Stafford-Allen, B, Dawnay, N, Hanson, EK, Ball, G, Gupta, A, Blackman, S, French, DJ, Duxbury, N, Ballantyne, J and Wells, S

Development of HyBeacon® probes for specific mRNA detection using body fluids as a model system.

Development of HyBeacon® probes for specific mRNA detection using body fluids as a model system

Beccy Stafford-Allen[a], Nick Dawnay[b], Erin K. Hanson[c], Glyn Ball[a], Ambika Gupta[d], Stephen Blackman[a], David J. French[e], Nicola Duxbury[a], Jack Ballantyne[c,f], Simon Wells[a].*

[a] Product Development Group, LGC Ltd, Culham Science Centre, Abingdon, OX14 3ED, UK

[b] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK

[c] National Center for Forensic Science, PO Box 162367, Orlando, FL 32816-2367, USA

[d] Department of Pharmacy and Forensic Science, King's College London, Faculty of Life Sciences and Medicine, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK

[e] Product Development Group, LGC Ltd, Queens Road, Teddington, TW11 0LY, UK

[f] Department of Chemistry, University of Central Florida, PO Box 162366, Orlando, FL 32816-2366, USA

** Corresponding author (Address correspondence to this author as detailed in 1, Email: paradna@lgcgroup.com)*

Abstract: HyBeacons are linear oligonucleotides which incorporate fluorescent dyes covalently linked to internal nucleotides. They have previously been used with PCR and isothermal amplification to interrogate SNPs and STRs in fields as diverse as clinical diagnostics, food authentication, and forensic DNA profiling. This work explores their use for the identification of expressed gene sequences through mRNA profiling. The use of mRNA is becoming increasingly common in forensic casework to identify body fluids on evidence items, as it offers higher specificity

and fewer false positives than current chemical presumptive testing methods. The work presented here details the development of a single-step one-tube RT-PCR assay to detect the presence of body fluids of forensic interest (saliva, blood, seminal fluid, vaginal fluid and menstrual blood) using HyBeacon® probes and melt curve analysis. Each assay shows a high degree of specificity to the target body fluid mRNA suggesting there is no requirement to remove genomic DNA prior to analysis. Of the five assays developed, four were able to detect between 10 and 100 copies of target cDNA, the fifth 1000 copies of target. The results presented here demonstrate that such an approach can be optimised for non-expert users and further areas of work are discussed.

Keywords: HyBeacons, body fluid, mRNA, RT-PCR, forensic

Introduction

Recent research has shown that HyBeacon probes offer a flexible and robust approach to nucleotide sequence detection across a variety of applications [1-8]. The versatility of the probe comes from the design which is specific to a complimentary target sequence. When hybridised to single-stranded target DNA they emit greater amounts of fluorescence than when un-bound. Detection is performed using melt curve analysis, and the temperature at which the probe dissociates from the target is determined by the degree of complementarity between the probe and the sequence to which is it bound, and can easily span a range of 30°C for the detection of mismatched or partially complementary target sequences [4,5]. This DNA based approach has allowed scientists to compare and match samples in a large number of applications ranging from Single Nucleotide Polymorphism

(SNP) analysis (Figure 1A) in food and medical diagnostic applications [7,8] to Short Tandem Repeat (STR) profiling (Figure 1B) in the analysis of forensic samples [4,5]. However, the detection of expressed gene products such as mRNA sequences, rather than DNA, is becoming increasingly important in a variety of fields. Research looking to measure and detect mRNA expression patterns in tissues often have a health focus such as disease diagnostics [9,10], susceptibility [11] and treatment [12], although there are other non-health applications which have no need to identify gene mutations or expression level and are simply concerned with the provenance of a biological sample.

Body Fluid Identification (BFID) forms part of the field of forensic genetics [14]. Investigations can often require activity level or cell source information which can allow the linking of a downstream DNA profile to a body fluid (and therefore an individual or action), or to allow discrimination between two versions of an event [13, 14]. Currently this information is mainly acquired using chemical tests (such as the Kastle-Meyer test for blood [15]) or microscopy (in the case of the identification of sperm [15]) to identify a body fluid. However, these tests can be time consuming, require expert interpretation, use hazardous chemicals, and are subject to a number of false positives. In addition such tests are generally not human-specific, although antibody based tests that have fewer false positives do exist for some sample types [16,17]. Research in this field has seen a steady progression towards assessing and understanding the utility of messenger RNA (mRNA), DNA methylation profiling [18-20], micro RNA (miRNA) [21-23] and microbial markers [14] to confirm the presence of forensically relevant body fluids on evidence items, such as swabs from sexual assault kits [14]. Today mRNA detection is an extensively researched method and a number of mRNA markers for forensically-relevant body fluids, such as saliva, seminal fluid, blood, menstrual blood and cervicovaginal fluid (CVF) have been identified [24-30]. A selection of these identified mRNA markers has been tested through European DNA Profiling Group (EDNAP) exercises [31-35]. Increasingly, forensic laboratories are beginning to offer mRNA profiling as a routine part of casework services [36, 37]. While these developments address the specificity and sensitivity issues of

simpler detection methods, new issues arise out of the more complex lab procedures required to isolate mRNA and remove any contaminating genomic DNA (gDNA), generate complementary DNA (cDNA) through Reverse Transcription (RT), amplify the resulting cDNA, and then differentiate the fragments, usually accomplished by capillary electrophoresis (CE) or high resolution melting (HRM)[38].

The specificity of a HyBeacon probe to its complementary sequence and its detection using melt curve analysis may solve many of the processing issues currently encountered by laboratories performing this service, and also serve to demonstrate the wider applicability of HyBeacon detection to mRNA. Positioning the probe such that it spans an exon:exon junction in mature mRNA (see Fig 1c) will allow differentiation between gDNA and mRNA. Where the intron is present in the gDNA, the probe will hybridise to a reduced number of nucleotides, resulting in a melt peak with a lower melting temperature (T_m) than one where the probe is fully hybridised to the target sequence. This allows the specific detection of mRNA targets in a sample where the gDNA is still present. The development of a one step approach to RT-PCR followed by melt curve detection, without further sample manipulation, further simplifies this process, increasing the usability for non-specialists.

The aim of this work was to develop a one-step RT-PCR process that would allow for the identification of body fluids from RNA extracted samples with gDNA still present, and that could be performed on a standard PCR and fluorescence detection platform. The single-step process would reduce the cost and time required for a result and would also allow for the RT and PCR steps to occur in a single tube, reducing the complexity of the process. The eventual scope of this work is to develop a simple system where a user can directly sample from a crime stain of interest and identify the body fluid present without any further manipulation of the sample or extract, similar to the generation of a DNA profile directly from evidence items using the ParaDNA Intelligence Test [4].

Here we present data on the initial assessment of the utility of HyBeacon probes for use in mRNA gene expression detection using the forensically relevant system of body fluid identification.

