Forensic Science International: Genetics 20 (2016) 61–70

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Degraded RNA transcript stable regions (StaRs) as targets for enhanced forensic RNA body fluid identification

Meng-Han Lin[a], Patricia P. Albani[a,b], Rachel Fleming[a,*]

[a] Forensic Research and Development, Institute of Environmental Science and Research Limited, Mount Albert, Auckland, New Zealand
[b] School of Chemical Sciences, University of Auckland, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 12 June 2015
Received in revised form 11 September 2015
Accepted 24 September 2015
Available online 3 October 2015

Keywords:
RNASeq
Transcriptomics
Forensic
Body fluid identification
Degraded RNA
Stable transcript regions

ABSTRACT

The detection of messenger RNA (mRNA) using reverse transcriptase PCR (RT-PCR) is becoming common practice for forensic body fluid identification. However, the degraded and scarce nature of RNA from forensic samples mean that mRNA transcripts are not consistently detected or remain undetected in practice. Conventional primer design for RT-PCR (and quantitative RT-PCR) includes targeting primers to span exon–exon boundaries or by having the primers on two separate exons, and satisfying common primer thermodynamic criteria. We have found that the conventional placement of primers is not always optimal for obtaining reproducible results from degraded samples. Using massively parallel sequencing data from degraded body fluids, we designed primers to amplify transcript regions of high read coverage, hence, higher stability, and compared these with primers designed using conventional methodology. Our findings are that primers designed for transcript regions of higher read coverage resulted in vastly improved detection of mRNA transcripts that were not previously detected or were not consistently detected in the same samples using conventional primers. We developed a new concept whereby primers targeted to transcript stable regions (StaRs) are able to consistently and specifically amplify a wide range of RNA biomarkers in various body fluids of varying degradation levels.

1. Introduction

The ability to accurately detect and quantify RNA transcript abundance is a core capability in molecular biology. The broad set of RNA detection methods currently available range from non-amplification methods (*in situ* hybridization, microarray and NanoString® nCounter®), to amplification (PCR) based methods (reverse transcriptase PCR (RT-PCR) and quantitative reverse transcriptase PCR (qRT-PCR)). With the exception of RNA massively parallel sequencing (MPS), a key prerequisite of all RNA detection technology is prior knowledge of the target transcript sequence. Target detection is facilitated by oligonucleotide sequences in both non-amplification methods (probe) and amplification-based methods (primers).

The ability to detect the RNA of interest is essential in many disciplines such as forensic RNA analysis, which is currently dominated by PCR-based detection methods such as the multiplex PCR-based body fluid assays [1–3]. Developments for improving marker detection often focus on PCR conditions and primer design. Methods for PCR primer design are always evolving [4,5] but remain based around the core criteria of specificity, thermodynamics, secondary structure, dimerization and amplicon length [6–10]. In addition to these criteria, RT-PCR primer design (for RNA transcript amplification) also considers exon boundary coverage to ensure amplification of only cDNA and to avoid amplification of genomic DNA [11]. Among other experimental factors [12–17], it is widely acknowledged that PCR primer design has critical implications for target amplification, detection and quantification [6,11,14,18–21].

Whilst improvements to primer design/amplification or probe design/hybridization conditions can yield performance improvements [6–10], the target molecule must also be considered. RNA is inherently unstable and easily degraded through a variety of pathways. In eukaryotes RNA 3′–5′ degradation occurs primarily through enzymatic cleavage of the poly(A) tail [22]; and 5′–3′ degradation through decapping, followed by 5′ exonuclease degradation or 3′ degradation by the exosome [23].

In forensic RNA analysis, RNA degradation has significant implications on the sensitivity and specificity of biomarker detection. Body fluid samples recovered for forensic testing may have been subjected to environmental insults or stored for a period of time and can be significantly degraded. Samples of such low

* Corresponding author. Fax.: +64 98496046.
E-mail address: rachel.fleming@esr.cri.nz (R. Fleming).

http://dx.doi.org/10.1016/j.fsigen.2015.09.012
1872-4973/

abundance and quality contradict conventional methodology that recommends sample RNA integrity (RIN) to be at least RIN 8 or above to ensure sensitive and specific target detection [24–27]. However, this is unavoidable in situations such as forensic, clinical, formalin-fixed paraffin-embedded (FFPE) and environmental sampling where real-world samples must be analyzed. The negative effects of RNA degradation on RNA detection and quantification are well documented [25,28–31]. Despite recent success demonstrated through RNA massively parallel sequencing (MPS) of degraded forensic RNA [32], currently there is no clear solution to this problem except to avoid analysing degraded RNA.

Here we present a new method for the sensitive and specific detection, and amplification of RNA markers from degraded forensic RNA samples. We postulated that the regions of high read coverage within target transcripts represent stable regions within the target transcripts; we therefore designated these regions as stable regions or "StaRs". Using RNA sequencing (MPS) read alignments of degraded body fluids, we are able to observe regions within an RNA transcript that are more stable and target these stable regions (StaRs) as the basis for target detection within a significantly degraded sample.

