BIOLOGY Methods & Protocols

Biology Methods and Protocols, 2018, 1–12

doi: 10.1093/biomethods/bpy005
Methods Manuscript

METHODS MANUSCRIPT

An array-based melt curve analysis method for the identification and classification of closely related pathogen strains

Arjang Hassibi, Jessica Ebert, Sara Bolouki, Alexander Anemogiannis, Gelareh Mazarei, Yuan Li, Kirsten A. Johnson, Tran Van, Pallavi Mantina, Taraneh Gharooni, Kshama Jirage, Lei Pei, Ruma Sinha, Arun Manickam, Amin Zia, Pejman Naraghi-Arani, Gary Schoolnik and Robert G. Kuimelis*

InSilixa, Inc., 1000 Hamlin Court, Sunnyvale, CA 94089, USA

*Correspondence address: InSilixa, Inc., 1000 Hamlin Court, Sunnyvale, CA 94089, USA. Tel: (408) 809-3302; E-mail: bob.kuimelis@insilixa.com

Abstract

PCR-based techniques are widely used to identify disease causing bacterial and viral pathogens, especially in point-of-care or near-patient clinical settings that require rapid results and sample-to-answer workflows. However, such techniques often fail to differentiate between closely related species that have highly variable genomes. Here, a homogenous (closed-tube) pathogen identification and classification method is described that combines PCR amplification, array-based amplicon sequence verification, and real-time detection using an inverse fluorescence fluorescence-resonance energy transfer technique. The amplification is designed to satisfy the inclusivity criteria and create ssDNA amplicons, bearing a nonradiating quencher moiety at the 5′-terminus, for all the related species. The array includes fluorescent-labeled probes which preferentially capture the variants of the amplicons and classify them through solid-phase thermal denaturing (melt curve) analysis. Systematic primer and probe design algorithms and empirical validation methods are presented and successfully applied to the challenging example of identification of, and differentiation between, closely related human rhinovirus and human enterovirus strains.

Keywords: melt curve, point-of-care, pathogen classification, microarray, enterovirus, rhinovirus, inverse fluorescence, quencher, FRET

Introduction

Infectious disease molecular diagnostics (MDx) promise rapid and accurate detection and quantification of pathogens at a nucleic acid (genome) level [1]. Modern clinical pathology labs routinely use MDx methods to analyze patient samples to identify known pathogens or the unique genetic markers within them [2]. Such capabilities have been made possible by substantial scientific advances in the field of genomics, the development of nucleic acid amplification testing (NAAT) [3], DNA microarrays [4, 5], and more recently, next-generation DNA sequencing (NGS) [6, 7] methods. Nevertheless, there still remains an urgent global need for even more comprehensive MDx tests that provide affordable, rapid, actionable results and compatibility with point-of-care (PoC) workflows [8, 9].

Today, PCR-based platforms dominate the PoC testing landscape [10, 11]. In many cases, these platforms offer sample-to-

Received: 17 January 2018; **Revised:** 23 April 2018; **Editorial decision:** 24 April 2018; **Accepted:** 8 May 2018

answer results with "hours" turnaround times. However, when compared to microarray or NGS platforms, they offer limited multiplexing and mutation (i.e. SNP and indel) detection capabilities. For example, by utilizing different fluorophores, one can increase the multiplexing up to 8, but hardly beyond [10, 12]. Also, by implementing allele-specific primers or fluorogenic probes, one can only check for sequence variations in a region (typically 3–6 nucleotides), but cannot report with high confidence the exact base change or the precise coordinates [13, 14]. Current NAAT PoC technologies also often fail to identify closely related pathogen species, or fall short when it comes to genotyping strains based on their drug resistance mutation profile. These are major shortcomings that limit the utility, adoption and ultimately the overall impact of NAAT technologies. It is important to note here that it has been shown that through addition of post-PCR amplicon analysis techniques (e.g. microarrays [15] and denaturing gradient gel electrophoresis (DGGE) methods [16, 17]), one can improve the base calling accuracy of NAAT methods; however such techniques, by and large, require additional complex steps and are therefore not compatible with PoC workflows.

Here, we present an NAAT method including primer/probe design algorithms and an array-based amplicon sequence verification technique that addresses the multiplexing and genotyping deficiencies of many PCR-based PoC tests, while retaining the core advantages of simple workflow and rapid turnaround time. This approach utilizes asymmetric multiplex PCR to amplify a related set of nucleic acids with primers that are designed to ensure amplification of all known sequence variations. This is followed by selective capture of the amplicons by solid-phase probes of an array within the same reaction chamber and discrimination among related sequences (e.g. two closely related species) via solid-phase thermal denaturing. Hence, the primers are designed to ensure inclusivity, and the solid-phase probes provide specificity. To allow a homogeneous and "closed tube" method (i.e. no wash steps or addition of reagents after amplification), we implement an inverse fluorescence technique [18, 19] that relies on fluorescence-resonance energy transfer (FRET) [20]. PCR primers are 5′-labeled with a nonradiating quencher and solid-phase immobilized probes are 3′-labeled with a fluorophore.

In this work, we demonstrate the feasibility and practicality of the proposed NAAT method by identifying and differentiating between human rhinoviruses (HRVs) and human enteroviruses (HEVs), which is a challenging task for conventional NAAT methods. Although both are significant respiratory pathogens, HEV infections can spread to secondary target organs and it is therefore imperative to differentiate them from HRV in immunocompromised or young patients. Today, the MDx method of choice for detection of these viruses is RT-PCR-based NAAT with primers targeting the 5′-UTR of the viral RNA, a relatively conserved region within all members of the genus HEV (of which, HRV are members) [21, 22]. Nevertheless, sequence similarities between HRVs and HEVs complicate their differential detection and cause significant cross-reactivity [11, 23]. In the following sections, we detail our algorithms, methodology and experimental results addressing the above challenges.

Material and methods

Concept

We developed the NAAT detection method depicted in Fig. 1 to differentiate between closely related species or strains. The premise is to PCR amplify a common genomic region among the species, then apply a solid-phase probe array to detect and characterize the amplicon sequence and thus identify the species. As illustrated in Fig. 1A, the identification region is flanked by two common and minimally diverse regions, which act as landing sites for "universal" PCR primers, denoted henceforth by p_R and p_F. To enhance the solid-phase probe-amplicon hybridization kinetics, excess ssDNA amplicons (associated with p_F) are produced at the later stage of PCR amplification. This is ensured by making the concentration of primers asymmetric, i.e. $[p_R] \ll [p_F]$ [24].

To enable homogeneous (closed-tube) amplification and array detection, we have employed an inverse fluorescence transduction (IFT) technique [18, 19]. As shown in Fig. 1B, p_F bears a dark quencher (nonradiating accepter) at its 5′-terminus, while the solid-phase oligonucleotide probes include a fluorophore (radiating donor) at their 3′-terminus. Based on this design, any ssDNA amplicon generated from elongation of p_F includes a quencher. Thus, upon hybridization of the amplicon to the solid-phase probe, the acceptor is brought into close proximity of the donor to allow fluorescence resonance energy transfer (FRET) and to "turn off" the radiating signal. FRET quenching is a reversible process that can be measured without any background interference from the PCR products or reagents in the reaction chamber [14]. Therefore, it can be used to monitor both the kinetics and the thermodynamics of the solid-phase hybridization reaction. Specifically, one can set up assays to observe

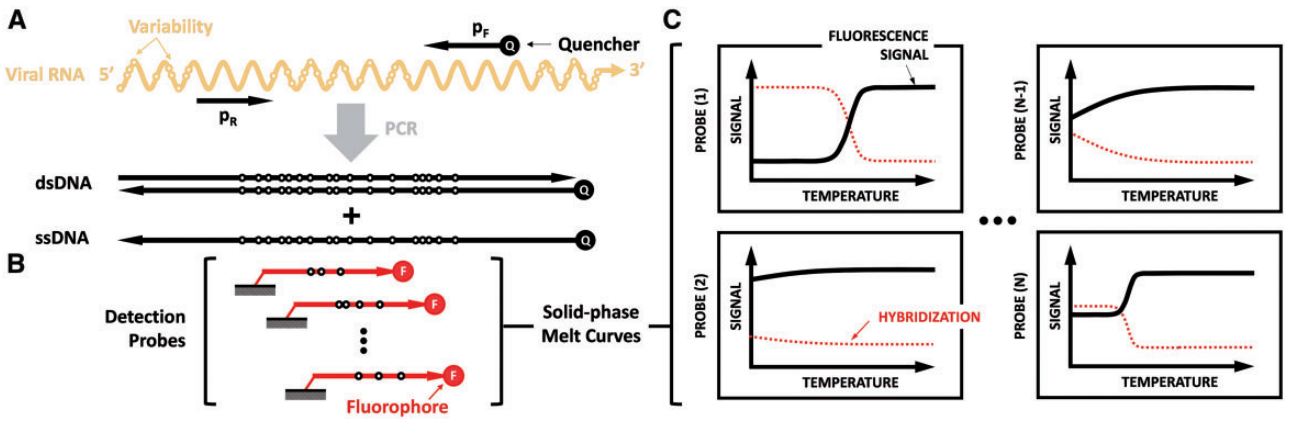

Figure 1: Our proposed detection and identification method, based on (**A**) amplifying all the target species using a universal primer pair; (**B**) introducing solid-phase IFT probes that are species-specific; and (**C**) performing solid-phase melt curve analysis to identify the sequence and the species.