Materials and methods

Samples used in study

Swabs of relevant body fluids were donated by volunteers after full informed consent was obtained following procedures approved by ethics review board. Vaginal and menstrual material was collected on low vaginal swabs using Bode SecurSwabs (Cat no: P08D72, Bode Technology, VA, USA). Blood, saliva and ejaculate were collected onto cotton-tipped swabs (Cat no: 11502483, Fisher, UK). All donations were anonymised at the point of collection by the donors and stored at -20°C until required. Total RNA was extracted from swabs using QIAamp RNA Blood Mini Kit (Cat no: 52304, Qiagen, Manchester, UK) following the manufacturers recommended conditions. Extracts were treated with DN*ase* following manufacturer's instructions during extraction to remove contaminating gDNA (RNase-free DNase set Cat no: 79254, Qiagen, Manchester, UK). Extracts were eluted in 50 µl of RNase-free water and stored at -20°C until needed. Total RNA concentration for extracts was determined using a NanoDrop® ND-1000 (ThermoScientific, UK). Several of the body fluids tested in this work include those with a high microbial load (saliva, cervicovaginal fluid, menstrual blood), so any co-purified microbial RNA will contribute to the total RNA concentration determined. Panels of purified DNA from human lymphoblast transformed cell lines were used as the gDNA source (Public Health England, Salisbury, UK). The gDNA samples were RN*ase* treated following manufacturer's instructions to ensure any non-specific amplification observed using the HyBeacon probe could be identified as gDNA (RN*ase* A Cat no: 19101, Qiagen, Manchester, UK).

Primer and HyBeacon® probe design:

Multiple primer sets were designed for each of the target mRNAs using AmplifX 1.7.0 software [39]. *In silico* specificity checks were performed using Primer-BLAST [40]. Primers were designed to flank an exon:exon junction and were supplied by Eurofins (Eurofins MWG Synthesis GmbH, Germany). Once optimal primer pairs had been selected, HyBeacon probes (Table 1) were designed and labelled with two fluorescein (FAM) dT monomers (Glen Research, Virginia, USA). Probes were supplied by LGC Biosearch Technologies, CA, USA. All oligonucleotides were supplied lyophilised, and were reconstituted in low-EDTA TE (IDT, Leuven, Belgium) and quantified using a NanoDrop® ND-1000.

Characterisation studies

RT-PCR was performed using a CFX96 Real-Time Detection System (Bio-Rad Laboratories) in Framestar® 96 well low-profile non-skirted qPCR plates sealed with optically clear caps (Cat nos: 4ti-0721 and 4ti-0751, 4titude, Surrey, UK). Qiagen OneStep RT-PCR kits (Cat no: 210212, Qiagen, Manchester, UK) were set up according to manufacturer's specifications, with asymmetric primer concentrations of 1 µM excess primer, 0.25 µM limiting primer and 0.3µM HyBeacon probe, and final reaction volume of 20µl. Sample volumes were two µL per well. Total input amounts of the different targets are detailed below. The RT-PCR conditions for this study were as follows: 50°C for 30 minutes, 95°C for 15 minutes, PCR cycling 94°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec, with a plate read at 54°C for 10 sec every cycle for real-time PCR when required by experimental design for assay evaluation (total cycle number = 50). Melting curve analysis was performed immediately after RT-PCR using a hold at 72°C for 10 minutes, cool to 30°C at 0.1°C/sec, melt 30°C to 80°C at 0.5°C/sec including a plate read.

The specificity of the assay to its target body fluid was assessed by performing triplicate RT-PCR amplification of mRNA extract from all other body fluids considered as part of this work, as well as extracted gDNA. Where the extract concentration was above 10ng/µl (CVF and menstrual extracts

only), samples were diluted to this concentration with DEPC-treated water (Cat. No: 95284, Sigma) before being added to reactions. Sample volume of two µl was added to the reactions, so the amount of total RNA being added to a reaction varied between 20ng and 0.8ng. The mRNA target would make up only a small proportion of this amount.

The sensitivity of the assay to its target body fluid was assessed by amplification of serial dilutions of plasmids of a known copy number, each containing a cDNA gene sequence fragment for an mRNA target. Plasmid construction for each target was outsourced (GenScript USA Inc, NJ, USA) whereby a 310-350bp cDNA gene sequence fragment was inserted into pBluescript II SK(+) . The sequence fragment for each target included the entire length of the target amplicon, including primer sites, so that the same reaction mix set-up could be used for plasmid work as for all other input template types. Plasmids were resolvated with DEPC-treated water (Sigma-Aldrich, Poole, UK) to a standard stock concentration in copies/µL, calculated from the manufacturer's stated yield in ng and the molecular weight of the plasmid. The use of a known number of copies of plasmid added to the reaction allowed an estimate of the total number of cDNA copies transcribed after RT-PCR to be determined (noted in Table 2) – this assumes 100% RT-PCR efficiency.

Variation between donors was assessed by extracting total RNA from multiple donors' target body fluid swabs (peripheral blood – 8 donors; saliva – 5 donors; seminal fluid – 7 donors; menstrual blood - 4 donors; cervicovaginal fluid – 5 donors). Total RNA from each extract was quantified and where the concentration exceeded 10ng/µl (CVF and menstrual swabs only), extracts were diluted to 10ng/µl using DEPC-treated water. For each body fluid assay, 2µl samples of the extracted target body fluid swab from each donor were amplified in duplicate. The peripheral blood marker (ALAS2) was tested with both peripheral blood and menstrual blood extracts as one donor was only represented by a menstrual sample. The vaginal secretions marker (CYP2B7P1) was likewise tested with both vaginal and menstrual extracts as it was expected to be present in both, and one donor was only represented by a menstrual swab sample.

SNP Screening

To identify and assess the impact of SNPs for each of the five mRNA markers, putative SNP sites were identified within the probe binding sites within each gene. This was done in two ways; firstly, the GenBank database was searched for currently identified genetic variations within each marker; secondly, 24 gDNA samples underwent DNA sequencing across the regions of interest. PCR product was generated using AmpliTaq Gold 360 DNA Polymerase (Cat no: 4398823, Applied Biosystems) at 1x, 25 µL reaction volume, 2.5mM $MgCl_2$, final primer concentrations of 0.5µM. The optimised PCR protocol was run on a CFX96 as follows: 95°C for 10 minutes, PCR cycling conditions 95°C/30 sec, 55-65°C/30 sec, 72°C/1 minute, total of 35 cycles. Final hold at 72°C for 7 mins. The resulting PCR product underwent ExoSAP clean up (Affymetrix, High Wycombe, UK), and was amplified using BigDye® Terminator v3.1 Cycle Sequencing Kit (LifeTechnologies) using the LabCycler (SensoQuest). CE was performed on an ABI 3730xl DNA Analyzer XL (LifeTechnologies) using polymer pop7 by LGC Genomics (Berlin, Germany). Sequence Data was examined using Chromas Lite 2.1.1 (Technelysium Pty Ltd).

If SNPs were identified under HyBeacon probe binding sites, reverse compliment (RC) oligos for the probe sequences were synthesised that included one of the identified SNPs (Eurofins MWG Synthesis GmbH, Germany). These were then added to enzyme-free reactions with the probe for the marker of interest and melting curve analysis was performed. Six replicate tests were performed for each RC target and any differences in peak T_m or height were noted and statistically examined for significance using Mann-Whitney U tests and Student's t-tests. While the conditions of the melt reaction did not precisely mimic the optimised RT-PCR conditions they provide an approximate assessment of the likely impact a genetic variation has on melt peak temperature.

