UNIVERSITY OF CALIFORNIA SAN DIEGO

Advancing Rapid Molecular Diagnostics using Dielectrophoretic and Machine Learning-Enabled On-Chip Platforms

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

NanoEngineering

by

Augustine Chidi Obirieze

Committee in charge:

Professor Michael J. Heller, Chair
Professor Todd P. Coleman
Professor Sadik C. Esener
Professor Jesse V. Jokerst
Professor Ying Shirley Meng
Professor Geert W. Schmid-Schönbein

2020

The dissertation of Augustine Chidi Obirieze is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California San Diego

2020

DEDICATION

This dissertation is dedicated to my family.

EPIGRAPH

"I have learned that success is to be measured not so much by the position that one has reached in life as by the obstacles which he has overcome while trying to succeed."

–Booker T. Washington

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ESCRT	endosomal sorting complexes required for transport
DEP	dielectrophoresis
CM	Clausius–Mossotti
LR	logistic regression
HRM	high resolution melt
dHRM	digital high resolution melt
U-dHRM	universal digital high resolution melt
AUC	area under the curve
ROC	receiver operating characteristic
AC	alternating current
DC	direct current
ACE	alternating current electrokinetic
pDEP	positive dielectrophoresis
nDEP	negative dielectrophoresis
TBI	traumatic brain injury

ACKNOWLEDGEMENTS

I owe a debt of gratitude to Dr. Michael Heller and Dr. Sadik Esener. After convincing me to consider doing a PhD, Dr. Esener took me into his laboratory when I knew very little about nanoengineering. He provided me with a welcoming laboratory environment, and was always available to provide professional and personal guidance. His mentorship and support continued even after I could not join the laboratory move to OHSU. His calmness, deep scientific and engineering insight, and entrepreneurial thinking are some of his attributes that I have strived to emulate. Dr. Heller has been more than an advisor to me. His scientific astuteness, innovative thinking, and kindness continue to motivate me. I am most grateful for his mentorship.

I am very grateful to Dr. Shirley Meng, who has been my Sloan Faculty mentor since the start of my PhD studies. To the rest of my committee members, Dr. Todd Coleman, Dr. Jesse Jokerst and Dr. Geert Schönbein, thank you for all your feedback, scientific guidance, and mentorship. Dana Jimenez, the Nanoengineering department graduate advisor, has been very helpful in making sure all administrative matters are taken care off in a timely manner, and I am grateful to her.

To the members of the Esener Lab and Heller Lab, with whom I not only enjoyed working alongside, but also learned a lot from, and formed friendships with, thank you! My memory may fail me as I try to list everyone here, but I would like to thank Dr. Stuart Ibsen, Dr. Jean Lewis, Dr. Jennifer Wright, Sareh Manouchehri, Dr. Benjamin Sarno, Dr. Augusta Modestino, Dr. Elaine Skowronski, Dr. Daniel Heineck, Dr. Sejung Kim, Dr. Alaleh Golkar Narenji, Dr. Ya-San Yeh, Dr. Mukanth Vaidyanathan, Dr. Negin Mokhtari, Dr. Ajay Sapre,

Justin Plaut, Selin Esener, Kyle Gustafson, Jason Pianalto, Dr. Jason Chen. I also enjoyed working with Dr. Mridu Sinha, Dr. Daniel Ortiz Velez, Lennart Langouche, Anshu Gupta, Kevin Chen, April Aralar, William Leineweber, Sural Ranamukhaarachchi, Tyler Goshia, Esra Tiftik, Maya Rowell, and Joe Yixu Yuan.

To my UC San Diego Sloan family, Dr. Kim Barrett, April Bjornsen, Shana Slebioda, Sinai Cota, and all Sloan Scholars, thank you for being a great source of inspiration, and for your support. The Newman Catholic Community at UCSD, and Our Mother of Confidence parish were sources of support for me as well.

My family is my backbone, and I would not be where I am today without their support. I am deeply grateful to my wife, Ngozi, daughters, Chinonyerem and Chinemerem, and son, Chinedum, for the many sacrifices they have had to make since I embarked on this PhD journey. My parents, Collins and Angela, raised me in an environment that placed a high value on education, they taught me resilience, hard work and discipline, and they continue to show me their steadfast love. To my siblings, Chioma, Stephanie, Samuel and Helen, I could not have asked for a better loving, and supportive team. I have also received enormous support from my extended family, especially my uncles, Valentine and Festus.

Last but not in any way least, I thank God Almighty for blessing me in more ways than I can count or imagine, and for making it possible for me to come this far. Thank you, Lord!

I also include the following acknowledgments below as required by the University of California San Diego:

Chapter 2, in full, is a reprint of material as it appears in *Frontiers in Neurology* journal. Jean M. Lewis, Sanjay Dhawan, Augustine C. Obirieze, Benjamin Sarno, Johnny

Akers, Michael J. Heller and Clark C. Chen. “Plasma Biomarker for Post-concussive Syndrome: A Pilot Study Using an Alternating Current Electro-Kinetic Platform”. The dissertation author was the secondary investigator and secondary author of the material.

Chapter 3, in part, is being revised for submission for publication of the material. Augustine C. Obirieze, Mridu Sinha, Lennart Langouche, Hannah Mack, William Leineweber, April Aralar, Todd P. Coleman, Stephanie I. Fraley. “A Probabilistic Approach to Melt Curve-Based DNA Profiling”. The dissertation author is the primary investigator and author of the material.

Chapter 4, in part, is being revised for submission for publication of the material. Augustine C. Obirieze, Mridu Sinha, Lennart Langouche, Hannah Mack, William Leineweber, April Aralar, Todd P. Coleman, Stephanie I. Fraley. “A Probabilistic Approach to Melt Curve-Based DNA Profiling”. The dissertation author is the primary investigator and author of the material.

Chapter 5, in part, is being prepared for submission for publication of the material. Augustine C. Obirieze, Benjamin G. Sarno, Michael J. Heller. “Rapid On-Chip Electrokinetic Isolation of Bacteria from Unprocessed Whole Blood”. In preparation. To be Submitted to Lab on a chip journal. The dissertation author is the primary investigator and primary author of the material.

VITA

2007	M.B.; B.S. (M.D.), Abia State University, Nigeria
2011	M.P.H. (concentration: Biostatistics and Epidemiology), Johns Hopkins University, Baltimore, MD
2017	M.S. in NanoEngineering, University of California San Diego
2020	Ph.D. in NanoEngineering with focus in Biomedical Nanotechnology, University of California San Diego

PUBLICATIONS

Lewis JM, Dhawan S, **Obirieze AC**, Sarno B, Akers J, Heller MJ, Chen CC. Plasma Biomarker for Post-concussive Syndrome: A Pilot Study Using an Alternating Current Electro-Kinetic Platform. *Front. Neurol.*, 14 July 2020.

Obirieze AC, Sinha M, Mack H, Langouche L, Leineweber W, Aralar A, Coleman TP, Fraley SI. A Probabilistic Approach to Melt Curve-Based DNA Profiling. *Nucleic Acids Research* 2020. Submitted. Under revision.

Obirieze AC, Sarno B, Heller MJ. Rapid On-Chip Electrokinetic Isolation of Bacteria from Unprocessed Whole Blood. In preparation. To be Submitted to *Lab on a chip* journal.

ABSTRACT OF THE DISSERTATION

Advancing Rapid Molecular Diagnostics using Dielectrophoretic and Machine Learning-Enabled On-Chip Platforms

by

Augustine Chidi Obirieze

Doctor of Philosophy in NanoEngineering

University of California San Diego, 2020

Professor Michael J. Heller, Chair

Technological platforms that enable rapid biomarker detection, disease diagnosis, prognosis, and treatment monitoring can significantly improve outcomes for patients. This is especially true in traumatic brain injury (TBI) and neonatal (newborn) sepsis, for which timely

and accurate recognition is important, and clinical signs alone are not sufficient for optimal management. This dissertation explores the clinical utility of TBI biomarkers using a novel electrokinetic microarray platform, as well as the application of dielectrophoresis for the isolation of bacteria from whole blood. In addition, this doctoral research work explores the application of machine learning algorithms for microbial pathogen identification, and for the detection of emerging infections, using a novel digital high resolution melt platform.

Chapter 1

Introduction

1.1 Exosome

Exosomes are extracellular vesicles measuring ~30-150 nm in diameter that have been shown to be secreted into the extracellular space by diseased and normal cells alike, and found to be present in body fluids [1, 2]. Early documentations of observations regarding exosomes dates back to the 1970s and 1980s (Figure 1.1) [3-5]. Formed by a process that involves inward budding of the plasma membrane and the endosomal sorting complex required for transport (ESCRT) machinery, these small lipid-bilayer-membrane-enclosed particles contain nucleic acids, proteins and lipids (Figures 1.2 and 1.3) [6, 7]. The biomolecular contents of exosomes have been shown to mirror those of their cells of origin [8]. Previously thought to be a mere cellular debris, evidence is accumulating that exosomes play important biological roles, such as in cell-to-cell communication, immune response, antigen presentation, tumorigenesis, and tumor progression, through the transfer of packaged bioactive molecules (including proteins, mRNA, microRNA, long non-coding RNA and mitochondria DNA) between cells [9-12]. Studies are emerging suggesting the utility of exosome biomarkers in the diagnosis, prognosis, and therapeutic monitoring for disease management [13-17]. This has led to the increasing interests in exosomes for

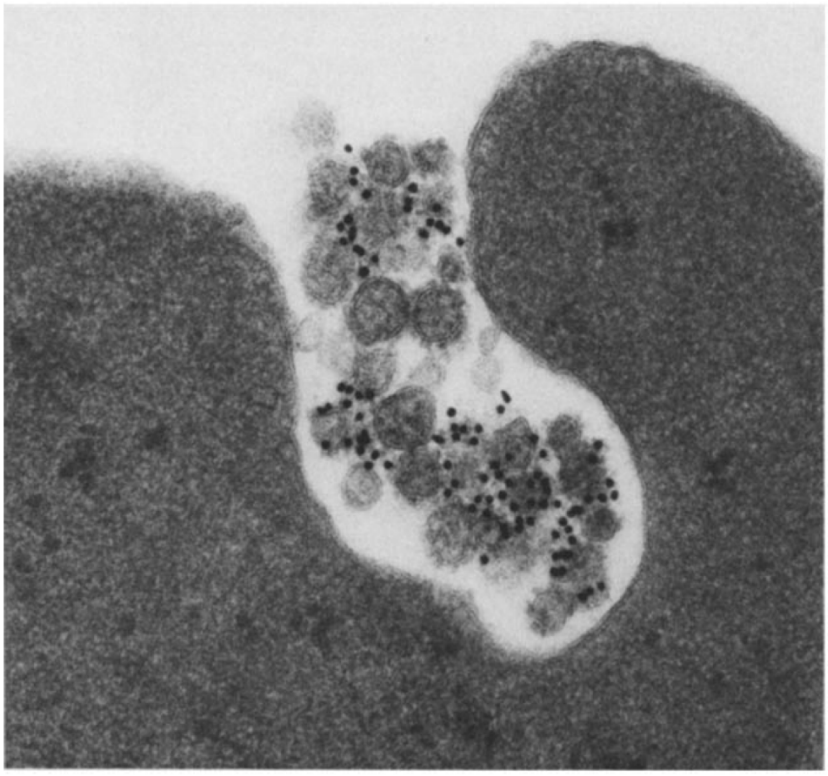


Figure 1. 1. Fusion of multivesicular elements (MVE) with the plasma membrane and release of round bodies. The figure shows exocytosis into the medium of small dense bodies labeled with gold label after a 3-h incubation at 37°C. The gold label is only present on the 50-nm bodies which are inside vesicles of 300-800 nm in diameter. The limiting membrane of the MVE is devoid of label. Reproduced with permission from [5]. Copyright: © 1985 Pan et al. Rockefeller University Press, *Journal of Cell Biology*.

liquid biopsy. In addition to oncology, interest is also growing in the application of exosome biomarkers in cardiovascular diseases [18], traumatic brain injury [19], infectious diseases [20, 21], inflammatory disorders [22], and neurodegenerative diseases [23]. Despite the growing interests and advancements in the field, optimal and effective methods for isolation, purification and analysis of exosomes have remained a challenge given their small size and the complex biological fluid in which they are present [24, 25].

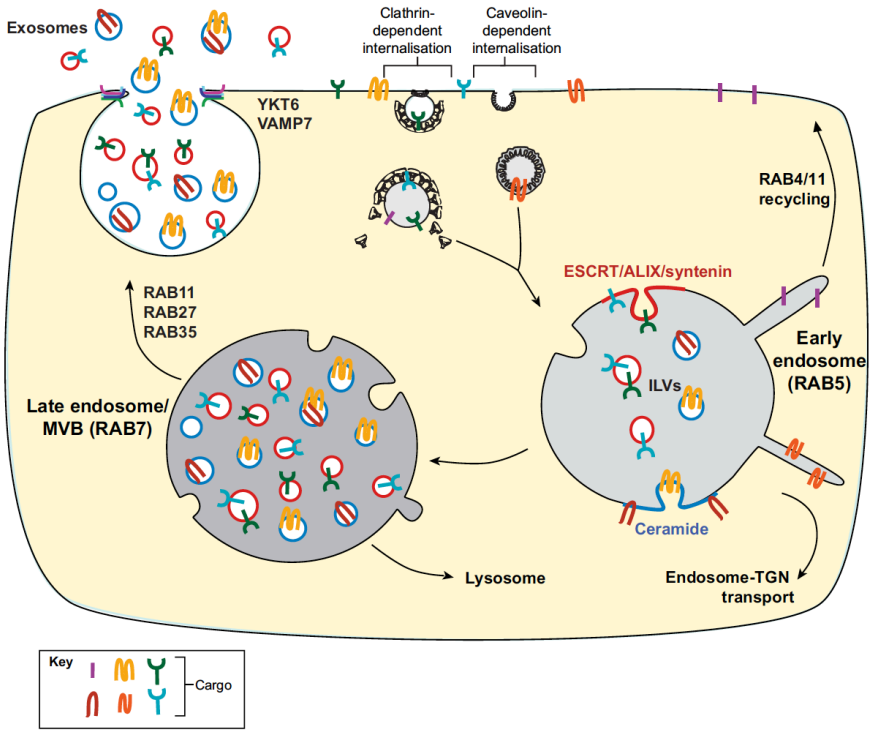


Figure 1. 2. Biogenesis of exosomes. Reproduced with permission from [6], Copyright: © 2016 McGough and Vincent. Company of Biologists, *Development*.

The standard method for exosome isolation is based on a multistep ultracentrifugation technique [26, 27]. Other exosome isolation methods include size-based techniques [28-31], affinity-based immunocapture [32-34], and microfluidics-based methods [35-38]. Harnessing the full diagnostic, as well as therapeutic and discovery potential of exosomes would require the development of technologies that allow for rapid, low-cost, high-yield, efficient isolation of uncontaminated exosomes in as close to their natural unchanged and undamaged state as possible.

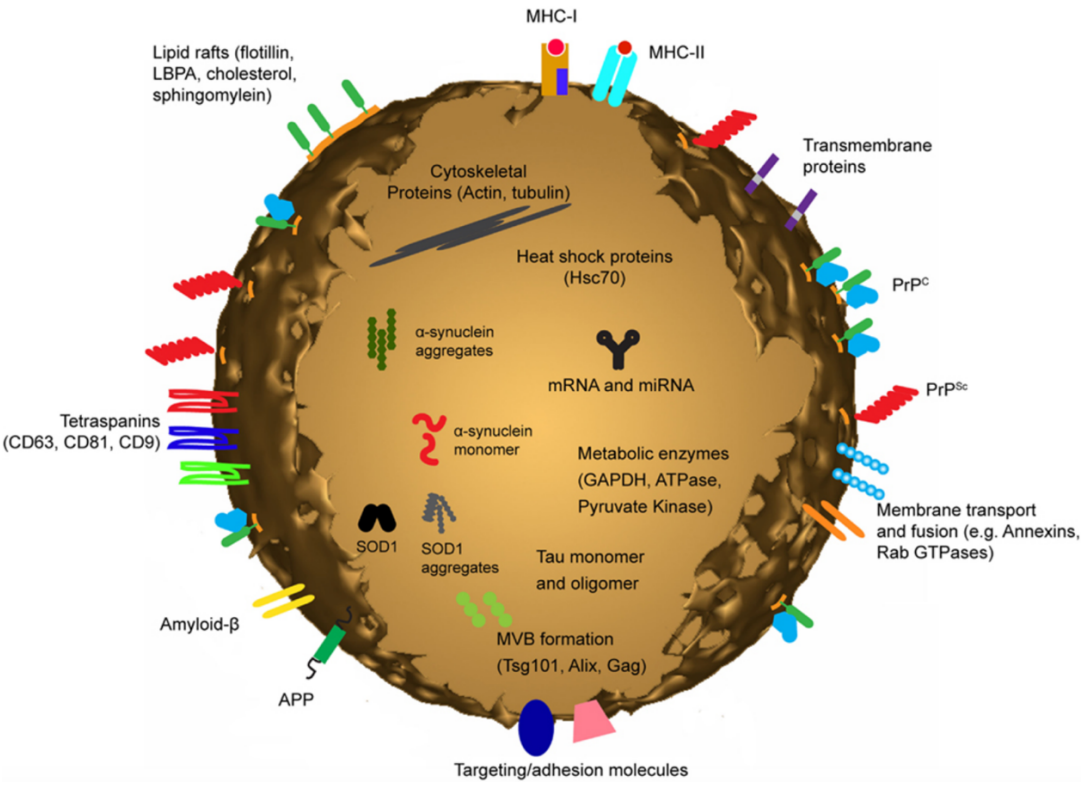


Figure 1. 3. Schematic representation of exosome membrane composition and biomolecular contents. Reproduced with permission from [7]. Copyright: © 2012 Bellingham et al. *Frontiers in Physiology*.

1.2 Dielectrophoresis

The ability to manipulate and separate micro- and nanoscale particles has long been of interest in biology and medicine. Over the years, a variety of techniques have been developed for this purpose, including optical [39-41], electrokinetic [42-45], mechanical [46, 47], hydrodynamic [48, 49], acoustic [50-52], and magnetic [53, 54]. Among these particle manipulation techniques, dielectrophoresis (DEP) has attracted growing attention in recent decades because of the advantages it offers over other methods, such as fully controllable and selective particle manipulation, potential for parallelization, miniaturization, ease of integration to existing electronics, and inexpensive fabrication [42, 43].

DEP, an electrokinetic phenomenon that was first named and described by Herbert Pohl in the 1950s [55, 56], is the translational motion of an uncharged polarizable particle in a nonuniform electric field (Figure 1.4) [57]. The particle motion results from the imposition on it of an external force, the dielectrophoretic force, following the induced polarization of the particle under the influence of the nonuniform electric field. Unlike electrophoresis, for which the electric field-induced particle motion relies on the particle having a net electrical charge, dielectrophoresis does not rely on intrinsic net charge of the particle, but instead relies on the dielectric properties of the particle and the suspending medium, thereby enabling the manipulation of an electrically neutral particle. The wide variety of DEP microdevices that have been developed and employed for the manipulation and separation of bioparticles (i.e. DNA, bacteria, viruses, proteins, mammalian cells) can be broadly classified based on the applied electric field into two:

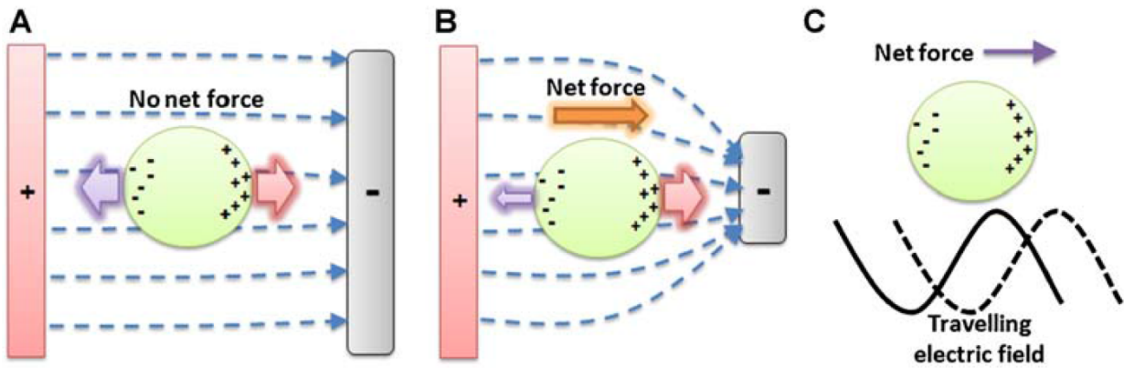


Figure 1. 4. Schematic of a particle in a suspending medium (A) experiencing zero net force while placed in a uniform electric field; (B) experiencing DEP force when placed in a non-uniform electric field; (C) experiencing DEP force because of the electric filed gradient. Reproduced with permission from [57]. Copyright: © 2009 Zhang et al. Springer Nature. *Analytical and Bioanalytical Chemistry*.

(i) alternating current (AC) DEP (also referred to as metal electrode-based DEP), and (ii) direct current (DC) DEP (also referred to as insulator-based DEP) [58]. In general, the design of an AC DEP microdevice involves embedding an array of metal electrodes, which generate spatially nonuniform electric field, inside the microdevice channels [59]. Advantages of the AC DEP include the ability to change the frequency of the applied electric field, thereby allowing for more efficient particle manipulation and separation. On the other hand, for DC DEP, spatially nonuniform electric field is generally created by placing insulating obstacles (hurdles) inside the microdevice, with the electrodes remotely positioned at the end of the microdevice channels [60, 61]. Advantages of the DC DEP over AC DEP include less fouling at the region of particle manipulation given that insulators are less prone to fouling compared to metal electrodes, reduced electrolysis (bubbling) in the microchannels due to remote positioning of electrodes, and simple device fabrication [61, 62]. However, a drawback with the DC DEP is the potential for Joule heating (and the resulting increase in temperature inside the microdevice) following

power loss in the DC electric field [63, 64]. For both designs, the DEP force acting on the polarized particle moves it toward (positive DEP, pDEP) or away from (negative DEP, nDEP) the region of strong electric field, depending upon the relative polarizability of the particle and the surrounding medium. A particle exposed to a nonuniform electric field experiences pDEP when its induced polarizability is higher than that of the medium within which it is suspended. Conversely, a particle subjected to an applied nonuniform electric field experiences nDEP when it is less polarizable than the surrounding medium (Figure 1.5) [58].

For a homogenous dielectric spherical particle in an AC electric field, the time-averaged DEP force is expressed as [65]:

$$F_{\mathrm{DEP}} = 2\pi\varepsilon_0\varepsilon_{\mathrm{m}} r^3 \mathrm{Re}[f_{\mathrm{cm}}] \nabla E^2 \tag{1.1}$$

where ε_0 is the permittivity of free space, ε_{m} is the relative permittivity of the surrounding medium, r is the particle radius, E is the root mean square (rms) value of the electric field, and $\mathrm{Re}[f_{\mathrm{cm}}]$is the real part of the Clausius–Mossotti (CM) factor, f_{cm}, given by

$$f_{\mathrm{cm}} = \frac{\varepsilon_{\mathrm{p}}^* - \varepsilon_{\mathrm{m}}^*}{\varepsilon_{\mathrm{p}}^* + 2\varepsilon_{\mathrm{m}}^*} \tag{1.2}$$

where ε^* is the complex permittivity, subscripts p and m denote particle and medium, respectively. The complex permittivity, ε^*, is defined by

$$\varepsilon^* = \varepsilon - j\frac{\sigma}{\omega} \tag{1.3}$$

where σ is the electrical conductivity, ω is the frequency of the applied electric field, and $j=\sqrt{-1}$. As shown in Equations (1.2) and (1.3), the CM factor, f_{cm}, is a complex function

of the frequency of the electric field, the permittivity and the electrical conductivity of the particle and the medium. For an isolated solid homogenous spherical particle, the real part of the CM factor, Re[f_{cm}], varies between -0.5 ($\varepsilon_p^* < \varepsilon_m^*$) and +1 ($\varepsilon_p^* > \varepsilon_m^*$) [66]. According to Equation (1.1), the sign of the CM factor, f_{cm}, determines the direction of the DEP force. When f_{cm} is positive, a pDEP is created. Conversely, when f_{cm} is negative, the particle experiences nDEP. When f_{cm} is zero, the DEP force is zero, and there is no particle movement. For a given particle and suspending medium, varying the frequency of the applied electric field can vary the Re[f_{cm}] (Figure 1.6) [67]. The electric field frequency at which the Re[f_{cm}] is zero is referred to as the crossover frequency.

In a study it was showed that, for the CM factor at low and high frequency limits, conductivity governs DEP behavior at low frequency, while permittivity has more influence on DEP behavior at high electric field frequency [68]. Thus, establishing two scenarios that govern the relationship between the applied electric field frequency and the Re[f_{cm}]: (i) this occurs when $\sigma_p < \sigma_m$ and $\varepsilon_p > \varepsilon_m$, making Re[$f_{cm}$] positive at high electric field frequencies and negative at low electric field frequencies; (ii) this occurs when $\sigma_p > \sigma_m$ and $\varepsilon_p < \varepsilon_m$, making Re[$f_{cm}$] positive at low electric field frequencies and negative at high electric field frequencies (Figure 1.6) [69].

