Enhancing Next-Generation Sequencing Library Preparation Efficiency by Integrating RNA-Based Methods and Comprehensive Automation Approaches

By

Jennifer N. Figueroa-Cruz

B.S., Polytechnic University of Puerto Rico, 2022

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 2024

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.

04/26/2024 Jennifer Figueroa-Cruz

Date Jennifer N. Figueroa-Cruz, 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.

04/26/2024 Jennifer Figueroa-Cruz

Date Jennifer N. Figueroa-Cruz, Author

This thesis by Jennifer N. Figueroa-Cruz is accepted in its present form by the Center of Biomedical Engineering as satisfying the thesis requirement for the degree of Master of Science.

Date 04/25/24

Dr. Anubhav Tripathi, Advisor

Recommended to the Graduate Council

Date 4/25/24

Dr. Theresa Raimondo, Reader

Date 04/25/24

Dr. Edith Mathiowitz, Reader

Approved by the Graduate Council

Date 04/25/2024

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

CURRICULUM VITAE

Jennifer N. Figueroa-Cruz

Education

Master of Science in Biomedical Engineering

Brown University, Providence, RI Expected Graduation Date May 2024

Relevant Courses:

Instrumentation Design Biomolecular Interactions

Implantable Devices Molecular Genetics

Advisor: Dr. Anubhav Tripathi

Bachelor of Science in Biomedical Engineering

Polytechnic University of Puerto Rico, Hato Rey, PR July 2022

Research and Engineering Experience

Automated Systems Engineering Team, Moderna, Inc. Norwood, MA

Automation Engineering Co-Op July 2024 – February 2023

- Developed, optimized, and troubleshot an array of methods on different types of Hamilton Liquid Handlers across different labs and divisions.
- Constantly coordinated with different teams to get their method requirements, manage tests for method changes, and get feedback before and after the changes were released.
- Used multiple Moderna in-house and outsourced software to carry out my everyday work.

Dr. Tripathi Lab Group, Brown University Providence, RI

Revvity-Sponsored Project July 2022 – Present

Research and Development Engineer

- Engineered NGS Library Preparation Assays for commercialization on the BioQule NGS System.
- Developed different DNA/RNA-Seq, and PCR-Free automated sample preparation protocols and helped in the process of WGS data analysis for them.
- Have developed and helped with different Python scripts to automate the assays.

  Also, used Bitbucket to store and share all the scripts with all of the groups

  involved in the project.

Dr. Tian Lab, Photonics Center, Boston University Boston, MA

NSF Research Experience for Undergraduate Students June-August 2021

- Automized and integrated a GigE Spinnaker Camera with a Thorlabs Kinesis DC Servo Stage using Python, Micro-Manager, and Pycro-Manager Software. This work led to a successful automatization of the data collection process while using miniature optics.
- After running the tests and gathering the necessary data, we put together a poster

  presentation for the end-of-program symposium.

Publications

- "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).

Presentations

- SLAS2024 Poster Presentation "Accelerating Discoveries: A Comprehensive

  RNA-Based Next Generation Sequencing Library Automation Approach"

Jennifer Figueroa Cruz, Khaliun Myagmar, Khulan Unurbuyan, Sabrina Tolppi, Olivia Ott, Anubhav Tripathi.

- SLAS2024 Poster Presentation “Optimization of Tagmentation Chemistries and Workflows for Automated Next-Generation Sequencing Library Preparation” Sabrina Tolppi, Khulan Unurbuyan, Khaliun Myagmar, **Jennifer Figueroa Cruz**, Olivia Ott, Anubhav Tripathi.
- SLAS2023 Poster 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.
- Boston University Photonics Center NSF REU in Integrated Nanomanufacturing 2021 Poster Presentation “Automated Data Collection for Testing Miniature Optics” **Jennifer Figueroa-Cruz**, Joseph Greene, Lei Tian.

ACKNOWLEDGMENTS

I would like to start by thanking Dr. Tripathi for being my research advisor and the mentorship provided. During my time at the lab your advice and encouragement towards me helped me grow both professionally and personally. I got lucky to get an advisor that apart from believing in me, pushed and challenged me to go above and beyond. Thank you for everything!

I would like to thank Dr. Gray for her support and advice about academics and department matters towards me. I would also like to thank Dr. Raimondo and Dr. Mathiowitz for agreeing to be members of my committee and supporting me throughout this process.

I would like to thank Sabrina Tolppi, Khaliun Myagmar, Khulan Unurbuyan, and Olivia Ott. They welcomed me with open arms into the lab and initially trained me to be up to par with them to carry out or work. Our collaboration and friendship made me grow as a person and researcher and for that I'm thankful. I want to thank Kathryn Whitehead who I had the privilege to collaborate with once she joined the lab in my second year. We have been able to teach each other about different engineering and bioinformatics topics which has helped me broaden my horizons and grow.

I would like to thank Revvity for their grant towards Brown University which helped for the development of this project.

Finally, I would like to thank my parents and sister Jannette Cruz Lugo, Jorge Figueroa Santos, and Janice Figueroa Cruz. I would have not been able to get this far without their unwavering and unconditional support on all of my endeavors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

ABSTRACT

The use of RNA for diagnosis and treatment becomes more rampant with each passing day. One of the main techniques used to analyze RNA is Next Generation Sequencing (NGS). NGS RNA-Seq workflows are known to be labor-intensive and at times difficult to manage because of the nature of the enzymes and RNA. This thesis sets out to find comprehensive ways to automate an array of NGS Library Preparation assays with an added focus on RNA-Seq workflows. The final automated workflows generated for each of the assays were able to produce standardized final libraries comparable to manual samples (positive controls). This was determined by doing quality control and analyzing an array of sequencing parameters for each of the final libraries. To achieve these results, several iterations of the process were tested. Parameters such as incubation times, reagent volumes, mixing techniques, and temperature were iterated until we obtained the results closest to those of the already validated manual ones. Lastly, challenges often encountered by smaller research centers with limited resources were taken into account to make these processes accessible to all.

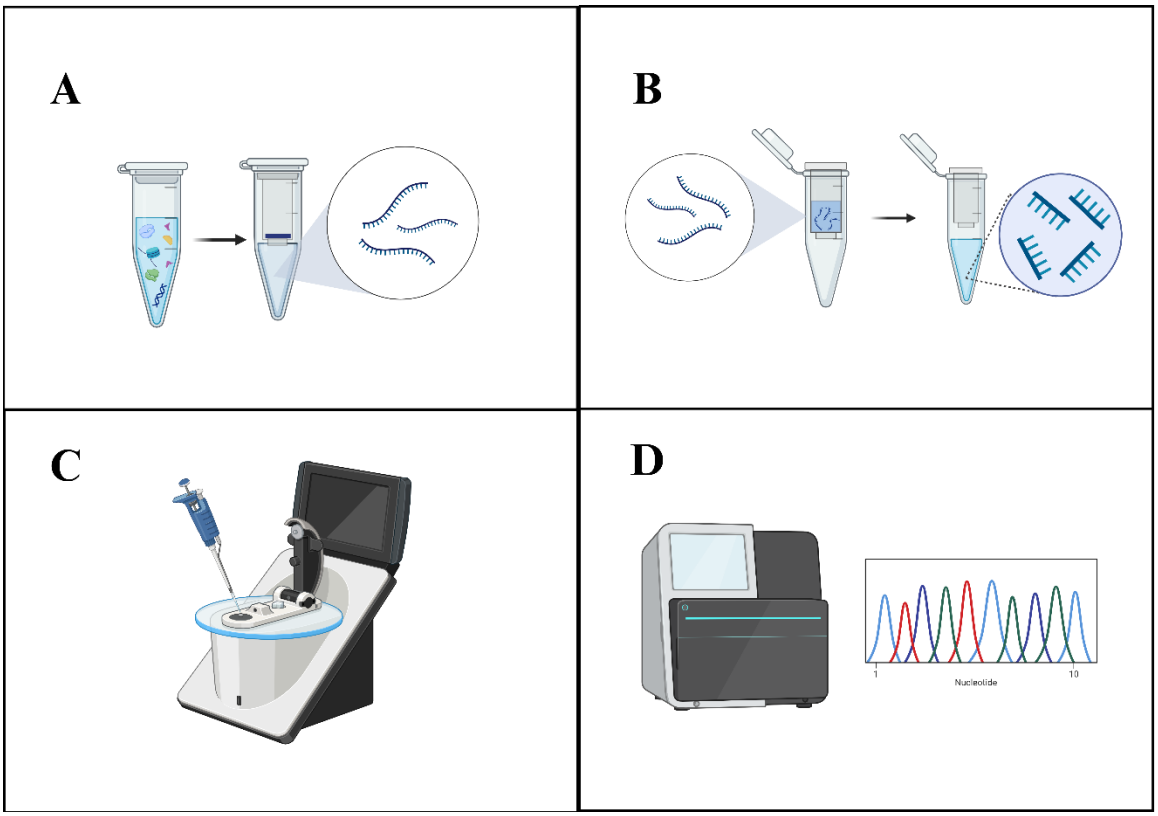
Chapter 1

Background

1.1 TOTAL RNA COMPOSITION AND ANALYSIS

Ribonucleic Acid (RNA) is a crucial molecule present in living organisms. Some of the biological functions that this nucleic acid is responsible for are cellular regulation, gene expression, and protein synthesis[1]. Total RNA is composed of transfer RNA (tRNA), messenger RNA (mRNA), ribosomal RNA (rRNA), and non-coding RNAs (ncRNAs). Because of how rich in information Total RNA is, analyzing it is of great importance to innovate clinical diagnostics and treatment techniques. To be able to do Total RNA analysis the researcher must first follow a series of steps which include the extraction, purification, quantification, and characterization of the RNA (Fig. 1.1). The different techniques to analyze Total RNA are microarray analysis, reverse transcription polymerase chain reaction (RT-PCR), and Next Generation Sequencing (NGS)[2]. Microarray analysis allows for the inspection of gene expression of thousands of mRNAs at the same time[3]. On the other hand, RT-PCR converts RNA into complementary DNA (cDNA) to be able to amplify it and do detection to quantify gene expressions[4]. Lastly, by employing RNA sequencing (RNA-seq) on NGS platforms we can identify different RNA species and splicing events and quantify gene expression[5].

Figure 1.1: Visual representation of the steps needed to do RNA analysis. (A) RNA extraction, (B) purification, (C) quantification, and (D) characterization.

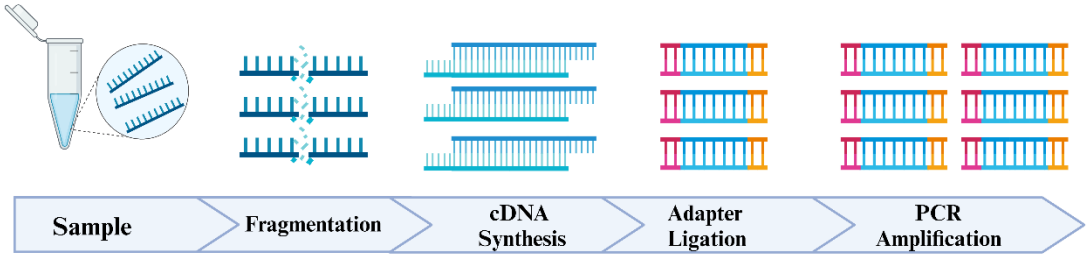


1.2 NEXT GENERATION SEQUENCING (NGS)

Next Generation Sequencing (NGS) is a technology that allows for the sequencing of hundreds of thousands of nucleic acid molecules in parallel[5]. This phenomenon is also known as massively parallel sequencing. NGS has stirred up the world of molecular research because of its efficiency, scalability, high throughput of information, and affordability. The fundamentals of an entire NGS run include library preparation, sequencing, and lastly data analysis. The first step of library preparation is the fragmentation of the nucleic acid (DNA or RNA) to turn it into shorter sequences. Next, adapters specific to sequencing platforms are ligated to the sequence fragments to be able to identify which nucleic acid sample is which. The library preparation's final step is

amplifying the fragments through PCR[6]. For RNA-Seq, an extra step is added between the fragmentation and adapter ligation steps. This step carries out a reverse transcription of the RNA fragments and turns them into complementary DNA (cDNA) fragments (Fig. 1.2)[7]. After the library preparation is done the samples are fed into a sequencing platform where typically millions of sequencing reads are created, and the product is raw sequencing data. Lastly, the raw sequencing data is passed through an array of software, some used for DNA-Seq and others for RNA-Seq, to generate the necessary plots, graphs, and charts to be able to analyze the data.

Figure 1.2: General workflow of a Next Generation Sequencing RNA-Seq Protocol



1.3 AUTOMATION IN MOLECULAR BIOLOGY

The introduction of automation into the laboratory space has transformed the way that research is carried out and opened the door for innovative diagnostic and treatment technologies in the molecular biology world. Some of the main and most important benefits of automation in research are increasing throughput, standardizing workflows, decreasing human error and manual labor, and strengthening experimental precision[8]. One of the most popular examples of where we can see automation being introduced into the laboratory is using robotics and automated systems such as liquid handlers to carry out processes such as PCR, nucleic acid isolation, and NGS[9]. For NGS library preparation workflows

automation is of great importance. As detailed before, NGS library preparation employs the use of several enzymatic steps, size selection, adapter ligation, and sample purification. This process is labor intensive and limiting in the number of samples that can be done precisely at once when done manually. This process is also sensitive to human error, the same sample could look different every time just because of the speed, intensity, and number of times the sample was pipetted. This fact is one of the main reasons why automating these protocols is of great importance.