Results and Discussion

HyBeacon Assay Designs

Previous research has identified a number of mRNA markers specific to body fluids of forensic relevance (summarised in Table 1). After an initial screen of primer candidates based on real-time PCR quantification cycle ($C_{q)}$ values and final melt peak appearance (data not shown), the primer combinations were reduced to a single forward and reverse primer (Table 1). After primers were selected, HyBeacon probes were designed for each of the markers to anneal across exon:exon junction regions. This enabled a design with 100% homology between the HyBeacon probe and mRNA strands with reduced homology between the HyBeacon and gDNA strand containing the intron (as shown in Figure 1c). It was expected that this reduced homology would cause a reduction in the melt temperature of peaks associated with gDNA, thereby providing a categorical differentiation between target mRNA and secondary gDNA. A first requirement of the HyBeacon assay design was that the melt curve generated from any target sample mRNA input was clearly differentiated from any melt curves generated from any co-amplified gDNA product. In addition, the gDNA fragment is expected to be larger than the target mRNA, due to the inclusion of the intron, leading to a preferential amplification of the smaller mRNA target. In all tests amplification of the target and the presence of a melting peak at the expected T_m (60°C for all markers) was seen in all positive control samples (extracted RNA and cDNA). In addition, the gDNA fragment would be larger than the target mRNA, due to the inclusion of the intron, potentially leading to preferential amplification of the smaller mRNA target. A small, lower T_m gDNA peak was identified in ALAS2, SEMG1, MMP10 and CYP2B7P1 assays (Figure 2, Table 2).

No gDNA peak was produced in the HTN3 assay. This is likely because the structure of the HTN3 mRNA allowed both primers as well as the probe to lie over at least one exon:exon junction each (one of which exceeds 2000bp), which strongly discourages the amplification of gDNA during PCR.

214

215 *Characterisation studies*

216 The second requirement of the HyBeacon assay was for the markers to be specific to the body fluid
217 in question. SEMG1, HTN3 and MMP10 peaks were only seen from the expected body fluid mRNA
218 template (Table 1). Peaks were seen in ALAS2 reactions for both peripheral and menstrual blood
219 mRNA (Figure 3), which is expected as peripheral blood is a component of menstrual blood. Peaks in
220 CYP2B7P1 reactions were also seen from menstrual mRNA, which again is expected as vaginal
221 material is present in menstrual blood. For both of these markers, purified mRNA from menstrual
222 blood was a better source of template (based on the lower mean C_q values seen) than purified
223 mRNA from either peripheral blood (for ALAS2 mean C_q 35.0 for peripheral blood, 33.6 for
224 menstrual) or vaginal fluid (for CYP2B7P1 mean C_t 41.42 for vaginal secretions, 31.33 for menstrual).
225 As all template types for one assay were run on the same plate, this is probably indicative of
226 differences in expression levels of different mRNA species between donors and tissues, or an
227 indication of the quality of purified mRNA that can be obtained from these sources. No peaks were
228 observed in either assay from the remaining body fluid types tested.

229 Forensic samples can be small in size or low in template, so it is important that any assay developed
230 is able to provide a result with a small amount of input template. The performance of the assays
231 developed in this work will be dependent on a number of factors: the efficiency of the assay itself,
232 differences in expression of the target marker in the sample being tested (which can vary between
233 donors, and within the same donor over time), and (particularly if crude sample analysis is required),
234 the presence of inhibitors. Primer and probe sets were initially tested for sensitivity using serial
235 dilution of plasmid targets to remove the impact of differences in amount of target present between
236 total mRNA extracts from samples of the same body fluid. All assays were run in duplicate against
237 10-fold dilutions from 100,000 copies/well to 1 copy/well. The limit of detection (LoD) was defined
238 as the input amount that gave discernible peaks (by eye) in both replicate samples (LoD ALAS2 = 10

copies, MMP10 = 10 copies, CYP2B7P1 = 1000 copies, HTN3 = 100 copies, SEMG1 = 10 copies). It is unclear why the CYP2B7P1 assay has a significantly higher LoD than the other assays, but it is thought this is related to lower primer efficiency which forms part of an ongoing study.

Variations between donors

It was necessary to determine if differences in gene sequences between donors (e.g. SNPs, or splice variants) would lead to differences in melt peak T_ms or assay performance. Expression of the markers of interest was likely to vary between donors, and within the same donor over time (e.g. expression of MMP10 during menstrual cycle).

There was a maximum of 3°C difference between melt peak T_ms across all donors tested within each assay (Table 2), and the spread of T_ms was only marginally greater than the differences seen in T_m between repeats from the same donor and was not statistically significant, suggesting that instrument variation played more of a role in peak T_m than donor-to-donor variation. All extracts tested against the marker panel gave the expected results. This suggests that, within this fairly limited donor pool, there were no SNPs or splice variants within the probe annealing regions, and no donors who did not express the target mRNA at sufficient levels for detection.

qPCR C_q values for samples varied between donors, but were consistent between repeats from the same donor extract (Table 2). CYP2B7P1 had the widest spread of C_q values across all of the samples tested possibly due to poor amplification of one of the vaginal samples. Further work is required to determine if this variation in performance is due to varying expression levels of CYP2B7P1 during the menstrual cycle, or sample degradation after collection.

SNP Screening

HyBeacon® melt analysis relies on complementarity between the probe and the target sequence. Most of the markers chosen for this work are based on genes which are expressed into functional proteins. The exception is the CYP2B7P1 marker, which is classed as a pseudogene as, although mRNA is transcribed from the genome, it is not further translated into protein [30]. Many of these mRNA species are believed to be functional, although the function, if any, of the CYP2B7P1 mRNA has yet to be determined [30]. Given the highly conserved nature of such coding regions it was expected that the number of SNPs observed in the screening study would be low. A sample screen would identity any SNPs within probe target sequences that would lead to changes in peak temperature, which would complicate the use of automated software to determine the presence or absence of the marker of interest based on melt peak T_ms. Another reason to screen for SNPs was the possibility that there may be disease traits linked to the variations in the gene sequences. For the purposes of the body fluid detection approach it would be important to avoid the release of sensitive health information during the analyses of a crime scene sample.

