A Bioengineering Approach to Develop New Technologies for Genomics-Based Healthcare

By
Olivia M. Ott
B.S., University of California San Diego, 2021

Thesis

Submitted in partial fulfillment of the requirements for the
Degree of Master of Science in Biomedical Engineering at
Brown University

PROVIDENCE, RHODE ISLAND

MAY 2023

AUTHORIZATION TO LEND AND REPRODUCE THE THESIS

As the sole author of this thesis, I authorize Brown University to lend it to other institutions or individuals for the purpose of scholarly research.

Date_____

Olivia Ott, Author

I further authorize Brown University to reproduce this thesis by photocopying or other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Date_____

Olivia Ott, Author

This thesis by Olivia M. Ott is accepted in its present form by the Center of Biomedical Engineering as satisfying the thesis requirements for the degree of Master of Science.

Date_____ _____

Dr. Anubhav Tripathi, Advisor

Recommended to the Graduate Council

Date_____ _____

Dr. Lawrence Larson, Reader

Date_____ _____

Dr. Anita Shukla, Reader

Date_____ _____

Dr. Edith Mathiowitz, Reader

Approved by the Graduate Council

Date_____ _____

Dr. Marissa Gray, Director of Biomedical Engineering Master's Program

SYNOPSIS AND IMPACT OF THIS THESIS WORK

The goal of this thesis work is to synthesize bioengineering and biotechnology fundamentals to improve access to genomics-based healthcare. To achieve this goal, I took a comprehensive approach that not only culminated in several technical breakthroughs but consisted of industry collaboration, mentoring, improving diversity, equity, and inclusion, and international, professional, and community engagement.

I. TECHNICAL BREAKTHROUGHS

This thesis work culminates in innovations in the genomics, molecular diagnostics, and sample preparation spheres that reduce NGS library preparation cost, error, and hands-on time, while increasing the accessibility and standardization of DNA sequencing and automation technologies. The automated library preparation platform, described in Chapter 2, consists of a simplistic and compact mechanical design aimed at cost-effectiveness that utilizes low-cost materials and optimizes bench space. The non-filter, 16-cannula polymer cartridge is designed to enhance laboratory sustainability by reducing filter-pipette tip plastic waste associated with lengthy and highly complex protocols such as library preparation. Assay design and engineering employ creative pre-plating and mixing procedures that optimize reagent use, enzyme efficiencies, bead-based purification steps, and minimize sample loss and hands-on plating time. Reaction heating is performed inside sealed capillaries to prevent sample loss from evaporation, cross-contamination, and index-hopping caused by free adapters post-ligation. This medium/low throughput method is desirable for smaller research and academic laboratories and offers a completely automated workflow that can be operated without highly sophisticated pipetting skills and training.

Publication: “Leveraging the fundamentals of heat transfer and fluid mechanics in microscale geometries for automated Next-Generation Sequencing library preparation” **Olivia Ott**, Sabrina Tolppi, Jennifer Figueroa Cruz, Khaliun Myagmar, Khulan Unurbuyan, Anubhav Tripathi. Submitted and under review by Nature Scientific Reports.

Publication: “Genomics in the Age of Capitalism: Is the “$200 genome” false advertising and implications for clinical researchers” **Olivia Ott** and Anubhav Tripathi. Opinion paper to Gen Biotechnology (under preparation).

Conference Presentation: “An Automated Next-Generation Sequencing Library Preparation Platform with Integrated Liquid Handling and Capillary Tube PCR” **Olivia Ott**, Jennifer Figueroa Cruz, Sabrina Tolppi, Khulan Unurbuyan, Khaliun Myagmar, Anubhav Tripathi. SLAS2023 Tony B. Poster Author, presented at the Society for Laboratory Automation and Screening International Conference and Exhibition in San Diego, February 2023 (SLAS2023).

II. INDUSTRY COLLABORATION

This thesis work is a direct result of industry collaboration with PerkinElmer-Applied Genomics and would not be possible without their funding through Brown University and their scientific insights and discussions. Throughout the course of this collaboration, I was able to interact directly with industry professionals, gain insights into the genomics field, including knowledge of genomics and laboratory automation market

needs, business insights, and learn the process of product development from initial discovery through commercialization and manufacturing. I gained an understanding of the different roles in the industry, including possible career paths. I took a leading role in this collaboration by presenting data, leading meetings, making connections, and learning from PerkinElmer. I was able to develop my leadership and soft skills and identify company strategies to incorporate diversity and accessibility into biomedical engineering while working on a project that is directly applicable to my career goals. In the end, I used these insights to design experiments to maximize this project's impact on the field.

III. MENTORING AND DIVERSITY, EQUITY, INCLUSION

This thesis work also provided research work for a diverse team of women scientists, including one undergraduate student and one master's student in Biomedical Engineering. Throughout this work, I directly mentored Sabrina Tolppi, a sophomore from Connecticut, as she began her research career with the goal of attending medical school upon graduation. We worked together on manuscript writing, programming with python, professional development, and planning and executing experiments. Furthermore, I was able to provide guidance and laboratory training to Jennifer Figueroa-Cruz, a first-year master's student from Puerto Rico. Together we approached the challenges of this project from unique perspectives, made design choices aimed at health equity, and engaged in active discussions of how to improve inclusivity in science and engineering. Not only was I able to provide my own advice and insights to Sabrina and Jennifer, but I actively learned from them and enhanced my leadership and communication skills.

IV. INTERNATIONAL ENGAGEMENT

Furthermore, this work allowed me the opportunity to collaborate and work directly with Khaliun Myagmar and Khulan Unurbuyan, visiting researchers from Mongolia. This collaboration facilitated the exchange of knowledge, culture, and ideas between scientists from vastly different backgrounds. From them, I gained clinical insights, important laboratory skills, molecular biology techniques, and a deeper understanding and appreciation of knowledge exchange. In return, I was able to provide insights into their professional development and community engagement, and I presented our research findings at a virtual summit for the Institute of Medical Sciences in Mongolia.

V. COMMUNITY AND PROFESSIONAL ENGAGEMENT

As a result of this work, I was able to actively engage with a large community of scientists in industry, genomics, automation, and analytical chemistry. In April of 2022, I attended the Mass Spectrometry & Advances in the Clinical Lab (MSACL) conference in Monterey, California alongside Ph.D. student Ramisa Fariha. In February of 2023, I presented a poster as a Tony B. Travel awardee at the Society for Laboratory Automation and Screening (SLAS2023) international conference and exhibition in San Diego, California, entitled “An Automated Next-Generation Sequencing Library Preparation Platform with Integrated Liquid Handling and Capillary Tube PCR”. These experiences allowed me to network with professionals and executives in the field, grow as a scientific leader and professional, acquire knowledge outside of my own expertise, and share my own research findings and passions.

Finally, during my first year at Brown University, I worked with Dr. Marissa Gray and Dr. Celinda Kofron as a CBME social media specialist while actively engaged in research, with the goal of interacting more closely with the Brown community and pursuing my interest in marketing and professional engagement. The purpose of this position was to increase the Center for Biomedical Engineering social media engagement, recruitment, and community awareness of events and opportunities for students, particularly master's students. During the summer of 2022, I instructed the Summer @ Brown course, *Introduction to Engineering and Design,* and engaged with high school students and their first experience with the engineering design process. As a result, I was able to engage with students and faculty outside of my own research lab and make a direct impact on the graduate program by highlighting student accomplishments, life, and opportunities.

This thesis is a result of my drive to engage with the community and bridge the gap between healthcare, science, and engineering**.**

ACKNOWLEDGEMENTS

This thesis work would not be possible without the help of so many wonderful people.

First, I want to thank my advisor Dr. Anubhav Tripathi for his mentorship and for providing me with the resources to grow into a scientific and professional leader. The opportunity to directly collaborate with industry professionals on a research project is unique, and was only made possible by your guidance, training, and leadership. Thank you for teaching me the ins and outs of science and innovation!

I would like to thank my teammates Sabrina Tolppi, Jennifer Figueroa-Cruz, Khulan Unurbuyan, and Khaliun Mayagmar for their comradery, encouragement, and inspiration throughout this work. Our accomplishments are a team effort, and I could not have succeeded alone.

I want to thank Dr. Lawrence Larson, Dr. Anita Shukla, and Dr. Edith Mathiowitz for supporting me as thesis committee members. I want to thank Dr. Marissa Gray for supporting me as the Director of the Biomedical Engineering master's program.

I would also like to thank the PerkinElmer group, Dr. Adam Snider, Christopher Tannock, Shreyas Shah, Spencer Curbelo, and Arjun Bansal for their helpful comments and scientific discussions. This thesis would not be possible without your continuous support and insights. I would also like to thank Dr. William Law for performing DNA sequencing, data-analysis and helping us troubleshoot. I want to thank PerkinElmer's grant to Brown University for enabling the research and development of this project.

Additionally, I want to thank Brown BME master's alumni Yasmin Karasu Benyes and Prutha Sameer Deshpande for taking me in and making me feel welcome in the lab. A special thank you to the entire Tripathi lab community for their valuable feedback and insights.

Finally, to my family and friends, thank you for believing in me and having faith every step of the way. Your love keeps me grounded, and for that, I am grateful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Genomics-based healthcare is a rapidly growing field, with use cases spanning molecular diagnostics and personalized medicine. Next-generation sequencing (NGS) is the underlying technology enabling innovations in genomics but remains limited by cumbersome and inefficient sample preparation. This thesis work explores an innovative bioengineering approach to develop cheaper and more accessible NGS sample preparation solutions by eliminating manual labor. The platform for automated NGS library preparation consists of a simplified mechanical design that exploits both macro- and microfluidic properties for optimizing heat transfer, reaction kinetics, mass transfer, fluid mechanics, adsorption-desorption rates, and molecular thermodynamics. The approach introduces a unique two-cannula cylindrical capillary system that integrates automatic reagent movement, mixing, and magnetic bead-based washing with capillary-based thermal cycling (capillary-PCR). The method was used to develop two library preparation assays, one incorporating mechanical and the other enzymatic DNA fragmentation. The platform successfully prepared eight libraries in parallel, generating sequencing data for both human and E. coli DNA libraries with negligible coverage bias compared to positive controls. The method achieved final library concentrations and size distributions comparable with the conventional manual approach, demonstrating compatibility with downstream sequencing and data analysis. This approach offers repeatability and consistency in the quality of DNA libraries, asserting the importance of mechanical design considerations that employ and optimize fundamental fluid mechanics and heat transfer properties.

Chapter 1

Background

1.1 IMPLICATIONS AND IMPACT OF MOLECULAR DIAGNOSTICS

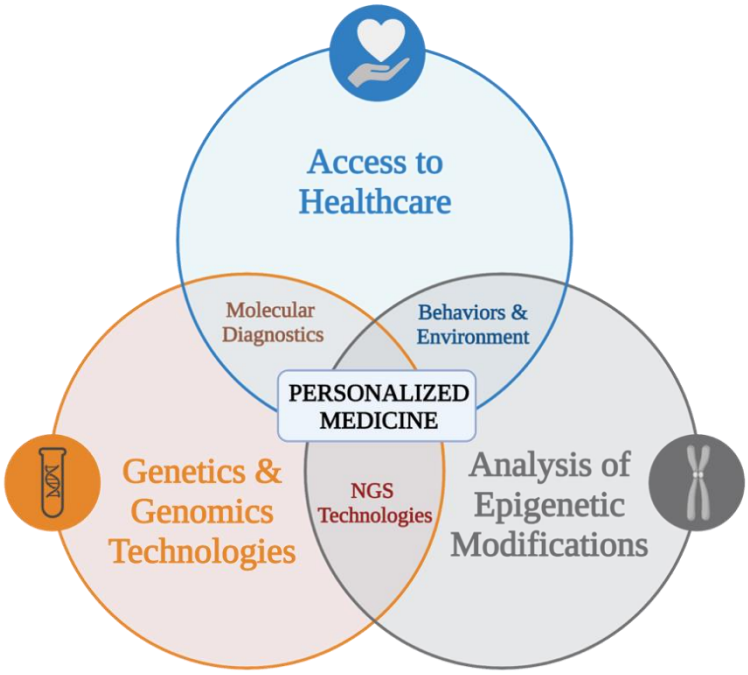
Health equity remains a key barrier to healthcare and diagnosis of disease globally and is crucial to improving global health, including women's and children's health, and pushing the medical field towards personalized medicine. The accessibility of medical and diagnostics technologies directly impacts the lives of more than 331 million people in the US[1] and 8 billion people globally[2]. The COVID-19 pandemic, beginning in December 2019, brought to light the drastic disparities in healthcare and the various bottlenecks associated with mass virus testing, emphasizing the importance of rapid diagnostics and evolving healthcare practices. According to the World Health Organization (WHO), as of January 14, 2023, there have been 6,844,267 COVID-19 related deaths globally, with low socioeconomic status and minority groups disproportionately affected in the US due to a lack of access to healthcare[34]. It is clear there is a desperate need to produce innovations in diagnostics and genomics to increase the accessibility of medical technologies, thereby enhancing the rapid and effective diagnosis of disease and ameliorating global health inequity.

Throughout the COVID-19 pandemic, genomic surveillance of the SARS-CoV-2 virus was vital to public health efforts to track and control infection, and enabled mass monitoring and identification of diagnostic biomarkers. However, surveillance of viruses such as SARS-CoV-2 would not be possible without recent advances in molecular diagnostics and genomics technologies such as Next-Generation Sequencing (NGS). NGS, particularly the availability of

fast and accurate whole-genome sequencing (WGS), enabled the initial identification of molecular targets for PCR-based SARS-CoV-2 detection assays and, subsequently, the identification of virus variants, mutations, and symptom severity predictions, in contrast to traditional virus monitoring molecular biology techniques such as RT-qPCR that only amplify specific sequence segments[56]. Obtaining the complete genomic picture of a virus is critical for identifying and tracking modes of transmission and mass transmission events[5] and will be essential for combating future pandemics amid the rising global population.

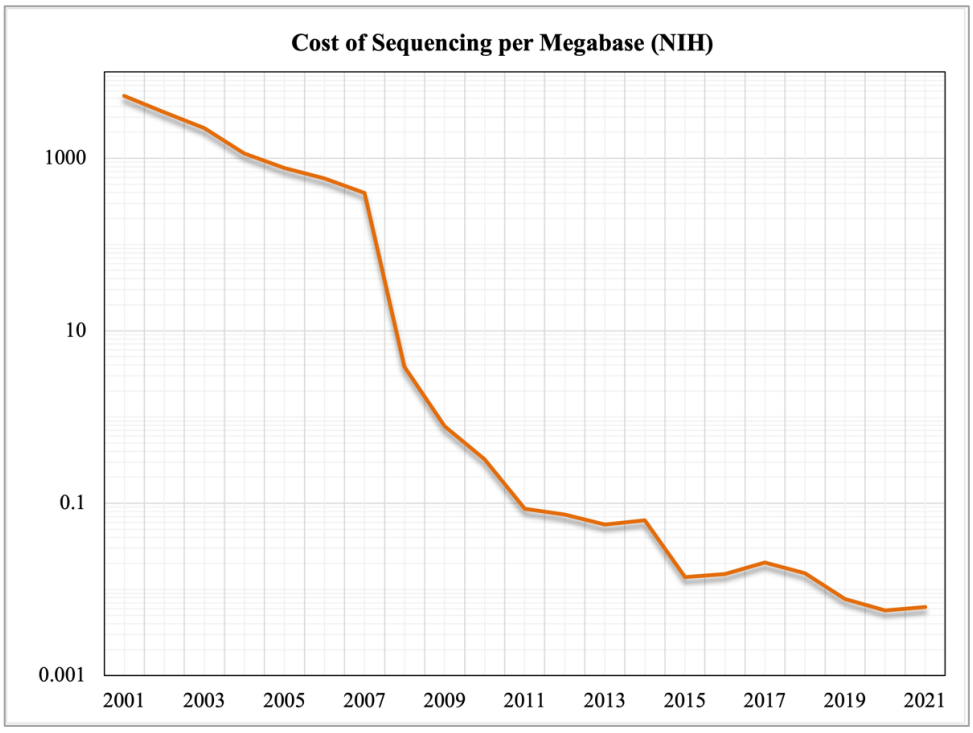
Not only is NGS vital for applications such as virus monitoring and detection, but this technology is a driver of personalized medicine for the treatment and diagnosis of gene-specific diseases, such as cancer, affecting ~1,918,030 people worldwide in 2022[7]. Genetics and genomics technologies, in conjunction with access to healthcare and analysis of epigenetic modifications, support the foundation of personalized medicine and must be integrated flawlessly to successfully assess and treat patients based on their genetic variations (Fig. 1.1)[8].

Figure 1. 1: Elements necessary for personalized medicine[8,9].



Data provided by NGS aids in the molecular detection and characterization of cancers and patient-specific mutations and can enhance the success of cancer drugs and therapeutics by aiding in the identification of efficacious drug targets[10]. This is critical, as the estimated clinical trial success rate for oncology drugs is only ~3.4%, while the median cost to develop an oncology drug is ~ $648.0 million[1112]. Furthermore, NGS technologies are essential for carrier detection, medical genetics, pharmacogenomics, prenatal diagnosis, and disease prognosis [10].

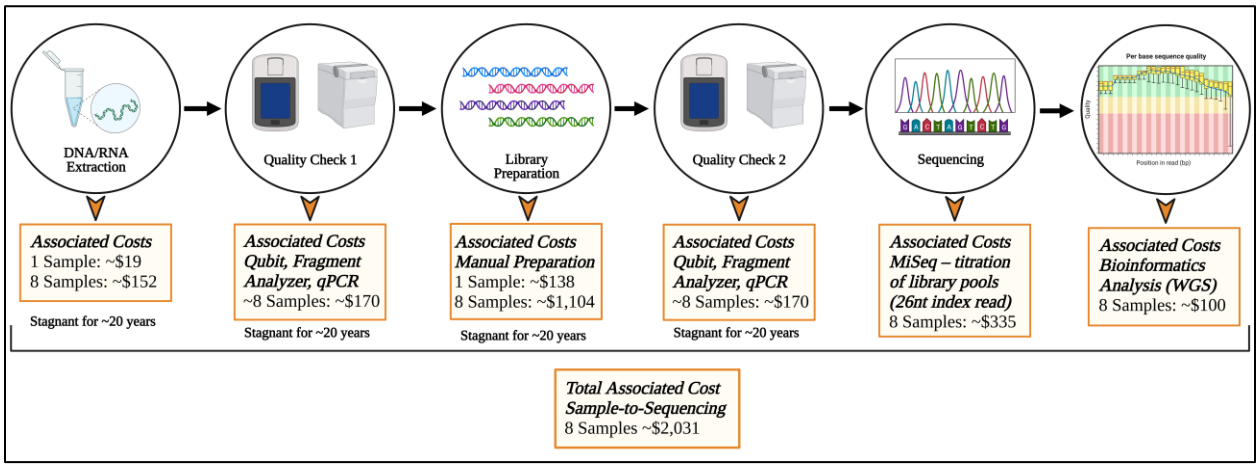
Figure 1. 2: Sequencing cost per megabase – 2021[13]. Adapted and reconstructed from the National Human Genome Research Institute[13].



