2008).

Since body fluids contain valuable DNA evidence, it is also possible to identify a perpetrator or proving a suspect is involved in a crime and exonerating an innocent (An et al., 2012; Budowle et al., 2009; Virkler & Lednev, 2009). Therefore, the analysis of forensic DNA evidence is the most important step that is always performed and that is

not lost with RNA based body fluid identification due to the possible co-extraction of both nucleic acids.

1.2.5. Advantages and disadvantages of mRNA profiling

The identification of the source of biological body fluids left at a crime scene plays a crucial role in the reconstruction of events that occurred and can link suspects and victims to a certain scenario providing probative value for forensic investigations (Kulstein & Wiegand, 2018; Z. Wang et al., 2013).

When a potential body fluid is found at a crime scene, a specific light or chemical techniques are needed to visualize or identify the stain (An et al., 2012). Consequently, presumptive tests are initially used to provide some indication to identify the substance and later, confirmatory tests are made to confirm the origin of the biological sample (An et al., 2012). Tests that rely on chemical or enzymatic assays are presumptive tests which are used as screening tests, and usually limited in specificity or sensitivity, such as luminol tests (chemiluminescent) for blood, and zinc paper strip (chemical) test for semen (Harbison & Fleming, 2016; Virkler & Lednev, 2009). Confirmatory tests will conclusively identify the presence or absence of body fluids since they depend on a microscopic, a spectroscopic or immunological assays, such as fluorescence tests (spectroscopy) for blood and prostate-specific antigen (PSA) tests (immunological) for semen (An et al., 2012; Harbison & Fleming, 2016; Virkler & Lednev, 2009).

These tests are a practice in forensic field to pre-analyse biological samples, however, they have limitations like incompatibility with DNA profiling or other individual identification assays, lack of sensitivity, low specificity, poor reproducibility, time consuming (long assays), high costs, instability of biomolecule assayed and high biological material consumption when it is already limited (loss of material) (An et al., 2012; Haas, Klessner, et al., 2009; Harbison & Fleming, 2016; Juusola & Ballantyne, 2003, 2007; Yang et al., 2013). According to the literature, there is no confirmatory test for the presence of saliva, menstrual blood, and vaginal secretions, which are very common body fluids found in a sexual assault crime scene (Juusola & Ballantyne, 2007; Setzer et al., 2008).

In a scenario where an ejaculation has occurred, tests mainly depend on the microscopic identification of spermatozoa, nevertheless due to aspermatic and old seminal stains resulting in negative results, the absence of spermatozoa is not indicative

of semen absence (Tamaki et al., 1989). Enzyme activity-based assays (acid phosphatase, γ -glutamyltransferase) and antigen-antibody reactions (PSA, semenogelin, SVSA) allow the identification of proteins present in seminal fluid which may be useful as a preliminary test, even in the absence of spermatozoa, when these methods are not specific for semen and have cross-reactivity with other body fluids and biological samples (Bauer & Patzelt, 2003). The advantages of an mRNA-based assay in comparison with conventional methods of analysis include greater specificity, simultaneous and semi-automatic analysis of multiple markers and tissue types through a common assay setup (for DNA analyses), the ability to co-extract DNA and RNA because of RNA compatibility with DNA extraction methodology, consequently, saving time and sample consumption and most importantly preserving DNA for identification of sample donor (Juusola & Ballantyne, 2003, 2007; Roeder & Haas, 2013).

For scientists, the biggest concern of mRNA usage is de facto that it seems highly unstable and with a fast degradation rate by the ubiquitous presence of ribonucleases, showing evident challenges for the forensic community (Nussbaumer et al., 2006; Sirker et al., 2016). Nevertheless, it's been demonstrated the possibility to isolate total RNA of sufficient quality and quantity from biological stains with many years of age (Zubakov et al., 2008). Several studies indicate high stability of RNA from old dried stains under controlled conditions (Haas, Klessner, et al., 2009; Kohlmeier & Schneider, 2012; Lindenberg et al., 2012; Setzer et al., 2008). According to Lindenberg et al., 2012, dried samples from specimens stored for many years originated positive identification, with both DNA and RNA profiling (for example: blood and semen with 28 years old, saliva with 6 years old, skin with 10 years old and menstrual secretion with 4 years old). So, RNA profiling not only allows the identification of body fluids and tissues but also plays an important role, for example, regarding determination of the age of a biological sample (Juusola & Ballantyne, 2007).

1.3. Forensic Genetics in sexual violence cases

During a crime scene investigation, the main purpose of forensic genetics is to analyse the biological material to obtain a DNA profile from the crime scene, so a genetic profile is linked to the crime. Forensic genetics field has evolved rapidly over the years with changing technologies leading to the establishment of DNA databases with DNA profiles from both suspects and convicted offenders, sustaining significant importance for convictions (Schneider, 2007). Therefore, if there is a potential suspect, it can have confirmatory results about the suspect's identity, as well as the exoneration of an innocent suspect. Biological evidence collected at a crime scene are used not only to genetically identify the donor of a stain using DNA identification but also to determine stains composition- the source (body fluid and/ or tissue type derived from the stain) with RNA analysis indicating the presence of blood, saliva, semen, etc. (Ingold et al., 2020; Sijen & Harbison, 2021). RNA is the link between the genetic information stored in the genome and the protein content of each cell and its expression and impact on the expression of cell type specific proteins have been extensively adopted as a laboratory-based method for body fluid identification in forensics (Sijen & Harbison, 2021).

Data from the World Health Organization (WHO) has estimated that 30% of women around the world has been exposed to sexual violence in their lifetime (WHO, 2021). Sexual violence is defined by any type of non-consensual of "sexual act, unwanted sexual comments or advances, or acts to traffic or otherwise directed against a person" including not only vaginal or anal penetration but also the physical force and use of coercion by anyone no matter their relationship to the victim and in any context (DeVore & Sachs, 2011; Mollen et al., 2012; WHO, 2012). This is a worldwide issue that violates women's rights and that can have severe health consequences (physical, reproductive, and psychological) including unwanted pregnancies, sexually transmitted infections (HIV for example), depression, and posttraumatic stress disorder (Jina & Thomas, 2013; Welch & Mason, 2007). According to Associação Portuguesa de Apoio à Vítima (APAV), an increase of 25.5% was noted in cases of sexual violence in the years between 2020 and 2022 in Portugal (APAV, 2022). In a total of 14688 sexual crimes committed, 77.7% of the victims were female and about 20.5% were male, and in 2595 crimes performed in children and young adults which 60.1% were also female (APAV, 2022).

To reconstruct the events of a crime scene, such as rape crimes, it is essential to have knowledge on the origin of a certain biological evidence sample (body fluids for example), being highly relevant for forensic casework (Kohlmeier & Schneider, 2012;

Lindenbergh et al., 2012). In rape crimes, the contact between the aggressor and the victim, or the environment of the aggression, (nearly) always leaves evidence whether from the aggressor to the victim, to the crime scene or vice-versa (Magalhães et al., 2015). According to Campbell & Fehler-Cabral, 2022, the majority of sexual assaults are never prosecuted and less than 10% of the cases, that are reported to the authorities, result in conviction or guilty plea since law enforcement people clear most cases without being forwarded to prosecutors for further evaluation and potential charges.

In cases of alleged sexual assault, the production of evidence relies on the forensic evidence, so the field of criminalistics is crucial and determinant to the successful outcome of forensic investigation (Magalhães et al., 2015). Recovered traces of body fluids in a sexual offence case (such as both peripheral or menstrual blood, semen, vaginal fluid, saliva, sweat, hair, pubic hair or other types, skin cells) is the most important type of evidence since it has valuable DNA evidence that can identify a suspect or a victim, as well as exonerate an innocent (Richard et al., 2012; Virkler & Lednev, 2009c). Forensic investigation in these cases relies on urgent forensic medical examinations (FME) of victims within 72 hours after the assault (Magalhães et al., 2015). The existence of physical evidence (transfer evidence) is an indicator of contact among objects that result in an exchange of biological materials between them so, it is crucial to collect the largest number of evidence and not only examine the victim (physical examination of victims' body) but also their clothing and other objects (Maras & Miranda, 2017). In order to success an investigation, it is necessary to handle evidence samples properly by preserving, storing and maintaining the chain of custody so it is ensured that evidence remains intact and preventing detrimental changes or loss while preserving its integrity (Maras & Miranda, 2017).

1.3.1. Semen in body fluid identification

Semen is a body fluid relevant for the forensic society, particularly in the sexual violence context, and it is a mixture of spermatozoa and other fluids from the seminal vesicles, the prostate, and the bulbourethral glands (Sijen, 2015). In ejaculation situations, the result of the investigation is largely dependent on the cytological detection and analysis of spermatozoa, being the most important biological evidence (semen) in these crimes (Bauer & Patzelt, 2003). However, the absence of spermatozoa is not an indicative of the absence of semen so immunological and enzymatic detection of semen specific proteins can be helpful in those scenarios (Bauer & Patzelt, 2003; Tamaki et al., 1989). In human semen, about 70% of the total ejaculated volume is composed of a yellowish viscous fluid rich in fructose and other substances that are originated in seminal vesicles. Semen samples are usually found in the form of dried stains on the victim's body (usually in the lower abdomen region) or clothes (such as underwear), in condoms, bedclothes or even at the crime scene itself (Johnson et al., 2012; Magalhães et al., 2015; Villaflores, 2018).

According to a study performed with mRNA markers for semen stains identification, specific mRNAs were successfully found in both semen and seminal fluid (azoospermic), confirming not only the presence of semen in stains deposited by healthy individuals, but also deposited by males affected with azoospermia or oligospermia (Albani & Fleming, 2018).

1.3.2. "In-use" semen markers

According to the literature, the genes protamine 1 and 2 (PRM1 and PRM2), transition protein 1 (TNP1), transglutaminase 4 (TGM4), semenogelin 1 and 2 (SEMG1 and SEMG2), prostate-specific antigen (PSA), kallikrein related peptidase 2 and 3 (KLK2 and KLK3), microseminoprotein beta (MSMB) and a prostatic tumour suppressor gene (NKX3-1) are the ones reported as among the most reliable markers for the identification of semen samples (Afolabi et al., 2017; Albani & Fleming, 2018; Dørum et al., 2018; Haas et al., 2013; Haas, Muheim, et al., 2009; E. Hanson et al., 2018; E. K. Hanson & Ballantyne, 2013; Ingold et al., 2018; Lindenbergh et al., 2012; Park et al., 2013; Richard et al., 2012; Roeder & Haas, 2013; van den Berge et al., 2014). Of all these markers, only PRM1 and PRM2 are specific mRNA sperm cells, while the others are conserved and abundant in seminal plasma; however, all of them have been demonstrated to be

possible markers for semen identification in forensic casework (Alshehhi & Haddrill, 2019; Sirker et al., 2013). Of the above-mentioned group of genes, PRM1 and SEMG1, were selected for the current project and therefore further details are given below.

PRM1 is known as a semen-specific mRNA marker and encodes protamine that substitutes histones in spermatozoa while PRM2 is a spermatozoa-specific gene that encodes protamine 2 which is cleaved to generate a group of protamine 2 peptides. (Lindenbergh et al., 2012; Nakanishi et al., 2014; Park et al., 2013). According to GTEx Portal (<https://gtexportal.org/home/>) and BioGPS database (<http://biogps.org/about/>), PRM1 tissue main expression is on testis, more specifically in seminiferous tubules, Leydig cells, testis interstitials and testis Germ cells (Figure 3). During the haploid phase of spermatogenesis, protamines serve as replacements for histones in sperm chromatin and bind to DNA in order to highly condensate the spermatid genome into a genetically inactive state (Balhorn, 2007; Bao & Bedford, 2016). So, protamines are proteins that play a crucial role in DNA and protein binding, spermatogenesis, chromosome condensation and organization and in sperm DNA condensation as well. PRM1 is only expressed in fertile man meaning that the complete absence of these genes is demonstrated in infertile patients (Lindenbergh et al., 2012; Schlicker et al., 1994). PRM1 is used as a specific marker in the forensic field to identify semen in biological evidence or stains. According to a study performed to compare the degradation process of stains, PRM1 was detected up to 70/71 weeks in dry conditions of storage and up to 20 weeks whenever in humid storage (time of detection varies depending on the amount of sample that was analysed); proving that PRM1 is a suitable sperm-specific marker not only for fresh ejaculated semen but also for RNA obtained from dried semen stains (Fleming & Harbison, 2010; Sirker et al., 2016). PRM1 is considered the most stable and reliable marker identifying a semen stain and in degraded samples since it was already possible the detection of cell-specific transcripts even in very small quantity and quality of sample material (Sirker et al., 2016). Despite that, PRM1 levels depend on the number of produced spermatozoa so, abnormalities in its expression or alteration in its structure can result in male infertility (marker is only expressed in cases of male fertility) (Aydos et al., 2018; Lindenbergh et al., 2012).

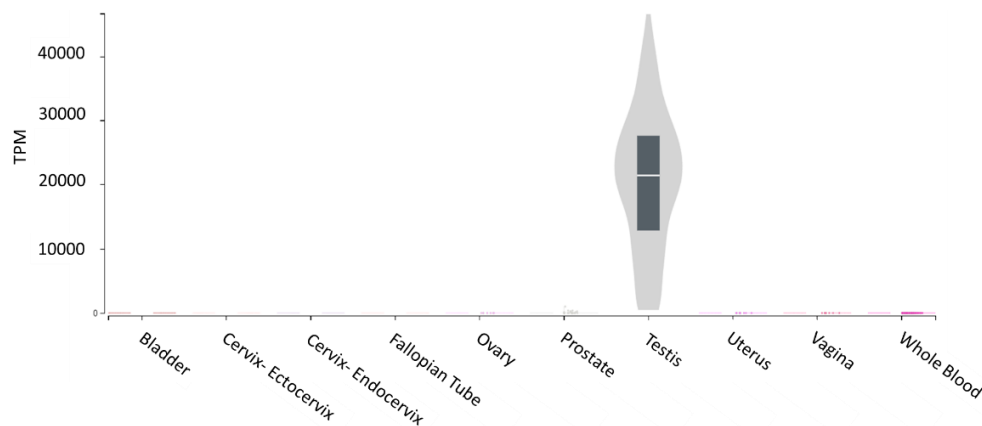


Figure 3 - Gene expression chart of PRM1 (ENSG00000175646.3). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: <https://gtexportal.org/home/gene/PRM1>.

SEMG1 is a seminal vesicle- and prostate-specific gene that encodes an abundant protein involved in the “gelatinous entrapment of ejaculated spermatozoa” (Lindenbergh et al., 2012; Nakanishi et al., 2014). Its tissue expression is mainly in the prostate (Figure 4). SEMG1 performs several important functions, such as regulation of sperm capacitation, and motility is involved in antibacterial response (Shuvalov et al., 2020). While PRM1 mRNA is exclusively expressed in fertile males, SEMG1 mRNA is detected in semen of both fertile and infertile/ vasectomized men (Lindenbergh et al., 2012). According to the literature, SEMG1 is detected in samples of vasectomized and azoospermic individuals being completely distinguished from “normozoospermic” samples, adding forensic value to this marker for body fluid identification (Haas, Muheim, et al., 2009; H. Li et al., 2012; Lindenbergh et al., 2012).

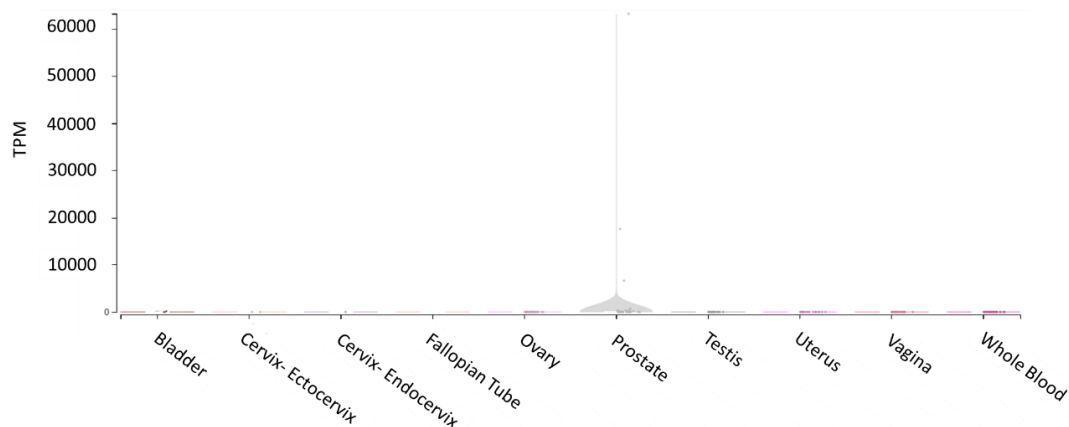


Figure 4 - Gene expression chart of SEMG1 (ENSG00000124233.11). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: <https://gtexportal.org/home/gene/SEMG1>.

1.3.3. “New” semen markers

Beyond PRM1 and SEMG1, other markers were selected for this study to identify semen in biological evidence. According to a study performed about the degradation of mRNA transcripts over time in biological crime scene traces, it is shown that outer dense fiber of sperm tails 1 (ODF1), outer dense fiber protein 2 (ODF2), PHD finger protein 7 (PHF7), sperm mitochondria associated cysteine rich protein (SMCP), T-complex 11 (TcP11) and transition protein 1 (TNP1) are potential mRNA markers for semen identification (Salzmann et al., 2021). These transcripts were included since they allowed the calculation of the dispersion coefficient in samples up to 6 months old (Salzmann et al., 2021).

1.3.3.1. ODF genes (ODF1 and ODF2)

According to NCBI database, the family of outer dense fibers (ODF) are cytoskeletal structures that surround the central core of the sperm tail in the middle piece and main piece of the sperm tail (ODF1 Outer Dense Fiber of Sperm Tails 1 [Homo Sapiens (Human)], 2023). ODF family are comprised of over twelve distinct proteins wherein only a few have been recognized, and, while not directly involved in the active movement, they play an important role maintaining the stability and flexibility of the sperm tail as well as supporting the flagellar motion (K. Yang et al., 2012). So, during the development of the sperm tail, these code proteins are connected along the axonemal tubuli doublets and preserve the elastic structure and recoil of the sperm tail while protecting it against

damage from external forces during epididymal transport and ejaculation (Ghafouri-Fard et al., 2010).

ODF1 is crucial for establishing a connection between the sperm head and tail, since its main molecular function is protein binding and this gene encodes the major protein of the ODF family of the sperm tail (Kleene & Bagarova, 2008; K. Yang et al., 2012). Furthermore, a loss of function of ODF1 could potentially explain instances of human infertility characterized by decapitated sperm heads (K. Yang et al., 2012). ODF1 tissue expression is limited to testis (Figure 5).

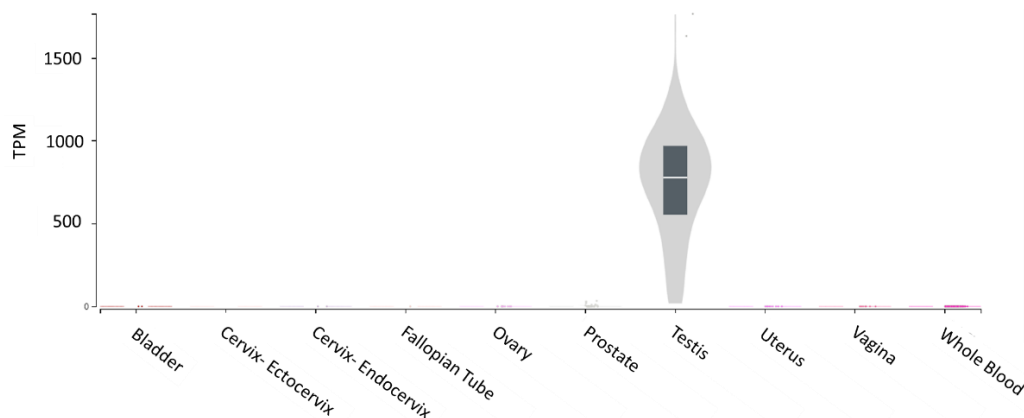


Figure 5 - Gene expression chart of ODF1 (ENSG00000155087.3). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: <https://gtexportal.org/home/gene/ODF1>.