GenBank data was interrogated to determine if reported SNPs existed within probe sites. Details of the numbers of SNPs identified for each marker and within each probe region are given in Table 3. None of the SNPs identified in the GenBank data were associated with disease states at the time of analysis. More than one SNP variant was identified at one locus under the HTN3 probe; this is highlighted in Table 3. There was no suggestion in the GenBank data of linkages between the SNPs of interest to our analysis, although there was no frequency data associated with any of the SNPs identified in the GenBank data. Data from the variation between donors study above suggested that none of the individuals tested possessed SNPs within the probe sites that affected probe T_m, as there was little variation in the melt peak T_ms. However, this pool of donors was relatively small and non-diverse, so sequencing of commercially available gDNA samples from 24 individuals from more ancestrally diverse populations was undertaken. Samples were taken from the panels of purified human lymphoblast transformed cell lines provided by Public Health England (Salisbury, UK). Samples were taken as follows: 10 random individuals from the UK Caucasian donors on the Human

Random Control DNA Panel HRC-5, two donors from each of the following ethnic groups from the Ethnic Diversity DNA Panel EDP-1; Japanese, Aborigine Australian, Thai, Oriental, Black African, Ashkenazi Jew and South American Indian. Populations were defined by the supplier of the DNA panels. This approach was practically easier than generating data directly from mRNA collected from multiple individuals given the personal nature of the samples and the scope of populations offered by using the EDP-1 panel. Where multiple introns occurred in the gDNA, a number of flanking primer sets were designed to amplify each region, with junctions identified and matched together afterwards. None of the samples sequenced contained any SNPs under probe binding sites (data not shown).

In order to determine the impact of the presence of any SNPs within the probe binding sites, RC oligos were generated as detailed in Table 3, and probes were melted with these RC templates. Melt curve data resulting from analyses of probes with RC sequences showed that in all instances the wild type RC/probe combination had the highest peak T_m (Table 4). Subsequent statistical analysis of the wild type and non-wild type peak T_ms (Mann-Whitney U test) and peak height (student t-test) indicated that the presence of a SNP within the probe annealing region led to a statistically significant decrease in melt peak T_m in all instances ($p \leq 0.05$). The maximum difference seen was a decrease of 7.1°C (CYP2B7P1 RC1). The CYP2B7P1 polymorphisms observed are located towards the centre of the probe, and are therefore more destabilising. RC1 is the most destabilising mismatch (C/A, C in the probe) whereas RC2 is an intermediate (T/C, T in the probe), and RC3 an even less destabilising mismatch (T/G, T in the probe). The HTN3 and SEMG1 SNPs are located towards the probe termini and are therefore less destabilising. Although the T_m decreases recorded from the SNPs were greater than the T_m variation observed between donors in the previous testing, none of the shifted peak T_ms corresponded to gDNA peaks seen in previous testing, and so they would not be confused with them when examining results from real samples. Due to a lack of population data associated with the reported SNPs we identified, it is not possible for us to determine the likelihood that these SNPs will be encountered if an expanded number of donors were to be tested. However,

314 if further testing suggests they are likely to be seen, they can be offset by incorporation of universal
315 bases if this is required. The use of universal bases is unlikely to prevent target detection. There was
316 no statistically significant difference in the peak heights from any of the RC oligos compared to the
317 wild type (t-test, $p \geq 0.05$).

318

319 Summary

320 The data presented demonstrates the feasibility of a HyBeacons® approach to RT-PCR detection of
321 mRNA using forensically relevant body fluids as a working example. The assays developed here have
322 sufficient sensitivity and specificity to detect body fluid mRNA extracted from swabs and to
323 differentiate between mRNA and gDNA through differences in melt peak T_m. Furthermore, this
324 research has demonstrated the potential of combining RT-PCR and HyBeacon probes for detecting
325 SNPs in expressed genes in the presence of gDNA, offering another molecular detection approach in
326 health related studies.

327 In their current form the assays could be slotted into existing forensic workflows to test extracted
328 nucleic acid material containing both gDNA and mRNA. The use of HyBeacon probes also offers the
329 possibility to multiplex reactions together and differentiate between targets using different dye
330 labels. This work will form a large part of the future development of this approach. These assays
331 were quick to design and offer a more rapid, sensitive and specific approach than many of the
332 current body fluid detection methods in use. However, it is important to note that some mRNA
333 markers identified in the wider literature are also expressed at a low level in non-target tissues [14].
334 Traditional RT-qPCR approaches are able to use C_q values to distinguish between high expression of
335 target mRNAs in target tissues versus lower expression in non-target tissues. However, primer
336 design, cycling conditions and enzyme choice can also significantly impact the sensitivity of an assay
337 to these lower-level expression profiles, and may account for some of the variation in reporting of

specificity in the existing literature. Expanding the specificity testing of the assays developed here to non-target body fluids that may be encountered in routine forensic case work (such as urine, tears, nasal mucosa or sweat [41]) will be an essential part of any future validation work. Any unexpected results suggesting low-level expression of markers in non-target tissues will be investigated, and assay designs can be altered to prevent the detection of low-level expression if required.

Acknowledgements

The authors would like to thank Monika Panasiuk and Rebecca Horrell for their assistance in reagent preparation. They would also like to thank the body fluid donors, without whom this work would not have been possible. This material is based upon work supported by the U. S. Army Research Office and the Defense Forensic Science Center under contract number W911NF-14-C-0097. The information contained within this report does not necessarily reflect the position or the policy of the Government, and no official endorsements should be inferred. The sponsor had no role in the study design; the collection, analysis or interpretation of data; the writing of the report; the decision to submit the article for publication.

Conflicts of interest

None.

Author contributions

All authors confirm that they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) critically important intellectual contribution to the conception,

359 design and/or analysis and interpretation; (b) drafting the manuscript or critically reading it; and (c)
360 through reading and final approval of the version to be published.

361

362 References

363 [1] D. French, C.L. Archard, T. Brown, D. McDowell. HyBeacon probes: a new tool for DNA sequence
364 detection and allele discrimination. Mol Cell Probes 15 (2001) 363-374.

365

366 [2] D. French, C.L. Archard, M. Andersen, D. McDowell. Ultra-rapid DNA analysis using HyBeacon
367 probes and direct PCR amplification from saliva. Mol Cell Probes 16 (2002) 319-326.

368

369 [3] D. French, N. Gale, T. Brown, D. McDowell, P. Debenham. HyBeacon probes for rapid DNA
370 sequence detection and allele discrimination, Methods Mol. Biol. 429 (2008) 171-85.

371

372 [4] S. Blackman, N. Dawnay, G. Ball, B. Stafford-Allen, N. Tribble, P. Rendell, K. Neary, E. Hanson, J.
373 Ballantyne, B. Kallifatidis, J. Mendel, D.K. Mills, S. Wells. Developmental validation of the ParaDNA
374 Intelligence System – a novel approach to DNA profiling, Forensic Science International: Genetics 17
375 (2015) 137-48.

376

377 [5] N. Dawnay, B. Stafford-Allen, D. Moore, S. Blackman, P. Rendell, E. Hanson, J. Ballantyne, B.
378 Kallifatidis, J. Mendel, D.K. Mills, R. Nagy, S. Wells. Developmental validation of the ParaDNA
379 Screening system – a presumptive test for the detection of DNA on forensic evidence items, Forensic
380 Science International: Genetics, 11 (2014) 73-9.

[6] N. Dawnay, R. Hughes, D. Syndercombe Court, N. Duxbury. Species detection using HyBeacon probe technology: working towards rapid onsite testing in non-human forensic and food authentification applications. Forensic Science International: Genetics 17 (2016) 103-11.