While the cost of DNA sequencing per megabase has drastically decreased over time[13], making the technology more widely available and desirable (Fig. 1.2), especially for small genome WGS

applications, NGS sample preparation remains a significant barrier to the accessibility of NGS screening strategies.

Figure 1. 3: Associated Costs of the Sample-to-Sequencing Workflow[9,14–17].



NGS sample preparation costs have remained largely stagnant over the last 20 years (Fig. 1.3)[14–17]. Manual library preparation is the most expensive step (~$1,104 per 8 samples), followed by quality checks (Qubit, Fragment Analyzer, qPCR) necessary post-extraction and post-library preparation (~$170 per 8 samples)[14–17]. The requisite reagent and instrument costs of NGS sample preparation limits the widespread use of molecular diagnostics-based healthcare; exacerbated by the lack of standardization among sample preparation assays, the highly technical and wet lab skills required to manually perform them, and the requisite time and cost of labor of experienced scientists and technicians[18]. To mitigate the bottleneck of NGS library preparation, our focus was to automate the workflow and reduce associated costs via innovations in biochemical reaction engineering and device design.

As a direct result of my desire to make an impact and help advance the field toward health equity and personalized medicine, in this work, I chose to dedicate myself to the development of accessible molecular diagnostic sample preparation technologies. This thesis bridges the gap

between effective genomics-based healthcare and accessible sample preparation, relying on bioengineering fundamentals to innovate, develop, verify, optimize, and validate accessible laboratory automation and molecular diagnostic sample preparation solutions.

1.2 NEXT-GENERATION SEQUENCING FUNDAMENTALS

NGS refers to the collection of high-throughput DNA sequencing technologies developed post-Sanger method, including but not limited to sequencing-by-synthesis, ion-conductor sequencing, and sequencing-by-ligation[19]. The most employed technique, popularized by Illumina, is sequencing-by-synthesis, that takes advantage of the natural DNA replication mechanism to sequence short strands of DNA that have been hybridized to a flow cell[19]. The technology heavily relies on sequencing instruments able to produce high-quality fluorescent images and yield up to 250 million unique reads in one run and remains largely automated due to high-throughput requirements[19]. As a result, Illumina NGS has become highly desirable for almost any application requiring massively parallel data collection. Though sequencing itself is quite streamlined, with instruments able to perform entire sequencing runs in the span of 1-2 days, NGS sample (library) preparation remains a significant bottleneck to the clinical utility of this technology. The basic sample preparation workflow consists of many sequential steps: DNA or RNA extraction, nucleic acid fragmentation, reverse transcription in the case of RNA, adapter ligation, and PCR amplification[20]. The NGS workflow culminates in library denaturation and hybridization to the NGS flow cell for cluster generation, followed by sequencing[20]. The final step of the workflow includes highly sophisticated and involved data analysis to determine sequence alignment to a reference genome, identify variants (variant calling), and analyze potential AT/GC bias introduced during the library preparation[21].

Critical to the success of library preparation are the size of the adapter-ligated DNA fragments in the final library, as Illumina platforms typically require fragment lengths between 200-500 bp[22]. As a result, most workflows incorporate magnetic bead-based purifications typically post-ligation and post-PCR, to remove any unwanted products, most notably small fragments like adapter and primer dimers that will interfere with sequencing data quality[23]. Some workflows incorporate dual-ended SPRI-bead size selection, in which both large fragments (>1000 bp) and small fragments (<200 bp) are removed so users can select for specific fragment lengths[20]. There are a wide variety of options for different types of library preparation, with hands-on experiment run-time ranging anywhere from ~3 hours for PCR-free and up to ~7 hours or more for target-capture assays[24].

There are many opportunities for error throughout the workflow. Each step requires a skilled professional with excellent pipetting skills and experienced scientists able to analyze data consisting of millions of reads. Issues most commonly arise during the nucleic acid fragmentation step, as each method of DNA fragmentation (enzymatic, physical, or chemical) can introduce fragmentation bias, commonly known as AT/GC bias[21]. Physical methods like sonication can over-fragment the AT-rich portions of the genome to create short fragments, due to the weak A-T bond, that are then selected-out during the library preparation, resulting in an overrepresentation of GC content in final sequencing results and interfering with genetic analysis[21]. Fragmentation enzymes have their own implicit bias that can be exacerbated if reaction conditions are not optimal[21]. There are two basic principles underlying enzymatic fragmentation. Tagmentation is the process in which partial adapters are inserted into genomic DNA by TN5 transposase, successfully integrating fragmentation, end-repair, adenylation, and adapter ligation into a single step[21]. The other method of enzymatic fragmentation employs DNA restriction endonucleases to randomly cut genomic DNA but often suffers from bias due to enzyme cut-site preference[21]. Furthermore, AT/GC bias

can be PCR-induced, as GC rich regions are more difficult to denature and therefore may not be equally represented in the final PCR product, especially if temperature conditions are not optimal within a tight tolerance[25].

Finally, lack of standardization continues to plague NGS library preparation workflows, as each method requires kit and vendor-specific reagents. Input amount requirements vary drastically between kits and can range anywhere from 1ng to 1 µg, making only certain methods clinically relevant[20]. Moreover, inefficiencies in bead purification steps can lead to sample loss, resulting in a push for optimized workflows and incredibly precise liquid handling automation solutions[23].

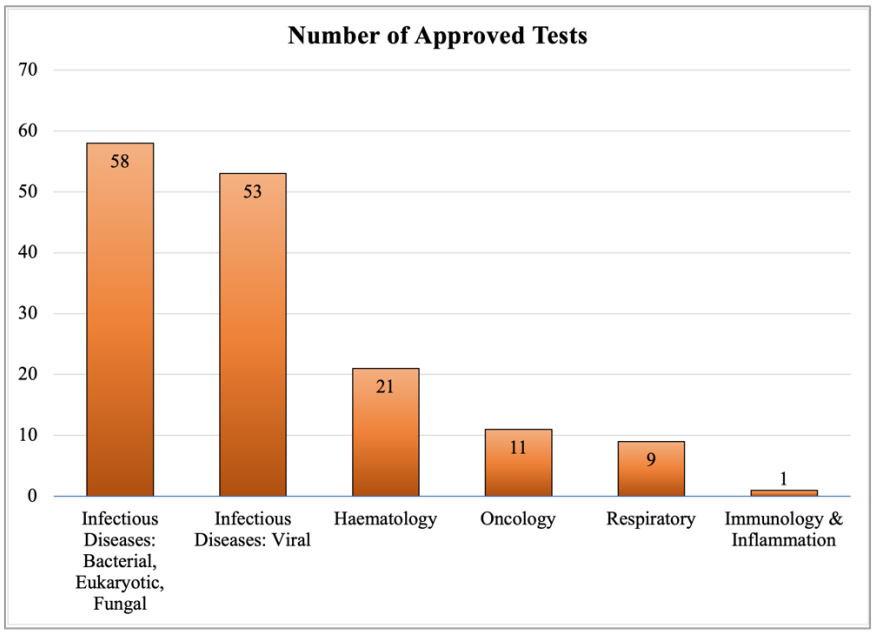
Clearly, successful sample-to-sequencing workflows require precise implementation of chemical, molecular, and bioengineering principles including fluid mechanics, molecular thermodynamics, adsorption-desorption kinetics, mass and heat transfer, and reaction kinetics. Molecular diagnostics sample preparation innovations will rely on the synthesis of each of these principles to democratize DNA-sequencing, making it more affordable and practical for clinical applications.

1.3 MOLECULAR DIAGNOSTICS AND GENOMICS MARKET NEED

Innovations in molecular diagnostics and genomics technologies are not only expected to directly impact the healthcare market and its buyers (patients), but the molecular diagnostics market itself is expected to grow at an annual growth rate of 10-15%[26], with infectious disease detection accounting for the largest market share (about 50%)[2627]. The key market drivers are a direct result of the pandemic as researchers have been pressured to create rapid and accurate pathogen detection strategies[27]. NGS technologies account for a large portion of this market as

DNA sequencing is becoming more affordable, headed by the launch of Illumina's NovaSeq platform and associated reagent kits that can reduce the cost of sequencing by up to 70%[27].

Figure 1. 4: Molecular tests approved in the United States[26]. Adapted and reconstructed from Morel et al.[26].



The oncology sector is also expected to grow, with 11 approved molecular tests in the US (Fig. 1.4) and is predicted to be a future driver of this market due to the global push for personalized medicine[26]. The global genomics market is also expected to grow and is predicted to reach $94.9 billion by 2030 at a CAGR of 16.4%[28]. The key market drivers are government and research organization support to analyze and understand whole genome data to better treat and diagnose cancer, genetic disorders, and understand molecular epidemiology and pathogen genomics[28]. Finally, the laboratory automation market is expected to surpass $7,194.1 million by 2030, driven by recent collaborations between key players because of the need for COVID-19 testing automation solutions[29].

Molecular diagnostics, genomics, and laboratory automation market analyses reveal a significant opportunity at the intersection of the three fields. New technologies that meet the rising global need for rapid and accurate disease/pathogen detection, incorporate the decreased requisite cost per genome of WGS, and reduce human error, while simultaneously enhancing the accessibility of genomics-based healthcare, will dominate the field and drive market growth in all three areas. This is a call for researchers to exploit this opportunity and make a direct impact on human health.

Chapter 2

Leveraging the fundamentals of heat transfer and fluid mechanics in microscale geometries for automated Next-Generation Sequencing library preparation

---

Leveraging the fundamentals of heat transfer and fluid mechanics in microscale geometries for automated Next-Generation Sequencing library preparation

Olivia Ott, Sabrina Tolppi, Jennifer Figueroa Cruz, Khaliun Myagmar, Khulan Unurbuyan, Anubhav Tripathi, submitted to Nature Scientific Reports, February 2023.

2.1 ABSTRACT

Next-generation sequencing (NGS) is emerging as a powerful tool for molecular diagnostics but remains limited by cumbersome and inefficient sample preparation. We present an innovative automated NGS library preparation system with a simplified mechanical design that exploits both macro- and microfluidic properties for optimizing heat transfer, reaction kinetics, mass transfer, fluid mechanics, adsorption-desorption rates, and molecular thermodynamics. Our approach introduces a unique two-cannula cylindrical capillary system connected to a programmable syringe pump and a Peltier heating element able to execute all steps with maximum efficiency. Automatic reagent movement, mixing, and magnetic bead-based washing with capillary-based thermal cycling (capillary-PCR) are completely integrated into a single platform. The manual 3-hour library preparation process is reduced to less than 15 minutes of hands-on time via optimally pre-plated reagent plates, followed by less than 6 hours of instrument run time during

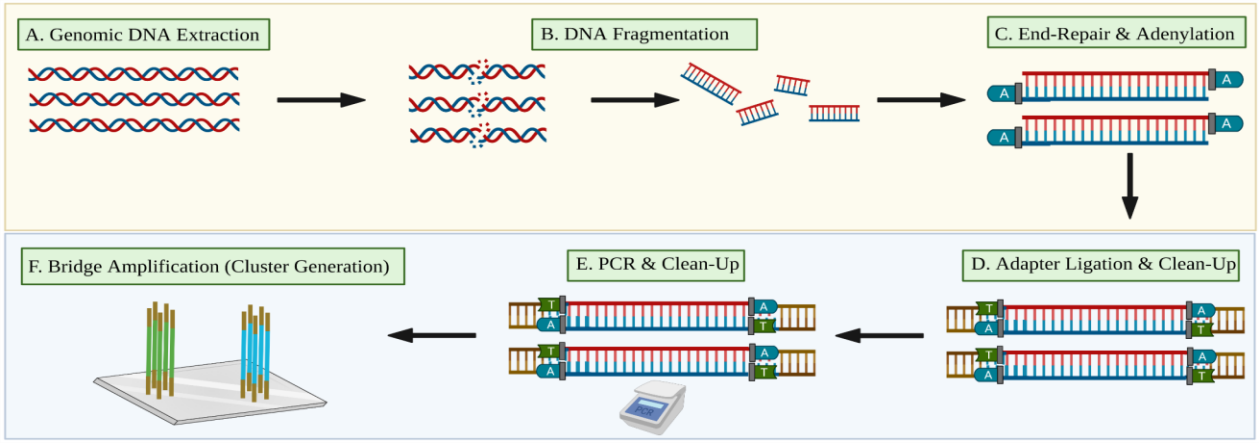
which no user interaction is required. We applied this method to two library preparation assays with different DNA fragmentation requirements (mechanical vs. enzymatic fragmentation), sufficiently limiting consumable use to one cartridge and one 384 well-plate per run. Our platform successfully prepared eight libraries in parallel, generating sequencing data for both human and E. coli DNA libraries with negligible coverage bias compared to positive controls. All sequencing data from our libraries attained Phred (Q) scores >30, mapping to reference genomes at 99% confidence. The method achieved final library concentrations and size distributions comparable with the conventional manual approach, demonstrating compatibility with downstream sequencing and subsequent data analysis. Our engineering design offers repeatability and consistency in the quality of sequence-able libraries, asserting the importance of mechanical design considerations that employ and optimize fundamental fluid mechanics and heat transfer properties. Furthermore, in this work, we provide unique insights into the mechanisms of sample loss within NGS library preparation assays compared with automated adaptations and pinpoint areas in which the principles of thermodynamics, fluid mechanics, and heat transfer can improve future mechanical design iterations.

2.2 INTRODUCTION

Next-generation sequencing (NGS) has transformed the biomedical landscape and ushered in a new age of discovery in genomics, significantly impacting disciplines such as cancer research[30–33], rare disorder diagnosis[34–36], and prenatal screening[37]. NGS is a markedly valuable tool for diagnostics and discovery because it generates millions of reads and is exceedingly sensitive to nucleic acid mutations[38–41]. Unfortunately, while innovations in sequencing technology have reduced requisite cost, time, and workload, sample (library) preparation has largely remained rudimentary and labor-intensive[23,42–44]. Even for applications requiring less data, such as microbial whole genome sequencing (WGS) or RNA-seq, the majority of the total sequencing cost can be attributed to library preparation, a critical upstream process that ultimately determines the success or failure of a sequencing run[45]. A typical NGS library preparation begins with the extraction of genomic DNA (gDNA)[46] (Fig. 2.1A) followed by mechanical or enzymatic fragmentation (Fig. 2.1B). Sufficiently large yield from nucleic acid extraction is required to ensure high library complexity, while a fragmentation method that does not introduce bias is critical[21,47]. The dsDNA fragments must then be ligated to indexed adapters or primers that will eventually enable hybridization to the NGS flow cell for cluster generation (Fig. 2.1F). To achieve this, the ends of the fragments are blunted, adenylated, and 5' phosphorylated (Fig. 2.1C)[44]. A ligase enzyme is then able to attach the indexed adapters to the A-overhangs (Fig. 2.1D). This reaction generates unwanted by-products, known as adapter-dimers, when adapters are ligated to one another rather than to the DNA inserts. Adapter-dimers can be problematic since they will readily hybridize and amplify on the sequencing flow cell, taking the place of DNA inserts and providing useless sequencing data. A post-ligation purification step is thus typically performed using solid phase reversible immobilization (SPRI), a method that employs paramagnetic beads

(MBs) to selectively bind larger DNA fragments[44]. Purification steps are sensitive and time-consuming, requiring large-volume washes and pipette mixing. Following purification, adapter-ligated DNA fragments are amplified via a polymerase chain reaction (PCR) to generate input library concentrations of >5 ng/µL for the NGS flow cell (Fig. 2.1E). A post-PCR purification step is necessary to remove unwanted primer and adapter dimers, nucleotides, and any DNA fragments less than 200 bp in length[23, 44].

Figure 2. 1: A typical NGS library preparation workflow. (A) Genomic DNA is extracted from blood or another sample matrix, (B) gDNA is mechanically, enzymatically, or chemically fragmented, (C) DNA fragments are blunted, adenylated, 5' phosphorylated, and (D) ligated to indexed adapters. Primer-dimers are removed in a SPRI bead purification step. (E) Adapter-ligated DNA fragments are PCR amplified and a second SPRI bead purification step removes fragments < 200 bp. (F) Final libraries undergo bridge amplification for cluster generation on the NGS flow cell.[9]



It is clear that NGS sample preparation requires many sequential steps using various pieces of laboratory equipment. Hence, NGS library preparation is expensive, cumbersome, and susceptible to contamination and pipettor error[48–51]. For small clinics and genomics laboratories, the use of NGS continues to hinge on the success of error-free manual library preparation done by an experienced scientist. Contamination remains the most persistent issue, but the risk can be mitigated by reducing human interaction with both reagents and samples via automation solutions.

Most automated NGS library prep methods focus on the implementation of robotic liquid handling for accurate pipetting of low and high-viscosity liquids[52], integration of thermal cycling units, and inclusion of magnetic capabilities enabling bead-based purification and size selection. Liquid handlers often rely on the one-to-one adaptation of manual processes using robotic arms, demanding additional hardware, such as plate grippers, motion sensors, pipette tip collectors, firmware controllers, etc., thereby significantly increasing the cost of library preparation per sample. In addition to substantial initial investment costs, liquid handling workstations require expert programmers to create scripts for the purpose of adapting new workflows onto specific platforms and workarounds to combat reagent evaporation and tackle high-viscosity liquids. As a result, pipetting workstations are suitable only for high throughput applications that require massively parallel data collection. It is important to note that robotic liquid handling systems generate sizable amounts of plastic waste (pipette tips, racks, and 96-well plates), which is not a sustainable laboratory practice[53].

Microfluidic approaches are another group of solutions suitable for low-throughput and single-use library preparation applications[23, 44, 50, 54–57]. They require a fraction of the investment cost of robotic liquid handlers and completely avoid any contamination concerns but lack scalability and manufacturing consistency. Applicable only at the point of care, single-use

microfluidic chips harbor risks for the environment, producing consequential waste and requiring safe disposal if toxic or dangerous chemicals are incorporated into chip design. Consequently, expensive robotic liquid handlers dominate research and clinical applications in core laboratories, and small laboratories suffer from the bottleneck of labor-intensive manual protocols. Automation innovation for small laboratories is necessary to accelerate NGS-based research, eliminate sources of human error, and reduce hands-on time[23,42,4344].

This work employs an innovative engineering design approach to develop a platform that integrates fluid mechanics, heat transfer, molecular thermodynamics, mass transport, and biochemical reaction kinetics. The platform utilizes a capillary cartridge that automates reagent movement, mixing, and magnetic bead-based washing with innovative capillary-based thermal cycling (capillary-PCR) that optimizes heat transfer and reaction kinetics, and limits sample loss. It also utilizes a reagent-plate layout in which the required reagents, stable at -20C, are optimally pre-plated into wells of a 384 deep-well plate. As a result, the manual 3-hour library preparation process is reduced to less than 15 minutes of hands-on time from the start of the procedure, followed by less than 6 hours of instrument run (walk-away) time. This approach allowed us to successfully prepare eight libraries in parallel in under 6 hours and obtain high-quality sequencing data for both human and E. coli DNA with negligible coverage bias. We applied this method to two library preparation assays with different DNA fragmentation methods (one mechanical and one enzymatic), sufficiently limiting consumable use to one cartridge (16 cannula or ‘pipette’ tips) and one 384-well plate per eight libraries. Our design overcomes the limitations of large pipetting workstations and microfluidics platforms with its simplified and low-cost design, enclosed sample environments to minimize sample loss via evaporation, limited plastic consumable waste, and repeatability and consistency in the quality of sequence-able libraries.