For decades pDEP has been utilized for capturing bioparticles contained in a suspending medium. However, this particle trapping has been limited by the requirement of a low conductivity media. Therefore, given that physiological media, such as blood and other body fluids, are high-conductance solutions, this has necessitated the dilution of the physiological sample prior to subjecting it to a nonuniform electric field [70-73].

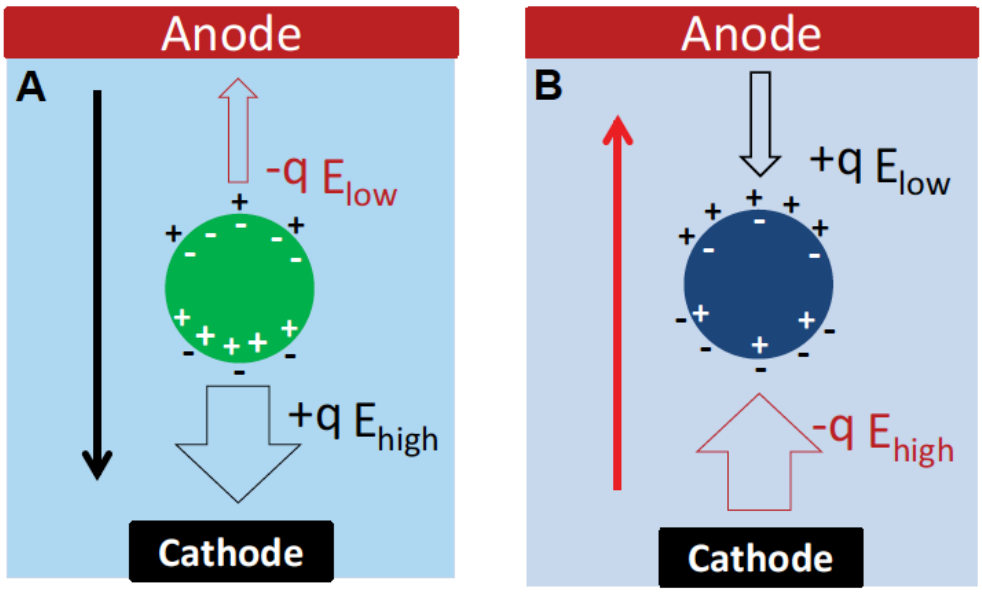


Figure 1. 5. Schematic showing the direction of the DEP force experienced by a polarized particle in a suspending medium under the influence of a nonunform electric field. (A) The particle is more polarizable than the suspending medium and it experiences net force toward the higher electric field (E_{high}) region. This process is known as pDEP. (B) The particle is less polarizable than the suspending medium, and the net force on the particle acts toward the lower electric field (El_{ow}) region. This type of particle motion is known as nDEP. Reproduced with permission from [58]. Copyright: © 2014 Jubery et al. John Wiley and Sons. *Electrophoresis*.

In addition, bioparticle manipulation and separation at the nanoscale were previously thought to be impossible owing to the effect of Brownian motion, and the need for high voltages that are impractical to generate at the electric field frequencies at which DEP operates [74, 75]. Alternative electrode design approaches have been explored to overcomes these limitations. Recently, an AC electrokinetic (ACE) microarray chip device have been developed and successfully employed for the manipulation and isolation of particles, such as DNA, exosomes, nanoparticles, and viruses, from high conductance buffer, undiluted whole blood, plasma, and serum [38, 76-80]. The platform consists of a microfluidic chamber with a multi-layered microarray electrode, fabricated

on a silicon base (Figure 1.8) [80]. With a small sample volume (25 - 50 μL), it has been demonstrated to have potential applications in cancer diagnosis, prognosis and therapy effect monitoring, drug delivery nanoparticle retrieval and analysis, and traumatic brain injury assessment (Figure 1. 9) [38].

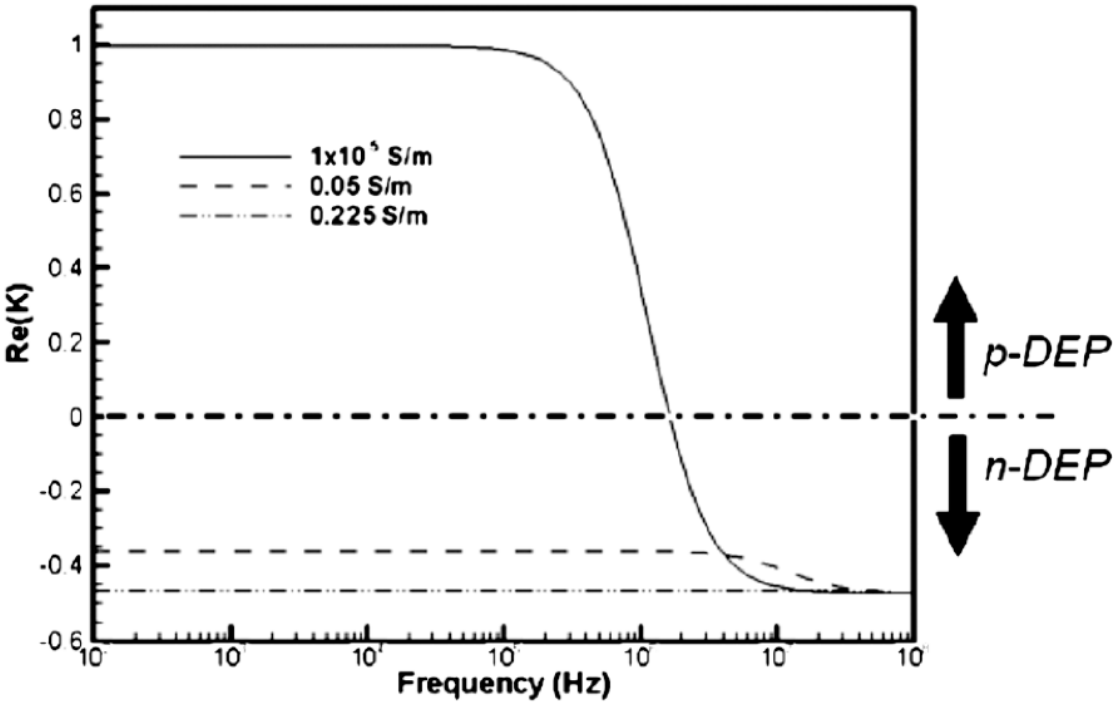


Figure 1. 6. The real part of the Clausius–Mossotti factor as a function of the applied electric field frequency. Data is shown for a 1 μm solid polystyrene spherical particle suspended in an aqueous medium of variable conductivity. The properties of the particle and the medium are σ_p = 0.01 S/m, ε_p = 2.55, ε_m = 78.5. For a suspending medium with very low conductivity (1 x 10-5 S/m), the polystyrene particles experienced pDEP up to ~ 2 MHz. Reproduced with permission from [67]. Copyright: © 2010 Koklu et al. AIP Publishing. *Biomicrofluidics*.

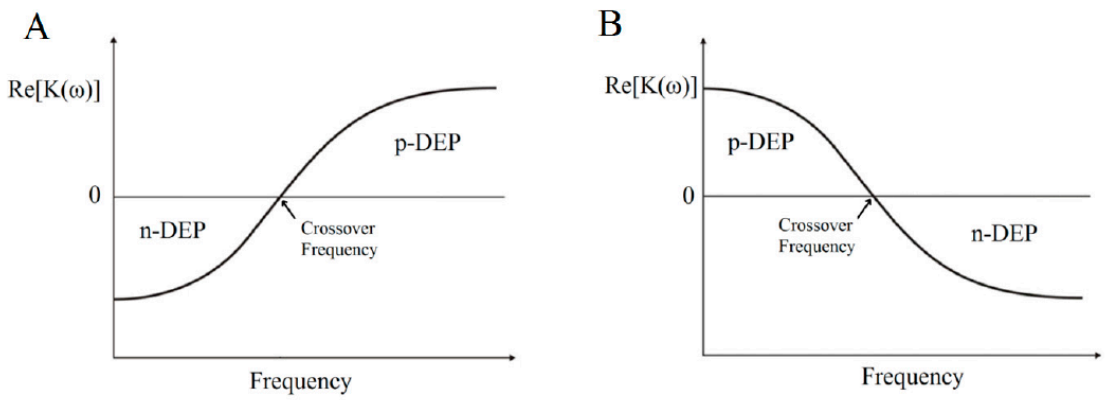


Figure 1. 7. An illustration of the relationship between applied electric field frequency and the Clausius–Mossotti factor with respect to the permittivity and conductivity of the particle and the suspending medium. (A) when $\sigma_p < \sigma_m$ and $\varepsilon_p > \varepsilon_m$; (B) when $\sigma_p > \sigma_m$ and $\varepsilon_p < \varepsilon_m$. Reproduced with permission from [69]. Copyright: © 2017 Abd Rahman et al. MDPI. *Sensors (Basel)*.

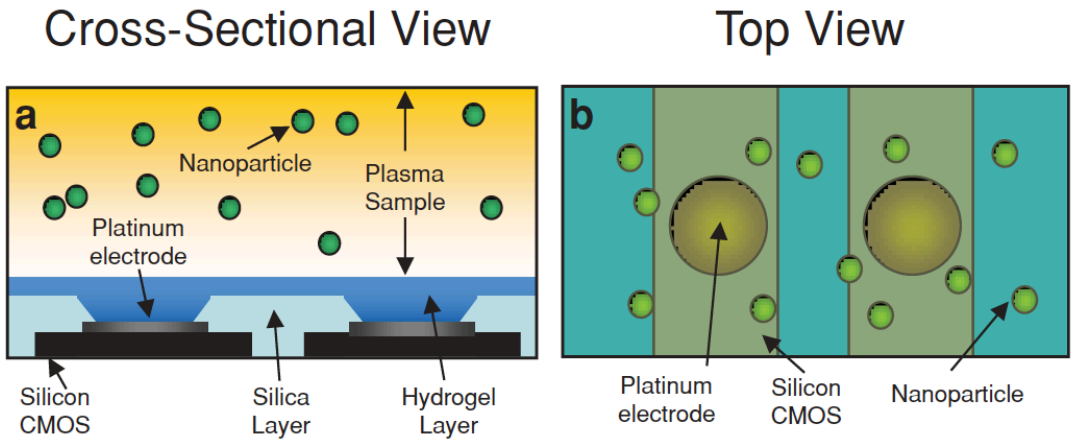


Figure 1. 8. Schematic of the cross-sectional (a) and top (b) views of an alternating current electrokinetic microelectrode array platform depicting application in nanoparticle isolation. Reproduced with permission from [80]. Copyright: © 2015 Ibsen et al. John Wiley and Sons. *Small.*

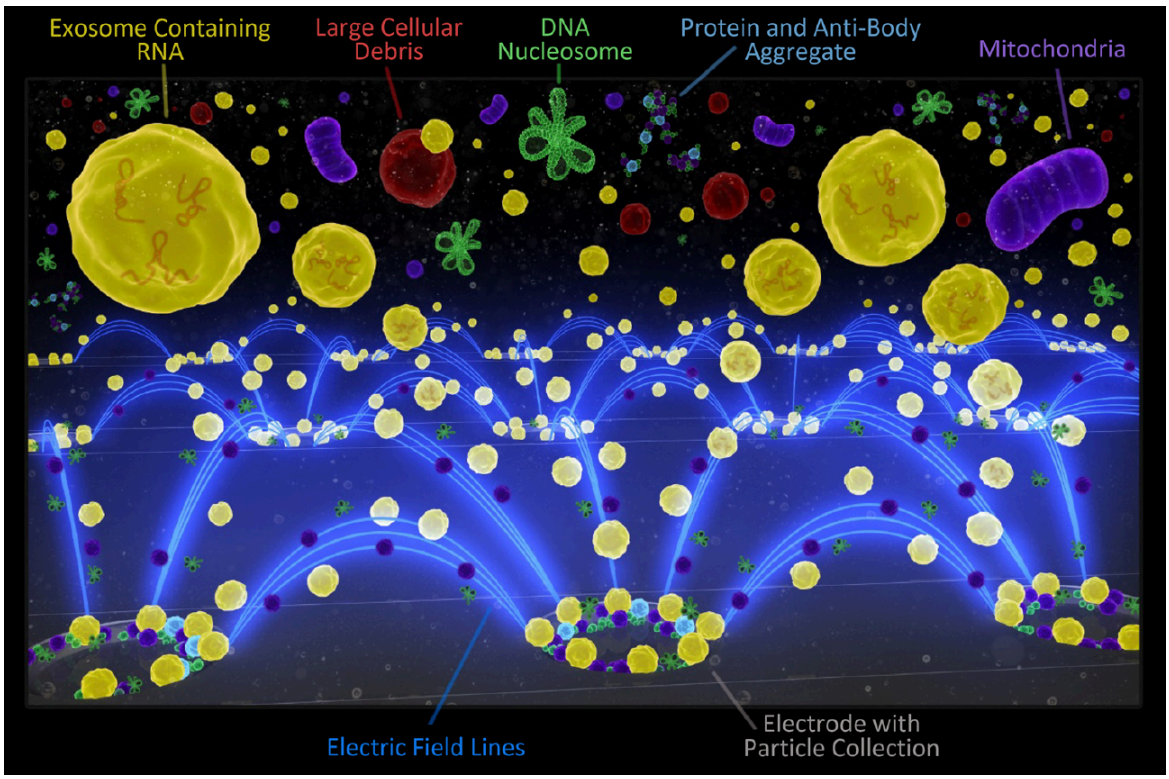


Figure 1. 9. Schematic representation of exosomes and other nanoparticles capture on the AC electrokinetic device microelectrodes. The electric field lines (blue) run between individual microelectrodes on the microarray and converge onto the edges of the microelectrodes, forming the DEP high-field regions. The exosomes, EVs, nucleosomes, and aggregated protein particles collect in these high-field regions around the microelectrode edges. Any cells or larger particles in the sample (blood, plasma, serum, etc.) are concentrated into the DEP low-field areas between the microelectrodes, while the lower molecular weight biomolecules are unaffected by the DEP electric fields. A fluid wash removes any cells and the other plasma materials, while the nanosize biomarkers (exosomes, etc.) remain concentrated in the DEP high-field regions. Reproduced with permission from [38]. Copyright: © 2017 Ibsen et al. American Chemical Society. *ACS Nano.*

1.3 High-resolution melt curve

The denaturation or melting of DNA with heat is a fundamental property of DNA that takes place when a double-stranded helical DNA (dsDNA) separates into random coils of single-stranded DNA (ssDNA). Conversely, hybridization or annealing occurs when there is a shift from the single-stranded random coil state to the double-stranded helical state. In both processes, the transition occurs in a narrow temperature interval about a melting temperature, Tm, the temperature at which half of the DNA is double-stranded, and the remaining half is single-stranded. The DNA melting process can be traced by using dsDNA-binding dyes that fluoresces in the presence of dsDNA but not in the presence of ssDNA. As the temperature increases, the fluorescence intensity decreases as the dsDNA gets denatured to ssDNA. This change in fluorescence as a function of temperature is captured by plotting fluorescence vs. temperature to give rise to unique DNA melting curve (Figure 1.10A) [81]. The Tm can be approximated by taking a negative derivative of the melting curve (Figure 1.10B).

Historically, fluorescent DNA melting analysis replaced UV absorbance methods for monitoring thermal DNA melting. In contrast to the UV absorbance method, which required large amounts (in microgram) of DNA and slower heating rates, fluorescence DNA melting is more sensitive, requiring much less amounts (in nanogram) of DNA, making it attractive for quantification, genotyping and variant scanning of polymerase chain reaction post-amplification products [82, 83]. Early fluorescent methods relied on labeled probes [84, 85]. However, with labeled probes, sequence variants under the probe are detected but not variants outside the region of the probe. The advent of intercalating dyes, such as SYBR Green, made it possible to scan for variants anywhere within the

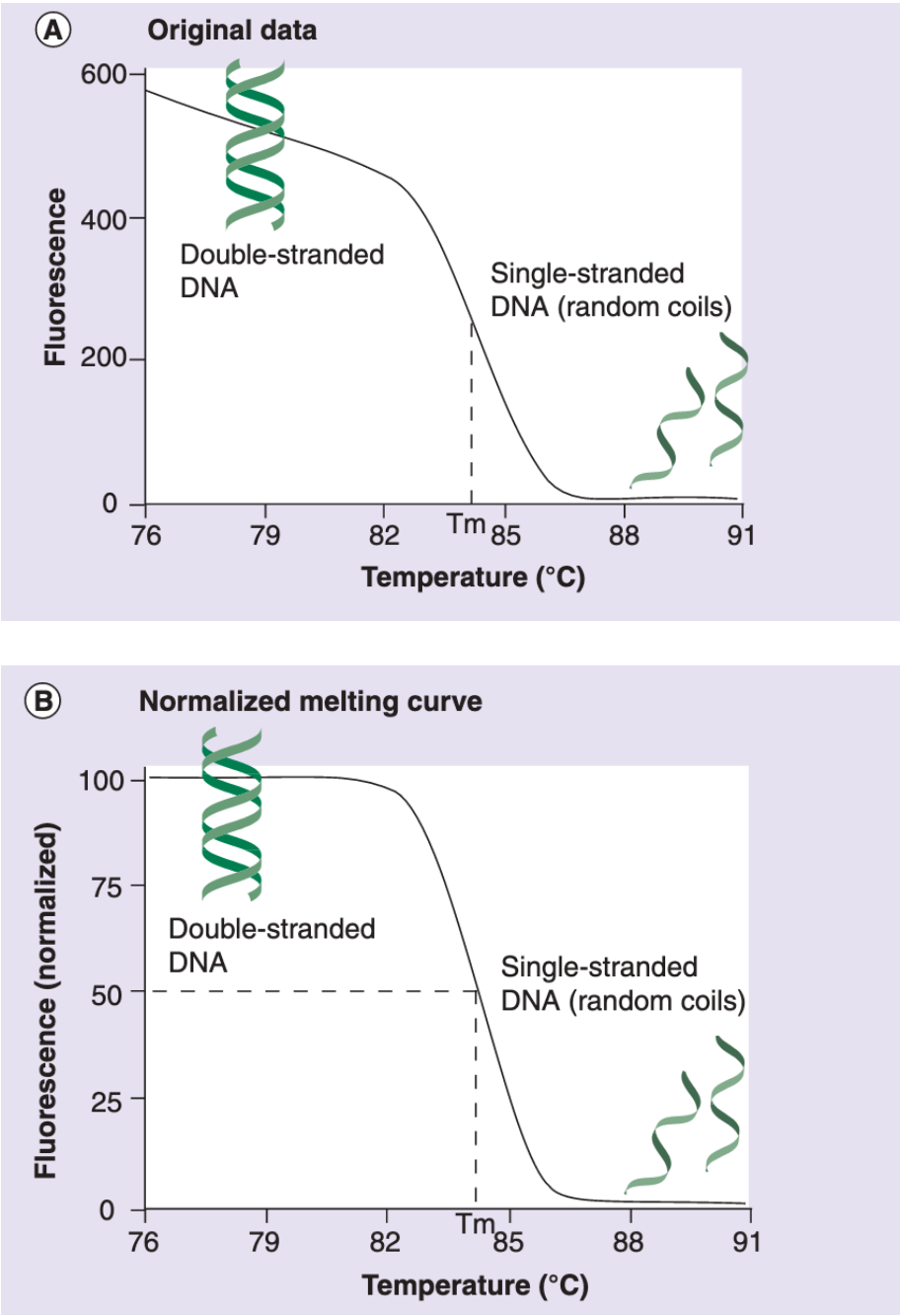


Figure 1. 10. Fluorescence of DNA-binding dye as a function of temperature. (A) Original fluorescence data showing a linear decrease of fluorescence followed by rapid decrease centered around the melting temperature (Tm). (B) The original data is normalized between 0 and 100% so that the curve is horizontal outside the transition. Reproduced with permission from [81]. Copyright: © 2007 Reed et al. Future Medicine Ltd. *Pharmacogenomics.*

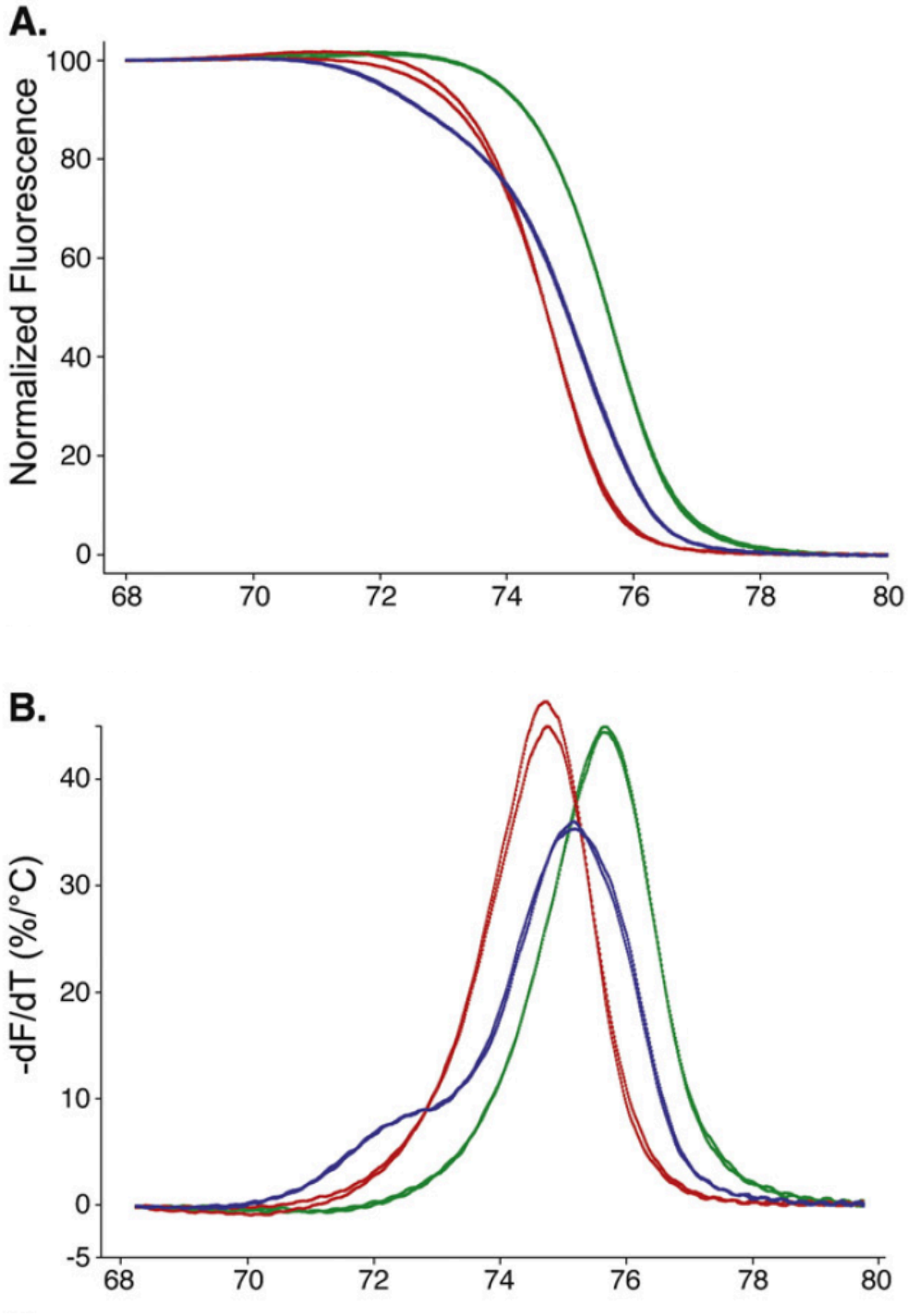


Figure 1. 11. Amplicon melting analyses for duplicate samples of factor V (Leiden) 1691 GNA wild-type (green), heterozygous (blue) and homozygous mutant (red) samples. (A) Normalized melting curves, (B) derivative plots. Reproduced with permission from [86]. Copyright: © 2008 Erali et al. Elsevier. *Experimental and Molecular Pathology.*

PCR product [87]. However, there were some limitations including [85, 88]: (i) the need for high concentrations of the dye, which inhibits PCR, (ii) requiring additional sample processing steps post-amplification for PCR product separation and amplicon purification prior, with the risk of introducing contaminants, and (iii) was able to differentiate homozygous variants that differ in Tm but it was difficult to detect heteroduplexes. High resolution melting (HRM) was developed to overcome these limitations.

HRM analysis was made possible by the discovery and synthesis of saturating DNA-binding dyes that are compatible with PCR in high concentrations, as well as advances in the resolution of melting instruments and data analysis [89-91]. HRM is a simple, inexpensive, rapid closed-tube homogenous technique that can be performed with just a DNA-binding dye in addition to the PCR reagents and the target nucleic acids, and does not require post-PCR sample processing, thereby reducing the risk of PCR product contamination [92]. The fluorescence data that is obtained during the DNA melting process can be analyzed based on the shape of the melting curve or on the melting temperature, Tm [89]. The applications for HRM analysis include genotyping [93, 94], mutation scanning [85, 95], and sequence matching [96]. HRM analysis has been gaining popularity as a research and clinical diagnostic tool.

