UNIVERSITY OF CALIFORNIA, SAN DIEGO

Massively Parallel Digital High Resolution Melt for Rapid and Absolutely Quantitative Sequence Profiling

A Thesis submitted in partial satisfaction of requirements for the degree of Master of Science

in

Bioengineering

by

Sinead Hawker

Committee in charge:

Professor Stephanie Fraley, Chair
Professor Todd Coleman
Professor Ratneshwar Lal
Professor Shelley Lawrence

2016

The Thesis of Sinead Hawker is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Chair

University of California, San Diego

2016

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

gDNA	Genomic DNA
LB	Luria-Bertani
NGS	Next Generation Sequenceing
OVO SVM	One Versus One Support Vector Machine
PDF	Probability Density Function
qPCR	Quantitative Polymerase Chain Reaction
U-dHRM	Universal digital High Resolution Melt

LIST OF SYMBOLS

λ	Average number of positive reactions/total reactions within a Poisson Distribution

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Chapter 1-4 in part, have been submitted for publication of the material as it may appear in Scientific Reports, 2016, Sinead Hawker, Daniel Ortiz Velez, Behnam Hedayatnia, Ninad Kulkarni, Yang Zhang, Shelley Lawrence, Stephanie I. Fraley. The dissertation/thesis author was the primary investigator and author of this paper.

ABSTRACT OF THE THESIS

Massively Parallel Digital High Resolution Melt for Rapid and Absolutely Quantitative Sequence Profiling

by

Sinead Hawker

Master of Science in Bioengineering

University of California, San Diego, 2016

Professor Stephanie Fraley, Chair

High Resolution Melt (HRM) analysis allows for the classification of DNA sequences based on the way the sequences melt when exposed to heat in the presence of an intercalating dye. While HRM has been shown to be effective in broad-based sequence identification, it has remained limited in its sensitivity and is unable to

provide information regarding absolute quantification of samples. Here, HRM is combined with digital PCR by isolating individual templates into separate wells and melting the end-point product, producing a highly sensitive technique capable of sequence identification and absolute quantification in multiplexed samples. By decreasing the reaction volume by more than 99% and increasing the number of reactions to 20,000, this novel platform offers new detection capabilities not previously seen in standard qPCR HRM. By optimizing temperature resolution, reagent concentrations, sampling rate, and microscope exposure settings, the platform's performance was shown to rival the resolution demonstrated by qPCR HRM. This resolution is of particular importance in the identification of bacteria based on their sequence-specific melt curves. In a proof of concept, by using the universal amplification of the 16s rRNA gene of bacteria, this dHRM platform is demonstrated to be capable of differentiating between two strains of bacteria based on the fingerprint melt curves generated from single-molecule amplification. By training a machine-learning support vector algorithm (SVM) coupled with an automated image analysis code, the platform successfully classified separate strains within a mixed culture of bacteria. These results suggest that this technology may have broad applications for sensitively and rapidly profiling bacterial populations, particularly in clinical cases of infection.

INTRODUCTION

Although there are more than 750,000 cases of sepsis a year in the United States (1-7), advances in treatment remain limited and mortality remains relatively high at 20-30% (1-7,1-26). Studies point to the importance of rapid treatment in patient prognosis (2), but without the identification of the specific strain causing infection treatment is limited to broad spectrum antibiotics, which prove largely ineffective (3). The problem is compounded by antibiotic resistant bacteria and polymicrobial infections, where nonspecific antibiotics may only partially effective and leave the bulk infection unaffected. The heterogeneity introduced by the variety of microbial infections that lead to sepsis remains the main hindrance in effective treatment as a broad-based treatment cannot account for every phenotype (3), but the identification of specific strains still relies heavily on blood culture, a slow process involving the culturing and phenotypic identification of the colonies on culture plates (4). This process takes between 72 and 96 hours, with a successful positive identification occurring only 30-40% of the time (4). Since without treatment during this time the patient would have died, most clinicians will treat sepsis with broad-spectrum antibiotics and only use the microbial culture results to fine-tune later antimicrobial treatments (4-8). Clearly a more rapid, sensitive method of microbial detection and identification is needed to improve patient prognosis.

Regardless of its limitations, blood culture has remained a “gold standard” of sepsis diagnostics due to the inherent difficulties associated with identifying bacteria from a blood sample. The vast number of strains with the potential to cause sepsis.

requires a diagnostic tool to be highly broad-spectrum, able to store and choose from a large library of clinically relevant strains. Polymerase chain reaction (PCR) based assays have been developed to detect specific strains, based on the principal that sequence-specific primers can identify the presence of genes specific to each bacteria. The need for a broad-based assay introduces a complication in PCR assays, as they require specific primers or probes to be able to detect specific bacterial strains. It is prohibitively difficult to design a separate assay for each bacteria potentially causing a septic infection, as the amount of bacterial DNA extracted from a septic patient is far to low to run a large series of assays on. Indeed, the concentration of bacteria in blood leading to a septic response is only 1 to 100 CFU/mL in adults (5) and less than 10 CFU/mL in neonatal babies (6). This requires that a diagnostic assay be highly sensitive and extremely accurate, as retests will not be possible. To further complicate the issue, some septic infections are caused by a polymicrobial infection (7), meaning an assay needs to be capable of detecting and reporting multiple species of bacteria simultaneously. Blood culture assays are able to meet these requirements, explaining their widespread use despite the limitations associated with them.

That said, quantitative PCR (qPCR) is gaining popularity as a diagnostic tool for sepsis as universal assays are developed to overcome the need for bacteria specific primers/probes. The 16s rRNA gene of bacteria contains conserved regions between different strains of bacteria, allowing for the design of primers universal to all bacteria (8,9). Within these conserved regions are variable regions, which can be used to distinguish between different strains based on variance in the sequences. Already, 16s

rRNA universal primers are used in combination with sequencing as a diagnostic tool, though again this type of assay cannot differentiate within polymicrobial infections (9). The nature of qPCR assays means that the bacteria with the highest starting concentration will overtake the reaction and low-level bacteria will not be detected in the final sequencing. Furthermore, standard sequencing reads are limited to approximately 500 base pairs, which does not encompass the full length needed to differentiate between strains of bacteria.

To combat this, assays using high resolution melt curve (HRM) analysis have been developed using 16 ribosomal DNA universal primers (8,10). HRM is not limited to short sequence reads, and is relatively inexpensive and fast. These assays are based on the principal that minor sequence differences lead to changes in the temperature that a DNA sequence melts at when slowly heated. By heating DNA in the presence of an intercalating dye, the fluorescence can be measured to generate a relationship between fluorescence and temperature; i.e. a melt curve. These melt curves have been shown to be capable of differentiating between different strains of bacteria, much the same way that sequencing does (8,10). Furthermore, by combining HRM with digital PCR, multiplexed detection within polymicrobial infection is possible (8, 10). Digital PCR involves the sequestering of individual template molecules into separate reactions allowing for absolute quantification and the amplification of pure templates even within mixed samples. While a very powerful technique, dPCR relies on sometimes faulty thresholds and does not allow for multiplexed detection. This is because current dPCR assays provide a final positive-reaction count but provide no

information about the sequences that amplified. However, by introducing HRM to a dPCR format, absolute quantification and identification within a mixed sample can be accomplished. DPCR allows for the amplification of single molecules of bacterial ribosomal DNA, while HRM allows for the classification of each positive reaction based on sequence differences in their variable regions.

While this combination, termed Universal digital High Resolution Melt (UdHRM), has been demonstrated previously (8,10), the purpose of this thesis is to demonstrate a drastic increase in dynamic range through decreasing reaction volume by 96% and increasing the number of reactions to 20,000. With this technology, thousands of bacteria can be simultaneously profiled and rapid diagnosis of polymicrobial septic infections can be achieved in a clinical setting. This thesis demonstrates the design of the UdHRM platform, the optimizations used to increase the resolution of the platform to the level of a standard qPCR machine, and a proof of concept demonstrating the ability of the platform to differentiate between multiplexed bacterial strains.

CHAPTER 1 – BACKGROUND

Universal digital High Resolution Melt (UdHRM) is based on the well established techniques of high resolution melt (HRM) and digital PCR (dPCR). HRM has gained popularity as a rapid, inexpensive, closed-tube post-PCR sequence characterization technique. By precisely heating and unwinding double stranded DNA amplicons in the presence of a fluorescent intercalating dye(12-14) or sloppy molecular probes(15, 16), loss-of-fluorescence melt curves are generated that are sequence-specific. Several groups have proposed the expansion of HRM from screening to profiling technology by preceding it with broad-based PCR(17). Using conserved priming sites that flank sites of genetic variation or mutation, sequence changes in a genetic locus can be identified by changes in the melt curve signature of the gene amplicon. This replaces the need for precise primer or probe hybridization-based discrimination of small sequence changes and relies only on the intrinsic melting properties of the amplified sequence itself.

The strength of a HRM platform lies in its ability to differentiate between single nucleotide differences between sequences. This relies on the resolution and repeatability of both the temperature ramp and the fluorescence intensity measurement. Standard HRM platform temperature ramp rates range from 0.005 to 0.3 °C/s, while sampling rate ranges from 10 to 67 data points/°C (11). A study into the genotyping accuracy across multiple HRM platforms determined that there is a slight positive correlation between increased ramp rate and genotyping accuracy, with a stronger correlation between increased sampling rate and genotyping accuracy (11).

Heterozygote sequence differences (pyrimidine to purine, or vice versa) had a universally high detection rate at 99.7%, whereas homozygote sequence differences (pyrimidine to pyrimidine or purine to purine) were more difficult to detect at 70.3% due to the similarity in bond strength and therefore melting temperature (11). Regardless, the study demonstrated the strength of HRM in its ability to differentiate down to single nucleotide differences under the right conditions and pointed toward the importance of ramp and sampling rate in the accuracy of a HRM platform.

It has been shown that HRM can enable single nucleotide specificity for one of the most difficult profiling tasks, the discrimination of microRNA in the Lethal-7 family(18). In this study, universal tags were ligated to synthetically synthesized microRNA to allow for universal primers to amplify all sequences equally. Differences in curve shape and melting temperature were used to differentiate between sequence differences as small as one nucleotide, demonstrating the potential for HRM as a profiling technique. Furthermore, the study showed that UdHRM is capable of differentiating between microRNA within a multiplexed sample of multiple sequences (18). While a powerful demonstration of this technique, the study is limited in the need for ligated tags in order for universal primers to be designed, as well as the use of synthetic sequences rather than microRNA extracted from cells.

Broad-based HRM methods have also been proposed for a variety of other applications, for example: to identify oncogenic mutations(19), gene methylation patterns(20, 21), and, most relevant for this thesis, for bacterial identification(22-27). Although generally reproducible melt curves are obtained, in-run template standards

are typically used to overcome run-to-run variability and enable curve matching. This has significantly limited the application of HRM by constraining the breadth of profiling to only a few sequence variants, forcing user intensive curve identification procedures, and reducing specificity since single nucleotide changes often manifest as slight temperature or curve shape changes. Likewise, if multiple sequence variants are present or when contaminating DNA is present in a high load, as often occurs in real samples, individual sequences cannot be identified in the conventional bulk HRM format(17, 28).