1.4 CHALLENGES AND OPPORTUNITIES IN TOTAL RNA ANALYSIS, NGS, AND AUTOMATION

With all the opportunities introduced by Total RNA analysis, NGS, and automation, come challenges for research centers with limited resources that need to be addressed. Firstly, while the amount of information that Total RNA possesses is a positive thing, it is also considered a challenge to analyze it thoroughly—the components found in Total RNA present different regulatory mechanisms, expression patterns, and functions[10]. To be able to differentiate between each component and quantify them correctly complex computational techniques and bioinformatics tools need to be introduced into the process. This makes the process of Total RNA analysis not user-friendly, labor intensive, and sometimes not cost-effective because of the extra personnel needed and the cost of all the programs combined. Secondly, NGS presents some similar challenges. Even though sequencing chemistries have been optimized throughout the years, the way that NGS works, it is receptive to biases and sequencing mistakes that make the analysis and interpretation of the generated data more complicated[11]. Also, introducing automated

systems into the laboratory requires a big investment upfront. This investment includes the purchase of the equipment/s, the space where this equipment will be located, and the training of personnel[12]. Because of these challenges most limited resources research centers cannot or choose to not integrate automation into their workflows. This work sets out to break these barriers and provide an innovative solution to all the challenges mentioned above.

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 high 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.

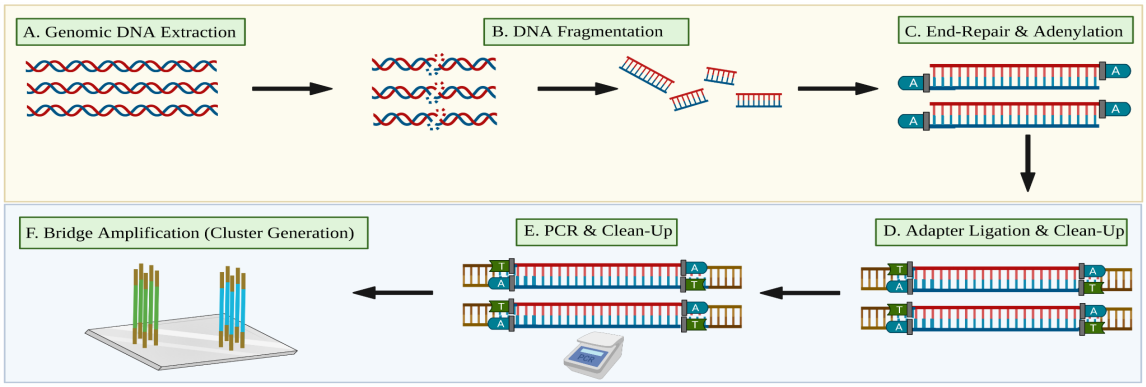
2.2 INTRODUCTION

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[13-16], rare disorder diagnosis[17-19], and prenatal screening[20]. NGS is a markedly valuable tool for diagnostics and discovery because it generates millions of reads and is exceedingly sensitive to nucleic acid mutations[21-24]. Unfortunately, while innovations in sequencing technology have reduced requisite cost, time, and workload,

sample (library) preparation has largely remained rudimentary and labor-intensive[25-28]. 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[29]. A typical NGS library preparation begins with the extraction of genomic DNA (gDNA)[30] (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[31, 32]. 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)[24]. 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[24]. 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 [22, 24].

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.



It is clear that NGS sample preparation requires many sequential steps using specialized laboratory equipment. Hence, NGS library preparation is expensive, cumbersome, and susceptible to contamination and pipettor error[33-36]. 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[37], 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[38].

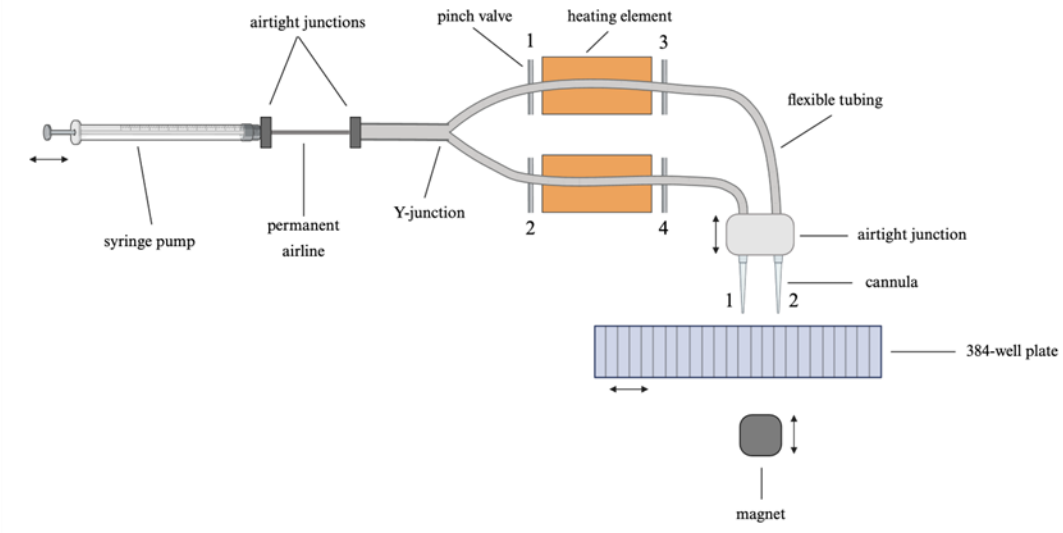
Expensive, high-throughput robotic liquid handlers dominate research and clinical applications in core laboratories, however small laboratories often have neither the funds nor the sample demand to warrant a high-throughput automation solution. Small laboratories are consequently often restricted to labor-intensive manual protocols that introduce many opportunities for sample contamination and manual pipetting errors due to various factors. Microfluidic approaches are one group of solutions often more suitable for low-throughput and single-use library preparation applications[22, 24, 31, 35, 39-42]. 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. Select microfluidic approaches, such as the nanoliter-scale column chromatography method described by Tan et. al., have scaled successfully to be high-throughput and

commercially viable[41]. Tan et. al. utilized a novel architecture, with a motif of two repeating fluidic circuits that employ a combination of commercially available reagents to achieve library preparation with minimal manual intervention. Still, full automation at the nanoscale for hundreds of parallel reactions demands low-tolerance chip manufacturing and integration with a high-throughput liquid handler, resulting in high investment costs for its associated equipment. 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[25-28]. Our previous work started with the innovative concept of integrating liquid mixing steps with thermodynamically driven steps such as ligation and amplification via a single capillary tube heating system[25-28]. This work utilized magnetic bead motion through a microchannel for purification step, as needed by steps C and D in Figure 1. Schneider et al achieved a high level of integration with minimal hands-on time, yet only one sample could be prepared at a time, with some concerns of microfluidics chip cost and manufacturability. Additionally, scalability of microfluidics chips for many library preparation chemistries, each requiring different amounts of bead solution, is sometimes operationally challenging. Hence, the initial proof-of-concept platform required further innovation for it to be globally adopted by researchers and clinicians with reduced complexity and cost, and a high degree of robustness and ease-of-use.

In this work, we present an innovative automated NGS library preparation system with a simplified mechanical design that exploits both macro- and microfluidic properties.

As shown in Figure 2.2, an air-displacement syringe pump is connected to two cannulas via permanent airlines then disposable flexible tubing that passes through pinch valves and a heating element. Cannula 1 goes into the desired wells of the 384-well reagent plate to aspirate, dispense, and mix reagents, then transport necessary mixtures to the heating element (as desired in steps B, D and E in Fig. 1) with pinch valves 2 & 4 in the closed position. Similarly, cannula 2 operates with pinch valves 1 & 3 in the closed position. A magnet, which is located beneath the 384-well plate, is used for magnetic bead-based purification of nucleic acids as needed by steps C and E in Fig. 1. The pinch valves are used to pressurize the air surrounding liquid plugs in the heating section undergoing ligation or PCR. Our modular platform integrates fluid mechanics, heat transfer, molecular thermodynamics, mass transport, and biochemical reaction kinetics.

Figure 2.2: Schematic of our platform hardware, simplified to illustrate components for preparation of one sample. The air pressure syringe pumps are placed on an x-axis motion stage, the cannulas are placed on a z-axis motion stage, the 384-well reagent plate is placed on an x-stage, and the magnet is put on a z-axis motion stage.



Our refined design allows for consistent operation of eight library preparations in parallel, while maintaining low manufacturing and operational costs without any need of microfluidic chip-based bead cleaning. The platform's use of a 384 deep-well plate for all liquid handling offers a centralized, efficient, and adaptable workflow. 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 allows 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. Although in this work we demonstrate NGS library preparation assay automation, our innovative platform could be easily programmed to perform nucleic acid extraction protocols, various end-point PCR assays, and protein detection assays utilizing the same reagent plate and cartridge consumables. Our platform combines the reliability of large pipetting workstations with the low cost and enclosed environment of microfluidic solutions, while striking a balance between integration and versatility.

2.3 RESULTS AND DISCUSSION

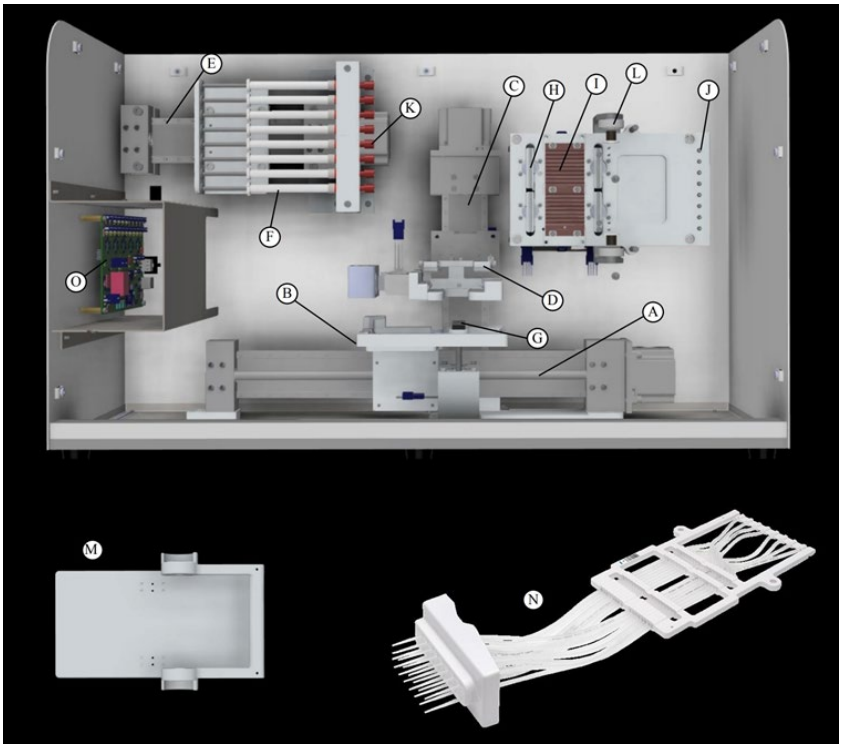
I. Optimization Of Mechanical Design

The principal innovation of our method is the simplified design layout of the platform (Fig. 2.2&2.3) based on the fundamental needs and constraints of fluidics, reactions, adsorption-desorption, and mixing. All NGS library preparation workflows (Fig. 2.1)

require several reagents and buffers at working volumes between 2µL – 200µL. To accommodate these working volumes 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). Narrow dimensions of wells offer faster diffusional ($\tau_d \sim w^2/2D$) and convective ($\tau_c \sim w^2 L/Q$) mixing. Here, D is the molecular diffusivity, and Q is the flow rate of liquid dispensing. A deep-well plate also offers adequate space to arrange and store reagents wells to avoid any carry-over contamination. Deep wells, in addition to a pierceable plate seal, help to prevent aerosol contamination of nearby reagent wells during liquid handling. Aerosol contamination of the airlines is prevented by using sufficiently long segments of cartridge tubing, as well as slow pump speeds to maintain volumetric accuracy of pipetting. 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 -20ºC. The platform enables movement of the 384-well plate in only a single dimension (x-stage for left-right motion) (Fig. 2.3A-B). A separate z-stage (for up-down motion) (Fig. 2.3C-D) holds two cannulas (Fig. 2.3N) 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. The two cannulas are connected via polyethylene capillary tubing (Fig. 2.3N) which pass through a thermal heating plate (Fig. 2.3I) and a set of two pinch valves (Fig. 2.3H-I). The pinch valves are used to select one of two cannulas by pinching the undesired cannula against the PCR door, allowing for use of a single syringe pump to control a bifurcated airflow. Our platform offers the option of two separate cannulas (Fig 2.3N) for various reactions in the library preparation workflow (Fig. 2.1). Our device uses a grooved copper plate that accommodates cartridge polyethylene tubing to maximize the thermal heating efficiency (Fig. 2.3I). Roughly one-

third of the tubing directly contacts the copper surface and is held securely in place by an insulated detachable door (Fig. 2.3M). 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. 2.3G). 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 2.3: Schematic of our platform. (A) X-stage for reagent plate motion, (B) 384 deep-well plate holder, (C) Z-stage for cannula motion, (D) Cartridge cannula spring-loaded clamp, (E) X-axis for pump syringe motion, (F) Syringes, (G) Magnet in raised position for bead-based purification, (H) Pinch valve, (I) Copper heating plate, (J) O-rings for cartridge-airline interface, (K) Airline connection point for plastic tubing (not pictured) to extend to (J), (L) Anchoring point for (M) insulated PCR door, (N) Disposable cartridge with cannula tips forms an airtight seal with (J) and is secured to the device via (M) and (D). (O) Printed circuit board for platform