the sequence-dependant probe-amplicon denaturing (melt) as a function of temperature. As shown in Fig. 1C, the shape and amplitude of such melt curves [19] from multiple probes can then be analyzed to identify the amplicon sequences and subsequently differentiate sequence and classify the species. Compared to solution-based methods, the array format is highly parallel and allows significant multiplex capability by virtue of (many) spatial coordinates, as opposed to (few) fluorophores and (few) instrument optical channels. Therefore, very complex and highly variable sequences can be identified using this NAAT method in a single reaction chamber.

Database generation

All available complete genome sequences for HEV and HRV strains capable of causing respiratory disease in humans were identified and downloaded from GenBank [25]. This search yielded 620 HEV and 37 HRV genomes (total = 657).

PCR primer design

MUSCLE [26] was used to produce a multiple sequence alignment of all 657 genomes. Shannon entropy metric [27] was calculated at each position for the aligned sequences. Possible coordinates for the universal p_R and p_F primers were then selected by identifying two regions of relatively low sequence entropy that would afford amplicons <100 bases in length.

To select the sequences and locations of the primers, all possible sequences between 20–30 bases in length were generated using the consensus sequence in the selected regions. Subsequently, extensive thermodynamic simulations were conducted with Oligonucleotide Modeling Platform (OMP) Developers Edition (DE) software package (DNA Software Inc, Ann Arbor, MI) [28], to identify thermodynamically optimal primer pair candidates (see Supplementary Data Section Table 1A and 1B for key parameters).

PCR primer verification

To verify the functionality and the performance of the p_R and p_F candidates, we tested eight commercially available HEV and HRV strains (see Supplementary Data Section Table 2) using conventional solution-based RT-qPCR methods. We required a PCR efficiency (p_e) [29, 30] of >90% for these eight HEV and HRV strains to establish inclusivity, and a threshold cycle (C_t) > 35 for the NTC (distilled water) samples. In addition, we tested both human background and non-HEV/HRV respiratory pathogens, and required $C_t > 40$ in all cases to establish exclusivity (see Supplementary Data Section Table 3).

RT-qPCR was performed with lyophilized ZipScript One-Step RT-qPCR reagent (Enzymatics, Beverly, MA) in a total reaction volume of 25 μl containing 1× EvaGreen® dye (Biotium, Fremont, CA), 300 nM of p_R and p_F primers (Integrated DNA Technologies, Coralville, IA), and 5 μl of template with the concentrations indicated below. Total viral DNA and RNA was isolated from NATtrol™ Molecular Controls (ZeptoMetrix, Buffalo, NY) using the Dynabeads™ SILANE Viral NA Kit (ThermoFisher Scientific, Waltham, MA), according to the manufacturer's instructions and quantified on a Qubit 3.0 Fluorometer (ThermoFisher Scientific, Waltham, MA). Total nucleic acid extracts were tested in 10-fold dilution series, with starting RNA concentration ranging from 0.1 ng/μl to 2 ng/μl and starting DNA concentration ranging from 0.01 ng/μl to 0.2 ng/μl. Quantified RNA for Enterovirus71 Strain BrCr (VR-1775DQ) was obtained from ATCC (Manassas, VA). Nucleic acids from 2 to 4 strains were pooled together for nontarget testing. All RT-qPCRs were performed in triplicate and cycled on a CFX96 Touch (Bio-Rad, Hercules, CA) at 55°C for 5 minutes, 95°C for 3 minutes, followed by 45 cycles of 95°C for 15 seconds and 65°C for 30 seconds. Reaction specificity was assessed by conventional solution-phase (EvaGreen) melt curve analysis from 55°C to 95°C at 0.5°C/10 seconds immediately following the RT-qPCR. uMELT [31] was used to predict the melt curve and melt temperature of the 122-bp amplicon and for comparison to experimental data. Data was analyzed with Bio-Rad CFX Manager 3.1 software.

Solid-phase probe design algorithm

The following three-step probe design algorithm was employed:

Selecting appropriate regions within the amplicon (step #1). Let us consider a set of N sequences (amplicons) composed of K species, each having n_k entries in the database, where $\sum_{k=1}^{K} n_k = N$. The first step for solid-phase probe design for species k is to find regions within the amplicon that are similar within all n_k sequence entries, but that differ from the other species. To do this, K separate multiple sequence alignments (one for each species) are created and a position probability matrix, denoted by $P_k \in \mathcal{R}^{5\times n_k}$, is created. The entries of the matrix denoted by $p_k(i,\ j)$, are defined by

$$p_k(i,j) = \text{The probability of observing nucleobase } i \text{ at coordinate } j \text{ within } n_k, \quad (1)$$

where $i = \{1,\ 2,\ 3,\ 4,\ 5$ corresponds to {A, C, G, T,"-"}, respectively.

One key characteristic of this specific matrix is that the inner-product of its columns for species m and n, i.e., $p_m(j)^T p_n(j)$, is a measure of similarity (correlation or cross-correlation) between the two species at coordinate j. One can further expand this concept and compute the correlation between the two species for a window of length w within the amplicons. This regional correlation function, denoted here by $C_{m,n}(x,w)$, is formulated as

$$C_{m,n}(x,w) = \sum_{j=x}^{x+w} p_m(j)^T p_n(j), \quad (2)$$

where x is the starting location of the window on the alignment. To find an optimal probe region for species m, one scans the regions to find the maximum $C_{m,m}(x,w)$ and sufficiently low $C_{m,n}(x,w)$'s.

Generating probe candidates (step #2). The next step is to create probe candidates for each region, to be down-selected later. We take advantage of the probability position matrix to create a super set of sequences that are likely to capture the sequences of the database. As exemplified in Fig. 2, the entries of the probability position matrix (Fig. 2A) are used to construct a permutation tree (Fig. 2B), with individual branches at sequence coordinates with multiple possible nucleobases [32, 33]. To do this, at the first coordinate the base with the highest probability is selected, then one moves to the second coordinate and selects the highest probability base there, moves to the next coordinate, and so forth. Later and through creating a new branch, the prior coordinates can be revisited to select the bases with the second highest probability by repeating the process. To ensure that probes capturing low abundance amplicons are not created, a probability threshold (0.2 for this work) is set and any base with lower probability is ignored.

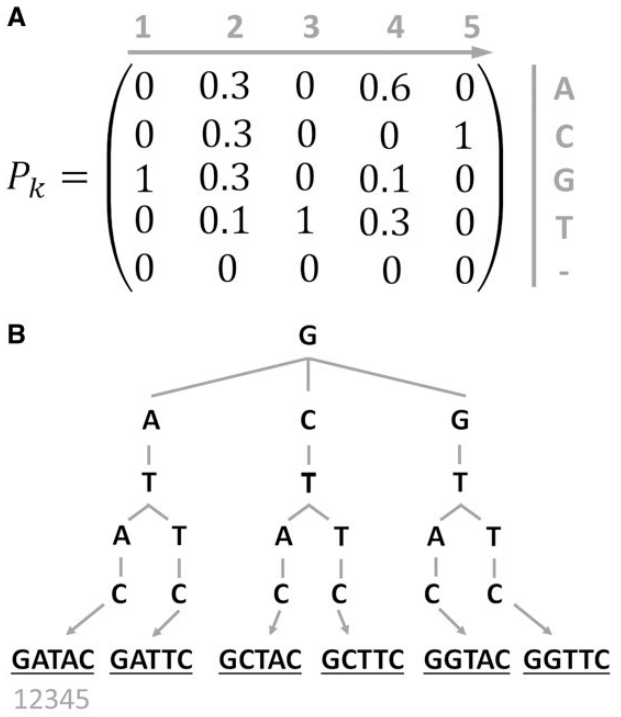

Figure 2: An example for (**A**) position probability matrix; and (**B**) the corresponding permutation tree used to create probe candidates.