These limitations on HRM profiling sensitivity, breadth, reproducibility, and specificity have been addressed by adapting and integrating limiting dilution digital PCR (dPCR) with universal amplification strategies and temperature calibrated HRM(18). This technique involves diluting the sample to the extent that it can be safely assumed that at most one template is present in each reaction. As dPCR relies on end-point amplification, it is more resilient to variations resulting from different primer efficiencies which plague qPCR assays. This same concept prevents variations resulting from the use of different master mixes from affecting the copy number determined from dPCR. Furthermore, dPCR provides absolute quantification rather than the relative quantification given by qPCR. Though a highly useful technique, dPCR is limited in that it relies on fluorescence thresholds to differentiate between positive and negative reactions, which can be faulty and lead to incorrect counts. Its ability to provide information about multiplexed samples is limited to the number of differentiable dyes linked to specific probes, which means it cannot be used as a

broad-based method of sequence identification. The specificity of dPCR is limited to the specificity of the primers and probes as misamplification will be detected as a positive reaction with no ability to check whether the correct sample was amplified. This problem is compounded by the large number of cycles needed to reach end-point amplification, as the more cycles used in PCR the more likely misamplification becomes. These limitations are addressed by combining dPCR with the sequence analysis of HRM (dHRM).

DHRM allows for sequence analysis following amplification, which enables a more accurate threshold to be applied based on melt curve, minimizing incorrect classification resulting from misamplification. Furthermore, the differentiation of multiplexed sample of a much greater magnitude is possible by sequestering the individual templates into separate reactions producing pure melt curves. These pure melt curves can then be identified based on a previously generated library. Similar to conventional HRM, dHRM is plagued by variations in melt curves due to variations in reaction-to-reaction temperature. This has been addressed through temperature calibration, which is accomplished through using synthetic sequences of a known melting temperature. Slight variations between experimental runs are accounted for by aligning the calibrator peaks to the same temperature point. Variations in reproducibility are minimized through a series of normalizations based on baseline and curve area. This allows for differences in curve shape and melt temperature to be maximized while differences in melt curves due to noise are minimized. Separately, a machine learning techniques using nested, linear kernel, one versus one support vector

machines (OVO SVM) has been developed to automatically identify sequences by their melt curve signatures despite inherent experimental variability(29, 30). SVMs work by determining the optimal hyperplane to separate data sets, and thereby determine whether a new point belongs to one data set or another. The optimal separating hyperplane is characterized by maximizing the margin between the training data sets. SVMs can be nested within each other to create layers of classification to differentiate between more complicated data. This is demonstrated by the multiple points that can be used to differentiate between melt curves, such as the peak melting temperature or the curve area. A linear kernel SVM defines the separating hyperplane as a linear plane, requiring that the data be well differentiated but decreasing computational time. A one vs. one SVM trains a separate classifier for each pair of classes, which while computationally expensive is able to handle unbalanced or complex data sets with more sensitivity. Using these types of SVM, a highly sensitive machine learning algorithm has been developed to reliably classify melt curves based on their shape and melt temperature. Using this SVM approach, universal digital HRM (U-dHRM) has been shown to be capable of automatically identifying multiple distinct and even novel genotypes in a mixture with single molecule sensitivity and single nucleotide specificity (7).

Others have also demonstrated the ability of dHRM to sensitively detect rare mutants/variants(31, 32) and importantly, novel variants(33). In one study, the heterogeneity of methylation in gene silencing was studied using dHRM. Methylation can be studied using melt curves due to different base compositions between

methylated and unmethylated DNA following bisulphite conversion. The study showed that while bulk HRM was useful in studying overall methylation in gene silencing, it was incapable of detecting heterogeneously methylated samples as the multiple PCR products cannot be differentiated in bulk (31). dHRM allowed for a single molecule to be amplified eliminating the competition between alternately methylated molecules, and thereby enabling the study of heterogeneous samples. The CDKN2B gene in particular was shown to contain a large range of methylation sites leading to a diverse spectrum of potential melt curves, which is of importance in the study of tumor suppression and malignancy (31). In another study, dHRM was shown to be significantly more sensitive than bulk HRM in the detection of gene mutations leading to colorectal cancer. They found that the detection limit of their bulk HRM was 10%, and so were able to detect single molecule mutations by limiting the number of template molecules to at most 10 per reaction (32). In yet another study, novel variants of the Lyme-disease bacteria *Borelia burgdorferi* were discovered by detecting abnormal melt curves that did not match the melt temperature and shape expected for each serotype. Once identified, the novel samples were sequenced and the corresponding phenotype analyzed (33). This demonstrates the powerful ability of dHRM as a rapid first level of detection when looking for novel variants, which can then be followed up with a deeper level of analysis. These findings suggest that U-dHRM could outperform microarray, qPCR array, and NGS sequence profiling if implemented at full digitization, where each single molecule in a sample could be

identified and quantified, while maintaining the ease of use and inexpensive benefits of bulk HRM.

Unlike digital PCR, the sensitivity and quantification power of dHRM profiling relies on full digitization of the sample, i.e. spreading the sequence mixture across so many reactions that each target molecule is isolated from all others. Since the process of loading DNA into wells is stochastic at limiting dilutions, the dynamic range of single molecule detection follows a Poisson distribution, requiring the total number of reactions to be approximately 10 to 100 times the number of sequence molecules. That is, the average occupancy (lambda) across all reactions must be 0.1 to 0.01 copies of DNA per well. The probability of DNA occupancy in any well, i.e. the fraction of wells having 1, 2, 3, etc. copies, is given by the Poisson probability distribution $P=(e^{-l}*l^{n})/n!$, where n is the total number of wells. DHRM is currently performed in traditional well plates using HRM enabled qPCR machines. In this format only about 9 molecules in a sample can be profiled per 96-well plate (Fig. 1A). Therefore, a major challenge to the advancement of dHRM profiling is the need for an exponential increase in the number of reactions to achieve scalability for realistic sample concentrations.

To combat this, several commercial platforms have been developed to incorporate tens of thousands of parallel PCR reactions. Several reports have demonstrated the use of microfluidic chambers or droplets for digital PCR, though these platforms have limitations preventing their extension to digital HRM. Fluidigm has introduced a series of microvalve-based digital PCR devices, which function

through a series of channels and valves embedded within a silicon substrate. The silicon is partitioned to allow parallel PCR amplification of up to an impressive 36, 960 reactions. While this platform further boasts the ability to capture fluorescence after each cycle allowing for quantitative dPCR, Fluidigm platforms do not have the high resolution heating blocks necessary for high resolution melt curve generation and moreover are not programmed to capture fluorescence during heat ramping. Microfluidic droplet-based digital PCR devices, such as Bio-rad's droplet dPCR, are based on a water-oil emulsion method of sequestering the PCR sample into 20,000 droplets. The droplets are then cycled as in conventional PCR, and analysis is performed in the form of endpoint PCR detection in a continuous flow format. This platform is limited in that analysis is performed without temperature control, one droplet at a time, preventing in-situ, real-time monitoring of fluorescence in droplets, as is needed by HRM. Moreover the quantitative capabilities of this platform are questionable due to the large dropout rate inherent in droplet dPCR. Current microfluidic chip-based digital PCR devices, such as Life Technologies' 20k dPCR chip, use wells formed from a silicon substrate with a hydrophilic treatment to draw the PCR master mix in with a high efficiency. These chips are then cycled and imaged for end-point PCR quantification. Unfortunately, as they are used currently they rely heavily on TaqMan probes and do not perform real-time monitoring of fluorescence, both of which prevent HRM analysis. DHRM requires a highly repeatable high resolution heating set-up as well as the ability to capture fluorescence intensity simultaneously with temperature ramping, which has not yet been demonstrated by

commercial platforms.

To address these challenges, this thesis presents a platform for massively parallelized microfluidic dHRM and integrated this with a U-dHRM approach and machine learning curve identification algorithm, achieving single molecule detection and absolute quantification of thousands of DNA molecules in a single sample in under 3 hours. DPCR is accomplished using a commercially available dPCR chip combined with a modified protocol to achieve consistent, full-length amplification. HRM is performed by using a custom built heating set-up coupled with a powerful fluorescent microscope to simultaneous heat and image the chip. A proof of principle using universal primers for bacterial 16 ribosomal DNA demonstrates that this platform enables absolute sensitivity and specific characterization of polymicrobial samples with the benefit of absolute load quantification of each species. Importantly, this proof of principal is completed at target concentrations relevant to clinical scenarios, in particular sepsis. This demonstrates the ability of this UdHRM platform as a clinical diagnostic tool for septic patients, as it is able to absolutely identify and quantify polymicrobial infections allowing for specific antibiotic treatments within hours of sample collection.

Acknowledgements

Chapter 1-4 in part, have been submitted for publication of the material as it may appear in Scientific Reports, 2016, Sinead Hawker, Daniel Ortiz Velez, Behnam Hedayatnia, Ninad Kulkarni, Yang Zhang, Shelley Lawrence, Stephanie I. Fraley. The dissertation/thesis author was the primary investigator and author of this paper.

CHAPTER 2 – EXPERIMENTAL DESIGN

High-Content dPCR chip

The first portion of the dHRM platform requires highly reproducible massively paralleled PCR reactions. In order to achieve high-content digital partitioning, we loaded the sample into a commercially available QuantStudio 3D Digital PCR 20K Chip v2 (Applied Biosystems). We chose to use a commercially manufactured chip for performance reliability and ease of scalability. The chip contains 20,000 picoliter-scale wells manufactured from silicon with a hydrophilic treatment to allow the PCR master mix to be drawn into each well with a high efficiency. A PCR-grade oil is spread over the loaded chip to prevent sample evaporation during cycling. We then sealed the chip with an adhesive lid containing an optical window, which allows for later imaging and the generation of melt curves. While we found the chips to be highly effective, the mastermix suggested by the manufacturer was inadequate for our needs. We therefore coupled the dPCR chip with our own custom designed master mix. The master mix is optimized to consistently amplify full length ~1,000bp templates of the 16s gene and produce high fluorescence signal intensity for melt curve analysis while maintaining optimal surface tension for easy loading. A 1,000bp template is longer than standard qPCR, requiring a higher dNTP to primer ratio in order to have full amplification. Furthermore, a polymerase capable of amplifying long templates was required. As the dPCR chip relies on hydrophobic/hydrophilic interactions to effectively sequester the master mix into separate reactions, the surface tension of the master mix must remain optimal. Using a droplet surface tension test our mastermix

was compared with the mastermix supplied by the manufacturer, allowing us to customize our mastermix while maintaining the required surface tension. The small reaction volume further complicates the master mix, as an increase of dye concentration is required in order to boost the fluorescent signal strength and decrease the signal-to-noise ratio. By using a buffer containing surfactant and increasing the concentration of dye, we were able to effectively load, amplify, and visualize the sample. We used a flatbed thermocycler (MJ Research PTC-200) for endpoint amplification following standard qPCR cycling protocols. The chips are loaded at 42^{o}C on the thermocycler to increase loading efficiency. The thermocycler is tilted at a 30-degree angle to collect the bubbles generated at high temperatures in the PCR-oil in an air pocket located outside of the chip's sample region, preventing sample evaporation from the small volume reactions. Once amplified, the chip is highly stable and can be stored at room temperature without degradation of the amplified DNA within.

Bacterial gDNA Isolation

In order to demonstrate the efficient amplification of our chip/master mix combination and provide a proof of concept for the differentiation of bacterial strains within a multiplexed sample, we isolated bacterial gDNA for further amplification. We used the Wizard Genomic DNA Purification Kit (Promega) to isolate gDNA from an overnight culture of bacteria following the manufacturer's protocol. Briefly, a culture of bacteria was grown up overnight, spun down and lysed, and then purified through RNase treatement and protein precipitation. The purity and concentration of

the DNA was determined using an Eppendorf Spectrofluorometer. This allowed us to estimate an approximate copy number of our sample, giving us the dilution necessary to digitize the sample as well as ensure that our sample had a high enough purity to prevent inhibition by carry over from the isolation process. The template was diluted in PCR water to the desired concentration based on this measured concentration.