2. Materials and methods

2.1. Body fluid sampling and aging (RNA degradation)

Samples used for PCR assessment of *Histatin 3* (*HTN3*), *Matrix metallopeptidase 11* (*MMP11*) and *Ubiquitin-conjugating enzyme E2D 2* (*UBE2D2*) StaR primer design were those samples used for RNAseq using massively parallel sequencing (MPS) and described by Lin et al. [32]. Fresh body fluid samples (oral mucosa/saliva (buccal), circulatory blood, vaginal material and menstrual blood) were collected on sterile Cultiplast® rayon swabs and aged at room temperature with exposure to ambient laboratory conditions, for $t = 0$, two and six weeks. Samples were collected from two individuals (one male, one female) for circulatory blood and buccal and from one individual for menstrual blood and vaginal material. Oral mucosa/saliva, vaginal material and menstrual blood samples were obtained by self-swabbing (by the participants themselves) while 50 μL of fresh circulatory blood was drawn using a sterile ACCU-CHEK® Safe-T-Pro Plus lancet (Roche Diagnostics USA, Indianapolis, IN, USA) and deposited onto each swab.

A separate set of fresh circulatory blood (50 μL) was drawn from two individuals using a sterile ACCU-CHEK® Safe-T-Pro Plus lancet (Roche Diagnostics USA, Indianapolis, IN, USA) and deposited onto each sterile Cultiplast® rayon swab. Swabs were dried under ambient laboratory conditions then stored frozen for 16 and 19 days respectively until extraction. Menstrual blood from two individuals was obtained by self-swabbing (by the participants themselves) and swabs were dried under ambient laboratory conditions then stored frozen for 7 and 13 days and one sample was stored frozen (−20 °C) for 12 days. Body fluid and storage conditions are detailed in Table 3. These samples were used for reproducibility testing.

2.2. RNA extraction

RNA extraction methodology is as described by Lin et al. [32]. Total RNA for all samples was extracted using the Promega® ReliaPrep™ RNA Cell Miniprep System (Promega Corporation, Madison, WI, USA) following the manufacturer's instructions. DNA was removed from extracted RNA using on-column DNase I treatment during the RNA extraction process. RNA was eluted in 50 μL elution buffer. Removal of human DNA to levels beyond detection was verified using the Quantifiler® Human DNA quantification kit (Life Technologies Corp., Carlsbad, CA, USA) using 1 μL of RNA in a 12.5 μL reaction. If human DNA was detected, samples were re-treated with Turbo™ DNase (Life Technologies Corp., Carlsbad, CA, USA) and re-quantified.

2.3. Library preparation and sequencing

cDNA libraries for RNAseq were prepared using Bioo Scientific NEXTFlex directional RNA-seq Kit (dUTP-Based) v2 48 (Bioo Scientific, Austin, TX, USA), as described by Lin et al. [32].

Barcoded libraries were sequenced across three lanes on an Illumina HiSeq2500 sequencer, with 2 × 100 bp paired-end chemistry.

2.4. Bioinformatics analysis

Bioinformatics analysis was undertaken as described by Lin et al. [32]. Sequencing read quality for all samples was analyzed using SolexaQA [33]. Data were preprocessed using DynamicTrim v1.9 [33]. Data were length-sorted and unpaired reads discarded using Lengthsort v1.9 [33]. Subsequent processed data consisted entirely of reads with <5% probability of error (or a Q score of >13), with pairs, and length >25 bp.

Tophat2 v2.0.12 [34] was used to align reads to the human genome hg19 (GRCh37) [35]. The "UCSC genes" annotation track of known genes was downloaded from the UCSC genome browser as hg19_UCSC_genes.gtf [35].

Final read alignments and transcript StaRs were viewed using Geneious v5.6.7 (Biomatters Ltd., Auckland, New Zealand). As Tophat2 is a splice-aware read alignment software, reads that span exon–exon junctions are split up. Geneious software v5.6.7 automatically stretches the existing exonic read coverage across the intronic regions to the next exon to continue to display read coverage. Therefore it may appear that there is read coverage across intronic regions, but there are actually no read sequences aligned within introns. Very occasionally a short read may be ambiguously (errant) mapped into an intronic region, but it is always an outlier.

2.5. cDNA synthesis

Complementary DNA used for PCR assessment of *HTN3*, *MMP11* and *UBE2D2* StaR primers was synthesized in a 40 μL reaction

Table 1
Primers designed using conventional methods.

Marker	Accession	Primer type	Forward primer (5′–3′)	Reverse primer (5′–3′)	Final concentration (μM)	Annealing temperature (°C)
Solute carrier family 4 (anion exchanger), member 1 (Diego blood group) (SLC4A1)	NM_000342.3	Conventional	CCTATACGCTTCCTCTTTGTGTT	TCTATGCGGAACACCCTCTC	0.1	65
Histatin 3 (HTN3)	NM_000200.2	Conventional	GGGGCATGATTATGGAGGTTT	CAGAAACAGCAGTGAAAACAGCTT	0.25	58
Ubiquitin-conjugating enzyme E2D 2 (UBE2D2)	NM_003339.2	Conventional	AATGATCTGGCACGGGACC	ATCGTAGAATATCAAGACAAATGCTGC	0.0125	58