Next, the number of probe candidates is reduced by identifying nonfunctional and outlier probes. OMP-DE thermodynamic simulations are performed to identify and remove any sequence having $<70\%$ random coil structure and melt temperature (T_m) $<55^\circ$C or $>75^\circ$C with its fully complementary sequence.

Finally, all probes capable of discriminating between HEV and HRV amplicons are identified through OMP-DE thermodynamic simulations and the rest of the probes are removed. The parameter that is computed and used for this phase is p_H, the heterodimer percentage, defined as the percentage of the probe in the duplex form with a given amplicon. To meet the inclusivity requirement, we set $p_H > 80\%$ for at least one of the database entries for the species to be detected by the probe, and for the exclusivity criteria we set $p_H < 2\%$ for all other species that are not to be detected. Any sequence not meeting both these criteria simultaneously are discarded.

Probe set selection (step #3). In the final stage of the algorithm, a greedy search and elimination heuristic is implemented to find a set of probes that can collectively capture and detect all (or a reasonable high percentage of) the amplicon targets. For this purpose, one starts with a sequence set composed of probes that capture the highest percentage of amplicons for each species. Next, a modified database is created by removing the amplicons that are captured by the set. The same process is executed for the modified database and new probes are added for each species to the probe set. This process is continued until no further probe is necessary to meet the required percentage or the modified database is emptied.

IFT

Figure 3A depicts our technique for maximizing the quenching probability (signal change) in the assay. The idea is to minimize the distance between the donor-acceptor FRET moieties and thereby promote contact quenching. The identification probe sequences include at their 3′-end a short sequence complementary to the 5′-end of the amplicon (p_F) and also a fluorophore construct (donor) at the 3′-terminus. As shown in Fig. 3B, after the amplicon hybridizes to the identification sequence, its 5′-dangling end can be captured by the "sticky end" of the probe to bring the fluorophore and quencher into proximity. It is important to recognize that the length of this "sticky-end" should be as short as possible since it can otherwise create undesired quenching if excess primers in the PCR are captured. OMP-DE thermodynamic simulations were performed to identify optimal "sticky-end" length and create IFT probes. To do this, thermodynamic parameters of the following three specific DNA–DNA hybridization reactions were computed: (i) IFT probe to p_F, (ii) identification region of the probe to the amplicon, and (iii) the "sticky-end" clamp structure of the probe to amplicon, with melt temperature designations T_m^p, T_m^i, and T_m^s, respectively (Fig. 3C). This third quantity is computed through thermodynamic simulation of a single oligonucleotide that joins the "sticky-end" portion of the probe to a segment of the amplicon that begins at its 5′-end and ends at the 3′-end of the identification region binding site. Next, all possible "sticky-ends" with lengths 7–14 bases and complementary to the 5′-terminus of the amplicon were evaluated to meet the following criteria

$$T_m^p \ll T_m^i < T_m^s. \tag{3}$$

Finally, candidates are ranked according to the highest difference between the melt temperatures of (3), and the top candidates are selected and graduated to experimental verification.

Array fabrication and probe printing

Bare array substrates diced from silicon wafers (25 × 76 × 0.625 mm with 3000 Å thermal SiO_2) (Addison Engineering, San Jose, CA) were treated with oxygen plasma (600W, 1 Torr, 120°C, 3 minutes, float plate) followed by water vapor (13 Torr, 120°C, 5 minutes) for surface activation and silanol repopulation. Without breaking vacuum, the substrates were then treated with gaseous (3-glycidoxypropyl) trimethoxysilane (0.5 Torr, 120°C, 30 minutes) (Gelest, Morrisville, PA). All steps were performed with a model 1224 P chemical vapor deposition instrument (Yield Engineering Systems, Livermore, CA). Coated substrates were cooled to room temperature in a controlled environment (20% RH drybox with continuous N_2 purge) and used within several days.

HPLC purified synthetic DNA oligonucleotide probes bearing hexylamine at the 5′-terminus and a dual fluorophore construct near the 3′-terminus were procured (LGC Biosearch, Petaluma, CA) and reconstituted to 50 μM in pH 8.5 phosphate buffer. The dual fluorophore construct was composed of tetramethylrhodamine (TAM) at the 3′-terminus and a thymidine-fluorescein (FAM) conjugate separated from TAM by six additional thymidine residues (Fig. 3B). Piezo type printing with a sciFLEXARRAYER SX (Scienion, Germany) was employed to deposit ~100pL droplets at 200 μm pitch. Covalent coupling of the DNA probe to the epoxy functionalized surface was accomplished with a humidity-controlled overnight incubation (room temperature, 50% RH). Excess and unattached DNA probe was removed by extensive dilution with aqueous buffer, and any unreacted surface epoxy groups were simultaneously rendered inert with a Tris-buffered solution containing ethanolamine. We estimate the final surface probe density to be approximately 20 pmole/cm^2 (unpublished results). Printed substrates were stored in a desiccated nitrogen environment at 4°C until use,

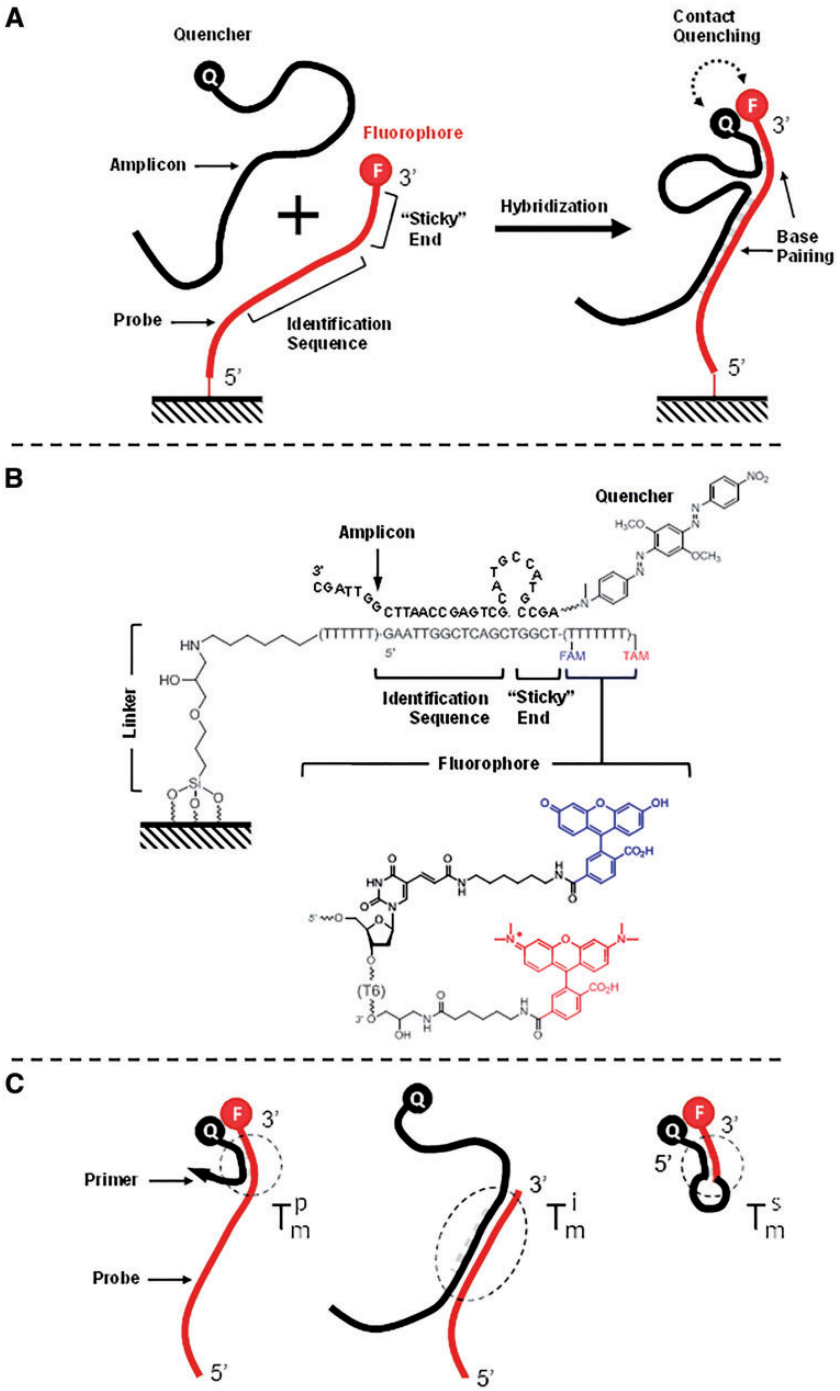

Figure 3: (**A**) IFT probe complexed structure to facilitate contact quenching; (**B**) IFT probe molecular structure; and (**C**) key constructs simulated for the design of the "sticky-end" length.

typically within ~6 months. To create a biochip module, a laminated, flow-through acrylic fluidic cap with a capacity of ~90μL was affixed to the substrate with silicone-based pressure sensitive adhesive.