II. 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 and corresponding capillary tubes to perform all liquid transfers and mixing functions for the entire library preparation assay. In development of the platform, we found that reuse of tubing following fragmentation or end repair and adenylation (ERA) reactions (Fig. 2.4B) negatively impacted subsequent steps, while adsorbed reagents from wash and PCR steps (Fig. 2.4C-F) were less likely to impair assay performance. As a result, one cannula is used for fragmentation or ERA reactions, while the other is reserved for all other steps (Fig. 2.4A,B,E). Our results below demonstrate that library preparation assays are surprisingly tolerant of surface reuse, allowing for design simplification as long as individual sample lanes do not contaminate each other. For a high-throughput liquid handler, similar surface reuse may be possible, though we anticipate that a full study of assay performance and cross-contamination would need to be conducted to ensure that there are no unintended consequences. We have optimized the use of the capillary tubes based on adsorption-desorption characteristics and by incorporating a cartridge surface passivation step that must be completed for every new run. Experiments were performed to optimize pre-treatment (PT) conditions using mixture components, fluid slug volume, slug routing path, temperature, and incubation time as variables. The pretreatment is comprised of water, Tween-20, ssDNA oligo, BSA protein, and ERA buffer (full mixture specified in S. Table 1A). 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

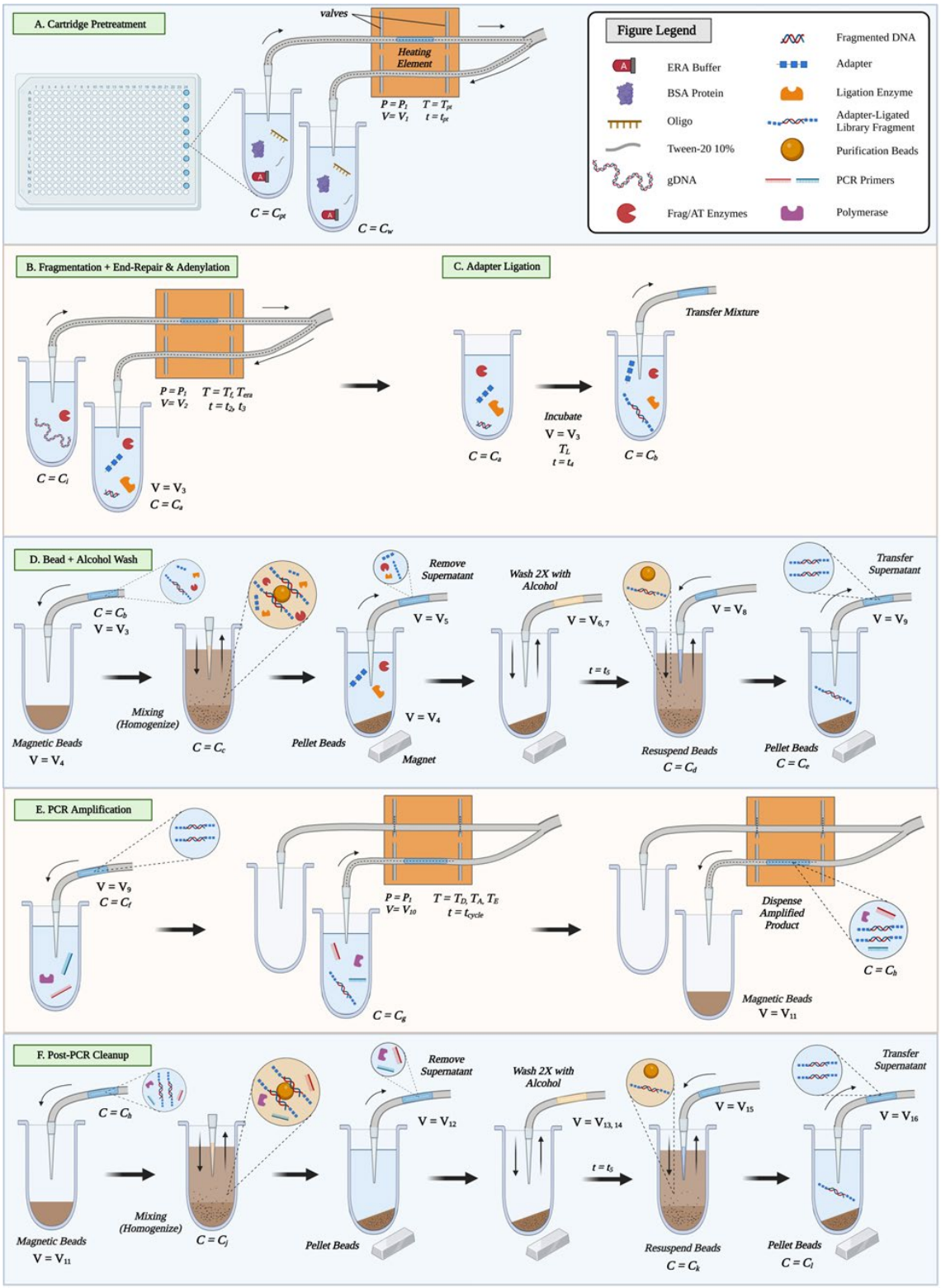
the heating element, undergoes heating for t_{PT} = 10 minutes at T_{PT} = 75 ℃, and is passed through the Y-junction of the cartridge and discarded (Fig. 2.4A). This step successfully prevents the adsorption of enzymes and DNA to the polymeric surface of the cartridge tubing. Without the optimized PT, final library concentrations for the mechanical fragmentation assay average 11.42±3.66 ng/μL compared to an average of 33.12±6.54 ng/μL with the PT. In addition to low yield, these samples have tall adapter dimer peaks (S. Fig. 5), highlighting the importance of PT for generating high quality libraries.

We optimized the fragmentation and ERA reaction volumes to be V_2 = 50 μL for both assays (Fig. 2.4B). These reactions contain input DNA, buffer, and the appropriate fragmentation or ERA enzymes. Fluid mixing in this step is performed in the 384-well plate using repeated pipetting up and down at moderately slow speeds (S. Table 1)(see Results Section III for further discussion of mixing techniques and pump speeds). The fluid slug is transported in the capillaries to the heating element to achieve optimum enzyme activity at the desired temperature (room temperature incubation during the PT step then t = 30 min at T_{era} = 65℃ for mechanical fragmentation, and t_2 = 15 min at T_f = 37 ℃, t_3 = 30 min at T_{era} = 65 ℃ for enzymatic fragmentation, Fig. 2.4B). Note that the fluid slug is constrained by pinch valves (Fig. 2.3H) and pressurized to applied pressure P_1 = 27.89 psi during heating to 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. 2.4C). The ligation reaction is mixed with a custom mixing technique (see Results Section III) and incubated at room temperature (T_L = 25 ℃) for time t_4 = 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_4 = 100 µL for mechanical fragmentation, V_4 = 60 µL for enzymatic fragmentation) (Fig. 2.4D). Several custom mixing techniques form the basis of the SPRI bead-binding and DNA elution reactions, with gentle alcohol washes in between. An optimized NGS library prep wash step is expected to remove any residual alcohol through bead drying and thorough removal of supernatant. The eluate of the wash 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_9, Fig. 2.4E). The fluid slug is again pressurized to optimize heat transfer, and after performing the PCR cycle for a programmed amount of time (specified in S. Table 2.1E), 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 remaining PCR reagents and small adapter-ligated DNA fragments < 200 bp (Fig. 2.4F). The final supernatant removal in this step containing the desired DNA library is dispensed as the final product in a specified plate column, ready for quantification and quality control (V_{16} = 30 µL for the mechanical fragmentation assay and V_{16} = 20 µL for the enzymatic fragmentation assay).

Figure 2.4: Details of our optimized method. Both the mechanical and enzymatic fragmentation assays follow the protocol depicted below, differing in select incubation times and reaction volumes (See S. Table 1). (A) Surface passivation (Pretreatment) of the cartridge tubing. (B) Fragmentation, end repair, and adenylation reaction at the heating element, (C) adapter ligation reaction, (D) post-ligation SPRI bead purification, (E) Capillary PCR amplification, (F) post-PCR SPRI bead purification. Variables T = temperature, t = time, P = pressure, V = volume, and C = concentration. Note that fragmentation only occurs on-platform for the enzymatic fragmentation assay. Cannula tip positions relative to the well are not exact,

as there are multiple z-axis positions per step.[48]



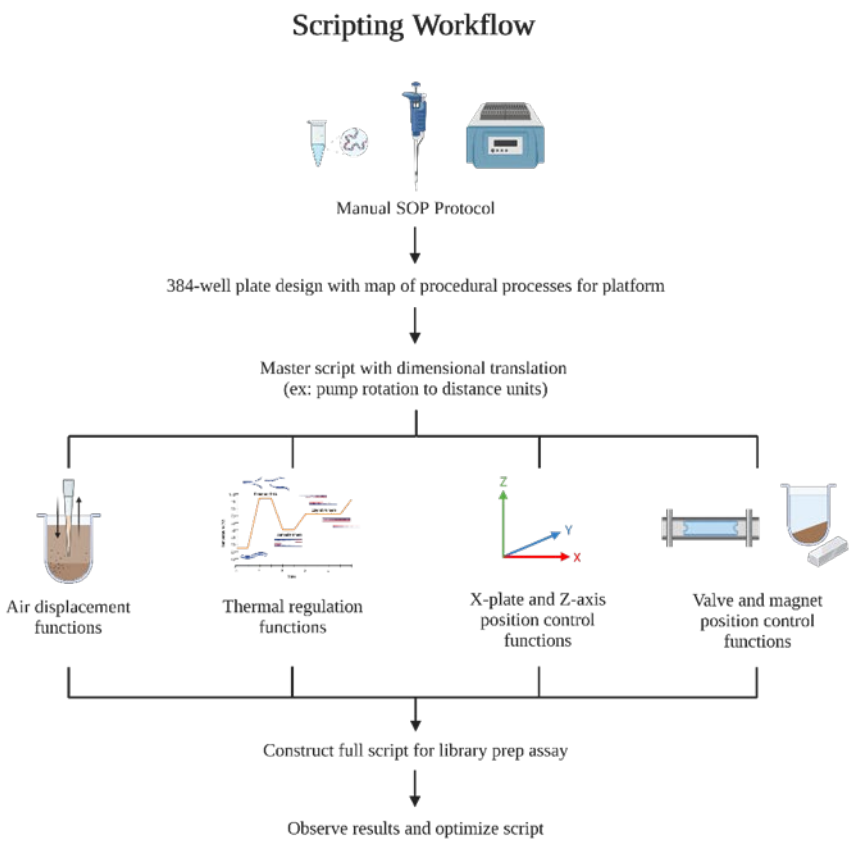
III. Analysis Of Fluid Mechanics And Platform Robotic Control

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. The pump motor utilizes two instructions from the assay script (a) set position and (b) set velocity. The final position determined the pressure and speed was selected based on viscosity of the liquid. Our analysis of volumetric accuracy and %CV (S. Fig. 2) 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-platform volumetric loss. Assay development begins with manually performed variations of the SOP protocol, followed by reaction fine-tuning to adapt variables such as reagent volumes, reaction temperatures, and incubation times to our platform (Fig. 2.5). Once these parameters have been determined, a plate layout may be implemented, where assay reagents are arranged into columns of the 384-well plate (S. Fig 1). 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 local scripts may call on for development of each individual assay (Fig. 2.5).

A preliminary assay script is written with the proposed plate layout, then liquid testing with dyed water in place of reagents is performed to identify potential mistakes, failures, or further points of iteration. The local script can be partitioned to perform liquid tests on individual 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 2.5: Scripting workflow. Standard operating procedure (SOP) for script and assay development.



Assay scripting is a highly informed process, with script writers aiming to control fluidic properties of the platform through refined motor movement. Relative reagent viscosity, estimated by manual pipetting, is used as a benchmark to select a pump speed for liquid transfers and mixing techniques. Operational pump speeds range from 1.67 µL/s (used only for liquid transfers of extremely viscous liquids) to 76.78 µL/s (used only for vigorous bead resuspension). Any newly implemented mixing or liquid transfer techniques must be assessed visually to ensure mixture homogeneity and precise aspiration and dispensation. High viscosity liquids must be aspirated and dispensed slowly, as they are prone to slow movement through cannula tips and cartridge tubing. Attempting high pump speeds with viscous liquids can result in volumetric inaccuracy or segmentation of fluid slugs. Moreover, techniques must avoid the introduction of bubbles in reaction mixtures, an indication of inappropriate pump speed or cannula tip positioning.

An example of these principles in practice is the custom ligation mixing technique. The ligation mixture of ligase enzyme, buffer, and adapters is highly viscous, and mixing must be done thoroughly yet carefully to prevent excessive adapter-dimer formation. Our mixing technique slowly aspirates a small volume of ligase enzyme/buffer, then a small volume of ERA/fragmentation mix combined with adapter, then repeats this process, alternating between the two mixtures. This technique increases surface area between the two mixtures, aiding diffusion while avoiding stringent mixing that may generate bubbles or adapter-dimers.

Bead-based purification steps have been a point of extensive script development, as these steps rely on liquid transfers orchestrated between many reagent wells (Fig. 2.4D,F) and custom mixing techniques that require precise movement of cannula tips. For

bead peletting, the magnet is positioned to the side of the well containing the bead pellet. Cannula tips are always located in the center of the well, thus an offset bead pellet allows for aspiration of supernatant from the bottom of the well without aspiration of SPRI beads (and subsequent loss of yield). Dispensation of alcohol onto bead pellets is performed at medium-low pump speeds (S. Table 1) to avoid disturbing the pellet. For binding and elution mixing techniques, pump speeds alternate between fast (vigorous) and slow (more gentle) motion (S. Table 1) to vary flow of SPRI beads throughout DNA-reagent mixtures. Resuspension of bead pellets requires high speed dispensation (S. Table 1) of elution buffer onto the pellet with cannula tips positioned deep in the well, followed by several rounds of up and down mixing. Mixing techniques use mixture volume as a script function parameter to ensure that the cannula tips are never positioned above the fluid surface while aspirating, preventing bubble formation.