Because of its bulk reaction format, traditional HRM analysis has an inherent limitation in resolving individual nucleic acid targets within a heterogenous sample containing a mixture of nucleic acids. This is because, with the bulk format, the melt curve represents an average of the individual melt curves of all nucleic acids targets within the sample [97]. Conventional approaches such as microarray, while allowing for semi-quantitative profiling, lack the sensitivity to detect individual nucleic acid

sequences at low concentrations and require prior knowledge of target nucleic acid sequences, making novel discoveries impossible. Highly sensitive and multiplexing technologies such as next generation sequencing, in addition to being costly, time consuming and not widely accessible, are prone to errors in base calling. By taking advantage of advances in digital PCR (dPCR), which allows for increased sensitivity, and progress in microfluidics technology, combined with universal (or broad-based) amplification, a novel universal digital high resolution melt (U-dHRM) platform has been developed [98, 99]. One application for which the utility of U-dHRM has been demonstrated is in microbial pathogen identification for infectious diseases diagnostics. As an example, by targeting the bacterial 16s rDNA gene using a universal primer, multiple bacterial nucleic acid targets within a sample can be amplified in partitioned dPCR reactions, and the melt curve of each target amplicon can be generated and analyzed.

1.4 Organization of the dissertation

This dissertation has two main areas of focus: (i) the application of an AC dielectrophoretic microelectrode array platform for the isolation and biomarker analysis of extracellular vesicles (exosomes) for the assessment of patients suffering traumatic brain injury; (ii) the development of machine learning algorithms for profiling high resolution DNA melt curves for microbial pathogen identification, and for the detection of novel (previously unknown) microbial DNA melt curves.

Chapter 2 explores the utility of an alternating current electrokinetic (ACE)

platform for the discovery and analysis of exosome biomarkers for traumatic brain injury. The levels of glial fibrillary acidic protein (GFAP), Tau, and ubiquitin C-terminal hydrolase L1 (UCH-L1) expression on exosomes, and the level of cell-free DNA (cfDNA) were assessed.

Chapter 3 discusses the development of probabilistic classification methods for universal digital high resolution melt curve analysis for application in microbial pathogen identification. The classification performance was assessed using DNA melt curves of ten bacterial species that are among the major causative organisms for bloodstream infections in the newborn.

Chapter 4 describes the development of machine learning algorithms for the detection of novel nucleic acid sequences using DNA melt curves generated from a digital high resolution melt platform. This has application in the identification of emerging pathogens.

Chapter 5 discusses future work that explores the utility of the ACE platform for the isolation of bacteria directly from unprocessed blood.

Chapter 2

Plasma Biomarker for Post-concussive Syndrome. A Pilot Study Using an Alternating Current Electro-Kinetic Platform

2.1 Abstract

Background: Technology platforms that afford biomarker discovery in patients suffering from traumatic brain injury (TBI) remain an unmet medical need. Here, we describe an observational pilot study to explore the utility of an alternating current electrokinetic (ACE) microchip device in this context.

Methods: Blood samples were collected from participating subjects with and without minor TBI. Plasma levels of glial fibrillary acidic protein (GFAP), Tau, ubiquitin C-terminal hydrolase L1 (UCH-L1), and cell-free DNA (cfDNA) were determined in subjects with and without minor TBI using ACE microchip device followed by on-chip

immunofluorescent analysis. Post-concussive symptoms were assessed using the Rivermead Post Concussion Symptoms Questionnaire (RPCSQ) at one-month follow-up.

Results: Highest levels of GFAP, UCH-L1, and Tau were seen in two minor TBI subjects with abnormality on head computed tomography (CT). In patients without abnormal head CT, Tau and GFAP levels discriminated between plasma from minor-TBI and non-TBI patients, with sensitivity and specificity of 64–72 and 50%, respectively. Plasma GFAP, UCH-L1, and Tau strongly correlated with the cumulative RPCSQ score. Plasma UCH-L1 and GFAP exhibited highest correlation to sensitivity to noise and light ($r = 0.96$ and 0.91, respectively, $p < 0.001$). Plasma UCH-L1 and Tau showed highest correlation with headache ($r = 0.74$ and 0.78, respectively, $p < 0.001$), sleep disturbance ($r = 0.69$ and 0.84, respectively, $p < 0.001$), and cognitive symptoms, including forgetfulness ($r = 0.76$ and 0.74, respectively, $p < 0.001$), poor concentration ($r = 0.68$ and 0.76, respectively, $p < 0.001$), and time required for information processing ($r = 0.77$ and 0.81, respectively, $p < 0.001$). cfDNA exhibited a strong correlation with depression ($r = 0.79$, $p < 0.01$) and dizziness ($r = 0.69$, $p < 0.01$). While cfDNA demonstrated positive correlation with dizziness and depression ($r = 0.69$ and 0.79, respectively, $p < 0.001$), no significant correlation was observed between cumulative RPCSQ and cfDNA ($r = 0.07$, $p = 0.81$).

Conclusion: We provide proof-of-principle results supporting the utility of ACE microchip for plasma biomarker analysis in patients with minor TBI.

2.2 Introduction

Each year, over 1.7 million people in the U.S. suffer traumatic brain injury (TBI), requiring medical attention [100]. Depending on the severity of the clinical presentation, TBI is classified into mild, moderate, or severe [101]. >80% of head trauma patients who present to the emergency room suffer from mild TBI. While the majority of these mild TBI patients are discharged from the emergency room on the same day and recover without detectable long-term sequelae, 30% of mild TBI patients will have persistent "post-concussive" symptoms that significantly compromise their quality of life, including headache, fatigue, as well as altered sensation, sleep, and attention span [102-104]. Notably, ~8% of mild TBI patients demonstrate visible injury to the cerebrum on computerized tomography (CT). Direct and indirect costs associated medical care and productivity loss associated with mild TBI exceeds $60 billion annually [105]. Mild TBI involves complex pathophysiologic processes associated with microscopic shearing of cells in the central nervous system secondary to traumatic biomechanical forces to the head [106]. Such shearing induces damage to the cell, resulting in the release of neuronal and astrocytic proteins or cell-free DNA (cfDNA) not normally found in the extracellular space [107], including glial fibrillary acidic protein (GFAP) [108], ubiquitin carboxyl-terminal hydrolase L1 (UCH-L1) [109], and Tau [110]. Tau is a microtubule stabilizing protein abundant in neurons [111], and released into the extracellular space upon neuronal damage [112]. GFAP is an intermediate filament protein highly abundant in astrocytes [113], and increased levels in blood or CSF are linked to axonal injury [114, 115]. UCH-L1 is a neuronal protease, with increased levels linked to brain injury [116-118]. Since ~20% of circulating blood volume perfuse the cerebrum and TBI often compromises the

blood-brain barrier, these released proteins can be detected in peripheral blood drawn from TBI patients [119]. The high sensitivity and specificity of GFAP and UCH-L1 as proxy for abnormal CT following TBI have paved way to clearance by the U.S. Food and Drug Administration (FDA) for their use in TBI work-up. While cfDNA has been shown to predict mortality in severe TBI patients, it becomes imperative to explore the utility of cfDNA in the management of mild TBI patients that constitute ~ 75% of the patients with TBI annually [120-122]. In contrast to the number of tools that afford study and characterization of TBI associated structural injuries, there are currently no standardized and well-established clinical criteria or biomarkers that afford identification of the minority of mild TBI patients who suffer long-term sequelae despite absence of detectable structural damage to the cerebrum [104, 123]. Currently, study of these symptoms relies on questionnaires, such as the Rivermead Post Concussion Symptoms Questionnaire (RPCSQ) [124, 125], that are administered at regular intervals after the injury. Development of predictive tools for likelihood of postconcussive symptoms can meaningfully impact the care of mild TBI patients and advance our understanding of the underlying pathophysiology. Here, we explore the utility of plasma GFAP, UCH-L1, Tau, and cell-free DNA in this context using an alternating current electro kinetic (ACE) microchip device.

ACE microarray chips use alternating current to isolate macro-molecular complexes from bio-fluids, such as blood or cerebrospinal fluid, in a highly efficient manner [38]. Of note, nearly all plasma GFAP, Tau, and UCH-L1 are found in these macro-molecular complexes [126, 127]. The chip-retained protein can be labeled with fluorescent antibodies or dyes specific for the biomarker of interest. Quantitative on-chip

fluorescent imaging analysis is then carried out Figure 2.1A [128, 129]. In this pilot study, we correlated the levels of ACE isolated GFAP, UCH-L1, Tau and cell-free DNA from blood drawn at the time of injury to Rivermead questionnaire results collected at one-month after head injury from mild TBI patients.

2.2 Methods

2.2.1 Study Design

Research protocol was approved by the University of California San Diego Institutional Review Board (Study subjects were patients who sought care at the Emergency Department of the University of California San Diego Hillcrest Medical Center. The hospital serves the endogenous population in the San Diego County. Inclusion criteria include: patients age > 18 and capable for consent based on- (i) determination of the treating physicians, and (ii) the University of California San Diego Brief Instrument for Assessing Decisional Capacity for Clinical Research (UBACC) 10 item scale administered by the trained clinical coordinator [130]. There were no explicit exclusion criteria.

To avoid study interference with the standard-of-care for patients in the emergency ward, the study is designed such that the study coordinator regularly checks in with the treating physician to identify potential candidate for consent. Treatment decisions, including indication for head CT, were made entirely by the treating physician.

Each patient underwent blood draw as per standard-of-care. Informed consent was obtained from each participating patient by a dedicated research assistant on the day of

presentation. Blood samples that remain after standard laboratory chemistry were collected, and no blood draw beyond the standard-of-care venipuncture was performed. The residual blood was collected from the chemistry lab the day after the presentation after adequate laboratory values were reported to the Electronic Medical Record. Hemolyzed samples were not collected. The volume of the residual blood ranged from 100 µl to 1.5ml. These samples were collected within 72 h of blood draw and stored in liquid nitrogen.

Samples were collected from consecutive patients who presented with history of minor TBI, defined by Glasgow Coma Scale (GCS) of > 13, and without history of loss of consciousness [131, 132]. The collection period extended between 2015 and 2016. In parallel, our protocol allowed collection of blood from patients who presented to the Emergency Ward with non-TBI and non-neurologic complaints. Patients who required major medical intervention, such as cardiac catheterization or surgical intervention were excluded from the study to minimize risk of adding stress of study participation to the patient. Residual blood after completion of standard chemistry was collected from these patients in the same manner as described above. Samples were stored in 4ºC refrigeration before transportation to the biorepository. After transportation to the biorepository, the samples were de-identified and stored as processed plasma by centrifugation at 1,300 rpm for 10min followed by 3,000 rpm for 10min. Samples were stored at −80ºC before analysis. Procedures for de-identification and human subject protection were performed in compliance to the hospital policy. The staff who analyzed the sample was blinded to the clinical history of the patients.

For all study subjects, participation in this study did not alter the standard of care,

including the routine one-month post-presentation follow-up for patients with minor TBI. For the patients who presented with minor TBI, they underwent standard-of-care work-up as determined by the attending Emergency Ward faculty physician, including non-contrast computed tomography (CT) scan of the head. The only exception to the above is that the Rivermead Post Concussion Symptoms Questionnaire (RPCSQ) was administered by a trained study coordinator at this one-month follow-up [124, 125]. During this assessment, the patients are asked to report only symptoms that were present prior to the minor TBI.

2.2.2 ACE-Based Processing of Plasma Samples

ACE chips were purchased from Biological Dynamics, Inc. (San Diego, CA). ACE-based processing of plasma samples has been demonstrated previously [133]. A syringe pump set to withdrawal mode served to regulate fluid flow across the ACE chip. Tygon tubing (inner diameter, 0.020 inches; outer diameter, 0.060 inches) was attached with superglue to either end of the chip, both ends were capped with syringe needles, and a 1ml syringe was attached to one end. Twenty-five µl of thawed plasma was drawn onto the chip. An alternating current (AC) electric field was applied to the chip for 10min at 14 volts peak-to-peak and 15 kHz to immobilize extracellular vesicles and other nanoparticles onto the microelectrode edges. The technique to isolate exosomes from plasma samples has been demonstrated previously [38]. With the AC field still on, the ACE chip was then washed with 200 µl of 0.5X PBS for an additional 10min. The time taken for the entire process was 20min for the ACE-based isolation, plus an additional 45–90min for antibody binding steps (Figure 2.1A).

2.2.3 On-Chip Immunofluorescent Analysis

Two biomarkers were tested simultaneously on each chip, using FITC and TRITC filter sets on the microscope. On-chip immunofluorescent analysis has been demonstrated previously [133]. The manufacturer and catalog number of the antibodies used are as follows: Rabbit anti-UCH-L1: Cell Signaling Technologies, clone D3T2E, #13179, diluted 1:800; Alexa Fluor 594-goat anti rabbit IgG, highly cross-adsorbed, Life Technologies #A11039, diluted 1:2000; Alexa Fluor 488 mouse anti GFAP: clone 1B4, BD Biosciences #560297, diluted 1:10; Mouse anti-Tau: clone TAU-5 (total-tau), Life Technologies #ABH0042, diluted 1:50; Alexa Fluor 594-goat anti mouse IgG, highly cross-adsorbed, Life Technologies #A11032, diluted 1:2000.

To enable access of the antibodies to proteins within the luminal space of the vesicles, EV membranes were permeabilized using 0.1% saponin for 10min. To label cfDNA, the selective dye YOYO-1 was added to a concentration of 1:5,000. Antibody incubations were performed for 45–90min at room temperature, or, if recommended by the manufacturer, overnight at 4ºC for optimal binding. For directly conjugated Alexa Fluor 488-anti-GFAP antibody (BD Pharmingen), samples were washed with PBS, then visualized and photographed for further analysis. For anti-Tau or anti-UCH-L1 (Life Technologies; Cell Signaling Technology), following the wash, Alexa Fluor 594-conjugated secondary antibody (Novex, Life Technologies) was incubated for an additional 60min at room temperature. Following an additional wash, samples were viewed on the microarray chips using an Olympus BX51W epifluorescence microscope with a 4X objective and imaged with Olympus software. All image acquisition

parameters were the same for the same fluorophore.

To quantify relative levels of fluorescent antibody-labeled Tau, GFAP, UCH-L1, and cfDNA for each sample, photographic images of each ACE-chip were imported to ImageJ ("FIJI"; National Institutes of Health). A circle was drawn around each of eight electrodes, and pixels measured. Background subtracted was the minimum number of pixels measured for each electrode, and averages and standard deviations were calculated. Direct 3D representations of the images were created using the "3D interactive viewer" plug-in for ImageJ.

Figure 2.1B shows representative examples of plasma sample analyses from the study cohort for the relative abundance of UCH-L1 and Tau.

2.2.4 Statistical Analysis

Models were used to predict the severity of injury with the probability of intracranial abnormality post TBI. The probability threshold was chosen as that which minimizes the Euclidean distance from point (0.1), or the upper-left corner, on the receiver operating characteristic (ROC) curve. ROC curves were calculated to determine the area under the curve (ROC-AUC) for different biomarkers. Based on the ROC-AUC, the optimal rIF values for discriminating minor-TBI plasma from non-TBI plasma were calculated for different biomarkers.

Model predictions were compared to observed diagnoses and performance metrics were calculated, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the area under the ROC curve (AUC). Heat maps

were constructed and Pearson correlation coefficients were calculated between GFAP, UCH-L1, Tau, cfDNA, and cumulative RPCSQ score. Correlation analysis between GFAP, UCH-L1, Tau, cfDNA, and individual RPCSQ symptoms was also performed. All analyses were carried out using open-source statistical analysis software (R version 3.5.0).

2.3 Results

2.3.1 Demographics and Clinical Course of the Study Cohort

The study enrolled 27 minor TBI subjects and 6 non-TBI subjects between December of 2015 and March of 2016. The demographic of the study population is as indicated in Table 2.1. The mean age of the minor-TBI and non-TBI cohort was 58.5 ± 16.4 and 34.6 ± 10.6 years, respectively. The male to female ratio were approximately 1.7:1 and 1:1 for the TBI and non-TBI cohort, respectively. All patients in the minor-TBI cohort underwent head CT as a part of their care. Except for the two patients (2/27 or 7%) who showed evidence of contusion on head CT (Figure 2.2A), all patients had negative head CTs and were discharged from the Emergency Ward on the day of the presentation. The two patients with abnormal head CT were admitted to the hospital for over-night observation. Both underwent interval surveillance imaging demonstrating stability of CT finding before discharge on the following day. The non-TBI patients presented with chest or abdominal discomfort or lower extremity pain. None of the patients in the non-TBI cohort underwent head CT. Diagnostic work-up were unremarkable and the patients were discharged on the day of intervention.

2.3.2 Biomaker Comparison Between Subjects With and Without Abnormal Head CT

Plasma samples from the study cohort were analyzed for the relative abundance of GFAP, Tau, UCH-L1, and cfDNA using the ACE microarray and on-chip immunofluorescence (IF) analysis. Relative immunofluorescence (rIF) level was determined for each sample. We first posed the question of whether any of the biomarkers were elevated in the two minor-TBI patients with abnormal head CT (subject 5 and subject 8) relative to all remaining patients. We performed this analysis in a qualitative manner since the sample size is too small for meaningful quantitative assessment. Consistent with the published utility of GFAP and UCH-L1 [134], plasma from subject eight showed significantly elevated levels of both proteins. In fact, this subject harbored the highest level of both GFAP and UCH-L1 for all study cohorts. Plasma from subject 5 also showed significantly elevated UCH-L1 (Figure 2.2B).

Notably, plasma from both subjects 5 and 8 also harbored significantly elevated Tau levels. These observations suggest the utility of plasma Tau as biomarker for structural brain injury after TBI. In contrast, the plasma level of cfDNA in subjects 5 and 8 were not significantly elevated relative to other study cohorts.

As a proof-of-principle study, these results support an association between elevated plasma GFAP, UCH-L1, and structural TBI demonstrated on head CT.

2.3.3 Biomarker Comparison Between Mild-TBI and Non-TBI Subjects

We wished to determine whether any of the biomarkers studied could discriminate the plasma collected from minor-TBI patients relative to non-TBI patients without abnormal head CT. We were only able to complete this analysis for plasma GFAP and Tau. We additionally excluded subjects 5 and 8 from this analysis since we were interested to characterize the value of these biomarkers in head CT negative TBI patients. Receiver operator characteristic curve were calculated to determine the area under the curve (ROC-AUC) for GFAP and Tau. Based on the ROC-AUC, the optimal rIF values for discriminating minor-TBI plasma from non-TBI plasma was 1.75 for Tau and 1.35 for GFAP (Table 2.2). Using these cut-offs, the sensitivity and specificity of discriminating plasma between minor-TBI and non-TBI patients was calculated and is shown in Table 2.3. In this analysis, Tau and GFAP performed similarly, with sensitivity of 72 and 64%, respectively, and specificity of 50% for both proteins. The combinations of GFAP and Tau did not significantly improve the sensitivity or specificity relative to the individual biomarker (Table 2.4).

2.3.4 Biomarker Comparison Between Subjects With and Without Post-concussive Symptoms

We next determined whether the presence of post-concussive symptoms is associated with elevated plasma GFAP, UCH-L1, Tau, or cfDNA. To this end, the Rivermead Post Concussion Symptoms Questionnaire (RPCSQ), a validated instrument for assessment of post-concussive symptoms following mild TBI [124, 125], was administered to study subjects by a trained study coordinator at 1-month follow-up. RPCSQ score was obtained for all 27 minor-TBI subjects. The highest cumulative

RPCSQ scores were observed in the two patients with CT imaging abnormalities. No significant correlation was observed between the cumulative RPCSQ and cfDNA. However, we observed significant, positive correlation between plasma GFAP, UCHL1 and Tau, and cumulative RPCSQ (Figure 2.3A). Specifically, the higher RPCSQ scores were associated with higher plasma biomarkers. The Pearson correlation between GFAP, UCH-L1, and Tau and cumulative RPCSQ were 0.68, 0.79, and 0.81, respectively (all $p < 0.01$). In a correlation matrix analysis, we found that plasma GFAP and UCH-L1 levels were highly correlated ($r = 0.95$, $p<0.001$) (Figure 2.3B). These results suggest plasma GFAP, UCH-L1, and Tau may be useful as predictive biomarker of post-concussive syndrome.

Pertaining to each of the individual items on the RPCSQ, Plasma UCH-L1, and GFAP exhibited the highest correlation to sensitivity to noise and light ($r = 0.96$ and 0.91, respectively, both $p < 0.001$). Plasma UCH-L1 and Tau showed highest correlation with headache ($r = 0.74$ and 0.78, respectively, both $p < 0.001$), sleep disturbance ($r = 0.69$ and 0.84, respectively, both $p <0.001$), and cognitive symptoms, including forgetfulness ($r = 0.76$ and 0.74, respectively, both $p < 0.001$), poor concentration ($r = 0.68$ and 0.76, respectively, both $p < 0.001$), and time required for information processing ($r = 0.77$ and 0.81, respectively, both $p<0.001$). In contrast, cfDNA exhibited a strong correlation with depression ($r = 0.79$, $p < 0.01$) and dizziness ($r = 0.69$, $p < 0.01$) (Figure 2.4).

2.4 Discussion

In this proof-of-principle study, we determined whether plasma isolated from

minor TBI patients can be analyzed using an ACE microarray platform (28). Because this platform had not been previously tested in this context, we selected TBI biomarkers with substantial literature supporting their utility in the study of TBI, including GFAP, UCH-L1, and Tau. We additionally tested whether cfDNA may serve as a useful biomarker. Our pilot data with the ACE-microarray platform support the utility of GFAP, UCH-L1, and Tau as plasma biomarker for TBI. Despite inherent constraints associated with the limited sample size, our pilot data are largely consistent with the previously published studies, including TRACK-TBI studies [134, 135]. As such, we believe our data support ACE-microarray as a platform for blood-based biomarker study in patients suffering from minor TBI.

In comparison to the currently available methods of plasma analysis, the ACE micro-array platform presents several major advantages. First, ~25 μl of plasma is required, in comparison to other analytic platforms that require larger volumes. It is important to note that the entirety of this study was performed using blood left-over from standard chemistry tests. As such, if validated, the ACE micro-array platform may be added to the standard chemistry set without additional blood collection. Additionally, the ACE-microarray platform minimizes the number of steps in terms of sample transfer, and thereby reduces the risk for cross-contamination or sample mix-up. The only sample transfer step in the ACE microarray platform was loading of the plasma onto the chip. In contrast, serial dilution of samples is typically required for sandwich enzyme-linked immunosorbent assays (ELISA). Finally, the chip can be subjected to multiplex immunofluorescent study to simultaneously assess biomarkers beyond Tau, GFAP, and UCH-L1.

A particularly intriguing result in this study involves the correlation between sub-domains of the RPCSQ and selected plasma biomarkers. Our results indicate that plasma UCH-L1, GFAP, Tau, and cfDNA levels correlated with different symptoms in the physical RPCSQ cluster domain. As a pilot analysis, the results should not be considered without scrutiny. For instance, meaningful quantitative assessment of this correlation is possible only in the context of the prevalence of the symptoms in the minor TBI population. Nevertheless, if the correlation reported here is validated in a future study, these results would suggest that the different aspects of post-concussive syndrome arise from pathophysiologic processes that ultimately lead to the release of the distinct biomarkers. A corollary of this hypothesis would suggest that medication that mitigates select patho-physiologic processes may be helpful to prevent or arrest post-concussive symptoms that compromise the patient's quality of life.

Admittedly, the predictive utility of the platform is suboptimal given the data presented. Improvement in study design is warranted for future studies. For instance, our study did not factor into consideration factors that affect serum concentration of GFAP, Tau, and UCH-L1, including extracranial injuries, neurological co-morbidities, and pre-injury functional status. Total Tau was tested in this study, which represents another limitation in this pilot analysis. Use of more specific and improved biomarkers like hyperphosphorylated Tau can improve the outcome prediction of TBI.

Additionally, given variability in the half-lives of these biomarkers, collection within 72 h of blood draw may present systematic bias in the correlative analyses. The reporting of RPCSQ at the time of follow-up is subject to recall bias. These and other potential confounding factors need to be considered in the design of future studies.

While this proof-of-principle study is, by definition, limited in its sample size and pilot in nature, the general demographic of the study population largely mirrors those of larger series, including the demographic of the study population and the proportion of minor TBI patients with abnormal head CT [136]. The recapitulation of the association of GFAP, UCH-L1, and Tau with various aspects of TBI reported elsewhere is also reassuring [137]. That said, we caution against definitive conclusions from this study beyond feasibility of ACE microarray as a potential biomarker platform. Studies reporting test-retest reliability of this assay are warranted in future. There are many steps ahead in clinical translation, including direct comparison of results derived from ACE-microarray platform with other established assays, such as the Quanterix assay [138]. Prospective age and gender matched cohorts [135, 139] as well as thoughtful consideration of extracranial injuries, neurological comorbidities, and pre-injury functional status will be needed in advancing this biomarker platform toward clinical application.