The printed arrays used a 18 × 14 grid pattern (252 spots) composed of 11 different IFT probes, each having 7 replicates (77 total). The remaining spots were taken up by reference probes for grid alignment, control probes for fluorophore signal calibration, process monitor probes, alternate probes that were not used and blank locations for pattern recognition (see Supplementary Data Section Figure 1 for additional details).

Measurement setup

The biochip module was mounted to a custom-built thermomechanical stage suitable for attachment to a microscope (Fig. 4A). The stage consisted of a heat spreader plate for mounting the substrates, a high-temperature rated VT-199-1.4-0.8

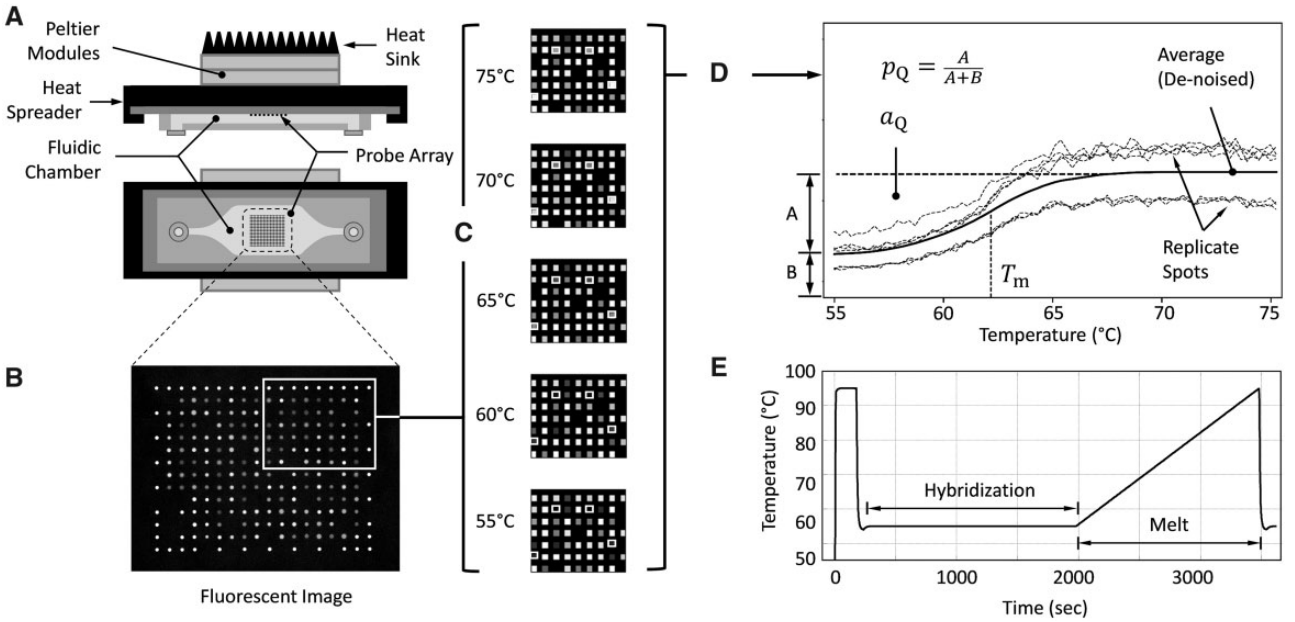

Figure 4: Measurement setup, with (**A**) the biochip module, mounted on a fluorescent microscope stage with heating and cooling capabilities. (**B**) and (**C**) The fluorescent images collected as a function of time/temperature to create (**D**) the melt curve when the temperature profile (**E**) is applied.

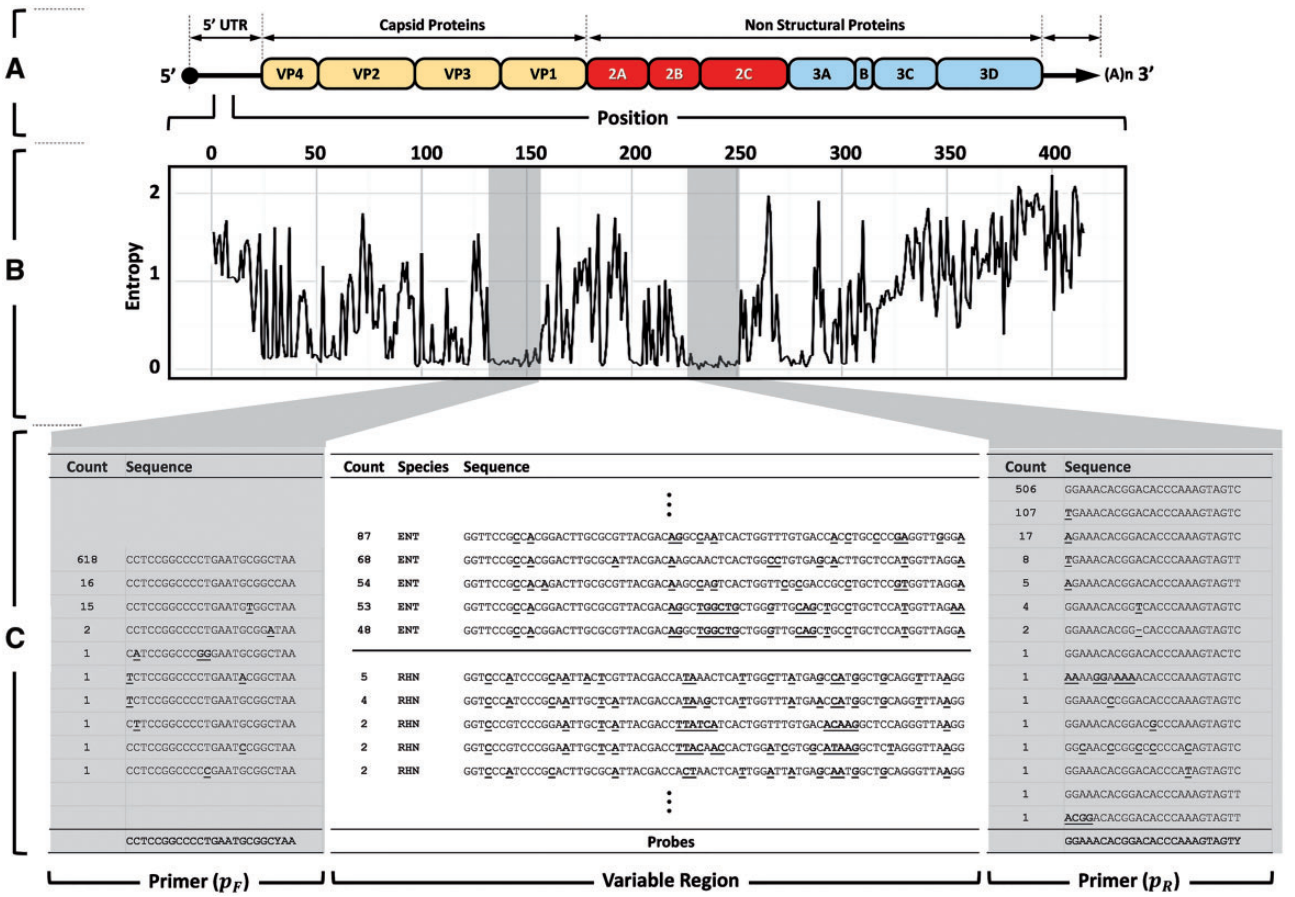

Figure 5: (**A**) Genomic structure of HEV and REV; (**B**) the sequence entropy plot for the selected target region for amplification, detection, and differentiation; and (**C**) representative sequences from the database used for the primers and probes design.

thermoelectric module with Q_{max}=172W (TETECH, Traverse City, MI), and a 4-2020XXUBFA copper heatsink (Cool Innovations, Concord, ON). The stage can perform rapid thermal cycling with heating and cooling rates exceeding ±4°C/second. Temperature accuracy within ±1°C is achieved using TR67 thermistors (Oven Industries, Camp Hill, PA).