PCR

The optimum PCR master mix for amplification on a chip combined with effective loading and high signal-to-noise ratio contained in a 14.5 uL reaction: 1X Phusion HF Buffer containing 1.5 mM MgCl2 (Thermo Fisher Scientific), 0.15 uM forward primer 5'-GYGGCGNACGGGTGAGTAA-3' (IDT), 0.15 uM reverse primer 5'-AGCTGACGACANCCATGCA-3' (IDT), 0.3 uM low temperature calibrator (IDT), 0.2 mM dNTPs (Invitrogen, Carlsbad, CA), 2.5x EvaGreen (Sigma), 2x ROX (Thermo Fisher Scientific), 0.02 U/uL of Phusion HotStart Polymerase (Thermo Fisher Scientific), 1 uL of sample, and ultra pure PCR water (Quality Biological, Gaithersburg, MD) to bring the total volume to 14.5 uL. The low calibrator at 0% GC was chosen to calibrate the chips as it had the greatest separation from the amplicon, minimizing its influence on the amplicon melt curve. The calibrator melts at a known temperature so that the melt curves from separate chips could be aligned to the same temperature, minimizing run-to-run variation in temperature readings. Phusion was chosen due to its ability for continuous amplification of long amplicons and because the phusion buffer contains a surfactant lowering the surface tension and enabling easy loading. The hot start adaptor on the enzyme allows for the loading of multiple chips

without the degradation of the template DNA or primers from the passive activity of the polymerase. The concentration of the dyes were chosen based on the highest concentration that could provide a strong fluorescent signal while still efficiently amplifying the full length template. The dPCR chip was cycled on a flatbed thermocycler with the following cycle: an initial enzyme activation (98 degrees C, 30 s), followed by 70 cycles (95 degrees C, 30 s, 59 degrees C, 30 s, 72 degrees C, 60 s). Temperature calibrator sequences with varying GC content used for system optimization are as follows: 0% GC (TTAAATTATAAAATATTTATAATATTAATTATATATATATAAATATAATA-C3), 12% GC (TTAATTATAAAGGTATTTATAATATTGAATTATACATATCTAATATAATC-C3), and 76% GC (GCGCGGCCGGCACCCGAGACTCTGAGCGGCTGCTGGAGGTGCGGAAGCG GAGGGGCGGG-C3)(7).

Chip Heating Device

After the chip is loaded and amplified, dHRM is performed to generate the melt curves for later analysis. DHRM is made possible through simultaneous fluorescent imaging and heating of the dPCR chip. The heating device consists of a high-temperature peltier chip (TE technology, inc. Traverse city, MI) powered by a 14V power source controlled via an Arduino UNO-based interface that uses pulse width modulation to generate a temperature ramp. Temperature ramp is controlled by step-increases in power followed by variable periods of holding at each power setting.

The power increases and periods of holding are set by a MATLAB code controlling the Arduino UNO. The peltier chip is in direct contact with a copper plate onto which the dPCR chips coated with thin layer of thermal paste are clamped. This allows for even heat distribution and optimal surface contact. On the reverse side of the peltier, another copper plate screwed to an aluminum heat sink is attached to enable fast excessive heat dissipation. To measure the temperature at each point during the temperature ramping, a type K thermocouple (OMEGA, Stamford, CT) is used. The thermocouple is fixed inside a surrogate chip, which is attached alongside the sample chip to the copper plate. The temperature readings are acquired by the microscope imaging software (NIS Elements) and are attached to the image file metadata for offline analysis. The complete chip-heating setup is placed in a custom designed 3D printed stage adapter to securely mount the device on the microscope for imaging.

Fluorescent Imaging

Fluorescent imaging is accomplished using a Nikon Eclipse Ti platform customized for our dHRM system. A Nikon Plan/Fluor 4X objective with a numerical aperture of 0.13 and a working distance of 16.5X minimizes the number of images and time required to scan the entire chip, while still giving a strong signal and clear image of the wells. A Lumencor LED Light Engine capable of producing 3-4W of visible white light from the range 380nm to 680nm is used as a light source. Images of the loading control dye, ROX, and melt curve intercalating dye, EvaGreen, are captured with 488/561nm and 405/488nm excitation/emission filters using an exposure time of 100 ms. Images are captured using a Hamamatsu digital camera, C11440 ORCA-

Flash4.0. NIS Elements Software is programmed to automatically image the chip as the heating device ramps using the following workflow: define the capture settings for the ROX and EvaGreen channels, set the stage area to the chip's sample area, generate points within that stage area to image, and run time lapse to image each location for every time point. A Prior NanoScanZ motorized stage is used to scan and image the entire chip automatically via the software. For every image, the microscope automatically records the temperature registered by our temperature probe within the metadata of the image. This allows for continuous scanning of the chip and recording of the fluorescence intensity in each well while concurrently heating the chip to generate 20,000 melt curves.

Image analysis & SVM

In order to generate melt curves from the acquired fluorescent images of the dPCR chip, we implement an automated image processing algorithm in MATLAB. The algorithm first generates a binary mask for each temperature point to identify the centroid corresponding to each digital reaction well in the field. The algorithm then records the pixel intensity of the 441 neighboring pixels from the images of both the Evagreen channel and the ROX channel. The average pixel intensity for each well is plotted against the measured temperature to generate the raw melt curve. Each melt curve is normalized to the ROX channel to account for any differences due to unequal loading. A threshold is applied to remove negative reactions and all incorrectly identified centroids. A Gaussian filter is then applied to smooth the curves and the derivative of the curves is taken with respect to temperature to obtain –

d(Fluoresence)/dT. The curves are normalized through area normalization, exponential normalization, and max-min normalization to minimize run-to-run differences. Finally, the curves are aligned via a temperature independent melt curve alignment at 0.1 –dF/dT. This allowed the differences in melt curve shape to be maximized for later identification using a previously developed one versus one (OVO) support vector machine (SVM) algorithm(30). Briefly, a OVO SVM creates a maximal margin separating hyperplane between two data classes (i.e. melt curve signatures) using the Least Squares Method. All combinations of OVO SVMs were created with the training data generated from melt curves of a known origin. During classification, a scoring method is applied and the most frequently called classification is chosen as the final melt curve identity.

Two approaches to thresholding reaction fluorescence for the identification and removal of negatives were compared. The typical dPCR approach of thresholding total reaction fluorescence proceeded as follows: The total fluorescence intensities, at room temperature, of all reactions on a chip were plotted as a histogram in MATLAB, and the probability density function (PDF) for a mixture of two normal distributions was applied to identify negative and positive reaction distributions. Then a threshold was applied at the lowest point of intensity where the two distributions intersected (Fig. 4B, top). This was performed for each chip individually, such that each chip underwent a unique threshold depending on its PDF.

A second approach was developed to identify a Tm threshold that separated off-target amplified reactions from true positives more accurately. First, raw melt

curves were converted to derivative melt curves. On fully loaded chips where all reactions were positive (2 training chips) all reactions contained 16S amplicons, which were observed to melt with an average Tm of 89^{o}C. On digitized chips (3 chips, testing data), off-target amplicons were observed to melt at a much lower temperature, average Tm of 81^{o}C, while positive 16S amplicons melted reproducibly in the same range as the training chips. For thresholding analysis, the maximum peak height (Tm) was found for each derivative melt curve between the range of 75^{o}C and 93^{o}C. Then a histogram of the Tms was plotted in MATLAB, and the PDF for a mixture of two normal distributions was applied (Fig. 4B, bottom). Finally, the Tm threshold was chosen at the minima between the two distributions. Reactions melting below this Tm threshold were identified as negatives, while those melting above the threshold were identified as positives. This threshold was determined using this method for each chip individually.

Cell Culture

Clinically isolated *Streptococcus pneumoniae* and *Listeria monocytogenes* were grown separately overnight in Luria-Bertani (LB) broth. Sterile conditions were used to ensure uncontaminated growth of each bacteria.

Acknowledgements

Chapter 1-4 in part, have been submitted for publication of the material as it may appear in Scientific Reports, 2016, Sinead Hawker, Daniel Ortiz Velez, Behnam Hedayatnia, Ninad Kulkarni, Yang Zhang, Shelley Lawrence, Stephanie I. Fraley. The dissertation/thesis author was the primary investigator and author of this paper.

CHAPTER 3 – RESULTS

Digital HRM Device Concept

In order to determine the number of reactions we would need for our digitizing chip, we first considered a model real-world application. In neonatal bacteremia cases, clinically relevant bacterial loads are estimated from culture techniques to be between 1 to ~2,000 cfu per blood sample (1ml), where 76% of samples have <50 cfu (42, 43). This load requires ~20,000 reactions to provide a dynamic range of detection up to 1,810 bacterial genomic DNA (gDNA) molecules at the single molecule level (Fig. 2A, pg 24). A digitizing chip fitting this scale is commercially produced for single target dPCR applications (see Methods), and as such was chosen as a robust and reliable digitizing device.

To identify digitized bacterial gDNA, universal primers targeting the 16S rRNA gene were used. The 16S harbors conserved sequence regions (priming sites) flanking hypervariable regions that are unique to different genus and species of bacteria(44) (Fig. 1). This gene was targeted to generate bacteria-specific amplicons

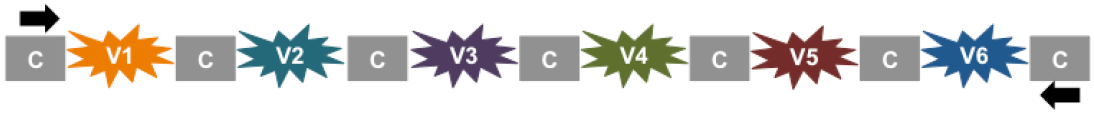


Figure 1: Bacteria 16s rRNA Gene. Regions conserved and therefore universal between all bacteria are shown in grey, with the primers marked as black arrows. The variable regions are colored and numbered to demonstrate the amplicon. This allows for universal amplification leading to amplicons specific to each bacterial strain.

for HRM profiling. Specifically, our long amplicon (~1,000bp) 16S bulk HRM assay(38) was adapted (see Methods) to enable successful digital amplification and

reliable dHRM in each of the 725 picoliter reactions on-chip, a 99.995% volume reduction compared to the typical qPCR format. To enable massively parallel dHRM across the 20,000 reactions, we developed a custom high resolution heating device and imaging system. A schematic of our design is shown in Figure 2B. Precise chip heating for HRM was accomplished using a thermoelectric peltier with arduino controller, power supply, and heat sink. A copper plate was attached between the peltier and the chip and between the heat sink and the peltier to evenly distribute heat. A custom adapter was designed to secure this chip-heating setup onto a microscope stage for imaging. Figure 2C shows an image of the integrated heating device, chip, and stage adapter. The microscope was equipped with a 4x objective, red and green fluorescence channels, and an automated x,y stage such that the 20,000 reactions could be rapidly imaged as four tiles stitched together at each temperature point during the dHRM heat ramp. An image analysis program was developed to stitch tiled images, align reaction centroids to overcome image drift due to thermal expansion, and extract raw fluorescence data from each reaction simultaneously (Fig. 2D). Our previously developed support vector machine algorithm (SVM) was adapted to classify and quantify dHRM curves after being trained on melt curves generated on-chip. The digital chip, chip heating device, fluorescent imaging system, control electronics, and analysis algorithms for image processing and SVM melt curve identification were integrated to enable massively parallel dHRM for absolutely quantitative bacterial profiling.