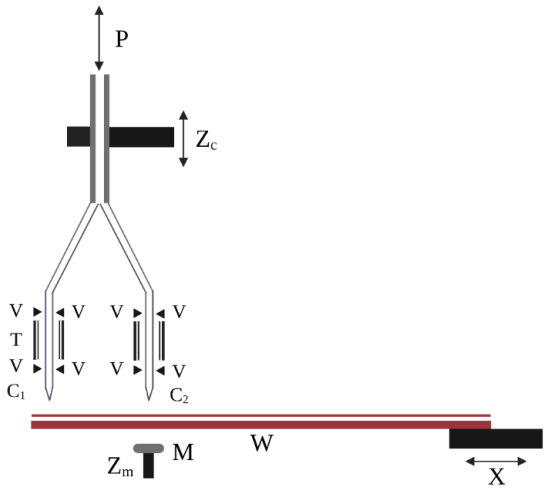
2.3 RESULTS AND DISCUSSION

VI. OPTIMIZATION OF MECHANICAL DESIGN

The principal innovation of our method is the simplified design layout of the platform (Fig. 2.2) based on the fundamental needs and constraints of fluidics, reactions, adsorption-desorption, and mixing. All NGS library preparation workflows (Fig.1) require several reagents and buffers at working volumes between 2µL – 200µL. Unlike the conventional use of 96-well or 384 shallow-well plates, we found that it is optimal to use a 384 deep-well plate with the well dimensions: 3.48mm(w) X 3.48mm(w) X 19.3mm(L) and a working volume between 2µL – 200µL. Narrow dimensions of wells offer faster diffusional ($\tau_d \sim \frac{w^2}{2D}$) and convective ($\tau_c \sim \frac{w^2 L}{Q}$) mixing. Here, D is the molecular diffusivity, and Q is the flow rate of liquid dispensing. A deep-well plate also offers many options to store repeat-use regents (such as cartridge surface blocking solution or PCR reagents) in multiple wells to avoid any carry-over contamination. Our design strategy allows users to use pre-plated reagents for eight samples (Rows A, C, E, G, I, K, M, O) in the 384-well plate stored at -20^{o}C. The platform enables movement of the 384-well plate in only a single dimension (x-stage for left-right motion). A separate z-stage (for up-down motion) that holds two cannulas (Fig. 2.2) able to move together perpendicular to the x-stage for liquid transfers from any of 24 wells in a single row of the 384-well plate, lends itself to the cost effective and robust design of our device. The two cannulas are connected via plastic capillary tubing which pass through a thermal heating section and a set of two pinch valves. The pinch valves are used to select one of two cannulas by pinching the undesired cannula to prevent air and liquid movement in that tubing. Hence, a single syringe pump is required for two tube paths that are joined via a Y-junction. Our platform offers the option of two separate locations (two cannulas) for various reactions (Fig. 2.1) in the library preparation workflow. This uniquely eliminates both inhibition of enzymatic

efficiency and promotion of side reactions due to the adsorption of previously used reagents onto the capillary walls. Our device uses a copper plate that provides spaces for the flexible polyethylene tubing to fit to maximize the thermal heating efficiency. Around one-third of the tubing directly contacts the copper surface and is held in place by an insulated door to make a tight fit so that heat can be transferred efficiently between the copper plate and the reagents within the tubing. Lastly, our device uses a permanent magnet that can be moved in proximity to the base of the 384-well plate for bead-based purification of DNA. The magnet is mounted to a separate z-stage for up-down motion (Fig. 2.2). To summarize, the psychology behind the design optimization was two-fold: (a) to keep the hardware simple, robust, and low-cost; (b) to use optimal fluidic and reaction time scales for integrating various steps needed in any library preparation workflow (Fig. 2.1).

Figure 2. 2: Schematic of our method: P = syringe pump, T = thermal heating and cooling, V= pinch valves, Zc = Z stage for cannula motion, Zm = Z stage for magnet motion, C1 = Cannula 1, C2 = Cannula 2, X = X-Stage, W= 384 deep well reagent plate, and M = Magnet for bead-based purification.[9]



VII. FUNDAMENTALS OF ASSAY DESIGN AND OPTIMIZATION OF VARIABLES

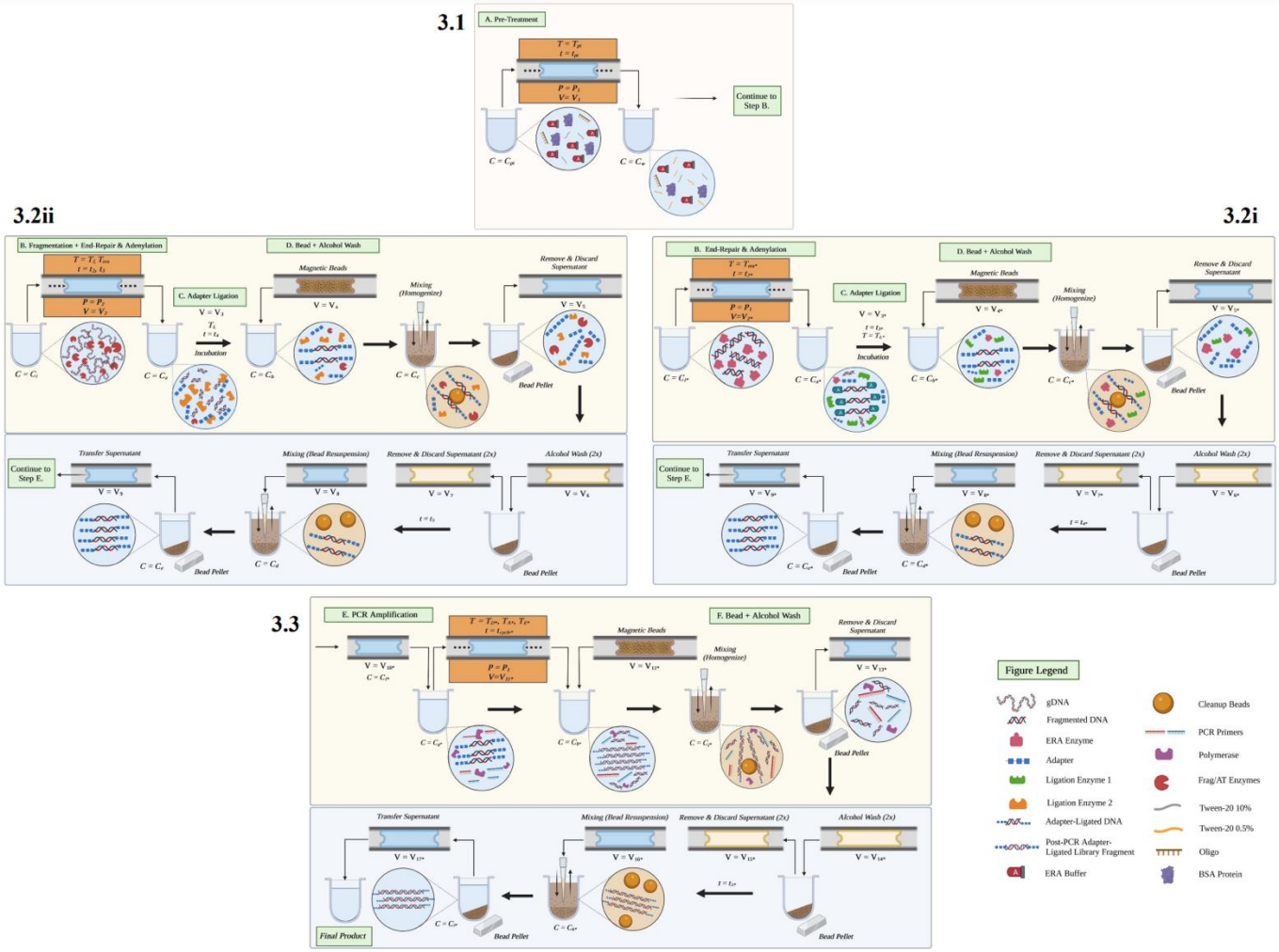
Our unique approach amalgamates multiple microfluidic and macrofluidic steps into one cohesive and comprehensive workflow. Unlike most liquid handling platforms, our design requires only two cannula tips per sample (8 tips in two parallel columns correlating to 8 samples) and corresponding capillary tubes to perform all liquid transfers and mixing functions for the entire library preparation assay. We have optimized the use of the capillary tubes based on adsorption-desorption characteristics and by incorporating a required cartridge surface passivation step. Pre-treatment (PT), fragmentation, and end-repair and adenylation (ERA) (Fig. 2.3.1, 2.3.2iB, 2.3.2iiB) are performed using one set of cannula tips, while adapter ligation, PCR, and both purification steps (Fig. 2.3.2iC-D, 2.3.2iiC-D, 2.3.3E-F) are performed using the other set of cannula tips. This ensures that each step of the assay avoids contamination from leftover enzymes, proteins, buffer components, or any other interfering reagents. PT is a required pre-assay surface (cannula and capillary) passivation step that must be completed for every new run. We performed experiments to optimize PT conditions using fluid slug volume, slug routing path, temperature, and incubation time as variables. Under optimal conditions, the PT reaction is mixed inside the 384-well plate, via up and down fluid motion in the cannula tips, before the entire reaction volume (V_1 = 50 µL) is transported inside the capillaries to the heating element, undergoes heating for t_{PT} = 35 minutes at T_{PT} = 75 °C, and is passed through the y-junction of the cartridge and discarded (Fig. 2.3.1A). This step successfully prevents the adsorption of enzymes and DNA to the polymeric surface of the cartridge tubing.

We optimized the fragmentation and ERA reaction volumes to be $V_2 = V_{2*}$ = 50 µL for both assays (Fig. 2.3.2iB, 2.3.2iiB). These reactions contain input DNA, buffer, and the appropriate fragmentation or ERA enzymes. Crucial mixing of this step is performed in the 384-

well plate using a custom mixing technique, and the fluid slug is transported in the capillaries to the heating element to maintain optimum enzyme activity at the desired temperature (room temperature incubation during the PT step then t_{2*} = 30 min at T_{era*} = 65℃ for mechanical fragmentation, Fig. 2.3.2iB and t_2 = 15 min at T_f = 37 ℃, t_3 = 30 min at T_{era} = 65 ℃ for enzymatic fragmentation, Fig. 2.3.2iiB). Note that the fluid slug is pressurized to applied pressure P = P_1 during heating to constrict fluid motion, eliminate any significant evaporation, and further optimize heat transfer. Thereafter, the fluid slug is transported to a specified well in the plate containing adapters and ligase (Fig. 2.3.2iC, 2.3.2iiC). The ligation reaction is mixed with a custom mixing technique and incubated at room temperature (T_L = T_{L*} = 25 ℃) for time t_4 = t_{3*} = 15 min. The adapter-ligated DNA sequences must then be purified to eliminate any non-ligated adapters and leftover enzymes. In this step, the liquid containing the adapter-ligated DNA fragments is transferred to a well containing SPRI beads (V_3 = 60 µL, V_{4*} = 100 µL) (Fig. 2.3.2iD, 2.3.2iiD). Several custom mixing techniques form the basis of the SPRI bead-binding and DNA elution reactions, with gentle alcohol washes in between. The eluate of the wash (purification) step is mixed with PCR master mix and primers, then transported in the capillaries to the heating element to undergo thermal cycling for amplification (V_{10*}, Fig. 2.3.3E). The fluid slug is again pressurized to optimize heat transfer, and after performing the PCR cycle for a desired amount of time (specified in S. Table 1-2.3), the slug containing sample DNA is dispensed back into the plate for the final purification step. Post-PCR purification follows the same principles as post-ligation purification, with the goal of removing adapter and primer dimers, as well as small adapter-ligated DNA fragments < 200 bp (Fig. 2.3.3F). The final supernatant removal in this step containing the desired DNA library is dispensed as the final product in a specified location on the plate, ready for

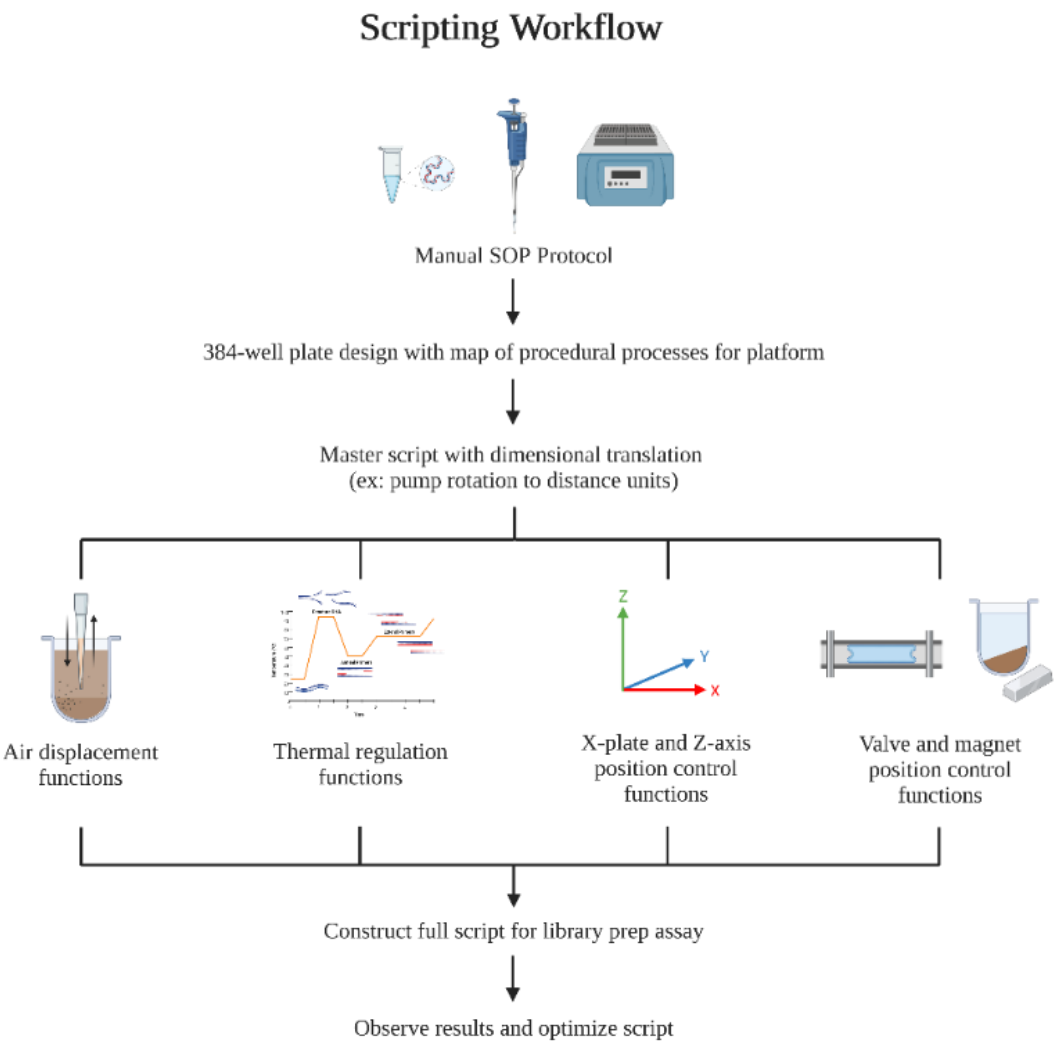
quantification and quality control (V_{17*} = 30 µL for the mechanical fragmentation assay and V_{17*} = 20 µL for the enzymatic fragmentation assay).

Figure 2. 3: Details of our optimized method. 3.1 (A) Surface passivation (Pretreatment) of the cartridge tubing. 3.2i-Mechanical fragmentation assay (B) End repair and adenylation reaction including heating, (C) adapter ligation reaction, (D) post-ligation SPRI bead purification. 3.2ii-Enzymatic fragmentation assay (B) Fragmentation, end repair and adenylation reaction including heating, (C) adapter ligation reaction, (D) post-ligation SPRI bead purification. 3.3 (E) Capillary PCR amplification, (F) post-PCR SPRI bead purification. Variables T = temperature, t = time, P = pressure, V = volume, and C = concentration.[9]



Conditions and variables were optimized for automation following a set process involving many experimental iterations. Assay development begins with manually performed variations of the SOP protocol, followed by plate mapping and reaction fine-tuning to adapt variables such as reagent volumes, reaction temperatures, and incubation times to our platform (Fig. 2.4). A master script, using Python as a programming language, establishes functions for air displacement, thermal regulation, X-plate control, Z-axis control, valve control, and magnet control that supplementary scripts may call on for development of each individual assay (Fig. 2.4). Following assay script scaffolding, liquid testing is performed to identify potential failures or further points of iteration. The local script can be partitioned to isolate specific steps of the assay, enabling the optimization of each step and function as individual segments. The script is then incrementally and experimentally adjusted until an optimized, robust, and automated assay is achieved.

Figure 2. 4: Scripting workflow. Standard operating procedure (SOP) for script and assay development.[9]



Many of these optimizations aim to refine the fluidic properties of our platform, with mixing techniques as a primary focus. Beyond accomplishing straightforward liquid transfers, the platform must be capable of homogenizing reagents of many distinct viscosities. Each reaction in an assay requires mixing techniques that take mixture volume, texture, and sensitivity to harsh pipetting into account. A prototypical fluid mixing with repeated up-down pipetting is generated

in a supplementary script, then tested with dye, water, and beads that provide a visual indication of homogeneity. With the intended reaction application in mind, the technique is assessed visually, then velocity, z-axis positioning, as well as volumes aspirated and dispensed are adjusted until the technique achieves a homogenous mixture. Finalized mixing techniques may be exported into the master script as functions for ease of future use in assay development. Many of these techniques depend on the precise displacement of fluid, which has been verified using fluid transfer tests and analyzed for consistency and volumetric accuracy (S. Fig. 1).