[7] D. French, D. Jones, D. McDowell, J. Thompson, P. Debenham. Analysis of multiple single nucleotide polymorphisms closely positioned in the ovine *PRNP* gene using linear fluorescent probes and melting curve analysis, BMC Infectious Diseases 7:90 (2007).

[8] N. Ben Gaied, J. Richardson, D. Singleton, Z. Zhao, D. French, T. Brown. End-capped HyBeacon probes for the analysis of human genetic polymorphisms related to warfarin metabolism, Organic & Biomolecular Chemistry 8 (2010) 2728-2734.

[9] Dahan, L., Huang, L., Kedmi, R., Behlke, M.A. and Peer, D., 2013. SNP Detection in mRNA in Living Cells Using Allele Specific FRET Probes. *PloS one*, *8*(9), p.e72389.

[10] Zuccato, C., Ciammola, A., Rigamonti, D., Leavitt, B.R., Goffredo, D., Conti, L., MacDonald, M.E., Friedlander, R.M., Silani, V., Hayden, M.R. and Timmusk, T., 2001. Loss of huntingtin-mediated BDNF gene transcription in Huntington's disease. *Science*, *293*(5529), pp.493-498.

401 [11] Su, J., Heng, J., Huang, T., Peng, L., Yang, C. and Li, Q., 2012. Identification, mRNA expression and
402 genomic structure of TLR22 and its association with GCRV susceptibility/resistance in grass carp
403 (Ctenopharyngodon idella). *Developmental & Comparative Immunology*, *36*(2), pp.450-462.

404

405 [12] Fell, J.M., Paintin, M., Arnaud-Battandier, F., Beattie, R.M., Hollis, A., Kitching, P., Donnet-
406 Hughes, A., MacDonald, T.T. and Walker-Smith, J.A., 2000. Mucosal healing and a fall in mucosal pro-
407 inflammatory cytokine mRNA induced by a specific oral polymeric diet in paediatric Crohn's disease.
408 *Alimentary Pharmacology and Therapeutics*, *14*(3), pp.281-290.

409

410 [13] M. Bauer. RNA in forensic science, Forensic Science International: Genetics 1 (2007) 69–74.

411

412 [14] T. Sijen. Molecular approaches for forensic cell type identification: On mRNA, miRNA, DNA
413 methylation and microbial markers, Forensic Science International: Genetics 18 (2014) 21–32.

414

415 [15] N. Watson. Chapter 15: Analysis of body fluids, in: P. C. White ed. Crime Scene to Court: The
416 Essentials of Forensic Science 3rd Edition (2010) p438-467, RSC Publishing.

417

418 [16] Abacus diagnostics Hematrace kits: http://www.abacusdiagnostics.com/hematrace.htm

419

420 [17] N. Frascione, R. Thorogate, B. Daniel, S. Jickells. Detection and identification of body fluid stains
421 using antibody-nanoparticle conjugates, Analyst 137 (2012) 508-12.

[18] BA. Vidaki, B. Daniel, D. Syndercombe Court. Forensic DNA methylation profiling—Potential opportunities and challenges, Forensic Science International 7 (2013) 499–507.

[19] Hwan Young Lee, Ja Hyun An, Sang-Eun Jung, Yu Na Oh, Eun Young Lee, Ajin Choi, Woo Ick Yang, Kyoung-Jin Shin. Genome-wide methylation profiling and a multiplex construction for the identification of body fluids using epigenetic markers, Forensic Science International: Genetics, 17 (2015) 17–24.

[20] Jong-Lyul Park, Oh-Hyung Kwon, Jong Hwan Kim, Hyang-Sook Yoo, Han-Chul Lee, Kwang-Man Woo, Seon-Young Kim, Seung-Hwan Lee, Yong Sung Kim. Identification of body fluid-specific DNA methylation markers for use in forensic science, Forensic Science International: Genetics 13 (2014) 147–153.

[21] Zheng Wang, Di Zhou, Yandong Cao, Zhen Hu, Suhua Zhang, Yingnan Bian, Yiping Hou, Chengtao Li. Characterization of microRNA expression profiles in blood and saliva using the Ion Personal Genome Machine® System (Ion PGM™ System), Forensic Science International: Genetics 20 (2016) 140-146.

[22] E.K. Hanson, J. Ballantyne. Identification of Forensically Relevant Body Fluids Using a Panel of Differentially Expressed microRNAs. Anal. Biochem. 387 (2009) 303-314.

444 [23] Sarah S. Silva, Cátia Lopes, A.L. Teixeira, M.J Carneiro de Sousa, R. Medeiros. Forensic miRNA:
445 Potential biomarker for body fluids? Forensic Science International: Genetics 14 (2014) 1–10.

446

447 [24] J. Juusola, J. Ballantyne. Messenger RNA profiling: a prototype method to supplant conventional
448 methods for body fluid identification, Forensic Science International 135 (2003) 85-96.

449

450 [25] J. Juusola, J. Ballantyne. Multiplex mRNA profiling for the identification of body fluids, Forensic
451 Science International 152 (2005) 1-12.

452

453 [26] A. Kraus, D. Patzelt, M. Bauer. Detection of epithelial cells in dried blood stains by reverse
454 transcriptase-polymerase chain reaction, J. Forensic Sci 44 (1999) 1232-1236.

455

456 [27] M. Setzer, J. Juusola, J. Ballantyne. Recovery and stability of RNA in vaginal swabs and blood,
457 semen and saliva stains, J. Forensic Sciences 53 (2008) 296-305.

458

459 [28] Mara L. Lennard Richard, Kathryn A. Harper, Rhonda L. Craig, Anthony J. Onorato, James M.
460 Robertson, Joseph Donfack. Evaluation of mRNA marker specificity for the identification of five
461 human body fluids by capillary electrophoresis, Forensic Science International: Genetics 6 (2012)
462 452-60.

463

464 [29] P. Danaher, R. Lynn White, E. K. Hanson, J. Ballantyne. Facile semi-automated forensic body
465 fluid identification by multiplex solution hybridization of NanoString® barcode probes to specific
466 mRNA targets, Forensic Science International: Genetics 14 (2014) 18–30.

467

468 [30] E.K. Hanson, J. Ballantyne, Highly specific mRNA biomarkers for the identification of vaginal
469 secretions in sexual assault investigations, Sci. Justice 53 (2013) 14–22.

470

471 [31] C. Haas, E. Hanson, W. Bär, R. Banemann, A.M. Bento, A. Berti, E. Borges, C. Bouakaze, A.
472 Carracedo, M. Carvalho, A. Choma, M. Dötsch, M. Durianciková, P. Hoff-Olsen, C. Hohofn, P.
473 Johansen, P.A. Lindenbergh, B. Loddenkötter, B. Ludes, O. Maroñas, N. Morling, H. Niederstätter, W.
474 Parson, G. Patel, C. Popielarz, E. Salata, P.M. Schneider, T. Sijen, B. Sviezená, L. Zatkalíková, J.
475 Ballantyne. mRNA profiling for the identification of blood – results of a collaborative EDNAP exercise,
476 Forensic Science International: Genetics 5 (2011) 21-26.