Fluorescence images from the biochip module's array were collected with a Leica DMi8 inverted fluorescence microscope outfitted with an X-Cite 120LED light source (Excilitas Technologies, MA, USA), 2.5× objective, FAM-TAM filter cube (Semrock, Rochester, NY) and QIClick CCD detector (QImaging, Canada) (Fig. 4B). The temperature of the biochip module was controlled using a custom Peltier temperature controller. The microscope and temperature controller were controlled using a custom LabVIEW-based (National Instruments, Austin, TX) instrument interface and data acquisition application software, which permitted programming specific temperature and imaging profiles and collecting the generated data.

The contrast-enhanced array images (Fig. 4C) were denoised with an adaptive Wiener filter [34] and a circular Hough transform [35] was used to identify array features. The median of the pixel intensities from each spot and its surrounding local area was then used to calculate the background intensity, and subsequently subtracted from the signal intensity for each spot.

As shown in Fig. 4D, three analysis metrics were then computed based on the measured and averaged melt curves from identical probes, using a custom MATLAB code. The first metric was p_Q, the quenching percentage, defined as the relative reduction in the fluorescent signal at the onset of the melt curve. The second metric was a_Q, the quenching area, which is the total area above the melt curve. Finally, the third metric was T_m, the observed melt temperature, which is computed by fitting two straight lines at the onset and plateau regions of the melt curve and computing the temperature at which the melt curve is in equal distances to the fitted lines [36].

Array melt curve experiments

Asymmetric RT-PCR was performed in 25 µl reactions using lyophilized ZipScript One-Step mastermix, 300 nM excess primer, 50 nM limiting primer, and 5 µl of template. All primers were BHQ2 labeled (LGC BioSearch, Petaluma, CA) at the 5′-terminus. Amplification was carried out on a CFX96 Touch (Bio-Rad, Hercules, CA) at 55°C for 5 minutes, 95°C for 3 minutes, followed by 30 cycles of 95°C for 15 seconds and 65°C for 30 seconds.

The two ATCC strains of purified RNA (VR-1775DQ and VR-1663D) were used with no additional treatment. The Zeptometrix RNA templates (NATRVP-IDI) were prepared by treating the cultured virus with Dynabeads Silane Viral NA Kit (Thermo Fisher p/n 37011D). Synthetic templates were obtained from Integrated DNA Technologies (Coralville, IA) as PAGE-purified Ultramers (lyophilized) and were resuspended in TE buffer. The concentration was then determined by OD260 on a DU-800 UV-Vis spectrophotometer (Beckman Coulter, Brea, CA). The stock was serially diluted down to 1 pM in 10 mM Tris buffer (pH 8.0 with 0.1% Tween-20), 5 µl of which was then added into 20 µL of the amplification reaction mixture.

For melt curve analysis, 90 µl of the pooled PCR product was added into the fluidic chamber of biochip module. Next, the module was mounted on the measurement setup. The thermal profile consisted of 3 minutes of denaturation at 95°C followed by 30 minutes at 55°C (Fig. 4E). The 10-second interval imaging was performed for 2 minutes at the beginning and the end of the hybridization, and 1-minute imaging interval during hybridization. For the melt step, the temperature was ramped from 55°C to 95°C over 25 minutes with imaging every 10 seconds.

Results and discussion

PCR amplification

As a demonstration of our method, we selected the long-standing problem of HEV and HRV differentiation. Figure 5A

Table 1. Primer and probe sequences

	Sequence (5′–>3′)
Probe-HEV-1	[linker]-GGCAAGTCTGTGGCGGAACCCGTGTTTCC-[F]
Probe-HEV-2	[linker]-GCCTGTCGTAACGCGCAAGTCGTGTTTCC-[F]
Probe-HEV-3	[linker]-ACGGGTAACTCTGCAGCGGAACCCGTGTTTCC-[F]
Probe-HEV-4	[linker]-GCGTAAGTCTGTGGCGGAACCCGTGTTTCC-[F]
Probe-HEV-5	[linker]-AACGGGAAAGTCTGTGGCGGAACCCGTGTTTCC-[F]
Probe-HEV-6	[linker]-GTGTGTCGTAATGAGCAATTCTCGTGTTTCC-[F]
Probe-HRV-1	[linker]-GAGTAATTGCGGGATGGGACCCGTGTTTCC-[F]
Probe-HRV-2	[linker]-AATGAGCAATTGCGGGATGGGACCCGTGTTTCC-[F]
Probe-HRV-3	[linker]-TGGTCGTAATGAGTAATTCTGGGACGTGTTTCC-[F]
Probe-HRV-4	[linker]-ACGGGCAATTCTGGGATGGGACCCGTGTTTCC-[F]
Probe-HRV-5	[linker]-GCAATTACGGGATGGGACCCGTGTTTCC-[F]
Primer (p_F)	[Q]-GTCCTCCGGCCCCTGAATGCGGCYAA
Primer (p_R)	[Q]-GGAAACACGGACACCCAAAGTAGTY

The underscored portion corresponds to the "sticky-end" section. [linker], [F], and [Q] correspond to the solid-phase linker moieties, fluorophore, and quencher, respectively.

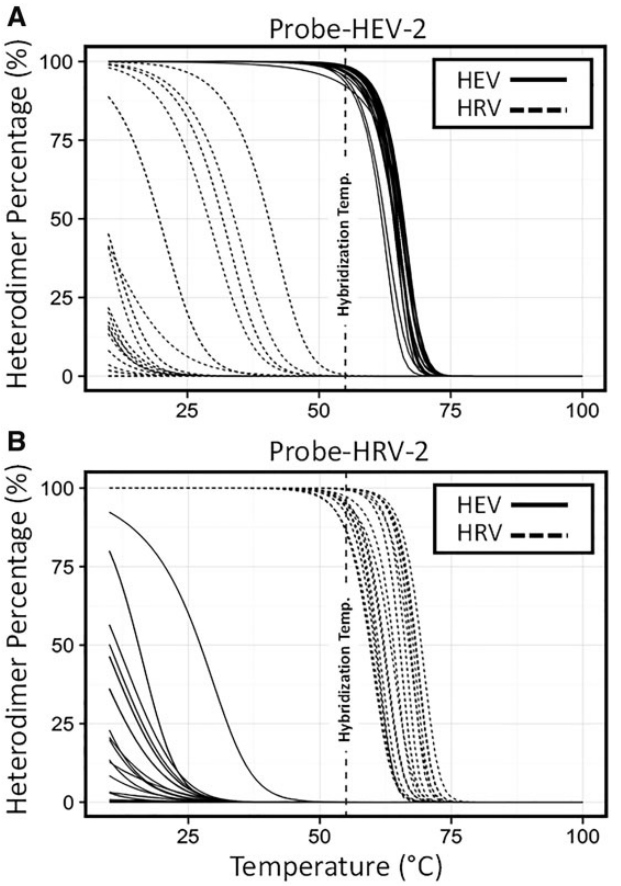

Figure 6: Example simulated melt curve and heterodimer percentage (p_H), showing the percentage bound for all database sequences at the hybridization (assay) temperature for (**A**) Probe-HEV2; and (**B**) Probe-HRV2.

shows the genomic structure of HEV and HRV, and Fig. 5B shows the sequence entropy of a section of their 5′-UTR, identified by our algorithms as an optimal coordinate for designing a universal HEV/HRV primer pair. Two very low-entropy regions were identified in this region for all 657 database sequences. These two regions are located ~70 bases apart which makes them particularly appropriate for the design of p_R and p_F due to the modest amplicon length.

As is apparent from the sequences listed in Fig. 5C, any selected primer in this region will unavoidably have mismatches with some of the database sequences. However, we designed five primer pairs with degenerate (mixed) bases using the previously described methods that were found to meet the *in-silico* PCR inclusivity criteria for the vast majority of sequences in the database, defined as >75% primer-template duplex formation in the PCR annealing conditions and >25% primer-template duplex formation in the PCR annealing conditions. These primer candidates were experimentally tested in solution and ranked based on the criteria discussed in the "Materials and Methods" section. The top candidate p_R and p_F are listed in Table 1 and all candidate primers are Supplementary Data Section Table 4.

The RT-qPCR dilution series for the top primer pair is shown as an example for ATCC HEV 71 Strain BrCr template. Using this data set, p_e, the PCR efficiency, was computed by comparing C_t values across dilutions (Supplementary Data Section Figure 2). The response of the top primer pair to the presence of human background, influenza nucleic acids and no template control (NTC) is also shown in Supplementary Data Section Figure 2.