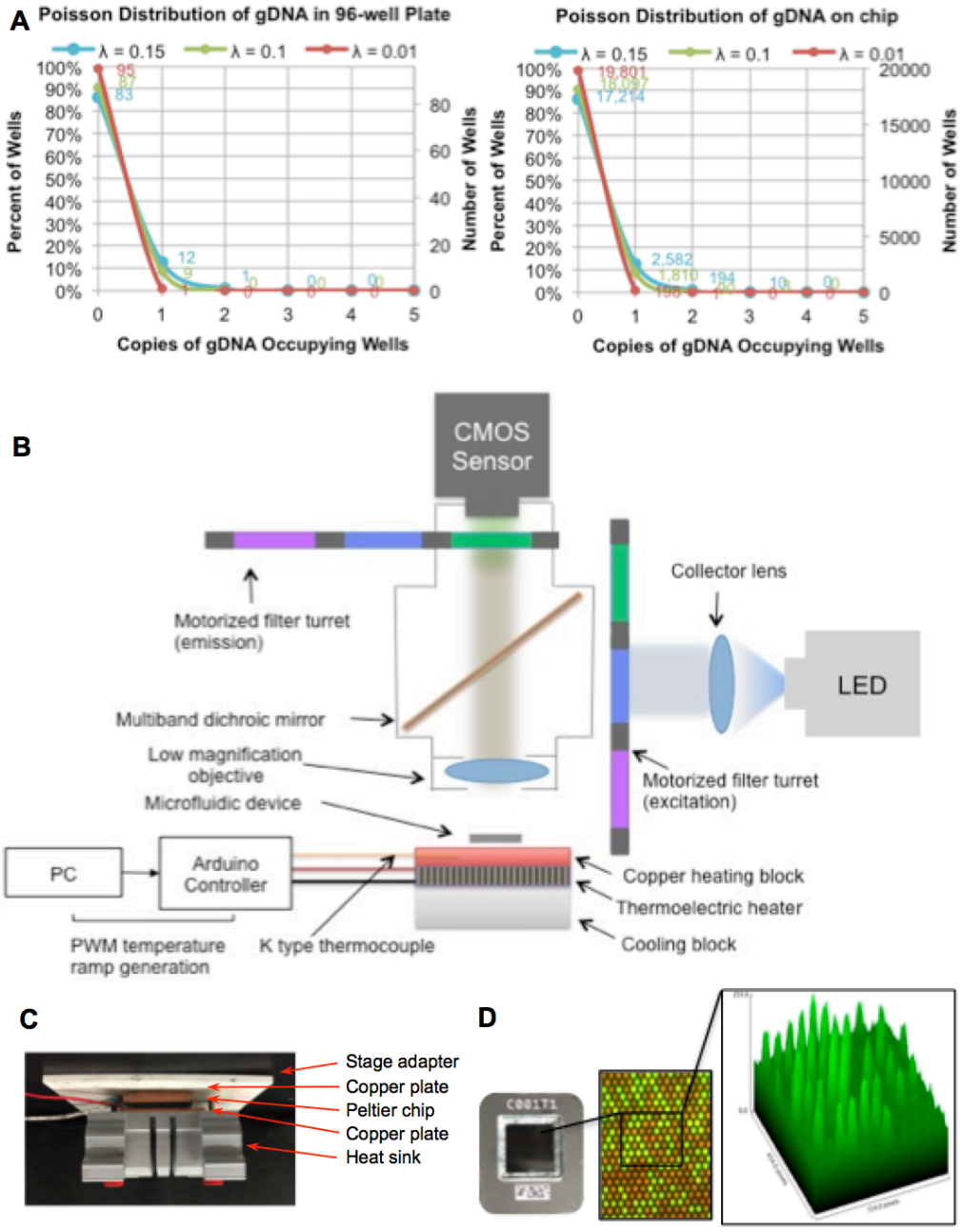


Figure 2: Massively Parallel Digital HRM Device. (A) Poisson distribution of gDNA in a 96-well plate versus a 20,000 well digital PCR chip, showing the distribution of molecules per well. **(B)** Schematic of the digital HRM platform. **(C)** Image of the actual dHRM platform. **(D)** Fluorescent image of a small portion of chip where background dye (red) and intercalating dye (green) are overlaid. 3D intensity plot of the green channel is shown in inset.

System Characterization & Optimization

The challenge of generating highly resolvable dHRM curves in picoliter-scale reactions was first approached by maximizing fluorescent intercalating dye concentrations to maximize signal. An Evagreen concentration of 2.5X was found to be the highest concentration that did not inhibit amplification on-chip. Next, the simultaneous imaging and heating process of melt curve generation (Fig. 3A) was tuned using three synthetic DNA sequences containing 0% GC, 12% GC, and 76% GC with known melting temperatures (Tms) (Fig. 3B). The greater the GC content, the higher the temperature required to melt the DNA due to higher bond strength. After loading mixtures of these three sequences onto a chip, we performed preliminary calibrations of our device, optimizing exposure time to minimize photobleaching while maintaining the highest possible signal to noise ratio. We also used these initial readings to develop our image analysis algorithm (see Methods). Figure 3B shows the normalized fluorescence versus temperature and derivative melt plots for the three calibrator sequences in qPCR and dHRM formats. The temperature calibrators are predicted to melt at 57.3°C, 62.8^{o}C, and 92.9^{o}C by melt curve prediction software, uMELT(22). The average Tms given by qPCR HRM were 56.9^{o}C, 67.4^{o}C, and 90.5^{o}C respectively, while the dPCR Tms were 55.5^{o}C, 64.6^{o}C, and 83.4^{o}C. These readings indicated that further temperature ramp optimization was necessary. Improved temperature resolution was achieved by varying the heating ramp rate until a linear and repeatable relationship between voltage and temperature could be maintained throughout our temperature ranges of interest, 45-95^{o}C. For highest accuracy,

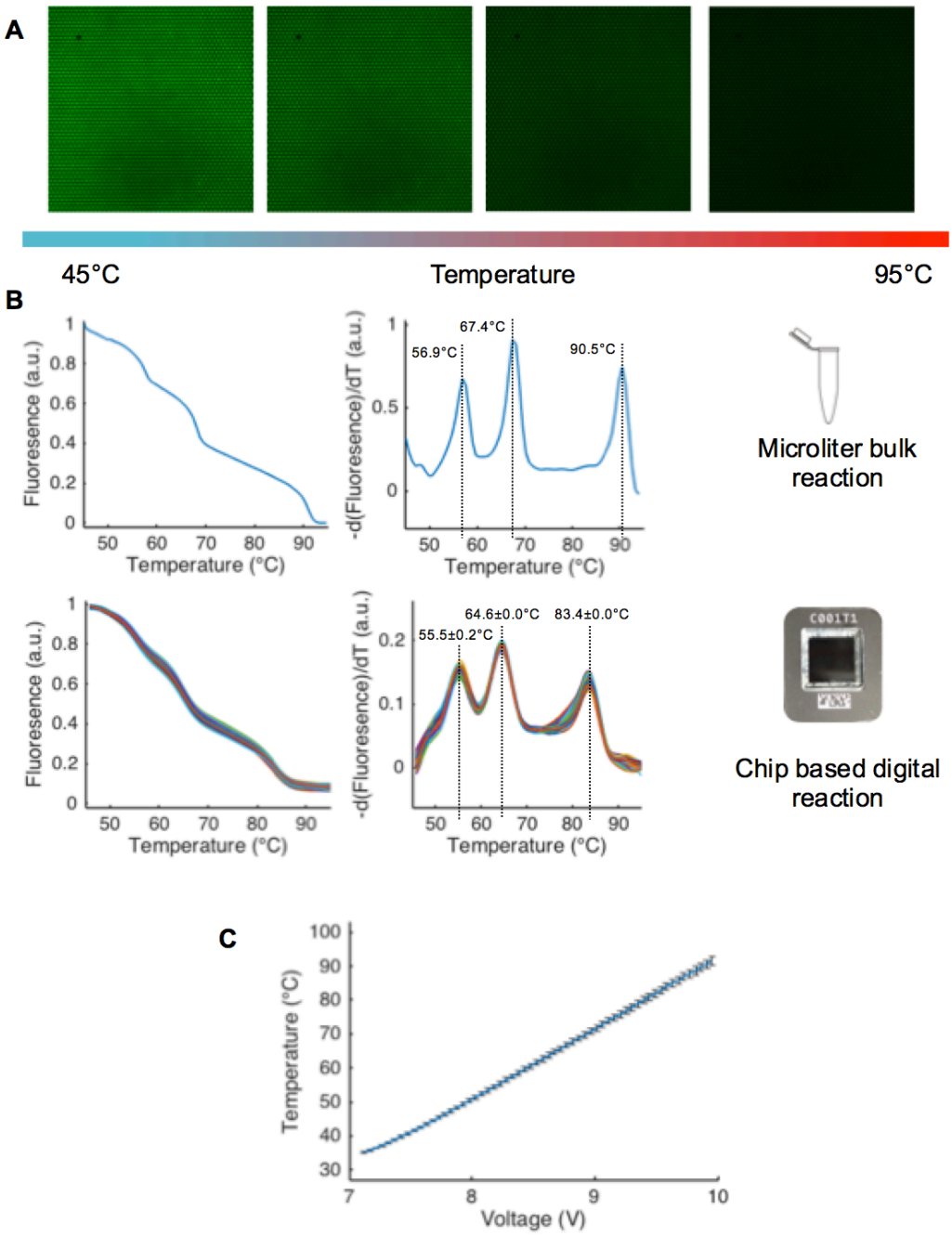


Figure 3: On-Chip dHRM Characterization and Optimization. **(A)** Image of a portion of the chip. Upon controlled heating, fluorescence is lost as the DNA denatures. **(B)** Melting of three synthetic temperature calibrator sequences containing different GC content at optimized ramp rate on-chip compared to bulk qPCR HRM. **(C)** A plot of the relationship between voltage and temperature for 5 runs, showing it to remain linear throughout the range of our temperature profile. The standard deviation has a maximum of 1.22 °C at 91.6 °C.

temperature was monitored during the ramp by placing a thermocouple inside a dummy dPCR chip and placing this chip beside the calibrator loaded chip. A ramp rate of 0.02 degrees/sec optimized the linearity and repeatability of our voltage and temperature relationship, with maximum standard deviation of 1.22°C occurring at a temperature of ~91.6°C over 5 runs (Fig. 3C).

Next, bacterial gDNA from clinical isolates of *Listeria monocytogenes* and *Streptococcus pneumoniae*, two common pathogens causing neonatal bacteremia(45), were used to further optimize signal to noise and melt curve shape (i.e. temperature) resolution. In the standard qPCR format, melt curve shape, a key discriminating feature of bacterial 16S melt curves(38), was found to be highly dependent on imaging rate. A low imaging rate averaged out key melt curve shape features (Fig. 4A, circle), but a faster imaging rate captures the small shape difference unique to this organism (Fig. 4B, circle). Using the heat ramp rate previously optimized for linearity and repeatability, we varied the imaging rate on-chip (Fig. 4, B and D). The low calibrator sequence (first peak from left in Fig. 4 melt curves) was included in all amplification reactions to align curves and overcome temperature variation across qPCR wells and digital chip wells. The on-chip imaging rate was adjusted such that it approximated the optimum imaging rate of the qPCR machine, which produces curves that are identifiable by our machine learning algorithm(38). In a standard qPCR machine, default settings typically generate HRM imaging data points every ~0.3 degrees. On our dHRM platform, imaging the chip every 15 sec at the optimal heat ramping rate of 0.02 degrees/sec allowed us to achieve this imaging rate. Melt curves generated from

these settings constitute our low imaging rate data in Figure 4A and 4B. With these settings, the average peak-to-baseline ratio of the 16S amplicon derivative melt curves (after min-max normalization of raw melt data) was 0.1096 ± 0.0024 on the qPCR machine versus 0.0660 ± 0.0034 for the dPCR. We then optimized the qPCR ramp rate to achieve the slowest ramp that maximized the signal to noise ratio: 0.05 degrees/sec with imaging data taken every ~0.1 degrees (Fig. 4C). Increasing the imaging rate on our dHRM system to match this gave our high imaging rate data, shown in Figure 4D.