In order to perform assay steps requiring heating, an elaborate pressurization and depressurization routine is employed. Scripts instruct the fluid slug to approach the heating element, then close the left valve (Fig. 2.3H). The pump is then instructed to push air against the fluid slug (now immobilized by the left valve), reaching 27.89 psi. Finally, the right valve closes to encapsulate the pressurized, centered fluid slug for heating. Depressurization occurs with incremental movement of individual valves to prevent sudden movement of the fluid slug.

IV. 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 °C and t_3 = 30 min at T_{era} = 65 °C, followed by a 4 °C hold (Fig. 2.4B). The maximum conductive heat flow rate through the cylindrical tube wall (Q) during the reaction is estimated

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

when the thermal conductivity[31] of the polyethylene tube wall is $k \sim 0.33 \frac{W}{m\,^oC}$, the thickness of the tube wall is $\Delta r \sim 0.8\,mm$, the maximum temperature difference between the outer and inner surfaces (the conductive resistance) of the capillary is $\Delta T \sim 73$ °C, the tube radius is $r \sim 1.6\,mm$, and length of the slug is $L \sim 24.9\,mm$. Similarly, the maximum conductive[32] 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 $k \sim 0.598 \frac{W}{m\,^oC}$. 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

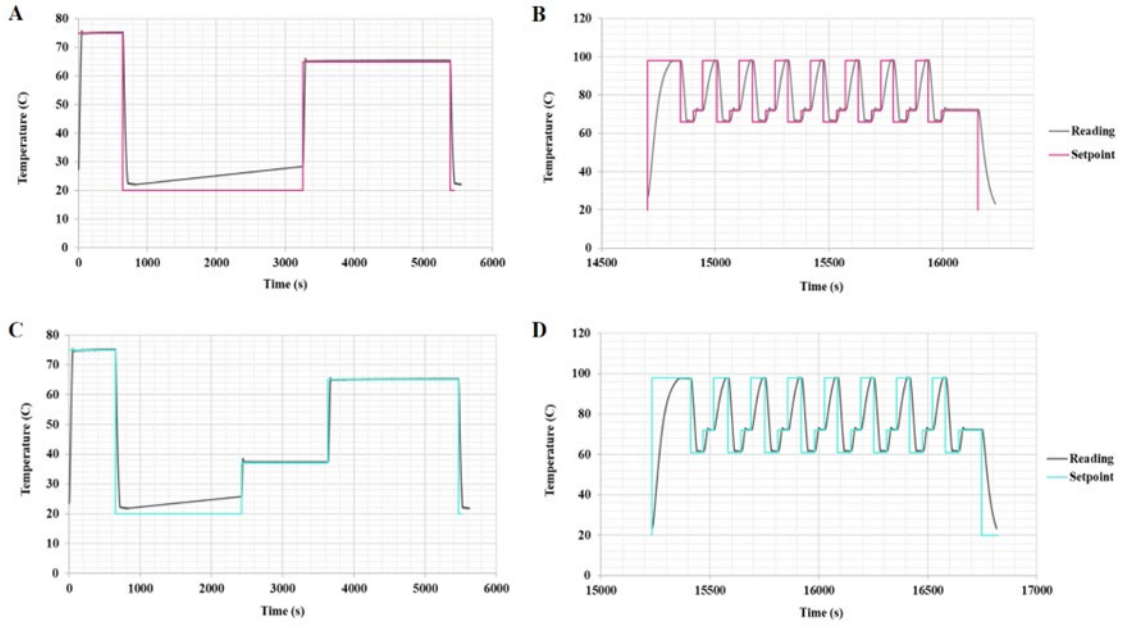
$$t_d \sim r^2/2\alpha \sim 2.2\,sec \quad (2)$$

owing to microscale diffusional length. Here, $\alpha = 0.15 \frac{mm^2}{s}$ is the thermal diffusivity[32] of the liquid, t_d is the thermal diffusion time, and $r = 0.8\,mm$ is the radius of the fluid slug. Furthermore, the thermal capacitance of the fluid slug $\sim \rho\pi r^2 L(4.18)\Delta T \sim 3.7\,J$ is small compared to the conductive heat flow $Q \sim$ 6.8 W. Hence, the time scale for heating the plug

is ~ $3.7/6.8 = 0.54s$, which is faster the timescale of resistive heat transfer. 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. Successful recreation of the sample environment allows our system to gather real-time, non-adjusted data of the reaction mixtures, which are slow to respond compared to the copper heating element. 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 6 shows the thermocouple temperature data compared against the scripted temperature setpoints for PT, fragmentation, ERA, and PCR during the mechanical fragmentation assay (Fig. 2.6A-B) and the enzymatic fragmentation assay (Fig. 2.6C-D). Total time for 8 cycles of PCR amplification is approximately 25 minutes (Fig. 2.6B,D), compared to 22 minutes on a standard thermal cycler. While our platform does have slightly slower ramp rates than a standard thermal cycler, we have not observed any significant primer/adapter dimer formation or negative impacts on amplification efficiency because of this slight difference. 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 2.6: 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. Note that the temperature drift (A, C) when the setpoint is 20°C is due to residual heat transferring to the

block containing the thermocouple. Reagents are in the 384-well plate during this incubation time, so there is no negative effect on assay performance.



The pressurization of capillary slugs between two pinching valves expands the capillary polymer tube between the copper heating plate on one side and thermal insulation on the other. It improves the thermal contact area and increases the heat transfer from the copper to the liquid slug as described above. Although the water vapor permeability through polymeric tubing increases at higher temperatures, the water loss in the system is minimal due to the lack of gradient water vapor concentration across the tube wall. Since the tubing is compressed between non-diffusive surfaces, the vapor quickly saturates the pores in the polymer tubing. A systematic study for quantifying evaporative loss at varying pressures is beyond the scope of this work, however, we did confirm that any volume loss was comparable to positive controls following heating steps. Note that pressurization likely prevents any bubble formation thatmay occur within fluid slugs, moreover, our data does not show any impact in amplification efficiency compared to positive controls. Given the

importance of maintaining pressure for achieving optimal heat transfer, 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 3) to acquire leak rates of the system and isolate specific airlines, syringes, or valves that may cause abnormal pressure readings.

Our thermal system design offered desired temperature profiles (as shown in figure 2.6) for PCR and ligation reactions as performed in a 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[45]; our sequencing results show no noticeable bias offered by thermal design and feedback system.

V. Library Preparation Efficiency

Library preparation efficiency is an important metric to quantify the success of our automated method compared to manually prepared positive controls (note that all positive controls discussed are prepared manually). Efficient library preparation has implications for sequencing success, clinical relevance, and conservation of resources such as cost and requisite input sample volume.

Our approach was to first prepare four on-platform plates, each with a separate set of manual positive controls: mechanical fragmentation assay with 1) human and 2) *E. coli* DNA input, and enzymatic fragmentation assay with 3) human and 4) *E. coli* DNA input. These four plates are analyzed below for final library concentration, library fragment distribution, as well as numerous sequencing metrics. After successful completion of these four plates, we gathered data across 7 on-platform plates for the mechanical fragmentation

assay and compared intra-run and inter-run consistency of final library concentration to positive controls to ensure robustness and repeatability of the instrument.

Controls: 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. 2.7A-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 $\pm$168.13 ng and 0.15 $\pm$ 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 $\geq$ 250 ng total, concentration $\geq$ 5 ng/µL, adapter-dimer $\leq$ 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. 2.7G). 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. 2.8A-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. 2.8G 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 2.7C and 2.8C 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 610.25 ± 128.93 ng and 0.44 ± 0.0041%, and 960.97 ± 176.47 ng and 0.28 ± 0.0056 %, respectively. An additional peak at ~150 bp corresponds to adapter dimers that formed during ligation[16]. 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. 2.8C)

suggests that some high molecular weight products remained post-fragmentation. However, the presence of this same hump in the positive controls (Fig. 2.8C, 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. 2.7C), but only for the libraries obtained on-platform. This could be a result of slightly more efficient manual purification steps, including bead-drying and alcohol wash removal, compared to on-platform purification in which the broad hump could be indicative of leftover enzymes or protein. Figure 2.7F shows an electropherogram overlay of *E. coli* positive control (red) and on-platform (blue) libraries generated by the mechanical fragmentation assay, while Figure 8F 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. 2.7A, 2.8A). 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 ± 105.62 ng and 0.51 ± 0.0022 %, and 2204.9 ± 493.78 ng and 0.23 ± 0.0034 %, 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. 2.8A). 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 2.7E shows an electropherogram overlay of human positive control (red) and on-platform (blue) libraries generated by the mechanical fragmentation assay, while Figure 2.8E 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 2.7: Mechanical fragmentation final library Quality Control. (A) Electropherograms of on-platform (Samples 1-6) and manual 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-platform (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-platform (blue) library overlay for yield and size distribution comparison. (F) E. coli DNA positive control (red) and on-platform (blue) library electropherogram overlay for yield and size distribution comparison. (G) Manual (red) and on-platform (blue) negative control electropherogram overlay.

All electropherograms consist of lower (first) and upper marker (last) peaks. Adapter-dimer peaks present at ~130bp (or ~55 sec). Other tiny peaks are either pertaining to primers or noise.

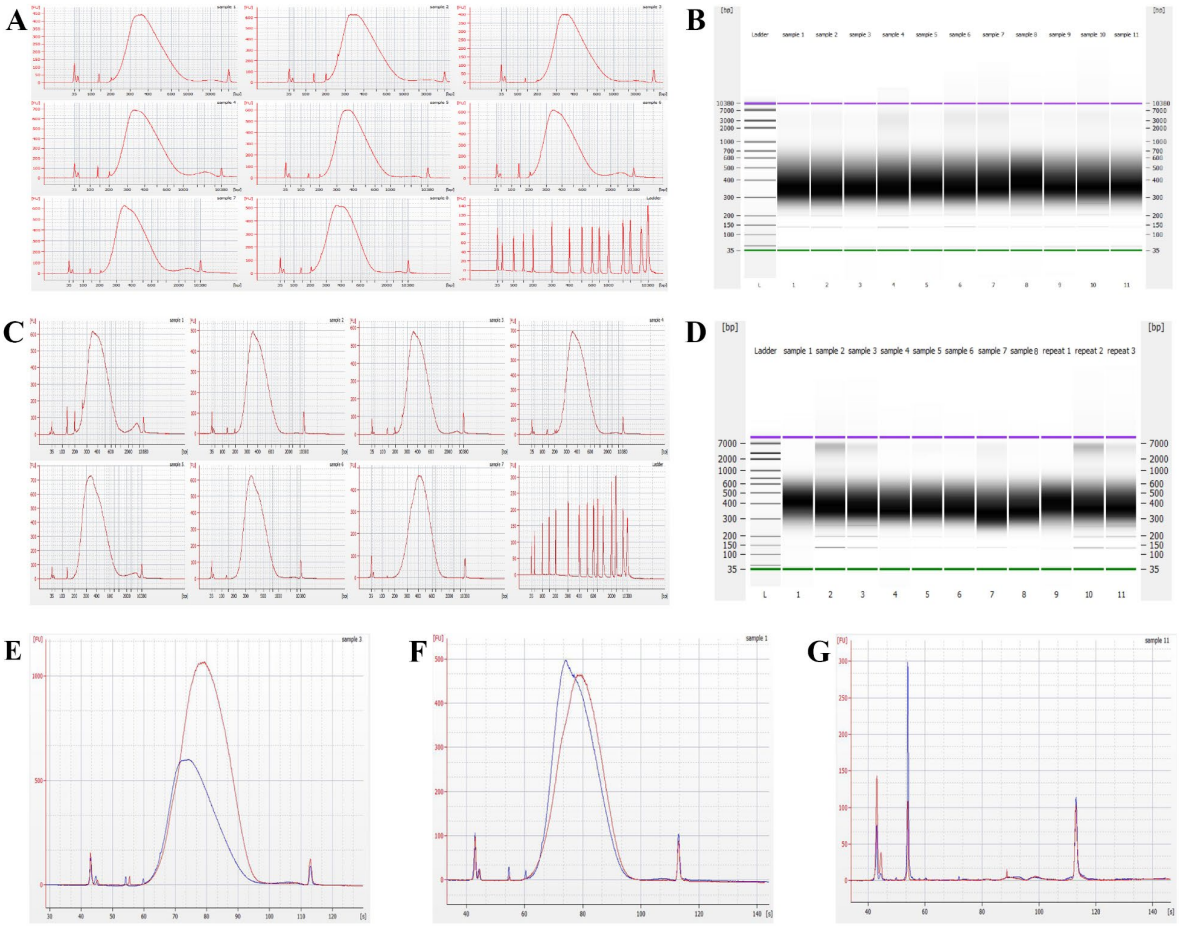
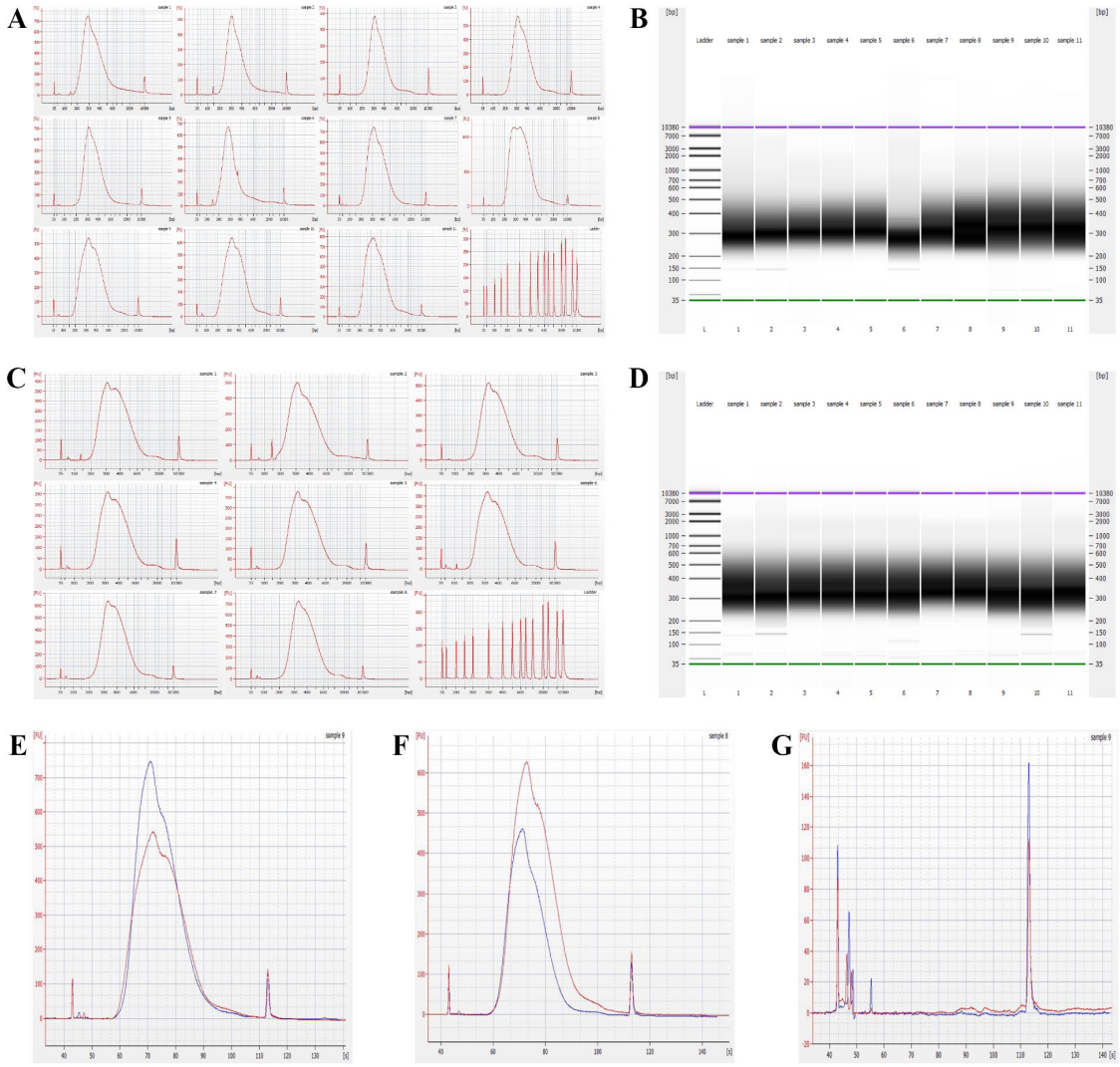