477

478 [32] C. Haas, E. Hanson, M.J. Anjos, W. Bär, R. Banemann, A. Berti, E. Borges, C. Bouakaze, A.
479 Carracedo, M. Carvalho, V. Castella, A. Choma, G. De Cock, M. Dötsch, P. Hoff-Olsen, P. Johansen, F.
480 Kohlmeier, P.A. Lindenbergh, B. Ludes, O. Maroñas, D. Moore, M.-L. Morerod, N. Morling, H.
481 Niederstätter, F. Noel, W. Parson, G. Patel, C. Popielarz, E. Salata, P.M. Schneider, T. Sijen, B.
482 Sviežena, M. Turanská, L. Zatkalíková, J. Ballantyne. RNA/DNA co-analysis from blood stains – Results
483 of a second collaborative EDNAP exercise. Forensic Science International: Genetics 6 (2012) 70-80.

484

485 [33] C. Haas, E. Hanson, M.J. Anjos, R. Banemann, A. Berti, E. Borges, A. Carracedo, M. Carvalho, C.
486 Courts, G. De Cock, M. Dötsch, S. Flynn, I. Gomes, C. Hollard, B. Hjort, P. Hoff-Olsen, K. Hríbiková, A.

Lindenbergh, B. Ludes, O. Maroñas, N. McCallum, D. Moore, N. Morling, H. Niederstätter, F. Noel, W. Parson, C. Popielarz, C. Rapone, A.D. Roeder, Y. Ruiz, E. Sauer, P.M. Schneider, T. Sijen, D. Syndercombe Court, B. Sviežená, M. Turanská, A. Vidaki, L. Zatkalíková, J. Ballantyne. RNA/DNA co-analysis from human saliva and semen stains – Results of a third collaborative EDNAP exercise, Forensic Science International: Genetics 7 (2013) 230-239.

[34] C. Haas, E. Hanson, M.J. Anjos, K.N. Ballantyne, R. Banemann, B. Bhoelai, E. Borges, M. Carvalho, C. Courts, G. De Cock, K. Drobnic, M. Dötsch, R. Fleming, C. Franchi, I. Gomes, G. Hadzic, S.A. Harbison, J. Harteveld, B. Hjort, C. Hollard, P. Hoff-Olsen, C. Hüls, C. Keyser, O. Maroñas, N. McCallum, D. Moore, N. Morling, H. Niederstätter, F. Noël, W. Parson, C. Phillips, C. Popielarz, A.D. Roeder, L. Salvaderi, E. Sauer, P.M. Schneider, G. Shanthan, D. Syndercombe Court, M. Turanská, R.A.H. van Oorschot, M. Vennemann, A. Vidaki, L. Zatkalíková, J. Ballantyne. RNA/DNA co-analysis from human menstrual blood and vaginal secretion stains: Results of a fourth and fifth collaborative EDNAP exercise. Forensic Science International: Genetics 8 (2014) 203-212.

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

[36] R.F. Fleming, H.M.R. Baker, M.V. Fallow, P.M. Simon, J.K. Stacey, R.J. Wivell, S.A. Harbison. Impementation of messenger RNA body fluid testing in forensic case work. Conference proceedings oral presentations ISHI 22 (2011). https://www.promega.com/~/media/files/resources/conference%20proceedings/ishi%2022/oral%20presentations/fleming.pdf

[37] BA. Lindenbergh, P. Maaskant, T. Sijen. Implementation of RNA profiling in forensic casework. Forensic Science International: Genetics, 7 (2013) 159-166.

[38] E. Hanson, J. Ballantyne. Rapid and inexpensive body fluid identification by RNA profiling-based multiplex High Resolution Melt (HRM) analysis. F1000Research 2:218 (2013).

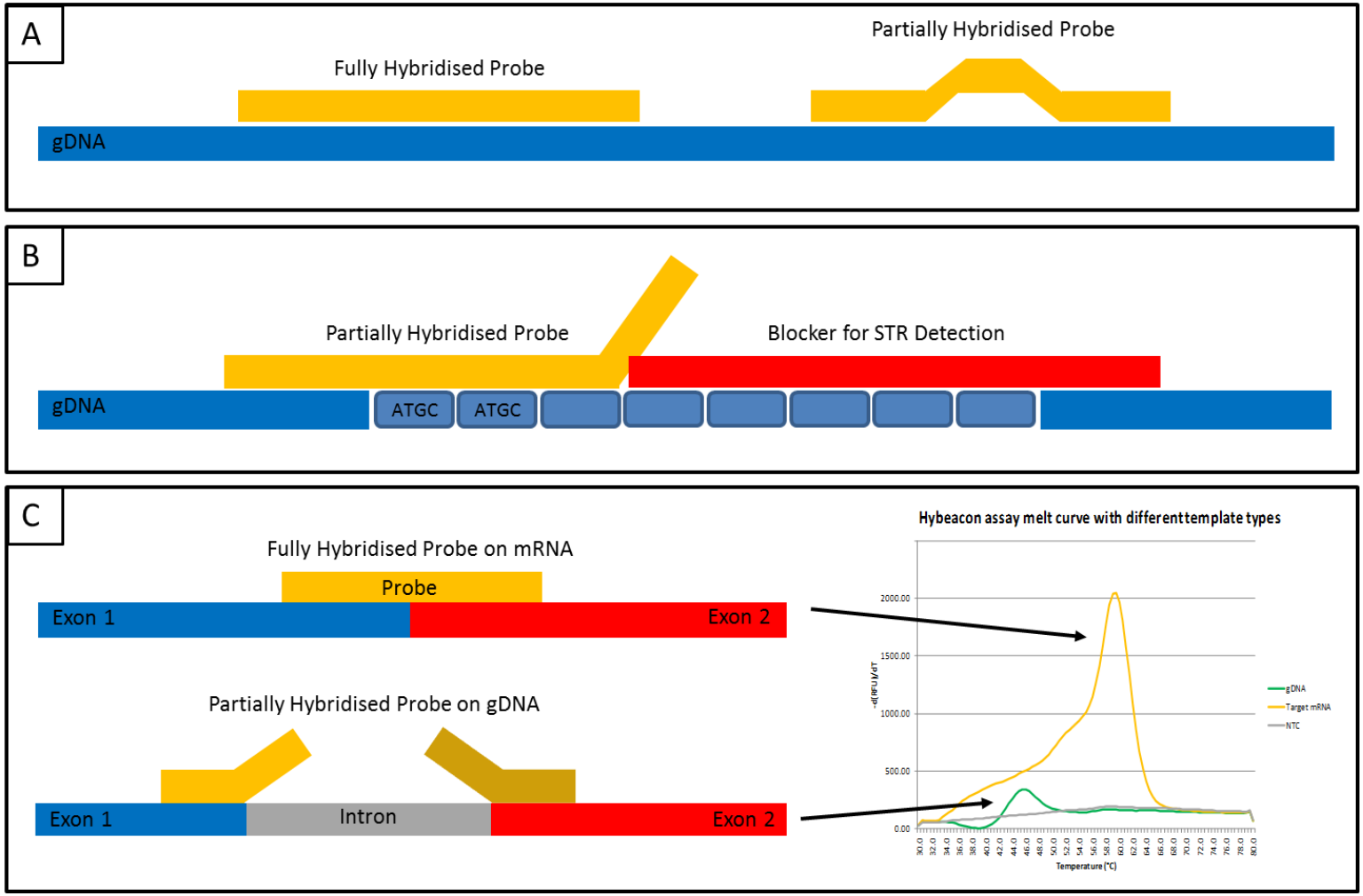
[39] AmplifX 1.7.0 by Nicolas Jullien, CNRS, Aix-Marseille Université. Available at: http://crn2m.univ-mrs.fr/pub/amplifx-dist. Accessed on 18 May 2015.