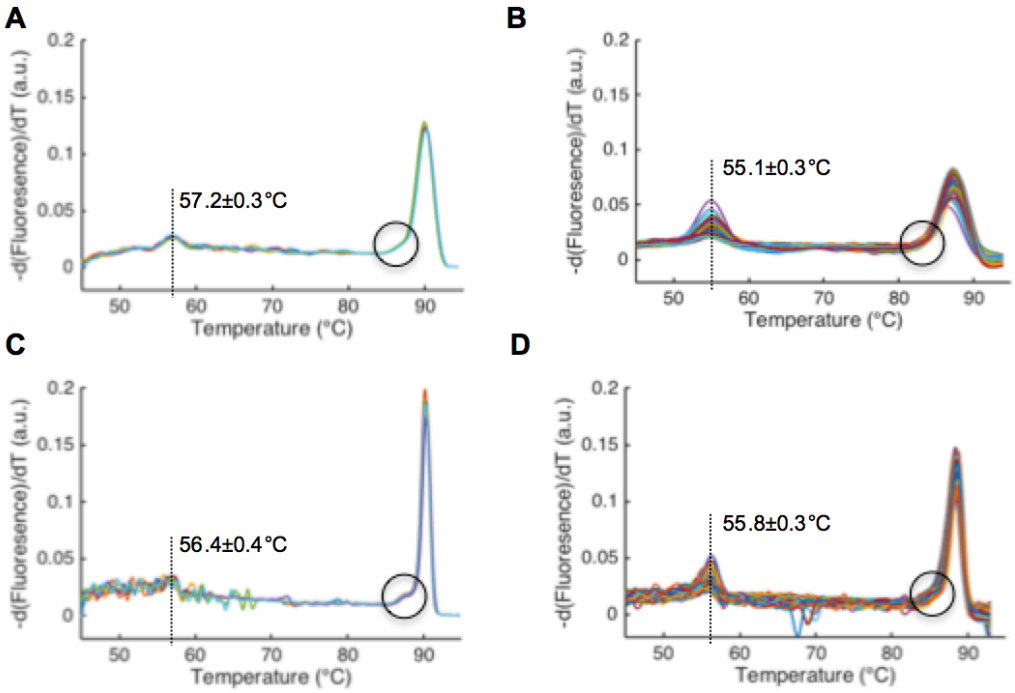


Figure 4: Sampling and Ramp Rate Optimization on the dHRM Chip. (A),(B) *L. monocytogenes* melt curves generated with a low sampling rate on qPCR and dPCR platforms respectively. **(C),(D)** *L. monocytogenes* melt generated using a high sampling rate on qPCR and dPCR platforms respectively. The synthetic temperature calibrator sequence melting temperature mean and standard deviation are shown in all. Black circle highlights key melt curve shape feature unique to *L. monocytogenes* 16S sequence that is dependent on sampling rate.

At the high imaging rate, the average peak-to-baseline ratio of the 16S amplicon derivative melt curves was 0.1759 ± 0.0073 on the qPCR machine versus 0.1225 ± 0.0066 for the dPCR, demonstrating that with these optimized settings, our device achieves comparable signal to noise performance. Importantly, these settings also enable a key feature of this particular bacterial melt curve to be observed (circled region of melt curves, Fig. 4). This feature is critical for the discrimination of *L. monocytogenes* from other bacterial 16S melt curves using our SVM algorithm. It is also important to note that Tm reproducibility was almost identical between the two optimized platforms, as demonstrated by the Tm standard deviation (~0.3^{o}C, Fig. 4) of the temperature calibrator sequence. Because this deviation still exists under optimized conditions, temperature calibrator sequences are included in all reactions for aligning melt curves prior to analysis.

Identification and Absolute Quantification of Single Cells in Polymicrobial Samples

We next integrated our automated OVO SVM melt curve identification approach with our dHRM platform to assess the sensitivity, specificity, and quantitative power of our technology within the range of clinical concentrations expected for neonatal bacteremia samples. A training database of bacterial melt curves was generated on-chip to enable automatic curve identification. Bacterial gDNA from *L. monocytogenes* and *S. pneumoniae* were loaded onto separate chips in excess, l of 223 and 141 respectively as calculated from spectrometer readings, such that each of the 20,000 reactions would be positive for amplification and could generate a unique

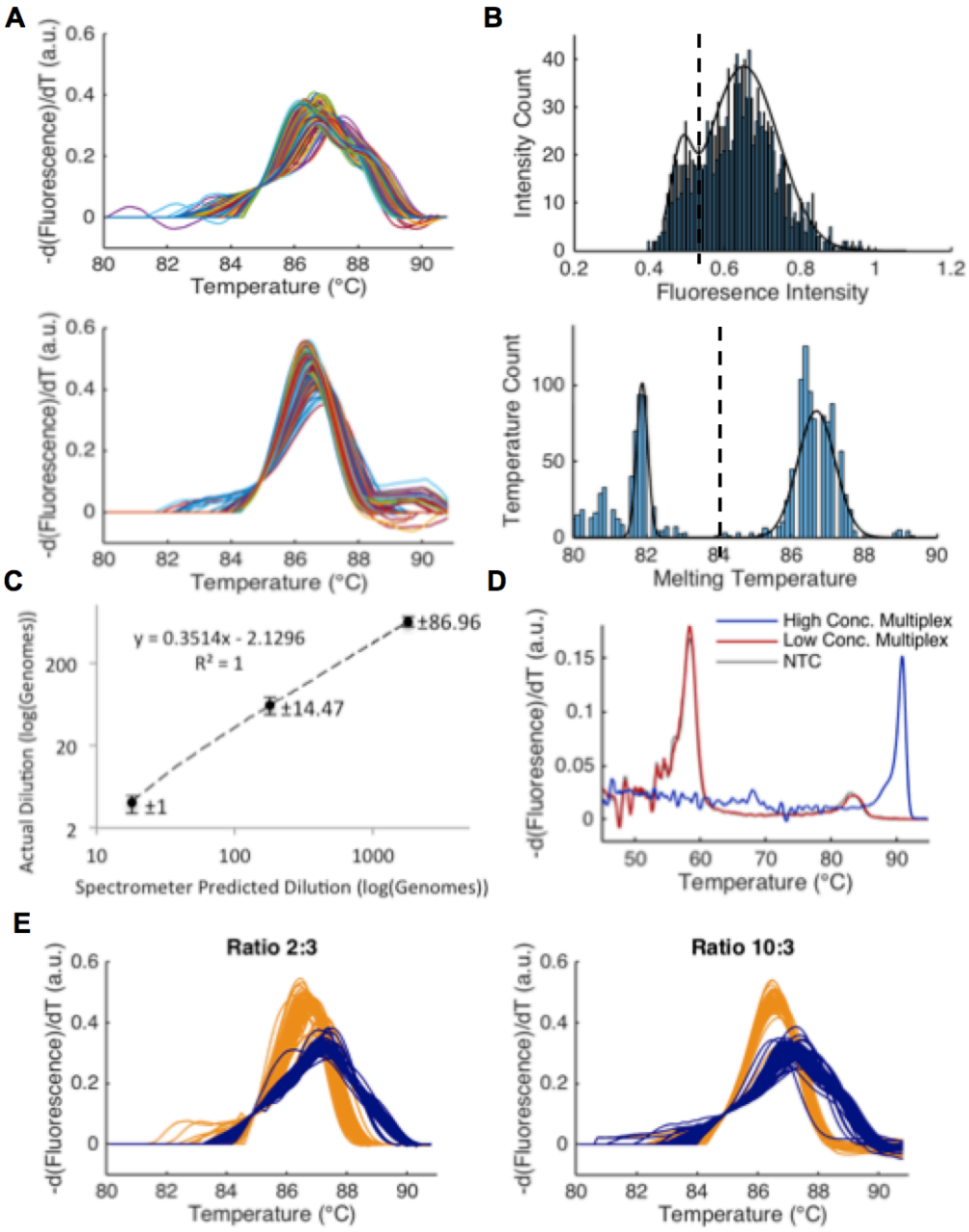
training melt curve for the bacterial isolate. Hypervariable regions 1-6 of the 16S gene, ~1000bp, were amplified on chip and underwent dHRM using the optimized ramp and imaging rates described above. Figure 5A shows the dHRM training curves generated on-chip for *S. pneumoniae* and *L. monocytogenes* after processing with our image analysis, normalization, and alignment algorithms (see Methods). The processed curves were then entered into our OVO SVM algorithm as training data (see Methods). Leave One Out Cross Validation (LOOCV) using 1,500 training curves for each organism resulted in a classification accuracy of 99.9% within the training dataset.

Next, we confirmed that our cycling protocol reached endpoint amplification for low values of l, to ensure that single molecules were fully amplified for absolute quantification. A ten-fold dilution series of monomicrobial gDNA samples of *L. monocytogenes* were applied on-chip and underwent dPCR for 70 cycles followed by dHRM and enumeration of the total number of positive and negative reactions at each dilution. Negative reactions were removed based on a total fluorescence threshold, according to typical dPCR protocols (see Methods). However, observation of the melt curve data above and below this threshold revealed a Type I (false positive) error rate of 22.6% and Type II (false negative) error rate of 1.19% (average across 3 chips), resulting in a lower limit of detection of ~238 with the standard approach. A no template control (NTC) chip was also run to confirm the melting patterns of off-target amplification (not shown). One reason for the total fluorescence threshold method performing poorly in our case could be that we thermocycle significantly longer than

most dPCR protocols recommend. Typical dPCR cycle number is kept to ~35, but we find that 70 cycles ensures full endpoint amplification from single molecules(*15*). While this extended cycling improves accuracy of single-molecule target detection, it also allows off-target amplification to fluoresce more prominently in negative reactions. We noted that those reactions which were identified falsely as positive based on a fluorescence intensity threshold displayed melt curves arising from off-target amplicons that melted much sooner during dHRM than our targeted 1,000bp amplicon. As such, we developed a new negative thresholding approach by identifying a Tm threshold (Fig. 5B) (see Methods), which is uniquely enabled by our platform. With this approach, the Type I and II error rates improved dramatically to 0.07% and 0.00% respectively (average across 3 chips), resulting in true single copy sensitivity. Tm thresholding resulted in a linear relationship across the monomicrobial gDNA dilution series having an r^2 value of 1 and high measurement precision demonstrated by the low sample standard deviations at each dilution (Fig. 5C).

We then compared the number of curves quantified by our Tm threshold with the sample gDNA concentrations calculated from spectrometer readings and qPCR cycle thresholds. Standard dilution qPCR curves generated for *S. pneumoniae* and *L. monocytogenes* gDNA samples in Figure 6, show that linearity and thus quantitative power are lost in the digital regime (below data point circled in red), where ~0-55 genomes are estimated to be present based on initial spectrometer readings. Although the slope of the standard curve is low, indicating low amplification efficiency, efficiency is inconsequential dHRM because endpoint cycling is used and

Figure 5: SVM Classification of L. monocytogenes and S. pneumoniae. (A) 2000 normalized *S. pneumoniae* **(top)** and *L. monocytogenes* **(bottom)** dPCR melt curves aligned to 0.1 –dF/dT, respectively. These curves are used to train the SVM to classify each bacteria. **(B)(top)** A histogram showing the fluorescence intensity values with the PDF overlayed and the intensity value chosen to classify positive from negative marked. **(B)(bottom)** A histogram showing the Tm of the peaks with the PDF overlayed and the Tm value chosen to classify positive from negative marked. **(C)** A dPCR dilution series of *L. monocytogenes* with the actual vs. predicted counts. The strong linearity of the relationship demonstrates the strong amplification of our amplicon down the single molecule level. The sample mean and sample standard deviation are reported. **(D)** In blue: qPCR melt curve generated from a 1:1 mix of 20 ng/μL of *S. pneumoniae* and *L. monocytogenes*. In red: qPCR melt curve generated from a 1:1 mix of 0.02 ng/μL of *S. pneumoniae* and *L. monocytogenes*. This concentration is the same as that used on a digital chip. In grey: qPCR melt curve generated from a negative template control (NTC) with no bacterial gDNA added. **(E)** SVM classification of *L. monocytogenes* and *S. pneumoniae* mixed at two concentrations on a chip, demonstrating the capability of the SVM to correctly identify multiplexed samples.