Figure 2.8: Enzymatic fragmentation final library Quality Control. (A) Electropherograms of on-platform (Samples 1-8) and positive control (Samples 9-11) human DNA libraries. The bottom right panel shows a DNA ladder for size comparison. (B) Gel image demonstrating fragment size distributions of the human DNA libraries shown in A. (C) Electropherograms of on-platform (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-platform (blue) library overlay for yield and size distribution comparison. (F) E. coli

DNA positive control (red) and on-platform (blue) library electropherogram overlay for yield and size distribution comparison. (G) Manual (red) and on-platform (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 step-wise 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[46-48]. We assume that fragmentation and ERA reaction efficiencies are close to 100%, considering the minimal presence of unfragmented gDNA in final libraries (Fig. 2.7).

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[49]. Ligation efficiency can range anywhere from 3-100%, with most ligases falling in the 15-40% range[50]. Our automated adaptation of the ligation reaction accounts for its high viscosity and sensitive nature, as discussed above. 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 largely dependent on bead formulation, including surface-coating consistency,texture, and buffer composition. 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[30]. 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 loss (discussed in efficiency analysis in terms of bead retention) via supernatant aspiration 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 2.9: Positive control and on-platform 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.

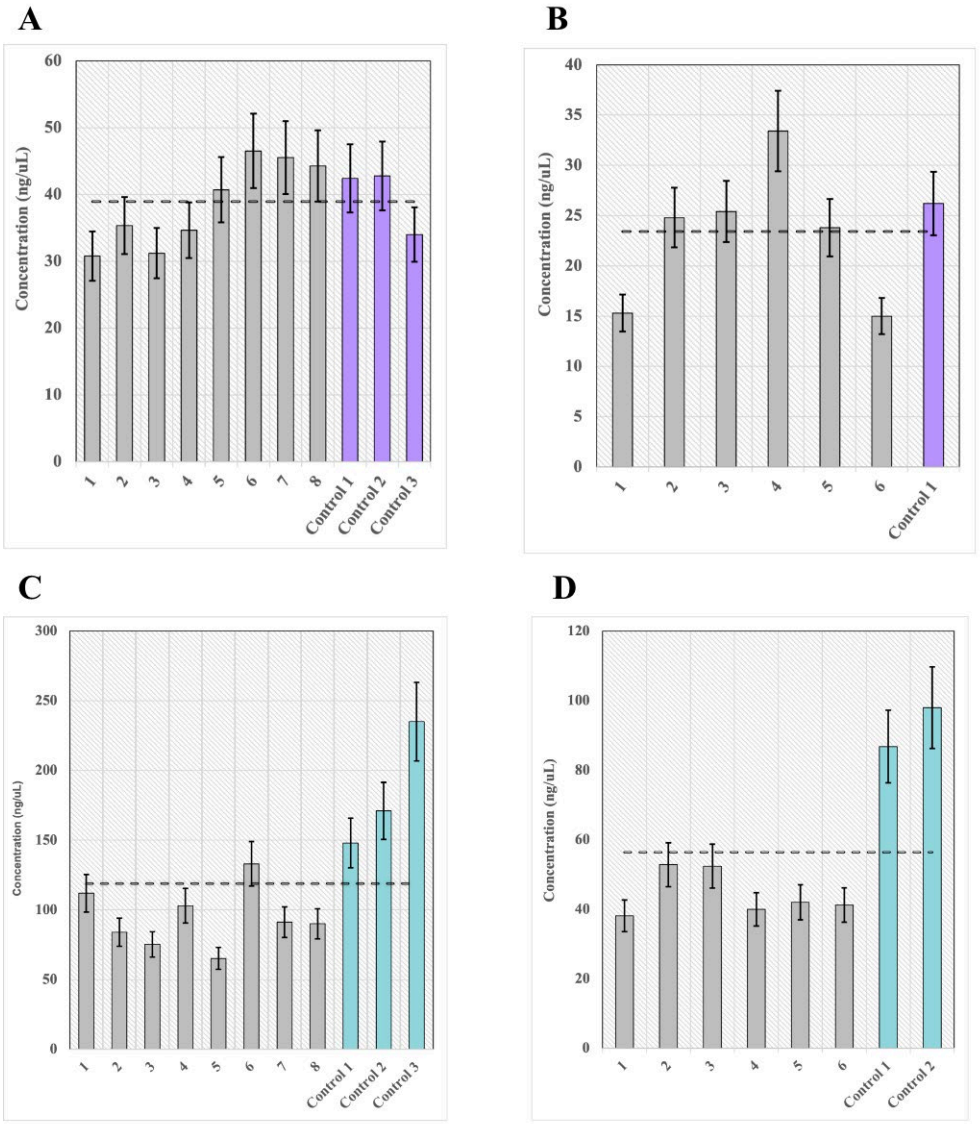
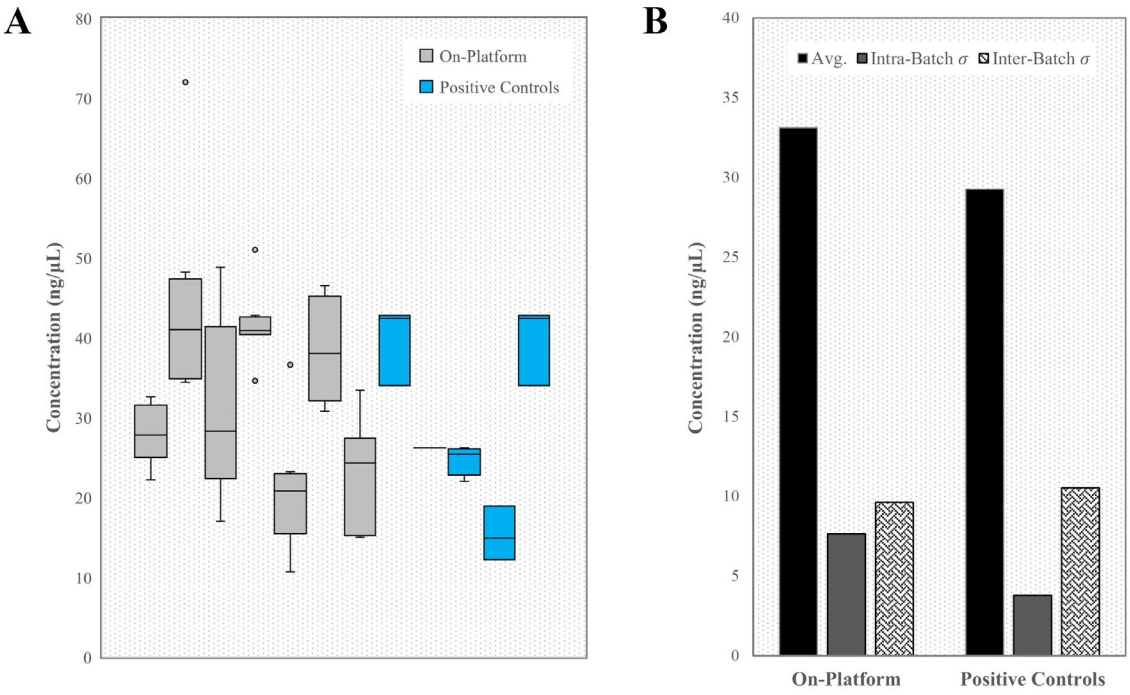


Figure 2.9 provides a visual summary of the final library concentrations for each of the initial four sample groups. On-platform libraries produced by the mechanical fragmentation assay averaged 38.6 ± 6.42 ng/µL and 23.0 ± 6.94 ng/µL for human and *E. coli* DNA, respectively, compared to 39.7 ± 4.97 ng/µL (human) and 26.2 ng/µL (*E. coli*) positive controls (Fig. 2.9 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%[39], and is a potential causative agent of the difference between measured and calculated concentrations. On-platform libraries produced by the enzymatic fragmentation assay averaged 94.2 ± 21.5 ng/µL and 44.4 ± 6.47 ng/µL for human and *E. coli* DNA, respectively, compared to 185 ± 45.1 ng/µL (human) and 92.4 ± 7.92 ng/µL (*E. coli*) positive controls (Fig. 2.9 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 (± 45.1 ng/µL vs. ± 21.5 ng/µL) for the enzymatic fragmentation assay.

Figure 2.10: Final library concentrations across multiple sample preparation sessions for mechanical fragmentation assay with assorted DNA input. (A) Automated plates compared to several batches of manually prepared positive controls. Each box represents a separate plate (for on-platform samples) or separate batch of manual samples (for positive controls). (B) Average library concentration across 54 on-platform samples and 14 manual positive controls. Sample-to-sample yield standard deviation, as well as standard deviation between average batch performances were determined to examine consistency between samples and between platform runs.



To further examine on-platform performance and consistency, we gathered final library concentrations from 7 different on-platform mechanical fragmentation assay runs conducted with mixed human and *E. coli* DNA, as well as 4 different batches of manual positive controls (Fig. 2.10A). All samples prepared passed sequencing criteria. On average, on-platform samples had a final concentration of 33.1 ng/µL. Average standard deviation across samples from the same plate was 7.64 ng/µL, while standard deviation across average plate concentrations was 9.61 ng/µL (Fig. 2.10B). Manual positive controls averaged 29.3 ng/µL, with an average intra-batch standard deviation of 3.80 ng/µL and a batch-to-batch standard deviation of 10.5 ng/µL (Fig. 2.10B).

Final library concentration is consistently comparable to manually prepared libraries, with similar batch-to-batch variation. Consistency in final library concentration is an indication that PCR amplification occurs reliably on the platform, achieving results comparable to a standard thermal cycler. The variability between 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. 2). Any differences in final concentrations observed 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[52]. 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.

VI. 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. As part of our sequencing data analysis, Phred was employed as a base calling program that reads sequence chromatogram files then assigns quality scores to each base call[41]. Quality scores are assigned on a logarithmic scale, where $Q = -log_{10}P$ (where P is the probability of a base being incorrectly called)[53]. Base calls with Q<20 (99% base call precision) are considered low quality and a standard indication of poor performance of the sequencing instrument or poor sample quality[53]. 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). Note that Q=30 corresponds to 99.9% base call precision[41]. Data quality was assessed as a measure of data yield and number of reads vs. DNA insert size (Fig. 2.11 A-B, D-E). Total sequencing data yield generated ~1,175 $\pm$ 0.25 megabases (Mb)

for the human group, averaging ~104 ± 0.25 Mb per sample (Fig. 11A) and ~755 ± 0.25 Mb for the *E. coli* group, averaging ~103 ± 0.25 Mb per sample (Fig. 11D). Positive controls averaged ~115 ± 0.25 Mb for the human group and ~135 ± 0.25 Mb for the *E. coli* group. Figure 2.11 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. 2.11B,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. 2.11C,F). However, both on-platform samples and positive controls have acceptable distributions, eliminating our platform as a perpetuator of sequencing bias. Note that GC bias is, in part, affected by size selection, where slight differences between manual and automated pipetting during bead-based cleanup steps can impact a sample's bias distribution. Still, no observed variation in GC bias between on-platform and manually prepared samples fell outside of passing criteria or became cause for concern.