[40] Primer BLAST – A tool for finding specific primers. Available at: http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome. Accessed on 3rd June 2015.

[41] M. van den Berge, B. Bhoelai, J. Harteveld, A. Matai, T. Sijen. Advancing forensic RNA typing: On non-target secretions, a nasal mucosa marker, a differential co-extraction protocol and the sensitivity of DNA and RNA profiling. Forensic Science International: Genetics 20 (2016) 119-129.

532

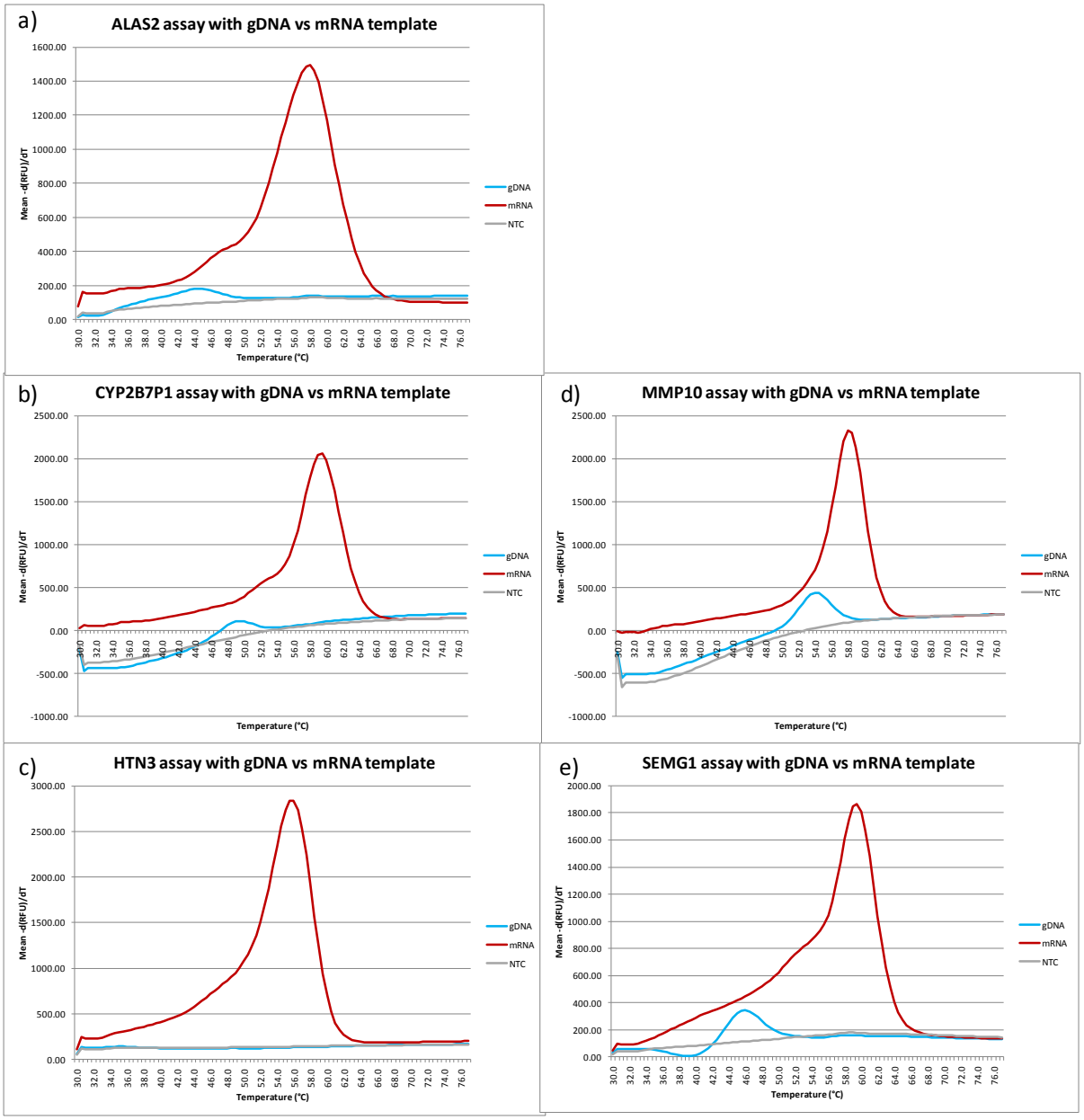
Figure 1: *Diagram showing different HyBeacon approaches for DNA identification A) SNP detection as used in medical and food authentication applications, B) STR detection with the use of a blocker oligo, and C) the proposed approach for detection and differentiation of gDNA and mRNA.*



536 ***Table 1:*** *Selected mRNA markers, primers and HyBeacon probe designs.*

mRNA marker	Target body fluid	GenBank Accession number	Oligo name	Sequence
ALAS2 *5'-aminolevulinate synthase 2*	Peripheral blood	NM_001037967.3	ALAS2 Fwd primer	GGCATGAGCCGACACCCTCAG
			ALAS2 Rev primer	CCTGAGATGTTGCGGGTGCCAC
			ALAS2 probe	CTGCAGGG-T(FAM)-CTCCTG-T(FAM)-GTGG-cap
CYP2B7P1 *Cytochrome P450, family 2, subfamily B, member 7, pseudogene*	CVF	NR_001278.1	CYP2B7P1 Fwd primer	CAAATCCTTTCTGAGGTTCCGAGA
			CYP2B7P1 Rev primer	GGTTTCCATTGGCAAAGAGCAT
			CYP2B7P1 probe	GCATGCCATA-T(FAM)-CCCTGG-T(FAM)-AGACT-cap
HTN3 *Histatin 3*	Saliva	NM_000200.2	HTN2 Fwd primer	TGGAGCTGATTCACATGCAAAGAGACAT
			HTN3 Rev primer	GCGAATTTGCCAGTCAAACCTCCATAATC
			HTN3 probe	GATG-T(FAM)-GAATGA-T(FAM)-GCTTTTCATGGA-cap
MMP10 *Matrix metalloproteinase stromelysin-2*	Menstrual blood	X07820.1	MMP10 Fwd primer	GTCACTTCAGCTCCTTTCCTGGCA
			MMP10 Rev primer	CTGTGTCCTGGGCCATCAA
			MMP10 probe	TTACATACAGGATTG-T(FAM)-GAATTA-T(FAM)-ACACCAG-cap
SEMG1 *Semenogelin I*	Seminal fluid	NM_003007.3	SEMG1 Fwd primer	CCAACATGGATCTCATGGGGGATTG
			SEMG1 Rev primer	AGCATGGGCAGGTGGTGTCAT
			SEMG1 probe	ACCAATATCAAGG-T(FAM)-GTCAGT-T(FAM)-GACC-cap