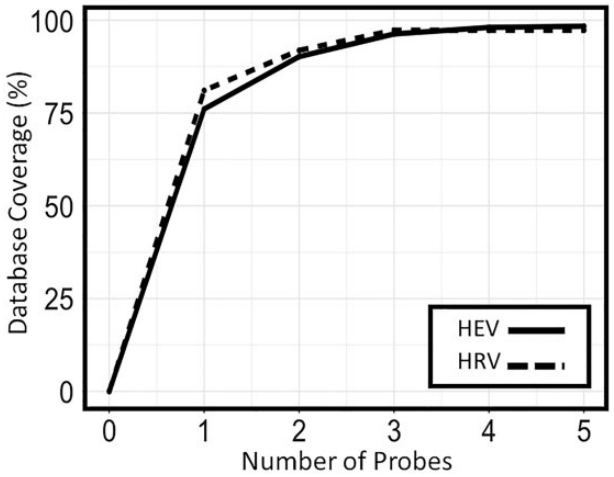

Figure 7: Number of probes vs. database coverage for both HEV and HRV.

Probe design

In Fig. 6, an example of simulated hybridization melt curves for two HEV and HRV probe candidates are shown as part of the probe design algorithm step #2. Clearly, most of the sequences in the database can be differentiated *in silico* using just these two probes; however there are strains that exhibit p_H values close to the threshold, and hence more probes are required to comprehensively differentiate all the sequences.

Figure 7 illustrates the results of the third step of the probe design algorithm. Significantly diminishing coverage returns are observed after increasing the number of probes beyond five for both HEV and HRV strains. With the selected probes listed in Table 1 (Probe-HEV-[1–6] and Probe-HRV-[1–5]), a coverage of 100% and 97.3% for HEV and HRV amplicons were achieved, respectively.

HEV and HRV array-based differentiation experiment

Table 2 lists the HEV and HRV strains selected for experimental validation of probes. In addition to three commercially available strains, 11 synthetic DNA templates were designed that collectively test the ability of the probe sets to detect the most challenging strains of the database. Specifically, to demonstrate the versatility of the method, we aimed to verify targets that are predicted to hybridize strongly to each probe (i.e. large inclusivity p_H), targets that are predicted to hybridize weakly (i.e. small inclusivity p_H), and targets from one species that are predicted to bind to probes for the other (i.e. high exclusivity p_H).

Table 2. HEV and HRV target sequences

Strain	Accession number/ catalog number	Prediction/purpose	Amplicon sequence excluding primer binding sites (5'–>3')
Target-HEV-A	GenBank: KJ170577.1	High p_H for Probe-HEV-[1, 2, 4, 5]	GGTTCCGCCACGGACTTGCGCGTTACGACAGGCTGGCTGCTGGGTTGCAGCTGCCTGCTCCATGGTTAGAA
Target-HEV-B	GenBank: AF465517.1	High p_H for Probe-HEV-3 Predicted cross-reactivity	GGTTCCGCTGCAGAGTTGCCCGTTACGACAGGCCACCCACTGGATTGTGAGCACCTGCTCCGCAGTTAAGA
Target-HEV-C	GenBank: KF311743.1	Low overall p_H	GGTTCCGCTGCGGAGTTGCCCATTACGACGCGCCACCTCCTGGTTTGGGGGCACGCGCTCCGCAGTTAGGA
Target-HEV-D	GenBank: EF174469.1	Low overall p_H	GGTTCCGCCGCAGAGTTGCCCATTACGACAGACTGCTCACTGGCGTGTGAGTATCTGCTCTGCGGTTAGGA
Target-HEV-E	GenBank: AY302559.1	High p_H for Probe-HEV-6	GGTTCCGCTGCGGAATTGCTCATTACGACACACTACCCCTTGGATTAGGGGCATGTGCTCCACAGTTAGGA
Target-HEV-F	GenBank: JX275008.1	Low overall p_H	GGTTCCGCCACGAACTTACGCGTTACGACAAGGTATGTGTTGGCTCACGCACCCTTGCTCCGTGGTTAGGA
Target-HEV-G	ATCC: VR-1775DQ	Commercial strain	na
Target-HRV-A	GenBank: DQ473501.1	High p_H for Probe-HRV-[1, 2]	GGTCCCATCCCGCAATTACTCATTACGACCAGCTACACACTGGTTTGTGCGCACTGGCTGCAGGGTTAAGG
Target-HRV-B	GenBank: EF173422.1	High p_H for Probe-HRV-2 Predicted cross-reactivity	GGTCCCGTCCCGGAATTGCTCATTACGACCTTATCATCACTGGTTTGTGACACAAGGCTCCAGGGTTAAGG
Target-HRV-C	GenBank: EF173420.1	Low overall pH	GGTCCCGTCCCGGAATTACTCATTACGACCTTGCCAACACTGGATTGTGTCTCAAGGCTCCAGGGTTTAGG
Target-HRV-D	GenBank: FJ445161.1	High p_H for Probe-HRV-[3, 4]	GGTCCCATCCCAGAATTGCTCATTACGACCTCCACAACATTGGATTATGCTGCAAGGCTCCGGGGTTAAGG
Target-HRV-E	GenBank: FJ445138.1	High p_H for Probe-HRV-[2, 5]	GGTCCCATCCCGTAATTGCTCGTTACGACTAGCCACACACTGGATTGTATGCACTAGCTACGGGTTTAAGG
Target-HRV-F	ATCC: VR-1663D	Commercial strain	na
Target-HRV-G	ZeptoMetrix: NATRVP-IDI	Commercial strain	na

Where indicated, na denotes samples whose sequences are unknown.

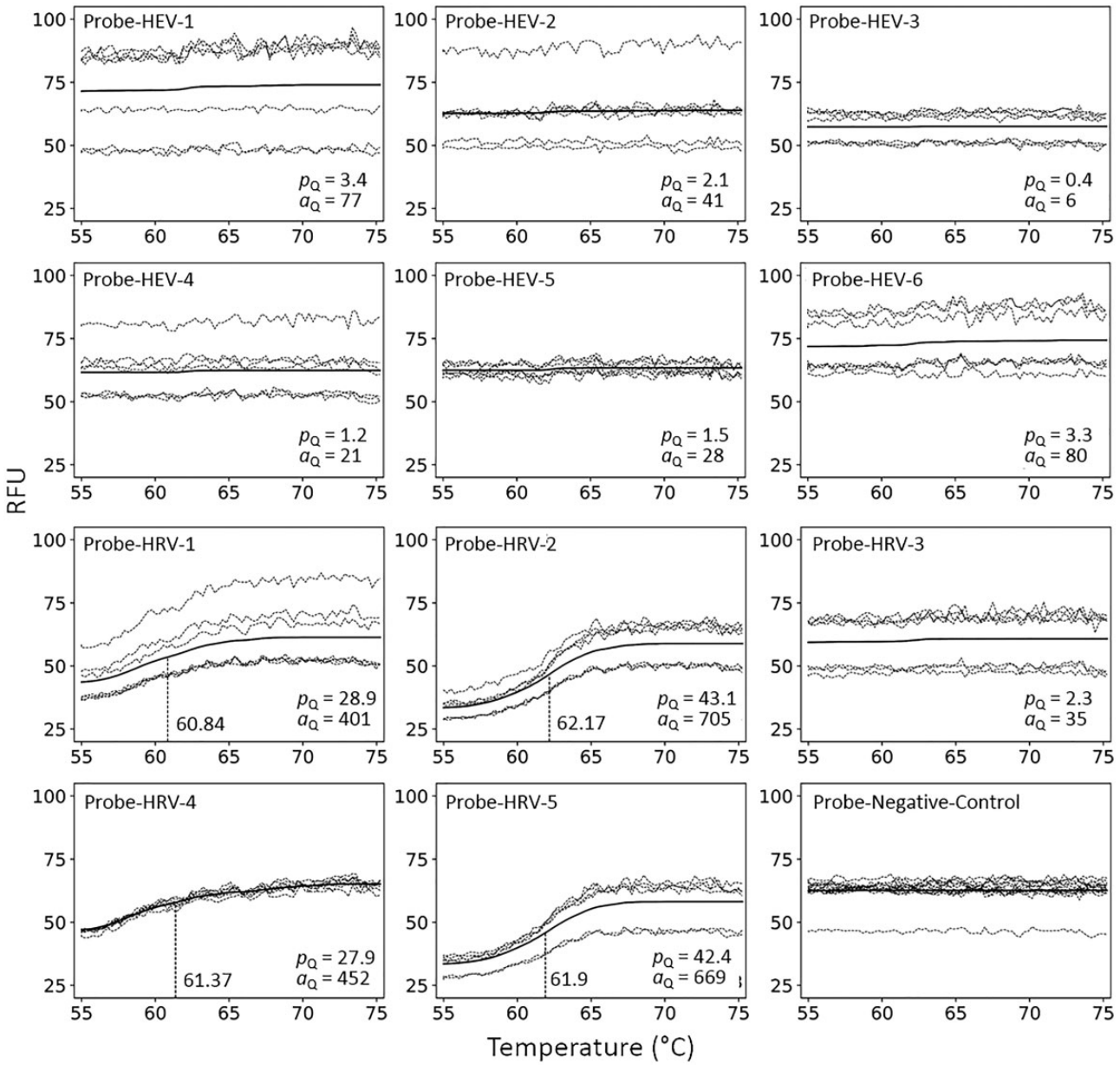

Figure 8: Array melt curves for all probes following Target-HRV-E amplicon hybridization. The dashed lines are from individual, replicate array spots (total seven). The solid line is the average of the replicate array spots after de-noising. The inset name corresponds to 1 of the 11 array probes. Also inset are the curve metrics (p_Q and a_Q) and the measured T_m, indicated with a vertical dotted line. RFU = Relative Fluorescence Units.