ODF1s interacting proteins have been recognized, such as ODF2 which is the main protein of the accessory fibers in the sperm tail, also expressed in testis (Figure 6) (Tapia Contreras & Hoyer-Fender, 2019). ODF2 is also a protein that participates in the development of cytoskeletal structures specific to spermatids and potentially acts as a binding protein as well (Tapia Contreras & Hoyer-Fender, 2019). Moreover, the scaffold of the ODF fiber might be established by these two primary ODF proteins, however it's not yet confirmed what proteins interact with ODF1 in order to form sperm specific structures (Tapia Contreras & Hoyer-Fender, 2019; K. Yang et al., 2012).

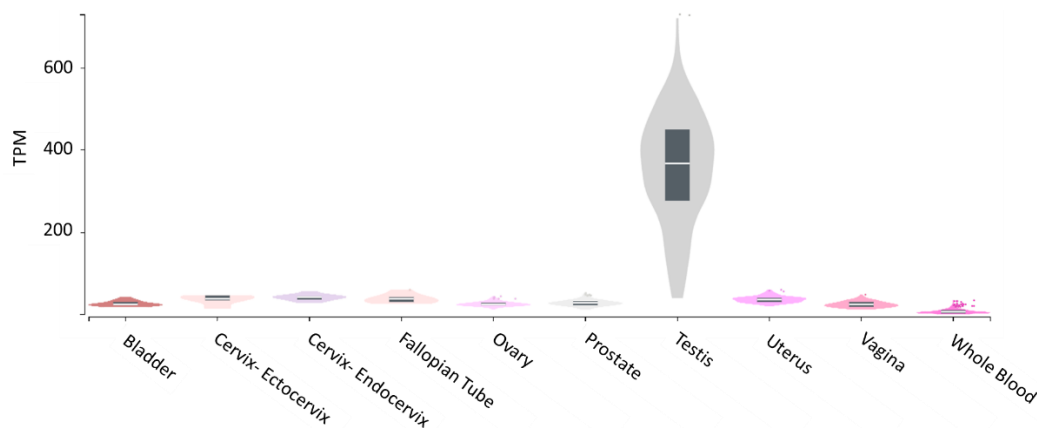


Figure 6 - Gene expression chart of ODF2(ENSG00000136811.16). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: <https://gtexportal.org/home/gene/ODF2>

Malfunctions in the development of the outer dense fibers results in abnormalities in sperm shape and infertility due to the sperm motility and morphology (K. Yang et al., 2012). It has been described that the proportion of ODF defects was higher in samples with asthenozoospermia (defects in motility) in comparison with “normal” samples, associating the proportion of axoneme defects and non-motile sperm (Zhao et al., 2018). Although it has been demonstrated that the lack of ODF proteins could impact the sperm motility resulting in ineffective sperm tails, it remains unknown the connection between ODFs and sperm movement as well as the potential role of ODF abnormalities in causing asthenozoospermia (Hashemitabar et al., 2015; Zhao et al., 2018).

1.3.3.2. PHF7

It is believed that PHF7 plays a role in the germ-cell development and spermatogenesis since it maintains the germline stem cells, facilitates gametogenesis in males and acts as a conserved epigenetic factor by triggering the initiation of the germline sexual cycle (Hering et al., 2014). Spermatogenesis is a complex process that is divided in three steps: mitotic cell division (multiplication of undifferentiated male germ cell); meiosis (reduction of the chromosome number) and spermiogenesis (transformation of the spermatid into the spermatozoon (motile sperm) structure (de Kretser et al., 1998). So, spermatogenesis is controlled by both internal and external factors, as well as cellular interactions involving interstitial cells of the testis, Sertoli cells, and germ cells (PHF7 PHD Finger Protein 7 [Homo Sapiens (Human)], 2023). PHF7 restricted expression is in the testis (Figure 7) in Sertoli cells. The gene's encoded protein includes plant homeodomain (PHD) finger domains which is believed to participate in transcriptional regulation (PHF7 PHD Finger Protein 7 [Homo Sapiens (Human)], 2023).

In *Drosophila*, PHF7 assumes a crucial role determining the male germline sexual identity (Ge et al., 2017). It's notable that it displays a selective expression in the early germ cells, germline stem cells and spermatogonia specific to males by contributing to male sexual determination and regulating sperm production, and its equivalent genes in humans can effectively correct fecundity issues in male fruit flies without this gene (Ge et al., 2017; X. R. Wang et al., 2017). It was also discovered that this gene is aberrantly regulated in males experiencing infertility indicating that its loss of function induces impaired sperm development during the late phase of spermatogenesis in mice (X. Wang et al., 2019).

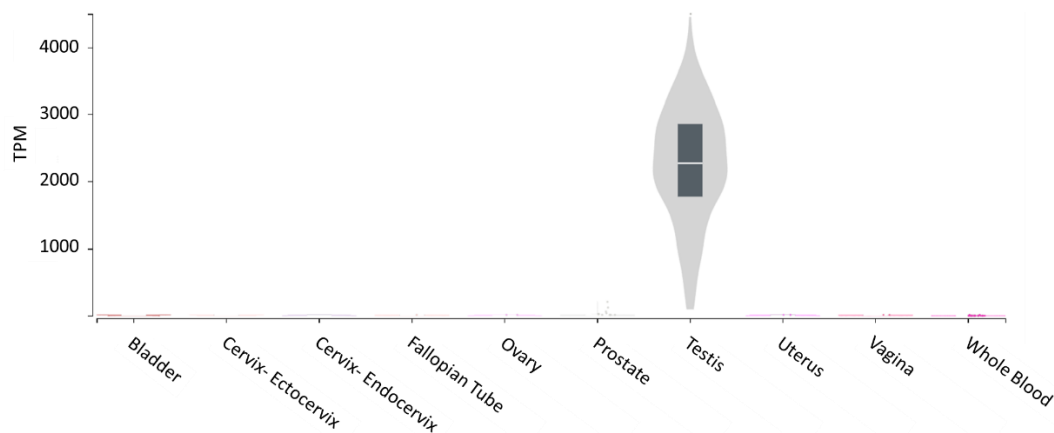


Figure 7 - Gene expression chart of PHF7 (ENSG00000010318.20). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: <https://gtexportal.org/home/gene/PHF7>.

1.3.3.3. SMCP

Mitochondrial sperm cells have distinct shapes and localization when compared to somatic cells and, in the midpiece of the mammalian spermatozoa flagellum, mitochondria are positioned in a circumference forming the mitochondrial sheath (Nayernia et al., 2002; Takahashi et al., 2013). A mitochondrial structure (capsule) is responsible for maintaining the proportions (size and shape) of the outer membrane of the mitochondria and the functioning of the arrangement and maintenance of the mitochondrial sheath (Nayernia et al., 2002; Takahashi et al., 2013). This mitochondrial capsule integrity is preserved by a cysteine and proline structural protein known as SMCP and it is believed that SMCP plays a role controlling and regulating structure of the mitochondrial sheath in spermatozoa (Ando et al., 2012; Takahashi et al., 2013).

SMCP is only significantly expressed in testis tissues (Figure 8) in human adults (after puberty) and in the sperm tail structures that are extremely differentiated (Herr et al., 1999). This protein is known to play an essential role in chromatin and protein binding and according to (Sharma et al., 2015), SMCP genes are “genetically regulated by epigenetic factors” observing that SMCP expression decreases as paternal age increases. Several studies have shown that a decreased expression of SMCP genes leads to a low-quality semen since it affects the development and motility of the flagellum’s sperm, even though still fertile (Abu-Halima et al., 2023; Lehti & Sironen, 2017; Sharma et al., 2015). Consequently, the lack of SMCP’s function results in

spermatozoa with decreased speed and a higher percentage of asthenozoospermia in males due to the reduction of energy into the flagellum (Nayernia et al., 2002).

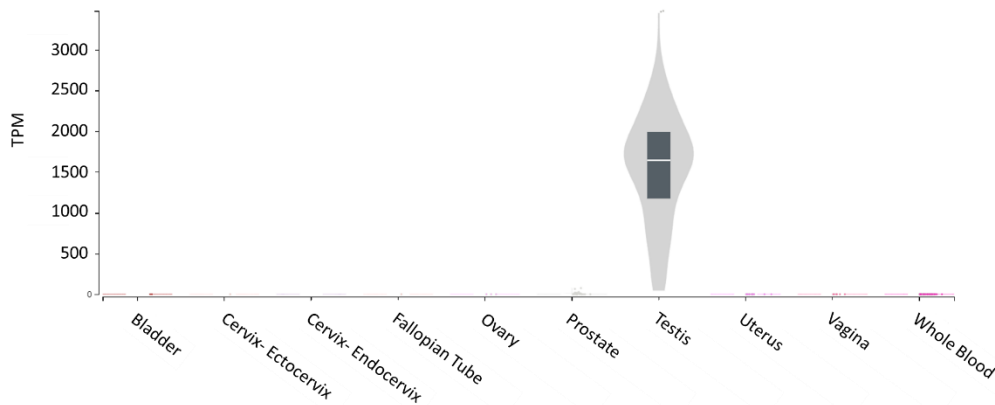


Figure 8 - Gene expression chart of SMCP (ENSG00000163206.5). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: <https://gtexportal.org/home/gene/SMCP>.

1.3.3.4. TcP11

In the t-complex, the major histocompatibility complex (MHC) in chromosome 6, encodes with genes that are involved in spermatogenesis and embryogenesis, affecting male fertility (Castaneda et al., 2020; Y. Liu et al., 2011). One of those genes of the t-complex, TcP11, reveals specific expression in testis tissues (Figure 9) (Castaneda et al., 2020). TcP11 can be found in both the sperm head and flagellum in spermatozoa, playing an important role in the functioning of sperm related to sperm capacitation, fertilization ability inhibiting the acrosome reaction, and motility (Fraser & Dudley, 1999; Y. Liu et al., 2011). Studies have shown that the human TcP11 gene encodes the receptor for the fertilization-promoting peptide (FPP) stimulating the capacitation and fertilizing capacity in spermatozoa (Adeoya-Osiguwa et al., 1998; Y. Liu et al., 2011). TcP11 also forms a complex with PKA (an enzyme responsible for the capacitation pathway) in sperm contributing the capacitation through the cAMP pathway (Castaneda et al., 2020; Stanger et al., 2016). So, it is believed that defects in the TcP11 gene can cause a weak bind of FPP leading to a male subfertility since FPP is crucial for the fertilization (Adeoya-Osiguwa et al., 1998; Fraser & Dudley, 1999).

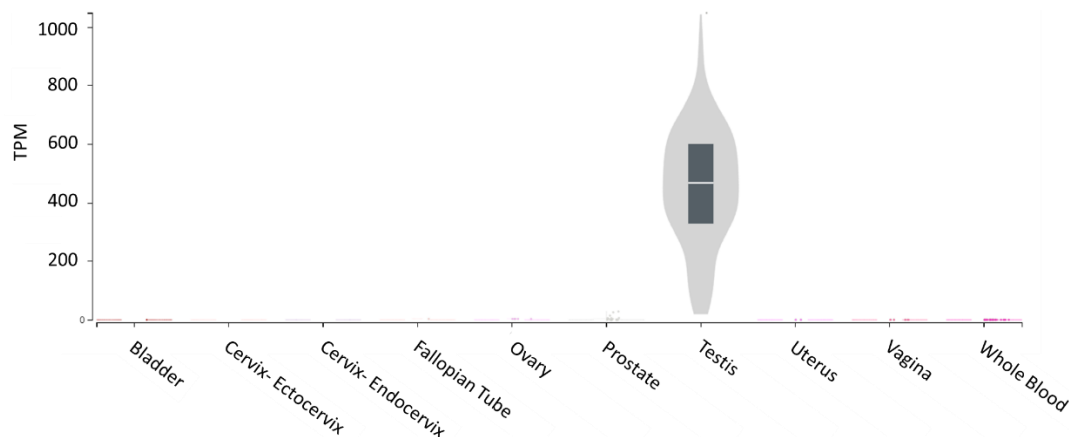


Figure 9 - Gene expression chart of TcP11 (ENSG00000124678.17). Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: <https://gtexportal.org/home/gene/TCP11>.

Examination of the TcP11 mRNA and protein expression revealed that TcP11 mRNA can be detected before meiosis, but the protein can only be found until the spermatids differentiation (just before the spermatids differentiation) (Fraser et al., 1997; Fraser & Dudley, 1999). T-complex DNAs is rearranged by inversions causing homologues sequences that do not line up in meiosis so if a mutation occurs in the t-complex, it will not be deleted by the usual events (gene conversion for example) (Fraser et al., 1997). Therefore, mutated spermatozoa have short tails or occasionally no tails, affecting sperm motility (Fraser & Dudley, 1999).

1.3.3.5. TNP1

The key proteins found in condensation of chromatin spermatids are the transition nuclear proteins (TNPs), mainly TNP1 and TNP2 (Yu et al., 2000). TNPs can be found in many mammals, including humans, that belongs to a small group of spermatid-specific nuclear proteins involved in the chromatin of condensing spermatids during the haploid stage of spermatogenesis (Heidarana et al., 1989; Jedrzejczak et al., 2007). TNP1 is a spermatid-specific protein that replaces histones in the haploid phase and is later replaced byprotamines (PRM1 and PRM2) in the mature sperm cells (Heidaran et al., 1988; Heidarana et al., 1989; Jedrzejczak et al., 2007). TNP1 gene is mostly expressed in the testis tissues (Figure 10) and is expressed in the late stage of sperm maturity (Jedrzejczak et al., 2007).

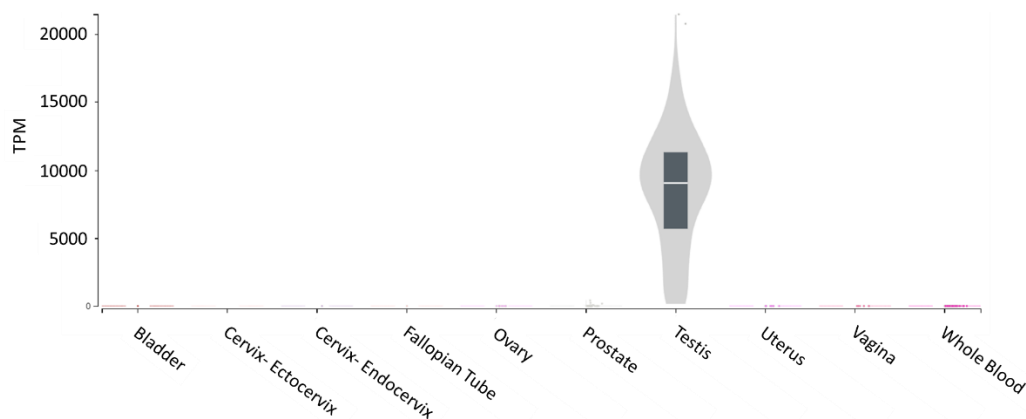


Figure 10 - Gene expression chart of TNP1 (ENSG00000118245.2) . Data source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2). Retrieved from: <https://gtexportal.org/home/gene/TNP1>.

Studies have demonstrated that males with a null TNP1 have shown reduced sperm motility since morphological sperm defects were discovered, decreased chromatin condensation, and a fertility rate lower than normal (around 60% of these males were infertile) (Hirenallur Maheshwarappa et al., 2019; Shirley et al., 2004; Yu et al., 2000). According to Yu et al., 2000), the lack of TNP1 can be balanced with TNP2 and PRM2 precursor therefore, not being essential for the histone replacement or the chromatin condensation but resulting in defected chromatin condensation pattern and still showing a decrease rate of fertility.

Further studies are needed about the TNP1 since it is known that it plays a crucial role regarding the spermiogenesis, however, it is not clear its function producing normal sperm (Miyagawa, 2005; Shirley et al., 2004).

1.3.4. PCR multiplex development for semen identification

The PCR technique has become the most currently used in forensic science since it demonstrates a high sensitivity and an effective multiplex system of highly polymorphic microsatellite loci within a single PCR tube and a high reproducibility in results (Zhu et al., 2020).

Furthermore, it has been shown that multiplex PCR, also with mRNA markers with fluorescence-labelled primers (after cDNA synthesis) and followed by capillary gel electrophoresis are a useful tool for the identification of human body fluids for forensic analysis. These systems can be built with selected genes that were previously

recognized for a specific tissue or body fluid expression (Juusola & Ballantyne, 2005; Lindenberg et al., 2012; Wang et al., 2013). The use of multiplex PCR system has helped mRNA profiling growth since it allows the simultaneous analysis of a group of body fluid specific markers in a single multiplex reaction and can be easily combined with DNA and RNA profiling using the same collected stain sample (due to co-extraction) (Cortellini et al., 2020; Haas et al., 2009). Consequently, the number of analyses that can be done with very limited sample material increases (Roeder & Haas, 2016). Several multiplexes for endpoint PCR allowing the identification of blood, saliva, semen, menstrual blood, and blood, for example, using up-to 19 body fluid-specific mRNAs has already been described, proving that PCR can be an appropriate method for the detection of transcripts with high tissue-specific expression, reducing sample consumption, in one single reaction (An et al., 2012; Haas et al., 2009; Lindenberg et al., 2012; Wang et al., 2013).

2. Objectives

The study of biological evidence collected from crime scenes are crucial since it can identify the stain donor by means of DNA testing. Additionally, and more recently, the research on RNA in forensic genetics has been increasing since it allows the identification of the biological source of a sample. This new area of research allows the identification of body fluids and tissues from samples submitted to forensic investigations and consequently, help in the reconstruction of the crime event. mRNA markers have shown to be highly stable and with tissue-specific expression patterns proving its efficiency for the identification of fluids or tissues in forensic casework.

The main objective of the present study was to develop a multiplex system with newly reported mRNA markers specific for semen identification and confirm its suitability in forensic genetics testing. For that reason, the following objectives were proposed:

- 1) To design a multiplex system with recently identified mRNA semen-specific markers as new potential candidates for body fluid identification in forensics;
- 2) To develop and optimize a multiplex mRNA system containing the final selected semen-specific markers for amplification in a single PCR;
- 3) To access the robustness and applicability of the final mRNA multiplex system for forensic cases by testing both sensibility and specificity of these candidate markers and the performance on older samples (≥ 17 months) collected at different time points.

3. Materials and methods

3.1. Collection of biological samples and RNA extraction

RNA extracts were already available as the semen samples used in this study were collected in the scope of a previous master thesis dissertation (Loureiro, 2019). RNA extractions were performed using the commercial kit ExtractME RNA & DNA (Extractme, BLIRT S.A., Poland) as described in the latter study following the manufacturer's recommendations (Loureiro, 2019). A short description is provided regarding the nature of the samples used: semen fresh samples were collected from supposedly healthy, unrelated individuals with ages between 20 and 60 years. Every participant provided informed written consent so the confidentiality of all the collected samples was granted. Samples were provided as a freshly ejaculated liquid in a sterile and sealed plastic tubes by 14 male volunteers. Subsequently, semen samples were stored in a freezer at -20°C in order to avoid nucleic acid degradation.

3.2. Coding and storage of biological samples

To avoid degradation due to "freeze-thaw" as well as to decrease contamination risks, RNA aliquots of 20 µl from semen samples were prepared and stored in a -20°C freezer. After anonymization of samples, each RNA sample was coded in accordance with the semen sample coding S1 to S14 (where "S" is referring to the semen sample) and male volunteers were randomly distributed numbers ranging from 1 to 14 denoting the replica number for each individual.

3.3. mRNA semen markers and new primer designs

In this study, a total of eight semen expressed genes (SEMG1, PRM1, ODF1, ODF2, TcP11, PHF7, SMCP and TNP1) were selected for semen identification. SEMG1 and PRM1 genes were obtained from Juusola & Ballantyne and Haas et al. studies (Haas, Muheim, et al., 2009; Juusola & Ballantyne, 2005). On the other hand, ODF1, ODF2, TcP11, PHF7, SMCP and TNP1 markers were selected from the work of Salzmann et al. as potential mRNA markers (Salzmann et al., 2021). Although other markers such as PRM2 and KLK3 are already described in the literature for semen identification, the objective was to validate and test new markers (B. Liu et al., 2020). Characteristics such

as tissue expression from each primer, gene localization and gene sequence are listed in Table 2.