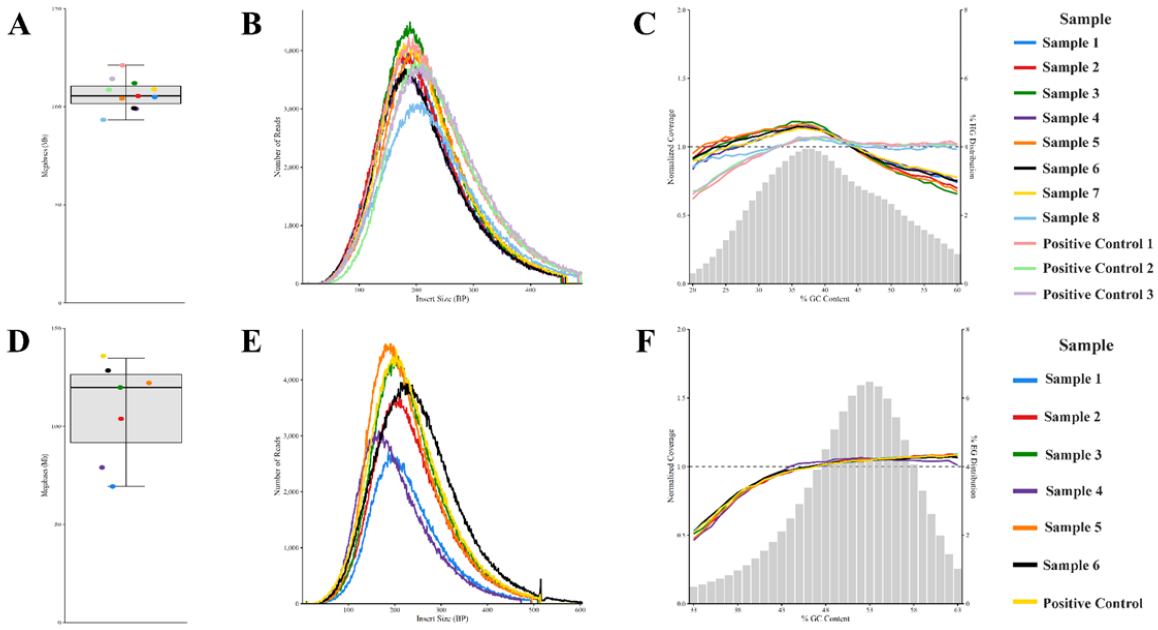
Mechanical fragmentation (sonication in this case), ERA, and ligation reactions are possible sources of bias present in both on-platform and positive control libraries. The grey bars represent the % GC content distributed across either the human or *E. coli* genome. The bar chart is centered around ~38 % GC content for human (Fig. 2.11C) and ~ 53 % GC

content for *E. coli* (Fig. 2.11F) DNA, representing the different amounts of GC in each genome (40.9% for human[54] and 50% for *E. coli*[55]). 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 %

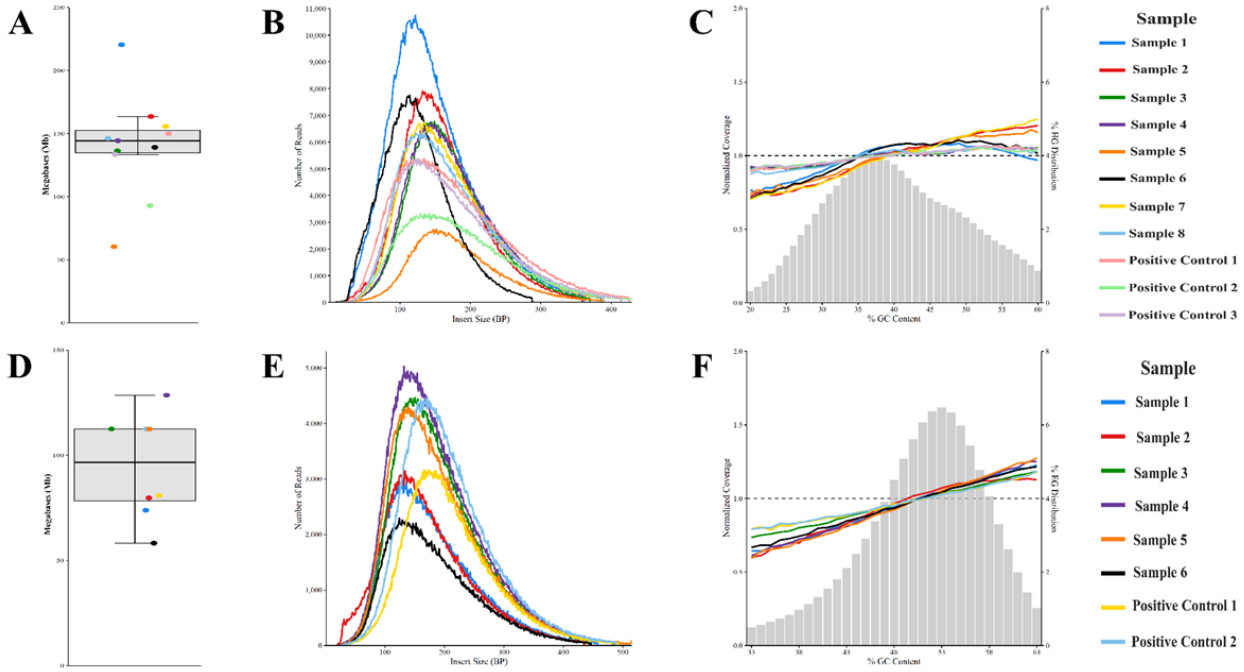
Figure 2.11: Sequencing data analysis for the mechanical fragmentation assay. Data yield per sample in Mb for (A) human DNA input and (D) E. coli DNA input. Number of reads vs. insert size for (B) human DNA input and (E) E. coli DNA input. Sequencing bias analysis in terms of normalized coverage vs. % 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. 2.12A), and ~761 ± 0.25 Mb for the *E. coli* group, averaging ~95 ± 0.25 Mb per sample (Fig. 2.12D). 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. Average read length was between 100-200 bp for both groups (Fig. 2.12 B,E), with on-platform libraries at ~11,000 reads for human (Fig. 2.12B) and ~5,000 reads for *E. coli* (Fig. 2.12E), 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. 2.12 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[56]. The grey bar chart is centered around ~38 % GC content for human (Fig. 2.12C) and ~ 53 % GC content for *E. coli* (Fig. 2.12F) DNA, representing the different amounts of GC in each genome, and aligns with known %GC content for the human and *E. coli* genomes.

Figure 2.12: 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[57, 58].

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. Select work in the field, such as performed by Shi, C. et al., has shown promising results for clinical applications of low-throughput NGS library preparation automation solutions[58]. Their cassette-dependent system, though different in design from our unified cartridge-plate system, was used to detect BRCA1 and BRCA2 mutations with high accuracy, high precision, and minimal cross-contamination, demonstrating the utility of platforms of this format in a

clinical research setting. As part of future studies, we look forward to testing our device with clinical samples and further examining its ability to prevent cross-contamination.

2.5 ACKNOWLEDGEMENTS

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

Competing Interests:

The authors declare no conflicts of interest.

AT is a paid scientific advisor/consultant for Revvity.

Data Availability:

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

2.6 EXPERIMENTAL PROCEDURES AND MATERIALS

I. 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 (isopropyl alcohol) 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 (2 μL each) were combined with 8 μL of water when placed in their respective wells to prevent evaporative loss and liquid handling errors. We consider a total volume of 10 μL in wells of the plate to be a large enough volume for

consistent plate preparation, while not exceeding the volume allowed at the heating element for PCR. Positive controls were performed with added water to ensure there were no negative results from ligation and PCR reaction dilution. 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). Note that room temperature and humidity are slightly variable from run-to-run of the platform, however we have not observed any negative effects from these variations.

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°C, then 30 minutes at 65°C in a thermal cycler. Adapters and ligation master mix were added to the fragmentation mixture. The mixture was incubated for 15 minutes at 20°C, 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°C, denaturation for 15 seconds at 98°C, annealing for 30 seconds at 60°C, and extension for 30 seconds at 72°C, with a final extension for 1 minute at 72°C. 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 to bring total well volume to 10 µL when placed in the reagent plate. No water was added to primer wells, as that volume was already 10 µL. 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).

II. 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. Human and *E. coli* genomic DNA was fragmented using the Covaris E210 Focused-ultrasonicator (Covaris, Woburn, MA).

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.

III. 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.

IV. 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 (Model 81320, Hamilton Company, Reno, NV) connected to a horizontal linear rail (SLW-BB-1040-06-100, igus Inc., Cologne, Germany). Airlines connect to a single-use custom cartridge (3D printed) comprised of flexible polymeric tubing (Flexelene brand, Eldon James, CO, USA) and joined to 16 cannula pipette tips (Revvity, Waltham, MA) 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 (igus Inc., Cologne, Germany) for movement along the z-axis. The reagent plate is inserted onto a horizontal aluminum frame (machined at Brown University lab), and attached to a linear rail (igus Inc., Cologne, Germany) for movement along the x-axis. A bar magnet (BZ084, K&J Magnetics, Pipersville, PA) moves along the z-axis to press against the reagent plate (269390, ThermoFisher Scientific, Waltham, MA) 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 (custom design, machined at Brown machine shop), held in place with a detachable insulated door (custom design, machined at Brown machine shop. A thermocouple (Digi-

Key Electronics, Thief River Falls, MN) epoxied within a segment of cartridge tubing and fixed to the heating element 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 (custom design, machined at Brown machine shop). The valves press firmly against the metal PCR door to seal samples inside the polymeric tubing and prevent vapor escape during heating processes. Proper maintenance of pressure, as demonstrated by low leak rates found in Supplementary Material (S. Fig. 3).

V. 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 16 10 µL pipette tips press-fit without dead volume into a custom cannula boat that fits firmly into the spring-loaded z-axis clamp. Eight separate lanes of flexible polymeric tubing interface with the pipette tips to directly connect them 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. Fluid volumes move through the cannula and tubing via Hagen-Poiseuille flow. Consistent cartridge manufacturing is key to the functional operation of the platform. Consequently, each cartridge is subjected to a strict quality control test, where pipette tip alignment, vertical positioning, and fluid aspiration and dispensation are evaluated in a sterile environment.

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 (IPA) used for the wash steps, then sealed with a 384-well pierceable plate seal (Azenta Life Sciences, Wotton, England).

VI. 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 cannulae, 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).

VII. Sequencing and Data Analysis

All samples were sequenced using the MiSeq System (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

Supplementary Table 1: Variables and values corresponding to Fig. 4. Mixing and aspiration velocities for each step can also be found in S. Table 1 as additional information for Fig. 4.

S. Table 1A

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
v_{mix}	16.7 µL/s
v_{asp}	4.32 µL/s

S. Table 1B

	Mechanical Fragmentation	**Enzymatic Fragmentation**
C_i	[sheared DNA] = 0.4 ng/µL [ERA buffer] [ERA enzyme]	[gDNA] = 1 ng/µL [Frag/AT buffer] [Frag/AT enzymes]
T_f	N/A	37°C

t_2	N/A	20 min
T_{era}	65℃	65℃
t_3	35 min	30 min
P_1	27.89 psi	27.89 psi
V_2	50 µL	50 µL
C_a	[sheared DNA] = 0.4 ng/µL [adapter] = 0.012 µM [ligase enzyme] [ligase buffer] [ERA buffer] [ERA enzyme]	[fragmented DNA] = 1 ng/µL [adapter] [ligation master mix] [Frag/AT buffer] [Frag/AT enzymes]
V_3	107.5 µL	80 µL
v_{mix}	16.7 µL/s	16.7 µL/s
v_{asp}	1.67 µL/s	1.67 µL/s

S. Table 1C

	Mechanical Fragmentation	**Enzymatic Fragmentation**
C_a	[sheared DNA] = 0.4 ng/µL [adapter] = 0.012 µM [ligase enzyme] [ligase buffer] [ERA buffer] [ERA enzyme]	[fragmented DNA] = 1 ng/µL [adapter] [ligation master mix] [Frag/AT buffer] [Frag/AT enzymes]

T_L	Room Temperature	Room Temperature
t_4	20 min	20 min
V_3	107.5 µL	80 µL
C_b	[adapter-ligated DNA] [non-ligated adapter] [ligase enzyme] [ligase buffer] [ERA buffer] [ERA enzyme]	[adapter-ligated DNA] [non-ligated adapter] [ligation master mix] [Frag/AT buffer] [Frag/AT enzymes]
v_{mix}	16.7 µL/s	16.7 µL/s
v_{asp}	1.67 µL/s	1.67 µL/s

S. Table 1D

	Mechanical Fragmentation	**Enzymatic Fragmentation**
V_3	107.5 µL	80 µL
C_b	[adapter-ligated DNA] [non-ligated adapter] [ligase enzyme] [ligase buffer] [ERA buffer] [ERA enzyme]	[adapter-ligated DNA] [non-ligated adapter] [ligation master mix] [Frag/AT buffer] [Frag/AT enzymes]
V_4	100 µL	60 µL
C_c	[Cleanup beads]	[Purification beads]

