Science & Justice xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Science & Justice

journal homepage: www.elsevier.com/locate/scijus

Novel messenger RNAs for body fluid identification

Patricia P. Albani[a,b], Rachel Fleming[b,*]

[a] School of Chemical Sciences, The University of Auckland, Auckland, New Zealand
[b] The Institute of Environmental Science and Research Ltd. (ESR), Auckland, New Zealand

ARTICLE INFO

Keywords:
Forensic science
Body fluid identification
Messenger ribonucleic acid
Transcriptome sequencing
Circulatory blood
Semen
Seminal fluid
Menstrual fluid
Body fluid markers
Polymerase chain reaction
Capillary electrophoresis
Gene expression patterns

ABSTRACT

In forensic investigations, the identification of the cellular or body fluid source of biological evidence can provide crucial probative information for the court. Messenger RNA (mRNA) profiling has become a valuable tool for body fluid and cell type identification due to its high sensitivity and compatibility with DNA analysis. However, using a single marker to determine the somatic origin of a sample can lead to misinterpretation as a result of cross-reactions. While false positives may be avoided through the simultaneous detection of multiple markers per body fluid, this approach is currently limited by the small number of known differentially expressed mRNAs.

Here we characterise six novel mRNAs, partly identified from RNA-Seq, which can supplement existing markers for the detection of circulatory blood, semen (with and without spermatozoa), and menstrual fluid: HBD and SLC4A1 for blood, TNP1 for spermatozoa, KLK2 for seminal fluid, and MMP3 and STC1 for menstrual fluid. Respective expression profiles were evaluated by singleplex endpoint reverse transcription polymerase chain reaction (RT-PCR). HBD, SLC4A1, and KLK2 were specific to their target body fluids. TNP1, MMP3, and STC1 each cross-reacted with two non-target samples; however, these signals were below 350 RFU, not reproducible, and likely resulted from large body fluid inputs. All candidates were more sensitive for the detection of their target body fluids than corresponding well-known mRNAs, in particular those for menstrual fluid. The increased sensitivities were statistically significant, except for KLK2. Thus, the new mRNAs introduced here are promising new targets for improved body fluid profiling.

1. Introduction

An important aspect in some criminal investigations may be the identification of the cellular or body fluid origin of a sample recovered from a crime scene or an item of interest. This collective knowledge offers vital details about the activity scenarios of a case [1–4]. For example, the presence of menstrual fluid indicates sexual activity, whereas circulatory blood may be the result of a traumatic injury. Conventional chemical, serological, and enzymatic body fluid tests, however, are often insufficient in terms of specificity and sensitivity and for some forensically relevant body fluids, no reliable tests exist [1–3].

Messenger RNA (mRNA) profiling based on unique gene expression patterns in cells and tissues has emerged as a method to overcome these limitations [1–4]. DNA/RNA co-extraction for combined STR and body fluid profiling is now an effective and comprehensive tool used by casework laboratories around the world. Yet since the introduction of differentially expressed mRNAs for forensic saliva analysis in 2003 [2], only a small set of 'core' markers has been used for multiplex design. These include histatin 3 (HTN3) and statherin (STATH) for saliva and buccal mucosa [1,3,5–7], protamines 1 and 2 (PRM1/2) for semen [1,3,5–7], transglutaminase 4 (TGM4) or semenogelin 1 (SEMG1) for seminal fluid [1,3], matrix metallopeptidases (MMPs) 7, 10, or 11 for menstrual fluid [1,3,5–7], as well as human beta-defensin 1 (HBD1), mucin 4 (MUC4), or *Lactobacilli crispatus* and *gasseri* for vaginal material [1,3,5–7]. Greater variability is seen in the use of circulatory blood markers. Commonly targeted transcripts include spectrin beta (SPTB), hydroxymethylbilane synthase (PBGD), 5′-aminolevulinate synthase 2 (ALAS2), glycophorin A (GYPA), adhesion molecule, interacts with CXADR antigen 1 (AMICA1), CD93 molecule, and haemoglobin beta (HBB) [1,3,5–7]. Other mRNA markers have been proposed, but are less frequently used due to inferior specificity and sensitivity in comparison to the above markers [8–13]. An exception to this is cytochrome P450 family 2, subfamily B, member 7, pseudogene (CYP2B7P), a useful marker for the detection of vaginal material [14].