2.5 Acknowledgement

This work was supported by a DOD grant, award number W81XWH-14-2-0192 and a National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP) award under Grant number DGE-1650112.

Chapter 2, in full, is a reprint of the material as it appears in Frontiers of Neurology, 2020. Lewis JM, Dhawan S, Obirieze AC, Sarno B, Akers J, Heller MJ, Chen CC. Plasma Biomarker for Post-Concussive Syndrome: A Pilot Study using an Alternating Current Electro-kinetic Platform. *Front. Neurol.*, 14 July 2020. The dissertation author is the secondary

investigator and author of the material. The dissertation author ran ACE on-chip extracellular vesicle isolation assays and immunofluorescent staining.

Table 2. 1. Summary statistics of demographics of the study population with comparative analysis in minor traumatic brain injury (TBI) and non-TBI patients

Variable	**Minor TBI cohort (*n* = 27)**	**Non- TBI cohort (*n* = 6)**	***p*-value**
Gender, *n* (%)			0.65
Male	17 (63)	3 (50)	
Female	10 (37)	3 (50)	
Age (years, mean ± SD)	58.5 ± 16.4	34.6 ± 10.6	0.002
Presentation: *n*	Fall: 16	Abdominal pain: 4	
	Syncope: 5	Chest pain: 1	
	MVA: 4	Lower extremity pain: 1	
	Others; 2		

MVA, motor vehicle accident; SD, standard deviation. Others include head injury and scalp laceration

Table 2. 2. Performance metrics of GFAP and Tau biomarkers

Area under curve (AUC) and cut-off relative immune fluorescence (rIF) value for GFAP and Tau		
Biomaker	**AUC**	**Cut-off rIF**
Tau	0.54 (0.24-0.83)	1.75
GFAP	0.48 (0.16-0.79)	1.36

Table 2. 3. Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) for GFAP and Tau

Biomarker	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Tau	72% (0.54-0.89)	50% (0.10-0.90)	85.7% (0.70-1.00)	30% (0.01-0.58)
GFAP	64% (0.45-0.82)	50% (0.10-0.90)	84.2% (0.67-1.00)	25% (0.01-0.49)

Table 2. 4. Comparison of Tau/GFAP combination receiver operating characteristic (ROC) curve with Tau-ROC and GFAP-ROC

With	***z*-value**	***p*-value**
ROC-Tau	0.25	0.80
ROC-GFAP	-0.23	0.81

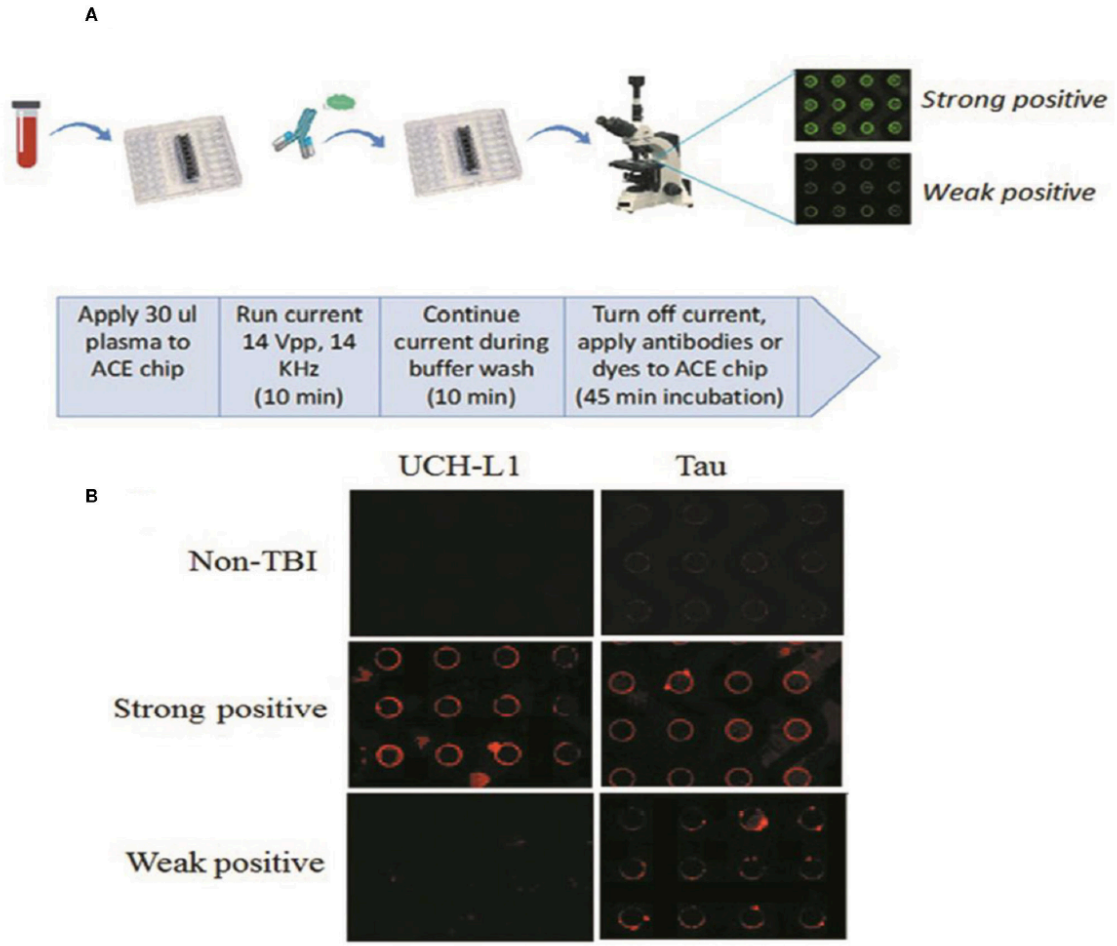


Figure 2. 1. (A) Overview of ACE microchip immunoassay workflow. Shown is a fluidics cartridge containing the 14 × 52 mm-sized ACE chip configured with eight sample chambers. When current is applied, extracellular vesicles (EVs), and other nanoparticles are drawn to the edges of the circular electrodes; the buffer wash serves to remove larger unbound cellular debris and smaller soluble plasma components from the chip. The current is then turned off, mixtures of fluorescent antibodies or dyes selective for each biomarker are added, and the concentration of biomarker proteins around the electrodes can be visualized as circular patterns of fluorescence. Image analysis provides a quantitative comparison of fluorescence intensities. (B) Representative examples showing relative abundance of UCH-L1 and Tau using the ACE microarray and on-chip immune-fluorescence (IF) analysis. ACE, alternating current electrokinetic. UCH-L1, ubiquitin C-terminal hydrolase L.

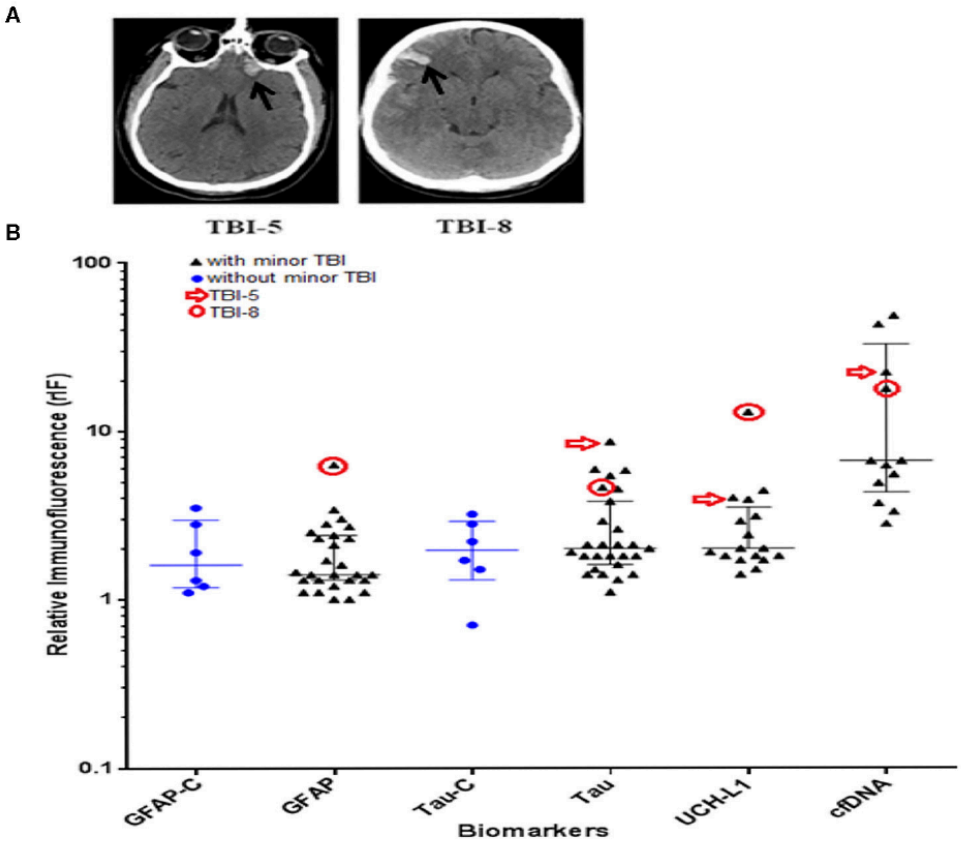


Figure 2. 2. (A) CT scan images showing contusion in patient #5 and #8 with minor traumatic brain injury (TBI) (black arrow). (B) Scattered dot plots illustrating distribution of relative immunofluorescence (rIF) (median with inter-quartile range) values for GFAP, Tau, UCH-L1, and cfDNA in TBI cohort, and GFAP and Tau in non-TBI cohort (represented as GFAP-C and Tau-C, respectively). Blue dots represent the rIF values for Tau and GFAP in patients with no history of head trauma. Black triangles represent the rIF values for Tau, GFAP, UCH-L1, and cfDNA in patients with history of head trauma. rIF values for patient #5 are denoted by red arrow; rIF values for patient #8 are encircled in red. GFAP, glial fibrillary acidic protein; UCH-L1, ubiquitin C-terminal hydrolase L1; cfDNA, cell-free DNA.

A

	Rivermead Cumulative Scores		
Biomarker	r	R^2 (adjusted)	*P* value
GFAP	0.68	0.18	0.01‡
Tau	0.81	0.63	< 0.001‡
UCHL1	0.79	0.59	< 0.001‡
cfDNA	0.07	-0.08	0.81

B

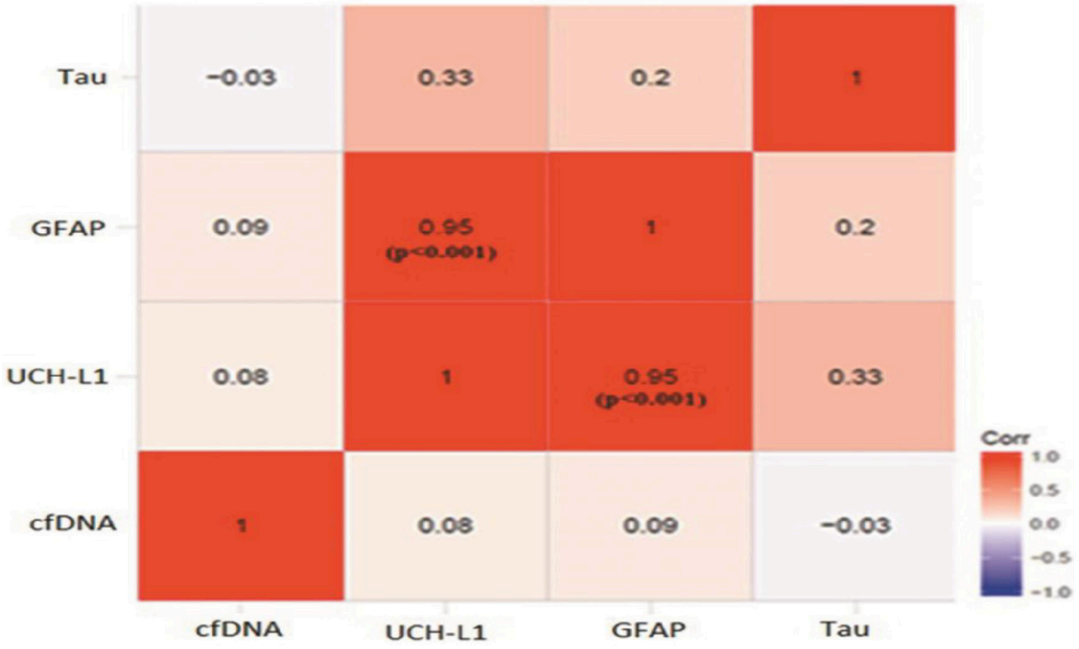


Figure 2. 3. (A) Correlation between different biomarkers and cumulative Rivermead symptom scores; $p \leq 0.01$ (B) Heat-map demonstrating the correlation between different biomarkers. Pearson correlation coefficient, r is mentioned in each box.

Biomarker / Symptom	UCH-L1	Tau	GFAP	cfDNA
Headache	**0.74**	**0.78**	**0.65**	-0.05
Nausea/Vomiting	0.15	**0.69**	0.14	0.32
Dizziness	-0.23	-0.43	-0.22	**0.69**
Noise sensitivity	**0.96**	0.12	**0.91**	0
Light sensitivity	**0.96**	0.12	**0.91**	0
Blurred vision	-	-	-	-
Double vision	-	-	-	-
Sleep disturbance	**0.69**	**0.84**	0.53	-0.05
Fatigue	**0.62**	**0.70**	0.41	-0.03
Irritable	**0.57**	**0.88**	0.48	-0.09
Depressed	0.45	0.21	0.35	**0.79**
Frustration	-	-	-	-
Restlessness	**0.77**	**0.82**	**0.68**	0.02
Forgetful	**0.76**	**0.74**	**0.69**	-0.11
Time for information processing	**0.77**	**0.81**	**0.63**	-0.01
Poor concentration	**0.68**	**0.76**	0.57	-0.13

Figure 2. 4. Correlation between different traumatic brain injury (TBI) biomarkers and Rivermead Questionnaire symptoms. Statistically significant ($p < 0.05$) Pearson co-efficient (r) is mentioned in bold format in box; - represents symptoms that were not recorded in minor TBI patients.

Chapter 3

Probabilistic Classification for Digital High Resolution Melt Curve

3.1 Abstract

High resolution melting (HRM) is a DNA analysis technique traditionally used to confirm specific amplification or detect a mutation in a PCR product. The potential of HRM to go beyond these functions and serve as a broad-based genotyping tool is gaining considerable popularity with the advances and availability of machine learning (ML) tools. Currently applied ML approaches are capable of matching HRM curves from a test sample to a database of HRM curves generated from known sequences. However, these algorithms do not allow for the assessment of the goodness of the classification. Herein, we develop probabilistic classification models of digital HRM curves to accomplish interpretable classification. Our algorithms accomplish an overall classification accuracy of 93% and an average area under the receiver operating characteristic curve (AUC) of 0.99, lending validity as a reliable HRM analysis tool.

3.2 Introduction

High Resolution Melting (HRM) analysis is a quick, inexpensive, and powerful post-amplification nucleic acid characterization technique that is increasingly being used to profile DNA sequences for research and clinical diagnostic applications [87, 140-143]. To accomplish HRM analysis, a DNA-binding dye is added to a sample of DNA, where it fluoresces upon intercalating into the double-stranded structure. Then the sample is heated. As the temperature increases, the double-stranded DNA denatures into single strands, releasing the intercalating dye and losing fluorescence. This loss in fluorescence with heating is recorded as a function of temperature, generating a melt curve.

Traditionally, the melting curves were matched to DNA sequence by either the melting point (Tm) or the difference in curve created by plotting the difference between a test curve and a reference curve [89, 144]. However, both approaches have their limitations. A single melt point Tm does not provide sufficient information to discriminate more than a few species. The difference curve, on the other hand, typically relies on temperature shifting and visual inspection at user's discretion to account for run-to-run or well-to-well variations. In addition, as melt curve reference library increases, the need for an automated curve matching method becomes inevitable. To overcome these limitations, recent studies have reported the use of machine learning algorithm for high resolution melt curve classification [145, 146]. Automation and incorporation of well-to-well and run-to-run variations within machine learning is imperative with digital HRM, where digital melt curves are generated simultaneously across many reactions. We

have previously demonstrated optimization of our universal digital high resolution melt (U-dHRM) platform to minimize run-to-run and melt-to-melt variation, for reliable and repeatable melt curve generation, eliminating the need for in-run template standard [99].

Machine learning approaches that have been employed include Naïve Bayes [147], Support Vector Machine (SVM) [145], k-Nearest Neighbor and Dynamic Time Warping [148]. Among these methods, SVM-based approaches have been more widely used and have been reported to be more robust [145], likely because of their ability to perform well on small datasets. Melt curve pattern recognition with SVM entails finding the best n-1 hyperplane in an n-dimensional space that maximizes the margin between the classes in the data. Even with small amounts of representative melt curve data to 'learn' from, SVM algorithms show excellent performance. However, with the formulation of the one-versus-one SVM (OVOSVM) that is being applied for HRM analysis, it is not possible to assess the goodness of the classification. This limitation can be overcome by probability classification methods such as multilayer perceptron (MLP), Gaussian Naïve Bayes, and multinomial logistic regression (LR). For its simplicity and ease of interpretation, we chose to investigate the utility of LR algorithms for HRM curve classification using a large dataset of melt curves generated from our U-dHRM platform.

Using neonatal (newborn) bacteremia as a test case, we evaluate the performance of multiclass LR models in classifying bacteria-derived DNA melt curves. The classes include ten bacterial species that comprise the majority of causative organisms for neonatal bacteremia. Universal primers are used to amplify hypervariable regions of the 16s rDNA gene in our dHRM system. This enables broad-based amplification of bacteria, while relying on melt curve and ML to specify the organism, and offers the ability to

detect individual organisms via the ‘digital’ design [98, 99].

3.3 Materials and Methods

3.3.1 Bacterial Strains

Table 3.1 lists the bacterial species included in the present study. These organisms make up the majority of the causative pathogens implicated in neonatal sepsis [149, 150]. The organisms were received as a generous gift from Dr. David Pride (University of California San Diego School of Medicine) or purchased from the American Tissue Culture Collection (ATCC, Old Town Manassas, VA). Bacteria were cultured in Lurie-Bertani (LB) broth or Tryptic Soy broth (TSB), as required, and incubated overnight at 37ºC.

3.3.2 Bacterial Genomic DNA Extraction and PCR

Following overnight culture, bacterial genomic DNA was extracted using Wizard Genome DNA Purification kit (Promega Corporation, Madison, WI). Spectrophotometric absorbance measurements were used to assess the quality and concentration of the extracted DNA, and sequencing was conducted to further confirm the identity of the species. Genomic DNA dilutions were prepared for use with dPCR. Commercially available QuantStudio 3D Digital PCR 20K chip v2 (Applied Biosystems, Foster City, CA) were used for amplification following the manufacturer’s recommended process, with the exception of reagents. As described previously, the dPCR master mix was

optimized for the recommended loading volume of 14.5 μL per reaction, and contained 1X Phusion HF Buffer containing 1.5 mM MgCl2 (Thermo Fisher Scientific, Waltham, MA), 0.15 μM forward primer 5′-GYGGCGNACGGGTGAGTAA-3′ (Integrated DNA Technologies, Coralville, IA), 0.15 μM reverse primer 5′-AGCTGACGACANCCATGCA-3′ (Integrated DNA Technologies, Coralville, IA), 0.2 mM dNTPs (Invitrogen, Carlsbad, CA), 2.5X EvaGreen (Biotium, Freemont, CA), 2X ROX (Thermo Fisher Scientific, Waltham, MA), 0.02 U/μL of Phusion HotStart Polymerase (Thermo Fisher Scientific, Waltham, MA), 1 μL of sample, and ultrapure PCR water (Quality Biological Inc., Gaithersburg, MD) to bring the total volume to 14.5 μL. To load the chip, a master mix reaction volume of 14.5μL was spread across following manufacturer's recommendation. After loading, the dPCR chip was cycled on a flatbed thermocycler with the following cycle settings: an initial enzyme activation (98 °C, 30 s), followed by 70 cycles (95 °C, 30 s, 59 °C, 30 s, 72 °C, 60 s) [98].

3.3.3 DNA Melt Curve Generation and Preprocessing

The U-dHRM device developed by our group has been previously described [98, 99]. Briefly, it comprises a copper plate on which the microfluidic dPCR chip is placed, thermoelectric heater/cooler (TE Technology Inc., Traverse City, MI), proportional-integral-derivative (PID) controller (Meerstetter Engineering GmbH, Rubigen, Switzerland), Class 1/3B resistance temperature detector (RTD) (Heraeus, Hanau, Germany) embedded in the copper block, K-type thermocouple (OMEGA Engineering, Stamford, CT), and heat sink. A thin layer of thermal grease added between the dPCR chip and copper block ensures efficient heat transfer. A custom-made adapter secures the

device on-stage for optimal fluorescent imaging. With heat ramping, simultaneous fluorescent images from the DNA-intercalating dye, EvaGreen (Ex/Em: 488 nm/561 nm) and the control dye, ROX (Ex/Em: 405 nm/488 nm) are acquired with a Nikon Eclipse Ti microscope (Nikon, Tokyo, Japan). Melt curves are generated by implementing an automated imaging processing algorithm in MATLAB. Although temperature measurement offset could occur from chip to chip, the linearity of the slope is maintained, allowing us to correct for such shift by using temperature calibrator sequences [99]. In the absence of calibrator sequences, we aligned all curves of the same organism by their peaks, adding jitter with expected standard deviation given that well to well variations follow a characteristic distribution. We used standard deviation of 0.1°C as previously reported [99]. All melt curves were normalized using area under the curve normalization. As an added preprocessing step to ensure high melt curve quality, melt curves on each chip were compared to a representative melt curve (which is an average of a sample of melt curves) on that chip. The similarity of each melt curve to the representative melt curve was assessed using Pearson correlation coefficient. For the purpose of the present study we included melt curves with a Pearson correlation coefficient of >0.90. Our choice of 0.9 cutoff was based on the rule of thumb for interpreting the size of a correlation coefficient, which classifies 0.9 to 1.0 as the highest correlation [151].

3.3.4 Logistic Regression Model Building

Logistic regression, a non-linear transformation of linear regression, is a probabilistic classification method used to classify a given input vector into one of two (binary logistic regression) or, by extension, more (multinomial/multiclass logistic

regression) classes. By using a set of observations with known classification, LR derives coefficients that are then used in the calculation of the predicted probability. For binary logistic regression, the equation is given as:

$$P(y_i = 1|X) = \frac{1}{1+ e^{-(\beta_o + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n)}} = \frac{1}{1+ e^{-(\beta_o + \Sigma \beta_i x_i)}} \quad (3.1)$$

where P ($y_i = 1|X$) is the conditional probability of the ith observation belonging to one of two classes, β_o is a constant term, β_i are derived regression coefficients, and x_i terms are the attributes (features) used in the classification. For the multinomial (multiclass) extension of binary LR, two approaches can be taken. First, the one-versus-rest LR model, in which a separate model is created for each class to predict the probability of an observation being in that class or the rest of the classes combined. This results in a set of independent binary LR models for all classes. Second, the multinomial model, in which the probability of an observation being in any class is estimated. With this approach, the equation is given as:

$$P(y_i = k|X) = \frac{e^{\beta_k x_i}}{\sum_{j=1}^{K} e^{\beta_j x_i}} \quad (3.2)$$

where P ($y_i = k|X$) is the probability that the ith observation belongs to the k class, and K is the total number of classes. With both approaches the test observation is then assigned the class to which it has the highest predicted probability. Although slightly more computationally costly, the predicted probabilities with multinomial LR model have been shown to be more reliable compared to the predicted probabilities with the one-versus-rest LR [152]. Both LR modeling approaches were used in the current study.

The classes for the present study include ten bacterial organism species. Attributes (features) of observations used in the model include negative first derivative of DNA melting fluorescence data obtained at given temperature points (range: 51.1 ºC – 92 ºC, interval: 0.1

ºC). Data pre-processing included standardization of features. To address the imbalance of classes in our data, we applied weights to the classes, that are inversely proportional to the class frequency [152]. Thus:

$$w_k = \frac{n}{n_k K} \tag{3.3}$$

where w_k is the weight applied to a class k, n is the total number of observations, n_k is the number of observations in class k, and K is the total number of classes. Applying class weights to Eq. (3.2) gives:

$$P(\mathrm{y_i} = k|X) = \frac{w_k e^{\beta_k x_i}}{\sum_{j=1}^{K} w_j e^{\beta_j x_i}} \tag{3.4}$$

3.3.5 Hyperparameter Tuning and Model Selection

In order to avoid overfitting, and to optimize the LR model hyperparameters, we evaluated the best values for the algorithm hyperparameters using cross-validation. The data was randomly split (65%:35%) into training set and test set using unique dPCR chip identifiers, i.e., all melt curves from a particular chip were either present in the training set or in the test set, but not both. Stratified 5-fold cross-validation was conducted using the training set. Hyperparameter values that generalize well across all folds were identified, and then used to evaluate the model performance on the held-out test set. Model performance was assessed using accuracy and area under the receiver operating characteristics (ROC) curve (AUC). The model with the best performance was then selected.