A
-d(Fluorescence)/dT (a.u.)
Temperature (°C)
B
Intensity Count
Fluoresence Intensity
Temperature Count
Melting Temperature
C
y = 0.3514x - 2.1296
R² = 1
±86.96
±14.47
±1
Actual Dilution (log(Genomes))
Spectrometer Predicted Dilution (log(Genomes))
D
High Conc. Multiplex
Low Conc. Multiplex
NTC
E
Ratio 2:3
Ratio 10:3

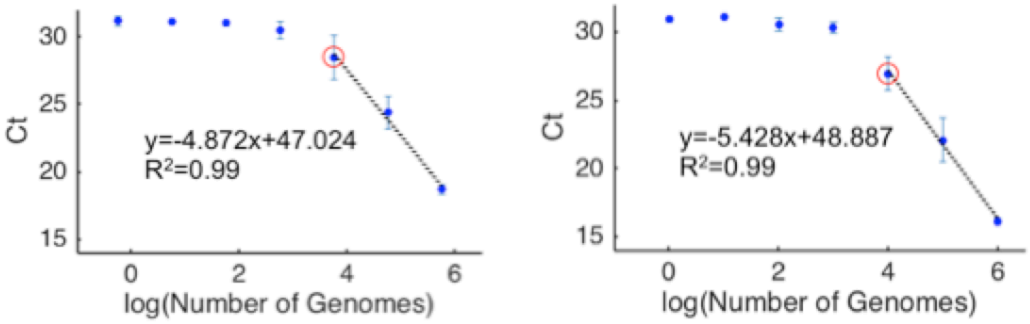


Figure 6. qPCR dilution series for *S. pneumoniae* and *L. monocytogenes* respectively. The primer efficiency slope is shown in black. A red circle marks the approximate Ct and concentration used on the dPCR chip.

quantification relies on counting melt curves. For highly multiplexed sequence profiling with U-dHRM, long sequence amplicons are more information rich(*27*) and as such are preferred, but their amplification efficiency is expected to be low. Regardless, Table 1 shows that our U-dHRM platform detects total DNA concentrations at similar levels as the other two technologies, but is capable of distinguishing target DNA from background/contaminant DNA. In the case of the *S. pneumonia*e sample, our results show that qPCR and spectrometry overestimated the target concentration by 5.5 and 4.8 fold respectively compared to the U-dHRM and OVO SVM confirmed amplicon count. Likewise for *L. monocytogenes*, overestimates were 4.8 and 4.1 fold respectively.

Next, mock samples of polymicrobial infection, which cannot be evaluated using standard HRM, were created by mixing *L. monocytogenes* and *S. pneumoniae* gDNA at two different ratios, 3:2 and 3:10. These were applied to the chip at

Table 1. Genome concentrations determined by 3 different quantification methods. The concentration of genomic DNA isolated from both *S. pneumoniae* and *L. monocytogenes* was measured using first an Eppendorf Biospectrometer, then estimated based on a qPCR standard curve, and finally directly quantified using dPCR.

Bacteria	Method of Quantification	Number of Genomes/µL
S. pneumoniae	Eppendorf Biospectrometer	5780
	qPCR Ct Calculated	6554
	dPCR Specific Curve Count	1200
	dPCR Nonspecific Curve Count	4260
L. monocytogenes	Eppendorf Biospectrometer	9160
	qPCR Ct Calculated	10839
	dPCR Specific Curve Count	2260
	dPCR Nonspecific Curve Count	5320

concentrations nearing the low and high end of the typical clinical pathogen load for neonatal bacteremia (50-2,000 copies), a dynamic range that cannot be assessed by any current HRM format (Fig. 2A). This is further demonstrated by Figure 5D, which shows the same 3:2 mixtures analyzed by qPCR HRM. Here, the resulting ensemble average melt curves fail to indicate the presence of two distinct species (blue curve) or, in cases of very low gDNA input, the presence of any bacteria at all (red curve) due to overwhelming background amplification that results in a melt curve matching the NTC melt curve. Figure 5E shows the melt curves resulting from U-dHRM of the two heterogeneous samples followed by automated Tm thresholding and OVO SVM

analysis. Yellow melt curves represent those identified by the OVO SVM as *L. monocytogenes* and blue as *S. pneumoniae*. Table 2 enumerates the bacterial composition of the sample reported by the OVO SVM output of total number of curves classified into each bacterial identity category.

Table 2. SVM classification of mixed genomic DNA samples from *S. pneumoniae* and *L. monocytogenes*. Two dPCR chips were loaded with mixed genomic DNA containing two different ratios of *S. pneumoniae* and *L. monocytogenes*. The predicted ratios were determined using an Eppendorf Biospectrometer, while the actual counts were determined using an SVM classifier on dHRM curves generated from a dPCR chip.

Ratio	Bacteria	Predicted Count	Actual Count
2:3	*S. pneumoniae*	289	60
	L. monocytogenes	458	113
10:3	*S. pneumoniae*	1445	238
	L. monocytogenes	458	119

In order to demonstrate the correct amplification of our amplicon as well as the misamplification generated by NTCs, we ran a gel comparing qPCR amplification of both the correct template and the NTCs (Fig 7). The correct amplicon is shown at approximately 1000 bp, while the NTCs generate bands at around 100-500 bp, as the lower melting temperature shown on the melt curves suggests. When template DNA is present the there is no misamplification present, but without a template to work off of the polymerase will sometimes generate shorter length misamplifed DNA. The presence of misamplification at high cycle numbers demonstrates the necessity for a

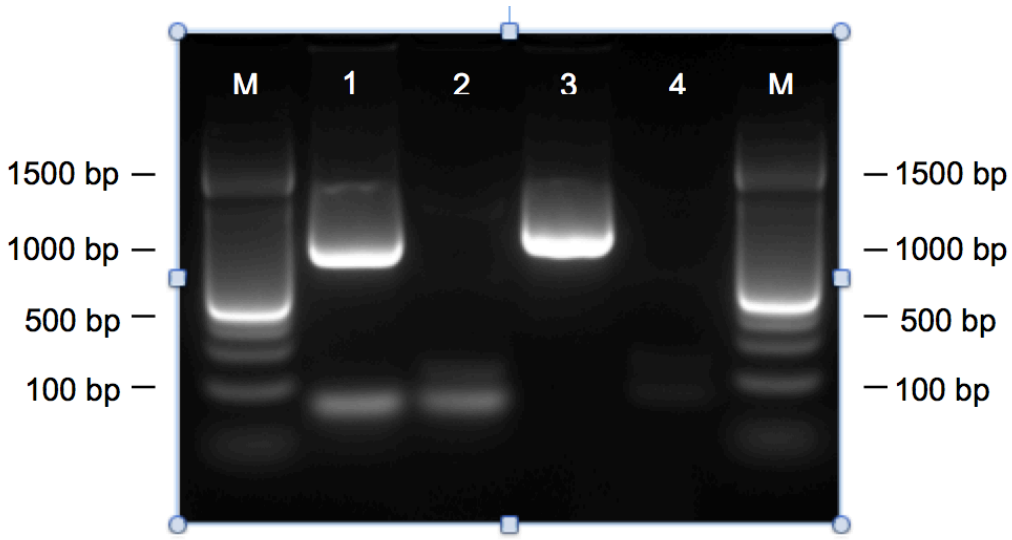


Figure 7. Gel Image of 16s Bacteria Amplicon. M: DNA ladder markers. 1: Correctly amplified DNA is shown as a band at 1000 bp. Calibrator sequences are shown as a band at 60 bp. 2: The NTC is loaded to show slight misamplification at approximately 100 bp, just above the calibrator sequences. 3: The template is again shown as a band at 1000 bp, but calibrator sequences are not loaded. 4: The NTC is loaded to show very faint misamplification bands at approximately 100 bp and 150 bp.

dPCR platform to distinguish between misamplification and the correct template, as the melt curve analysis of our platform uniquely allows with high accuracy. The gel in Fig. 7 is shown with and without calibration sequences to emphasize the bands due to misamplification as opposed to those due to the calibration sequences (seen at 60 bp).

Acknowledgements

Chapter 1-4 in part, have been submitted for publication of the material as it may appear in Scientific Reports, 2016, Sinead Hawker, Daniel Ortiz Velez, Behnam Hedayatnia, Ninad Kulkarni, Yang Zhang, Shelley Lawrence, Stephanie I. Fraley. The dissertation/thesis author was the primary investigator and author of this paper.

CHAPTER 4 - DISCUSSION

The absolute quantification and identification of numerous genotypes in a heterogeneous sample at realistic clinical sample concentrations is made possible with our chip-based U-dHRM, Tm thresholding, and OVO SVM system. This technology advances broad-based HRM profiling by accomplishing high resolution melt curve generation in reaction volumes that are 0.005% of the traditional HRM volume, and by massive parallelization, where curve generation occurs simultaneously in 20,000 reactions. Nucleic acid samples are partitioned across the 20,000 picoliter-scale reactions such that each reaction holds only one or zero microbial genomes. Reduction in the size of reactions allows less reagents to be used while maintaining optimal reagent concentrations. This overcomes the problem of reagent and environmental microbial DNA contamination by spatially diluting, i.e. contaminating DNA and target cells/DNA are partitioned from each other to enable discrimination (45). Subsequent amplification and dHRM yield singular products and distinct melt profiles for each digital reaction. Incorporating reference temperature calibrator sequences into each reaction normalizes against reaction condition variations, allowing highly robust sequence-specific melt curves to be obtained. This enables accurate identification to the species level by matching to database melt curves using our custom SVM machine learning algorithm (46,47). Building on the concept of digital PCR, which allows single molecule sensitivity and eliminates template amplification competition bias, dHRM confers single nucleotide resolution, enabling highly specific identification and quantification of all species in a mixed population (45).

Our novel dHRM platform achieves a clinically relevant dynamic range of detection while simultaneously reducing reagent costs. It also permits the generation of an exponentially larger SVM training curve database for each genotype with only a single chip run. This high-throughput format in combination with the single nucleotide specificity of our universal HRM technique(26, 38) will allow for rapid expansion of HRM databases of identifiable genotypes. The integration of universal dPCR for single molecule amplification, U-dHRM sequence fingerprinting with generic intercalating dye, and melt curve processing and identification with machine learning completes a inexpensive system for rapidly, automatically, and accurately identifying a large number of genotypes in under 3 hours with minimal hands-on time and technical skill. The low cost of the chips and reagents and the use of generic intercalating dye as a universal reporter yield a total materials cost per sample of approximately $10.

The combined capabilities of universal dHRM could offer significant advantages over qPCR, microarray, and even NGS profiling approaches, especially for infectious disease detection where breadth of detection, timing, and cost are critical factors. The ability to quantify even low level organisms in polymicrobial samples within hours could significantly impact clinical microbiology diagnostic practice, where patients suspected of bloodstream infection suffer from an hourly increase in mortality risk due to lack of diagnostic information and inaccurate treatments. Moreover, retrospective studies suggest that absolute quantification of bacterial genomic load in patients may be useful to assess severity of infection and to predict

prognosis in sepsis cases(46). Detection of microbial DNA in clinical samples like blood poses unique challenges(46-48). In this case, human DNA often vastly outnumbers that of the infecting microbe, necessitating targeted microbial DNA amplification. Since a vast array of pathogens can cause infection, a broad-based approach for amplifying all microbial DNA and a high level of specificity in post-PCR sequence characterization techniques are required. Highly related organisms may have few DNA sequence differences, requiring single nucleotide specificity(44). Moreover, the number of microbial genomes present in the sample may be extremely low and/or the sample may be comprised of several different microbes, a polymicrobial infection, necessitating single cell level analysis. These requirements challenge the abilities of current molecular detection technologies(15, 41), including NGS. Recent studies report that several NGS sequencing platforms for microbial detection approach the analytical sensitivity of standard qPCR assays(15).