Unambiguous specificity to the intended target body fluid is one of the most important qualities of a valuable mRNA marker. Gene expression, however, varies in response to a plethora of extrinsic and

* Corresponding author at: Institute of Environmental Science and Research Ltd. (ESR), 120 Mt Albert Road, Auckland 1025, New Zealand.
E-mail address: rachel.fleming@esr.cri.nz (R. Fleming).

http://dx.doi.org/10.1016/j.scijus.2017.09.002
Received 15 January 2017; Received in revised form 20 August 2017; Accepted 23 September 2017

intrinsic factors, including age, gender, environmental, and physiological changes [4]. Numerous studies have reported that cross-reactions with non-target body fluids, although often sporadic and non-reproducible, can occur with all of the aforementioned mRNAs [8–13,15,16]. In particular, HBD1 and MUC4 are well-known for their presence in both vaginal and buccal mucosa [10,15]. Adjusting RNA and cDNA input amounts, primer concentrations, and PCR conditions in combination with post-PCR clean-up steps and suitable interpretation guidelines can help reduce the frequency of such false positives [4,10–13]. Nevertheless, true low-level expression in non-target tissues remains problematic, particularly in the resolution of mixtures of body fluids.

Another approach to increasing the confidence in observed peaks of low height and reducing the probability of obtaining false positives can be to detect multiple mRNAs per body fluid simultaneously [4,10]. This is currently limited by the small number of known markers for most body fluid types. Previous attempts to identify novel differentially expressed mRNAs were mostly based on database and literature searches [1,2,5,6,10,17] or microarray analysis [18–20]. Problems can arise from the low sensitivity of microarrays for small copy numbers [21], which in the past led to the identification of body fluid candidates that were not specific when tested on body fluids. In addition, both methods rely on previous transcript characterisation, i.e. determination of possible physiological function and mRNA sequence [21].

Recently, transcriptome sequencing (RNA-Seq) has gained increasing attention as a powerful approach for the discovery of novel body fluid markers and the enhancement of detection sensitivity [14,21–23]. Other than hybridisation-based methods, RNA-Seq determines base sequences and therefore screens both known and unknown mRNAs, as well as transcript variants [21]. The simultaneous high-throughput assessment of the entire transcript pool in a sample also enables measuring the relative abundance of each mRNA [14,21–23]. Previously, this laboratory reported the successful sequencing of degraded body fluid transcriptomes [22]. Using the same data set and comparing different body fluid types, we identified four novel body fluid-specific mRNAs: haemoglobin delta (HBD) and solute carrier family 4 (anion exchanger), member 1 (Diego blood group) (SLC4A1) for circulatory blood, as well as MMP3 and stanniocalcin 1 (STC1) for menstrual fluid. Two further candidates were obtained from an extensive database search: transition protein 1 (TNP1) for spermatozoa and kallikrein-related peptidase 2 (KLK2) for seminal fluid which does not contain spermatozoa.

2. Materials and methods

2.1. Identification of body fluid-specific candidate genes

Candidate mRNAs for the detection of circulatory blood and menstrual fluid were identified from interrogation of RNA-Seq data of degraded body fluids as published previously [22]. Whole transcriptome paired-end sequencing (2 × 100 bp) of circulatory blood (two donors) and menstrual fluid (one donor) was performed in order to select highly expressed biomarkers possibly exclusive to each body fluid type [22]. Processed and merged sequencing reads were aligned to the human reference sequence assembly hg19 (GRCh37) and normalised to generate fragments per kilobase of exon per million fragments mapped (FPKM) values for each detected transcript [22]. Data were sorted by maximum FPKM values and compared between sample types to exclude genes that were expressed in multiple body fluids, and identify highly abundant and possibly specific body fluid markers. Four mRNA candidates were selected from this data set: HBD and SLC4A1 for circulatory blood, as well as MMP3 and STC1 for menstrual fluid. Table 1 lists the FPKM values of these transcripts compared to the known mRNA markers GYPA and MMP11.