Table 2 - Genes, tissue expression and genomic location (BUILD GRCh38/hg38) of the selected genes (data retrieved from www.genecards.org).

Markers	Tissue expression	Gene localization
SEMG1	Prostate gland	<i>chr20:45,207,033-45,209,768</i>
PRM1	Testis	<i>chr16:11,280,841-11,281,330</i>
ODF1	Testis	<i>chr8:102,551,589-102,561,018</i>
ODF2	Testis	<i>chr9:128,455,185-128,501,292</i>
TcP11	Testis	<i>chr6:35,118,071-35,148,610</i>
SMCP	Testis	<i>chr1:152,878,322-152,885,047</i>
TNP1	Testis	<i>chr2:216,859,458-216,860,064</i>
PHF7	Testis	<i>chr3:52,410,660-52,423,641</i>

Primer sequences for each selected mRNA marker, such as its product sizes and melting temperatures are listed in Table 3. Primer sequences were designed with Primer3 software (Koressaar & Remm, 2018) and assuring that at least one primer (within a given primer pair) spans an exon-exon junction, which is useful for limiting the amplification only to mRNA. Fluorescently labeled mRNA primers were only ordered after primer performance of each marker were tested in singleplex amplifications. The fluorochromes used (6-FAM and YelYak) are called Freedom™ Dyes (from IDT-Integrated DNA Technologies) which are fluorophores with no licensing restrictions from IDT or third-party companies. Freedom dyes include IDT proprietary fluorophores that are available for commonly used dye wavelengths.

Table 3 - mRNA markers with the respective primers sequences, melting temperature, 5' Fluorochrome, expected amplicon sizes and References: a) (Juusola & Ballantyne, 2005), b) (Haas et al., 2009).

Gene	Primer Sequences	Tm*	Label 5'	Size (bp)	References
SMCP	f: AGGCTGCTGAGGAAGATCAAT	59.17	6-FAM	105	This study
	r: GAACACTTGGAGGTACTGGTC	57.96			
TcP11	f: TCACAGGTCTGACAGAAACCG	59.93	YelYak**	117	This study
	r: CCTTGCCTTCCAGACTGCTT	60.25			
ODF1	f: TCGCGCCTGTGTGATCTCTA	60.97	YelYak**	143	This study
	r: GTTGTCTTCTCAGTTTGGCAAG	58.65			
ODF2	f: GACCGGGTGTGCGAAAAGG	60.67	YelYak**	163	This study
	r: ACACGATGGAAACCAGGAGC	60.32			
TNP1	f: AAGCCCTCATTTTGGCAGA	59.89	6-FAM	180	This study
	r: GTAATTGCGATTGGCGTCAT	57.24			
PHF7	f: GCAAAGAACTGAAGTTTCAAGCTG	59.50	YelYak**	200	This study
	r: AGGCATAGCCAGCAAACAGG	60.68			
PRM1	f: GCCAGGTACAGTGCTGTCGCAG	65.86	6-FAM	153	a
	r: TTAGTGTCTTCTACATCTCGGTCT	58.50			
SEMG1	f: TCGGTAACCATGTGAAAGGA	56.47	6-FAM	120	b
	r: GTGTCATCCATGGACCAAGA	57.21			

Primers were ordered in a lyophilized state with 100 µM, so primer solutions had to be prepared. First, they were all submitted to a quick spin for 15 seconds (to ensure no oligo product remains stuck to the cap), then H₂O was added according to each primer using the manufacturer's volumes (Table 4a and 4b), then centrifuged again and finally a water bath at 50°C with shaking for 10 minutes to obtain full dissolution of oligos.

Table 4 - a) Quantity of H₂O added to markers without the label 5'; b) quantity of H₂O added to markers with the label 5'

a)	Markers	H ₂ O (µl)	b)	Markers (F)	H ₂ O (µl)
	SMCP (F*)	259		SMCP (6-FAM)	585
	SMCP (R**)	256		TcP11 (YelYak)	484
	TcP11 (F)	263		ODF1 (YelYak)	821
	TcP11 (R)	272		TNP1 (6FAM)	697
	ODF1 (F)	320			
	ODF1 (R)	226			
	ODF2 (F)	284			
	ODF2 (R)	316			
	TNP1 (F)	327			
	TNP1 (R)	266			
	PHF7 (F)	218			
	PHF7 (R)	258			

F*: referring to primer forward; **R****: referring to primer reverse

3.4. Reverse transcription reaction: cDNA synthesis

To allow the amplification of mRNA markers specific to the body fluid, it is needed to generate complementary DNA (cDNA) to act as a template in the endpoint of the PCR reaction. So, RNA is converted into cDNA by an enzyme (reverse transcriptase). Reverse transcription (RT) of the RNA sample was performed for each RNA sample of semen (14 samples) that were obtained in this study. cDNA synthesis was made with the commercial “SensiFAST cDNA Synthesis kit” (BIRLT S.A., Poland) in a final volume of 10 µl. The reagents of RT reaction were added according to Table 5, and it was reproduced for all 14 samples. RT negative (reaction without the enzyme Reverse Transcriptase) was also include as control for unwanted DNA amplification.

Thermal cycle conditions for the cDNA synthesis reaction were 25°C for 10 minutes (primer annealing), 42°C for 15 minutes (reverse transcription) and 85°C for 5 minutes (inactivation). All samples of cDNA reaction products were further diluted in a concentration ratio of 1:10 and stored at -20°C.

Table 5 - Reagents and respective volumes used for one sample for the cDNA reaction.

Reagents	Volume per sample
5x TransAmp Buffer	2 µl
Dnase/Rnase free-water	2.5 µl
Reverse transcriptase	0.5 µl
TOTAL: 5 µl Reaction Mix + 5 µl RNA	

3.5. PCR of mRNA markers (single amplifications)

Polymerase chain reactions were performed for each gene individually, before multiplex development, to ensure successful performance of each new primer set that was designed. A reaction mix was made containing primers (forward and reverse), along with the amplification kit HotStarTaq MasterMix (Qiagen, Hilden, Germany) and H₂O, for amplification of each of the 14 samples, as listed in Table 6. 1 µl of cDNA was added to the reaction mix of each study samples. Thermocycling conditions for each PCR reaction are listed in Table 7.

Table 6 - Reagents and respective volumes per sample for single amplification assays.

Reagents	Volume per sample (µl)
H ₂ O	2.2
Primer forward (5 µM)	0.4
Primer reverse (5 µM)	0.4
Amplification kit	4
Total of 8 µl: 7 µl reaction mix + 1 µl diluted (1:10) cDNA	

Table 7 - Thermocycling conditions each PCR reaction.

PCR phase	Temperature (°C)	Time	Cycles
Initial denaturation	95	15 min	1x
Denaturation	94	20 sec	
Annealing	60	30 sec	35x
Extension	72	40 sec	
Final extension	72	15 min	1x

3.6. Polyacrylamide gel electrophoresis (PAGE)

To verify the efficacy of amplification of the eight primers (SMCP, TcP11, ODF1, ODF2, PHF7, TNP1, SEMG1 and PRM1), in the 14 semen samples, polyacrylamide gel electrophoresis (PAGE) was performed from both single and multiplex assays (development of multiplex described below) by visualizing the amplified fragments on a gel. Polyacrylamide gel electrophoresis is a method that relies on the separation of fragments depending on their molecular size. Negatively charged fragments migrate through the polyacrylamide gel toward the positive pole, when applying an electric current. Because the mobility of these molecules is inversely proportional to their size, smaller fragments migrate faster through the gel in comparison with larger ones, so when looking at a gel membrane, small fragments are at the bottom ("front" of the run) and large at the top of the gel. To visualize the bands of fragments, silver nitrate is used in this case to stain the gel.

First, a polyacrylamide solution was arranged with 25 ml of gel buffer solution (1.5M Tris-HCL), 20 ml of acrylamide (Bis acrylamide 19:1 40%), 7 ml of glycerol and 43 ml of deionized water. Then, to prepare the gel, 6 ml of the polyacrylamide solution, 340 µl of

APS and 14 μ l of Temed was added. This final solution was poured into the frame glass and left for 20 minutes to polymerize.

For the electrophoresis run, 1 μ l of amplified cDNA (from Table 5 reaction set-up) was loaded onto the gel. A negative control was also loaded into the wells of a polyacrylamide gel. All samples ran through the gel using an Amersham Biosciences Multitemp III thermostatic circulator and a consort EV 243 power supply (BioRad, USA) with an electric current set at 195 V, 35 mA and 20 W and adding a 100 bp ladder as a reference for fragment length.

After the run, band visualization was performed with the silver staining method that involves the accumulation of metallic silver onto the gel's surface, specifically targeting the areas corresponding to protein bands, revealing the amplified products. The gel was immersed inside a hood fume due to toxic reagents following the respective steps:

- 1) Ethanol 10% for 10 minutes;
- 2) Nitric acid 1% for 5 minutes with agitation;
- 3) Wash with distilled water for 10 seconds;
- 4) Silver nitrate 0.2% for 20 minutes in the dark with agitation;
- 5) Wash with distilled water for 10 seconds;
- 6) Solution with 6 g of sodium carbonate (0.28M), 1 ml of formaldehyde 0.02% and 200 ml of distilled water poured until the appearance of the bands is noted;
- 7) Acetic acid 10% for 2 minutes;
- 8) Wash with distilled water.

3.7. Multiplex development and optimization

3.7.1. Hexaplex mRNA system

To validate primers performance for semen secretions, a final hexaplex PCR multiplex reaction was developed for six of the eight initially selected semen-specific primers: SMCP, TcP11, ODF1, TNP1, SEMG1 and PRM1. All 14 RNA extracts were used for this testing process.

The multiplex set-up was as following: a Primer Mix was performed with all primers at a starting concentration of 2 μ M in a final volume of 200 μ l. The PCR reactions were conducted in a total volume of 8 μ l with 1 μ l of cDNA; the reagents per sample are listed

in Table 8 and the final concentration of each primer in the PCR was 0.25 μ M. A negative control was included in all PCR reactions lacking cDNA. The thermocycling conditions for the PCR reactions used were the same as the ones listed in Table 7.

Table 8 - Reagents and respective volume per sample for developed hexaplex.

Reagents	Volume per sample (μ l)
H2O	2
Primer Mix	1
Amplification kit	4
Total of 8 μ l: 7 μ l reaction mix + 1 μ l diluted (1:10) cDNA	

3.7.2. Multiplex optimization process

Several attempts of PCR reactions with different conditions were tested as optimization procedures to optimize peak imbalance among markers. A total of four Primer Mixes were tested with different primer concentrations as listed in Table 9.

Both primers (forward and reverse) of each marker always had the same concentration in the different Primer Mixes. Besides the different mixes that were tested, PCR conditions were also analysed such as: cDNA dilution 1:20 (PCR 1) and 32 cycles in the PCR reaction.

Table 9 - Primer concentrations (μ M) in the respective Primer Mix (1-4) during multiplex optimization. In bold are highlighted the major differences in primer concentrations among primer mixes optimization.

	mRNA markers concentration in Primer Mix (μ M)						mRNA markers concentration in PCR reaction (μ M)					
	SMCP	TcP11	ODF1	TNP1	SEMG1	PRM1	SMCP	TcP11	ODF1	TNP1	SEMG1	PRM1
P Mix 1	2	2	2	2	2	2	0.25	0.25	0.25	0.25	0.25	0.25
P Mix 2	1	-	-	-	-	1	0.125	-	-	-	-	0.1254
P Mix 3	-	-	-	-	-	2	-	-	-	-	-	0.25
P Mix 4	-	2.5	2.5	-	-	1.5	-	0.3125	0.3125	-	-	0.1875

3.8. Capillary Electrophoresis (CE)

Capillary electrophoresis is an analytical technique relying on the separation of DNA amplicons according to their molecular size enabling the detection of these fragments through a fluorescent labelled marker. Therefore, PCR products migrate over a capillary containing a polymer by an electronic current. The migration speed of DNA fragments varies based on their lengths, resulting in different detection times (Lee et al., 2012). So, the previously labelled fluorescent fragments are detected as they go through the laser at the column's end.

Therefore, after the PCR procedure, Post PCR was performed in a final volume of 11 µl: 10 µl of a mix solution containing 0.4 µl of GeneScan 500 LIZ dye Size Standard (Applied Biosystems, Thermo Fischer Scientific, USA) and 9.6 µl of HiDi formamide (Applied Biosystems) added to 1 µl of PCR product. The process of loading samples onto the CE and running the instrument was executed following the guidelines and instructions provided by the manufacturer.

Each sample was submitted to capillary electrophoresis analysis with the system ABI 3500 Genetic Analyser for Human Identification (Applied Biosystems, Thermo Fisher Scientific, USA) product detection. This procedure was conducted in-house at the Genomics Core Facility Gen-Core and the obtained electropherograms were analysed with GeneMapper ID-X 1.4 Software (Applied Biosystems, Thermo Fisher Scientific, USA).

3.9. Sensibility testing of mRNA semen-specific hexaplex

To evaluate the detection threshold, a sensibility assay of the developed hexaplex system a dilution series method for the semen samples was assessed. The established detection limits of the semen samples were determined with the smallest sample concentration required to detect the specific mRNA markers. Therefore, five quantifications of one sample (S6) were performed with the NanoDrop One microvolume UV-Vis spectrophotometer (Thermo Scientific Scientific, USA). The average of these five quantifications was made to obtain the concentration of S6.

For this assay, the different dilutions rates were prepared with cDNA samples assuming that 1 ng of RNA is reverse transcribed into 1 ng of cDNA. As a result, the selected sample was submitted to a dilution covering a range of concentration of 5 ng/µl

to 0.02 ng/μl (5 dilutions were performed) as outlined in Table 10. All samples were amplified with the hexaplex system under the same conditions as the ones used in the multiplex assay (section 3.5 described in Table 7). A negative control was included in the PCR reaction lacking cDNA and two positive controls were also added (S2 and S4) at 1:20.

Table 10 - Concentration of cDNA fore respective dilutions prepared for sensibility testing.

Dilution	Dilution ratio	Concentration (ng/ μl)
D1	1:4	5
D2	1:4	1.25
D3	1:4	0.31
D4	1:4	0.08
D5	1:4	0.02

3.10. Specificity testing of mRNA semen-specific hexaplex

To assess the specificity of the six selected semen-specific markers for identification of semen in body fluids, analysis of other different biologically relevant fluids of forensic significance were conducted. Consequently, one vaginal secretion sample, one blood sample, one menstrual blood sample and one saliva sample were incorporated and subjected to the developed hexaplex system. A negative control was also included lacking cDNA and a positive control was added (S4, a previously positively tested sample). All RNA extracted from the samples for this assay were synthesized to cDNA and diluted to 1:10, except for the semen sample (1:20).

3.11. Degradation assessment of mRNA semen-specific hexaplex

3.11.1. Samples coding and storage

In this study, six semen samples that were collected at the “Instituto Nacional de Medicina Legal e Ciências Forenses- INMLCF- Delegação do Norte” in another study were also submitted to analysis to test the robustness of the semen markers by assessing the degradation performance over time. Coding of semen samples and their collection date are listed in Table 11. Samples were collected and stored in darkness at room temperature since they were collected (swabs). Time of degradation corresponds to the time since samples were collected to the day they were extracted (Table 11).

Table 11 - Coding samples and corresponding time of degradation among samples

Semen samples	Time of degradation (days)					
ML-S1	502	}	*	}	**	}
ML-S2	506					
ML-S3	509	}		}	***	}
ML-S4	511					
ML-S5	538	}		}	****	}
ML-S6	542					

*4 days; ** 7 days; *** 9 days; **** 36 days; ***** 40 days

3.11.2. Co-extraction of DNA and RNA

To extract both DNA and RNA from the semen samples the ExtractMe RNA & DNA Kit (Extractme, BLIRT S.A., Poland) was used. This commercial kit was designed to enable quick and simultaneous isolation of high-quality total RNA and genomic DNA. Guidelines were followed provided by the manufacturer; however, a few changes were necessary to adapt to the study samples. To remove impurities and substances that inhibit the enzyme’s activity the manufacturer protocol is divided into two stages: the Isolation Protocol and the Total Purification Protocol.

3.11.2.1. Isolation Protocol

1) Biological material (swab of each semen sample) was placed in a 1.5 ml Eppendorf tube. 600 µl of Lysis Buffer and 20 µl of DTT were added and vortexed for 60 s. An incubation over night at 56°C at 350 rpm was performed.

2) Samples were centrifuged for 120 s at 15000 x g.

3) The supernatant was transferred into a DNA Purification Column (red column) placed in a collection tube and was centrifuged for 30 s at 12000 x g. Filtrate was kept and used in Step 4 and the DNA Purification Column was saved in a new collection tube for further DNA Purification.

4) 600 µl 70% ethanol was added to the filtrate from Step 3 and was mixed well by pipetting.

5) 700 µl of the obtained mixture was transferred into an RNA Purification Column (violet) placed in a new collection tube and centrifuged for 15 s at 12000 x g. The filtrate was then discarded and the RNA column with the collection tube was saved for further RNA Purification. The remaining mixture was transferred into the same purification minicolumn and was centrifuged for 15 s at 12000 x g. Filtrate was discarded and placed in a new collection tube.

3.11.2.2. Total RNA Purification

1) 700 µl of RNA Wash Buffer 1 (RW1) was added into the RNA Purification Column (violet) and centrifuged for 15s at 12000 x g. Filtrate was discarded and collection tube reused.

2) 500 µl of RNA Wash Buffer 2 (RW2) (with 100% ethanol) was added to the same RNA Purification Column and was centrifuged for 15 s at 12000 x g. Filtrate was discarded and collection tube reused. This step was repeated one more time.

3) RNA Purification Column was centrifuged for 90 s at 15000 x g. Both collection tube and filtrate were discarded and the purification minicolumn was carefully transferred to a sterile RNase-free 1.5 ml Eppendorf tube (it was crucial to remove all the alcohol from the minicolumn before elution).

4) Purified RNA was eluted in 50 µl of RNA Elution Buffer (REB) and was centrifuged for 60 s at 12000 x g. Another 30 µl RNA elution was performed with same columns.

3.11.2.3. Genomic DNA Purification

1) 700 µl of DNA Wash Buffer 1 (DW1) was added into a DNA Purification Column (red) and centrifuged for 15 s at 12000 x g. Filtrate was discarded and collection tube was reused.

2) 500 µl of Wash Buffer 2 (W2) (with 100% ethanol) was added into the same DNA Purification column. Filtrate was discarded and collection tube was reused. This step was repeated one more time.

3) DNA Purification column was centrifuged for 90 s at 15000 x g. Both collection tube and filtrate were discarded and the purification minicolumn was carefully transferred to a sterile nuclease-free 1.5 ml Eppendorf tube (it was crucial to remove all the alcohol from the minicolumn before elution).

4) Purified DNA was eluted in 50 µl of DNA Elution Buffer (DEB) and was centrifuged for 60 s at 12000 x g.

Finally, after completing the co-extraction procedure, both purified DNA and RNA extracts were stored at a -20°C freezer.

3.11.3. Degradation assessment of hexaplex of mRNA semen-specific markers

Samples were analysed in a hexaplex system following the amplification procedures previously mentioned (section 1.5 Table 6). A negative control was added without cDNA and two positive controls from initial samples of this work (S2 and S4).

4. Results and discussion

4.1. Multiplex design: mRNA marker selection and final hexaplex mRNA semen specific assay

The main aim of the present work was to develop a multiplex system containing semen specific identification mRNA markers and evaluate its suitability for forensic applications. The multiplex design consisted in several steps:

1) Marker selection: in an initial stage of this study, eight genes (SEMG1, PRM1, ODF1, ODF2, TcP11, PHF7, SMCP and TNP1) were selected as semen-specific markers for body fluid identification. SEMG1 and PRM1 genes were obtained from Juusola & Ballantyne and Haas et al. (Haas, Muheim, et al., 2009; Juusola & Ballantyne, 2005) and ODF1, ODF2, TcP11, PHF7, SMCP and TNP1 markers were selected from the work of Salzmänn et al. (Salzmänn et al., 2021). The latter genes represent six “new” mRNA candidates for semen specific body fluid identification.