Figure 8 shows the individual array-based melt curves for all probes for Target-HRV-E. The analysis parameters p_Q, a_Q, and T_m are also shown. Clearly in this example, the HEV probes did not exhibit a melt curve, while four out of five HRV probes produced readily discernible melt curves, correctly predicting that the amplified target was in fact a HRV strain. Melt curves for the remaining 13 targets are shown in Supplementary Data Section Figures 3.1–3.13.

A summary of experimental results for all 14 targets of Table 2 is shown in Table 3. In this table, all three analysis metrics are reported. Furthermore, probe responses for each metric that corresponds to the presence of a melt curve are highlighted. The probes all behave as expected and correctly differentiate the selected HRV and HEV strains. For example, according to Table 2, one should expect strong affinity of Target-HRV-E to Probe-HRV-2 and Probe-HRV-5, and weaker binding to Probe-HRV-1 and Probe-HRV-4.

Finally, a scoring parameter is defined, denoted here as S, to represent the collective result of the solid-phase melt curve experiment. S is formulated by

$$S = \log_{10}\left(\sum a_{Q-HEV} / \sum a_{Q-HRV}\right), \quad (4)$$

where $S > 0$ corresponds to the detection of HEV, while $S < 0$ corresponds to the detection of HRV strains. In Fig. 9, the computed S of all 14 experiments vs. the template list is reported and shows the correct species differentiation for all examples.

The array-based melt curve method described here successfully detected and differentiated a challenging collection of

Table 3. Measurement result table for all targets vs. all probes

	Strain	Solid-phase probes										
		HEV-1	HEV-2	HEV-3	HEV-4	HEV-5	HEV-6	HRV-1	HRV-2	HRV-3	HRV-4	HRV-5
Quenching Percentage (p_Q)	Target-HEV-A	45.6	38.7	1.0	46.5	31.1	4.0	1.7	0.9	4.6	1.4	1.5
	Target-HEV-B	24.5	5.6	57.2	10.7	15.4	3.3	1.9	1.2	3.8	2.1	1.4
	Target-HEV-C	14.2	1.9	37.9	9.4	6.8	3.6	1.5	1.2	3.0	1.2	1.4
	Target-HEV-D	37.7	0.9	31.7	25.6	24.4	5.0	1.1	0.8	7.4	3.0	0.8
	Target-HEV-E	6.1	0.0	14.2	3.7	3.3	28.6	0.7	0.8	7.6	1.6	0.5
	Target-HEV-F	21.1	6.9	2.5	31.8	16.7	2.1	0.7	1.2	1.5	0.2	0.3
	Target-HEV-G	18.7	1.8	46.0	7.4	14.8	2.1	1.9	2.5	2.9	2.0	1.6
	Target-HRV-A	2.4	1.1	0.8	1.8	1.5	3.3	51.5	46.3	23.1	21.7	24.1
	Target-HRV-B	4.5	3.2	1.2	2.6	3.7	19.3	18.8	40.8	31.7	35.3	28.1
	Target-HRV-C	5.9	4.8	3.1	6.7	7.5	13.9	27.2	13.8	32.3	15.4	10.4
	Target-HRV-D	2.7	1.5	0.6	2.5	1.9	20.8	18.9	37.3	33.5	51.1	19.0
	Target-HRV-E	3.4	2.1	0.4	1.2	1.5	3.3	28.9	43.1	2.3	27.9	42.4
	Target-HRV-F	2.1	1.0	0.4	1.2	0.7	1.7	26.6	12.7	38.6	12.9	5.7
	Target-HRV-G	2.9	1.7	1.1	2.1	2.8	1.9	50.9	44.5	10.3	12.9	15.0
Quenching Area (a_Q)	Target-HEV-A	1027	863	25	937	510	159	33	8	144	24	34
	Target-HEV-B	403	75	1498	262	218	109	49	18	110	39	35
	Target-HEV-C	412	39	751	240	172	84	31	13	57	25	40
	Target-HEV-D	760	23	532	629	582	167	16	5	210	52	17
	Target-HEV-E	140	1	348	93	97	425	11	7	79	31	19
	Target-HEV-F	462	83	50	437	346	32	7	6	17	1	7
	Target-HEV-G	559	73	1830	326	445	107	81	88	142	63	60
	Target-HRV-A	43	22	14	29	21	90	1281	888	327	508	471
	Target-HRV-B	139	105	33	76	125	476	233	725	591	643	443
	Target-HRV-C	135	123	72	146	172	517	321	195	475	301	204
	Target-HRV-D	43	25	11	46	30	319	401	565	511	1118	337
	Target-HRV-E	77	41	6	21	28	80	401	705	35	452	669
	Target-HRV-F	80	55	16	52	21	81	590	363	1327	480	146
	Target-HRV-G	114	63	41	69	109	84	1949	1501	216	535	410
Melting Temperature (T_m)	Target-HEV-A	62.4	63.5	na	61.9	60.6	na	na	na	na	na	na
	Target-HEV-B	59.8	na	64.9	64.6	59.5	na	na	na	na	na	na
	Target-HEV-C	65.6	na	62.2	na	na	na	na	na	na	na	na
	Target-HEV-D	62.2	na	61.4	67.0	64.0	na	na	na	na	na	na
	Target-HEV-E	na	na	65.7	na	na	59.8	na	na	na	na	na
	Target-HEV-F	65.7	na	na	60.3	67.0	na	na	na	na	na	na
	Target-HEV-G	60.6	na	63.8	na	60.0	na	na	na	na	na	na
	Target-HRV-A	na	na	na	na	na	na	64.6	62.7	59.8	67.8	62.7
	Target-HRV-B	na	na	na	na	na	64.1	60.1	61.9	61.4	61.7	60.3
	Target-HRV-C	na	na	na	na	na	70.2	59.8	63.2	61.4	64.8	67.8
	Target-HRV-D	na	na	na	na	na	60.6	65.4	60.8	60.8	64.3	63.2
	Target-HRV-E	na	na	na	na	na	na	60.8	62.2	na	61.4	61.9
	Target-HRV-F	na	na	na	na	na	na	59.5	62.4	61.4	64.3	na
	Target-HRV-G	na	na	na	na	na	na	63.3	61.6	59.3	67.5	61.9

Highlighted cells correspond to positive calls and na denotes the absence of an observable melt curve.

closely related HEV and HRV viral strains. To our knowledge, this is the first report of such PCR-based HEV/HRV differentiation. Systematic *in silico* primer/probe design identified high-performing and experimentally validated sequences. Using the approaches described here, we have recently developed a 26-plex assay broadly focused on clinically relevant upper respiratory pathogens (unpublished results). We expect that these methods will be generalizable to other diverse NAAT applications. The "sticky end" probe construct and IFT method enables a homogenous (closed tube) assay not requiring post-amplification addition of reagents or wash steps. Although the closed-tube nature of the assay is particularly important for clinical diagnostic PoC applications, the algorithms and methods described here will also find use in research applications that require a high degree of multiplex and specificity compared to conventional solution-phase hydrolytic probe (i.e. TaqMan) chemistry. The array-based format allows for significant multiplexing and parallel melt curve data collection for each probe: amplicon complex represented on the array. Embedded within each melt curve is rich thermodynamic information that

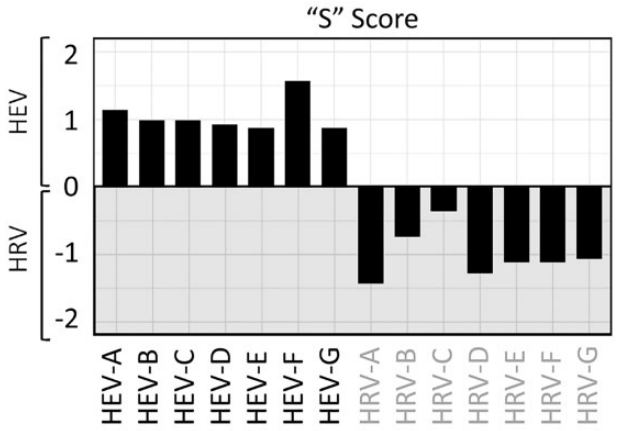

Figure 9: S score for seven HEV and seven HRV targets.

is a biophysical attribute of the specific complex. Although work presented here employed readily available equipment and slide-based printed arrays with $14 \times 18 = 252$ individual features, future work will translate this method to a miniaturized biochip with $32 \times 32 = 1024$ individual features, capable of performing PCR amplification within the array chamber and real-time fluorescence data acquisition.