3.3.6 Model Performance Evaluation

Following hyperparameter optimization and model selection, we evaluated the performance of the model using the entire dataset. We used leave-one-group-out cross-validation method, with the unique chip identifiers as the group IDs. This ensures that each individual melt curve was in the test set only once, i.e., during the cross-validation fold in which the chip containing that melt curve is held out as a test set. Model performance evaluation metrics used include accuracy, precision, recall (or sensitivity), specificity, F1 score and AUC. All algorithms were built and implemented using the scikit-learn package within Python programming language [153].

3.4 Results

3.4.1 Dataset

We developed our classification algorithms using a total of 18,893 melt curves in our database, generated from 10 bacterial species (Table 3.1). These bacteria species were chosen because of their clinical relevance to neonatal bacteremia [149, 150]. Figure 3.1 shows example melt curves for each of the ten bacteria species. For each species, the average melt curve (of all the melt curves for that species) is shown in Figure 3.2.

3.4.2 Multiclass Classification

We tested the utility of a probabilistic classification LR model for the identification and discrimination of bacteria species based on bacteria DNA melt curves. The quality of a melt curve from a given chip was assessed by comparing the melt curve to a representative curve on that chip using Pearson correlation coefficient. We included melt curves with

correlation coefficient of >0.90. To optimize the parameters of our LR models we used the grid search and the random search approaches with 5-fold cross-validation, trained on two-third and validated on one-third of melt curves. To handle imbalanced classes, we set the class weight parameter in the model to 'balanced', which adjusts class weights to be inversely proportional to class frequencies in the input data (Eqs. (3.3) and (3.4)). We evaluated the performance of our LR classification model using leave-one-group-out cross-validation, by which we train the model on melt curves from all but one chip, and then test the model on melt curves from the left-out chip. This approach was to ensure that melt curves from any one chip were not concurrently present on both the training and test sets, as this would give rise to 'leaking' of information from the training set to the test set, resulting in erroneously high performance. Figure 3.3 shows the confusion matrix of our classification algorithm. Overall, accuracy, precision, recall and F1 score are 93.2%, 93.4%, 93.2% and 93.3%, respectively. At the class-specific level, *C. koseri* has the lowest accuracy of 85%, followed by *E. faecium* with a classification accuracy of 86%. The remainder of the classes have accuracy of 90% or greater, with the highest accuracy (99%) seen with Group B *Streptococcus* (GBS), *H. influenza* and *S. pneumonia*. Other performance metrics for each class is shown in Table 2. Sensitivity and specificity ranged from 85% to 99% and 99% to 100%, respectively. We also assessed the discriminating ability of our model using the ROC curve, and determined the area under the ROC curve (AUC) (Figure 3.4). The overall class sample size-weighted average AUC across all species is 0.99. Species-specific AUC range from 0.97 (*E. coli*) to 1.00 (*Strep. pneumonia*, *H influenza*, *S. gallolyticus*, and *L. monocytogenes*).

3.5 Discussion

Herein we have demonstrated the utility of a probabilistic classification algorithm for DNA profiling using melt curves generated from a dHRM platform. Our algorithm was able to discriminate among ten bacterial organism species with 93% overall classification accuracy, and an average AUC of 0.99. Bacterial species-specific classification accuracy ranged from 85% to 99%, while specificity was 99% or greater for the species. This work demonstrates the utility of a probabilistic classification algorithm in resolving multiple bacterial organism melt curves. The large amount of dPCR chip-generated melt curve data enabled the development of probabilistic classifier, which distinguishes this study from previous studies applying non-probabilistic methods to small datasets of curves [154, 155]. Moreover, our approach of limiting all melt curves from a single chip to either training set or test set, but not both, at any one time, ensures accurate assessment of the model performance. This leave-one-group-out methodological approach is critical in evaluating the predictive performance of machine learning tools when applied to melt curves generated across distinct runs. Testing prediction models on melt curves generated on the same chip as the melt curves that the model was trained on will result in falsely high performance, as some information about the test melt curve had already been learned by virtue of the same-chip melt curves present in the training set. The implication will be high generalizability error with future test melt curves

Currently, the most widely used classification algorithm for HRM and dHRM analysis is SVM [154], which is a non-probabilistic algorithm. There are several advantages of probabilistic classification methods over non-probabilistic classifiers. First, unlike probabilistic classification models, with non-probabilistic models, uncertainty in classification cannot be modeled. This is important in high-stake situations, such as clinical

disease diagnosis, where decisions rely strongly on the certainty of the classification [156]. As interest in the utilization of HRM technology in the clinical setting grows, probabilistic classification methods will likely also become increasingly necessary for these cost-sensitive situations, where the certainty of classification has to be sufficiently high for clinical management decisions to be made. Second, probabilistic classifiers can be more effectively combined with other classifiers within a large machine learning framework [157]. Although more advanced probabilistic classifiers could be explored, we chose LR classifier for its interpretability and simplicity.

In conclusion, advances in machine learning and 'big data' generation are opening up more opportunities for the advancement of HRM, where the speed, low cost and simplicity of HRM was already attractive. Probabilistic algorithms can serve as robust tools for HRM analysis, which will open up opportunities for the use of HRM technology as a discovery tool as well as profiling technology, and further advance HRM technology towards its application in research and clinical diagnostics.

3.6 Acknowledgement

This work was supported by a grant from the NIH/NIAID, award number R01AI134982, a Burroughs Wellcome Fund Career Award at the Scientific Interface award number 1012027, and a National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP) award under Grant number DGE-1650112.

Chapter 3, in part, is currently being revised for submission for publication of the material. Obirieze AC, Sinha M, Langouche L, Mack H, Leineweber W, Aralar A, Coleman

TP, Fraley SI. “A Probabilistic Approach to Melt Curve-Based DNA Profiling”. The dissertation author is the primary investigator and primary author of the material.

Table 3. 1. Training data generated per bacterial species.

Bacterial organism	Number of melt curves
Group B *Streptococcus*	4167
Escherichia coli	803
Haemophilus influenzae	2343
Listeria monocytogenes	2736
Methicillin-sensitive *Staphylococcus aureus* (MSSA)	1792
Citrobacter koseri	1410
Streptococcus pneumoniae	1079
Enterococcus faecium	1468
Streptococcus gallolyticus	983
Streptococcus sanguinis	2112

Table 3. 2. Multiclass logistic regression classification performance

Bacteria species	Recall/sensitivity	Specificity	Precision/PPV	F1 score
Group B *Streptococcus*	0.99	0.99	0.96	0.98
Escherichia coli	0.91	0.99	0.76	0.82
Haemophilus influenzae	0.99	0.99	0.96	0.98
Listeria monocytogenes	0.95	0.99	0.96	0.96
Staphylococcus aureus	0.90	1.00	0.99	0.94
Citrobacter koseri	0.85	0.99	0.92	0.89
Streptococcus pneumoniae	0.99	1.00	0.97	0.98
Enterococcus faecium	0.86	1.00	0.95	0.90
Streptococcus gallolyticus	0.92	1.00	0.96	0.94
Streptococcus sanguinis	0.95	0.99	0.91	0.93

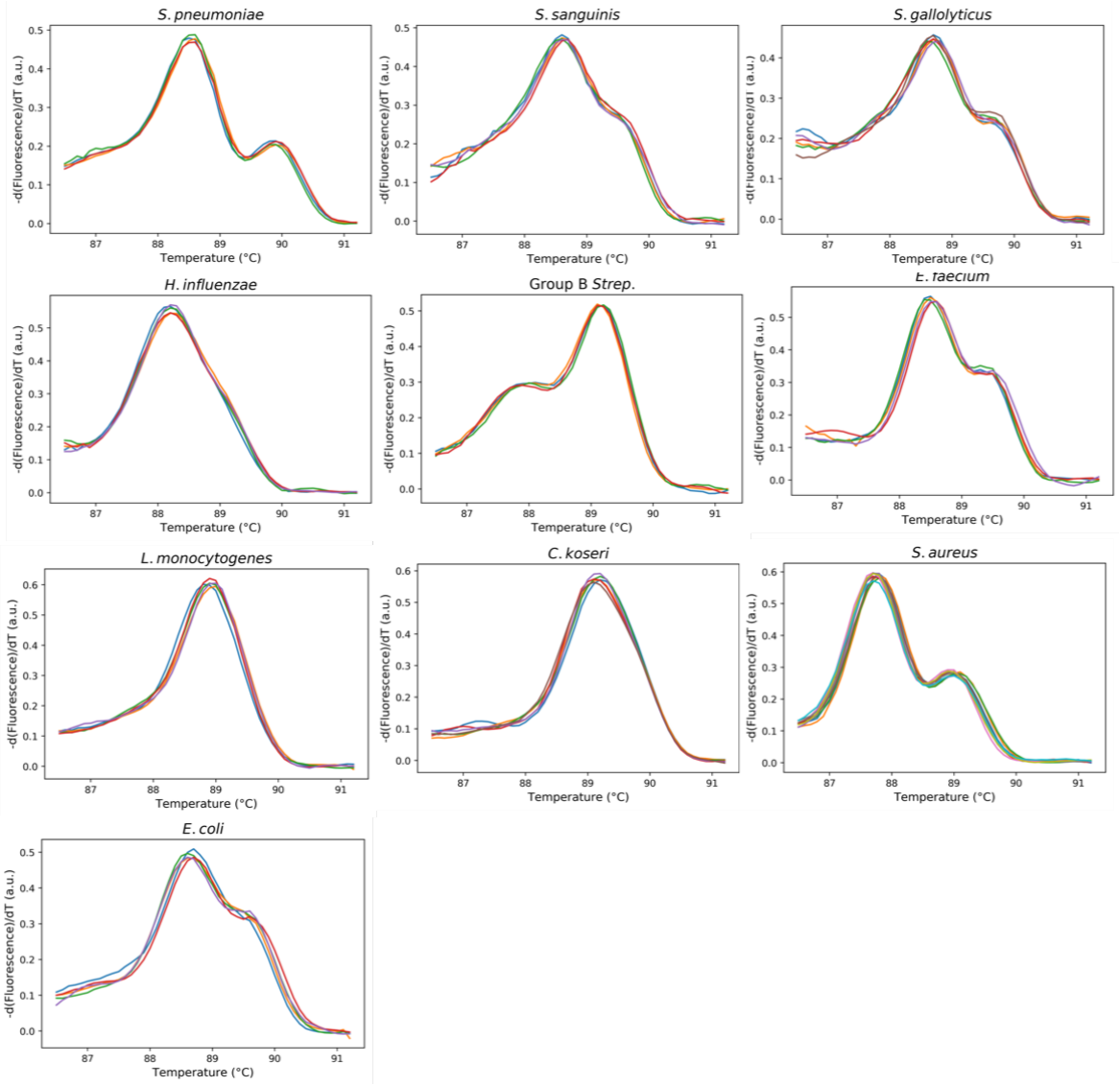


Figure 3. 1. Sample DNA melt curves generated from 10 bacteria species

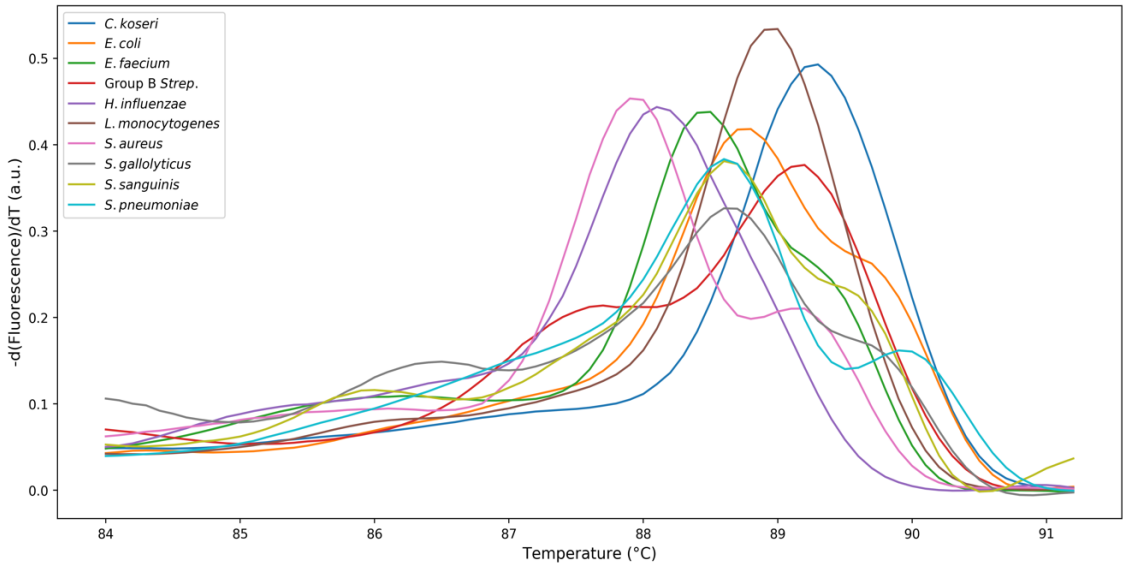


Figure 3. 2. Average melt curve for each of the 10 bacteria species

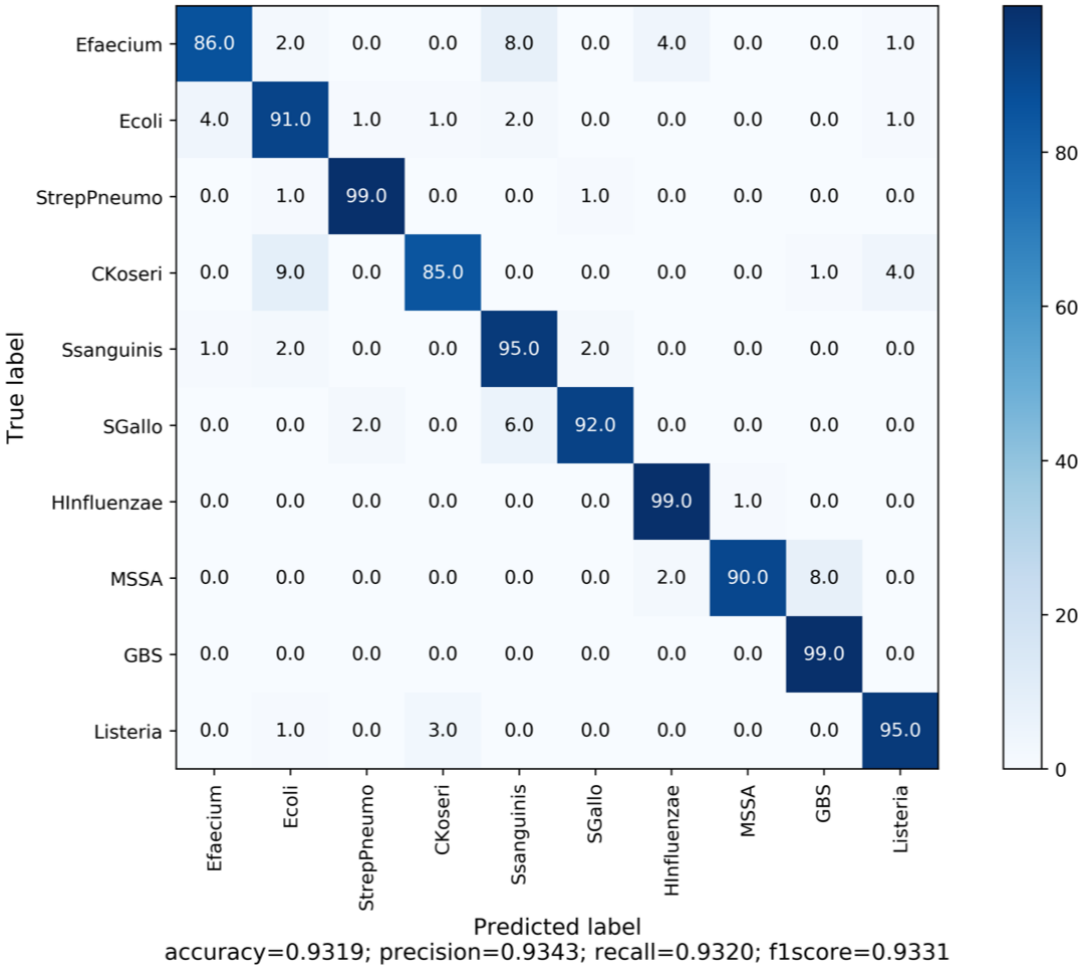


Figure 3. 3. Confusion matrix for multiclass logistic regression classification following leave-one-group-out cross-validation

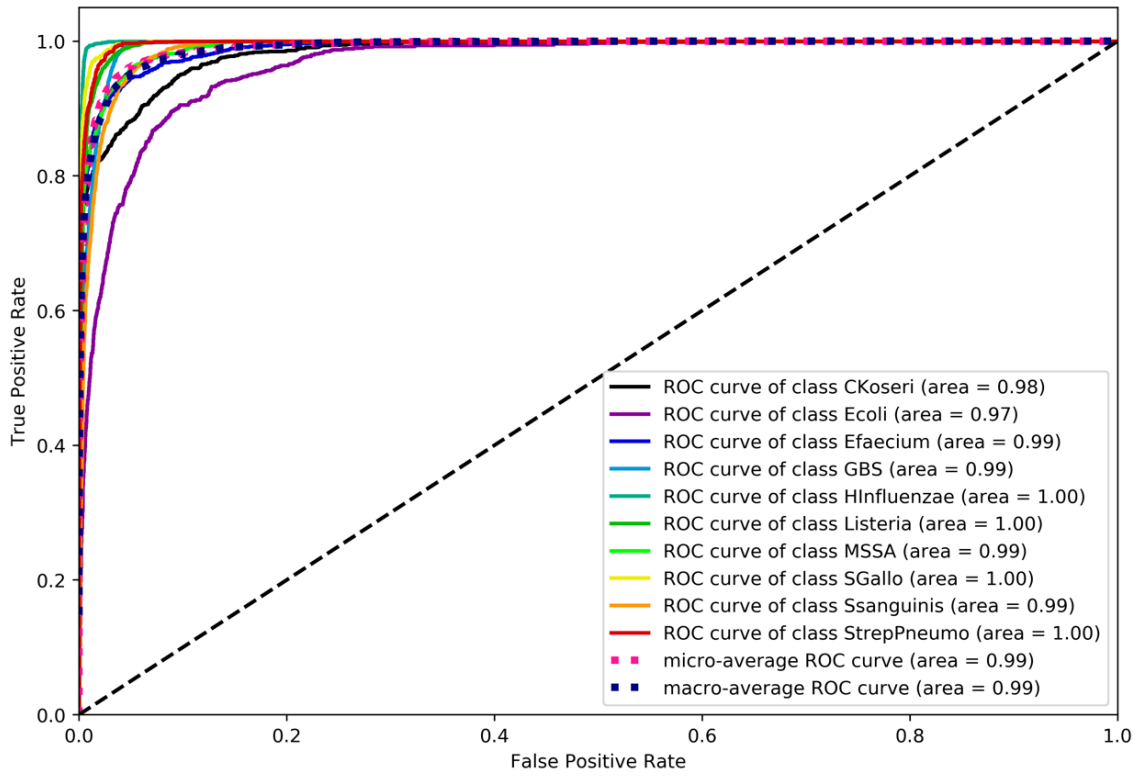


Figure 3. 4. Receiver operating characteristic (ROC) curves showing logistic regression model predictability of 10 bacterial species

Chapter 4

Novelty Detection for Digital High Resolution Melt Curve

4.1 Abstract

The rapid detection of novel genotypes remains a challenge for many basic research, clinical, and industrial applications. High resolution melting (HRM) is a DNA analysis technique with the potential to serve as a rapid and broad-based genotyping tool through the use of machine learning (ML). Current ML algorithms are capable of matching HRM curves from a test sample to a database of HRM curves generated from known sequences. However, they falsely classify melt curves of novel sequences that are not represented in the database. Here, we develop probabilistic models of sequence-defined HRM curves to accomplish novelty detection. This approach is uniquely enabled by the large melt curve datasets generated using our high-throughput digital HRM technique. As an example application, we demonstrate the potential of this approach to rapidly detect both known and novel microbial pathogens. Using a Shannon entropy-based method and a modified one-class detection experimental approach, our models achieve AUCs of 0.92 and greater for novelty detection.

Our results support the feasibility of HRM-based DNA profiling.

4.2 Introduction

Advances in high resolution melt (HRM) technology have taken this analysis from a simple check on amplification product homogeneity to a tool for mutation detection. Then, heat transfer and reaction engineering have improved HRM such that a melt curve could be used as a signature unique to the DNA sequence in question. Efforts to harness the potential of HRM as a broad-based sequence profiling tool have necessitated the use of machine learning for HRM curve analysis. Among the machine learning approaches that have been employed such as Naïve Bayes [147], Support Vector Machine (SVM) [145], k-Nearest Neighbor and Dynamic Time Warping [148], SVM-based algorithms have been most widely used.

SVM methods, which entails finding the best n-1 hyperplane in an n-dimensional space that maximizes the margin between the classes in the data, have been reported to be robust in HRM analysis. However, like other non-probabilistic ML methods, with the one-vs-one SVM (OVOSVM), a problem emerges when the melt curve to be predicted has no representation in the melt curve database. For example, with a database containing melt curves from known pathogens, the currently used OVOSVM algorithm will erroneously classify test melt curves from an emerging pathogen (that is not represented in the database) into a known pathogen class. For HRM to reach its potential as a broad-based profiling tool, there is the need for machine learning classification algorithms that enable the identification of such 'novel', emerging DNA melt curves that are not yet represented in the database.

Probabilistic classification methods are suitable for such tasks. By outputting classification probabilities in addition to class predictions, these methods allow for the assessment of the degree of uncertainty of the classification, which can then be used for novelty detection. One such novelty detection approach that employs the use of classification probabilities is the Shannon entropy method [158]. Entropy-based anomaly detection methods are widely used in other fields, computer network security in particular [159-161]. In information theory, Shannon entropy is a measure of the uncertainty associated with the probability distribution of a random variable [162]. Within the context of classification, Shannon entropy can be used to assess the level of confidence underlying a classification method's decision. The magnitude of the entropy can, therefore, be a reflection of the uncertainty of the classification. This can then be used to form the basis of confidence scoring for the goodness of classification, and allows for a novelty detection threshold to be set.

While there are methods for generating probabilities from SVM-based approaches [163-165], these approaches are computationally expensive and require large amounts of data. Moreover, even when used, the probabilities generated may not correlate with the classification results. Therefore, we chose to investigate the utility of more traditional probabilistic classification methods, such as logistic regression (LR), in novelty detection. In addition to the LR model parameters being easy to interpret, unlike SVM-based models (which have hinge loss that is generally non-differentiable), the differentiable LR loss function allows for the application of regularization methods [166, 167].

Our approach is to evaluate the utility of the LR algorithm for novelty detection. This approach requires large amount of melt curve data than previously reported. The generation of thousands of melt curves for our training datasets is powered by our digital high-resolution

melt (dHRM) platform, which uses a custom heat transfer and imaging system to reliably melt thousands of digital PCR (dPCR) reactions simultaneously. We focus on one application as a test case: the detection of known and novel bacterial pathogens relevant to the diagnosis of bacteremia in neonates. Although a small number of bacterial organism species are implicated in the majority of neonatal bacteremia cases, opportunistic infections and emerging pathogens can occur [149]. We demonstrate the ability to automatically detect both causative bacterial pathogens as well as rare organisms through their 'anomalous' melt curve signatures, lending further promise to the concept of using HRM as a profiling technology.

4.3 Materials and Methods

The bacteria strains, genomic DNA extraction, PCR, melt curve generation and processing are as described in the previous chapter. Briefly, bacteria species, that were either purchased or received as a gift, were cultured in Lurie-Bertani (LB) broth or Tryptic Soy broth (TSB), as required, and incubated overnight at 37ºC. Following culture, bacteria genomic DNA was extracted using the Wizard Genome DNA Purification kit (Promega Corporation, Madison, WI). Digital PCR was performed using commercially available QuantStudio 3D Digital PCR 20K chip v2 (Applied Biosystems, Foster City, CA), following the manufacturer's recommended process, with the exception of reagents. DNA melt curves were generated from the U-dHRM device developed by our group as previously described [98, 99]. Processing of the melt curve data included curve alignment and normalization using area under the curve. Figure 4.1 shows the machine learning workflow.