2) Although eight markers were initially selected, after the primer design and single assays PCR testing, both ODF2 and PHF7, were excluded due to several reasons, but mainly due to the need of cost reduction. PHF7 was eliminated due to the fact of being the largest amplicon (allowing the size of the multiplex to be more suitable for degraded RNA). On the other hand, ODF2 was eliminated because there was already another gene candidate (ODF1) from the same gene family (outer dense fiber proteins from human sperm tail).

3) New primers were designed (except for SEMG1 and PRM1 as primers were already available for use) considering general recommendations for primer design and multiplexing requests for forensic use:

- Gene sequences were retrieved using the Genome Browser Ensemble (<https://www.ensembl.org/index.html>).
- Primer sequences were designed with Primer3 software (Koressaar & Remm, 2018) and assuring that at least one primer (within a given primer pair) spans an exon-exon junction, which is useful for limiting the amplification only to mRNA.
- Preference was given to small length amplicons (useful for degraded samples).
- Selection of similar T_m among primers (to allow one single PCR) with optimal T_m for increasing primer specificity (54°C to 65°C).

- PCR primers selected should have a length between 18 and 24 nucleotides and the GC content of a primer should be between 40% and 60%.
- Prevention of primer-dimer and hairpin formation was taken into account by using a reliable primer design tool and keeping the parameters “self-complementarity” and “self-3'-complementarity” low [Primer3 (Koressaar & Remm, 2018) and Autodimer softwares (Vallone & Butler, 2004)].
- Avoid overlap of marker size (even when labeled with different fluorochromes) to avoid stochastic effects among different dye channels as also a criteria for design.

In an initial stage of this study, eight candidate genes (SEMG1, PRM1, ODF1, ODF2, TcP11, PHF7, SMCP and TNP1) were selected as semen-specific markers for body fluid identification. An image (Figure 11) was generated with the peaks of these markers representing the expected product sizes and fluorescent labels (Table 3 in “Material and Methods”). as: SMCP with 105 bp, TcP11 with 117 bp, SEMG1 with 120 bp, ODF1 with 143 bp, PRM1 with 153 bp, ODF2 with 163 bp, TNP1 with 180 bp and PHF7 with 200 bp.

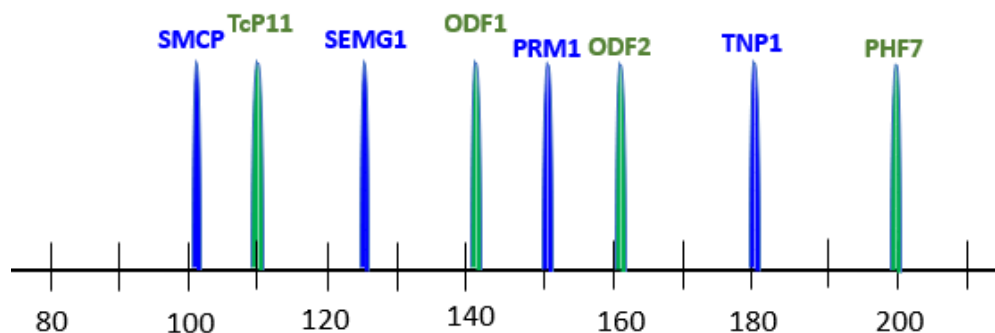


Figure 11 - Representation of the initially designed in hexaplex system with the eight selected markers. Numbers (80-180) represent the bps and peaks are located in the approximate expected amplicon sizes of primers.

A final Hexaplex mRNA marker system was projected as shown in the image below (Figure 12). The expected product sizes are represented in the image: SMCP with 105 bp, TcP11 with 117 bp, SEMG1 with 120 bp, ODF1 with 143 bp, PRM1 with 153 bp, and TNP1 with 180 bp.

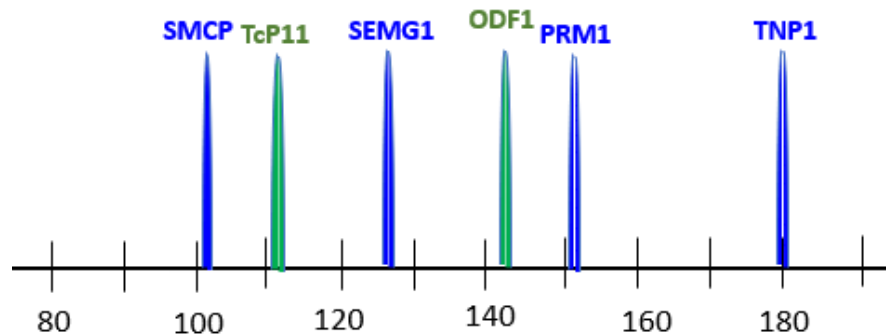


Figure 12 - Representation of the final hexaplex system designed with the six selected markers after elimination of ODF2 e PHF7. Numbers (80-180) represent the bps and peaks are located in the approximate expected amplicon sizes as designed.

4.2. Multiplex development and optimization

4.2.1. Amplification performance of mRNA markers in singleplex

To be able to assess the efficacy of the selected mRNA markers in semen samples, before the multiplex development, all semen-specific primers (even the excluded ones) were analysed by single PCR assays and detected with PAGE (methodology is described in “Materials and Methods” section). PAGE singleplex was performed in order to verify the successful individual PCR reactions and to ensure that no issues were identified, confirming the presence of the amplified amplicons and thus the successful performance of the new primers.

All initial six “new” mRNA semen specific marker candidates were tested in all 14 semen samples in singleplex PCRs. Before the exclusion of ODF2 and PHF7, all candidates were analysed. SEMG1 and PRM1 were not tested because primers for these markers were already used in a previous study (Loureiro, 2019) and therefore amplification performance had already been assessed.

In Figure 13, samples from 1 to 7 were tested with a negative control included. In Figure 13a SMCP, TcP11 and ODF1 markers are represented; it is possible to observe that all PCR products were amplified. ODF2, TNP1 and PHF7 PCR products were also amplified, however in ODF2 the majority of fragments had a weak amplification signal (Figure 13b).

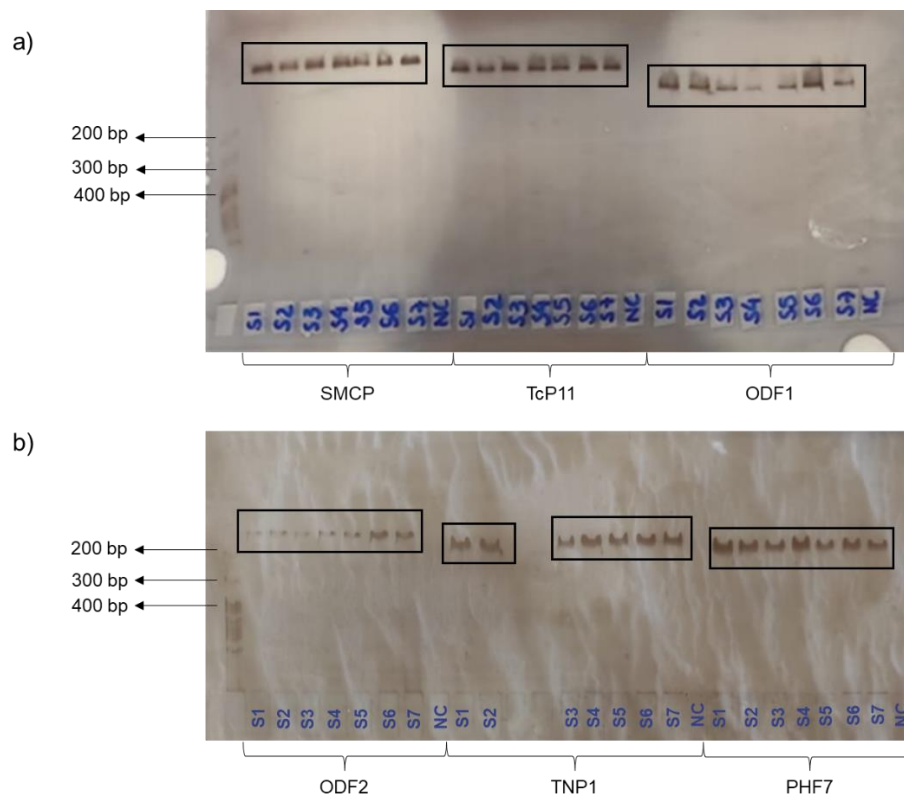


Figure 13 - Polyacrylamide gel of singleplex assays of SMCP, TcP11, ODF1, ODF2, TNP1 and PHF7 semen-specific markers for samples S1 to S7. a) SMCP, TcP11 and ODF1 b) ODF2, TNP1 and PHF7

In Figure 14, semen samples from 8 to 14 were tested also with a negative control. Both samples 10 and 11 were not amplified in SMCP as well as sample 14 in ODF1 (Figure 14a). Regarding TNP1 and PHF7 markers all PCR products amplified, but in ODF2 sample 9, 10 and 13 did not amplify, and samples 11 and 14 had a weak amplification signal (Figure 14b). Overall, all markers had a good amplification signal except for ODF2, showing a weaker signal comparing with other markers. This did not represent a problem as ODF2 was one of the two excluded markers that were tested. In addition, no unspecific products were detected, only “clean” bands were observed which supports a good design of the primers, specifically regarding specificity.

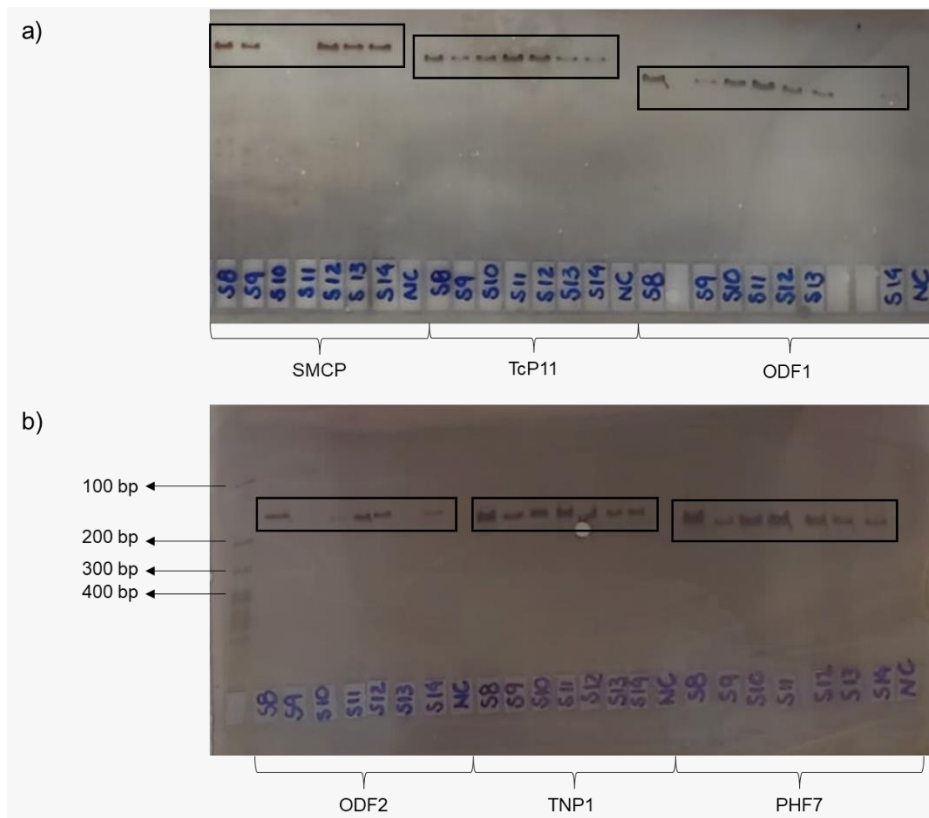


Figure 14 - Polyacrylamide gel of singleplex assays of SMCP, TcP11, ODF1, ODF2, TNP1 and PHF7 semen-specific markers for samples S8 to S14. a) SMCP, TcP11 and ODF1 b) ODF2, TNP1 and PHF7

4.2.2. Amplification performance of hexaplex mRNA system

After establishing the efficacy of the selected mRNA markers in semen samples, in singleplex assays, a final multiplex with six semen specific mRNA markers (hexaplex)- SMCP, TcP11, SEMG1, ODF1, PRM1 and TNP1 was created. As mentioned above, although eight markers were initially selected, after the primer design and single assays PCR testing, both ODF2 and PHF7, were excluded due to several reasons, but mainly due to the need for cost reduction. The hexaplex system allows the amplification in one single reaction of six semen specific markers which is very useful for forensic use as many times the amount of DNA to be tested is very low. The hexaplex mRNA system was tested in the 14 semen samples and simultaneous analysis of all markers in one single reaction was performed.

An initial analysis of the hexaplex was performed with PAGE technique with a primer mix (Primer Mix 1) developed and containing all unmarked primers (forward and reverse) as described in methodology in the “Materials and Methods” section. Comparing the PCR

amplification products, it is possible to note that the markers TcP11 (in all samples) and ODF1 (in sample 1 to 7) had a weaker amplification signal than in the singleplex as demonstrated in Figure 15. Samples 10 and 11 with SMCP, and 9, 10, 13 and 14 with ODF2 were amplified with the hexaplex but not in the singleplex reactions meaning that an error in the singleplex PCR might have occurred. In sample 7, neither TcP11 nor ODF1 amplified, as well as samples 12 and 13 (also very weak amplification of ODF1) (Figure 4).

In addition, a group of nonspecific bands were also observed, represented with green in Figure 15. Some of these nonspecific bands had a product size greater than 200 bp, not interfering with the wanted fragments since the higher wanted bp size from selected markers is 200 bp. However, it was also observed that unwanted nonspecific fragments showing as an extra band in TcP11.

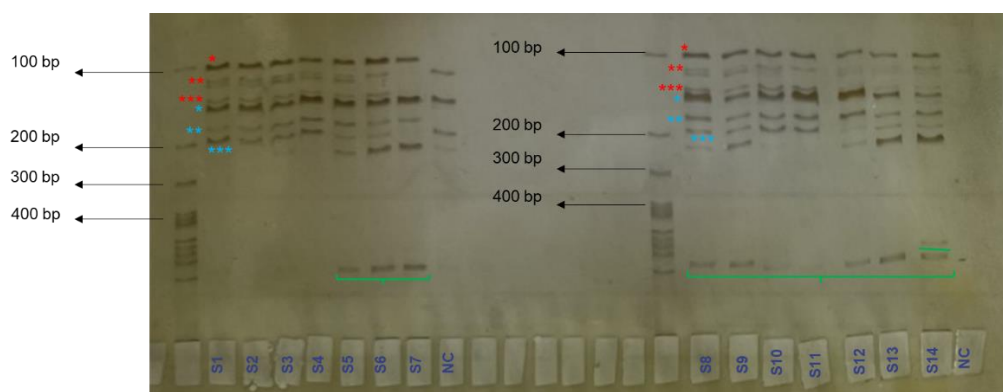


Figure 15 - Polyacrylamide gel hexaplex assay for the first attempt of multiplex development. The PCR included the selected semen-specific markers amplified for all samples (1 to 14). Label: (*) SMCP, (**) TcP11, (***) ODF1, (*) ODF2, (**) TNP1, (***) PHF7, (green) nonspecific bands.

To confirm the results obtained from the PAGE technique, an analysis by capillary electrophoresis (CE) was performed. Overall, the results obtained from bands detected in the polyacrylamide gel were also noted with the electropherograms. A table was made with a comparison of the results between both techniques (Table 12).

Table 12 - Results comparison between PAGE and CE techniques in hexaplex assay. Label: (*) weak SEMG1 amplification, (**) weak TcP11 and ODF1 amplification, (***) weak TcP11 amplification, (****) weak ODF1 amplification.

Samples	PAGE	CE
S1	All samples amplified	All samples amplified
S2	All samples amplified	All samples amplified (*)
S3	All samples amplified	All samples amplified
S4	All samples amplified (**)	All samples amplified
S5	All samples amplified (**)	All samples amplified
S6	All samples amplified	All samples amplified (*)
S7	No TcP11, ODF1 nor PHF7	No SEMG1 (**)
S8	All samples amplified	All samples amplified (***)
S9	All samples amplified (***)	All samples amplified (**)
S10	All samples amplified	All samples amplified (*)
S11	All samples amplified	No SEMG1
S12	No TcP11 nor ODF1	All samples amplified (**)
S13	No TcP11 (****)	No TcP11 (****)
S14	All samples amplified (**)	All samples amplified (**)

It is also important to note that because of the exclusion of the two markers (ODF2 and PHF7), they were not ordered with a label 5' fluorochrome and, consequently, not present in the CE assay, as well as neither SEMG1 nor PRM1 were in the PAGE assay as these had already been tested in another study as mentioned.

The electropherogram results highlighted the markers expression showing peaks in a range of approximately 100 bp to 180 bp as expected. The electropherogram of sample 1 is represented in Figure 5 as an example of the obtained electropherograms where it is possible to identify the six selected semen-specific markers and the development of the semen mRNA hexaplex system. All peaks represent the amplification of PCR products of the respective markers, except for the peak with the YelYak dye next to TcP11 that is a pull-up of SMCP due to the strong signal (Figure 16).

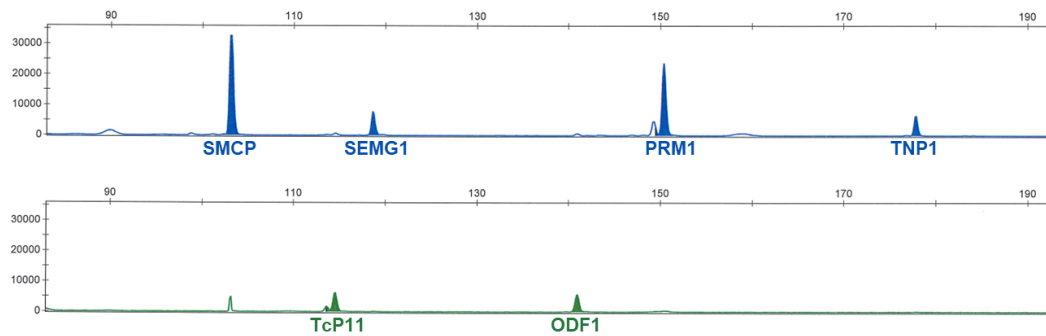


Figure 16 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 1.

The observed fragment sizes (from sample 1) were obtained using the GeneMapper ID-X version 1.4 Software (Applied Biosystems, Thermo Fisher Scientific) and are listed in Table 13, as well as the expected peak values obtained from the primer designs. Due to weak amplification signals from PCR products in TcP11 and ODF1, in both PAGE and CE assays, plus weak amplification in SEMG1 in the CE assay, an optimization of the hexaplex was needed.

Table 13 - Expected and observed product sizes for respective markers for semen identification (sample 1).

Markers	Expected product size (bp)	Observed product size (bp)
SMCP	105	103
TcP11	117	115
SEMG1	120	119
ODF1	143	141
PRM1	153	150
TNP1	180	178

4.2.3. Multiplex optimization procedures

As a result of a weak amplification signal of SEMG1, TcP11 and ODF1 in most samples, or no amplification at all, new multiplex conditions were tested in order to improve peak imbalance. In this optimization process only 3 samples were tested (sample 1 to 3).

A new primer mix (Primer Mix 2) was made with adjustments in the concentrations of the primers of SMCP and PRM1 markers. Both concentrations were reduced to half, as described in methodology of “Materials and Methods”, since peaks were too strong which could probably interfere with the amplification of other peaks. As expected, SMCP and PRM1 peaks decreased in all tested samples and SEMG1, TNP1 and ODF1 increased their amplification signal when compared with Primer Mix 1 results (Figure 17). However, PRM1 signal was too weak this time and did not even amplify in one sample (sample 3) so another concentration adjustment was made to avoid dropout of PRM1 marker.