537



538

539 ***Figure 2:*** *Results from HyBeacon assays run with either target body fluid mRNA, purified gDNA or*

540 *NTC. Results shown are the mean of 3 repeats for each input type. Target body fluid mRNA was*

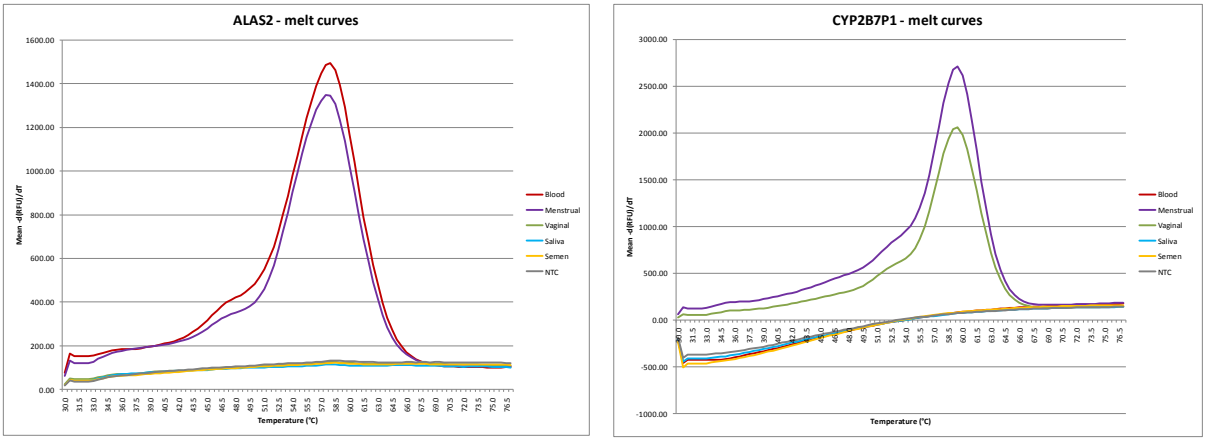
541 *peripheral blood (ALAS2, image a), vaginal fluid (CYP2B7P1, image b), saliva (HTN3, image c),*

542 *menstrual blood (MMP10, image d) or seminal fluid (SEMG1, image e).*

543

544

545



546

547 ***Figure 3:*** *Specificity of ALAS2 (blood marker) and CYP2B7P1 (vaginal marker) to target mRNA.*
548 *Template mRNA inputs are coloured as follows: red – peripheral blood, purple – menstrual blood,*
549 *green – vaginal secretions, blue – saliva, orange – seminal fluid.* C_q *values for the different template*
550 *inputs varied: ALAS2 peripheral blood (red trace, mean* C_q *32.37, s.d. 0.07), menstrual blood (purple*
551 *trace, mean* C_q *31.3, s.d. 0.14), CYP2B7P1 vaginal (green trace, mean* C_q *37.47, s.d. 0.53), menstrual*
552 *blood (purple trace, mean* C_q *30.47, s.d. 0.25). Means are calculated from 3 repeats of the same*
553 *extract.*

554

555 ***Table 2:*** *Mean peak* T_m*s and* C_q *values achieved from target body fluid extracts from multiple donors*
556 *and mean peaks from gDNA input. Italicised numbers in brackets below mean* T_m*s indicate standard*
557 *deviations. * indicates that one of the donors included is represented only by a menstrual swab*
558 *rather than a target body fluid swab. Donor swab extractions run in duplicate. Estimated number of*
559 *copies of target mRNA species in 1ng total RNA extracted from swabs estimated based on*
560 *comparison of* C_q *values with plasmid dilutions. Total RNA will include any microbial RNA extracted*
561 *from the swab – this will represent a considerable contribution in body fluids such as saliva,*
562 *cervicovaginal fluid and menstrual blood. As indicated in Figure 2, no gDNA peak is produced in the*
563 *saliva (HTN3) assay.*

564

mRNA target	Number of donors tested	Mean target peak T_m (°C)	Mean C_q from mRNA target	Mean gDNA peak T_m (°C)	Estimated # copies/1ng total RNA
ALAS2 *Peripheral blood*	9*	57.4 (*0.3*)	30.8 (*1.3*)	44.5 (*0.5*)	7409.0
CYP2B7P1 *CVF*	6*	58.8 (*0.3*)	33.2 (*3.4*)	49.0 (*0*)	12.3
HTN3 *Saliva*	5	55.7 (*0.8*)	33.0 (*1.7*)	n/a	16763.2
MMP10 *Menstrual blood*	4	57.3 (*0.3*)	33.5 (*2.0*)	54.1 (*0.29*)	178.5
SEMG1 *Seminal fluid*	7	59.1 (*0.5*)	28.9 (*1.7*)	45.8 (*0.58*)	1251.6

***Table 3**: SNPs identified within markers of interest from both GenBank data analysis and sequencing of individuals from diverse populations. Where multiple SNP variants have been identified at the same location the number of SNP sites is given, in addition to the total number of SNP variants identified (HTN3 only). More information on the location of the SNPs within the probe sequences is in Table 4.*

mRNA marker	Target fluid	Genomic location	Exonic amplicon length	Total number of SNPs identified within amplicon	SNPs occurring within probe sites
ALAS2	Blood	Xp11.21	92	2	0
MMP10	Menstrual blood	11q22.3	226	27	0
HTN3	Saliva	4q13	158	19	3 (2 loci)
SEMG1	Semen	20q12-q13.2	255	23	1
CYP2B7P1	Vaginal fluid	19q13.2	203	14	3

***Table 4**: Reverse complement oligonucleotides for testing of known SNPs at probe sites. Underlined sections in the wild type RCs indicate base situated opposite a fluorophore once the probe is bound. Bold bases in the RC sequence indicate SNPs. Mean melt peak T_m calculated from six replicates.*

Oligo Name	Sequence	Mean melt peak T_m (°C)
CYP2B7P1 Wild Type	5′ AGTCTACCAGGGATATGGCATGC 3′	**66.48**
CYP2B7P1 RC1	5′ ----------**A**---------- 3′	59.44
CYP2B7P1 RC2	5′ ----------**C**---------- 3′	61.68
CYP2B7P1 RC3	5′ ----------**G**--- 3′	64.10
HTN3 Wild Type	5′ TCCATGAAAAGCATCATTCACATC 3′	**61.56**
HTN3 RC1	5′ –**T**---------- 3′	59.29
HTN3 RC2	5′ ----------**A** 3′	60.55
HTN3 RC3	5′ ----------**T** 3′	60.65
SEMG1 Wild Type	5′ GGTCAACTGACACCTTGATATTGGT 3′	**64.78**
SEMG1 RC1	5′ ----------**A**- 3′	62.20

577

578