Supplementary data

Supplementary data are available at *Biology Methods and Protocols* online.

Acknowledgements

The authors thank Professor John SantaLucia Jr and the staff of DNA Software, Inc for their technical support and suggestions related to thermodynamic simulations for primer and probe design.

Conflict of interest statement. All co-authors are employees of InSilixa, Inc., a company developing CMOS biosensors and associated assays for infectious disease testing.

References

1. Buchan BW, Ledeboer NA. Emerging technologies for the clinical microbiology laboratory. *Clin Microbiol Rev* 2014;**27**:783–822.
2. Choffnes ER, Olsen L, Wizemann T. *The Science and Applications of Microbial Genomics: Workshop Summary*. *Institute of Medicine (US) Forum on Microbial Threats*. Washington (DC): National Academies Press (US), 2013.
3. Niemz A, Ferguson TM, Boyle DS. Point-of-care nucleic acid testing for infectious diseases. *Trends Biotechnol* 2011;**29**:240–50. doi: 10.1016/j.tibtech.2011.01.007.
4. Lin B, Wang Z, Vora GJ *et al*. Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. *Genome Res* 2006;**16**:527–35. doi: 10.1101/gr.4337206.
5. Simoes EAF, Patel C, Sung W-K *et al*. Pathogen chip for respiratory tract infections. *J Clin Microbiol* 2013;**51**:945–53. doi: 10.1128/JCM.02317-12.
6. Abril MK, Barnett AS, Wegermann K *et al*. Diagnosis of Capnocytophaga canimorsus sepsis by whole-genome next-generation sequencing. *Open Forum Infect Dis* 2016;**3**:ofw144. doi.org/10.1093/ofid/
7. Lefterova MI, Suarez CJ, Banaei N *et al*. Next-generation sequencing for infectious disease diagnosis and management—a report *of the Association for Molecular Pathology*. *J Mol Diagn* 2015;**17**:623–34. edx.doi.org/10.1016/j.jmoldx.2015.07.004.
8. Dincer C, Bruch R, Kling A *et al*. Multiplexed point-of-pare testing–xPOCT. *Trends Biotechnol* 2017;**35**:728. dx.doi.org/10.1016/j.tibtech.2017.03.013.
9. Gubala V, Harris LF, Ricco AJ *et al*. Point of care diagnostics: status and future. *Anal Chem* 2012;**84**:487–515.
10. Ioannidis P, Papaventsis D, Karabela S *et al*. Cepheid GeneXpert MTB/RIF assay for Mycobacterium tuberculosis detection and rifampin resistance identification in patients with substantial clinical indications of tuberculosis and smear-negative microscopy results. *J Clin Microbiol* 2011;**49**:3068–70.
11. Popowitch EB, O'Neill SS, Miller MB. Comparison of the Biofire FilmArray RP, Genmark eSensor RVP, Luminex xTAG RVPv1, and Luminex xTAG RVP fast multiplex assays for detection of respiratory viruses. *J Clin Microbiol* 2013;**51**:1528–33.
12. Estrade C, Sahli R. Comparison of Seegene Anyplex II HPV28 with the PGMY-CHUV assay for human papillomavirus genotyping. *J Clin Microbiol* 2014;**52**:607–12.
13. Gaudet M, Fara AG, Beritognolo I *et al*. Allele-Specific PCR in SNP genotyping. In: Komar A. (ed.), *Single Nucleotide Polymorphisms*. Methods in Molecular Biology (Methods and Protocols), Vol **578**, Totowa, NJ: Humana Press, 2009, 415–24.
14. Mhlanga MM, Malmberg L. Using molecular beacons to detect single-nucleotide polymorphisms with real-time PCR. *Methods* 2001;**25**:463–71.
15. Chandler DP, Bryant L, Griesemer SB *et al*. Integrated amplification microarrays for infectious disease diagnostics. *Microarrays* 2012;**1**:107–24.
16. Muyzer G. DGGE/TGGE a method for identifying genes from natural ecosystems. *Curr Opin Microbiol* 1999;**2**:317–22.
17. O'Sullivan LA, Webster G, Fry JC *et al*. Modified linker-PCR primers facilitate complete sequencing of DGGE DNA fragments. *J Microbiol Method* 2008;**75**:579–81.
18. Hassibi A, Vikalo H, Riechmann JL et al. Real-time DNA microarray analysis. *Nucleic Acids Res* 2009;**37**:e132.
19. Vikalo H, Hassibi B, Hassibi A. Modeling and estimation for real-time microarrays. *IEEE J Sel Top Signal Process* 2008;**2**:286–96.
20. Johansson MK. Choosing reporter-quencher pairs for efficient quenching through formation of intramolecular dimers. *Methods Mol Biol* 2006;**335**:17–29.
21. Palmenberg AC, Rathe JA, Liggett SB. Analysis of complete genome sequences of human rhinovirus. *J Allergy Clin Immunol* 2010;**125**:1190–201. doi: 10.1016/j.jaci.2010.04.010.
22. Oberste MS, Maher K, Williams AJ *et al*. Species-specific RT-PCR amplification of human enterovirus: a tool for rapid species identification of uncharacterized enteroviruses. *J Gen Virol* 2006;**87**:119–28.
23. McAllister SC, Schleiss MR, Arbefeville S *et al*. Epidemic 2014 enterovirus D68 cross-reacts with human rhinovirus on respiratory molecular diagnostic platform. *PLoS One* 2015;**10**:e0118529. doi: 10.1371/journal.pone.0118529.
24. Sanchez JA, Pierce KE, Rice JE *et al*. Linear-After-The-Exponential (LATE)–PCR: an advanced method of asymmetric PCR and its uses in quantitative real-time analysis. *Proc Natl Acad Sci USA* 2004;**101**:1933–38.
25. Benson DA, Cavanaugh M, Clark K *et al*. GenBank. *Nucleic Acids Res* 2012;**41**:D36–42.
26. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;**32**:1792–97.

27. Schneider TD, Stormo GD, Gold L *et al*. Information content of binding sites on nucleotide sequences. *J Mol Biol* 1986;**188**:415–31.
28. SantaLucia J. Jr. Physical principles and Visual-OMP Software for optimal PCR design. In: Yuryev A (ed.), *Primer Design. Methods in Molecular Biology*. Totowa, NJ: Humana Press, 2007.
29. Ruijter JM, Ramakers C, Hoogaars WMH *et al*. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res* 2009;**37**:e45.
30. Stolovitzky G, Cecchi G. Efficiency of DNA replication in the polymerase chain reaction. *Proc Natl Acad Sci USA* 1996;**93**: 12947–52.
31. Dwight Z, Palais R, Wittwer CT. uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. *Bioinformatics* 2011;**27**: 1019–20.
32. Stormo GD, Schneider TD, Gold L *et al*. Use of the 'Perceptron' algorithm to distinguish translational initiation sites in E. coli. *Nucleic Acids Res* 1982;**10**:2997–3011.
33. Stormo GD. DNA binding sites: representation and discovery. *Bioinformatics* 2000;**16**:16–23.
34. Jin F, Fieguth P, Winger L *et al*. Adaptive Wiener filtering of noisy images and image sequences. In: *Proceedings of 2003 International Conference on Image Processing, Barcelona, Spain*, September 14–17, 2003, Vol-3, III-349. NJ, USA: IEEE.
35. Rizon M, Haniza Y, Puteh S *et al*. Object detection using circular Hough transform. *Am Jo F Appl Sci* 2005;**2**:1606–9.
36. SantaLucia J. Jr (2000) The use of spectroscopic techniques in the study of DNA stability. In: Gore MG (ed.), *Spectrophotometry and Spectrofluorimetry: A Practical Approach*. New York: Oxford University Press, 329–56.