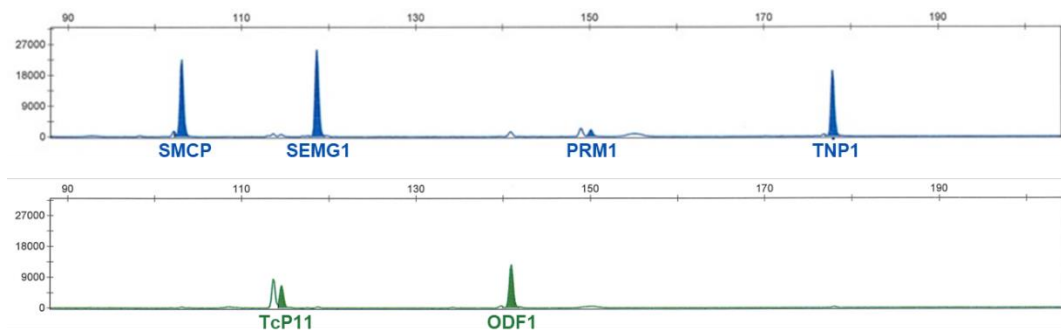


Figure 17 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 1 with Primer Mix 2. PRM1 has very weak and nearly “drops-out”.

Therefore, another primer mix (Primer Mix 3) was tested where SMCP marker remained with the same concentration as in Primer Mix 2 (half the concentration in Primer Mix 1) and PRM1 with the same concentration as used in Primer Mix 1 (initial concentration). All samples amplified and Figure 18 serves as an example with sample 1. A better balance among peaks is observed: SMCP and PRM1 had higher peaks and TcP11 and ODF1 peaks were still weak but stronger than other primer mixes tested.

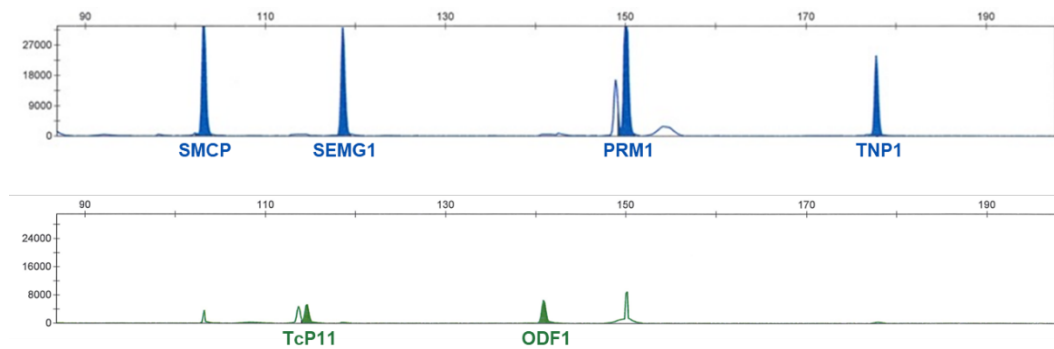


Figure 18 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 1 with Primer Mix 3.

To continue the optimization and achieve an even better balance of amplification among markers, another (final) primer mix (Primer Mix 4) was analysed with different marker concentrations as described in the “Materials and Methods” methodology. The aim this time was about lowering PRM1 concentration and increasing TcP11 and ODF1 concentrations with the same PCR conditions as previous assays (PCR 1). Besides that, additional PCR conditions were also changed: cDNA dilution with 1:20 instead of 1:10 (PCR 2) and cDNA dilution with 1:10 but lowering the PCR cycles from 35 to 32 (PCR 3). In PCR 1 all samples amplified for the exception of SEMG1 in sample 1. An expected increase of TcP11 and ODF1 was noted in all samples and TNP1 peak became too strong (Figure 19).

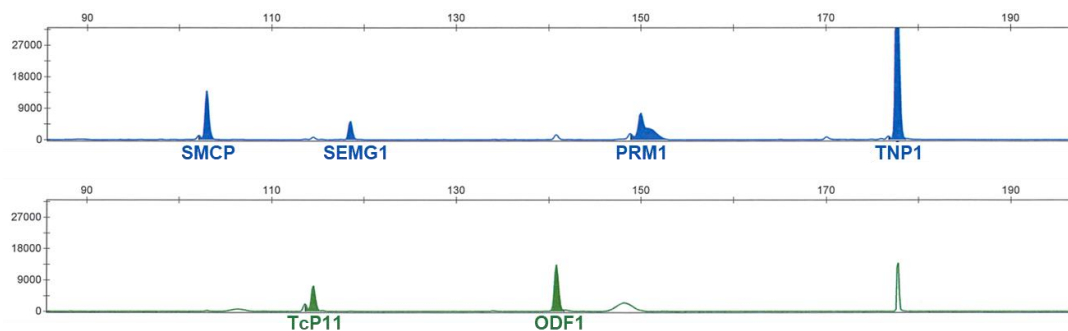


Figure 19 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 3 with Primer Mix 4 (PCR 1 conditions).

PCR 2 results (1:10 dilution instead of 1:20 of cDNA) were not expected since no amplification signal was stated probably due to too low cDNA as it shows in Figure 20.

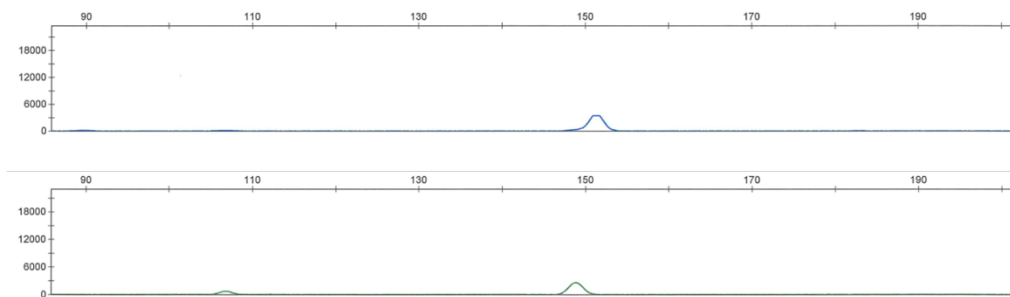


Figure 20 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 3 with Primer Mix 4 (PCR 2 conditions).

Finally, with PCR 3 conditions (32 cycles instead of 35 cycles), all samples amplified but with a lower peak of SEMG1 compared to the original PCR conditions (Figure 21).

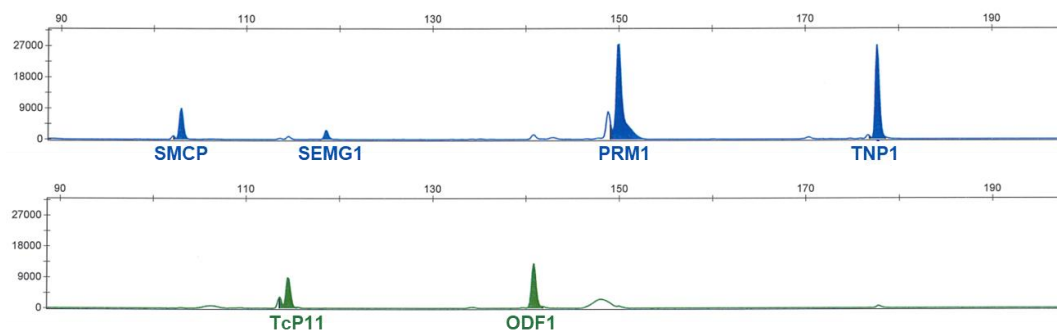


Figure 21 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 2 with Primer Mix 4 (PCR 3 conditions).

Primer Mix 1 was tested with 32 cycles because even though a weak peak of SEMG1 was observed with PCR 3 conditions, results seemed promising with Primer Mix 1. Figure 22 is the obtained electropherogram with sample 1 where it is possible to identify all selected markers peaks and it was noted in all samples used in this assay, being the selected Primer Mix and PCR conditions for further analysis (1:10 cDNA dilution and 32 cycles thermocycling conditions).

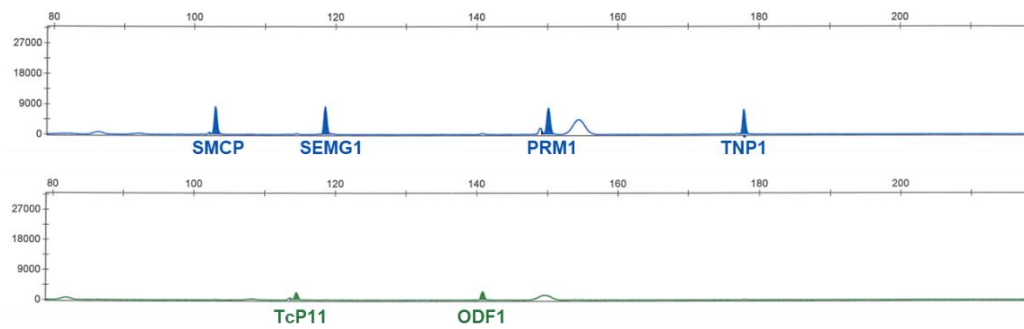


Figure 22 - Results of hexaplex assay with capillary electrophoresis. Representative electropherograms of mRNA semen-specific markers in sample 1 with Primer Mix 1 (PCR 3 conditions).

It is important to consider that used primers were fluorescein analogues. Yakima Yellow® (λ max= 524 nm) is an alternative dye to VIC® (λ max= 538 nm) dye (<https://www.idtdna.com/site/Catalog/Modifications/Product/1894>), while 6-FAM (Fluorescein) (λ max= 495 nm) is a single isomer derived of fluorescein (<https://www.idtdna.com/site/Catalog/Modifications/Product/1108>). So, all electropherograms show artifacts since alternative dyes have matrixes generated with the common dyes and causes spectral issues such as “pull-ups” (Haas et al., 2013).

4.3. Applicability of the mRNA semen-specific hexaplex in forensics

To further study the applicability and robustness of the developed semen mRNA hexaplex system in forensic settings, several features must be tested such as sensibility, specificity, and usefulness in degraded samples. Therefore, several experiments were performed to analyse the performance of the selected mRNA semen specific markers in body fluid identification in forensics.

4.3.1. Sensibility testing of the mRNA semen-specific hexaplex

The sensibility assay was needed to determine the concentration limit of a specific semen sample (i.e., until detection is possible). For that reason, different concentrations (from 5 ng/μl to 0.020 ng/μl) of one sample were tested. Sample 6 was chosen for this assay since it had a high RNA concentration rate (88.80 ng/μl) as well as good quality control parameters. Its nucleic acid concentration and absorbances are listed in Table 14. The sample was quantified five times (for a better estimation of its concentration) with the NanoDrop One microvolume UV-Vis spectrophotometer (Thermo Scientific Scientific, USA). For quality controls parameters, the A260/280 ratio is recommended at ~ 2.0 for high quality RNA, while the A260/230 is at 2.0 to 2.3 in order to be considered “pure” and if lower than those values it may indicate the presence of inhibitors or contaminants. So, Table 14 shows that the chosen sample has high quality RNA.

Table 14 - Quantification of sample 6 with respective nucleic acid concentrations and absorbances.

Sample 6	Nucleic Acid (ng/μl)	A260/280	A260/230
1	88.803	2.046	2.330
2	93.917	2.050	2.306
3	93.254	2.043	2.313
4	90.036	2.054	2.300
5	94.187	2.045	2.036
Average concentration of nucleic acid: 92.039 ng/μl			

According to the methodology in Materials and Methods (Table 10), five dilutions were made using sample 6 (D1= 5 ng/μl to D5= 0.02 ng/μl) and using as estimate for RNA concentration 92 ng/μl (average concentration of the 5 Nanodrop readings, Table 14). For the final sensibility hexaplex assay assessment, 3 replicas were performed for each dilution just to have more reliable results (Table 15).

Table 15 - Representation of the height (RFU) of each amplified marker in the electropherogram results from the sensibility testing obtained in the software GeneMapper ID-X 1.4 Software 3500 CE system.

Diluted Samples	Markers					
	SMCP	SEMG1	PRM1	TNP1	TcP11	ODF1
D1.1	32341	26731	32612	30894	30718	24041
D1.2	32153	15420	32584	32712	31408	32714
D1.3	32006	21675	32495	32711	32639	32705
D2.1	32458	7797	32588	23638	16618	29008
D2.2	-	-	-	-	-	-
D2.3	32318	9165	32545	24140	26251	27597
D3.1	32554	1904	32692	1658	1825	9156
D3.2	-	-	-	-	-	-
D3.3	32480	11914	32660	8360	-	6705
D4.1	28374	2073	17620	8675	-	8770
D4.2	32701	-	22602	8644	-	-
D4.3	23569	5122	29550	-	7380	-
D5.1	9248	-	4123	-	13393	6465
D5.2	8781	-	7001	-	1273	-
D5.3	4987	-	3821	-	-	-

Observations of these results are shown in Table 15, with peaks heights (RFU) for each marker and respective replica:

- As expected, in the first dilution (D1= 5 ng/μl), all the six selected markers were noticed with a strong peak amplification, i.e., all the semen markers amplified in all 3 replicas (D1.1, D1.2 and D1.3).
- In D2 (1.25 ng/μl) all markers also amplified, except for replica D2.2, probably due to a manual/technical error; the same applies to replica D3.2 consequently results for these two replicas were ignored (Table 15). In the two replicas that did work (D2.1 and D2.3) although all markers had strong amplification, it is possible to observe a decrease in amplification strength for SEMG1, TNP1 and TcP11 at this concentration.
- In D3 (0.31 ng/μl), peaks were still positive for all markers, but with intensity decreasing for the same markers as mentioned above: SEMG1, TNP1 and TcP11.
- TcP11 was detected in only one sample at the concentration of 0.08 ng/μl (D4) and the same observed for TNP1 and ODF1. One dropout was also observed at SEMG1 for replica D4.2. At this concentration the intensity of peaks decreased for all markers, including SMCP and PRM1.
- Finally, at D5 dilution (Table 15), SEMG1 and ODF1 were not detected in any replica, wherein only SMCP, PRM1 and TcP11 markers were identified at 0.02 ng/μl (D5) concentration. ODF1 was only detected in one of the replicas (D5.1, Table 15).

Overall, an expected decrease in peak sizes is noted in all replicas as the sample concentration also decreases, i.e., as the RNA concentration in the sample decreases, the peak signal decreased as well (Table 15). It is possible to assume that the least robust markers are the SEMG1, TNP1 and ODF1 since they were not detected at a concentration of 0.02 ng/μl (D5) and the more robust (higher sensitivity to low RNA concentration) are the SMCP and PRM1. Further studies would be needed to know the detection limit to amplify these markers (SMCP and PRM1). Nevertheless, the preliminary data supports the use of these markers, in particular, SMCP and PRM1, in forensics due to still high and positive detection at 0.02 ng/μl of RNA input.

4.3.2. Specificity testing of mRNA semen-specific hexaplex

In this study, the specificity testing was important to measure how well the markers amplify in other fluids besides semen (“cross-reactivity”), i.e., if they are highly specific to this body fluid or not. The aim is to have reliable results for the semen-specific markers when identifying body fluids and low or ideally absent amplification signals in others. Therefore, samples from vaginal secretion (VS), menstrual blood (MB), blood (B) and saliva (SAL) were tested including one sample of semen (S4) as a positive control (CP).

NOTE: RNA samples from blood and saliva were already available from the previous study of Loureiro, S., 2019 and were tested. However, because there were no amplification signals, new samples were extracted to exclude problems with the older samples. Therefore, two samples from both Saliva (SAL1 and SAL2) and Blood (B1 and B2) were collected and extracted in the lab using the same protocol as described in 3.11.2. in “Materials and Methods”. SAL1, SAL2, B1, B2, VS and MB samples were submitted to the NanoDrop One microvolume UV-Vis spectrophotometer quantification in order to measure the RNA concentration of each sample (Table 16). Because of a too low sample concentration and bad quality parameters, SAL1, B1 and B2 were excluded from this assay. Therefore, blood was not included in the specificity testing since the kit used for the RNA extraction probably doesn't work with blood, so a specific kit for blood extraction would be needed and no financial resources were available. Therefore, cross reactivity was tested for one saliva sample (SAL2), one menstrual blood (MB) and one vaginal secretion (VS).

Table 16 - Quantification results of the different body fluids with respective nucleic acid concentrations and absorbances. Grey columns highlight the tested samples.

Samples	Nucleic Acid (ng/μl)	A260/280	A260/230
SAL1	2.5	0.9	0.19
SAL2	42.0	2.0	1.31
B1	0.4	0.28	0.02
B2	1.0	0.57	0.01
MB	117.1	2.07	0.19
VS	16.3	2.02	0.46

As expected in the S4 sample, all six semen-specific markers amplified with very strong peaks (in fact excess sample profiles were observed). In all other samples of the different tested body fluids no amplification was noticed, meaning that the most likely markers tested are in fact only specific to semen (Figure 23). Then, when analysing a sample with a mixture of body fluids, these markers will only identify semen in that sample. However, more samples of the same body fluid type from different individuals need to be tested to support these preliminary results.

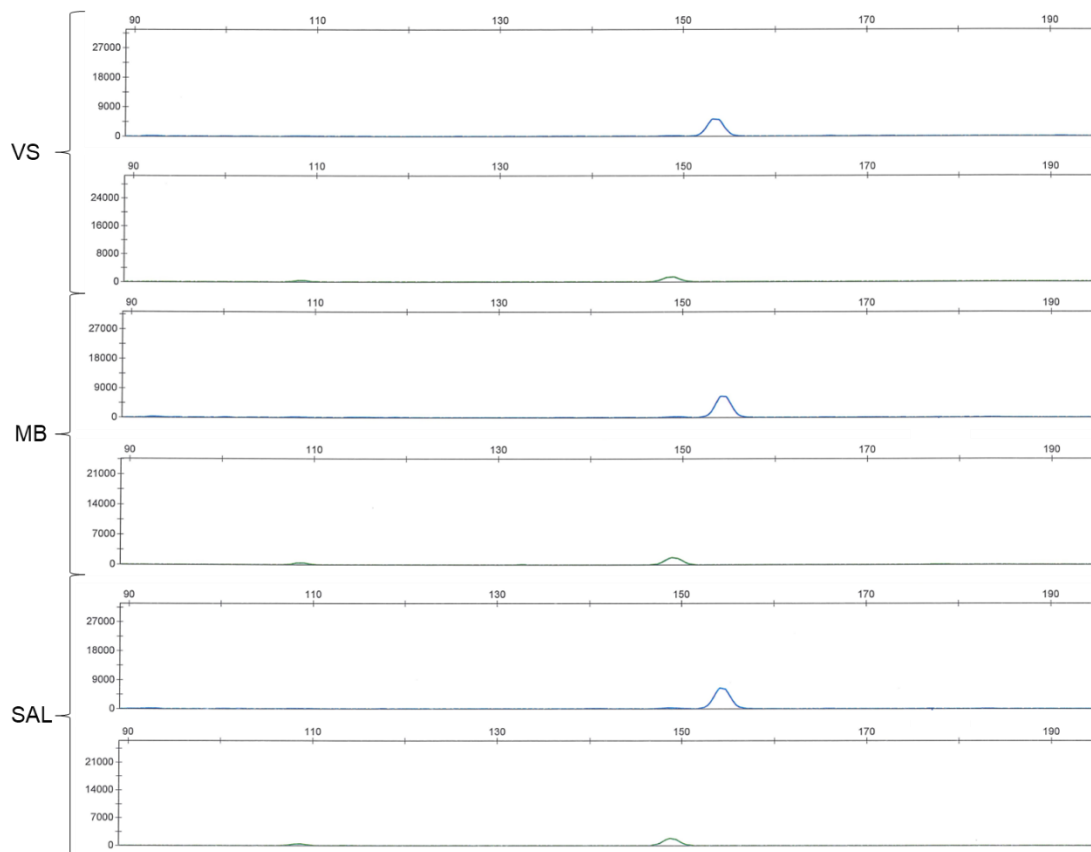


Figure 23 - Specificity results of the VS, MB, and SAL samples.