4.3.1 Shannon Entropy-based Novelty Detection

We identified novel curves by evaluating the uncertainty of the predicted probabilities, using the concept of entropy. Our entropy-based measure of choice is Shannon entropy [158]. Shannon entropy is calculated by:

$$H\ (P) = -\sum_{k=1}^{K} p_k log(p_k) \quad (4.1)$$

where P is the probability distribution of a discrete random outcome variable, p_k is the probability mass function for outcome k using previously defined classification methods. For instance, using Eq. (3.4), we have K hypotheses and thus, given melt curve data X and coefficients beta that have already been fitted, we can calculate $p_1 \ldots p_K$. The magnitude of entropy depends on the randomness of the variable, with maximum entropy attained when the probability distribution is uniform. Entropy for each melt curve is calculated over the bacteria species classes. To assess the ability of our model to detect novel melt curves, i.e., melt curves belonging to organism classes not previously represented in our database, we carried out a leave-one-class-out experiment similar to an approach that has been previously reported [168]. Using melt curves from all ten bacterial species, representing ten classes, we held out melt curves for each of the ten classes in turn. Melt curves of the remaining nine classes were then randomly split into training and test sets (at a ratio of 80:20). The left-out class was then mixed with the test set to make up the new test set. The machine learning model was built with the training set following the steps described in the previous chapter. The model is then

evaluated on its ability to correctly identify the melt curves belonging to the left-out (novel) class from the mixture comprising the 20% split of the remaining nine classes and the withheld class. The left-out class was regarded as the positive class. This was done ten times, for the ten classes. Figure 4.2 presents a schematic of the novelty detection experimental approach. We assessed the performance of the novelty detection model using AUC. We also determined the optimal entropy threshold from the ROC curve using Youden's index (threshold at which [sensitivity + specificity -1] is maximum) [169]. We built and implemented all algorithms using the scikit-learn package within Python programming language [153].

4.4 Results

4.4.1 Novelty/Anomaly Detection

We developed a novelty detection algorithm based on the probabilistic classification models and the Shannon entropy method, for the detection of melt curves not represented in a melt curve database. An overview of the experimental approach taken for the Shannon entropy-based novelty detection is shown in Figure 4.2. With melt curves of one class of bacterial species left out (novel class) in turn, the melt curves for the remaining nine classes were randomly split (80:20) into training and test sets. The left-out class was then mixed with the test set to make up the final test set. This process was carried out ten times for all ten classes of bacteria species in the study. The novelty detection task was to use the entropy value derived from the model prediction probabilities to correctly identify melt curves belonging to the novel class. We evaluated the performance of our novelty detection model using AUC. Figure 4.3a shows box plots of Shannon entropy values for left-out (novel) class

and remainder (known) classes for one of the ten rounds. The ROC curve for that round is shown in Figure 4.3b. Table 4.1 shows AUCs for all ten rounds of novelty prediction. All rounds achieved novelty detection AUC of greater than 0.92. From the ROC curve of each round, we then determined the optimal entropy threshold using Youden's index [169], i.e., the cut-off that gives the maximum of (sensitivity + specificity – 1). Sensitivity and specificity at the optimal threshold is shown in Table 4.1. At optimal entropy threshold (determined by Youden's index), sensitivity and specificity range from 82% to 96% and 85% to 97%, respectively. It is possible to optimize the threshold depending on the goal at hand (i.e., threshold can be optimized for sensitivity, specificity, etc.).

4.5 Discussion

Current widely used ML methods for dHRM analysis [154] are unsatisfactory for melt curve novelty detection, i.e., detecting when test melt curves are not represented in the melt curve database. This work demonstrates the utility of a probabilistic classification algorithm in identifying previously unknown (novel) melt curves that are not represented in the database. Our novelty detection models achieved AUC in the range of 0.92 to 0.99. To our knowledge, this is the first report of the development of novelty detection methods for HRM analysis. The choice of a probabilistic algorithm approach to novelty detection has advantages. Unlike non-probabilistic classification models, probabilistic classification algorithms allow for the modeling of uncertainty in classification, which is critical in situations where decisions rely strongly on the degree of confidence in the classification.

Potential applications include emerging pathogen identification, and discovery

research. Emerging infectious diseases have been on the rise, and resulted in significant public health and economic burden [170]. With the increasing threats of epidemics and pandemics, and concerns about bioterrorism, HRM can serve as a simple, inexpensive, and powerful tool for early identification of these threats. Although there has been progress made in reducing the cost of next generation sequencing (NGS), it remains inaccessible to many clinical and research laboratories. In addition, it is time consuming and requires skilled personnel. As interest in the utilization of HRM technology in the clinical setting grows, advances in dHRM technology coupled with novelty detection frameworks such as demonstrated in the present study will bring HRM technology closer to the realization of its potential as a DNA profiling tool.

4.6 Acknowledgement

This work was supported by a grant from the NIH/NIAID, award number R01AI134982, a Burroughs Wellcome Fund Career Award at the Scientific Interface, award number 1012027, and a National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP) award under Grant number DGE-1650112.

Chapter 4, in part, is currently being revised for submission for publication of the material. Obirieze AC, Sinha M, Langouche L, Mack H, Leineweber W, Aralar A, Coleman TP, Fraley SI. “A Probabilistic Approach to Melt Curve-Based DNA Profiling”. The dissertation author is the primary investigator and primary author of the material.

Table 4. 1. Summary of novelty detection model performance

Left-out class	AUC (%)	Sensitivity at optimal entropy threshold (%)	Specificity at optimal entropy threshold (%)
Group B *Streptococcus*	95.3	95.4	95.1
Escherichia coli	95.5	91.5	86.6
Haemophilus influenzae	97.1	92.7	90.8
Listeria monocytogenes	97.2	89.2	93.7
Staphylococcus aureus	93.1	91.9	85.4
Citrobacter koseri	92.2	82.3	88.6
Streptococcus pneumoniae	95.8	86.2	92.9
Enterococcus faecium	99.1	93.8	96.8
Streptococcus gallolyticus	94.3	89.5	87.5
Streptococcus sanguinis	98.6	96.3	93.8

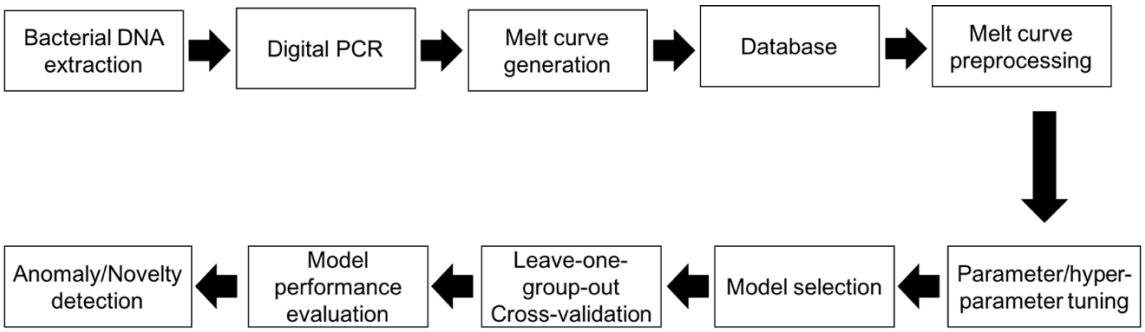


Figure 4. 1. Workflow for novelty detection

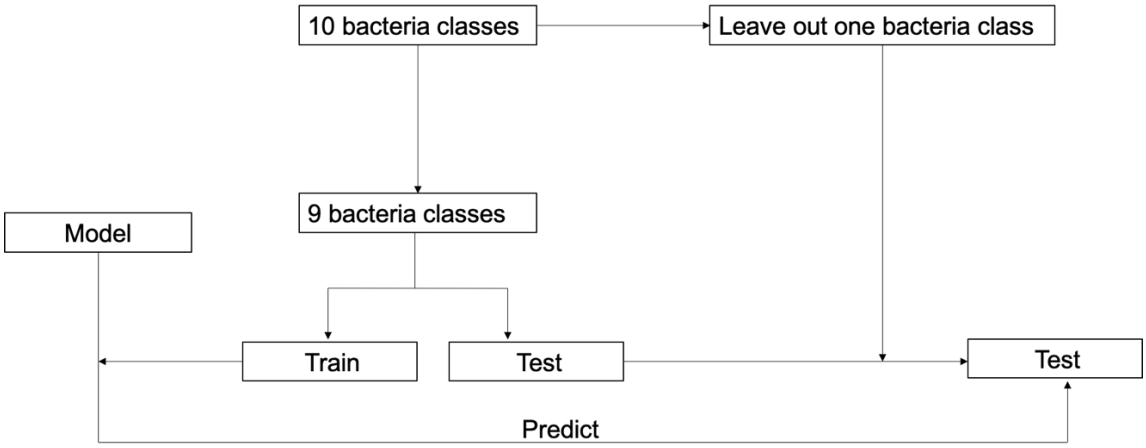


Figure 4. 2. Novelty detection experimental overview. Melt curves for each class of bacteria species was withheld in turn and mixed with the test set (randomly selected 20% of the remaining non-held-out melt curves) to make up the final test set. The left-out melt curves were assigned the novel class. The task of the model is to correctly identify the melt curves belonging to the novel class within the test set mixture.

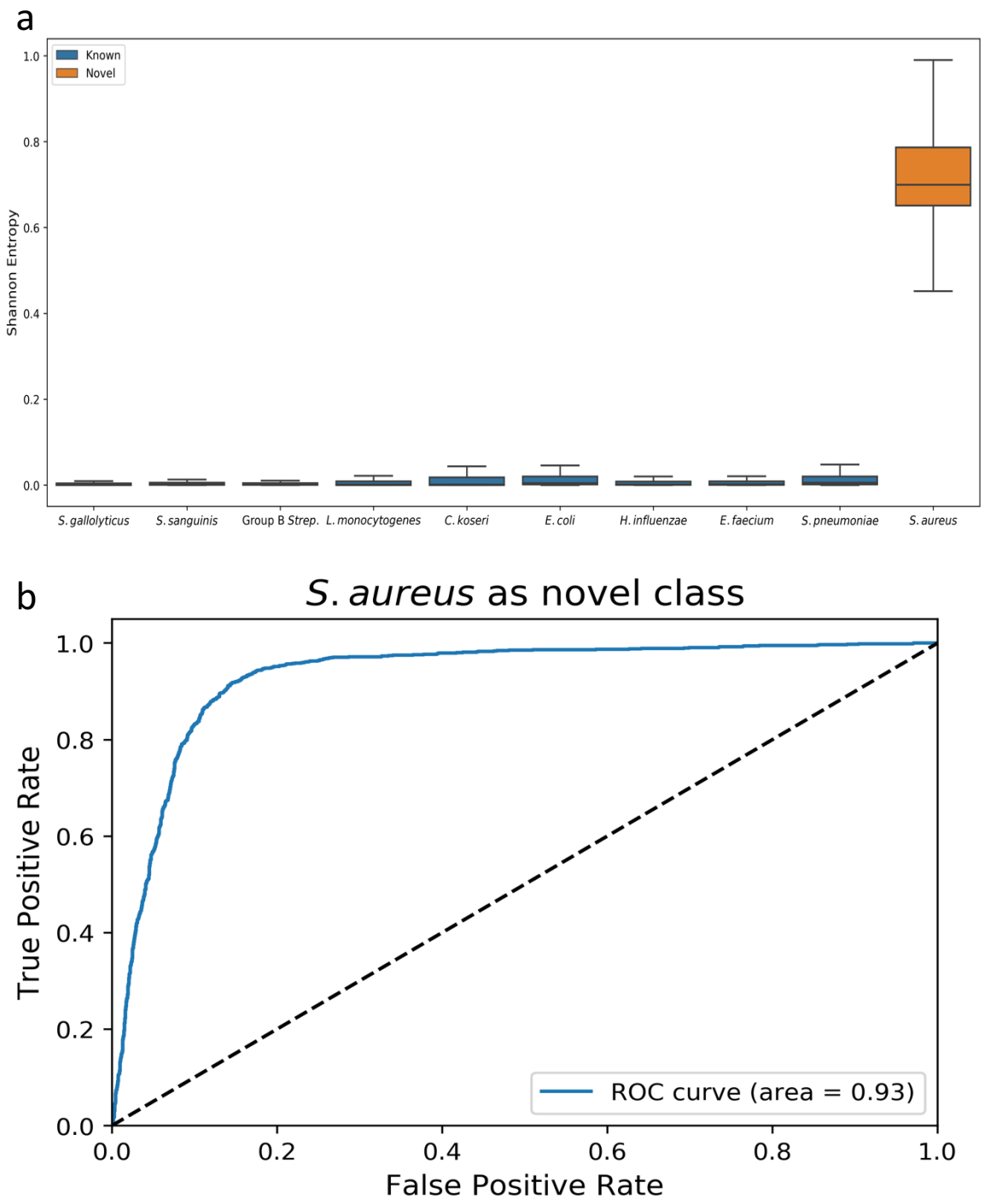


Figure 4. 3. Shannon entropy-based novelty detection. a) Box plots for entropy measure of left-out (novel) class and remainder (known) classes during one round of novelty detection leave-one-class-out experimental approach. b) ROC curve for novelty prediction with S. aureus as left-out (novel) class.

Chapter 5

Future Directions

5.1 Enabling the Integration of Host Immune Response and Pathogen Identification for Sepsis Diagnosis

Sepsis is a dysregulated immune system response to an invading pathogen that can progress to multiorgan dysfunction and hemodynamic instability (septic shock) [171]. It is a leading cause of morbidity and mortality worldwide [172]. In the United states, sepsis is the most common cause of death in the hospital [173], and the single most expensive condition to treat, accounting for nearly $24 billion in annual costs [174]. The timely and accurate diagnosis of sepsis is critical in its management and patient outcomes. On the other hand, non-infectious systemic inflammation can arise as a nonspecific response to trauma, surgery, or other noninfectious conditions. Thus, differentiating sepsis from noninfectious systemic inflammation is clinically challenging.

There has been growing interest in finding biomarkers with sufficient accuracy and reliability in differentiating sepsis from sterile inflammation. Several immune response biomarkers for sepsis have been reported, including inflammatory markers such as C-reactive

protein (CRP) [175], procalcitonin (PCT) [176], cytokines (such as IL-6 and IL-8) [177, 178], immune cell surface markers such as CD64 [179] and TREM [180], and gene expression (RNA) biomarkers (such as CEACAM4, LAMP1, PLA2G7, and PLAC8) [181-183]. Alone, however, host immune response does not provide sufficient information for the optimal management of the septic patient. The prompt and accurate determination of the microbial etiology of sepsis is important for implementing pathogen-specific targeted therapy. Although molecular microbiology tests are growing in popularity, the current gold standard clinical microbiologic tests are limited in speed and sensitivity. Therefore, in the presence of a systemic inflammation, a delayed or false negative microbiologic test may lead clinicians to presume noninfectious inflammation, and initiate inappropriate therapy. On the other hand, using empiric antibiotics when sepsis is suspected, but there is lack of information on the offending pathogen, worsens the antibiotics resistance problem. Given these challenges, there is the need for technological approaches that integrate pathogen identification and host immune response profiling for sepsis diagnosis and stratification.

Critical to the approach of integrating host immune response profiling and pathogen identification, is sample processing methods that will enable efficient separation of host immune cellular and biomolecular (DNA, RNA) components from microbial cellular and biomolecular components within the test sample (e.g., blood). A procedure for isolating bacteria from blood culture broth using density gradient solution has been reported [184]. However, this method requires relatively large volume of blood (with consequent limited use in the pediatric patient population) and suffer from low sensitivity. In addition, commercially available human DNA-depleting reagents that serve to improve the recovery of bacteria DNA from samples, require lengthy processes, show widely variable efficiencies, and are not

optimized for host DNA/RNA recovery [185, 186].

5.2 Rapid On-Chip Electrokinetic Isolation of Bacteria from Unprocessed Blood

With the advances in molecular diagnostics for infectious disease, there is an increasing need for sample processing methods that enable rapid, efficient separation of bacteria from human cell in clinical samples, to facilitate downstream processes like PCR. Microfluidic platforms have emerged as attractive diagnostic tools which offer advantages such as small sample volume and reagent consumption, high sensitivity, short assay time, and opportunity for automation [187, 188]. We explored, and herein present preliminary findings on, the feasibility of dielectrophoretic bacteria isolation from high conductance solution using an AC electrokinetic microarray device.

5.3 Materials and Methods

5.3.1 Bacterial Strain and Growth

DH5-Alpha *Escherichia coli* (*E. coli*) was chemically transformed using two different plasmids (sfGFP-N1 pET-39b(+) and pET-28), each offering direct fluorescent imaging of collected bacteria from resulting cultures. sfGFP-N1 pET-39b(+) was cloned from psfGFP-N1 stab purchased from Addgene (Watertown, Massachusetts) and a pET-39b(+) plasmid from Novagen/Thermo Fisher (Waltham, Massachusetts). Bacterial streaks for both variants showed Green Fluorescent Protein (GFP) fluorescence in

colonies. Bacteria were cultured in a standard Lurie-Bertani (LB) broth, with the addition of kanamycin at 50 μg/mL, at 37°C. While in the log phase of growth, an inducing agent was added at I μL/mL. It was incubated overnight at 37°C. Following overnight culture, freshly prepared bacterial suspension was adjusted to an optical density (OD 600) of 0.5, and serial dilutions (10-fold, 100-fold, 1000-fold) were made in phosphate-buffered saline (PBS).

5.3.2 On-Chip Dielectrophoretic Bacterial Capture

Alternating current electrokinetic (ACE) chips were purchased from Biological Dynamics, Inc. (San Diego, CA). The ACE-based sample processing has been described in chapter 2. Tygon tubing (inner diameter, 0.020 inches; outer diameter, 0.060 inches) is attached with superglue to either end of the chip, with both ends capped with syringe needles, and a 1ml syringe was attached to one end. A syringe pump set to withdrawal mode served to control fluid flow across the ACE chip. The chip was pretreated with 0.5X PBS, and alternating current (AC) electric field was applied to the chip for 5 minutes at 3 volts peak-to-peak and 14 kHz. The buffer was evacuated from the chip, and 25 μL of bacterial suspension was loaded onto the chip. AC electric field was then applied at 2 volts peak-to-peak and 2 KHz. The syringe pump was set at 3 μl/min.

Fluorescent imaging of GFP-expressing *E. coli* on the microarray chips was carried out using an Olympus BX51W epifluorescence microscope with a 4X (and 10X) objective and imaged with Olympus software.

5.4 Preliminary Results and Discussion

Figure 5.1 shows brightfield images of ACE chip before and after AC electric field was applied. Fluorescent images of DEP-isolated GFP-expressing bacteria is shown in figure 5.2. This demonstrates the feasibility of isolating bacteria from a high conductance solution using the ACE chip. A recent study reported the application of a DEP microfluidic platform for human and bacterial cell separation from whole blood [189]. However, the process of the DEP-based bacterial cell separation required additional sample preparation steps and reagents which could potentially affect both the duration and the quality of the isolate. While our preliminary results demonstrate the feasibility of isolating bacteria from a high conductance solution using a DEP platform, experiments to further quantify the bacteria isolate and assess the efficiency of the human and bacterial cell separation are necessary. This has the potential to improve the sensitivity of microbial identification especially for low volume clinical samples with low abundance microbes (as could be seen in the pediatric patient population, and in the early stages of bloodstream infections), and enable more streamlined integration of host immune response profiling and pathogen identification for the diagnosis of sepsis.

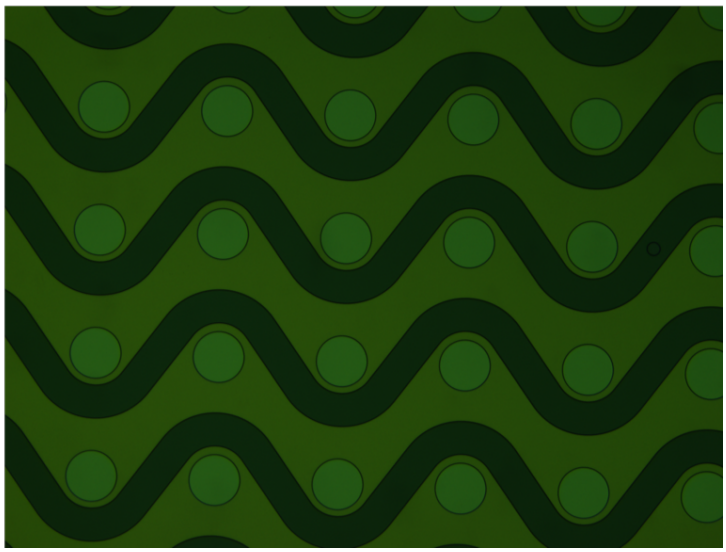
5.5 Acknowledgement

This work was supported by a research supplement grant from the NIH/NIAID, award number 3R01AI134982-02S1, and the UC San Diego Zable Endowed Chair Fund.

Chapter 5, in part, is being prepared for submission for publication of the material. Obirieze AC, Sarno B, Heller MJ. Rapid On-Chip Electrokinetic Isolation of Bacteria from Unprocessed Whole Blood. In preparation. The dissertation author is the primary investigator

and primary author of the material.

a



b

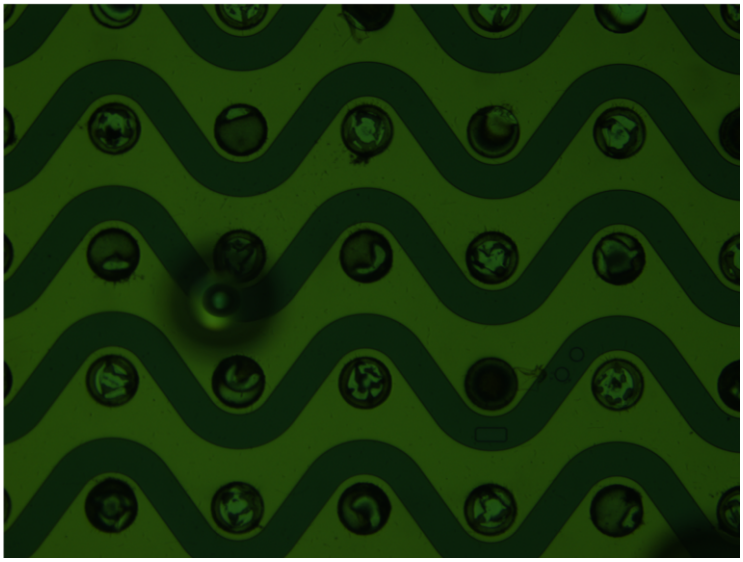


Figure 5. 1. Brightfield image of alternating current electrokinetic chip (a) before and (b) after AC electric field is applied

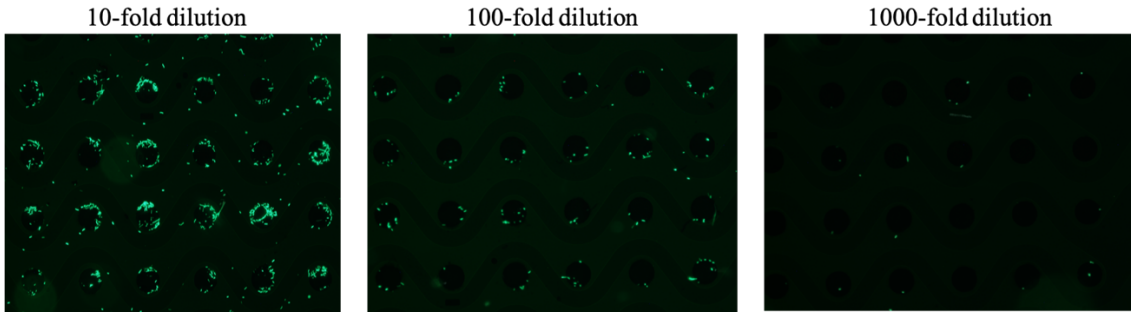


Figure 5. 2. Fluorescent images of respective dilutions of GFP-expressing *E. coli* on alternating current electrokinetic microarray chip. Starting concentration of bacteria suspension (OD 600): 0.5. GFP, green fluorescent protein.

References

[1] G. Raposo and W. Stoorvogel, "Extracellular vesicles: exosomes, microvesicles, and friends," *J Cell Biol,* vol. 200, pp. 373-83, Feb 18 2013.

[2] C. Théry, L. Zitvogel, and S. Amigorena, "Exosomes: composition, biogenesis and function," *Nature Reviews Immunology,* vol. 2, pp. 569-579, 2002.

[3] C. Friend, W. Marovitz, G. Henie, W. Henie, D. Tsuei, K. Hirschhorn, J. G. Holland, and J. Cuttner, "Observations on cell lines derived from a patient with Hodgkin's disease," *Cancer Res,* vol. 38, pp. 2581-91, Aug 1978.

[4] B. T. Pan and R. M. Johnstone, "Fate of the transferrin receptor during maturation of sheep reticulocytes in vitro: selective externalization of the receptor," *Cell,* vol. 33, pp. 967-78, Jul 1983.

[5] B. T. Pan, K. Teng, C. Wu, M. Adam, and R. M. Johnstone, "Electron microscopic evidence for externalization of the transferrin receptor in vesicular form in sheep reticulocytes," *J Cell Biol,* vol. 101, pp. 942-8, Sep 1985.

[6] I. J. McGough and J. P. Vincent, "Exosomes in developmental signalling," *Development,* vol. 143, pp. 2482-93, Jul 15 2016.

[7] S. A. Bellingham, B. B. Guo, B. M. Coleman, and A. F. Hill, "Exosomes: vehicles for the transfer of toxic proteins associated with neurodegenerative diseases?," *Front Physiol,* vol. 3, p. 124, 2012.