	[adapter-ligated DNA] [non-ligated adapter] [ligase enzyme] [ligase buffer] [ERA buffer] [ERA enzyme]	[adapter-ligated DNA] [non-ligated adapter] [ligation master mix] [Frag/AT buffer] [Frag/AT enzymes]
V_5	~205 µL	~138 µL
V_6	198 µL	200 µL
V_7	~196 µL	~198 µL
t_5	~3 min	~3 min
V_8	28 µL	23 µL
C_d	[Cleanup beads] [adapter-ligated DNA]	[Purification beads] [adapter-ligated DNA]
C_e	[adapter-ligated DNA]	[adapter-ligated DNA]
V_9	27 µL	22 µL
v_{mix}	Alternating between 16.7 µL/s and 76.8 µL/s	Alternating between 16.7 µL/s and 76.8 µL/s
$v_{asp,\ supernatant}$	3.34 µL/s	3.34 µL/s
$v_{disp,\ IPA}$	20.0 µL/s	20.0 µL/s

S. Table 1E

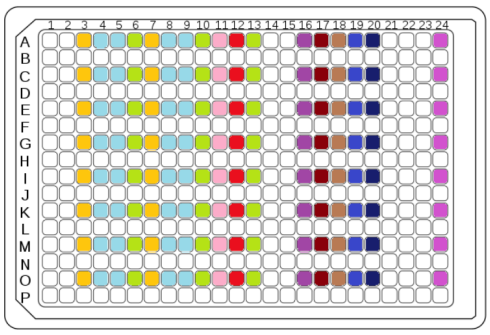
	Mechanical Fragmentation	**Enzymatic Fragmentation**
C_f	[adapter-ligated DNA]	[adapter-ligated DNA]
V_9	27 µL	22 µL
V_{10}	62 µL	57 µL
C_g	[adapter-ligated DNA] [PCR primers] [PCR master mix]	[adapter-ligated DNA] [PCR primers] [PCR master mix]
T_D	98°C	98°C
T_A	65°C	60°C
T_E	72°C	72°C
t_{cycle}	Initial 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
C_h	[amplified DNA] [PCR primers] [PCR master mix]	[amplified DNA] [PCR primers] [PCR master mix]
v_{mix}	16.7 µL/s	16.7 µL/s
v_{asp}	1.67 µL/s	1.67 µL/s

S. Table 1F

	Mechanical Fragmentation	**Enzymatic Fragmentation**
V_{10}	62 µL	57 µL
C_h	[amplified DNA] [PCR primers] [PCR master mix]	[amplified DNA] [PCR primers] [PCR master mix]
V_{11}	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_{12}	~105 µL	~105 µL
V_{13}	198 µL	200 µL
V_{14}	~196 µL	~198 µL
t_5	~3 min	~3 min
V_{15}	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_{16}	33 µL	25 µL
C_e	[adapter-ligated DNA]	[adapter-ligated DNA]
V_9	27 µL	22 µL

v_{mix}	Alternating between 16.7 µL/s and 76.8 µL/s	Alternating between 16.7 µL/s and 76.8 µL/s
$v_{asp,\ supernatant}$	3.34 µL/s	3.34 µL/s
$v_{disp,\ IPA}$	20.0 µL/s	20.0 µL/s

Supplementary Figure 1: Example 384-well plate layout for library preparation, including all pre-loaded reagents.

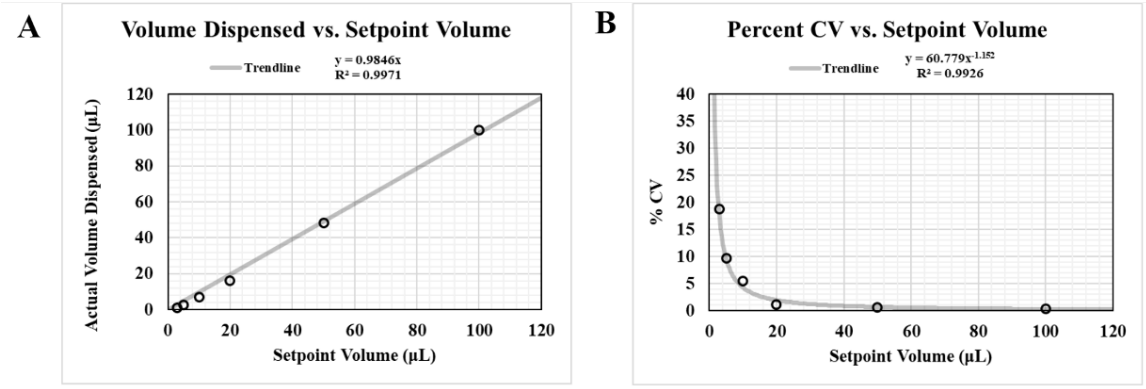


Column 3 - Resuspension buffer
Column 4 - 70% IPA
Column 5 - 70% IPA
Column 6 - 0.5% Tween-20
Column 7 - Resuspension buffer
Column 8 - 70% IPA
Column 9 - 70% IPA
Column 10 - 0.5% Tween-20
Column 11 - Adapter mix
Column 12 - Ligation master mix
Column 13 - 0.5% Tween-20
Column 16 - Pretreatment solution
Column 17 - Cleanup beads
Column 18 - Cleanup beads
Column 19 - Twist UDI primer
Column 20 - Equinox Library Amp Mix (2X)
Column 24 - Fragmentation mix

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).

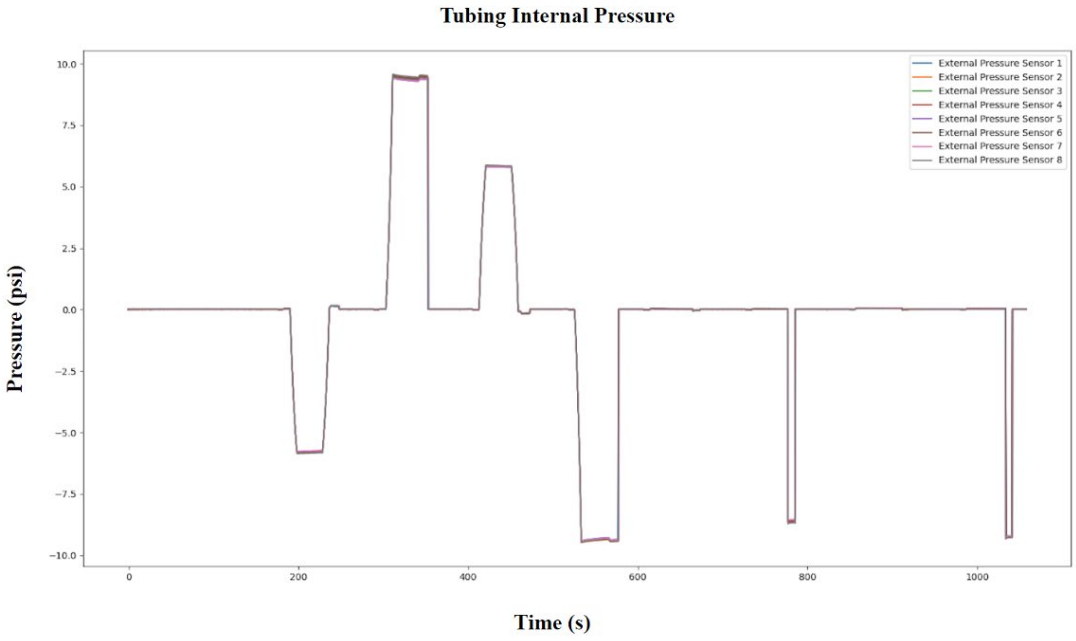
Volume accuracy of liquid transfers improves with increasing volume, while % CV decreases (Fig 5). The platform has an approximate working range of 10 to 200 µL, which allows it to complete liquid transfers and perform proper mixing of plated reagents. For this volumetric range, % CV remains low, with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

Supplementary Figure 2: Volumetric accuracy of the pump assembly. (A) Actual volume dispensed vs. scripted setpoint volumes, and (B) Percent CV vs. scripted setpoint volumes.



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. For a fully functioning platform, leak rates are expected to be insignificant, often no more than 1 psi of pressure lost per hour. 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).

Supplementary Figure 3: Plot of pressure readings over time during pump pressure evaluation.



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 $P_1V_1 = P_2V_2$. 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.

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

Component	**Volume (mm^3)**
Tubing from fluid slug to Y-junction	132.61
Tubing from -junction to cartridge frame	138.55
Interfaces between cartridge and airline	40.43
Airline	296.90

Interfaces between airline and syringes	159.26
Syringe barrel (post-pressurization in parentheses)	846.61 (83.00)
Total (post-pressurization in parentheses)	1614.36 (850.36)

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.

$$y = \frac{V_{RsB} \sum_{i=1}^{n} C}{l_{max}}$$

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.

$$\% \, d = \frac{d}{y + d}$$

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).

Supplementary 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	965.25	0.92	0.10%

Control 2			
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	Mass	Mass Dimer	% Dimer	Dilution

gDNA	Library/Yield (ng)	(ng)		
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	2457.9	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	1085.6	2.52	0.21%	-
Undiluted Std.	274.70	6.05	0.0049%	-

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).

Supplementary Table 4: Final library concentrations.

A: Mechanical Fragmentation

Mechanical Fragmentation			
Human sheared DNA		E.coli sheared DNA	
Sample	Concentration (ng/µL)	Sample	Concentration (ng/µL)
1	30.7	1	15.3
2	35.3	2	24.8
3	31.2	3	25.4
4	34.6	4	33.4
5	40.7	5	23.8
6	46.5	6	15.0
7	45.5	Positive Control	26.2
8	44.3	Average	23.4
Positive Control 1	42.4	Std.	6.46
Positive Control 2	42.8		
Positive Control 3	34		
Average	38.9		
Std.	5.84		

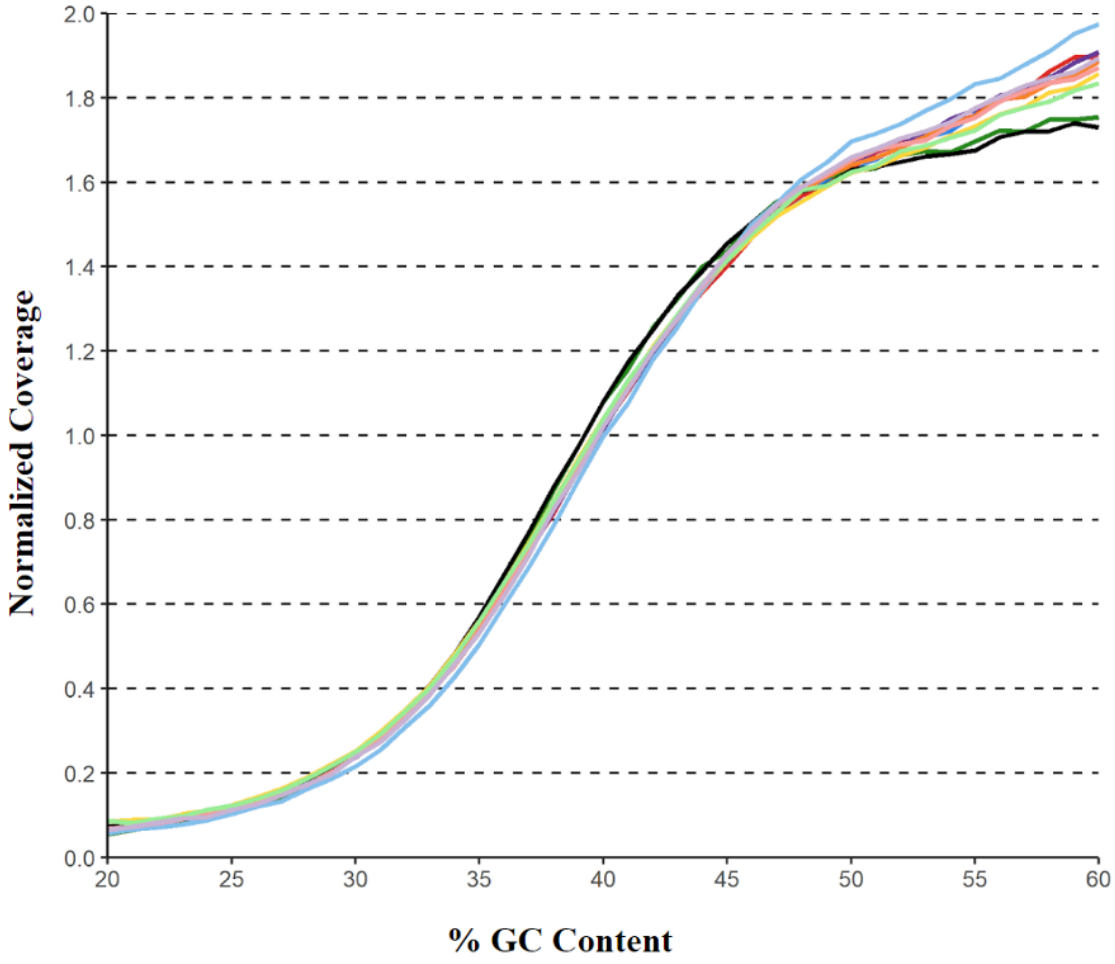
B: Enzymatic Fragmentation

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	-	56.4
Positive Control 2	1:10	17.1	Undiluted Std.	-	23.1
Positive Control 3	1:10	23.5			
Average Undiluted	-	112			

Library		
Undiluted Std.	-	50.1

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. 4. 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.

Supplementary Figure 4: 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.