VIII. ANALYSIS OF CAPILLARY HEAT TRANSFER, PRESSURIZATION, AND DIFFUSION

All reactions requiring heating are performed inside the cartridge capillaries in the form of cylindrical fluid slugs. For example, during enzymatic fragmentation, the 50 µL reaction mix is transported to the heating element at a specified pump velocity and subject to applied air pressure and temperature cycling: t_2 = 15 min at T_f = 37 ℃ and t_3 = 30 min at T_{era} = 65 ℃, followed by a 4 ℃ hold (Fig. 2.3.2iiB). The maximum conductive heat flow rate through the cylindrical tube wall (Q) during the reaction is estimated

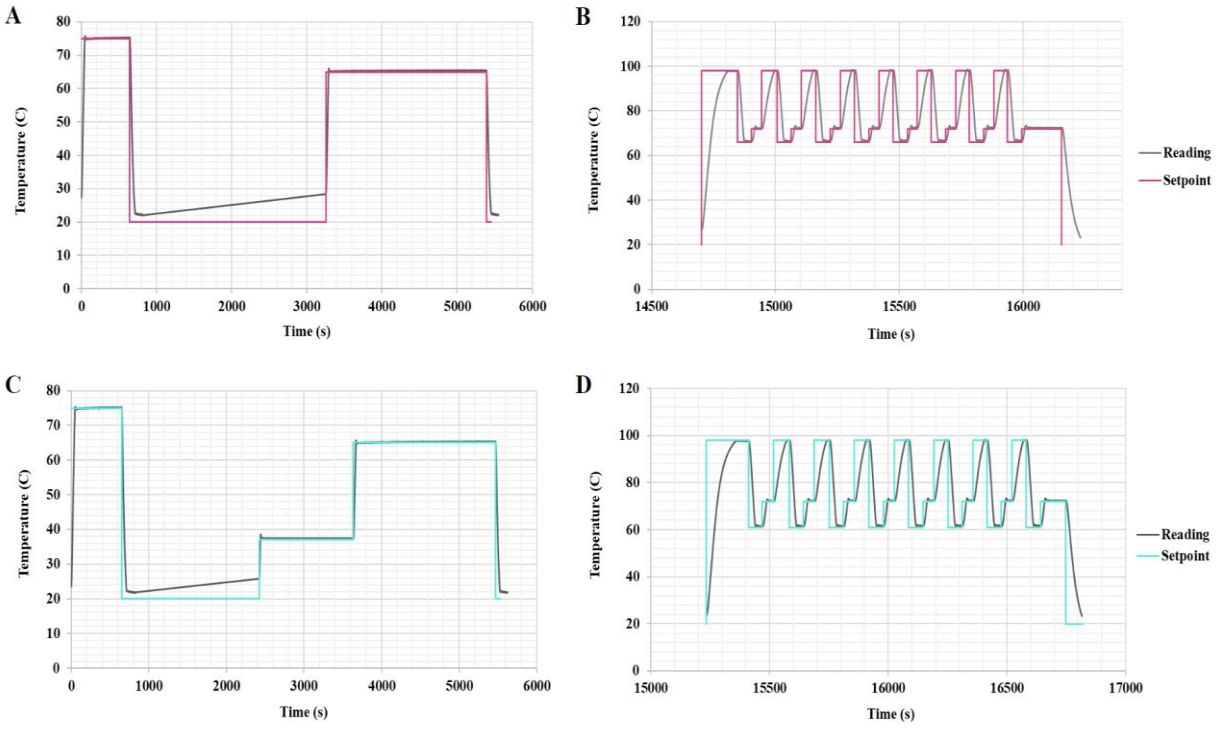
$$Q = 2\pi krL\left(\frac{\Delta T}{\Delta r}\right) \sim 7.8\ W \qquad (1)$$

when the thermal conductivity[58] of the polyethylene tube wall is $k \sim 0.34\ \frac{W}{moC}$, the thickness of the tube wall is $\Delta r \sim 0.8\ mm$, the maximum temperature difference between the outer and inner surfaces of the capillary is $\Delta T \sim 73$ ℃, the tube radius is $r \sim 1.6\ mm$, and length of the slug is $L \sim 24.9\ mm$. Similarly, the maximum conductive[59] heat flow rate inside the fluid slug itself is $Q \sim 2\pi krL\left(\frac{\Delta T}{\Delta r}\right) \sim 6.8$ W, where $r \sim 0.8\ mm$ and $k \sim 0.598\ \frac{W}{moC}$. Hence, the heat flow rates are approximately matched to achieve a rapid steady state between the tube and fluid slugs. The cylindrical shape of the fluid slug itself accelerates heat transfer,

$$t_d \sim r^2/2\alpha \sim 2.2 \text{ sec} \quad (2)$$

owing to microscale diffusional length. Here, $\alpha = 0.14558 \frac{mm^2}{s}$ is the thermal diffusivity[59] of the liquid, t_d is the thermal diffusion time, and $r = 0.8$ mm is the radius of the fluid slug. We purposely designed the Peltier heating element with a thermocouple embedded in a segment of cartridge tubing to mimic the thermal environment of the sample. Hence, our system logs temperature in real-time and compares measured temperatures against pre-scripted setpoints for reactions that require heating. A PID feedback mechanism ensures that the temperature of the fluid slug reaches the setpoint. The data, collected and stored in a text file, enables monitoring of system functionality and constant improvement of the device. Figure 5 shows the thermocouple temperature data compared against the scripted temperature setpoints for PT, fragmentation, ERA, and PCR during the mechanical fragmentation assay (Fig. 2.5A-B) and the enzymatic fragmentation assay (Fig. 2.5C-D). Statistical analysis following each run allows us to observe thermal profiles, calculate ramp rates of the heating element, and target specific modes of failure when developing new library preparation assays.

Figure 2. 5: Comparison of thermocouple temperature readings and scripted setpoint temperatures vs. time for the duration of the mechanical fragmentation assay (A-B) and the enzymatic fragmentation assay (C-D). (A) Temperature vs. time for PT and ERA, and (B) temperature vs. time for PCR amplification. (C) Temperature vs. time for PT, fragmentation, and ERA, and (D) temperature vs. time for PCR amplification.



Pressurization of the capillary during heating serves a secondary purpose of counteracting a common mechanism of sample loss in conventional liquid handling systems: evaporative loss. Sample and reagent evaporation is a persistent issue with platforms that use a 96-well plate apparatus for thermal cycling, often due to difficulties with conventional plate seals[60]. Logically, the relationship between temperature during thermal cycling and the amount of sample that is lost as evaporate can be studied using the Clausius Clapeyron equation[23]

$$\ln(P) = -\frac{\Delta H_{vap}}{RT} \qquad (3)$$

as it directly relates temperature to vapor pressure. Here, P is the vapor pressure, ΔH_{vap} is the molar enthalpy of vaporization, and R is the ideal gas constant. Extending the analysis, vapor pressure and applied atmospheric pressure are correlated in terms of liquid boiling point, as this occurs when the two pressures are equal. When the capillary is pressurized, the applied force restricts the Brownian motion of molecules at the liquid surface, resulting in less vibration, increasing ΔH_{vap}, and making a phase transition from liquid to gas energetically unfavorable even at higher temperatures. Given the importance of maintaining pressure for preventing evaporation, improving heat transfer, and achieving volumetrically accurate liquid handling, we conduct frequent analysis of system performance using a pressure sensor-equipped cartridge. This evaluation involves moving the syringe assembly to various positions to simulate high-pressure and vacuum conditions. Pressure is plotted over time (S. Figure 2) to acquire leak rates of the system and isolate specific airlines, syringes, or valves that may cause abnormal pressure readings. Optimized heat transfer results in refined kinetics for enzyme-dependent reactions, yielding enzyme efficiencies and biases that are comparable to the conventional manual method. PCR-induced bias, a common mode of failure in traditional automated platforms, is often a result of poor temperature and ramp rate regulation[61]; a problem bypassed via our unique thermal design and feedback system.

IX. ANALYSIS OF FLUID MECHANICS

When referring to positive controls, the amount of liquid lost during manual pipetting is assumed to be insignificant. Hence, our design is required to offer automated but reliable liquid transfers. However, automated liquid transfers typically result in inefficiencies at low volumes. Instead of adopting the expensive and very sophisticated pump assembly used by conventional

liquid handlers, our design relies on a simple air-pressure syringe pump. Our analysis of volumetric accuracy and %CV (S. Fig. 1) suggests that only volumes less than 5µL resulted in a %CV higher than 10%. The error present is attributed to the limited accuracy and precision of the linear-rail motor driving syringe movement at small fractions of a shaft rotation. To mitigate this limitation, our platform is programmed to always transfer volumes greater than 5 µL, while the 384-well plate layout is designed so that low-volume reagent wells become liquid transfer destinations rather than liquid transfer sources. This method takes advantage of the pre-plating process for handling of critical low-volume reagents such as adapters, primers, and enzymes, avoiding on-deck volumetric loss.

X. V. LIBRARY PREPARATION EFFICIENCY

Library preparation efficiency is an important metric to quantify the success of our automated method compared to controls. Efficient library preparation has implications for sequencing success, clinical relevance, and conservation of resources such as cost and requisite input sample volume.

Controls: We performed several controls with each experiment to verify successful library preparation. Manual positive controls consisting of 20 ng of sheared DNA (E. coli and human, average 300 bp) were constructed for genomic libraries intended for whole genome sequencing (WGS), to confirm the validity of our on-platform mechanical fragmentation assay. Successful positive controls indicate that input DNA and reagents are within the concentrations and shelf life specified for the kit. Electropherograms and gel images illustrate the final library fragment size distribution for each sample and positive control (Fig. 2.6A-D), with positive control average fragment size ~ 400 bp for human libraries and ~ 420 bp for E. coli libraries. Positive control

average yield and percent adapter-dimer were 912.22 ±168.13 ng and 0.15 ± 0.00040 % for human DNA and 445.69 ng and 0.10% for E. coli DNA, respectively (S. Table 3A). This data is consistent with our passing criteria: yield ≥ 250 ng total, concentration ≥ 5 ng/μL, adapter-dimer ≤ 5%, and fragment size distribution between 200 and 1000 bp, with an average close to 400 bp. A manual negative control was performed without input DNA, to evaluate any potential reagent contamination, while another on-platform negative control without DNA input assessed potential system contamination (Fig. 2.6G). The negative control electropherogram overlay shows a single peak between the upper and lower markers around 150 bp corresponding to adapter-dimer, consistent with the lack of input. No erroneous peaks were observed, indicating that the reagents and system are contamination-free.

Manual positive controls consisting of 50 ng gDNA (E. coli and human) were constructed using the enzymatic fragmentation assay intended for WGS. Supplementary Table 3B shows the yield and percent adapter-dimer for each sample and positive control corresponding to the enzymatic fragmentation assay and either human or E. coli gDNA input. Electropherograms and gel images further confirm that our positive control libraries pass the criteria in terms of the size distribution (Fig. 2.7A-D) (average fragment size ~ 350 bp for both human and E. coli DNA). Positive control average yield and percent adapter-dimer were 3132.4 ± 718.14 ng and 0% for the human DNA libraries and 1459.3 ± 3.54 and 0% for the E. coli DNA libraries, respectively (S. Table 3B), consistent with passing criteria. Manual and on-platform negative controls were performed without input DNA to assess reagent and platform contamination. No erroneous peaks were observed; the electropherogram overlays shown in Fig. 2.7G highlight the presence of a single adapter-dimer peak as expected. Thus, we concluded that the reagents and platform were not contaminated.

Libraries from Escherichia coli bacterial DNA: We evaluated the size distribution of libraries generated from our device and compared them to positive controls to gauge the effectiveness of our method for preparing sequencing-ready E. coli DNA libraries. Figures 6C and 7C show broad bands in the ~200-1000 bp region, corresponding to the size distribution of adapter-bearing, indexed, DNA fragments. The average yield and % adapter-dimer for on-platform E. coli libraries generated by the mechanical and enzymatic fragmentation assays were 610.245 $\pm$ 128.93 ng and 0.44 $\pm$ 0.0041%, and 960.97 $\pm$ 176.47 ng and 0.28 $\pm$ 0.0056 %, respectively. An additional peak at ~150 bp corresponds to adapter dimers that formed during ligation[44]. The lack of erroneous peaks highlights the success of the post-PCR SPRI bead purification step in removing primer dimers. Indexed fragments mainly fell within the desired 200-1000 bp size range, though a small, broad hump in the baseline past 1000 bp for the enzymatic fragmentation assay (Fig. 2.7C) suggests that some high molecular weight products remained post-fragmentation. However, the presence of this same hump in the positive controls (Fig. 2.7C, Samples 7-8) signifies the success of our method in replicating results obtained from the gold-standard practice. Conversely, the broad peak past 1000 bp is also present in the mechanical fragmentation assay electropherograms (Fig. 2.6C), but only for the libraries obtained on deck. This could be a result of slightly more efficient manual purification steps, including bead-drying and alcohol wash removal, compared to on-deck purification in which the broad hump could be indicative of leftover enzymes or protein. Figure 6F shows an electropherogram overlay of E. coli positive control (red) and on-platform (blue) libraries generated by the mechanical fragmentation assay, while Figure 7F shows an electropherogram overlay of E. coli positive control (red) and on-platform (blue) libraries generated by the enzymatic fragmentation assay, for visual comparison.

Libraries from Human DNA: Human DNA libraries obtained via our automated method were evaluated for their size distribution. Again, the fragment size distribution fell within the desired ~200-1000 bp region (Fig. 2.6A, 7A). Adapter-dimer can be seen at ~150 bp, but no significant primer-dimer is present. The average yield and % adapter-dimer for on-platform human libraries generated by the mechanical and enzymatic fragmentation assays were 677.20 $\pm$ 105.62 ng and 0.51 $\pm$ 0.0022 %, and 2204.91 $\pm$ 493.78 ng and 0.23 $\pm$ 0.00344 %, respectively. Similar to the E. coli libraries, the electropherograms of the human libraries produced by the enzymatic fragmentation assay show a trailing edge past 1000 bp, highlighting the presence of high molecular weight DNA fragments and possible gDNA contamination remaining from the fragmentation reaction (Fig. 2.7A). Slight high molecular weight DNA contamination is expected as we did not perform double-sided size selection for any of the human or E. coli libraries produced by the enzymatic fragmentation assay and therefore only removed small fragments including adapter and primer-dimers in both purification steps. Like the E. coli libraries produced by the mechanical fragmentation assay, the human libraries also show a small, broad hump past 1000 bp, while the positive controls do not. Therefore, we can say that this broad hump is not dependent on the input sample but perhaps a function of purification step precision and bead binding efficiency. Figure 6E shows an electropherogram overlay of human positive control (red) and on-platform (blue) libraries generated by the mechanical fragmentation assay, while Figure 7E shows an electropherogram overlay of human positive control (red) and on-platform (blue) libraries generated by the enzymatic fragmentation assay, for visual comparison. It is important to note that all final libraries produced via enzymatic fragmentation (E. coli and human) were diluted appropriately to be within the concentration range of the Bioanalyzer dsDNA High Sensitivity Assay (necessary due to 50 ng input and 8 PCR cycles) (S. Table 3B).

Figure 2. 6: Mechanical fragmentation final library Quality Control. (A) Electropherograms of on-deck (Samples 1-6) and positive control (Samples 7-8) human DNA libraries. The bottom right panel shows a DNA ladder for size comparison. (B) Gel image demonstrating fragment size distributions of the human DNA libraries shown in A. (C) Electropherograms of on-deck (Samples 1-6) and positive control (Sample 7) E. coli DNA libraries. The bottom right panel shows a DNA ladder for size comparison. (D) Gel image demonstrating fragment size distributions of the E. coli DNA libraries shown in C. (E) Human DNA positive control (red) and on-deck (blue) library overlay for yield and size distribution comparison. (F) E. coli DNA positive control (red) and on-deck (blue) library electropherogram overlay for yield and size distribution comparison. (G) Manual (red) and on-deck (blue) negative control electropherogram overlay.

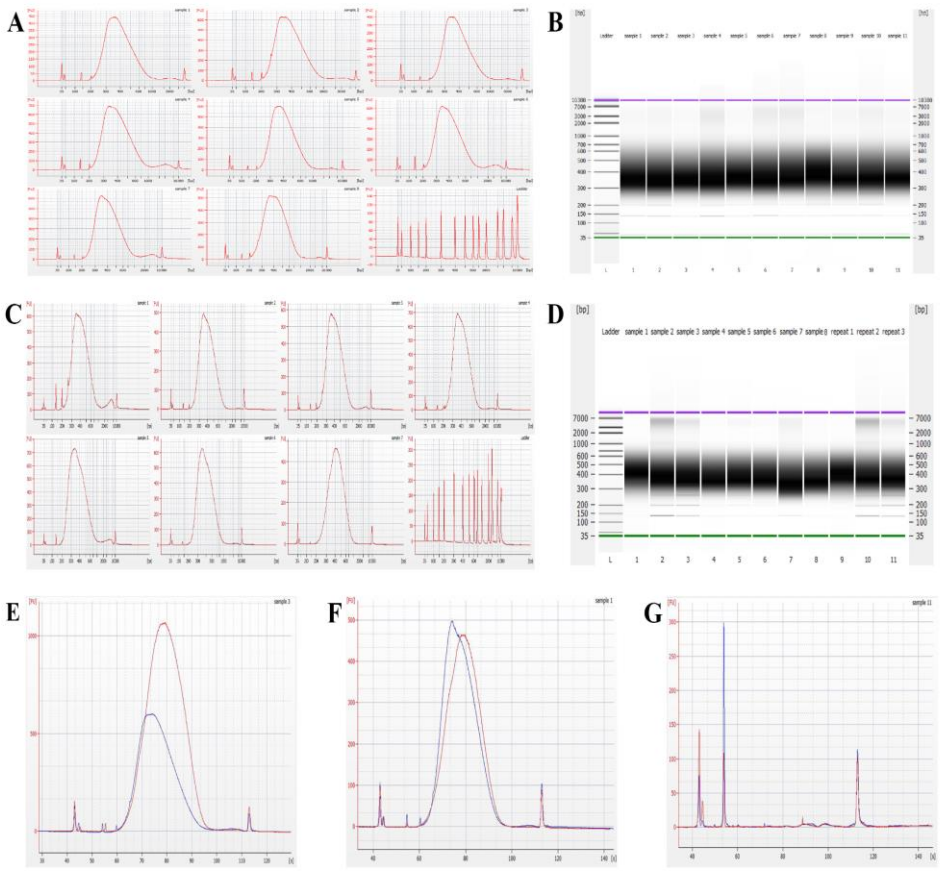
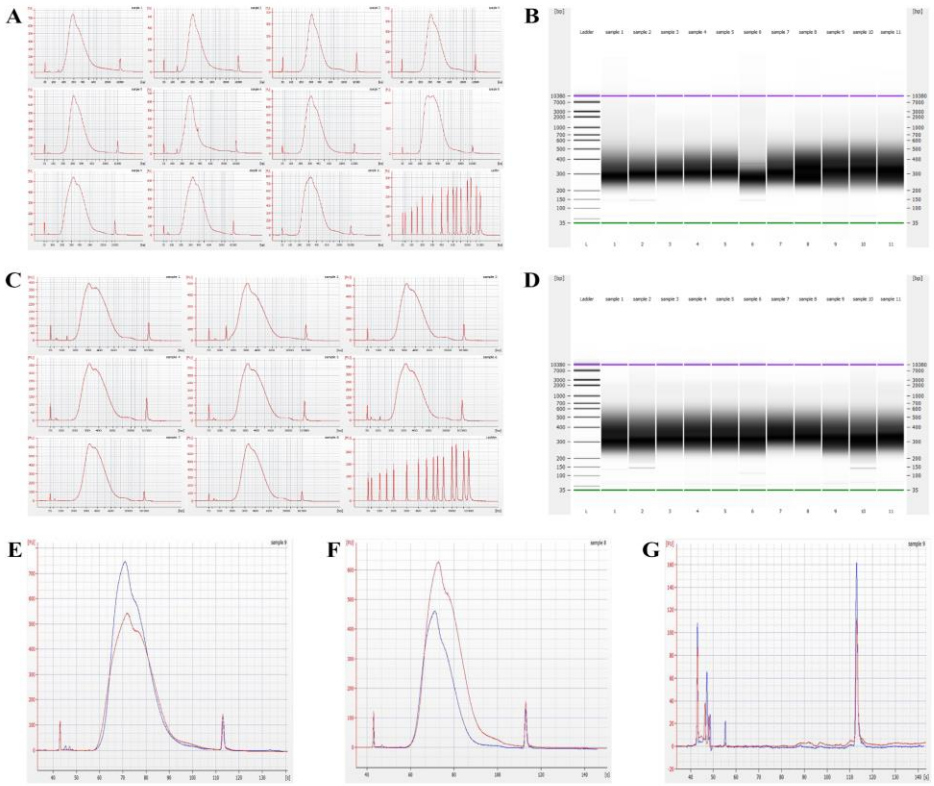


Figure 2. 7: Enzymatic fragmentation final library Quality Control. (A) Electropherograms of on-deck (Samples 1-8) and positive control (Samples 9-11) human DNA libraries. The bottom right panel shows a DNA ladder for size comparison. (B) Gel image demonstrating fragment size distributions of the human DNA libraries shown in A. (C) Electropherograms of on-deck (Samples 1-6) and positive control (Sample 7-8) E. coli DNA libraries. The bottom right panel shows a DNA ladder for size comparison. (D) Gel image demonstrating fragment size distributions of the E. coli DNA libraries shown in C. (E) Human DNA positive control (red) and on-deck (blue) library overlay for yield and size distribution comparison. (F) E. coli DNA positive control (red) and on deck (blue) library electropherogram overlay for yield and size distribution comparison. (G) Manual (red) and on-deck (blue) negative control electropherogram overlay.