[8] C. Kahlert, S. A. Melo, A. Protopopov, J. Tang, S. Seth, M. Koch, J. Zhang, J. Weitz, L. Chin, A. Futreal, and R. Kalluri, "Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer," *J Biol Chem,* vol. 289, pp. 3869-75, Feb 14 2014.

[9] V. Muralidharan-Chari, J. W. Clancy, A. Sedgwick, and C. D'Souza-Schorey, "Microvesicles: mediators of extracellular communication during cancer progression," *J Cell Sci,* vol. 123, pp. 1603-11, May 15 2010.

[10] G. Camussi, M. C. Deregibus, S. Bruno, V. Cantaluppi, and L. Biancone, "Exosomes/microvesicles as a mechanism of cell-to-cell communication," *Kidney Int,* vol. 78, pp. 838-48, Nov 2010.

[11] H. Valadi, K. Ekstrom, A. Bossios, M. Sjostrand, J. J. Lee, and J. O. Lotvall, "Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells," *Nat Cell Biol,* vol. 9, pp. 654-9, Jun 2007.

[12] C. Thery, M. Ostrowski, and E. Segura, "Membrane vesicles as conveyors of immune responses," *Nat Rev Immunol,* vol. 9, pp. 581-93, Aug 2009.

[13] R. Kalluri, "The biology and function of exosomes in cancer," *J Clin Invest,* vol. 126, pp. 1208-15, Apr 1 2016.

[14] S. A. Melo, L. B. Luecke, C. Kahlert, A. F. Fernandez, S. T. Gammon, J. Kaye, V. S. LeBleu, E. A. Mittendorf, J. Weitz, N. Rahbari, C. Reissfelder, C. Pilarsky, M. F. Fraga, D. Piwnica-Worms, and R. Kalluri, "Glypican-1 identifies cancer exosomes and detects early pancreatic cancer," *Nature,* vol. 523, pp. 177-82, Jul 9 2015.

[15] D. D. Taylor and C. Gercel-Taylor, "MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer," *Gynecol Oncol,* vol. 110, pp. 13-21, Jul 2008.

[16] J. Nilsson, J. Skog, A. Nordstrand, V. Baranov, L. Mincheva-Nilsson, X. O. Breakefield, and A. Widmark, "Prostate cancer-derived urine exosomes: a novel approach to biomarkers for prostate cancer," *Br J Cancer,* vol. 100, pp. 1603-7, May 19 2009.

[17] J. Skog, T. Wurdinger, S. van Rijn, D. H. Meijer, L. Gainche, M. Sena-Esteves, W. T. Curry, Jr., B. S. Carter, A. M. Krichevsky, and X. O. Breakefield, "Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers," *Nat Cell Biol,* vol. 10, pp. 1470-6, Dec 2008.

[18] W. Zhao, X. L. Zheng, and S. P. Zhao, "Exosome and its roles in cardiovascular diseases," *Heart Fail Rev,* vol. 20, pp. 337-48, May 2015.

[19] D. D. Taylor and C. Gercel-Taylor, "Exosome platform for diagnosis and monitoring of traumatic brain injury," *Philos Trans R Soc Lond B Biol Sci,* vol. 369, Sep 26 2014.

[20] H. S. Chahar, X. Bao, and A. Casola, "Exosomes and Their Role in the Life Cycle and Pathogenesis of RNA Viruses," *Viruses,* vol. 7, pp. 3204-25, Jun 19 2015.

[21] M. R. Anderson, F. Kashanchi, and S. Jacobson, "Exosomes in Viral Disease," *Neurotherapeutics,* vol. 13, pp. 535-46, Jul 2016.

[22] E. I. Buzas, B. György, G. Nagy, A. Falus, and S. Gay, "Emerging role of extracellular vesicles in inflammatory diseases," *Nat Rev Rheumatol,* vol. 10, pp. 356-64, Jun 2014.

[23] J. Howitt and A. F. Hill, "Exosomes in the Pathology of Neurodegenerative Diseases," *J Biol Chem,* vol. 291, pp. 26589-26597, Dec 23 2016.

[24] J. Wahlgren, L. K. T. De, M. Brisslert, F. Vaziri Sani, E. Telemo, P. Sunnerhagen, and H. Valadi, "Plasma exosomes can deliver exogenous short interfering RNA to monocytes and lymphocytes," *Nucleic Acids Res,* vol. 40, p. e130, Sep 1 2012.

[25] S. Keller, J. Ridinger, A. K. Rupp, J. W. Janssen, and P. Altevogt, "Body fluid derived exosomes as a novel template for clinical diagnostics," *J Transl Med,* vol. 9, p. 86, 2011.

[26] C. Théry, S. Amigorena, G. Raposo, and A. Clayton, "Isolation and characterization of exosomes from cell culture supernatants and biological fluids," *Curr Protoc Cell Biol,* vol. Chapter 3, p. Unit 3.22, Apr 2006.

[27] P. Li, M. Kaslan, S. H. Lee, J. Yao, and Z. Gao, "Progress in Exosome Isolation Techniques," *Theranostics,* vol. 7, pp. 789-804, 2017.

[28] A. Cheruvanky, H. Zhou, T. Pisitkun, J. B. Kopp, M. A. Knepper, P. S. Yuen, and R. A. Star, "Rapid isolation of urinary exosomal biomarkers using a nanomembrane ultrafiltration concentrator," *Am J Physiol Renal Physiol,* vol. 292, pp. F1657-61, May 2007.

[29] M. L. Heinemann, M. Ilmer, L. P. Silva, D. H. Hawke, A. Recio, M. A. Vorontsova, E. Alt, and J. Vykoukal, "Benchtop isolation and characterization of functional exosomes by sequential filtration," *J Chromatogr A,* vol. 1371, pp. 125-35, Dec 5 2014.

[30] I. Lozano-Ramos, I. Bancu, A. Oliveira-Tercero, M. P. Armengol, A. Menezes-Neto, H. A. Del Portillo, R. Lauzurica-Valdemoros, and F. E. Borràs, "Size-exclusion chromatography-based enrichment of extracellular vesicles from urine samples," *J Extracell Vesicles,* vol. 4, p. 27369, 2015.

[31] J. Z. Nordin, Y. Lee, P. Vader, I. Mäger, H. J. Johansson, W. Heusermann, O. P. Wiklander, M. Hällbrink, Y. Seow, J. J. Bultema, J. Gilthorpe, T. Davies, P. J. Fairchild, S. Gabrielsson, N. C. Meisner-Kober, J. Lehtiö, C. I. Smith, M. J. Wood, and S. El Andaloussi, "Ultrafiltration with size-exclusion liquid chromatography for high yield isolation of extracellular vesicles preserving intact biophysical and functional properties," *Nanomedicine,* vol. 11, pp. 879-83, May 2015.

[32] N. Zarovni, A. Corrado, P. Guazzi, D. Zocco, E. Lari, G. Radano, J. Muhhina, C. Fondelli, J. Gavrilova, and A. Chiesi, "Integrated isolation and quantitative analysis of exosome shuttled proteins and nucleic acids using immunocapture approaches," *Methods,* vol. 87, pp. 46-58, Oct 1 2015.

[33] W. Nakai, T. Yoshida, D. Diez, Y. Miyatake, T. Nishibu, N. Imawaka, K. Naruse, Y. Sadamura, and R. Hanayama, "A novel affinity-based method for the isolation of highly purified extracellular vesicles," *Sci Rep,* vol. 6, p. 33935, Sep 23 2016.

[34] K. Koga, K. Matsumoto, T. Akiyoshi, M. Kubo, N. Yamanaka, A. Tasaki, H. Nakashima, M. Nakamura, S. Kuroki, M. Tanaka, and M. Katano, "Purification, characterization and biological significance of tumor-derived exosomes," *Anticancer Res,* vol. 25, pp. 3703-7, Nov-Dec 2005.

[35] R. T. Davies, J. Kim, S. C. Jang, E. J. Choi, Y. S. Gho, and J. Park, "Microfluidic filtration system to isolate extracellular vesicles from blood," *Lab Chip,* vol. 12, pp. 5202-10, Dec 21 2012.

[36] Z. Wang, H. J. Wu, D. Fine, J. Schmulen, Y. Hu, B. Godin, J. X. Zhang, and X. Liu, "Ciliated micropillars for the microfluidic-based isolation of nanoscale lipid vesicles," *Lab Chip,* vol. 13, pp. 2879-82, Aug 7 2013.

[37] M. He, J. Crow, M. Roth, Y. Zeng, and A. K. Godwin, "Integrated immunoisolation and protein analysis of circulating exosomes using microfluidic technology," *Lab Chip,* vol. 14, pp. 3773-80, Oct 7 2014.

[38] S. D. Ibsen, J. Wright, J. M. Lewis, S. Kim, S. Y. Ko, J. Ong, S. Manouchehri, A. Vyas, J. Akers, C. C. Chen, B. S. Carter, S. C. Esener, and M. J. Heller, "Rapid Isolation and

Detection of Exosomes and Associated Biomarkers from Plasma," *ACS Nano,* vol. 11, pp. 6641-6651, Jul 25 2017.

[39] A. Ashkin, J. M. Dziedzic, and T. Yamane, "Optical trapping and manipulation of single cells using infrared laser beams," *Nature,* vol. 330, pp. 769-71, Dec 24-31 1987.

[40] W. H. Wright, G. J. Sonek, Y. Tadir, and M. W. Berns, "Laser Trapping in Cell Biology," *Ieee Journal of Quantum Electronics,* vol. 26, pp. 2148-2157, Dec 1990.

[41] C. Monat, P. Domachuk, and B. J. Eggleton, "Integrated optofluidics: A new river of light," *Nature Photonics,* vol. 1, pp. 106-114, Feb 2007.

[42] P. R. Gascoyne and J. Vykoukal, "Particle separation by dielectrophoresis," *Electrophoresis,* vol. 23, pp. 1973-83, Jul 2002.

[43] M. P. Hughes, "Strategies for dielectrophoretic separation in laboratory-on-a-chip systems," *Electrophoresis,* vol. 23, pp. 2569-82, Aug 2002.

[44] B. Roda, A. Zattoni, P. Reschiglian, M. H. Moon, M. Mirasoli, E. Michelini, and A. Roda, "Field-flow fractionation in bioanalysis: A review of recent trends," *Anal Chim Acta,* vol. 635, pp. 132-43, Mar 9 2009.

[45] Y. Kang and D. Li, "Electrokinetic motion of particles and cells in microchannels," *Microfluidics and Nanofluidics,* vol. 6, pp. 431-460, 2009.

[46] P. Sethu, A. Sin, and M. Toner, "Microfluidic diffusive filter for apheresis (leukapheresis)," *Lab Chip,* vol. 6, pp. 83-9, Jan 2006.

[47] S. Zheng, H. Lin, J. Q. Liu, M. Balic, R. Datar, R. J. Cote, and Y. C. Tai, "Membrane microfilter device for selective capture, electrolysis and genomic analysis of human circulating tumor cells," *J Chromatogr A,* vol. 1162, pp. 154-61, Aug 31 2007.

[48] L. R. Huang, E. C. Cox, R. H. Austin, and J. C. Sturm, "Continuous particle separation through deterministic lateral displacement," *Science,* vol. 304, pp. 987-90, May 14 2004.

[49] J. A. Davis, D. W. Inglis, K. J. Morton, D. A. Lawrence, L. R. Huang, S. Y. Chou, J. C. Sturm, and R. H. Austin, "Deterministic hydrodynamics: taking blood apart," *Proc Natl Acad Sci U S A,* vol. 103, pp. 14779-84, Oct 3 2006.

[50] F. Petersson, A. Nilsson, C. Holm, H. Jonsson, and T. Laurell, "Separation of lipids from blood utilizing ultrasonic standing waves in microfluidic channels," *Analyst,* vol. 129, pp. 938-43, Oct 2004.

[51] F. Petersson, L. Aberg, A. M. Swärd-Nilsson, and T. Laurell, "Free flow acoustophoresis: microfluidic-based mode of particle and cell separation," *Anal Chem,* vol. 79, pp. 5117-23, Jul 15 2007.

[52] S. Kapishnikov, V. Kantsler, and V. Steinberg, "Continuous particle size separation and size sorting using ultrasound in a microchannel," *Journal of Statistical Mechanics-Theory and Experiment,* Jan 2006.

[53] N. Pamme and C. Wilhelm, "Continuous sorting of magnetic cells via on-chip free-flow magnetophoresis," *Lab Chip,* vol. 6, pp. 974-80, Aug 2006.

[54] K. H. Han and A. B. Frazier, "Diamagnetic capture mode magnetophoretic microseparator for blood cells," *Journal of Microelectromechanical Systems,* vol. 14, pp. 1422-1431, Dec 2005.

[55] H. A. Pohl, "The Motion and Precipitation of Suspensoids in Divergent Electric Fields," *Journal of Applied Physics,* vol. 22, pp. 869-871, 1951.

[56] H. A. Pohl, "Some Effects of Nonuniform Fields on Dielectrics," *Journal of Applied Physics,* vol. 29, pp. 1182-1188, 1958.

[57] C. Zhang, K. Khoshmanesh, A. Mitchell, and K. Kalantar-Zadeh, "Dielectrophoresis for manipulation of micro/nano particles in microfluidic systems," *Anal Bioanal Chem,* vol. 396, pp. 401-20, Jan 2010.

[58] T. Z. Jubery, S. K. Srivastava, and P. Dutta, "Dielectrophoretic separation of bioparticles in microdevices: a review," *Electrophoresis,* vol. 35, pp. 691-713, Mar 2014.

[59] K. Khoshmanesh, S. Nahavandi, S. Baratchi, A. Mitchell, and K. Kalantar-zadeh, "Dielectrophoretic platforms for bio-microfluidic systems," *Biosens Bioelectron,* vol. 26, pp. 1800-14, Jan 15 2011.

[60] S. K. Srivastava, A. Gencoglu, and A. R. Minerick, "DC insulator dielectrophoretic applications in microdevice technology: a review," *Anal Bioanal Chem,* vol. 399, pp. 301-21, Jan 2011.

[61] K. H. Kang, Y. Kang, X. Xuan, and D. Li, "Continuous separation of microparticles by size with direct current-dielectrophoresis," *Electrophoresis,* vol. 27, pp. 694-702, Feb 2006.

[62] S. K. Srivastava, J. L. Baylon-Cardiel, B. H. Lapizco-Encinas, and A. R. Minerick, "A continuous DC-insulator dielectrophoretic sorter of microparticles," *J Chromatogr A,* vol. 1218, pp. 1780-9, Apr 1 2011.

[63] B. G. Hawkins and B. J. Kirby, "Electrothermal flow effects in insulating (electrodeless) dielectrophoresis systems," *Electrophoresis,* vol. 31, pp. 3622-33, Nov 2010.

[64] S. Sridharan, J. J. Zhu, G. Q. Hu, and X. C. Xuan, "Joule heating effects on electroosmotic flow in insulator-based dielectrophoresis," *Electrophoresis,* vol. 32, pp. 2274-2281, Sep 2011.

[65] D. F. Chen, H. Du, and W. H. Li, "Bioparticle separation and manipulation using dielectrophoresis," *Sensors and Actuators a-Physical,* vol. 133, pp. 329-334, Feb 12 2007.

[66] B. Cetin and D. Li, "Dielectrophoresis in microfluidics technology," *Electrophoresis,* vol. 32, pp. 2410-27, Sep 2011.

[67] M. Koklu, S. Park, S. D. Pillai, and A. Beskok, "Negative dielectrophoretic capture of bacterial spores in food matrices," *Biomicrofluidics,* vol. 4, Sep 2010.

[68] Y. Ghallab and W. Badawy, "Sensing methods for dielectrophoresis phenomenon: From bulky instruments to lab-on-a-chip," *IEEE Circuits and Systems Magazine,* vol. 4, pp. 5 - 15, 2004.

[69] N. Abd Rahman, F. Ibrahim, and B. Yafouz, "Dielectrophoresis for Biomedical Sciences Applications: A Review," *Sensors (Basel),* vol. 17, Feb 24 2017.

[70] F. F. Becker, X. B. Wang, Y. Huang, R. Pethig, J. Vykoukal, and P. R. Gascoyne, "Separation of human breast cancer cells from blood by differential dielectric affinity," *Proc Natl Acad Sci U S A,* vol. 92, pp. 860-4, Jan 31 1995.

[71] J. Cheng, E. L. Sheldon, L. Wu, A. Uribe, L. O. Gerrue, J. Carrino, M. J. Heller, and J. P. O'Connell, "Preparation and hybridization analysis of DNA/RNA from E. coli on microfabricated bioelectronic chips," *Nat Biotechnol,* vol. 16, pp. 541-6, Jun 1998.

[72] A. Alazzam, I. Stiharu, R. Bhat, and A. N. Meguerditchian, "Interdigitated comb-like electrodes for continuous separation of malignant cells from blood using dielectrophoresis," *Electrophoresis,* vol. 32, pp. 1327-36, Jun 2011.

[73] I. Ermolina, J. Milner, and H. Morgan, "Dielectrophoretic investigation of plant virus particles: Cow Pea Mosaic Virus and Tobacco Mosaic Virus," *Electrophoresis,* vol. 27, pp. 3939-48, Oct 2006.

[74] N. G. Green, H. Morgan, and J. J. Milner, "Manipulation and trapping of sub-micron bioparticles using dielectrophoresis," *J Biochem Biophys Methods,* vol. 35, pp. 89-102, Sep 25 1997.

[75] H. A. Pohl, *Dielectrophoresis: The Behavior of Neutral Matter in Nonuniform Electric Fields*: Cambridge University Press, 1978.

[76] R. Krishnan, B. D. Sullivan, R. L. Mifflin, S. C. Esener, and M. J. Heller, "Alternating current electrokinetic separation and detection of DNA nanoparticles in high-conductance solutions," *Electrophoresis,* vol. 29, pp. 1765-74, May 2008.

[77] R. Krishnan and M. J. Heller, "An AC electrokinetic method for enhanced detection of DNA nanoparticles," *J Biophotonics,* vol. 2, pp. 253-61, Apr 2009.

[78] R. Krishnan, D. A. Dehlinger, G. J. Gemmen, R. L. Mifflin, S. C. Esener, and M. J. Heller, "Interaction of Nanoparticles at the DEP Microelectrode Interface under High Conductance Conditions," *Electrochem commun,* vol. 11, pp. 1661-1666, Aug 2009.

[79] A. Sonnenberg, J. Y. Marciniak, R. Krishnan, and M. J. Heller, "Dielectrophoretic isolation of DNA and nanoparticles from blood," *Electrophoresis,* vol. 33, pp. 2482-90, Aug 2012.

[80] S. Ibsen, A. Sonnenberg, C. Schutt, R. Mukthavaram, Y. Yeh, I. Ortac, S. Manouchehri, S. Kesari, S. Esener, and M. J. Heller, "Recovery of Drug Delivery Nanoparticles from Human Plasma Using an Electrokinetic Platform Technology," *Small,* vol. 11, pp. 5088-96, Oct 2015.

[81] G. H. Reed, J. O. Kent, and C. T. Wittwer, "High-resolution DNA melting analysis for simple and efficient molecular diagnostics," *Pharmacogenomics,* vol. 8, pp. 597-608, Jun 2007.

[82] C. N. Gundry, P. S. Bernard, M. G. Herrmann, G. H. Reed, and C. T. Wittwer, "Rapid F508del and F508C assay using fluorescent hybridization probes," *Genet Test,* vol. 3, pp. 365-70, 1999.

[83] C. T. Wittwer, M. G. Herrmann, C. N. Gundry, and K. S. Elenitoba-Johnson, "Real-time multiplex PCR assays," *Methods,* vol. 25, pp. 430-42, Dec 2001.

[84] M. J. Lay and C. T. Wittwer, "Real-time fluorescence genotyping of factor V Leiden during rapid-cycle PCR," *Clin Chem,* vol. 43, pp. 2262-7, Dec 1997.

[85] C. N. Gundry, J. G. Vandersteen, G. H. Reed, R. J. Pryor, J. Chen, and C. T. Wittwer, "Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes," *Clin Chem,* vol. 49, pp. 396-406, Mar 2003.

[86] M. Erali, K. V. Voelkerding, and C. T. Wittwer, "High resolution melting applications for clinical laboratory medicine," *Exp Mol Pathol,* vol. 85, pp. 50-8, Aug 2008.

[87] M. Liew, R. Pryor, R. Palais, C. Meadows, M. Erali, E. Lyon, and C. Wittwer, "Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons," *Clin Chem,* vol. 50, pp. 1156-64, Jul 2004.

[88] R. H. Lipsky, C. M. Mazzanti, J. G. Rudolph, K. Xu, G. Vyas, D. Bozak, M. Q. Radel, and D. Goldman, "DNA melting analysis for detection of single nucleotide polymorphisms," *Clin Chem,* vol. 47, pp. 635-44, Apr 2001.

[89] C. T. Wittwer, G. H. Reed, C. N. Gundry, J. G. Vandersteen, and R. J. Pryor, "High-resolution genotyping by amplicon melting analysis using LCGreen," *Clin Chem,* vol. 49, pp. 853-60, Jun 2003.

[90] M. G. Herrmann, J. D. Durtschi, L. K. Bromley, C. T. Wittwer, and K. V. Voelkerding, "Amplicon DNA melting analysis for mutation scanning and genotyping: cross-platform comparison of instruments and dyes," *Clin Chem,* vol. 52, pp. 494-503, Mar 2006.

[91] M. Li, L. Zhou, R. A. Palais, and C. T. Wittwer, "Genotyping accuracy of high-resolution DNA melting instruments," *Clin Chem,* vol. 60, pp. 864-72, Jun 2014.

[92] J. L. Montgomery, L. N. Sanford, and C. T. Wittwer, "High-resolution DNA melting analysis in clinical research and diagnostics," *Expert Rev Mol Diagn,* vol. 10, pp. 219-40, Mar 2010.

[93] R. Graham, M. Liew, C. Meadows, E. Lyon, and C. T. Wittwer, "Distinguishing different DNA heterozygotes by high-resolution melting," *Clin Chem,* vol. 51, pp. 1295-8, Jul 2005.

[94] R. A. Palais, M. A. Liew, and C. T. Wittwer, "Quantitative heteroduplex analysis for single nucleotide polymorphism genotyping," *Anal Biochem,* vol. 346, pp. 167-75, Nov 1 2005.

[95] C. Willmore, J. A. Holden, L. Zhou, S. Tripp, C. T. Wittwer, and L. J. Layfield, "Detection of c-kit-activating mutations in gastrointestinal stromal tumors by high-resolution amplicon melting analysis," *Am J Clin Pathol,* vol. 122, pp. 206-16, Aug 2004.

[96] L. Zhou, J. Vandersteen, L. Wang, T. Fuller, M. Taylor, B. Palais, and C. T. Wittwer, "High-resolution DNA melting curve analysis to establish HLA genotypic identity," *Tissue Antigens,* vol. 64, pp. 156-64, Aug 2004.

[97] S. I. Fraley, J. Hardick, B. J. Masek, P. Athamanolap, R. E. Rothman, C. A. Gaydos, K. C. Carroll, T. Wakefield, T. H. Wang, and S. Yang, "Universal digital high-resolution melt: a novel approach to broad-based profiling of heterogeneous biological samples," *Nucleic Acids Res,* vol. 41, p. e175, Oct 2013.

[98] D. O. Velez, H. Mack, J. Jupe, S. Hawker, N. Kulkarni, B. Hedayatnia, Y. Zhang, S. Lawrence, and S. I. Fraley, "Massively parallel digital high resolution melt for rapid and absolutely quantitative sequence profiling," *Sci Rep,* vol. 7, p. 42326, Feb 8 2017.

[99] M. Sinha, H. Mack, T. P. Coleman, and S. I. Fraley, "A High-Resolution Digital DNA Melting Platform for Robust Sequence Profiling and Enhanced Genotype Discrimination," *SLAS Technol,* vol. 23, pp. 580-591, Dec 2018.

[100] M. Faul, L. Xu, M. M. Wald, and V. G. Coronado. (2010). *Traumatic brain injury in the united states: emergency department visits, hospitalizations and deaths 2002–2006.* Available: https://www.cdc.gov/traumaticbraininjury/pdf/blue_book.pdf

[101] K. E. Saatman, A. C. Duhaime, R. Bullock, A. I. Maas, A. Valadka, and G. T. Manley, "Classification of traumatic brain injury for targeted therapies," *J Neurotrauma,* vol. 25, pp. 719-38, Jul 2008.

[102] K. McInnes, C. L. Friesen, D. E. MacKenzie, D. A. Westwood, and S. G. Boe, "Mild Traumatic Brain Injury (mTBI) and chronic cognitive impairment: A scoping review," *PLoS One,* vol. 12, p. e0174847, 2017.

[103] S. Marshall, M. Bayley, S. McCullagh, D. Velikonja, and L. Berrigan, "Clinical practice guidelines for mild traumatic brain injury and persistent symptoms," *Can Fam Physician,* vol. 58, pp. 257-67, e128-40, Mar 2012.

[104] C. Prince and M. E. Bruhns, "Evaluation and Treatment of Mild Traumatic Brain Injury: The Role of Neuropsychology," *Brain Sci,* vol. 7, Aug 17 2017.