Our U-dHRM platform meets or exceeds each of these critical factors at bacterial concentrations relevant to clinical scenarios, in particular, neonatal sepsis. Standard blood culture methods require 3-10 mL from two or more sites in older children and adults and 1 mL from a single site in neonates(42). Sample limitations secondary to blood volume restrictions, particularly in very low birth weight premature infants, can cause false negative test results when bacteremia is low or if the infant was exposed to intrapartum empiric antibiotics. DHRM resolves these common problems through its single molecule sensitivity and ability to identify both replicating and non-replicating bacteria in minimal blood sample volume. Further work to

automate DNA extraction and couple it to automated chip loading would result in a nearly hands-free process that is down-scalable for portability. Improved chip design could reduce the chip's thermal mass to generate results at an even faster pace.

The ability to accomplish dHRM on chip also enables our platform to overcome detection limitations of traditional dPCR that arise from false positive and negative reactions, where off-target background amplification produces fluorescent intensity near that of a true positive reaction. The typical approach of applying an intensity threshold to remove false positives but retain all true positives left a significant number of reactions as false positive and false negative based on melt curve observation. Our approach using Tm thresholding to identify positive versus negative reactions led to a significant improvement in the Type I and II error rate of dPCR such that true single molecule sensitivity was attained for optimal lower limit of detection.

Integration of our OVO SVM approach for melt curve signature identification and absolute quantification enables broad-based, automated identification of bacterial organisms in heterogeneous samples. However, some foreseeable limitations of our technology exist. Improvements to the temperature ramp reliability in the ~70-95^{o}C temperature range will be critical to ensuring that a larger database of melt curves can be used reliably for dHRM OVO SVM. Here, temperature data was ignored by aligning the melt curves to at their derivative fluorescence value of 0.1 for shape comparison. However, temperature shift can be an important feature for melt curve discrimination. Insulation from environmental temperatures as well as improved chip design with lower thermal mass will likely to overcome this issue. Likewise, there is a

need for even more reactions for versatile application of the technology. For example, adult bloodstream infection levels can reach 100,000 cfu/ml, requiring millions of digital reactions to partition the sample for single organism level identification. Moving away from a fixed well chip design towards droplet and flow-based processing would also significantly advance dHRM technology by enabling an almost unlimited number of reactions to be used and analyzed in a high-throughput fashion. Finally, computational approaches for anomaly detection are being explored by our group to identify true bacterial melt curves that are not yet represented in our database. Currently, a 16S amplicon that melts above the Tm threshold will be automatically classified by our OVO SVM as an organism whose melt curve that it most closely resembles. For undefined samples, where significantly more organisms may arise and even unexpected emerging pathogens could be present, the ability to identify whether a melt curve is a poor match to the database curves will be crucial. Indeed, other groups have identified novel bacteria by observing alterations in bulk HRM curves by eye(*30*). Automation of this ability would represent a significant advancement for dHRM technology, as novel bacteria could be rapidly and inexpensively discovered by our universal 16S dHRM and further interrogated by whole genome sequencing approaches.

In summary, our dHRM platform offers a robust means of differentiating genotypes in heterogeneous samples and quantifying their absolute abundance based on their finger-print melt curves. Many other applications would benefit from such a platform, which can even be used to discover novel mutations or novel species.

Acknowledgements

Chapter 1-4 in part, have been submitted for publication of the material as it may appear in Scientific Reports, 2016, Sinead Hawker, Daniel Ortiz Velez, Behnam Hedayatnia, Ninad Kulkarni, Yang Zhang, Shelley Lawrence, Stephanie I. Fraley. The dissertation/thesis author was the primary investigator and author of this paper.

CONCLUSION

In order to develop our platform for digital HRM, we started by optimizing a commercially available dPCR chip with 20k PCR reactions. While the chip itself contains highly reproducible wells capalble of taking in PCR mastermix with high efficiency, we found the manufacturer supplied mastermix to be ineffective at amplifying full-length amplicons with high reliability. To combat this, we developed our own mastermix using EvaGreen as the intercalating dye and ROX as the passive reference dye, using a PCR buffer containing surfactant to maintain a low surface tension for easy loading. We further modififed the manufactuer's protocol by changing the cycling from 40 cycles of a two-step PCR to 70 cycles of a 3-step PCR. In this way, we had a highly reproducible way to generate full-length amplification of our region of interest using an easily obtainable dPCR platform.

Having successfully developed the first part of our platform, we moved on to building a stable temperature ramp in order to generate reproducible melt curves. We controlled a Peltier chip sandwiched between two copper plates, a heat sink, and the dPCR chips to uniformly heat our dPCR reactions. We used synthetic sequences of a known melting temperature to test the prototype of our dHRM platform, allowing us to determine how we could control the temperature ramp rate based on the power we fed into the Peltier chip. This let us find what ranges our platform could stably produce a linear relationship between voltage and temperature, meaning what ranges we could generate a smooth, linear temperature ramp. Furthermore, using the synthetic sequences we determined the best microscope objective to use (namely a 4x objective)

and what settings to program into the microscope's automated software. Finally, we developed the beginnings of our image analysis code using these sequences in order to pull the correct melt curves from the microscope images based on both the *in silico* prediction of the synthetic sequence melt curves and the qPCR melt curves.

With our prototype successfully melting synthetic sequences, we moved onto amplifying the bacterial 16s rRNA gene and optimizing the melt curves to match the best qPCR melt curves we could generate. We obtained pure genomic DNA and amplified it with our optimized mastermix and cycling protocol within the chip. We then altered the imaging rate of our microscope while holding our temperature ramp rate constant in order to study how sampling rate affected the melt curves we obtained. We found that by increasing the sampling rate on our qPCR macine we were able to get cleaner, more differentiable melt curves. This translated to our dHRM platform as we increased sampling rate by imaging more frequently. We found that 1 data point per 0.15 degrees maximized our signal to noise ratio and gave highly differentiable melt curves.

Having optimized our hardware and sampling protocol, we further developed our image analysis code. We took the melt curves generated from each well and applied a Gaussian smoothing filter to them, before normalizing to area and performing exponential normalization. Due to the high number of cycles needed to get end-point amplification of a single molecule, some misamplification occurs in the NTCs. We were able to separate misamplifiaction from correct amplification using the melting temperature of the curves, as the NTC's melt at a much lower temperature

than full length amplication due to their shorter length. We then used dPCR chips overloaded with pure bacterial DNA from *L. monocytogenes* and *S. pneumoniae* to generate the melt curves needed to train our SVM code for a proof of concept classification. We then applied our SVM code to a chip of mixed *L. monocytogenes* and *S. pneumoniae* DNA diluted to the single molecule level and found it to be highly accurate in differentiating between the two bacterial species and providing correct counts of each.

This proof of concept demonstrates the potential our dHRM platform has to detect and classify polymicrobial septic infections. Our platform is demonstrated to detect a single molecule of DNA within a microliter of sample, making it sensitive enough to detect the low quantity of pathogenic bacterial DNA found in a septic patient. Furthermore, we demonstrated that our platform is capable of identifying and differentiating between multiple bacterial species, such as is necessary in polymicrobial infection. Theoretically this platform could detect up to 2000 different bacterial species with one chip at the single molecule level. Given that there could be bacteria present within the blood unrelated to the septic response, it is necessary for the platform to be capable of pulling out the bacteria of interest. As identification of clinically relevant bacteria based on their 16s rRNA gene has already been shown in bulk qPCR reactions and the resolution of our dHRM platform has been shown to be on par with that of qPCR machines, we believe our platform capable of differentiating between many more than the two bacteria that we demonstrated.

Future work will focus on expanding our dPCR database. By working with clinically isolated strains we can develop a database specific to clinical strains of bacteria. This is of particular importance if the 16s sequences of the bacteria differ between clinical strains and thus generate varying melt curves. After developing a database based on pure clinical samples, we will then test our database on blood samples obtained from septic patients. While at first our database will be used as a confirmation of standard diagnostic techniques, when the accuracy is confirmed it can be tried as a diagnostic tool.

The platform itself can be improved through increasing the resolution of the melt curves, through even more reproducible temperature ramping and by imaging the sample more often. This can be done through improving the heat sink to ensure that its cooling is more reproducible and less dependent on environmental temperature. Insulating the platform would further protect the platform from day-to-day temperature variations. The most beneficial improvement to the platform would be to export it to a portable system, such that it could be easily transported to hospitals or other clinical settings without requiring a large microscope and heating apparatus. Improved portability would greatly improve the dHRM platform's ease of use and make its applicability to a clinical setting a reality.

While this thesis demonstrates that our dHRM platform has great potential as a clinical diagnostic tool that overcomes the problems associated with both bulk HRM and standard dPCR, it is still early in the work needed to make this accessible in a more realistic clinical setting. There is clearly great potential in the ability to pull

identifiable melt curves from single molecules of DNA, as well as profile up to 2000 sequences as once. Once a much larger database has been developed and applied to actual clinical samples and the platform is made to be portable, the our dHRM analysis could be the next standard in septic diagnostics.

REFERENCES

1. D. Angus, T. van der Poll, Severe sepsis and septic shock. *The New England Journal of Medicine* **369**, 840-851 (2013).

2. A. Kumar, D. Roberts, K.E. Wood, B. Light, J.E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, A. Kumar, M. Cheang, Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* **34**, 1589-1596 (2006).

3. A. Kotsaki, E.J. Giamerellos-Bourboulis, Emerging therapies for the treatment of sepsis. *Expert Opin Emerg Drugs* **17**, 379-391 (2012).

4. M.M.P. Faria, J.M. Conly, M.G. Surette, The development and application of a molecular community profiling strategy to identify polymicrobial bacterial DNA in the whole blood of septic patients. *BMC Microbiology* **15** (2015).

5. P. Yagupsky, F.S. Nolte, Quantitative aspects of septicemia. *Clin Microbiol Rev* **3**, 269-279 (1990).

6. T. Reier-Nilsen, T. Farstad, B. Nakstad, V. Laurvrak, M. Steinbakk, Comparison of broad rang 16s rDNA PCR and conventional blood culture for diagnosis of sepsis in the newborn: a case control study. *BMC Pediatr* **9**, 5 (2009).

7. R.P. Peters, M.A. van Agtmael, S.A. Danner, P.H. Savelkoul, C.M. Vandenbroucke-Grauls, New developments in the diagnosis of bloodstream infections. *Lancet Infect Dis* **4**, 751-760 (2004).

8. S. I. Fraley, P. Athamanolap, B. J. Masek, J. Hardick, K.C. Carroll, Y.H. Hsieh, R. E. Rothman, C.A. Gaydos, T.H. Wang, S. Yang, Nested machine learning facilitates increased sequence content for large-scale automated high resolution melt genotyping. *Nature Scientific Reports* **6** (2016)

9. R.M. Hassan, M. G. El Enany, H. H. Rizk, Evaluation of broad-range 16S rRNA PCR for the diagnosis of bloodstream infections: two years of experience. *J Infect Dev Ctries* **8**, 1252-1258 (2014).