Library Preparation Efficiency Analysis: We performed a library preparation efficiency analysis to further quantify and compare libraries produced via our automated method to those generated via manual preparation. All library preparation protocols mandate an input mass (in g) of DNA, m_{in}, for optimal efficiency and quality of sequence-able fragments. Hence, the number of input DNA fragments, N_{DNA}, are represented by,

$$N_{DNA} = \frac{m_{in} N_A}{650 \langle l_{bpi} \rangle}, \quad (4)$$

where N_A is the Avogadro's number, $\langle l_{bpi} \rangle$ is the average length of the input DNA, and 650 g/mol is the molar mass of a base pair. Proceeding stepwise throughout the library preparation assay, the efficiency of ligation η_{lig}must be taken into account before bead purification efficiency to obtain the number of adapter-ligated DNA fragments present post-ligation (N_{AL}).

$$N_{AL} = \eta_{lig} N_{DNA} \quad (5)$$

The efficiency of the first purification (wash 1) can be attributed to an assembly of bead binding efficiency (η_{bb}), bead retention efficiency (η_{br}) owing to bead loss throughout the assay, and bead elution efficiency (η_e): $\eta_{wash} = \eta_{bb}\eta_e\eta_{br}$. However, for the sake of simplicity, we will refer to the overall wash efficiency η_{wash}.

The number of post-PCR sequence-able DNA fragments (N_{PCR}) depend on the efficiency (η_{PCR}), of ($n = 8$) PCR cycles and (η_{wash}) , as described below.

$$N_{PCR} = 2^n \eta_{wash} \eta_{AL} \eta_{PCR} N_{AL} \quad (6)$$

The post-PCR purification step generates the same efficiency as the first purification step since both are the same protocol with different volumes. Therefore, the number of sequence-able DNA fragments in the final library (N_{out}) is estimated via

$$N_{out} = N_{PCR} \eta_{wash} (1 - F_{ads})(1 - F_{LT}), \quad (7)$$

where F_{ads} and F_{LT} represent the fraction of DNA lost to adsorption and during liquid transfers, respectively. The intent of our method is to keep these fractions close to zero. To better quantify and analyze library preparation efficiency, we convert the number of output DNA molecules to sequence-able DNA mass (m_{out}) in g, by accounting for the length and subsequent mass change of the DNA fragments post-adapter ligation as

$$m_{out} = 650 \frac{N_{out}}{N_A} \frac{\langle l_{bpi} \rangle}{\langle l_{bpi} \rangle + 2\langle l_A \rangle} \langle l_{bpi} \rangle \quad (8)$$

Here, $\langle l_A \rangle$ is the length of a single adapter in bp. Final library concentration in ng/μL (C_f) can be determined by dividing the output mass by the volume of the resuspension buffer.

$$C_f = \frac{m_{out}}{V_{RsB}} \quad (9)$$

Table 1 lists the calculated and constant values for each of the mentioned variables for both assays. Ligation and PCR efficiency values are founded on previously established data[62,63,64]. We assume that fragmentation and ERA reaction efficiencies are close to 100%, considering the presence of minuscule gDNA contamination in final libraries (Fig. 2.6).

Table 1 : Library Preparation Efficiency Analysis.

Variable Name	Definition	Mechanical Fragmentation	Enzymatic Fragmentation
m_{in}	Input mass (g)	20 ng	50 ng
$\langle l_{bpi} \rangle$	Length of input DNA fragment (bp)	260 bp	50 kbp → average 250 bp

			fragments
N_{DNA}	Number of input DNA molecules	7.19×10^{10}	1.87×10^{11}
η_{lig}	Ligation efficiency	60%	65%
N_{AL}	Number of adapter-ligated DNA molecules	4.31×10^{10}	1.21×10^{11}
η_{wash}	Total wash step efficiency	73%	77%
η_{PCR}	PCR efficiency	95%	95%
N_{PCR}	Number of adapter-ligated DNA molecules post-PCR	7.62×10^{12}	2.27×10^{13}
F_{ads}	Fraction of DNA lost to adsorption	0.05	0.05
F_{LT}	Fraction of DNA lost during liquid transfers	0.15	0.15
N_{out}	Number of DNA molecules in output (final library)	4.47×10^{12}	1.41×10^{13}
m_{out}	Theoretical total output DNA mass	809 ng	2360 ng
$\langle l_A \rangle$	Length of 1 adapter (bp)	70	75
V_{RsB}	Final resuspension buffer volume	33 µL	22 µL
C_f	Theoretical final library concentration	24.5 ng/µL	107 ng/µL

Our calculations demonstrate that the majority of the input DNA is lost during the ligation and purification steps. Adapter-ligation reactions are known to be sensitive and highly reliant on

the specific ligase enzyme and critical mixing of the ligase-adapter-DNA reaction[65]. Ligation efficiency can range anywhere from 3-100%, with most ligases falling in the 15-40% range[66]. Our automated adaptation of the ligation reaction exploits the fluidic properties associated with different viscosity liquids to create an optimized and custom mixing technique. Slow aspiration speeds allow for sufficient mixing without facilitating adapter-dimer formation, so ligation efficiency is presumed to be comparable on- and off-deck as the reaction conditions are optimized in both cases.

Purification step efficiency is a function of SPRI bead binding adsorption-desorption kinetics, a reaction entirely dependent on bead formulation, including surface-coating consistency and texture. SPRI beads are generally made of polystyrene covered in a layer of magnetite and positively charged carboxyl groups, allowing negatively charged DNA to bind to the beads[46]. If the bead's surface coating is uneven, DNA will not bind uniformly, and fragments will be lost during the purification steps, decreasing wash step efficiencies. Bead retention is a parameter unique to our automated method and can be regarded as negligible when performing library preparation manually. Though our approach produces high-quality DNA libraries, bead loss is more significant on-platform as a result of limited cannula tip motion in the y-direction and bead adherence to the cartridge tubing. This challenge is addressed via carefully crafted mixing techniques that take advantage of fluid flow and convective mixing in the vertical cannula tips to maximize DNA-bead adsorption and elution, thereby enhancing final library yield.

PCR is well-established in literature as highly efficient. We assumed PCR efficiency to be the same across both assays and positive controls. DNA adsorption to the cartridge tubing is another parameter unique to our automated platform, as it is negligible in the conventional manual

approach. Though surface passivation greatly reduces DNA adsorption rates, a small fraction (~5%) is lost when liquid is transferred inside the cartridge tubing.

Figure 2. 8: Positive control and on-deck final library concentrations in ng/µL for (A) 20 ng input human DNA libraries produced by the mechanical fragmentation assay, (B) 20 ng input E. coli DNA libraries produced by the mechanical fragmentation assay, (C) 50 ng input human DNA libraries produced by the enzymatic fragmentation assay, and (D) 50 ng input E. coli DNA libraries produced by the enzymatic fragmentation assay. Error bars represent Qubit standard error (12%), and dashed lines represent the average library concentration across samples and positive controls for each group.

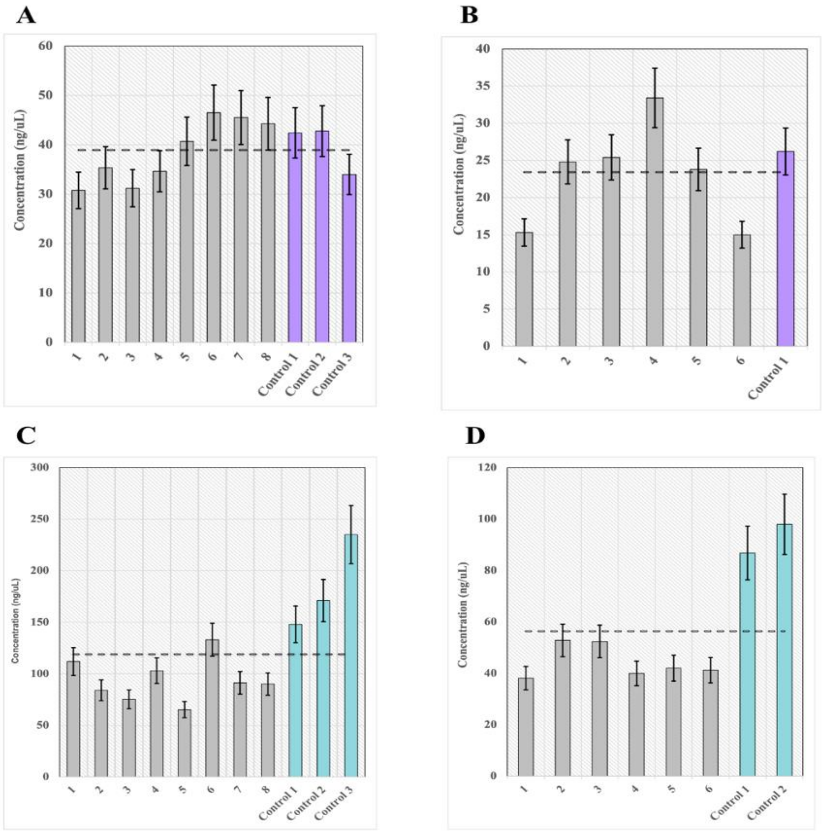


Figure 2.8 provides a visual summary of the final library concentrations for each of the four groups. On-platform libraries produced by the mechanical fragmentation assay averaged 38.62 ± 6.42 ng/µL and 22.95 ± 6.94 ng/µL for human and E. coli DNA, respectively, compared to 39.73 ± 4.97 ng/µL (human) and 26.2 ng/µL (E. coli) positive controls (Fig. 2.8 A-B). The theoretical final library concentration was calculated as ~ 24.5 ng/µL (Table 1). Positive control and on-platform final library concentrations for the human DNA libraries differed by ~1 ng/µL, while the E. coli libraries differed from positive controls by ~ 3 ng/µL, with slightly higher positive control concentrations in both cases, as expected. The average on-platform E. coli final library concentration fell within ~ 1.5 ng/µL of the calculated theoretical value, while the average on-platform human final library concentration was ~ 38% higher than calculated. It is important to note that the Qubit Flex Fluorometer has a known error of ± 12%[67] and is a potential causative agent of the difference between measured and calculated concentrations. On-platform libraries produced by the enzymatic fragmentation assay averaged 94.21 ± 21.48 ng/µL and 44.42 ± 6.47 ng/µL for human and E. coli DNA, respectively, compared to 184.66 ± 45.08 ng/µL (human) and 92.4 ± 7.92 ng/µL (E. coli) positive controls (Fig. 2.8 C-D). The theoretical final library concentration was calculated as ~ 107 ng/µL (Table 1). Positive control and on-platform final library concentrations for the human DNA libraries differed by ~90 ng/µL, while the E. coli libraries differed from positive controls by ~ 48 ng/µL, with more concentrated positive controls in both cases, as expected. The average on-platform E. coli final library concentration was ~12% lower than calculated, while the average on-platform human final library concentration was ~ 42% higher than calculated. Enzymatic fragmentation human and E. coli libraries had large standard deviations both on-platform and manually. The human DNA positive control standard deviation is significantly larger than the on-platform standard deviation (± 45.08 ng/µL vs. ± 21.48 ng/µL)

for the enzymatic fragmentation assay. The variability between the concentrations of libraries produced in the same run can be attributed to inconsistent adsorption throughout capillaries in the same cartridge (due to uneven surface morphology contributing to variable adsorption rates and ultimately uneven surface passivation), and variable liquid transfer efficiency between cannula tips (S. Fig. 1). The difference in final concentrations between on-platform and positive control libraries is due to the difference in purification step efficiencies on and off-deck (Table 1), as well as DNA adsorption, and liquid transfer efficiency compared to manual pipetting. Inconsistency in human vs. E. coli DNA final library concentrations produced via the same assay with the same amount of input nucleic acid material is accredited to differences in initial DNA dilutions, including pipetting and quantification discrepancies. Despite the inconsistencies described above, it is critical to realize that both on-platform and positive control libraries pass the criteria for sequencing success in terms of concentration, as the minimum sequence-able concentration is 5 ng/µL[68]. Overall, our library prep efficiency analysis provides insights into the mechanisms of sample loss within NGS library preparation assays compared with automated adaptations and pinpoints areas of future research, improvement, and optimization. Of course, our analysis is simply an estimate useful for contextualizing experimental data and calculated final library concentrations are expected to differ from experimentally obtained results simply as a consequence of error and approximation propagation.

XI. SEQUENCING OF HUMAN AND E. COLI DNA

Mechanical fragmentation: Libraries produced by the mechanical fragmentation assay were subject to WGS for both human and E. coli DNA. Mean Phred (Q) scores were >36 for all sequencing data generated by the mechanical fragmentation assay, with ~99% mean alignment

(sequence similarity to a reference genome) for the human DNA reads and ~97% mean alignment for the E. coli DNA reads (Table 2). It is important to note that Q=20 corresponds to 99% base call precision, while Q=30 corresponds to 99.9% base call precision[69]. Data quality was assessed as a measure of data yield and number of reads vs. DNA insert size (Fig. 2.9 A-B, D-E). Total sequencing data yield generated ~1,175.05 ± 0.25 megabases (Mb) for the human group, averaging ~103.73 ± 0.25 Mb per sample (Fig. 2.9A) and ~755 ± 0.25 Mb for the E. coli group, averaging ~103 ± 0.25 Mb per sample (Fig. 2.9D). Positive controls averaged ~115.08 ± 0.25 Mb for the human group and ~135 ± 0.25 Mb for the E. coli group. Though samples 8, Positive Control 1 in the human group, and 1,4 in the E. coli group were outliers, the average data yield was comparable to positive controls. Figure 9 B, E is a visual representation of the number of sequencing reads vs. DNA insert size in bp for both DNA input groups. Average read length was between 100-200 bp for both groups, with the highest number of reads generated from on-platform libraries (~11,000 reads for human and ~5,000 reads for E. coli), with human and E. coli positive controls falling between 3,000-5,000 and 3,000-4,500 reads, respectively (Fig. 2.9B, E). Sample-to-sample variation in the number of reads vs. insert size is likely due to library pooling and uneven cluster generation due to varying input library concentrations.

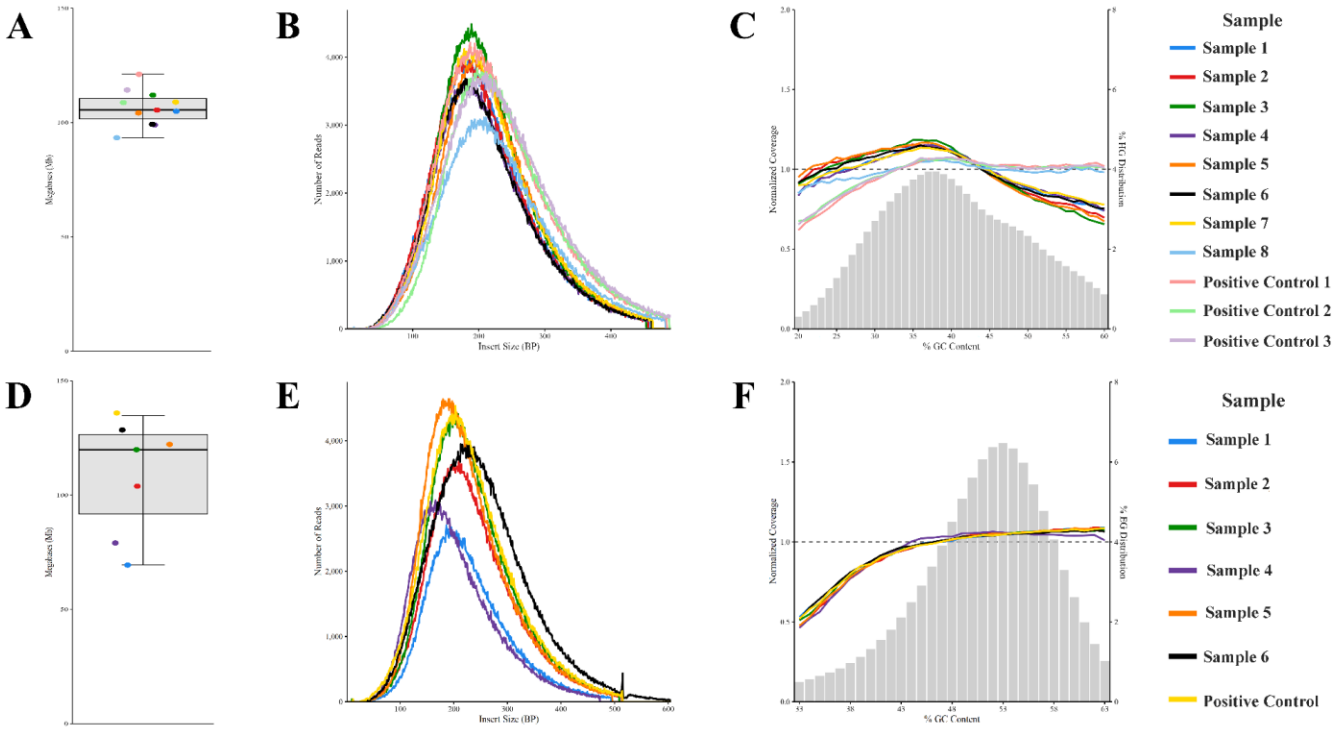
Genome coverage analysis was performed to generate normalized coverage vs. % GC content plots. The human and E. coli groups show normalized coverage close to 1, with slight underrepresentation of AT-rich regions and overrepresentation of GC-rich regions of the two genomes (Fig. 2.9C, F). However, positive controls have almost identical distributions, eliminating our platform as a perpetuator of sequencing bias. Mechanical fragmentation (sonication in this case), ERA, and ligation reactions are possible sources of bias present in both on-platform and positive control libraries. The grey bars represent the % GC content distributed across either the

human or E. coli genome. The bar chart is centered around ~38 % GC content for human (Fig. 2.9C) and ~ 53 % GC content for E. coli (Fig. 2.9F) DNA, representing the different amounts of GC in each genome (40.9% for human[70] and 50% for E. coli[71]). Plots were generated using Picard, and normalized coverage was calculated by way of the GC bin method in which the reference genome is categorized into bins corresponding to GC content between 20 and 60 %, then the number of reads in each GC bin is divided by the average number of reads for all GC bins. Percent GC is calculated based on a set read length of 100 bp known as 'window-size'.