[105] V. Y. Ma, L. Chan, and K. J. Carruthers, "Incidence, prevalence, costs, and impact on disability of common conditions requiring rehabilitation in the United States: stroke, spinal cord injury, traumatic brain injury, multiple sclerosis, osteoarthritis, rheumatoid arthritis, limb loss, and back pain," *Arch Phys Med Rehabil,* vol. 95, pp. 986-995.e1, May 2014.

[106] A. C. Mckee and D. H. Daneshvar, *The neuropathology of traumatic brain injury*. London: Elsevier B.V.

[107] F. Mouliere, R. Mair, D. Chandrananda, F. Marass, C. G. Smith, J. Su, J. Morris, C. Watts, K. M. Brindle, and N. Rosenfeld, "Detection of cell-free DNA fragmentation and copy number alterations in cerebrospinal fluid from glioma patients," *EMBO Mol Med,* vol. 10, Dec 2018.

[108] M. Brenner, "Role of GFAP in CNS injuries," *Neurosci Lett,* vol. 565, pp. 7-13, Apr 17 2014.

[109] H. Liu, M. E. Rose, X. Ma, S. Culver, C. E. Dixon, and S. H. Graham, "In vivo transduction of neurons with TAT-UCH-L1 protects brain against controlled cortical impact injury," *PLoS One,* vol. 12, p. e0178049, 2017.

[110] K. Tomita, T. A. Nakada, T. Oshima, T. Motoshima, R. Kawaguchi, and S. Oda, "Tau protein as a diagnostic marker for diffuse axonal injury," *PLoS One,* vol. 14, p. e0214381, 2019.

[111] J. Rogg, H. Spader, B. J. Wilcox, A. Ellermeier, S. Correira, A. Chodobski, J. Szmydynger-Chodobska, N. Raukar, J. T. Machan, J. J. Crisco, W. C. LaFrance, Jr., C. Brown University Traumatic Brain Injury Research, J. Rogg, W. C. LaFrance, Jr., J. Robson, J. J. T. Crisco, B. Wilcox, S. Correia, M. Worden, N. Raukar, A. Chodobski, J. Szmydynger-Chodobska, A. Ellermeier, H. Spader, E. Morrell, W. Smith, J. T. Machan, G. L. Baird, G. Tung, S. Mernoff, P. Quesenberry, M. Hulstyn, P. Fadale, R. Fiore, S. Deoni, W. Heindel, S. Mather, and P. Lieberman, "The Brown University Traumatic Brain Injury Research Consortium and the Norman Prince Neurosciences Institute," *Rhode Island medical journal (2013),* vol. 97, pp. 22-6, 2014.

[112] D. Kanmert, A. Cantlon, C. R. Muratore, M. Jin, T. T. O'Malley, G. Lee, T. L. Young-Pearse, D. J. Selkoe, and D. M. Walsh, "C-Terminally Truncated Forms of Tau, But Not Full-Length Tau or Its C-Terminal Fragments, Are Released from Neurons Independently of Cell Death," *J Neurosci,* vol. 35, pp. 10851-65, Jul 29 2015.

[113] E. M. Hol and M. Pekny, "Glial fibrillary acidic protein (GFAP) and the astrocyte intermediate filament system in diseases of the central nervous system," *Curr Opin Cell Biol,* vol. 32, pp. 121-30, Feb 2015.

[114] F. Notturno, M. Capasso, A. DeLauretis, M. Carpo, and A. Uncini, "Glial fibrillary acidic protein as a marker of axonal damage in chronic neuropathies," *Muscle Nerve,* vol. 40, pp. 50-4, Jul 2009.

[115] F. Notturno, C. M. Caporale, A. De Lauretis, and A. Uncini, "Glial fibrillary acidic protein: a marker of axonal Guillain-Barrè syndrome and outcome," *Muscle Nerve,* vol. 38, pp. 899-903, Jul 2008.

[116] I. N. Day and R. J. Thompson, "UCHL1 (PGP 9.5): neuronal biomarker and ubiquitin system protein," *Prog Neurobiol,* vol. 90, pp. 327-62, Mar 2010.

[117] P. Jackson and R. J. Thompson, "The demonstration of new human brain-specific proteins by high-resolution two-dimensional polyacrylamide gel electrophoresis," *J Neurol Sci,* vol. 49, pp. 429-38, Mar 1981.

[118] S. Mondello, A. Linnet, A. Buki, S. Robicsek, A. Gabrielli, J. Tepas, L. Papa, G. M. Brophy, F. Tortella, R. L. Hayes, and K. K. Wang, "Clinical utility of serum levels of ubiquitin C-terminal hydrolase as a biomarker for severe traumatic brain injury," *Neurosurgery,* vol. 70, pp. 666-75, Mar 2012.

[119] A. Chodobski, B. J. Zink, and J. Szmydynger-Chodobska, "Blood-brain barrier pathophysiology in traumatic brain injury," *Transl Stroke Res,* vol. 2, pp. 492-516, Dec 2011.

[120] E. M. Rodrigues Filho, D. Simon, N. Ikuta, C. Klovan, F. A. Dannebrock, C. Oliveira de Oliveira, and A. Regner, "Elevated cell-free plasma DNA level as an independent predictor of mortality in patients with severe traumatic brain injury," *J Neurotrauma,* vol. 31, pp. 1639-46, Oct 1 2014.

[121] S. Ohayon, M. Boyko, A. Saad, A. Douvdevani, B. F. Gruenbaum, I. Melamed, Y. Shapira, V. I. Teichberg, and A. Zlotnik, "Cell-free DNA as a marker for prediction of brain damage in traumatic brain injury in rats," *J Neurotrauma,* vol. 29, pp. 261-7, Jan 20 2012.

[122] J. L. Gerberding and S. Binder. (2003). *Report to Congress on Mild Traumatic Brain Injury in the United States: Steps to Prevent a Serious Public Health Problem*. Available: https://www.cdc.gov/traumaticbraininjury/pdf/mtbireport-a.pdf

[123] Z. S. Gan, S. C. Stein, R. Swanson, S. Guan, L. Garcia, D. Mehta, and D. H. Smith, "Blood Biomarkers for Traumatic Brain Injury: A Quantitative Assessment of Diagnostic and Prognostic Accuracy," *Front Neurol,* vol. 10, p. 446, 2019.

[124] N. S. King, S. Crawford, F. J. Wenden, N. E. Moss, and D. T. Wade, "The Rivermead Post Concussion Symptoms Questionnaire: a measure of symptoms commonly experienced after head injury and its reliability," *J Neurol,* vol. 242, pp. 587-92, Sep 1995.

[125] M. Lannsjö, J. Borg, G. Björklund, J. L. Af Geijerstam, and A. Lundgren-Nilsson, "Internal construct validity of the Rivermead Post-Concussion Symptoms Questionnaire," *J Rehabil Med,* vol. 43, pp. 997-1002, Nov 2011.

[126] D. V. Agoston, A. Shutes-David, and E. R. Peskind, "Biofluid biomarkers of traumatic brain injury," *Brain Inj,* vol. 31, pp. 1195-1203, 2017.

[127] P. Bishop, D. Rocca, and J. M. Henley, "Ubiquitin C-terminal hydrolase L1 (UCH-L1): structure, distribution and roles in brain function and dysfunction," *Biochem J,* vol. 473, pp. 2453-62, Aug 15 2016.

[128] J. M. Lewis, A. D. Vyas, Y. Qiu, K. S. Messer, R. White, and M. J. Heller, "Integrated Analysis of Exosomal Protein Biomarkers on Alternating Current Electrokinetic Chips Enables Rapid Detection of Pancreatic Cancer in Patient Blood," *ACS Nano,* vol. 12, pp. 3311-3320, Apr 24 2018.

[129] A. Sonnenberg, J. Y. Marciniak, E. A. Skowronski, S. Manouchehri, L. Rassenti, E. M. Ghia, G. F. Widhopf, 2nd, T. J. Kipps, and M. J. Heller, "Dielectrophoretic isolation and detection of cancer-related circulating cell-free DNA biomarkers from blood and plasma," *Electrophoresis,* vol. 35, pp. 1828-36, Jul 2014.

[130] D. V. Jeste, B. W. Palmer, P. S. Appelbaum, S. Golshan, D. Glorioso, L. B. Dunn, K. Kim, T. Meeks, and H. C. Kraemer, "A new brief instrument for assessing decisional capacity for clinical research," *Arch Gen Psychiatry,* vol. 64, pp. 966-74, Aug 2007.

[131] G. Teasdale and B. Jennett, "Assessment of coma and impaired consciousness. A practical scale," *Lancet,* vol. 2, pp. 81-4, Jul 13 1974.

[132] J. F. Malec, A. W. Brown, C. L. Leibson, J. T. Flaada, J. N. Mandrekar, N. N. Diehl, and P. K. Perkins, "The Mayo classification system for traumatic brain injury severity," *Journal of Neurotrauma,* vol. 24, pp. 1417-1424, Sep 2007.

[133] J. Lewis, A. A. Alattar, J. Akers, B. S. Carter, M. Heller, and C. C. Chen, "A Pilot Proof-Of-Principle Analysis Demonstrating Dielectrophoresis (DEP) as a Glioblastoma Biomarker Platform," *Sci Rep,* vol. 9, p. 10279, Jul 16 2019.

[134] R. Diaz-Arrastia, K. K. Wang, L. Papa, M. D. Sorani, J. K. Yue, A. M. Puccio, P. J. McMahon, T. Inoue, E. L. Yuh, H. F. Lingsma, A. I. Maas, A. B. Valadka, D. O. Okonkwo, and G. T. Manley, "Acute biomarkers of traumatic brain injury: relationship between plasma levels of ubiquitin C-terminal hydrolase-L1 and glial fibrillary acidic protein," *J Neurotrauma,* vol. 31, pp. 19-25, Jan 1 2014.

[135] R. C. Gardner, R. Rubenstein, K. K. W. Wang, F. K. Korley, J. K. Yue, E. L. Yuh, P. Mukherje, A. B. Valadka, D. O. Okonkwo, R. Diaz-Arrastia, and G. T. Manley, "Age-Related Differences in Diagnostic Accuracy of Plasma Glial Fibrillary Acidic Protein and Tau for Identifying Acute Intracranial Trauma on Computed Tomography: A TRACK-TBI Study," *J Neurotrauma,* vol. 35, pp. 2341-2350, Oct 15 2018.

[136] B. Lee and A. Newberg, "Neuroimaging in traumatic brain imaging," *NeuroRx,* vol. 2, pp. 372-83, Apr 2005.

[137] T. Bogoslovsky, J. Gill, A. Jeromin, C. Davis, and R. Diaz-Arrastia, "Fluid Biomarkers of Traumatic Brain Injury and Intended Context of Use," *Diagnostics (Basel),* vol. 6, Oct 18 2016.

[138] F. K. Korley, J. K. Yue, D. H. Wilson, K. Hrusovsky, R. Diaz-Arrastia, A. R. Ferguson, E. L. Yuh, P. Mukherjee, K. K. W. Wang, A. B. Valadka, A. M. Puccio, D. O. Okonkwo, and G. T. Manley, "Performance Evaluation of a Multiplex Assay for Simultaneous Detection of Four Clinically Relevant Traumatic Brain Injury Biomarkers," *J Neurotrauma,* vol. 36, pp. 182-7, Jul 23 2018.

[139] L. Ho, W. Zhao, K. Dams-O'Connor, C. Y. Tang, W. Gordon, E. R. Peskind, S. Yemul, V. Haroutunian, and G. M. Pasinetti, "Elevated plasma MCP-1 concentration following traumatic brain injury as a potential "predisposition" factor associated with an increased risk for subsequent development of Alzheimer's disease," *J Alzheimers Dis,* vol. 31, pp. 301-13, 2012.

[140] G. H. Reed and C. T. Wittwer, "Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis," *Clin Chem,* vol. 50, pp. 1748-54, Oct 2004.

[141] P. Bidet, S. Liguori, C. Plainvert, S. Bonacorsi, C. Courroux, C. d'Humieres, C. Poyart, A. Efstratiou, and E. Bingen, "Identification of group A streptococcal emm types commonly associated with invasive infections and antimicrobial resistance by the use of multiplex PCR and high-resolution melting analysis," *Eur J Clin Microbiol Infect Dis,* vol. 31, pp. 2817-26, Oct 2012.

[142] A. L. Roth and N. D. Hanson, "Rapid detection and statistical differentiation of KPC gene variants in Gram-negative pathogens by use of high-resolution melting and ScreenClust analyses," *J Clin Microbiol,* vol. 51, pp. 61-5, Jan 2013.

[143] M. R. Zianni, M. R. Nikbakhtzadeh, B. T. Jackson, J. Panescu, and W. A. Foster, "Rapid discrimination between Anopheles gambiae s.s. and Anopheles arabiensis by High-Resolution Melt (HRM) analysis," *J Biomol Tech,* vol. 24, pp. 1-7, Apr 2013.

[144] S. Yang, P. Ramachandran, R. Rothman, Y. H. Hsieh, A. Hardick, H. Won, A. Kecojevic, J. Jackman, and 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,* vol. 47, pp. 2252-5, Jul 2009.

[145] P. Athamanolap, V. Parekh, S. I. Fraley, V. Agarwal, D. J. Shin, M. A. Jacobs, T. H. Wang, and S. Yang, "Trainable high resolution melt curve machine learning classifier for large-scale reliable genotyping of sequence variants," *PLoS One,* vol. 9, p. e109094, 2014.

[146] 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, and S. Yang, "Nested Machine Learning Facilitates Increased Sequence Content for Large-Scale Automated High Resolution Melt Genotyping," *Sci Rep,* vol. 6, p. 19218, Jan 18 2016.

[147] N. Andini, B. Wang, P. Athamanolap, J. Hardick, B. J. Masek, S. Thair, A. Hu, G. Avornu, S. Peterson, S. Cogill, R. E. Rothman, K. C. Carroll, C. A. Gaydos, J. T. Wang, S. Batzoglou, and S. Yang, "Microbial Typing by Machine Learned DNA Melt Signatures," *Sci Rep,* vol. 7, p. 42097, Feb 6 2017.

[148] S. Lu, G. Mirchevska, S. S. Phatak, D. Li, J. Luka, R. A. Calderone, and W. A. Fonzi, "Dynamic time warping assessment of high-resolution melt curves provides a robust metric for fungal identification," *PLoS One,* vol. 12, p. e0173320, 2017.

[149] G. Klinger, I. Levy, L. Sirota, V. Boyko, B. Reichman, L. Lerner-Geva, and N. Israel Neonatal, "Epidemiology and risk factors for early onset sepsis among very-low-birthweight infants," *Am J Obstet Gynecol,* vol. 201, pp. 38 e1-6, Jul 2009.

[150] B. J. Stoll, N. I. Hansen, P. J. Sanchez, R. G. Faix, B. B. Poindexter, K. P. Van Meurs, M. J. Bizzarro, R. N. Goldberg, I. D. Frantz, 3rd, E. C. Hale, S. Shankaran, K. Kennedy, W. A. Carlo, K. L. Watterberg, E. F. Bell, M. C. Walsh, K. Schibler, A. R. Laptook, A. L. Shane, S. J. Schrag, A. Das, R. D. Higgins, H. Eunice Kennedy Shriver National Institute of Child, and N. Human Development Neonatal Research, "Early onset neonatal sepsis: the burden of group B Streptococcal and E. coli disease continues," *Pediatrics,* vol. 127, pp. 817-26, May 2011.

[151] D. E. Hinkle, W. Wiersma, and S. G. Jurs, *Applied Statistics for the Behavioral Sciences*, 5th ed.: Houghton Mifflin, 2003.

[152] C. Albon, *Machine Learning with Python Cookbook*. California, USA: O'Reilly Media, 2018.

[153] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research,* vol. 12, pp. 2825-2830, Oct 2011.

[154] P. Athamanolap, K. Hsieh, and A. T. Wang, "Integrated Bacterial Identification and Antimicrobial Susceptibility Testing for Polymicrobial Infections Using Digital PCR and Digital High-Resolution Melt in a Microfluidic Array Platform," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2018, pp. 5346-5349, Jul 2018.

[155] C. M. O'Keefe and T. L. Wang, "Digital High-Resolution Melt Platform for Rapid and Parallelized Molecule-by-Molecule Genetic Profiling," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2018, pp. 5342-5345, Jul 2018.

[156] C. E. Bianca Zadrozny, "Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers," in *Proceedings of the 18th International Conference on Machine Learning*, 2001, pp. 609-616.

[157] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, pp. 226-239, Mar 1998.

[158] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal,* vol. 27, pp. 379-423, 1948.

[159] W. Lee and D. Xiang, "Information-theoretic measures for anomaly detection," *2001 Ieee Symposium on Security and Privacy, Proceedings,* pp. 130-143, 2001.

[160] A. Lakhina, M. Crovella, and C. Diot, "Mining anomalies using traffic feature distributions," *Acm Sigcomm Computer Communication Review,* vol. 35, pp. 217-228, Oct 2005.

[161] P. Berezinski, B. Jasiul, and M. Szpyrka, "An Entropy-Based Network Anomaly Detection Method," *Entropy,* vol. 17, pp. 2367-2408, Apr 2015.

[162] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed.: Wiley & Sons, 2012.

[163] J. C. Platt, *Probabilistic outputs for support vector machines and comparison to regularized likelihood methods*. Cambridge, Massachusetts: MIT Press, 2000.

[164] P. Sollich, "Probabilistic methods for Support Vector Machines," *Advances in Neural Information Processing Systems 12,* vol. 12, pp. 349-355, 2000.

[165] H. H. Chen, P. Tino, and X. Yao, "Probabilistic Classification Vector Machines," *Ieee Transactions on Neural Networks,* vol. 20, pp. 901-914, Jun 2009.

[166] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net (vol B 67, pg 301, 2005)," *Journal of the Royal Statistical Society Series B-Statistical Methodology,* vol. 67, pp. 768-768, 2005.

[167] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society Series B-Statistical Methodology,* vol. 68, pp. 49-67, 2006.

[168] A. Sokolov, E. O. Paull, and J. M. Stuart, "One-Class Detection of Cell States in Tumor Subtypes," *Pac Symp Biocomput,* vol. 21, pp. 405-16, 2016.

[169] W. J. Youden, "Index for rating diagnostic tests," *Cancer,* vol. 3, pp. 32-5, Jan 1950.

[170] K. E. Jones, N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak, "Global trends in emerging infectious diseases," *Nature,* vol. 451, pp. 990-3, Feb 21 2008.

[171] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J. D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J. L. Vincent, and D. C. Angus, "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)," *Jama,* vol. 315, pp. 801-10, Feb 23 2016.

[172] K. E. Rudd, S. C. Johnson, K. M. Agesa, K. A. Shackelford, D. Tsoi, D. R. Kievlan, D. V. Colombara, K. S. Ikuta, N. Kissoon, S. Finfer, C. Fleischmann-Struzek, F. R. Machado, K. K. Reinhart, K. Rowan, C. W. Seymour, R. S. Watson, T. E. West, F. Marinho, S. I. Hay, R. Lozano, A. D. Lopez, D. C. Angus, C. J. L. Murray, and M. Naghavi, "Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study," *Lancet,* vol. 395, pp. 200-211, Jan 18 2020.

[173] C. Rhee, T. M. Jones, Y. Hamad, A. Pande, J. Varon, C. O'Brien, D. J. Anderson, D. K. Warren, R. B. Dantes, L. Epstein, and M. Klompas, "Prevalence, Underlying Causes, and Preventability of Sepsis-Associated Mortality in US Acute Care Hospitals," *JAMA Netw Open,* vol. 2, p. e187571, Feb 1 2019.

[174] C. Torio and B. Moore. (2016). *National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2013*. Available: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb204-Most-Expensive-Hospital-Conditions.pdf.

[175] P. Povoa, "C-reactive protein: a valuable marker of sepsis," *Intensive Care Med,* vol. 28, pp. 235-43, Mar 2002.

[176] B. Uzzan, R. Cohen, P. Nicolas, M. Cucherat, and G. Y. Perret, "Procalcitonin as a diagnostic test for sepsis in critically ill adults and after surgery or trauma: a systematic review and meta-analysis," *Crit Care Med,* vol. 34, pp. 1996-2003, Jul 2006.

[177] R. Berner, C. M. Niemeyer, J. U. Leititis, A. Funke, C. Schwab, U. Rau, K. Richter, M. S. Tawfeek, A. Clad, and M. Brandis, "Plasma levels and gene expression of granulocyte colony-stimulating factor, tumor necrosis factor-alpha, interleukin (IL)-1beta, IL-6, IL-8, and soluble intercellular adhesion molecule-1 in neonatal early onset sepsis," *Pediatr Res,* vol. 44, pp. 469-77, Oct 1998.

[178] S. Lusyati, C. V. Hulzebos, J. Zandvoort, H. Sukandar, and P. J. Sauer, "Cytokines patterns in newborn infants with late onset sepsis," *J Neonatal Perinatal Med,* vol. 6, pp. 153-63, 2013.

[179] B. H. Davis, S. H. Olsen, E. Ahmad, and N. C. Bigelow, "Neutrophil CD64 is an improved indicator of infection or sepsis in emergency department patients," *Arch Pathol Lab Med,* vol. 130, pp. 654-61, May 2006.

[180] S. Gibot, M. N. Kolopp-Sarda, M. C. Bene, A. Cravoisy, B. Levy, G. C. Faure, and P. E. Bollaert, "Plasma level of a triggering receptor expressed on myeloid cells-1: its diagnostic accuracy in patients with suspected sepsis," *Ann Intern Med,* vol. 141, pp. 9-15, Jul 6 2004.

[181] L. McHugh, T. A. Seldon, R. A. Brandon, J. T. Kirk, A. Rapisarda, A. J. Sutherland, J. J. Presneill, D. J. Venter, J. Lipman, M. R. Thomas, P. M. Klein Klouwenberg, L. van Vught, B. Scicluna, M. Bonten, O. L. Cremer, M. J. Schultz, T. van der Poll, T. D. Yager, and R. B. Brandon, "A Molecular Host Response Assay to Discriminate Between Sepsis and Infection-Negative Systemic Inflammation in Critically Ill Patients: Discovery and Validation in Independent Cohorts," *PLoS Med,* vol. 12, p. e1001916, Dec 2015.

[182] J. J. Zimmerman, E. Sullivan, T. D. Yager, C. Cheng, L. Permut, S. Cermelli, L. McHugh, D. Sampson, T. Seldon, R. B. Brandon, and R. A. Brandon, "Diagnostic Accuracy of a Host Gene Expression Signature That Discriminates Clinical Severe Sepsis Syndrome and Infection-Negative Systemic Inflammation Among Critically Ill Children," *Crit Care Med,* vol. 45, pp. e418-e425, Apr 2017.

[183] R. R. Miller Iii, B. K. Lopansri, J. P. Burke, M. Levy, S. Opal, R. E. Rothman, F. R. D'Alessio, V. K. Sidhaye, N. R. Aggarwal, R. Balk, J. A. Greenberg, M. Yoder, G. Patel, E. Gilbert, M. Afshar, J. P. Parada, G. S. Martin, A. M. Esper, J. A. Kempker, M. Narasimhan, A. Tsegaye, S. Hahn, P. Mayo, T. van der Poll, M. J. Schultz, B. P. Scicluna, P. Klein Klouwenberg, A. Rapisarda, T. A. Seldon, L. C. McHugh, T. D. Yager, S. Cermelli, D. Sampson, V. Rothwell, R. Newman, S. Bhide, B. A. Fox, J. T. Kirk, K. Navalkar, R. F. Davis, R. A. Brandon, and R. B. Brandon, "Validation of a Host Response Assay, Septicyte LAB, for Discriminating Sepsis from SIRS in the ICU," *Am J Respir Crit Care Med,* Apr 6 2018.

[184] J. Tan, B. D. Lee, L. Polo-Parada, and S. Sengupta, "Kinetically limited differential centrifugation as an inexpensive and readily available alternative to centrifugal elutriation," *Biotechniques,* vol. 53, pp. 104-8, Aug 2012.

[185] M. R. Hasan, A. Rawat, P. Tang, P. V. Jithesh, E. Thomas, R. Tan, and P. Tilley, "Depletion of Human DNA in Spiked Clinical Specimens for Improvement of Sensitivity of Pathogen

Detection by Next-Generation Sequencing," *J Clin Microbiol,* vol. 54, pp. 919-27, Apr 2016.

[186] H. P. Horz, S. Scheer, M. E. Vianna, and G. Conrads, "New methods for selective isolation of bacterial DNA from human clinical specimens," *Anaerobe,* vol. 16, pp. 47-53, Feb 2010.

[187] J. W. Hong and S. R. Quake, "Integrated nanoliter systems," *Nat Biotechnol,* vol. 21, pp. 1179-83, Oct 2003.

[188] G. M. Whitesides, "The origins and the future of microfluidics," *Nature,* vol. 442, pp. 368-73, Jul 27 2006.

[189] L. D'Amico, N. J. Ajami, J. A. Adachi, P. R. Gascoyne, and J. F. Petrosino, "Isolation and concentration of bacteria from blood using microfluidic membraneless dialysis and dielectrophoresis," *Lab Chip,* vol. 17, pp. 1340-1348, Mar 29 2017.