Two further candidate genes were selected from two gene expression databases (TiGER, PaGenBase) [24,25] based on their putative physiological function in the human body: TNP1 for spermatozoa and KLK2 for seminal fluid which may be free of spermatozoa.

2.2. Primer design

Primers for HBD, SLC4A1, MMP3, and STC1 were designed to target transcript stable regions (StaRs) as described previously [23] using the OligoAnalyzer 3.1 online tool (Integrated DNA Technologies, Inc., Coralville, IA, USA). Sequencing coverage maps were viewed using the Geneious v.5.6.7 software (Biomatters Ltd., Auckland, New Zealand) and regions of high coverage were selected for primer design. Primers for TNP1 and KLK2 were designed using conventional primer design strategy. The specificity of all primers to their intended mRNA targets was verified using Primer-BLAST [26]. Primer sequences and expected amplicon sizes are listed in Table 2.

2.3. Collection of body fluid samples

Six samples each of 50 µL circulatory blood, semen, and seminal fluid (azoospermic), as well as saliva/buccal mucosa, menstrual, and non-menstrual vaginal swabs were obtained from healthy, consenting volunteers, as approved by the University of Auckland Human Participants Ethics Committee (application no. 011733). Blood was drawn using a sterile AKKU-CHEK® Safe-T-Pro Plus lancet (Roche Diagnostics USA, Indianapolis, IN, USA). Blood, semen, and seminal fluid aliquots were deposited onto sterile Cultiplast® rayon swabs. Buccal, menstrual, and vaginal samples were obtained by volunteers themselves using sterile swabs. All samples were allowed to dry overnight at ambient laboratory conditions and subsequently extracted as described below.

2.4. RNA extraction and purification

Total RNA from body fluid samples was prepared as described previously [22,23] using the Promega® DNA IQ™ and ReliaPrep™ RNA Cell Miniprep Systems (Promega Corporation, Madison, WI, USA) following the manufacturer's instructions. Genomic DNA was removed by incorporating an on-column DNase I treatment during the RNA extraction process. RNA was eluted in 45 µL nuclease-free water. The absence of genomic DNA was verified by real-time PCR using the Quantifiler® Human DNA quantification kit (Life Technologies™ by Thermo Fisher Scientific, Inc., Waltham, MA, USA) with 1 µL purified RNA in a 12.5 µL reaction. Samples, which contained residual DNA, were treated with TURBO™ DNase (Invitrogen™ by Thermo Fisher Scientific, Inc.) and re-quantified until no DNA was detectable.

2.5. cDNA synthesis

Complementary DNA (cDNA) was prepared using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems™ by Thermo Fisher Scientific, Inc.) according to the manufacturer's instructions. Ten microliters of DNA-free RNA were subjected to reverse transcription in a 20 µL reaction. Synthesis was performed on a GeneAmp PCR System 9700 thermal cycler (Applied Biosystems™ by Thermo Fisher Scientific, Inc.) using the following program: 25 °C for 10 min, 37 °C for 120 min, followed by 85 °C for 5 min and hold at 4 °C.

2.6. Polymerase chain reaction (PCR)

2.6.1. PCR reactions

Body fluid cDNA samples were amplified using the QIAGEN® Multiplex PCR Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. Two microliters of cDNA was was amplified in a 25 uL reaction containing 12.5 µL of 2 × PCR master mix. Optimised primer concentrations for specificity testing were as follows: 0.05 µM (HBD), 0.03 µM (SLC4A1), 0.08 µM (TNP1), 0.4 µM (KLK2),