Table 2: Mean Phred (Q) Scores and % Alignment.

Data Group	Average Q Score	Std. Dev.	Average % Alignment	Std. Dev.
Mechanical Fragmentation Human	~36.20	± 0.47	~99.00 %	± 0.14 %
Mechanical Fragmentation E. coli	~36.25	± 0.43	~97.00 %	± 0.06 %
Enzymatic Fragmentation Human	~36.45	± 0.47	~98.71 %	± 0.52 %
Enzymatic Fragmentation E. coli	~35.76	± 0.65	~96.57 %	± 0.10 %

Figure 2. 9: Sequencing data analysis for the mechanical fragmentation assay. Data yield per sample in Mb for (A) human DNA input and (D) E. coli DNA input. Number of reads vs. insert size for (B) human DNA input and (E) E. coli DNA input. Sequencing bias analysis in terms of normalized coverage vs. % GC content for (C) human DNA input and (F) E. coli DNA input.

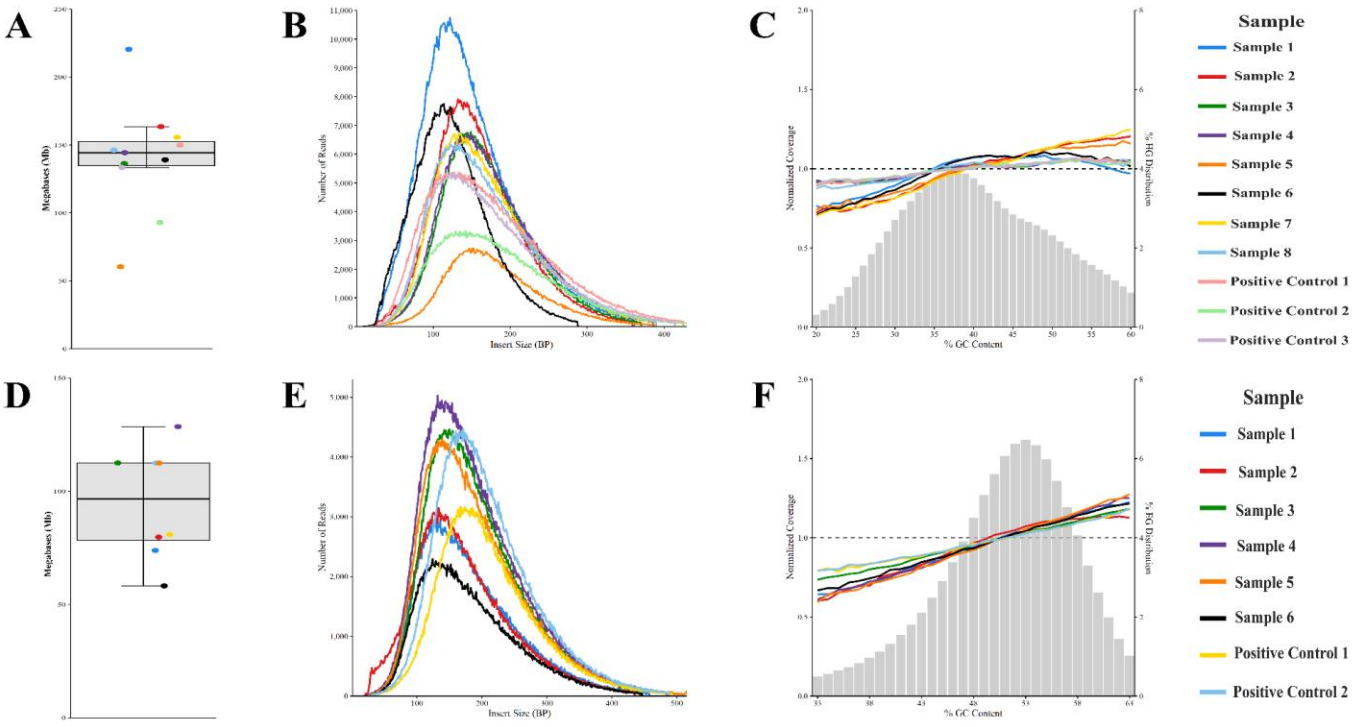


Enzymatic fragmentation: Libraries produced by the enzymatic fragmentation assay underwent WGS for both human and E. coli DNA. Mean Q scores were >35 for all sequencing data produced by the enzymatic fragmentation assay, with ~98.7% mean alignment (sequence similarity to a reference genome) for the human DNA reads and ~96.5% mean alignment for the E. coli DNA reads (Table 2). Sequencing generated ~ 1,449 ± 0.25 Mb of raw data for the human group, averaging ~147 ± 0.25 Mb per sample (Fig. 2.10A), and ~761 ± 0.25 Mb for the E. coli group, averaging ~95 ± 0.25 Mb per sample (Fig. 2.10D). Positive controls averaged ~ 92 ± 0.25 Mb per sample for the human group and ~97 ± 0.25 Mb per sample for the E. coli group. Though samples 4 and 6 (Fig. 2.10D) were outliers, data yield was comparable to positive controls. Average read length was between 100-200 bp for both groups (Fig. 2.10 B, E), with on-platform libraries at ~11,000 reads for human (Fig. 2.10B) and ~5,000 reads for E. coli (Fig. 2.10E), and

positive controls falling between 3,000-5,000 and 3,000-4,500 reads, respectively. Again, sample to sample variation is due to varying input library concentrations.

Both the human and E. coli groups demonstrate normalized coverage close to 1, with an underrepresentation of AT-rich regions and an overrepresentation of GC-rich regions, as expected (Fig. 2.10 C, F). Positive controls have comparable coverage distributions, indicating negligible sources of bias generated by our instrument. However, enzymatic fragmentation is an added overarching source of potential bias in this case, as fragmentation enzymes are notorious for inconsistent fragmentation among different % GC regions of the genome[72]. The grey bar chart is centered around ~38 % GC content for human (Fig. 2.10C) and ~ 53 % GC content for E. coli (Fig. 2.10F) DNA, representing the different amounts of GC in each genome, and aligns with known %GC content for the human and E. coli genomes.

Figure 2. 10: Sequencing data analysis for the enzymatic fragmentation assay. Data yield per sample in Mb for (A) human DNA input and (D) E. coli DNA input. Number of reads vs. insert size for (B) human DNA input and (E) E. coli DNA input. Sequencing bias analysis in terms of normalized coverage vs. % GC content for (C) human DNA input and (F) E. coli DNA input.



By simplifying various steps in the library preparation and lowering the input DNA requirements, our system enables surveillance and identification of potential disease outbreaks at early stages, isolating sources of a pathogen by molecular genotyping of bacterial isolates[73, 74].

2.4 CONCLUSIONS

As NGS library preparation remains a significant bottleneck to the entire genomic research and clinical community, it is essential that innovative engineering technologies automate the process and meet standards for accurate sequencing and diagnostic screening. Our design

innovation presents a platform that combines all elements of DNA NGS library preparation by employing simplified micro and macroscale geometries to efficiently perform purification, reaction heating and mixing, and fluidic transport. DNA libraries prepared using our platform met the same standard as libraries prepared manually in terms of coverage bias, yield, and fragment size. Our scientific approach uniquely integrates automation into applications in areas such as precision medicine, genomic surveillance, and antibiotic resistance tracking. Our method is ideal for smaller laboratories where high-throughput robotic liquid handlers are not a feasible option for automating library preparation. By adopting our method, the research community is expected to save time and resources as well as more effectively process clinical samples.

2.5 ACKNOWLEDGEMENTS

We would like to thank the PerkinElmer group (Dr. Adam Snider, Christopher Tannock, Shreyas Shah, Spencer Curbelo, and Arjun Bansal) for their helpful comments and scientific discussions. We would also like to thank Dr. William Law for sequencing all DNA libraries and analyzing the sequencing data.

The authors would like to thank PerkinElmer's grant to Brown University for carrying out this research and development of the platform.

XII. Competing Interests:

The authors declare no conflicts of interest.

AT is a paid scientific advisor/consultant for PerkinElmer.

XIII. Data Availability:

The datasets generated and analyzed during the current study are available in the NCBI Sequence Read Archive (SRA) repository (BioProject accession PRJNA935804), https://www.ncbi.nlm.nih.gov/bioproject/PRJNA935804.

2.6 EXPERIMENTAL PROCEDURES AND MATERIALS

XIV. Library Preparation Methods

We applied our method to two library preparation kits, one utilizing mechanical DNA fragmentation (sonication) and the other utilizing enzymatic fragmentation.

Mechanical fragmentation: We constructed positive controls and on-platform libraries using the NEXTFLEX Rapid DNA-Seq Kit 2.0 for Illumina Platforms (Perkin Elmer Inc., Waltham, MA) to assess mechanical fragmentation library preparation. The manufacturer-issued protocol was used to construct all mechanically fragmented positive controls. Fragmented DNA, water, ERA buffer, and ERA enzyme were combined and incubated for 30 minutes at 20℃ and 30 minutes at 65℃ using the MiniAmp™ Thermal Cycler (ThermoFisher Scientific, Waltham, MA). Adapters were diluted with a 1:40 dilution and then homogenized with the ERA mixture, ligase enzyme, and ligase buffer. The mixture was incubated for 15 minutes at 20℃, then purified with cleanup beads and 70% IPA per the kit instructions. DNA was eluted in resuspension buffer, and the supernatant was combined with primer and PCR master mix. 8 cycles of PCR were performed, with an initial denaturation of 30 seconds at 98℃, denaturation for 15 seconds at 98℃, annealing for 30 seconds at 65℃, and extension for 30 seconds at 72℃, with a final extension for 2 minutes

at 72℃. A second wash was performed with cleanup beads and 70% IPA per kit instructions to yield the final library of adapter-ligated DNA.

Automated adaptation: For the on-platform adaptation of the protocol, a pretreatment solution was added for surface passivation of the cartridge tubing, comprised of ERA buffer from the library preparation kit, 20 mg/mL BSA (New England Biolabs, Ipswich, MA), 10% Tween-20 (Bio-Rad Laboratories, Hercules, CA), 30 ng/μL oligo (Integrated DNA Technologies, Coralville, OH), and nuclease-free water (Integrated DNA Technologies, Coralville, OH). Adapters and primers were combined with 8 μL of water when placed in their respective wells to prevent evaporative loss and liquid handling errors. 28 μL of resuspension buffer was used for elution at the end of the post-ligation cleanup and 35 μL at the end of the post-PCR cleanup. All other volumes remained the same in the on-platform adaptation. All temperatures and incubation times were consistent with the manual protocol (S. Table 1).

Enzymatic fragmentation: Additional positive controls and on-platform libraries were constructed with the Library Preparation Enzymatic Fragmentation Kit 2.0 (Twist Bioscience, South San Francisco, CA) to assess enzymatic fragmentation library preparation. The manufacturer-issued protocol was used to construct all enzymatically fragmented positive controls. Genomic DNA, water, frag/AT buffer, and frag/AT enzymes were combined and incubated for 20 minutes at 37℃, then 30 minutes at 65℃ in a thermal cycler. Adapters and ligation master mix were added to the fragmentation mixture. The mixture was incubated for 15 minutes at 20℃, then purified with purification beads and 70% IPA per the kit instructions. DNA was eluted in nuclease-free water, and the supernatant was combined with primer and PCR master mix. 8 cycles of PCR were

performed, with an initial denaturation of 45 seconds at 98℃, denaturation for 15 seconds at 98℃, annealing for 30 seconds at 60℃, and extension for 30 seconds at 72℃, with a final extension for 1 minute at 72℃. A second wash was performed with purification beads and 70% IPA per kit instructions to yield the final library of adapter-ligated DNA.

Automated adaptation: For the on-platform adaptation of the protocol, we used the same blocking pretreatment as the mechanical fragmentation assay. Adapter was combined with 5 µL water when placed in the reagent plate. 23 µL of resuspension buffer was used for elution at the end of post-ligation cleanup and 27 µL at the end of post-PCR cleanup. All other volumes remained the same in the in-platform adaptation. All temperatures and incubation times were consistent with the manual protocol (S. Table 1).

XV. Nucleic Acid Starting Material

20 ng total (concentration 2 ng/µL) of fragmented Escherichia coli (E. coli) DNA (average 280 bp) and fragmented human DNA (Female, average 260 bp, Promega, Madison, WI) obtained from Bioo Scientific (Austin, TX) were input into the mechanical fragmentation assay using the NEXTFLEX Rapid DNA-Seq Kit 2.0 for Illumina Platforms.

E. coli gDNA (Thermo Fisher Scientific, Waltham, MA) and Human gDNA (Mixed, 50 kbp, Promega, Madison, WI) were input into the enzymatic fragmentation assay using the Library Preparation Enzymatic Fragmentation Kit 2.0. The Twist Library Preparation Enzymatic Fragmentation Kit 2.0 assay required 50 ng total of starting nucleic acid material per the recommendations in the protocol.

XVI. Library Quantification and Quality Control

All starting material and final libraries were quantified using an 8-channel Qubit Flex Fluorometer and the dsDNA High Sensitivity (HS) assay (Invitrogen, Waltham, MA). Input DNA and final libraries were assessed for quality control using the High Sensitivity DNA assay for dsDNA on an Agilent 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA). Bioanalyzer and Qubit input DNA library concentrations ranged from 5-50 ng/µL. Libraries were diluted appropriately to concentrations within the range of the Bioanalyzer High Sensitivity DNA assay prior to chip loading and the HS Qubit assay prior to quantification.

XVII. Mechanical Device Assembly

Our device utilizes air displacement to achieve accurate reagent aspiration and dispensation. Eight parallel lanes displace liquid via a unified motion of eight syringes connected to a horizontal linear rail. Airlines connect to a single-use custom cartridge comprised of flexible polymeric tubing and joined to 16 cannula tips that insert into the wells of a 384-well reagent plate in two parallel columns of 8. Cannula tips are secured to a vertical linear rail for movement along the z-axis. The reagent plate is inserted onto a horizontal aluminum frame and attached to a linear rail for movement along the x-axis. A bar magnet moves along the z-axis to press against the reagent plate during bead pelleting steps. All components move along their respective linear rails via step motor motion control.

The polymeric tubing of the cartridge inserts directly into a grooved Peltier heating element, held in place with a detachable insulated door. A thermocouple secured to the heating element via heat-

resistant epoxy allows for real-time logging of thermal data during thermal cycling. Valves are present on both sides of the heating element with the freedom to move along the y-axis, able to seal samples inside the polymeric tubing and prevent vapor escape.

XVIII. Device Consumables

The disposable cartridge is single-use and therefore replaced after each preparation of 8 libraries to avoid contamination and ensure accurate liquid handling. The cartridge consists of eight separate lanes of flexible polymeric tubing joined directly to the device airlines and, subsequently, the syringe pumps via tightly sealed O-rings. Each lane contains a y-junction to allow for fluid displacement in two separate segments of tubing and subsequent cannula tips situated in parallel columns of eight.

Every new library preparation (use of the device) requires a new, sterile Thermo Scientific™ Nunc™ DeepWell™ 384-Well Plate (Thermo Fisher Scientific, Waltham, MA). Reagents are loaded into specific columns of the plate by the user, so the platform can transfer them appropriately throughout each assay. Every other row of the plate corresponds to the preparation of a separate indexed library, which allows for successful multiplexing. Plates are centrifuged briefly prior to loading of 70% isopropyl alcohol used for the wash steps, then sealed with a 384-well pierceable plate seal (Azenta Life Sciences, Wotton, England).

XIX. Scripting Resources

Python was chosen as the main scripting language for assay development, software, and firmware on the platform. The majority of scripting functions are defined in a single master script that sets

device landmarks in terms of pump rotations and establishes a framework for mechanical feedback on-platform. The master script translates from pump rotations to linear movement along each axis, allowing for a more user-friendly assay development process. The script contains arrays of column positions for each of the two cannulas, as well as classes for speeds and various consumable reagent plates. Functions exist for motor homing, logging of procedure execution and regulation, setting the temperature of the heating element, generating thermal data for post-run analysis, and refined, customizable X-plate, z-axis, and pump movement.

Assay development occurs in supplementary scripts that import functions from the master script. Each liquid displacement function contains parameters for cannula being used, column number on the plate, speed of pump movement, and, in some cases, a quantity of liquid to be aspirated or dispensed. There are methods defined for mixing at various intensities, with a number of loop repetitions delineated locally in the supplementary scripts. For instance, to resuspend a bead pellet during a library wash step, the assay developer might first use the "move_all" function to transfer resuspension buffer to the column with the bead pellet, then use a de-peletting mixing technique with 40 repetitions to homogenize the SPRI beads and resuspension buffer. Each step of an assay protocol is achieved through these fundamental building blocks; subsystems are called on sequentially to manipulate reagents and heat boluses to desired temperatures. All supplementary scripts are stored in an assay development repository, hosted, and managed via Bitbucket (Atlassian, Sydney, Australia).

XX. Sequencing and Data Analysis

All samples were sequenced using MiniSeq and MiSeq Systems (Illumina, San Diego, CA). Data analysis was performed using Trimmomatic, Bowtie2, Samtools, Picard, and custom bash and R scripts. R functions used to generate figures were ggplot2, reshape2, tidyverse, scales, pals, RColorBrewer, and gridExtra.

2.7 SUPPLEMENTARY MATERIAL

S Table 1: Variables and values corresponding to Fig. 2.