10. S. I. Fraley, J. Hardick, B.J. Masek, P. Athamanolap, R.E. Rothman, C.A. Gaydos, K.C. Carroll, T. Wakefield, T.H. Wang, S. Yang,Universal digital high-resolution melt: a novel approach to broad-based profiling of heterogeneous biological samples. *Nucleic Acids Research* **41**, e175 (2013).

11. M. Li, L. Zhou, R.A. Palais, C.T. Wittwer, Genotyping Accuracy of High Resolution DNA Melting Instruments. *Clinical Chemistry* **60**, 6 (2014).

12. M. Erali, R. Palais, C. Wittwer, SNP genotyping by unlabeled probe melting analysis. *Methods Mol Biol* **429**, 199-206 (2008).

13. G. H. Reed, C. T. Wittwer, Sensitivity and Specificity of Single-Nucleotide Polymorphism Scanning by High-Resolution Melting Analysis. *Clinical Chemistry* **50**, 1748-1754 (2004).

14. Z. Dwight, R. Palais, C. T. Wittwer, uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. (2011).

15. S. Chakravorty, J. Seok Lee, E. Jin Cho, S. Roh, L. E. Smith, J. Lee, C.T. Kim, L.E. Via, S.N. Cho, C.E. Barry III, D. Alland, Genotypic susceptibility testing of Mycobacterium tuberculosis isolates for amikacin and kanamycin resistance by use of a rapid sloppy molecular beacon-based assay identifies more cases of low-level drug resistance than phenotypic Lowenstein-Jensen testing. *J Clin Microbiol* **53**, 43-51 (2015).

16. H. H. El-Hajj, S.A.E. Marras, S. Tyagi, E. Shashkina, M. Kamboj, T.E. Kiehn, M.S. Glickman, F.R. Kramer, D. Alland, Use of sloppy molecular beacon probes for identification of mycobacterial species. *J Clin Microbiol* **47**, 1190-1198 (2009).

17. J. T. den Dunnen, R. H. A. M. Vossen, E. Aten, A. Roos, High-Resolution Melting Analysis (HRMA)-More Than Just Sequence Variant Screening. *Hum Mutat* **30**, 860-866 (2009).

18. S. I. Fraley, J. Hardick, B.J. Masek, P. Athamanolap, R.E. Rothman, C.A. Gaydos, K.C. Carroll, T. Wakefield, T.H. Wang, S. Yang, Universal digital high-resolution melt: a novel approach to broad-based profiling of heterogeneous biological samples. *Nucleic Acids Research* **41**, e175 (2013).

19. N. A. Mohamed Suhaimi, Y.M. Foong, D.Y. Lee, W.M. Phyo, I. Cima, E.X. Lee, W.L. Goh, W.Y. Lim, K.S. Chia, S.L. Kong, M. Gong, B. Lim, A.M. Hillmer, P.K. Koh, J.Y. Ying, M.H. Tan, Non-invasive sensitive detection of KRAS and BRAF mutation in circulating tumor cells of colorectal cancer patients. *Mol Oncol* **9**, 850-860 (2015).

20. P. Athamanolap, D. J. Shin, T. H. Wang, Droplet Array Platform for High-Resolution Melt Analysis of DNA Methylation Density. *J Lab Autom* **19**, 304-312 (2013).

21. J. S. Castresana, W. Mueller, E. Urdangarin, P. Lazcoz, A. von Deimling, J.S. Castresana, Detection of methylation in promoter sequences by melting curve analysis-based semiquantitative real time PCR. *Bmc Cancer* **8**, (2008).

22. V. Gürtler, D. Grandob, B. C. Mayalla, J. Wanga, S. Ghaly-Deriasa, A novel method for simultaneous Enterococcus species identification/typing and van genotyping by high resolution melt analysis. *Journal of Microbiological Methods* **90**, 167–181 (2012).

23. M. H. Hjelmsø, M.J. Hjelsmo, L.J. Hansen, J. Baelum, L. Feld, W.E. Holben, C. S. Jacobsen, High Resolution Melt analysis for rapid comparison of bacterial community composition. *Applied and Environmental Microbiology*, (2014).

24. J. Hardick, H. Won, K. Jeng, Y.H. Hsieh, C.A. Gaydos, R.E. Rothman, S. Yang, Identification of Bacterial Pathogens in Ascitic Fluids from Patients with Suspected Spontaneous Bacterial Peritonitis by Use of Broad-Range PCR (16S PCR) Coupled with High-Resolution Melt Analysis. *Journal of Clinical Microbiology* **50**, 2428-2432 (2012).

25. K. Jeng, S. Yang, H. Won, C.A. Gaydos, Y.H. Hsieh, A. Kecojevic, K.C. Carroll, J. Hardick, R.E. Rothman, Application of a 16S rRNA PCR–High-Resolution Melt Analysis Assay for Rapid Detection of Salmonella Bacteremia. *Journal of Clinical Microbiology* **50**, 1122-1124 (2012).

26. B. J. Masek, J. Hardick, H. Won, S. Yang, Y.H. Hsieh, R.E. Rothman, C.A. Gaydos, Sensitive detection and serovar differentiation of typhoidal and nontyphoidal Salmonella enterica species using 16S rRNA Gene PCR coupled with high-resolution melt analysis. *J Mol Diagn* **16**, 261-266 (2014).

27. S. Yang, P. Ramachandran, R. Rothman, Y.H. Hsieh, A. Hardick, H. Wong, A. Kecojevic, J. Jackman, C. Gaydos, Rapid identification of biothreat and other clinically relevant bacterial species by use of universal PCR coupled with high-resolution melting analysis. *J Clin Microbiol* **47**, 2252-2255 (2009).

28. J. B. Fan, M. S. Chee, K. L. Gunderson, Highly parallel genomic assays. *Nat Rev Genet* **7**, 632-644 (2006).

29. P. Athamanolap, W. Parekh, S.I. Fraley, V. Agarwal, D.J. Shin, M.A. Jacobs, T.H. Wang, S. Yang, Trainable high resolution melt curve machine learning classifier for large-scale reliable genotyping of sequence variants. *PLoS One* **9**, e109094 (2014).

30. S. I. Fraley, P. Athamanolap, B.J. Masek, J. Hardick, K.C. Carroll, Y.H. Hsieh, R.E. Rothman, C.A. Gaydos, T.H. Wang, S. Yang, Nested Machine Learning

Facilitates Increased Sequence Content for Large-Scale Automated High Resolution Melt Genotyping. *Sci Rep* **6**, 19218 (2016).

31. I. L. Candiloro, T. Mikeska, P. Hokland, A. Dobrovic, Rapid analysis of heterogeneously methylated DNA using digital methylation-sensitive high resolution melting: application to the CDKN2B (p15) gene. *Epigenetics Chromatin* **1**, 7 (2008).

32. H. Zou., W.R. Taylor, J.J. Harrington, F.T. Hussain, X. Cao, C.L. Loprinzi, T.R. Levine, D.K. Rex, D. Ahnen, K.L. Knigge, P. Lance, X. Jiang, D.I. Smith, D.A. Ahlquist, High detection rates of colorectal neoplasia by stool DNA testing with a novel digital melt curve assay. *Gastroenterology* **136**, 459-470 (2009).

33. B. S. Pritt., P.S. Mead, D.K. H. Johnson, D.F. Neitzel, L.B. Respicio-Kingry, J.P. Davis, E. Schiffman, L.M. Sloan, L.M. Sloan, M.E. Shriefer, A.J. Replogle, S.M. Paskewitz, J.A. Ray, J. Bjork, C.R. Steward, A. Deedon, X. Lee, L.C. Kingry, T.K. Miller, M.A. Feist, E.S. Theel, R. Patel, C.L. Irish, J. M. Petersen, Identification of a novel pathogenic Borrelia species causing Lyme borreliosis with unusually high spirochaetaemia: a descriptive study. *Lancet Infect Dis*, (2016).

34. D. E. Dietzman, G. W. Fischer, F. D. Schoenknecht, Neonatal Escherichia coli septicemia--bacterial counts in blood. *J Pediatr* **85**, 128-130 (1974).

35. J. A. Kellogg, F.L. Ferrentino, M.H. Goodstein, J. Liss, S.L. Shapiro, D.A. Bankert, Frequency of low level bacteremia in infants from birth to two months of age. *Pediatr Infect Dis J* **16**, 381-385 (1997).

36. S. Chakravorty, D. Helb, M. Burday, N. Connell, D. Alland, A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiologial Methods* **69**, 330-339 (2007).

37. T. Lisboa, G. Waterer, J. Rello, We should be measuring genomic bacterial load and virulence factors. *Crit Care Med* **38**, S656-662 (2010).

38. R. P. Dellinger, R.P. Dellinger, M.M. Levy, J.M. Carlet, J. Bion, M.M. Parker, R. Jaeschke, K. Reinhart, D.C. Angus, C. Brun-Buisson, R. Beale, T. Calandra, J.F. Dhainaut, H. Gerlach, M. Harvey, J.J. Marini, J. Marshall, M. Ranieri, G. ramsay, J. Sevransky, B.T. Thompson, S. Townsend, J.S. Vender, J.L. Zimmerman, J.L. Vincent, Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Crit Care Med* **36**, 296-327 (2008).

39. K. L. McGowan, J. A. Foster, S. E. Coffin, Outpatient pediatric blood cultures: time to positivity. *Pediatrics* **106**, 251-255 (2000).

40. O. Opota, A. Croxatto, G. Prod'hom, G. Greub, Blood culture-based diagnosis of bacteraemia: state of the art. *Clin Microbiol Infect* **21**, 313-322 (2015).

41. L. Song, D. Shan, M. Zhao, B.A. Pink, K.A. Minnehan, L. York, M. Gardel, S. Sullivan, A.F. Phillips, R.B. Hayman, D.R. Walt, D.C. Duffy, Direct detection of bacterial genomic DNA at sub-femtomolar concentrations using single molecule arrays. *Anal Chem* **85**, 1932-1939 (2013).

42. R. Patel, MALDI-TOF MS for the diagnosis of infectious diseases. *Clin Chem* **61**, 100-111 (2015).

43. K. G. Frey, J. E. Herrera-Galeano, C.L. Redden, T.V. Luu, S.L. Servetas, A.J. Mateczun, V.P. Mokashi, K.A. Bishop-Lilly, Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics* **15**, 96 (2014).

44. P. J. Sykes, M.J. Brisco, E. Hughes, J. Condon, A.A. Morley, Quantitation of targets for PCR by use of limiting dilution. *Biotechniques* **13**, 444-449 (1992).

45. B. Vogelstein, K. W. Kinzler, Digital PCR. *Proc Natl Acad Sci U S A* **96**, 9236-9241 (1999).

46. B. J. Masek, J. Hardick, H. Wong, S. Yang, Y.H. Hsieh, R.E. Rothman, C.A. Gaydos, Sensitive Detection and Serovar Differentiation of Typhoidal and Nontyphoidal Salmonella enterica Species Using 16S rRNA Gene PCR Coupled with High-Resolution Melt Analysis. *The Journal of Molecular Diagnostics* **16**, 261-266 (2014).

47. S. Yang, P. Ramachandran, A. Hardick, Y.H. Hsieh, C. Quianzon, M. Kuroki, J. Hardick, A. Kecojevic, A. Abeygunawardena, J. Zenilman, J. Melendez, V. Doshi, C. Gaydos, R.E. Rothman, Rapid PCR-Based Diagnosis of Septic Arthritis by Early Gram-Type Classification and Pathogen Identification. (2008).

48. S. Yang, P. Ramachandran, R. Rothman, Y.H. Hsieh, A. Hardick, H. Won, A. Kecojevic, J. Jackman, C. Gaydos, Rapid Identification of Biothreat and Other Clinically Relevant Bacterial Species by Use of Universal PCR Coupled with High-Resolution Melting Analysis. (2009).