4.3.3. Degradation assessment of mRNA semen-specific hexaplex

An additional small study was performed with the six semen-specific selected markers using six samples collected with swabs from the “Instituto Nacional de Medicina Legal e Ciências Forenses- INMLCF- Delegação do Norte”. The aim was not only to confirm the selected markers but also to assess the degradation of samples over time wherein the collection of Sample 1 was 502 days (~17 months) and is 40 days apart from the collection of Sample 6 (~18 months). Time of degradation corresponds to the time since samples were collected to the day they were extracted (Table 10 in M&M).

In this study, it was possible to observe how time affected the semen samples so a decrease of size peaks was expected since samples would be more degraded. However, that was not seen with all the selected markers as all markers were detected even after 542 days (~18 months) (Table 17).

Table 17 - Representation of the height of each amplified marker in the electropherogram results from the degradation assay with samples from Instituto de Medicina Legal do Porto.

Samples	Markers					
	SMCP	SEMG1	PRM1	TNP1	TcP11	ODF1
ML-S1*	32440	1200	32403	19533	2626	7186
ML-S2	32439	7372	31826	10051	4273	11362
ML-S3	2280	-	-	-	-	-
ML-S4	32395	2853	31851	12459	5231	13859
ML-S5	32581	835	32447	-	1973	1846
ML-S6	23062	-	32420	214	1937	5591

*ML-S1: Medicina Legal Sample 1

ML-S3 only showed one peak at SMCP, that most likely is due to a technical error (sample collection, RNA extraction or PCR) because regarding “degradation time” it stands in the middle of the other tested samples. Similar to the sensibility testing SMCP and PRM1 also revealed to be the most “robust” markers over time as amplification was always present with “strong” peaks (except sample ML-S3). SEMG1 seems to be the most susceptible marker to degradation as low intensity peaks were observed and a dropout was seen in one sample (ML-S6). TNP1 semen specific marker also dropped out in sample ML-S5.

The degradation results of this (very) preliminary study support the use of the mRNA markers in forensics due to the degraded conditions of samples often found in sexual assault cases as in general in other forensic scenarios. Other studies have confirmed the reliability of both PRM1 and SEMG1 for the identification of semen stains with reduced sample quantities and in old samples (Haas, Klessner, et al., 2009; Sirker et al., 2013). On the contrary to the instability for which RNA is known, it has been demonstrated that in old dried stains (10 to 20 years old) it is possible to detect some levels of several transcripts allowing the approach of forensic when involving stains of unknown age (at least under certain environmental conditions) (Fu & Allen, 2019).

5. Conclusions

The study on mRNA semen-specific markers developed during this project has provided new data contributing this way to the field of body fluid and tissue identification in forensic genetics. The main conclusions are summarized in the following points:

1) A PCR multiplex was successfully designed based on the expected product sizes from each selected primer for each mRNA marker and that aim to be specific for semen identification in forensic samples.

2) The (final) created hexaplex mRNA system was tested first with the PAGE method (without SEMG1 and PRM1) and showed immediate promising results since all markers amplified for the majority of samples (11 semen RNA samples out of the 14 total). To confirm these results, an analysis with CE was performed. Results obtained from the CE method also showed that the six selected markers amplified in all samples, except SEMG1 in both sample 7 and 11 and TcP11 in sample 13. A reasonable balance among peaks of the six markers was already observed with the first primer mix developed. Both methods for fragment separation and detection (PAGE and CE) shared similar results. However, CE was the method chosen to continue the multiplex optimization as this is the same technique used for DNA profiling allowing easiness in the implementation of the multiplex in forensic laboratories.

3) Several conditions and concentrations were tested in the developed multiplex in order to optimize it. Nevertheless, Primer Mix 1 was the final selected mix and with 32 cycles in the thermocycling PCR condition, since it was the result with a more balanced marker peak showing the amplification of all markers in all tested samples.

4) Sensitivity, specificity and degradation testing were performed to evaluate if the mRNA markers selected are robust and sensitive enough for the forensic field. Forensic samples are usually low in quantity, degraded and with mixtures of different fluids.

Sensitivity results have shown that SMCP and PRM1 were the more robust markers of this study since they still amplified in a sample concentration of 0.02 ng/μl.

The results about specificity obtained have shown that the six markers, selected in this work, are in fact specific for semen since they did not amplify in any sample of vaginal secretions, menstrual blood nor saliva, which could be useful in cases of mixtures.

5) The six mRNA markers were also tested in RNA extracted from older samples, with ≥17 months old, that were collected at different time points to access the performance of markers along a possible degradation scenario. This small “degradation”

study reinforced the potential use in forensics, since all primers amplified in the last time point of sample collection, 542 days (almost 18 months), except for SEMG1.

This study has generated only preliminary data on the applicability in forensics of SMCP, SEMG1, PRM1, TNP1, TcP11 and ODF1, in particular, on what regards sensitivity, specificity and degradation analyses. Therefore, further studies are necessary since the obtained multiplex must be further evaluated as co-expression in other tissues is not fully excluded and much older samples should be included. To conclude, the mRNA semen specific markers evaluated in the present study seem to be reliable and stable enough for the identification of body fluid and tissue in forensic casework.

6. Future perspectives

For future work, it would also be important to test the specificity more in-depth of the present semen markers: more samples from different fluids and test other fluids (peripheral blood, for example). In cases of sexual violence, it is important to differ menstrual blood from peripheral blood since this last one may indicate an injury.

It would be important to have financial support in order to study the excluded primers in this work, both ODF2 and PHF7, to provide additional results that could match the other mRNA markers mentioned in this work, as well as more tested samples.

It would also be crucial to include markers for other body fluids, in a final multiplex, with sexual context that could be used specifically in mixtures from sexual assault cases such as vaginal secretions, saliva, menstrual blood and peripheral blood.

Body fluid and tissue identification using molecular markers such as mRNA is a still field in forensic genetics with many possibilities for further future development and applications.

7. References

- Abu-Halima, M., Becker, L. S., Al Smadi, M. A., Abdul-Khaliq, H., Raeschle, M., & Meese, E. (2023). Sperm Motility Annotated Genes: Are They Associated with Impaired Fecundity? *Cells*, 12(9), 1239. <https://doi.org/10.3390/cells12091239>
- Adeoya-Osiguwa, S. A., Dudley, R. K., Hosseini, R., & Fraser, L. R. (1998). FPP modulates mammalian sperm function via TCP-11 and the adenylyl cyclase/cAMP pathway. *Molecular Reproduction and Development*, 51(4), 468–476. [https://doi.org/10.1002/\(SICI\)1098-2795\(199812\)51:4<468::AID-MRD14>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1098-2795(199812)51:4<468::AID-MRD14>3.0.CO;2-6)
- Afolabi, O. A., Roeder, A. D., Iyengar, A., & Hadi, S. (2017). Evaluation of genetic markers for forensic identification of human body fluids. *Forensic Science International: Genetics Supplement Series*, 6, e241–e243. <https://doi.org/10.1016/j.fsigss.2017.09.098>
- Albani, P. P., & Fleming, R. (2018). Novel messenger RNAs for body fluid identification. *Science & Justice*, 58(2), 145–152. <https://doi.org/10.1016/j.scijus.2017.09.002>
- Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalki, E., & Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *The EMBO Journal*, 5399–5410.
- Alshehhi, S., & Haddrill, P. R. (2019). Estimating time since deposition using quantification of RNA degradation in body fluid-specific markers. *Forensic Science International*, 298, 58–63. <https://doi.org/10.1016/j.forsciint.2019.02.046>
- Alvarez, M., Juusola, J., & Ballantyne, J. (2004). An mRNA and DNA co-isolation method for forensic casework samples. *Analytical Biochemistry*, 335(2), 289–298. <https://doi.org/10.1016/j.ab.2004.09.002>
- An, J.-H., Shin, K.-J., Yang, W.-I., & Lee, H.-Y. (2012). Body fluid identification in forensics. *BMB Reports*, 45(10), 545–553. <https://doi.org/10.5483/BMBRep.2012.45.10.206>
- Ando, M., Yamaguchi, K., Chiba, K., Miyake, H., & Fujisawa, M. (2012). Expression of VASA mRNA in Testis as a Significant Predictor of Sperm Recovery by Microdissection Testicular Sperm Extraction in Patient With Nonobstructive Azoospermia. *Journal of Andrology*, 33(4), 711–716. <https://doi.org/10.2164/jandrol.110.012278>

APAV. (2022). *Estatísticas APAV: Relatório Anual 2022*.

Aregger, M., & Cowling, V. H. (2016). RNA Biology Regulation of mRNA capping in the cell cycle Regulation of mRNA capping in the cell cycle. <https://doi.org/10.1080/15476286.2016.1251540>

Aydos, O. S. E., Hekmatshoar, Y., Altınok, B., Özkan, T., Şakirağaoğlu, O., Karadağ, A., Kaplan, F., Ilgaz, S., Taşpınar, M., Yükselen, I., Sunguroğlu, A., & Aydos, K. (2018). Genetic Polymorphisms in PRM1 , PRM2 , and YBX2 Genes are Associated with Male Factor Infertility. *Genetic Testing and Molecular Biomarkers*, 22(1), 55–61. <https://doi.org/10.1089/gtmb.2017.0040>

Bachellerie, J.-P., Cavaille, J., & Huttenhofer, A. (2002). The expanding snoRNA world. *Biochimie*, 775–790.

Bai, P., Deng, W., Wang, L., Long, B., Liu, K., Liang, W., & Zhang, L. (2013). Micro RNA profiling for the detection and differentiation of body fluids in forensic stain analysis. *Forensic Science International: Genetics Supplement Series*, 4, 216–217. <https://doi.org/10.1016/j.fsigss.2013.10.111>

Balhorn, R. (2007). The protamine family of sperm nuclear proteins. *Genome Biology*, 8(9), 227. <https://doi.org/10.1186/gb-2007-8-9-227>

Bamberg, M., Bruder, M., Dierig, L., Kunz, S. N., Schwender, M., & Wiegand, P. (2022). Best of both: A simultaneous analysis of mRNA and miRNA markers for body fluid identification. *Forensic Science International: Genetics*, 59, 102707. <https://doi.org/10.1016/j.fsigen.2022.102707>

Bao, J., & Bedford, M. T. (2016). Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. *Reproduction*, 151(5), R55–R70. <https://doi.org/10.1530/REP-15-0562>

Barton, A. J. L., Pearson, R. C. A., Najlerahim, A., & Harrison, P. J. (1993). Pre-and Postmortem Influences on Brain RNA. *Journal of Neurochemistry*, 61(1), 1–11. <https://doi.org/10.1111/j.1471-4159.1993.tb03532.x>

Bauer, M. (2007). RNA in forensic science. *Forensic Science International: Genetics*, 1(1), 69–74. <https://doi.org/10.1016/j.fsigen.2006.11.002>

Bauer, M., & Patzelt, D. (2003). Protamine mRNA as molecular marker for spermatozoa in semen stains. *International Journal of Legal Medicine*, 117(3), 175–179. <https://doi.org/10.1007/s00414-002-0347-2>

Bauer, M., Polzin, S., & Patzelt, D. (2003). Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of bloodstains? *Forensic Science International*, 138(1–3), 94–103. <https://doi.org/10.1016/j.forsciint.2003.09.008>

Ben-Shem, A., De Loubresse, N. G., Melnikov, S., Jenner, L., Yusupova, G., & Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, 334(6062), 1524–1529. <https://doi.org/10.1126/SCIENCE.1212642>

Bhansali, R. (2021). A Transcriptomically Guided Computational Quantification of the SRR8671434 RNA Sequence A Transcriptomically Guided Computational Quantification of the SRR8671434 RNA Sequence View project Determining the Correlation Between the TP53 Gene and Kidney Cancer Survival Through the Cox Proportional-Hazards Regression Model View project A TRANSCRIPTOMICALLY GUIDED COMPUTATIONAL QUANTIFICATION OF THE SRR8671434 RNA SEQUENCE A PREPRINT-FEBRUARY 2, 2021. <https://doi.org/10.6084/m9.figshare.13697431>

Blackman, S., Stafford-Allen, B., Hanson, E. K., Panasiuk, M., Brooker, A.-L., Rendell, P., Ballantyne, J., & Wells, S. (2018). Developmental validation of the ParaDNA® Body Fluid ID System—A rapid multiplex mRNA-profiling system for the forensic identification of body fluids. *Forensic Science International: Genetics*, 37, 151–161. <https://doi.org/10.1016/j.fsigen.2018.08.012>

Bowden, A., Fleming, R., & Harbison, S. (2011). A method for DNA and RNA co-extraction for use on forensic samples using the Promega DNA IQ™ system. *Forensic Science International: Genetics*, 5(1), 64–68. <https://doi.org/10.1016/j.fsigen.2009.11.007>

Brosius, J., & Raabe, C. A. (2016). What is an RNA? A top layer for RNA classification. *RNA Biology*.

Budowle, B., Onorato, A. J., Callaghan, T. F., Manna, A. Della, Gross, A. M., Guerrieri, R. A., Luttman, J. C., & McClure, D. L. (2009). Mixture Interpretation: Defining the Relevant Features for Guidelines for the Assessment of Mixed DNA Profiles in Forensic Casework. *Journal of Forensic Sciences*, 54(4), 810–821. <https://doi.org/10.1111/j.1556-4029.2009.01046.x>

Campbell, R., & Fehler-Cabral, G. (2022). “Just Bring Us the Real Ones”: The Role of Forensic Crime Laboratories in Guarding the Gateway to Justice for Sexual Assault Victims. *Journal of Interpersonal Violence*, 37(7–8), NP3675–NP3702. <https://doi.org/10.1177/0886260520951303>

Castaneda, J. M., Miyata, H., Archambeault, D. R., Satouh, Y., Yu, Z., Ikawa, M., & Matzuk, M. M. (2020). Mouse t-complex protein 11 is important for progressive motility in sperm†. *Biology of Reproduction*, 102(4), 852–862. <https://doi.org/10.1093/biolre/ioz226>

Chalbatani, G. M., Dana, H., Memari, F., Gharagozlou, E., Ashjaei, S., Kheirandish, P., Marmari, V., Mahmoudzadeh, H., Mozayani, F., Maleki, A. R., Sadeghian, E., Nia, E. Z., Miri, S. R., Nia, N. zainali, Rezaeian, O., Eskandary, A., Razavi, N., Shirkhoda, M., & Rouzbahani, F. N. (2019). Biological function and molecular mechanism of piRNA in cancer. *Practical Laboratory Medicine*, 13. <https://doi.org/10.1016/J.PLABM.2018.E00113>

Clancy, S., & Brown, W. (2008). Translation: DNA to mRNA to Protein . Nature Education . https://www.fcav.unesp.br/Home/departamentos/tecnologia/marcostuliooliveira/translati on_-dna-to-mrna-to-protein-_-learn-science-at-scitable.pdf

Coding and Non-Coding RNA. (2023). <https://www.bio-rad.com/en-pt/applications-technologies/coding-non-coding-rna?ID=Q1070M70KWE7>

Copley, S. D., Smith, E., & Morowitz, H. J. (2007). The origin of the RNA world: Co-evolution of genes and metabolism. *Bioorganic Chemistry*, 35(6), 430–443. <https://doi.org/10.1016/J.BIOORG.2007.08.001>

Cortellini, V., Brescia, G., Cerri, N., & Verzeletti, A. (2020). Simultaneous DNA and RNA profiling in a case of sexual assault in a 3-year-old child: Forensic genetics solves the crime. *Legal Medicine*, 47, 101727. <https://doi.org/10.1016/j.legalmed.2020.101727>

Dahariya, S., Paddibhatla, I., Kumar, S., Raghuwanshi, S., Palapati, A., & Kumar Gutti, R. (2019). Long non-coding RNA: Classification, biogenesis and functions in blood cells. <https://doi.org/10.1016/j.molimm.2019.04.011>

de Kretser, D. M., Loveland, K. L., Meinhardt, A., Simorangkir, D., & Wreford, N. (1998). Spermatogenesis. *Human Reproduction*, 13(suppl 1), 1–8. https://doi.org/10.1093/humrep/13.suppl_1.1

Deniz, E., & Erman, B. (2017). Long noncoding RNA (lincRNA), a new paradigm in gene expression control. *Functional & Integrative Genomics*, 17(2–3), 135–143. <https://doi.org/10.1007/s10142-016-0524-x>

DeVore, H. K., & Sachs, C. J. (2011). Sexual Assault. *Emergency Medicine Clinics of North America*, 29(3), 605–620. <https://doi.org/10.1016/j.emc.2011.04.012>

Dørum, G., Ingold, S., Hanson, E., Ballantyne, J., Snipen, L., & Haas, C. (2018). Predicting the origin of stains from next generation sequencing mRNA data. *Forensic Science International: Genetics*, 34, 37–48. <https://doi.org/10.1016/j.fsigen.2018.01.001>

Ehrlich, R., Davyt, M., López, I., Chalar, C., & Marín, M. (2021). On the Track of the Missing tRNA Genes: A Source of Non-Canonical Functions? *Frontiers in Molecular Biosciences*, 8. <https://doi.org/10.3389/fmolb.2021.643701>

Enwerem, I. I., Velma, V., Broome, H. J., Kuna, M., Begum, R. A., & Hebert, M. D. (2014). Coilin association with Box C/D scaRNA suggests a direct role for the Cajal body marker protein in scaRNP biogenesis. *Biology Open*, 3(4), 240–249. <https://doi.org/10.1242/bio.20147443>

Faghihi, M. A., & Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. *Nature Reviews Molecular Cell Biology*, 10(9), 637–643.

Fasken, M. B., & Corbett, A. H. (2005). Process or perish: quality control in mRNA biogenesis. *Nature Structural & Molecular Biology*, 12(6), 482–488. <https://doi.org/10.1038/nsmb945>

Finger, J. M., Mercer, J. F. B., Cotton, R. G. H., & Danks, D. M. (1987). Stability of protein and mRNA in human postmortem liver-analysis by two-dimensional gel electrophoresis. *Clinica Chimica Acta*, 170(2–3), 209–218. [https://doi.org/10.1016/0009-8981\(87\)90130-6](https://doi.org/10.1016/0009-8981(87)90130-6)

Firdaus-Raih, M., Hamdani, H. Y., Nadzirin, N., Ramlan, E. I., Willett, P., & Artymiuk, P. J. (2014). COGNAC: a web server for searching and annotating hydrogen-bonded base interactions in RNA three-dimensional structures. *Nucleic Acids Research*, 42. <https://doi.org/10.1093/nar/gku438>

Fleming, R. I., & Harbison, S. (2010). The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. *Forensic Science International: Genetics*, 4(4), 244–256. <https://doi.org/10.1016/j.fsigen.2009.10.006>

Fraser, L. R., & Dudley, K. (1999). New insights into the t-complex and control of sperm function. *BioEssays*, 21(4), 304–312. [https://doi.org/10.1002/\(SICI\)1521-1878\(199904\)21:4<304::AID-BIES6>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1521-1878(199904)21:4<304::AID-BIES6>3.0.CO;2-H)

Fraser, L. R., Hosseini, R., Hanyalogou, A., Talmor, A., & Dudley, R. K. (1997). TCP-11, the product of a mouset-complex gene, plays a role in stimulation of capacitation and inhibition of the spontaneous acrosome reaction. *Molecular Reproduction and Development*, 48(3), 375–382. [https://doi.org/10.1002/\(SICI\)1098-2795\(199711\)48:3<375::AID-MRD11>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1098-2795(199711)48:3<375::AID-MRD11>3.0.CO;2-V)

Fu, J., & Allen, R. W. (2019). A method to estimate the age of bloodstains using quantitative PCR. *Forensic Science International: Genetics*, 39, 103–108. <https://doi.org/10.1016/j.fsigen.2018.12.004>

Fujimoto, S., Manabe, S., Morimoto, C., Hamano, Y., & Tamaki, K. (2017). Effect of the absence of spermatozoa on microrna-based semen identification. *Forensic Science International: Genetics Supplement Series*, 6, e238–e240. <https://doi.org/10.1016/j.fsigss.2017.09.099>

Garcia, A., Dunoyer-Geindre, S., Fish, R. J., Neerman-Arbez, M., Reny, J.-L., & Fontana, P. (2021). Methods to Investigate miRNA Function: Focus on Platelet Reactivity. *Thrombosis and Haemostasis*, 409–421.