S. Table 1-2.1

T_{pt}	75°C
t_{pt}	10 min
P_1	27.89 psi
V_1	50 µL
C_{pt}	[BSA] = 0.4 µg/µL; [Oligo] = 0.6 ng/µL; [Tween-20] = 1.1 ng/µL; [ERA Buffer]
C_w	[BSA] = 0.105 µg/µL; [Oligo] = 0.158 ng/µL; [Tween-20] = 4.316 ng/µL; [ERA Buffer] in waste well

S. Table 1-2.2i

C_{i*}	[sheared DNA] = 0.4 ng/µL [ERA buffer] [ERA enzyme]
T_{era*}	65°C
t_{2*}	35 min
P_1	27.89 psi

S. Table 1-2.2ii

C_i	[gDNA] = 1 ng/µL [Frag/AT buffer] [Frag/AT enzymes]
T_f	37°C
t_2	20 min
T_{era}	65°C

V_{2*}	50 µL		t_3	30 min
C_{a*}	[sheared DNA] = 0.4 ng/µL [adapter] = 0.012 µM [ligase enzyme] [ligase buffer] [ERA buffer] [ERA enzyme]		P_1	27.89 psi
T_{L*}	Room Temperature		V_2	50 µL
t_{3*}	20 min		C_a	[fragmented DNA] = 1 ng/µL [adapter] [ligation master mix] [Frag/AT buffer] [Frag/AT enzymes]
V_{3*}	107.5 µL		T_L	Room Temperature
C_{b*}	[adapter-ligated DNA] [non-ligated adapter] [ligase enzyme] [ligase buffer] [ERA buffer] [ERA enzyme]		t_4	20 min

V_{4*}	100 µL
C_{c*}	[Cleanup beads] [adapter-ligated DNA] [non-ligated adapter] [ligase enzyme] [ligase buffer] [ERA buffer] [ERA enzyme]
V_{5*}	~205 µL
V_{6*}	198 µL
V_{7*}	~196 µL
t_{4*}	~3 min
V_{8*}	28 µL
C_{d*}	[Cleanup beads] [adapter-ligated DNA]

V_3	80 µL
C_b	[adapter-ligated DNA] [non-ligated adapter] [ligation master mix] [Frag/AT buffer] [Frag/AT enzymes]
V_4	60 µL
C_c	[Purification beads] [adapter-ligated DNA] [non-ligated adapter] [ligation master mix] [Frag/AT buffer] [Frag/AT enzymes]
V_5	~138 µL
V_6	200 µL
V_7	~198 µL
t_5	~3 min

C_{e*}	[adapter-ligated DNA]
V_{9*}	27 µL

V_8	23 µL
C_d	[Purification beads] [adapter-ligated DNA]
C_e	[adapter-ligated DNA]
V_9	22 µL

S. Table 1-2.3

	Mechanical Fragmentation	Enzymatic Fragmentation
C_{f*}	[adapter-ligated DNA]	[adapter-ligated DNA]
V_{9*}	27 µL	22 µL
C_{g*}	[adapter-ligated DNA] [PCR primers] [PCR master mix]	[adapter-ligated DNA] [PCR primers] [PCR master mix]
T_{D*}	98°C	98°C
T_{A*}	65°C	60°C
T_{E*}	72°C	72°C

t_{cycle*}	Initial Denaturation: 30 sec Denaturation: 15 sec Annealing: 30 sec Extension: 30 sec Final Extension: 2 min	Initial Denaturation: 45 sec Denaturation: 15 sec Annealing: 30 sec Extension: 30 sec Final Extension: 1 min
P_1	27.89 psi	27.89 psi
V_{11*}	62 µL	57 µL
C_{h*}	[amplified DNA] [PCR primers] [PCR master mix]	[amplified DNA] [PCR primers] [PCR master mix]
V_{12*}	45 µL	50 µL
C_{j*}	[Cleanup beads] [amplified DNA] [PCR primers] [PCR master mix]	[Purification beads] [amplified DNA] [PCR primers] [PCR master mix]
V_{13*}	~105 µL	~105 µL
V_{14*}	198 µL	200 µL
V_{15*}	~196 µL	~198 µL
t_{5*}	~3 min	~3 min

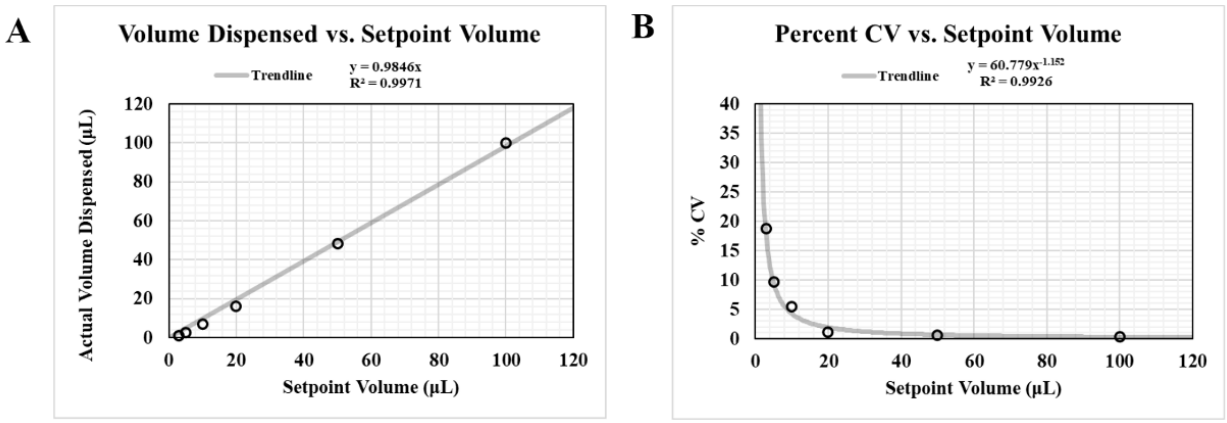
V_{16*}	35 µL	27 µL
C_{k*}	[Cleanup beads] [amplified DNA library]	[Purification beads] [amplified DNA library]
C_{l*}	[amplified DNA library]	[amplified DNA library]
V_{17*}	33 µL	25 µL

XXI. Liquid handling accuracy:

The platform was evaluated for volumetric accuracy during liquid handling. Liquid transfer tests were conducted, with three trials performed for transfers of 3 µL, 5 µL, 10 µL, 20 µL, 50 µL, and 100 µL of water. Final dispensed volume was determined using the density of water and an AL104 Analytical Balance (Mettler Toledo, Columbus, OH).

Volume accuracy of liquid transfers improves with increasing volume, while % CV decreases (Fig 5). The platform has an approximate working range of 10 to 200 µL, which allows it to complete liquid transfers and perform proper mixing of plated reagents. For this volumetric range, % CV remains low, with high accuracy.

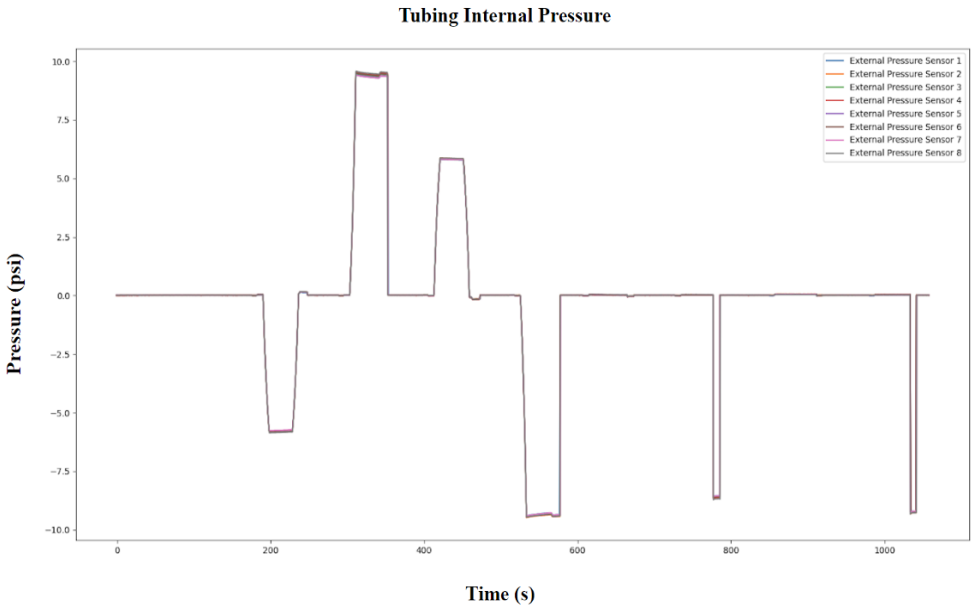
S Figure 1: Volumetric accuracy of the pump assembly. (A) Actual volume dispensed vs. scripted setpoint volumes, and (B) Percent CV vs. scripted setpoint volumes.



XXII. Evaluation of Pump Pressure:

The pump's ability to hold pressure is evaluated using a rig with embedded pressure sensors. The rig is placed on the platform with the syringe plungers at position 0, and the second cannula selected by closing the two valves on the other cannula. The syringe assembly moves to position 30 (maximum extension of the plungers) to apply a vacuum for a set period of time. The rig is removed to relieve the vacuum, then replaced. The syringe plungers move to position 0 to apply pressure and are held for a set period of time to monitor leak rates. This process is repeated for the first cannula. The first cannula pressures are negative compared to actual pressures due to the directionality of pressure connections within the set of sensors. The last portion of the test assesses leak rates of individual valves, where the two spikes represent a rapid release of pressure between valve tests (S. Fig. 2).

S Figure 2: Plot of pressure readings over time during pump pressure evaluation.



XXIII. Calculation of Pressure in Fluid Slug During Heating:

To quantify the pressure a fluid slug is placed under during heating, the initial and final volume of air between the fluid slug and the syringe plunger were calculated. Boyle's Law states that pressure and volume are inversely proportional, thus, with an initial volume of 1614.3626 mm^3, a post-pressurization volume of 850.7514 mm^3 (S. Table 2), and an initial pressure of 1 atm, the fluid slug experiences a pressure of approximately 27.89 psi.

S Table 2: Dead volumes of pump-airline assembly components.

Component	Volume (mm^3)
Tubing from fluid slug to y-junction	132.6148
Tubing from y-junction to cartridge frame	138.5528

Interfaces between cartridge and airline	40.4268
Airline	296.8989
Interfaces between airline and syringes	159.2569
Syringe barrel (post-pressurization in parentheses)	846.6124 (83.0012)
Total (post-pressurization in parentheses)	1614.3626 (850.3626)

XXIV. Determination of Library Yield and Percent Adapter-Dimer:

Raw data obtained from the Agilent 2100 Bioanalyzer instrument was exported as a .csv file and imported into Microsoft Office Excel. Mass libraries (y in ng) were determined by taking the sum of all concentrations corresponding to a fragment size range from 200 to 1000 bp, multiplying by the final resuspension buffer volume (V_{RsB}), and dividing by the upper limit fragment size (l_{max} = 1000 bp). For enzymatic fragmentation the V_{RsB} value is 27 µl for on-deck and 22 µl for the positive manual control process. On the other hand, for mechanical fragmentation, the V_{RsB} value is 35 µl for on-deck and 33µl for the positive manual control process.

$$y = \frac{V_{RsB} \sum_{i=1}^{n} C}{l_{max}} \quad \text{(S.1)}$$

Mass of the adapter-dimer (d) was calculated using the same equation but for a range between 120 and 150 bp for enzymatic fragmentation and 120-144 bp for mechanical fragmentation (owing to different adapter lengths for each assay). Percent dimer was determined by simply dividing the mass dimer by the sum of mass library and mass dimer.

$$\% d = \frac{d}{y+d} \quad \text{(S.2)}$$

Mass library, mass dimer, and % dimer were calculated for all libraries, on-deck, and positive controls across both assays for Human and E. coli DNA. Samples were considered "passing" (i.e., sequence-able) if they exhibited a mass library ≥ 250 ng and % dimer < 5%. Our method produced libraries with comparable yield and % dimer to positive controls across the board (S. Table 3A, B).

S Table 3: Bioanalyzer Data and Calculations.

A: Mechanical Fragmentation

Mechanical Fragmentation			
Human sheared DNA	Mass Library/Yield (ng)	Mass Dimer (ng)	% Dimer
1	646.79	3.66	0.56%
2	625.12	3.54	0.56%
3	587.02	1.75	0.30%
4	890.93	7.01	0.78%
5	718.22	2.26	0.31%
6	755.29	6.68	0.88%
7	600.78	1.90	0.32%
8	593.43	2.25	0.38%

Positive Control 1	1047.45	1.82	0.17%
Positive Control 2	965.25	0.92	0.10%
Positive Control 3	723.97	1.84	0.17%
Average	741.30	3.06	0.41%
Std.	159.73	2.03	0.25%
E. coli sheared DNA	Mass Library/Yield (ng)	Mass Dimer (ng)	% Dimer
1	559.26	7.14	1.26%
2	480.08	1.40	0.29%
3	500.04	1.57	0.31%
4	624.23	1.36	0.22%
5	828.81	3.51	0.42%
6	669.05	0.89	0.13%
Positive Control	445.69	0.43	0.10%

Average	586.74	2.33	0.39%
Std.	133.12	2.33	0.40%

B: Enzymatic Fragmentation

Enzymatic Fragmentation				
Human gDNA	Mass Library/Yield (ng)	Mass Dimer (ng)	% Dimer	Dilution
1	296.15	0.88	0.30%	1:10
2	228.43	1.89	0.82%	1:10
3	188.23	0.00	0.00%	1:10
4	245.92	0.00	0.00%	1:10
5	326.59	0.00	0.00%	1:4
6	250.85	1.75	0.69%	1:10
7	492.24	0.00	0.00%	1:4
8	567.04	0.00	0.00%	1:4
Positive Control 1	268.99	0.00	0.00%	1:10

Positive Control 2	274.63	0.00	0.00%	1:10
Positive Control 3	396.10	0.00	0.00%	1:10
Undiluted Average	2457.86	4.11	0.16%	-
Undiluted Std.	679.34	7.45	0.0031%	-
E. coli gDNA	Mass Library/Yield (ng)	Mass Dimer (ng)	% Dimer	Dilution
1	260.33	0.72	0.27%	1:4
2	304.57	4.32	1.40%	1:4
3	263.05	0.00	0.00%	1:4
4	188.27	0.00	0.00%	1:4
5	227.50	0.00	0.00%	1:4
6	197.74	0.00	0.00%	1:4
Positive Control 1	364.20	0.00	0.00%	1:4

Positive Control 2	365.45	0.00	0.00%	1:4
Undiluted Average	1085.56	2.52	0.21%	-
Undiluted Std.	274.70	6.05	0.0049%	-

XXV. Final Library Quantification and Percent Error:

To assess experimental concentration in terms of theoretical concentration, calculate percent error, and quantify libraries for sequencing, concentration values were obtained for each library under the four conditions (Human vs. E. coli DNA, mechanical fragmentation vs. enzymatic fragmentation) using a Qubit Flex Fluorometer. There is some variability between the concentrations of libraries produced in the same run, which can be partially attributed to Qubit's 12% error rate, as well as inconsistent adsorption throughout tubes in the same cartridge (due to uneven surface passivation), and variable liquid transfer efficiency between cannula tips. Calculated standard deviations quantify this variance between (S. Table 4).

S Table 4: Final library concentrations.

A: Mechanical Fragmentation

Human sheared DNA		E. coli sheared DNA	
Sample	Concentration (ng/µL)	Sample	Concentration (ng/µL)

1	30.7933	1	15.3
2	35.3402	2	24.8
3	31.2222	3	25.4
4	34.6391	4	33.4
5	40.6902	5	23.8
6	46.5167	6	15.0
7	45.5093	Positive Control	26.2
8	44.2569	Average	23.4
Positive Control 1	42.4	Std.	6.46
Positive Control 2	42.8		
Positive Control 3	34		
Average	38.92		
Std.	5.84		

B: Enzymatic Fragmentation

Human gDNA	E. coli gDNA

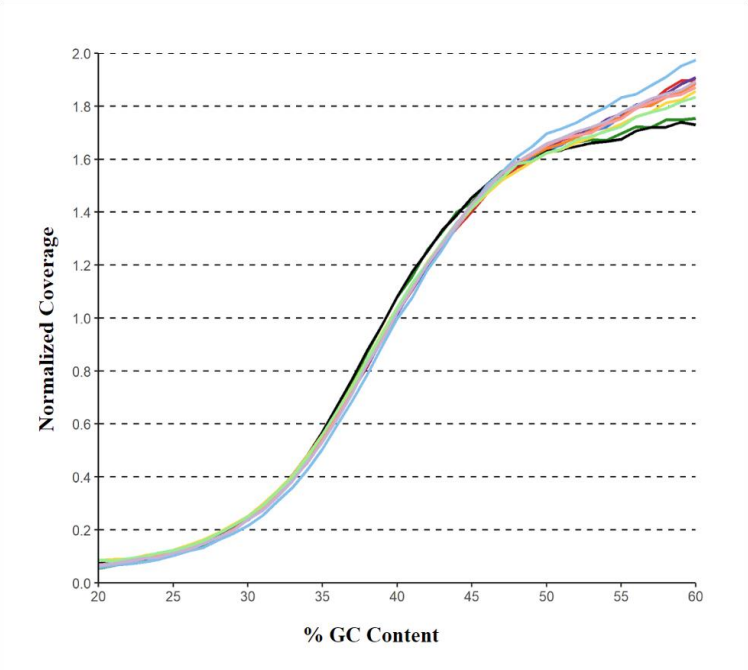
Sample	Dilution	Concentration (ng/µL)	Sample	Dilution	Concentration (ng/µL)
1	1:10	11.2	1	1:4	9.53
2	1:10	8.4	2	1:4	13.2
3	1:10	7.53	3	1:4	13.1
4	1:10	10.3	4	1:4	10.0
5	1:4	16.3	5	1:4	10.5
6	1:10	13.3	6	1:4	10.3
7	1:4	22.8	Positive Control 1	1:4	21.7
8	1:4	22.5	Positive Control 2	1:4	24.5
Positive Control 1	1:10	14.8	Average Undiluted Library	-	56.42
Positive Control 2	1:10	17.1	Undiluted Std.	-	23.07
Positive Control 3	1:10	23.5			

Average Undiluted Library	-	111.61
Undiluted Std.	-	50.14

XXVI. Sequencing bias:

Significant GC bias resulted from the sonication parameters listed in S. Table 5. These sonication parameters over-fragmented the adenine-thymine (AT)-rich portions of DNA. This occurred because of the weaker A-T bond (two hydrogen bonds) compared to the much stronger G-C bond (three hydrogen bonds). Size selection occurs throughout both library preparation assays to remove small and inutile DNA fragments, in this case disproportionately removing AT-rich portions of the sample genome. At the time of sample sequencing, we observed drastic pro-GC bias as seen in S. Fig. 3. The normalized coverage should ideally be a flat line at 1. We ultimately used different input DNA with more optimized sonication parameters, however these results represent a significant problem with DNA fragmentation methods and NGS library preparation in general. This problem may be addressed by lowering the sonication peak incident power and reducing the number of cycles per burst or duration of the cycle to fragment the AT-rich regions to a lesser extent.

S Figure 3: A compilation of positive control and on-deck mechanically fragmented human DNA libraries exhibiting significant pro-GC bias. Ideal normalized coverage is a flat line at 1, with variation between 1.4 and 0.4.



S Table 5: DNA sonication parameters that produced significant pro-GC bias. DNA was sonicated using a Covaris S220 Focused-ultrasonicator (Covaris, Woburn, MA).

Parameter	Value
Peak Incident Power	210
Duty Factor	10%
Cycles per burst	200 count
Duration	150 sec
Water level	12
Temperature	7°C

Chapter 3

Conclusions and Future Directions

This work has focused on the creation of an automation platform and the engineering, verification, optimization, and validation of DNA sequencing sample preparation assays, using a wholistic bioengineering approach. This method is intended for practical applications utilizing whole-genome sequencing in molecular diagnostics, genomics-based screening strategies, genomic surveillance, non-invasive prenatal testing, and personalized medicine. There is a clear need, outlined and discussed in Chapter 1, for bioengineering solutions to the clinical bottleneck of NGS library preparation. The introduction of an accessible method, founded on the principles of cost effectiveness, ease of use, and optimization of reagents and materials, to the market is crucial to democratize laboratory automation and DNA sequencing technologies. To date, several key players monopolize the NGS market, including companies like Illumina, contributing to the inaccessibility and higher price points of NGS technologies. Competition in the NGS market is essential for accessibility, lowering requisite cost of sample-to sequencing workflows for smaller laboratories, and diversity in product design and applications.

3.1 FUTURE APPLICATIONS

XXVII. RNA

Future applications of this method include RNA library preparation, in conjunction with rRNA depletion and PolyA enrichment protocols. RNA library preparation procedures are lengthy, difficult, and susceptible to degradation and contamination. However, clinical and research applications of RNA sequencing are rapidly expanding as a supplemental approach to DNA

sequencing diagnosis of genetic disorders and can identify aberrant RNA phenotypes like mono-allelic gene expression, aberrant gene splicing, and aberrant gene expression[75]. In fact, it has been reported that RNA-seq in conjunction with genomic data can lead to a 15% uplift in patient molecular diagnosis, translating to a 15% improvement in diagnostic accuracy[75]. Not only is whole transcriptome sequencing from input total RNA gaining popularity, but single cell RNA-seq is rapidly gaining foothold in the oncology and personalized medicine spheres. Though the library preparation process for single cell sequencing includes additional steps such as tissue sample processing, cell lysis, and cell separation, the basic library preparation steps post-sample processing remain the same[76]. This method contains the essential functions (liquid handling, magnetic bead pelleting, and heating/cooling) to completely automate single cell and RNA-seq sample preparation processes from reverse transcription through amplification. Furthermore, this approach has the potential to incorporate mechanical or enzymatic cell lysis procedures, bead-based target capture and/or depletion, and bead-based nucleic acid extraction protocols from whole blood or other sample matrices.

XXVIII. TARGET ENRICHMENT

Target enrichment, including probe hybridization, is another application with widespread use in cancer diagnostics and mutation analysis of congenital disease[77]. Probe hybridization procedures in conjunction with indexed library preparation protocols are especially important for analysis of regions with low copy numbers and targeted sequencing methods/data analysis and have clinical significance for the characterization of infectious disease[77,78]. Probe hybridization/hybrid capture protocols typically consist of the NGS library preparation core steps (fragmentation, adapter ligation, purification) but include biotin-labeled probes that hybridize to

the DNA fragments and are subsequently captured and enriched using streptavidin-magnetic beads[78]. This method provides a framework for these reactions that require liquid handling, heating and cooling steps, and magnetic bead separation.

XXIX. PROTEINS

As the diagnostics field moves from DNA diagnostics to RNA diagnostics, and finally to protein diagnostics, it is important to consider the flexibility of diagnostic platforms and methods to the quickly evolving requirements of the field. Though personalized, precision medicine is currently limited to genetic analysis, there is a wealth of protein analyses and biomarkers that must be exploited. Proteins and small molecule arrays allow for kinome and interactome profiling from a diverse collection of sample matrices, and while mass spectrometry dominates the proteomics field[79], there is a push to develop more accessible methods of protein and antibody capture such as bead based ELISAs. Bead-based immunoassays integrated with microfluidics approaches have the potential to impact point of care diagnosis in terms of sample multiplexing and sample processing time[80]. The bioengineering approach outlined in this work provides a framework for reaction engineering and assay development that can be adapted to protein applications and future innovations in the field.

3.2 IMPLICATIONS FOR CLINICAL RESEARCHERS

As NGS technologies continue to develop, it is important to employ a comprehensive bioengineering approach to innovation. Historically, advances in NGS library preparation have been component-specific, focusing on the optimization of enzymes and specific reactions. While this has resulted in improvements such as reduced coverage bias, lower starting material

requirements, and reduced error rates, it has not resolved accessibility issues surrounding NGS reagents and labor expenses. The bioengineering approach outlined in this work demonstrates a comprehensive analysis of each component of the library preparation workflow, focusing on simultaneous optimizations of each step culminating in a cheaper and more efficient overall workflow. Furthermore, the mathematical approach to pinpoint areas of sample loss within the DNA preparation process outlined in this work provides future researchers with a template to identify reactions and processes that may require the most optimization and improvement, while illustrating a molecular understanding of the assay-instrument interface. These innovations have significant implications for clinical researchers, as clinical labs are historically underfunded and undervalued as sources of innovation and revenue and are expected to aid in the proliferation of NGS-based screening strategies and personalized medicine tests.

REFERENCES

1. Population Clock. Preprint at https://www.census.gov/popclock/.
2. Population Clock: World. Preprint at https://www.census.gov/popclock/world.
3. Who Coronavirus (COVID-19) dashboard. Preprint at https://covid19.who.int.
4. Tai, D. B. G., Shah, A., Doubeni, C. A., Sia, I. G. & Wieland, M. L. The Disproportionate Impact of Covid-19 on Racial and Ethnic Minorities in the United States. *Clin Infect Dis* **72**, 703–706 (2021).
5. Beckett, A. H., Cook, K. F. & Robson, S. C. A Pandemic in the Age of Next-Generation Sequencing. *Biochem (Lond)* **43**, 10–15 (2021).
6. John, G. *et al.* Next-Generation Sequencing (NGS) in Covid-19: A Tool for Sars-Cov-2 Diagnosis, Monitoring New Strains and Phylodynamic Modeling in Molecular Epidemiology. *Curr Issues Mol Biol* **43**, 845–867 (2021).
7. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA Cancer J Clin* **72**, 7–33 (2022).
8. Goetz, L. H. & Schork, N. J. Personalized Medicine: Motivation, Challenges and Progress. *Fertil Steril* **109**, 952–963 (2018).
9. Created with Biorender.com.
10. Guan, Y.-F. *et al.* Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. *Chin J Cancer* **31**, 463–470 (2012).
11. Wong, C. H., Siah, K. W. & Lo, A. W. Estimation of clinical trial success rates and related parameters. *Biostatistics* **20**, 273–286 (2019).
12. Prasad, V. & Mailankody, S. Research and Development Spending to Bring a Single Cancer Drug to Market and Revenues After Approval. *JAMA Intern Med* **177**, 1569–1575 (2017).
13. DNA Sequencing Costs: Data. *Genome.gov* Preprint at https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.
14. Institute for Genomic Medicine - UC San Diego Health Sciences. Preprint at http://igm.ucsd.edu/genomics/rates.shtml.
15. Pricing. Preprint at https://www.cgm.northwestern.edu/cores/nuseq/pricing.html.
16. Bo, R. CGI Service Rates Center for Genome Innovation. Preprint at https://cgi.uconn.edu/cgi_service_rates/ (2015).

17. Pricing Biotech. Preprint at https://biotech.illinois.edu/htdna/pricing.
18. Gladman, N., Goodwin, S., Chougule, K., Richard McCombie, W. & Ware, D. Era of gapless plant genomes: innovations in sequencing and mapping technologies revolutionize genomics and breeding. *Curr Opin Biotechnol* **79**, 102886 (2023).
19. Muzzey, D., Evans, E. A. & Lieber, C. Understanding the Basics of NGS: From Mechanism to Variant Calling. *Curr Genet Med Rep* **3**, 158–165 (2015).
20. Head, S. R. *et al.* Library construction for next-generation sequencing: Overviews and challenges. *Biotechniques* **56**, 61–77 (2014).
21. Ribarska, T., Bjørnstad, P. M., Sundaram, A. Y. M. & Gilfillan, G. D. Optimization of enzymatic fragmentation is crucial to maximize genome coverage: a comparison of library preparation methods for Illumina sequencing. *BMC Genomics* **23**, (2022).
22. Quail, M. A., Swerdlow, H. & Turner, D. J. Improved protocols for the illumina genome analyzer sequencing system. *Curr Protoc Hum Genet* (2009) doi:10.1002/0471142905.hg1802s62.
23. Schneider, L., Fraser, M. & Tripathi, A. Integrated magneto–electrophoresis microfluidic chip purification on library preparation device for preimplantation genetic testing for aneuploidy detection. *RSC Adv* **11**, 14459–14474 (2021).
24. Saraiya, A. *Making NGS library prep practical for low-throughput labs*. https://www.tecan.com/blog/making-ngs-library-prep-practical-for-low-throughput-labs.
25. Benjamini, Y. & Speed, T. P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res* **40**, e72–e72 (2012).
26. Morel, C. *et al. Overview of the diagnostics market*. (European Observatory on Health Systems and Policies, 2016).
27. Growth Plus Reports, P. N. Molecular Diagnostics Market Size Worth USD 31.28 billion by 2030| Global Molecular Diagnostics Industry Updates, Leading. Preprint at (2022).
28. Genomics Global Market to Reach $94.9 Billion by 2030 at a CAGR of 16.4%. Preprint at https://www.yahoo.com/now/genomics-global-market-reach-94-191500229.html (2022).
29. Lab Automation Market to Surpass US$ 7,194.1 Million by 2030, Says Coherent Market Insights (CMI). Preprint at https://nz.news.yahoo.com/lab-automation-market-surpass-us-140000667.html (2023).

30. Christofyllakis, K. *et al.* Cost-effectiveness of precision cancer medicine-current challenges in the use of next generation sequencing for comprehensive tumour genomic profiling and the role of clinical utility frameworks (Review). *Mol Clin Oncol* **16**, 21 (2021).
31. Onoyama, T., Ishikawa, S. & Isomoto, H. Gastric cancer and genomics: review of literature. *J Gastroenterol* **57**, 505–516 (2022).
32. Schweiger, M. R., Kerick, M., Timmermann, B. & Isau, M. The power of NGS technologies to delineate the genome organization in cancer: from mutations to structural variations and epigenetic alterations. *Cancer Metastasis Rev* **30**, (2011).
33. Servetto, A. *et al.* A review of the use of next generation sequencing methodologies to identify biomarkers of resistance to CDK4/6 inhibitors in ER+/HER2- breast cancer. *Crit Rev Oncol Hematol* **157**, 103191 (2021).
34. Cinque, L. *et al.* Novel pathogenic variants of the AIRE gene in two autoimmune polyendocrine syndrome type i cases with atypical presentation: Role of the NGS in diagnostic pathway and review of the literature. *Biomedicines* **8**, 1–12 (2020).
35. Schnitzler, L. J. *et al.* Sporadic late-onset nemaline myopathy: clinico-pathological characteristics and review of 76 cases. *Orphanet J Rare Dis* **12**, 86 (2017).
36. Suhaimi, S., Zulkipli, I., Ghani, H. & Abdul-Hamid, M. Applications of next generation sequencing in the screening and diagnosis of thalassemia: A mini-review. *Front Pediatr* (2022).
37. Alberry, M. S., Aziz, E., Ahmed, S. R. & Abdel-Fattah, S. Non invasive prenatal testing (NIPT) for common aneuploidies and beyond. *Eur J Obstet Gynecol Reprod Biol* **258**, 424–429 (2021).
38. Craig, D. J. *et al.* Technical advance in targeted NGS analysis enables identification of lung cancer risk-associated low frequency TP53, PIK3CA, and BRAF mutations in airway epithelial cells. *BMC Cancer* **19**, 1081 (2019).
39. Ebner, S. *et al. Sequencing of BRCA1/2-alterations using NGS-based technology: annotation as a challenge*. *Oncotarget* vol. 13 www.oncotarget.com (2022).
40. Endris, V. *et al.* NGS-based BRCA1/2 mutation testing of high-grade serous ovarian cancer tissue: results and conclusions of the first international round robin trial. *Virchows Archiv* **468**, 697–705 (2016).

41. Petrackova, A. *et al.* Standardization of Sequencing Coverage Depth in NGS: Recommendation for Detection of Clonal and Subclonal Mutations in Cancer Diagnostics. *Front Oncol* **9**, (2019).
42. Schneider, L., Cui, F. & Tripathi, A. Isolation of target DNA using synergistic magnetic bead transport and electrokinetic flow. *Biomicrofluidics* **15**, 24104 (2021).
43. Schneider, L., Usherwood, T. & Tripathi, A. A microfluidic platform for high-purity cell free DNA extraction from plasma for non-invasive prenatal testing. *Prenat Diagn* **42**, 240–253 (2022).
44. Snider, A., Nilsson, M., Dupal, M., Toloue, M. & Tripathi, A. A Microfluidics Workflow for Sample Preparation for Next-Generation DNA Sequencing. *SLAS Technol* **24**, 196–208 (2019).
45. Gordon, L. G. *et al.* Estimating the costs of genomic sequencing in cancer control. *BMC Health Serv Res* **20**, 492 (2020).
46. Lee, K. & Tripathi, A. Parallel DNA Extraction From Whole Blood for Rapid Sample Generation in Genetic Epidemiological Studies. *Front Genet* **11**, 374 (2020).
47. Bronner, I. F. & Quail, M. A. Best Practices for Illumina Library Preparation. *Curr Protoc Hum Genet* **102**, (2019).
48. Chiniquy, J., Garber, M. E., Mukhopadhyay, A. & Hillson, N. J. Fluorescent amplification for next generation sequencing (FA-NGS) library preparation. *BMC Genomics* **21**, 85 (2020).
49. Dallavilla, T., Marceddu, G., Casadei, A., de Antoni, L. & Bertelli, M. A fast, reliable and easy method to detect within-species DNA contamination. *Acta Biomedica* **91**, 1–8 (2020).
50. Hess, J. F. *et al.* Library preparation for next generation sequencing: A review of automation strategies. *Biotechnol Adv* **41**, 107537 (2020).
51. Huang, J. & Li, C. A modified protocol with less clean-up steps increased efficiency and product yield of sequencing library preparation. *3 Biotech* **12**, 111 (2022).
52. Muscarella *et al.* Automated Workflow for Somatic and Germline Next Generation Sequencing Analysis in Routine Clinical Cancer Diagnostics. *Cancers (Basel)* **11**, 1691 (2019).

53. Tegally, H., San, J. E., Giandhari, J. & de Oliveira, T. Unlocking the efficiency of genomics laboratories with robotic liquid-handling. *BMC Genomics* **21**, 729 (2020).
54. Hess, J. F. *et al.* Automated library preparation for whole genome sequencing by centrifugal microfluidics. *Anal Chim Acta* **1182**, 338954 (2021).
55. Kim, H. *et al.* A Microfluidic DNA Library Preparation Platform for Next-Generation Sequencing. *PLoS One* **8**, e68988 (2013).
56. Tan, S. J. *et al.* A Microfluidic Device for Preparing Next Generation DNA Sequencing Libraries and for Automating Other Laboratory Protocols That Require One or More Column Chromatography Steps. *PLoS One* **8**, e64084 (2013).
57. Murphy, T. W., Hsieh, Y.-P., Zhu, B., Naler, L. B. & Lu, C. Microfluidic Platform for Next-Generation Sequencing Library Preparation with Low-Input Samples. *Anal Chem* **92**, 2519–2526 (2020).
58. Plastics - Thermal Conductivity Coefficients. Preprint at https://www.engineeringtoolbox.com/thermal-conductivity-plastics-d_1786.html.
59. Material Properties. Preprint at https://www.engineeringtoolbox.com/material-properties-t_24.html.
60. Smit, M. L. *et al.* Automated extraction and amplification of DNA from whole blood using a robotic workstation and an integrated thermocycler. *Biotechnol Appl Biochem* **32**, 121 (2000).
61. Aird, D. *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* **12**, R18 (2011).
62. Ávila-Ríos, S. *et al.* Next-generation sequencing for HIV drug resistance testing: Laboratory, clinical, and implementation considerations. *Viruses* vol. 12 Preprint at https://doi.org/10.3390/v12060617 (2020).
63. Zhang, L. *et al.* One-Step Ligation on RNA Amplification for the Detection of Point Mutations. *The Journal of Molecular Diagnostics* **17**, 679–688 (2015).
64. Zhuang, F., Fuchs, R. T., Sun, Z., Zheng, Y. & Robb, G. B. Structural bias in T4 RNA ligase-mediated 3′-adapter ligation. *Nucleic Acids Res* **40**, e54–e54 (2012).
65. Li, T. W. & Weeks, K. M. Structure-independent and quantitative ligation of single-stranded DNA. *Anal Biochem* **349**, 242–246 (2006).

66. Bauer, R. J. *et al.* Comparative analysis of the end-joining activity of several DNA ligases. *PLoS One* **12**, e0190062 (2017).
67. Comparison of fluorescence-based quantitation with UV absorbance measurements. Preprint at https://www.k-state.edu/igenomics/services/documents/comparisson%20between%20qubit.pdf.
68. Best practices for manually normalizing library concentrations. Preprint at https://knowledge.illumina.com/library-preparation/general/library-preparation-general-reference_material-list/000001252.
69. Quality Scores for Next-Generation Sequencing . Preprint at https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf.
70. Piovesan, A. *et al.* On the length, weight and GC content of the human genome. *BMC Res Notes* **12**, 106 (2019).
71. Matsumura, Y., Yamamoto, M., Nakano, S. & Nagao, M. Complete Genome Sequence of Escherichia coli ME8067, an Azide-Resistant Laboratory Strain Used for Conjugation Experiments. *Genome Announc* **6**, e00515–18 (2018).
72. Tanaka, N. *et al.* Sequencing artifacts derived from a library preparation method using enzymatic fragmentation. *PLoS One* **15**, e0227427 (2020).
73. Nutman, A. & Marchaim, D. How to: molecular investigation of a hospital outbreak. *Clinical Microbiology and Infection* **25**, 688–695 (2019).
74. Quainoo, S. *et al.* Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. *Clin Microbiol Rev* **30**, 1015–1063 (2017).
75. Peymani, F., Farzeen, A. & Prokisch, H. RNA sequencing role and application in clinical diagnostic. *Pediatric Investigation* vol. 6 29–35 Preprint at https://doi.org/10.1002/ped4.12314 (2022).
76. Li, L. *et al.* What are the applications of single-cell RNA sequencing in cancer research: a systematic review. *Journal of Experimental and Clinical Cancer Research* vol. 40 Preprint at https://doi.org/10.1186/s13046-021-01955-1 (2021).
77. Liyanapathirana, V. *et al.* Application of a target enrichment-based next-generation sequencing protocol for identification and sequence-based prediction of pneumococcal serotypes. *BMC Microbiol* **14**, (2014).

78. Singh, R. R. Target Enrichment Approaches for Next-Generation Sequencing Applications in Oncology. *Diagnostics* vol. 12 Preprint at https://doi.org/10.3390/diagnostics12071539 (2022).
79. Duarte, T. T. & Spencer, C. T. Personalized proteomics: The future of precision medicine. *Proteomes* vol. 4 Preprint at https://doi.org/10.3390/proteomes4040029 (2016).
80. Lim, C. T. & Zhang, Y. Bead-based microfluidic immunoassays: The next generation. *Biosensors and Bioelectronics* vol. 22 1197–1204 Preprint at https://doi.org/10.1016/j.bios.2006.06.005 (2007).