Ge, L.-Q., Xia, T., Huang, B., Gu, H.-T., Song, Q.-S., Yang, G.-Q., Liu, F., & Wu, J.-C. (2017). PHF7, a novel male gene influences female fecundity and population growth in *Nilaparvata lugens* Stål (Hemiptera: Delphacidae). *Scientific Reports*, 7(1), 11611. <https://doi.org/10.1038/s41598-017-11524-2>

Geisler, S., & Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. <https://doi.org/10.1038/nrm3679>

Ghafouri-Fard, S., Ousati Ashtiani, Z., Sabah Golian, B., Hasheminasab, S.-M., & Modarressi, M. H. (2010). Expression of Two Testis-specific Genes, SPATA19 and LEMD1, in Prostate Cancer. *Archives of Medical Research*, 41(3), 195–200. <https://doi.org/10.1016/j.arcmed.2010.04.003>

Glynn, C. L. (2020). Potential applications of microRNA profiling to forensic investigations. *RNA*, 26, 1–9. <https://doi.org/10.1261/rna>

Gomes, I., Strohbücker, B., Rothschild, M. A., & Schneider, P. M. (2013). Evaluation of mRNA specific markers using a pentaplex system for the identification of skin and saliva from contact trace evidence. *Forensic Science International: Genetics Supplement Series*, 4(1), e180–e181. <https://doi.org/10.1016/j.fsigss.2013.10.093>

Green, M. R. (1986). PRE-mRNA SPLICING. *Annual Review of Genetics*, 20(1), 671–708. <https://doi.org/10.1146/annurev.ge.20.120186.003323>

Guy, M. P., Young, D. L., Payea, M. J., Zhang, X., Kon, Y., Dean, K. M., Grayhack, E. J., Mathews, D. H., Fields, S., & Phizicky, E. M. (2014). Identification of the determinants of tRNA function and susceptibility to rapid tRNA decay by high-throughput in vivo analysis. *Genes & Development*, 28(15), 1721–1732. <https://doi.org/10.1101/gad.245936.114>

Haas, C., Hanson, E., Anjos, M. J., Banemann, R., Berti, A., Borges, E., Carracedo, A., Carvalho, M., Courts, C., De Cock, G., Dötsch, M., Flynn, S., Gomes, I., Hollard, C., Hjort, B., Hoff-Olsen, P., Hríbková, K., Lindenbergh, A., Ludes, B., ... Ballantyne, J. (2013). RNA/DNA co-analysis from human saliva and semen stains – Results of a third collaborative EDNAP exercise. *Forensic Science International: Genetics*, 7(2), 230–239. <https://doi.org/10.1016/j.fsigen.2012.10.011>

Haas, C., Hanson, E., Anjos, M. J., Bär, W., Banemann, R., Berti, A., Borges, E., Bouakaze, C., Carracedo, A., Carvalho, M., Castella, V., Choma, A., De Cock, G., Dötsch, M., Hoff-Olsen, P., Johansen, P., Kohlmeier, F., Lindenbergh, P. A., Ludes, B., ... Ballantyne, J. (2012). RNA/DNA co-analysis from blood stains—Results of a second collaborative EDNAP exercise. *Forensic Science International: Genetics*, 6(1), 70–80. <https://doi.org/10.1016/j.fsigen.2011.02.004>

Haas, C., Hanson, E., Bär, W., Banemann, R., Bento, A. M., Berti, A., Borges, E., Bouakaze, C., Carracedo, A., Carvalho, M., Choma, A., Dötsch, M., Durianciková, M., Hoff-Olsen, P., Hohoff, C., Johansen, P., Lindenberg, P. A., Loddenkötter, B., Ludes, B., ... Ballantyne, J. (2011). mRNA profiling for the identification of blood—Results of a collaborative EDNAP exercise. *Forensic Science International: Genetics*, 5(1), 21–26. <https://doi.org/10.1016/j.fsigen.2010.01.003>

Haas, C., Hanson, E., Morling, N., & Ballantyne, J. (2011). Collaborative EDNAP exercises on messenger RNA/DNA co-analysis for body fluid identification (blood, saliva, semen) and STR profiling. *Forensic Science International: Genetics Supplement Series*, 3(1), e5–e6. <https://doi.org/10.1016/j.fsigss.2011.08.002>

Haas, C., Klessner, B., Maake, C., Bär, W., & Kratzer, A. (2009). mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. *Forensic Science International: Genetics*, 3(2), 80–88. <https://doi.org/10.1016/j.fsigen.2008.11.003>

Haas, C., Muheim, C., Kratzer, A., Bär, W., & Maake, C. (2009). mRNA profiling for the identification of sperm and seminal plasma. *Forensic Science International: Genetics Supplement Series*, 2, 534–535. <https://doi.org/10.1016/j.fsigss.2009.08.109>

Hanson, E., Ingold, S., Haas, C., & Ballantyne, J. (2018). Messenger RNA biomarker signatures for forensic body fluid identification revealed by targeted RNA sequencing. *Forensic Science International: Genetics*, 34, 206–221. <https://doi.org/10.1016/j.fsigen.2018.02.020>

Hanson, E. K., & Ballantyne, J. (2013). Highly specific mRNA biomarkers for the identification of vaginal secretions in sexual assault investigations. *Science & Justice*, 53(1), 14–22. <https://doi.org/10.1016/j.scijus.2012.03.007>

Hanson, E. K., Lubenow, H., & Ballantyne, J. (2009). Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. *Analytical Biochemistry*, 387(2), 303–314. <https://doi.org/10.1016/j.ab.2009.01.037>

Harbison, S., & Fleming, R. (2016). Forensic body fluid identification: state of the art. *Research and Reports in Forensic Medical Science*, 11. <https://doi.org/10.2147/RRFMS.S57994>

Hashemitabar, M., Sabbagh, S., Orazizadeh, M., Ghadiri, A., & Bahmanzadeh, M. (2015). A proteomic analysis on human sperm tail: comparison between normozoospermia and asthenozoospermia. *Journal of Assisted Reproduction and Genetics*, 32(6), 853–863. <https://doi.org/10.1007/s10815-015-0465-7>

Heidaran, M. A., Showman, R. M., & Kistler, W. S. (1988). A cytochemical study of the transcriptional and translational regulation of nuclear transition protein 1 (TP1), a major chromosomal protein of mammalian spermatids. *The Journal of Cell Biology*, 106(5), 1427–1433. <https://doi.org/10.1083/jcb.106.5.1427>

Heidarana, M. A., Kozak, C. A., & Kistler, W. S. (1989). Nucleotide sequence of the Stp-1 gene coding for rat spermatid nuclear transition protein 1 (TP1): homology with protamine P1 and assignment of the mouse Stp-1 gene to chromosome 1. *Gene*, 75(1), 39–46. [https://doi.org/10.1016/0378-1119\(89\)90381-8](https://doi.org/10.1016/0378-1119(89)90381-8)

Hering, D. M., Oleński, K., Ruś, A., & Kaminski, S. (2014). Genome-wide association study for semen volume and total number of sperm in Holstein-Friesian bulls. *Animal Reproduction Science*, 151(3–4), 126–130. <https://doi.org/10.1016/j.anireprosci.2014.10.022>

Herr, J. C., Thomas, D., Bush, L. A., Coonrod, S., Khole, V., Howards, S. S., & Flickinger, C. J. (1999). Sperm Mitochondria-Associated Cysteine-Rich Protein (SMCP) Is an Autoantigen in Lewis Rats¹. *Biology of Reproduction*, 61(2), 428–435. <https://doi.org/10.1095/biolreprod61.2.428>

Hirenallur Maheshwarappa, Y., Kumar, S., Chaudhary, R., Mishra, C., Ayyar, S., Kumar, A., Chauhan, A., Ghosh, S. K., Panigrahi, M., & Mitra, A. (2019). Identification of sperm motility markers in bovine transition protein genes. *Reproduction in Domestic Animals*, 54(2), 365–372. <https://doi.org/10.1111/rda.13369>

Holley, C. L., & Topkara, V. K. (2011). An Introduction to Small Non-coding RNAs: miRNA and snoRNA. *Cardiovascular Drugs and Therapy*, 25(2), 151–159. <https://doi.org/10.1007/s10557-011-6290-z>

Hornos, J. E. M., & Hornos, Y. M. M. (1993). Algebraic model for the evolution of the genetic code. *Physical Review Letters*, 71(26), 4401–4404. <https://doi.org/10.1103/PHYSREVLETT.71.4401>

Ingold, S., Dørum, G., Hanson, E., Ballantyne, J., & Haas, C. (2020). Assigning forensic body fluids to donors in mixed body fluids by targeted RNA/DNA deep sequencing of coding region SNPs. *International Journal of Legal Medicine*, 134(2), 473–485. <https://doi.org/10.1007/s00414-020-02252-w>

Ingold, S., Dørum, G., Hanson, E., Berti, A., Branicki, W., Brito, P., Elsmore, P., Gettings, K. B., Giangasparo, F., Gross, T. E., Hansen, S., Hanssen, E. N., Kampmann, M.-L., Kayser, M., Laurent, F.-X., Morling, N., Mosquera-Miguel, A., Parson, W., Phillips, C., ... Haas, C. (2018). Body fluid identification using a targeted mRNA massively parallel sequencing approach – results of a EUROFORGEN/EDNAP collaborative exercise. *Forensic Science International: Genetics*, 34, 105–115. <https://doi.org/10.1016/j.fsigen.2018.01.002>

Jackson, A. L., & Linsley, P. S. (2010). Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nature Reviews Drug Discovery*, 9(1), 57–67.

Jarroux, J., Morillon, A., & Pinskaya, M. (2017). History, Discovery, and Classification of lncRNAs . *Long Non Coding RNA Biology*, 1–46.

Jawad, M., Noreen, S., Ashraf, MF., Asif, M., Asif, HM., Ali, S., Rakha, A. Application of ribonucleic acid (RNA) in Forensic sciences. *Biomedica*. 2020; 36 (3). 232-47

Jedrzejczak, P., Kempisty, B., Bryja, A., Mostowska, M., Depa-Martynow, M., Pawelczyk, L., & Jagodzinski, P. P. (2007). Quantitative Assessment of Transition Proteins 1, 2 Spermatid-Specific Linker Histone H1-Like Protein Transcripts in Spermatozoa from Normozoospermic and Asthenozoospermic Men. *Archives of Andrology*, 53(4), 199–205. <https://doi.org/10.1080/01485010701426430>

Jina, R., & Thomas, L. S. (2013). Health consequences of sexual violence against women. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 27(1), 15–26. <https://doi.org/10.1016/j.bpobgyn.2012.08.012>

Johnson, D., Peterson, J., Sommers, I., & Baskin, D. (2012). Use of Forensic Science in Investigating Crimes of Sexual Violence. *Violence Against Women*, 18(2), 193–222. <https://doi.org/10.1177/1077801212440157>

Juusola, J., & Ballantyne, J. (2003). Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification. *Forensic Science International*, 135(2), 85–96. [https://doi.org/10.1016/S0379-0738\(03\)00197-X](https://doi.org/10.1016/S0379-0738(03)00197-X)

Juusola, J., & Ballantyne, J. (2005). Multiplex mRNA profiling for the identification of body fluids. *Forensic Science International*, 152(1), 1–12. <https://doi.org/10.1016/j.forsciint.2005.02.020>

Juusola, J., & Ballantyne, J. (2007). mRNA Profiling for Body Fluid Identification by Multiplex Quantitative RT-PCR. *Journal of Forensic Sciences*, 0(0), 070917231752009-??? <https://doi.org/10.1111/j.1556-4029.2007.00550.x>

Kapranov, P., Hubé, F., Wolf, D. A., Liu, C., & Li, J. (2019). Coding or Noncoding, the Converging Concepts of RNAs. <https://doi.org/10.3389/fgene.2019.00496>

Karijolich, J., & Yu, Y.-T. (2010). Spliceosomal snRNA modifications and their function. *Special Focus Review RNA Biology*, 7(2), 192–204. <https://doi.org/10.4161/rna.7.2.11207>

Khatter, H., Myasnikov, A. G., Natchiar, S. K., & Klaholz, B. P. (2015). Structure of the human 80S ribosome. *Nature*, 520(7549), 640–645. <https://doi.org/10.1038/nature14427>

Kim, T.-K., Hemberg, M., & Gray, J. M. (2015). Enhancer RNAs: A Class of Long Noncoding RNAs Synthesized at Enhancers: Figure 1. *Cold Spring Harbor Perspectives in Biology*, 7(1), a018622. <https://doi.org/10.1101/cshperspect.a018622>

Kleene, K. C., & Bagarova, J. (2008). Comparative genomics reveals gene-specific and shared regulatory sequences in the spermatid-expressed mammalian Odf1, Prm1, Prm2, Tnp1, and Tnp2 genes. *Genomics*, 92(2), 101–106. <https://doi.org/10.1016/j.ygeno.2008.05.001>

Kohlmeier, F., & Schneider, P. M. (2012). Successful mRNA profiling of 23 years old blood stains. *Forensic Science International: Genetics*, 6(2), 274–276. <https://doi.org/10.1016/j.fsigen.2011.04.007>

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. *International Journal of Legal Medicine*, 125(4), 573–580. <https://doi.org/10.1007/s00414-011-0578-1>

Koressaar, T., & Remm, M. (2018). Enhancements and modifications of primer design program Primer3. *Bioinformatics*. <https://primer3.ut.ee/>

Kristensen, L. S., Andersen, M. S., Stagsted, L. V. W., Ebbesen, K. K., Hansen, T. B., & Kjems, J. (2019). The biogenesis, biology and characterization of circular RNAs. *Nature Reviews Genetics*, 20(11), 675–691. <https://doi.org/10.1038/s41576-019-0158-7>

Kulstein, G., & Wiegand, P. (2018). Comprehensive examination of conventional and innovative body fluid identification approaches and DNA profiling of laundered blood- and saliva-stained pieces of cloths. *International Journal of Legal Medicine*, 132(1), 67–81. <https://doi.org/10.1007/s00414-017-1691-6>

Lee, P. Y., Costumbrado, J., Hsu, C.-Y., & Kim, Y. H. (2012). Agarose Gel Electrophoresis for the Separation of DNA Fragments. *Journal of Visualized Experiments*, 62. <https://doi.org/10.3791/3923>

Lehti, M. S., & Sironen, A. (2017). Formation and function of sperm tail structures in association with sperm motility defects†. *Biology of Reproduction*, 97(4), 522–536. <https://doi.org/10.1093/biolre/iox096>

Li, H., Wu, C., Gu, X., & Xiong, C. (2012). A novel application of cell-free seminal mRNA: non-invasive identification of the presence of germ cells or complete obstruction in men with azoospermia. *Human Reproduction*, 27(4), 991–997. <https://doi.org/10.1093/humrep/der481>

Li, S., & Mason, C. E. (2014). The Pivotal Regulatory Landscape of RNA Modifications. *Annu. Rev. Genomics Hum. Genet*, 15, 127–150. <https://doi.org/10.1146/annurev-genom-090413-025405>

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 Science International: Genetics*, 6(5), 565–577. <https://doi.org/10.1016/j.fsigen.2012.01.009>

Lindenbergh, A., Maaskant, P., & Sijen, T. (2013). Implementation of RNA profiling in forensic casework. *Forensic Science International: Genetics*, 7(1), 159–166. <https://doi.org/10.1016/j.fsigen.2012.09.003>

Lindenbergh, A., van den Berge, M., Oostra, R.-J., Cleypool, C., Bruggink, A., Kloosterman, A., & Sijen, T. (2013). Development of a mRNA profiling multiplex for the inference of organ tissues. *International Journal of Legal Medicine*, 127(5), 891–900. <https://doi.org/10.1007/s00414-013-0895-7>

Liu, Y., Jiang, M., Li, C., Yang, P., Sun, H., Tao, D., Zhang, S., & Ma, Y. (2011). Human t-Complex Protein 11 (TCP11), a Testis-Specific Gene Product, Is a Potential Determinant of the Sperm Morphology. *The Tohoku Journal of Experimental Medicine*, 224(2), 111–117. <https://doi.org/10.1620/tjem.224.111>

Loureiro, S. (2019). Identification of the source of biological fluids recovered from samples collected at Sexual Assaults scenarios using mRNA markers. Faculty of Medicine of the University of Porto.

Loureiro, S., Amorim, A., Cainé, L., Silva, B., & Gomes, I. (2019). Evaluation of two DNA/RNA co-extraction methods for body fluid identification in forensics. *Forensic Science International: Genetics Supplement Series*, 7(1), 250–252. <https://doi.org/10.1016/j.fsigss.2019.09.096>

Lu, M. (2020). Circular RNA: functions, applications and prospects. *ExRNA*, 2(1), 1. <https://doi.org/10.1186/s41544-019-0046-5>

Magalhães, T., Dinis-Oliveira, R. J., Silva, B., Corte-Real, F., & Nuno Vieira, D. (2015). Biological Evidence Management for DNA Analysis in Cases of Sexual Assault. *The Scientific World Journal*, 2015, 1–11. <https://doi.org/10.1155/2015/365674>

Maras, M.-H., & Miranda, M. D. (2017). Overlooking forensic evidence? A review of the 2014 International Protocol on the Documentation and Investigation of Sexual Violence in Conflict. *Global Security: Health, Science and Policy*, 2(1), 10–21. <https://doi.org/10.1080/23779497.2017.1281088>

Mathews, D. H., Sabina, J., Zuker, M., & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*, 288(5), 911–940. <https://doi.org/10.1006/JMBI.1999.2700>

Meng, J., Khandelia, P., Yu, Y.-T., Morais, P., & Adachi, H. (2021). Spliceosomal snRNA Epitranscriptomics. *Frontiers in Genetics | Www.Frontiersin.Org*, 1, 652129. <https://doi.org/10.3389/fgene.2021.652129>

Meng, S., Zhou, H., Feng, Z., Xu, Z., Tang, Y., Li, P., & Wu, M. (2017). CircRNA: Functions and properties of a novel potential biomarker for cancer. *Molecular Cancer*, 16(1), 1–8. <https://doi.org/10.1186/S12943-017-0663-2/FIGURES/2>

miRBase. (n.d.). Retrieved August 17, 2022, from <https://www.mirbase.org/>

Miyagawa, Y. (2005). Single-Nucleotide Polymorphisms and Mutation Analyses of the TNP1 and TNP2 Genes of Fertile and Infertile Human Male Populations. *Journal of Andrology*, 26(6), 779–786. <https://doi.org/10.2164/jandrol.05069>

Molecular Biology Review. (n.d.). Retrieved August 13, 2022, from <https://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/MolBioReview/mrna.html>

Mollen, C. J., Goyal, M. K., & Frioux, S. M. (2012). Acute Sexual Assault. *Pediatric Emergency Care*, 28(6), 584–590. <https://doi.org/10.1097/PEC.0b013e318258bfea>

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. (2012). Determination of an Effective Housekeeping Gene for the Quantification of mRNA for Forensic Applications*. *Journal of Forensic Sciences*, 57(4), 1051–1058. <https://doi.org/10.1111/j.1556-4029.2012.02086.x>

Nakanishi, H., Hara, M., Takahashi, S., Takada, A., & Saito, K. (2014). Evaluation of forensic examination of extremely aged seminal stains. *Legal Medicine*, 16(5), 303–307. <https://doi.org/10.1016/j.legalmed.2014.04.002>

Nayernia, K., Adham, I. M., Burkhardt-Göttges, E., Neesen, J., Rieche, M., Wolf, S., Sancken, U., Kleene, K., & Engel, W. (2002). Asthenozoospermia in Mice with Targeted Deletion of the Sperm Mitochondrion-Associated Cysteine-Rich Protein (Smcp) Gene. *Molecular and Cellular Biology*, 22(9), 3046–3052. <https://doi.org/10.1128/MCB.22.9.3046-3052.2002>

Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M. K., Bonnefond, L., Matsumoto, N., Nishizawa, T., Nakanaga, K., Aoki, J., Ishitani, R., Siomi, H., Siomi, M. C., & Nureki, O. (2012). Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature*. <https://doi.org/10.1038/nature11509>

Nussbaumer, C., Gharebaghi-Schnell, E., & Korschineck, I. (2006). Messenger RNA profiling: A novel method for body fluid identification by Real-Time PCR. *Forensic Science International*, 157(2–3), 181–186. <https://doi.org/10.1016/j.forsciint.2005.10.009>

O`Leary, K. R., & Glynn, C. L. (2018). Investigating the Isolation and Amplification of microRNAs for Forensic Body Fluid Identification. *MicroRNA*, 7(3), 187–194. <https://doi.org/10.2174/2211536607666180430153821>

ODF1 outer dense fiber of sperm tails 1 [Homo sapiens (human)]. (2023). <https://www.ncbi.nlm.nih.gov/gene/4956#reference-sequences>

Ohshima, T. (2000). Forensic wound examination. *Forensic Science International*, 113(1–3), 153–164. [https://doi.org/10.1016/S0379-0738\(00\)00269-3](https://doi.org/10.1016/S0379-0738(00)00269-3)

Palazzo, A. F., Lee, E. S., Mekhail, K., & Miura, P. (2015). HYPOTHESIS AND THEORY ARTICLE Non-coding RNA: what is functional and what is junk? <https://doi.org/10.3389/fgene.2015.00002>

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. (2013). Genome-wide mRNA profiling and multiplex quantitative RT-PCR for forensic body fluid identification. *Forensic Science International: Genetics*, 7(1), 143–150. <https://doi.org/10.1016/j.fsigen.2012.09.001>

Patil, V. S., Zhou, R., & Rana, T. M. (2014). Gene regulation by non-coding RNAs. *Critical Reviews in Biochemistry and Molecular Biology*, 49(1), 16–32. <https://doi.org/10.3109/10409238.2013.844092>

Pelley, J. W. (2012). RNA Transcription and Control of Gene Expression. Elsevier's Integrated Review Biochemistry, 137–147. <https://doi.org/10.1016/B978-0-323-07446-9.00016-7>

PHF7 PHD finger protein 7 [Homo sapiens (human)]. (2023). <https://www.ncbi.nlm.nih.gov/gene/51533#gene-expression>

Pierce, B. (2012). *Genetics: A Conceptual Approach* (S. Winslow, L. Samols, G. Lee Hadler, P. Zimmerman, T. Szczepanski, A. Gass, & A. Bristow, Eds.; 4th ed.). Kate Ahr Parker.

Poór, V. S., Lukács, D., Nagy, T., Rácz, E., & Sipos, K. (2016). The rate of RNA degradation in human dental pulp reveals post-mortem interval. *International Journal of Legal Medicine*, 130(3), 615–619. <https://doi.org/10.1007/s00414-015-1295-y>

Porro, A., Cioce, M., Neugebauer, K. M., Farnebo, M., & Henriksson, S. (2015). On the road with WRAP53β: guardian of Cajal bodies and genome integrity. *Frontiers in Genetics | Www.Frontiersin.Org*, 6, 91. <https://doi.org/10.3389/fgene.2015.00091>

Ramaswamy, G., & Slack, F. J. (2002). siRNA: A Guide for RNA Silencing . *Chemistry & Biology*, 9(10), 1053–1055.

Ransohoff, J. D., Wei, Y., & Khavari, P. A. (2017). The functions and unique features of long intergenic non-coding RNA. Nature Publishing Group, 19. <https://doi.org/10.1038/nrm.2017.104>

Richard, M. L. L., Harper, K. A., Craig, R. L., Onorato, A. J., Robertson, J. M., & Donfack, J. (2012). Evaluation of mRNA marker specificity for the identification of five human body fluids by capillary electrophoresis. Forensic Science International: Genetics, 6(4), 452–460. <https://doi.org/10.1016/j.fsigen.2011.09.007>

Roeder, A. D., & Haas, C. (2013). mRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification. International Journal of Legal Medicine, 127(4), 707–721. <https://doi.org/10.1007/s00414-012-0794-3>

Roeder, A. D., & Haas, C. (2016). Body Fluid Identification Using mRNA Profiling (pp. 13–31). https://doi.org/10.1007/978-1-4939-3597-0_2

Rosikiewicz, W., & Makałowska, I. (2016). Biological functions of natural antisense transcripts. Biochimica Polonica, 63(4), 665–673. https://doi.org/10.18388/abp.2016_1350

Runnels, C. M., Lanier, K. A., Krish Williams, J., Bowman, J. C., Petrov, A. S., Hud, N. V., & Dean Williams, L. (2018). Folding, Assembly, and Persistence: The Essential Nature and Origins of Biopolymers. Journal of Molecular Evolution, 86, 598–610. <https://doi.org/10.1007/s00239-018-9876-2>

Salzmann, A. P., Russo, G., Kreutzer, S., & Haas, C. (2021). Degradation of human mRNA transcripts over time as an indicator of the time since deposition (TsD) in biological crime scene traces. Forensic Science International: Genetics, 53, 102524. <https://doi.org/10.1016/j.fsigen.2021.102524>

Schlicker, M., Schnülle, V., Schnepfel, L., Vorob'ev, V. I., & Engel, W. (1994). Genetics: Disturbances of nuclear condensation in human spermatozoa: search for mutations in the genes for protamine 1, protamine 2 and transition protein 1. Human Reproduction, 9(12), 2313–2317. <https://doi.org/10.1093/oxfordjournals.humrep.a138444>

Schneider, P. M. (2007). Scientific standards for studies in forensic genetics. Forensic Science International, 165(2–3), 238–243. <https://doi.org/10.1016/j.forsciint.2006.06.067>

Schöniger, C., & Arenz, C. (2013). Perspectives in targeting miRNA function . *Bioorganic & Medicinal Chemistry*, 6115–6118.

Scott, M. S., & Ono, M. (2011). From snoRNA to miRNA: Dual function regulatory non-coding RNAs. *Biochimie*, 93(11), 1987–1992. <https://doi.org/10.1016/J.BIOCHI.2011.05.026>

Scrivano, S., Sanavio, M., Tozzo, P., & Caenazzo, L. (2019). Analysis of RNA in the estimation of post-mortem interval: a review of current evidence. *International Journal of Legal Medicine*, 133(6), 1629–1640. <https://doi.org/10.1007/s00414-019-02125-x>

Sen, S. K. (2023, September 22). Messenger RNA (MRNA). <https://www.genome.gov/genetics-glossary/messenger-rna>

Setzer, M., Juusola, J., & Ballantyne, J. (2008). Recovery and Stability of RNA in Vaginal Swabs and Blood, Semen, and Saliva Stains. *Journal of Forensic Sciences*, 53(2), 296–305. <https://doi.org/10.1111/j.1556-4029.2007.00652.x>

Sharma, R., Agarwal, A., Rohra, V. K., Assidi, M., Abu-Elmagd, M., & Turki, R. F. (2015). Effects of increased paternal age on sperm quality, reproductive outcome and associated epigenetic risks to offspring. *Reproductive Biology and Endocrinology*, 13(1), 35. <https://doi.org/10.1186/s12958-015-0028-x>

Shibata, C., Otsuka, M., Kishikawa, T., Ohno, M., Yoshikawa, T., Takata, A., & Koike, K. (2015). Diagnostic and therapeutic application of noncoding RNAs for hepatocellular carcinoma . *World Journal of Hepatology*, 7(1).

Shirley, C. R., Hayashi, S., Mounsey, S., Yanagimachi, R., & Meistrich, M. L. (2004). Abnormalities and Reduced Reproductive Potential of Sperm from Tnp1- and Tnp2-Null Double Mutant Mice1. *Biology of Reproduction*, 71(4), 1220–1229. <https://doi.org/10.1095/biolreprod.104.029363>

Shuvalov, O., Kizenko, A., Petukhov, A., Fedorova, O., Daks, A., Bottrill, A., Snezhkina, A. V., Kudryavtseva, A. V., & Barlev, N. (2020). SEMG1/2 augment energy metabolism of tumor cells. *Cell Death & Disease*, 11(12), 1047. <https://doi.org/10.1038/s41419-020-03251-w>

Sijen, T. (2015). Molecular approaches for forensic cell type identification: On mRNA, miRNA, DNA methylation and microbial markers. *Forensic Science International: Genetics*, 18, 21–32. <https://doi.org/10.1016/j.fsigen.2014.11.015>

Sijen, T., & Harbison, S. (2021). On the Identification of Body Fluids and Tissues: A Crucial Link in the Investigation and Solution of Crime. *Genes*, 12(11), 1728. <https://doi.org/10.3390/genes12111728>

Silva, S. S., Lopes, C., Teixeira, A. L., Sousa, M. J. C. de, & Medeiros, R. (2015). Forensic miRNA: Potential biomarker for body fluids? *Forensic Science International: Genetics*, 14, 1–10. <https://doi.org/10.1016/j.fsigen.2014.09.002>

Sirker, M., Gomes, I., Rothschild, M. A., & Schneider, P. M. (2013). A 1-year time course study of human RNA degradation in body fluids under dry and humid environmental conditions. *Forensic Science International: Genetics Supplement Series*, 4(1), e164–e165. <https://doi.org/10.1016/j.fsigss.2013.10.085>

Sirker, M., Schneider, P. M., & Gomes, I. (2016). A 17-month time course study of human RNA and DNA degradation in body fluids under dry and humid environmental conditions. *International Journal of Legal Medicine*, 130(6), 1431–1438. <https://doi.org/10.1007/s00414-016-1373-9>

Sloan, K. E., Warda, A. S., Sharma, S., Entian, K.-D., Lafontaine, D. L. J., & Bohnsack, M. T. (2016). RNA Biology Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. <https://doi.org/10.1080/15476286.2016.1259781>

Sponer, J., Bussi, G., Krepl, M., Banas, P., Bottaro, S., Cunha, R. A., Gil-Ley, A., Pinamonti, G., Poblete, S., Jurečka, P., Walter, N. G., & Otyepka, M. (2018). RNA structural dynamics as captured by molecular simulations: A comprehensive overview. *Chemical Reviews*, 118(8), 4177–4338. <https://doi.org/10.1021/ACS.CHEMREV.7B00427>

Stanger, S. J., Law, E. A., Jamsai, D., O'Bryan, M. K., Nixon, B., McLaughlin, E. A., Aitken, R. J., & Roman, S. D. (2016). A novel germ cell protein, SPIF (sperm PKA interacting factor), is essential for the formation of a PKA/TCP11 complex that undergoes conformational and phosphorylation changes upon capacitation. *The FASEB Journal*, 30(8), 2777–2791. <https://doi.org/10.1096/fj.201500136R>

Sweeney, B. A., Roy, P., & Leontis, N. B. (2015). An introduction to recurrent nucleotide interactions in RNA. *Wiley Interdisciplinary Reviews: RNA*, 6(1), 17–45. <https://doi.org/10.1002/WRNA.1258>

Takahashi, A., Hirohashi, Y., Torigoe, T., Tamura, Y., Tsukahara, T., Kanaseki, T., Kochin, V., Saijo, H., Kubo, T., Nakatsugawa, M., Asanuma, H., Hasegawa, T., Kondo, T., & Sato, N. (2013). Ectopically Expressed Variant Form of Sperm Mitochondria-Associated Cysteine-Rich Protein Augments Tumorigenicity of the Stem Cell Population of Lung Adenocarcinoma Cells. *PLoS ONE*, 8(11), e69095. <https://doi.org/10.1371/journal.pone.0069095>

Takamura, A., Watanabe, K., & Akutsu, T. (2017). Advanced forensic validation for human spermatozoa identification using SPERM HY-LITERTM Express with quantitative image analysis. *International Journal of Legal Medicine*, 131(4), 933–939. <https://doi.org/10.1007/s00414-017-1536-3>

Tamaki, K., Fujisawa, K., Okajima, H., Sato, K., & Katsumata, Y. (1989). Identification of semen in stain by determination of the specific activity of L-tartrate-inhibitable acid phosphatase. *Zeitschrift Für Rechtsmedizin*, 102(7). <https://doi.org/10.1007/BF00204073>

Tang, G. (2005). siRNA and miRNA: an insight into RISCs. *Trends in Biochemical Sciences*, 30(2), 106–114.

Tapia Contreras, C., & Hoyer-Fender, S. (2019). CCDC42 Localizes to Manchette, HTCA and Tail and Interacts With ODF1 and ODF2 in the Formation of the Male Germ Cell Cytoskeleton. *Frontiers in Cell and Developmental Biology*, 7. <https://doi.org/10.3389/fcell.2019.00151>

Tian, H., Lv, M., Li, Z., Peng, D., Tan, Y., Wang, H., Li, Q., Li, F., & Liang, W. (2018). Semen-specific miRNAs: Suitable for the distinction of infertile semen in the body fluid identification? *Forensic Science International: Genetics*, 33, 161–167. <https://doi.org/10.1016/j.fsigen.2017.12.010>

Tomita, H., Vawter, M. P., Walsh, D. M., Evans, S. J., Choudary, P. V, Li, J., Overman, K. M., Atz, M. E., Myers, R. M., Jones, E. G., Watson, S. J., Akil, H., & Bunney, W. E. (2004). Effect of agonal and postmortem factors on gene expression profile: quality control in microarray analyses of postmortem human brain. *Biological Psychiatry*, 55(4), 346–352. <https://doi.org/10.1016/j.biopsych.2003.10.013>

Torres, A. G. (2019). Enjoy the Silence: Nearly Half of Human tRNA Genes Are Silent. *Bioinformatics and Biology Insights*, 13, 117793221986845. <https://doi.org/10.1177/1177932219868454>

Uhlenbeck, O. (1995). Keeping RNA happy. *RNA Society*, 1(1), 4–6.

Vallone, P. M., & Butler, J. M. (2004). AutoDimer: a screening tool for primer-dimer and hairpin structures. <https://bio.tools/autodimer>

van den Berge, M., Carracedo, A., Gomes, I., Graham, E. A. M., Haas, C., Hjort, B., Hoff-Olsen, P., Maroñas, O., Mevåg, B., Morling, N., Niederstätter, H., Parson, W., Schneider, P. M., Court, D. S., Vidaki, A., & Sijen, T. (2014). A collaborative European exercise on mRNA-based body fluid/skin typing and interpretation of DNA and RNA results. *Forensic Science International: Genetics*, 10, 40–48. <https://doi.org/10.1016/j.fsigen.2014.01.006>

van den Berge, M., & Sijen, T. (2017). A male and female RNA marker to infer sex in forensic analysis. *Forensic Science International: Genetics*, 26, 70–76. <https://doi.org/10.1016/j.fsigen.2016.10.018>

Verduci, L., Tarcitano, E., Strano, S., Yarden, Y., & Blandino, G. (2021). CircRNAs: role in human diseases and potential use as biomarkers. *Cell Death & Disease*, 12(5), 468. <https://doi.org/10.1038/s41419-021-03743-3>

Virkler, K., & Lednev, I. K. (2009). Analysis of body fluids for forensic purposes: From laboratory testing to non-destructive rapid confirmatory identification at a crime scene. *Forensic Science International*, 188(1–3), 1–17. <https://doi.org/10.1016/j.forsciint.2009.02.013>

Vos, P. D., Leedman, P. J., Filipovska, A., & Rackham, · Oliver. (2019). Modulation of miRNA function by natural and synthetic RNA-binding proteins in cancer miRNA · Designer RNA-binding proteins · Pentatricopeptide repeat · PUF domain · Synthetic biology · RNA-protein interactions · RNA interference. *Cellular and Molecular Life Sciences*, 76, 3745–3752. <https://doi.org/10.1007/s00018-019-03163-9>

Wang, X., Kang, J.-Y., Wei, L., Yang, X., Sun, H., Yang, S., Lu, L., Yan, M., Bai, M., Chen, Y., Long, J., Li, N., Li, D., Huang, J., Lei, M., Shao, Z., Yuan, W., Zuo, E., Lu, K., ... Li, J. (2019). PHF7 is a novel histone H2A E3 ligase prior to histone-to-protamine exchange during spermiogenesis. *Development*. <https://doi.org/10.1242/dev.175547>

Wang, X. R., Ling, L. Bin, Huang, H. H., Lin, J. J., Fugmann, S. D., & Yang, S. Y. (2017). Evidence for parallel evolution of a gene involved in the regulation of spermatogenesis. *Proceedings of the Royal Society B: Biological Sciences*, 284(1855), 20170324. <https://doi.org/10.1098/rspb.2017.0324>

Wang, Z., Zhang, J., Luo, H., Ye, Y., Yan, J., & Hou, Y. (2013). Screening and confirmation of microRNA markers for forensic body fluid identification. *Forensic Science International: Genetics*, 7(1), 116–123. <https://doi.org/10.1016/j.fsigen.2012.07.006>

Wang, Z., Zhang, S., Zhou, D., Zhao, S., & Li, C. (2013). Messenger RNA Profiling for Forensic Body Fluid Identification: Research and Applications. *Journal of Forensic Medicine*, 29(5).

Weick, E.-M., & Miska, E. A. (2014). piRNAs: from biogenesis to function. <https://doi.org/10.1242/dev.094037>

Welch, J., & Mason, F. (2007). Rape and sexual assault. *BMJ*, 334(7604), 1154–1158. <https://doi.org/10.1136/bmj.39211.403970.BE>

WHO. (2012). Understanding and addressing violence against women: sexual violence .

WHO. (2021, March 9). Violence against women. <https://www.who.int/news-room/fact-sheets/detail/violence-against-women>

Wight, M., & Werner, A. (2013). The functions of natural antisense transcripts. *Essays In Biochemistry*, 54, 91–101.

Wingo, T. S., Loreni, F., Yao, B., Guang, S., Feng, X., Yan, Q., & Zhu, C. (2019). The Functions of Non-coding RNAs in rRNA Regulation. <https://doi.org/10.3389/fgene.2019.00290>

Wu, J., Xiao, J., Zhang, Z., Wang, X., Hu, S., & Yu, J. (2014). Ribogenomics: The Science and Knowledge of RNA. *Genomics, Proteomics and Bioinformatics*, 12(2), 57–63. <https://doi.org/10.1016/J.GPB.2014.04.002>

Yang, H., Zhou, B., Deng, H., Prinz, M., & Siegel, D. (2013). Body fluid identification by mass spectrometry. *International Journal of Legal Medicine*, 127(6), 1065–1077. <https://doi.org/10.1007/s00414-013-0848-1>

Yang, K., Meinhardt, A., Zhang, B., Grzmil, P., Adham, I. M., & Hoyer-Fender, S. (2012). The Small Heat Shock Protein ODF1/HSPB10 Is Essential for Tight Linkage of Sperm Head to Tail and Male Fertility in Mice. *Molecular and Cellular Biology*, 32(1), 216–225. <https://doi.org/10.1128/MCB.06158-11>

Yasojima, K., McGeer, E. G., & McGeer, P. L. (2001). High stability of mRNAs postmortem and protocols for their assessment by RT-PCR. *Brain Research Protocols*, 8(3), 212–218. [https://doi.org/10.1016/S1385-299X\(01\)00119-2](https://doi.org/10.1016/S1385-299X(01)00119-2)

Yu, Y. E., Zhang, Y., Unni, E., Shirley, C. R., Deng, J. M., Russell, L. D., Weil, M. M., Behringer, R. R., & Meistrich, M. L. (2000). Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. *Proceedings of the National Academy of Sciences*, 97(9), 4683–4688. <https://doi.org/10.1073/pnas.97.9.4683>

Zhao, W., Li, Z., Ping, P., Wang, G., Yuan, X., & Sun, F. (2018). Outer dense fibers stabilize the axoneme to maintain sperm motility. *Journal of Cellular and Molecular Medicine*, 22(3), 1755–1768. <https://doi.org/10.1111/jcmm.13457>

Zhu, H., Zhang, H., Xu, Y., Laššáková, S., Korabečná, M., & Neužil, P. (2020). PCR past, present and future. *BioTechniques*, 69(4), 317–325. <https://doi.org/10.2144/btn-2020-0057>

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 time-wise degraded samples. *International Journal of Legal Medicine*, 122(2), 135–142. <https://doi.org/10.1007/s00414-007-0182-6>

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. *International Journal of Legal Medicine*, 123(1), 71–74. <https://doi.org/10.1007/s00414-008-0249-z>