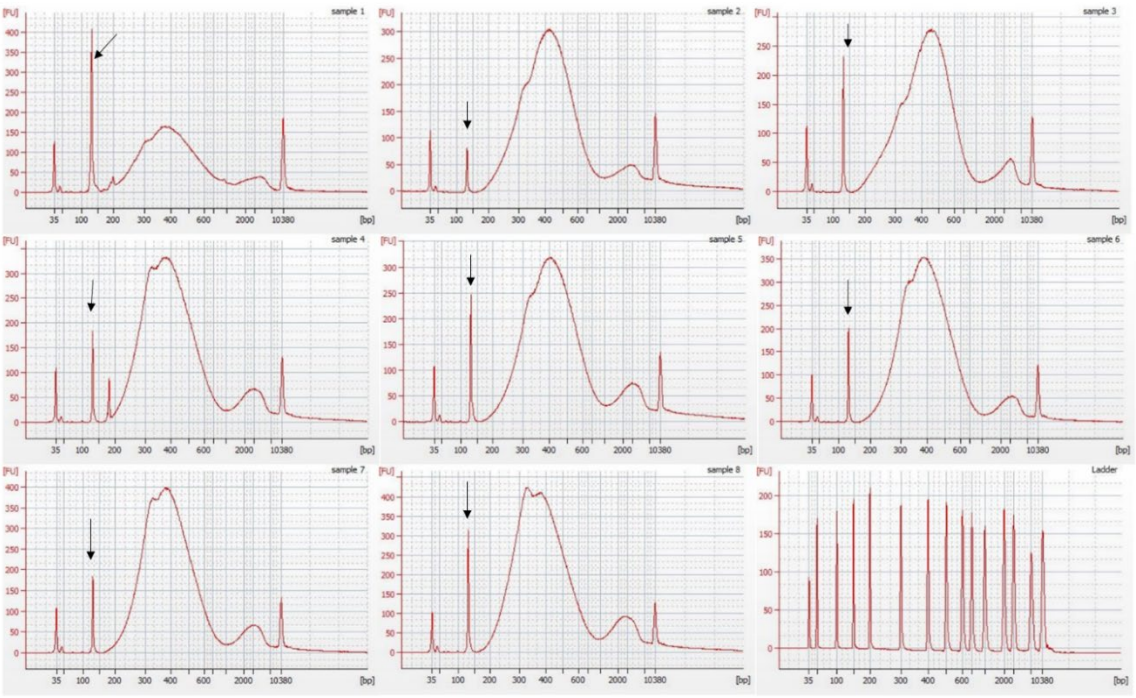


Supplementary Table 5: DNA sonication parameters that produced significant pro-GC bias. DNA was sonicated using a Covaris S220 Focused-ultrasonicator (Covaris, Woburn, MA).

Parameter	**Value**
Peak Incident Power	210
Duty Factor	10%

Cycles per burst	200 count
Duration	150 sec
Water level	12
Temperature	7°C

Supplementary Figure 5: Electropherograms of samples prepared on-platform without pretreatment. Note high adapter-dimer peaks (shown by arrows) at roughly 130bp, related to insufficient surface passivation.



Chapter 3

Accelerating Discoveries: A Comprehensive RNA-Based Next Generation Sequencing Library Automation Approach

---

Accelerating Discoveries: A Comprehensive RNA-Based Next Generation Sequencing Library Automation Approach

Jennifer Figueroa-Cruz*, Kathryn Whitehead*, Khaliun Myagmar, Khulan Unurbuyan, Sabrina Tolppi, Anubhav Tripathi**, Under final verification to be submitted to a journal

Center for Biomedical Engineering, School of Engineering, Brown University

[*These authors contributed equally.]

[**To whom correspondence should be addressed to. Email: anubhav_tripathi@brown.edu]

3.1 ABSTRACT

With the rise of RNA-based therapeutics, automating methods that validate them is of great importance. RNA-based Next Generation Sequencing (NGS) library preparation assays are an example of these methods. Current RNA NGS methods offer the opportunity to analyze previously extracted Total RNA; Ribosomal RNA (rRNA); Poly (A) RNA; and Formalin-Fixed, Paraffin-Embedded (FFPE) RNA. These assays could be divided into two (2) main parts. The first part is a cDNA Synthesis workflow. This part consists of the fragmentation of the extracted RNA, followed by First Strand Synthesis, and Second Strand Synthesis which ends with a cDNA product. The second part is the Library Preparation Workflow of

the obtained cDNA. This part consecutively does the samples' Adenylation, Adapter Ligation, and Primer Amplification. After this last step, the samples are ready to be sequenced and analyzed for their intended application. We have successfully automated two common Total RNA-based library preparation assays. These protocols are the Nextflex Rapid Directional RNA-Seq Kit 2.0 and Watchmaker RNA Library Prep Kit. The protocols were designed for the user to fully walk away from the instrument and would only need to interact with the process when loading the plate into the automated platform. Lastly, our final libraries are comparable to the manual positive controls of the protocol. This was determined by doing quality control of the samples and analyzing their sequencing results.

3.2 INTRODUCTION

Next Generation Sequencing (NGS) has revolutionized genomics by enabling rapid and cost effective sequencing by using massively parallel short reading. The improved platforms and decreasing costs of NGS enables clinicians and researchers to delve deeper into the genetic underpinnings of diseases, identify mutations, and tailor precise treatments. The four main steps in the NGS process are sample extraction, library preparation, sequencing, and bioinformatic analysis. Of these steps, library preparation is the most time and work intensive process. Next Generation Sequencing (NGS) library preparation is a novel research process used to prepare DNA or RNA for sequencing. This involves the ligation of adapters to the DNA or RNA sequence that will allow samples to bind to a flow cell during sequencing and facilitate post sequencing identification of samples[59].

With the advent of NGS technology, RNA sequencing has gained prominence and has quickly become an indispensable tool in both research and clinical diagnostics, offering insights into gene expression dynamics, transcriptomics, cellular regulation, and tumorigenesis[60]. In recent years, researchers have demonstrated different RNA-based techniques to diagnose and treat pathogens and diseases such as Alzheimer's Diseases (AD) and different types of cancer [61,62]. RNA-based procedures have also been proven to be useful for detecting chemical contamination in water, which is one of the main concerns for public health[63]. However, because handling RNA presents challenges related to the instability and complexity of RNA molecules, as well as the need for specialized preservation procedures, library preparation for sequencing RNA samples remains a bottleneck in research and clinical diagnostics [61,64]. Therefore, this study describes

automation of two different Total RNA-based NGS library preparation assays to help advance discoveries in the field.

Total RNA is composed of Ribosomal RNA (rRNA) Messenger RNA (mRNA), Transfer RNA (tRNA), and a variety of non-coding RNAs[65]. Because total RNA represents all of the types of RNA, it allows for a comprehensive analysis of the entire transcriptomes, including non coding regions[66]. This allows for a more in depth analysis and understanding of regulatory regions, global expression levels of transcripts, splicing patterns, and to identify exons, introns, and their boundaries.

The standard library preparation methods for Total RNA involve RNA extraction and fragmentation, complementary DNA (cDNA) synthesis, adapter ligation, pcr amplification, and purification[67]. cDNA synthesis involves the creation of DNA from the RNA template via reverse transcription. This occurs in two major steps, the first being first strand synthesis. First strand synthesis is facilitated by the annealing of a primer to the RNA template. Once the primer is bound, a reverse transcriptase enzyme can synthesize a cDNA strand along the template, adding deoxyribonucleotide triphosphates (dNTPs) to the cDNA chain. After the RNA molecule has been transcribed, seconds strand synthesis can occur. This involves the degradation of the original RNA templates and synthesis of a more stable second cDNA strand using a polymerase enzyme (resulting in double stranded cDNA) (Fig. 3.1). Following cDNA synthesis, adapter ligation, pcr amplification, and cDNA purification occur resulting in sequencing ready samples (Fig. 3.2).

Figure 3.1: Visual representation of a standard RNA-Seq Next Generation Sequencing (NGS) complementary DNA (cDNA) synthesis workflow.

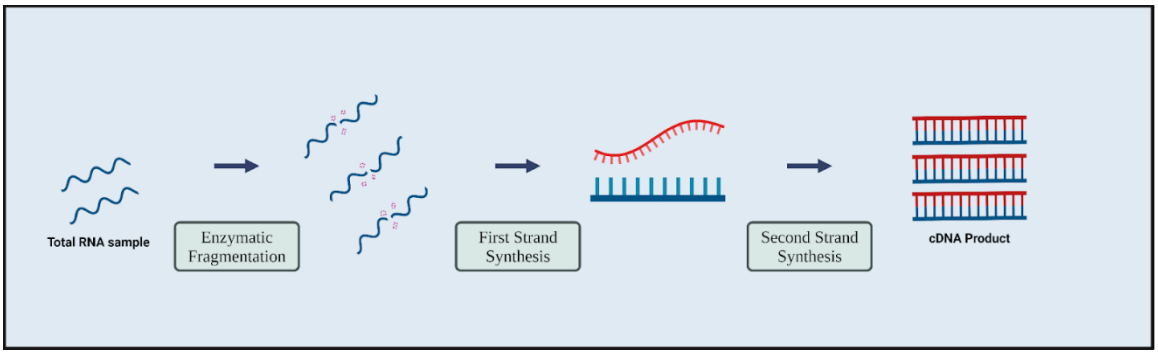
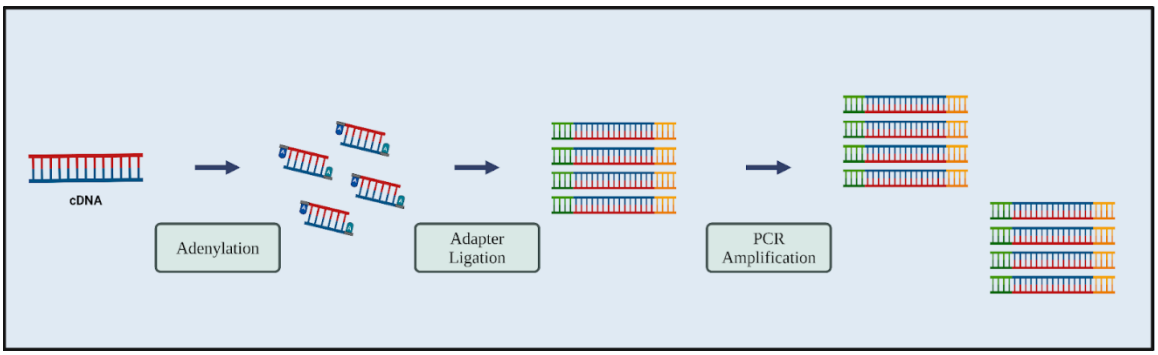


Figure 3.2: Visual representation of a standard RNA-Seq Next Generation Sequencing (NGS) library preparation workflow.



To develop and test automated workflows for Total RNA sequencing, two common library preparation kits were chosen, Nextflex Rapid Directional RNA-Seq Kit 2.0 and Watchmaker RNA Library Prep Kit. Using an automated device with liquid handling, thermocycling, and magnetic purification capabilities, workflows were developed and optimized to prevent RNA degradation, enhance sample purity, and eliminate sequencing bias.

3.3 RESULTS AND DISCUSSION

I. Automated Workflow

Our automated workflows did not differ step-wise from the manufacturer's manual protocols (Fig. 3.3 & Fig. 3.4). While several of the protocol parameters were adjusted to accommodate automation, the overall procedure remained the same. The main challenges that we encountered while automating both assays were RNA degradation and minimal sample loss. Because of this, we had to increase the minimal sample input concentration for both assays for the final libraries to pass the quality criterias. For the NextFlex Rapid Directional RNA-Seq Kit 2.0, the minimal sample input concentration was determined to be 200 ng. On the other hand, for the Watchmaker RNA Library Prep Kit, minimal sample input concentration was determined to be 100 ng. Both of the automated workflows generate 8 sequenceable libraries in parallel where user intervention is only 30 minutes for the NextFlex Rapid Directional RNA-Seq Kit 2.0 and 15 minutes for the Watchmaker RNA Library Prep Kit. This intervention is performed at the beginning of the run when the user loads the reagents into the plate and then loads it into the automated device.

Figure 3.3: Visual representation of the NextFlex Rapid Directional RNA-Seq Kit 2.0 Workflow.

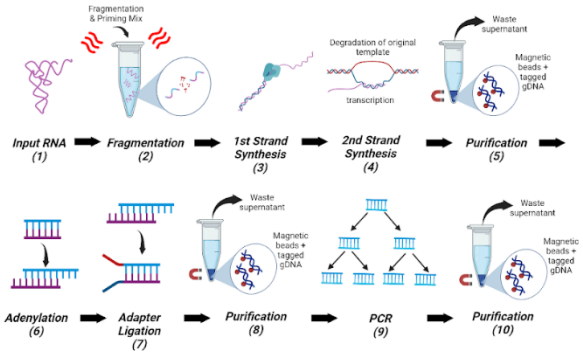
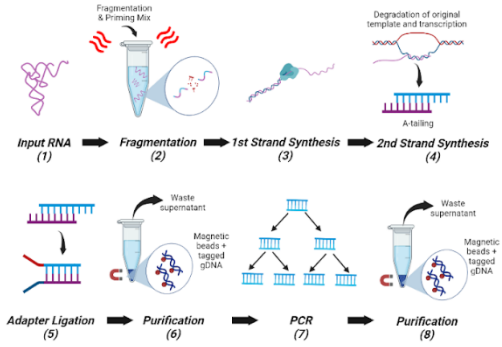


Figure 3.4: Visual representation of the Watchmaker RNA Library Prep Kit Workflow.

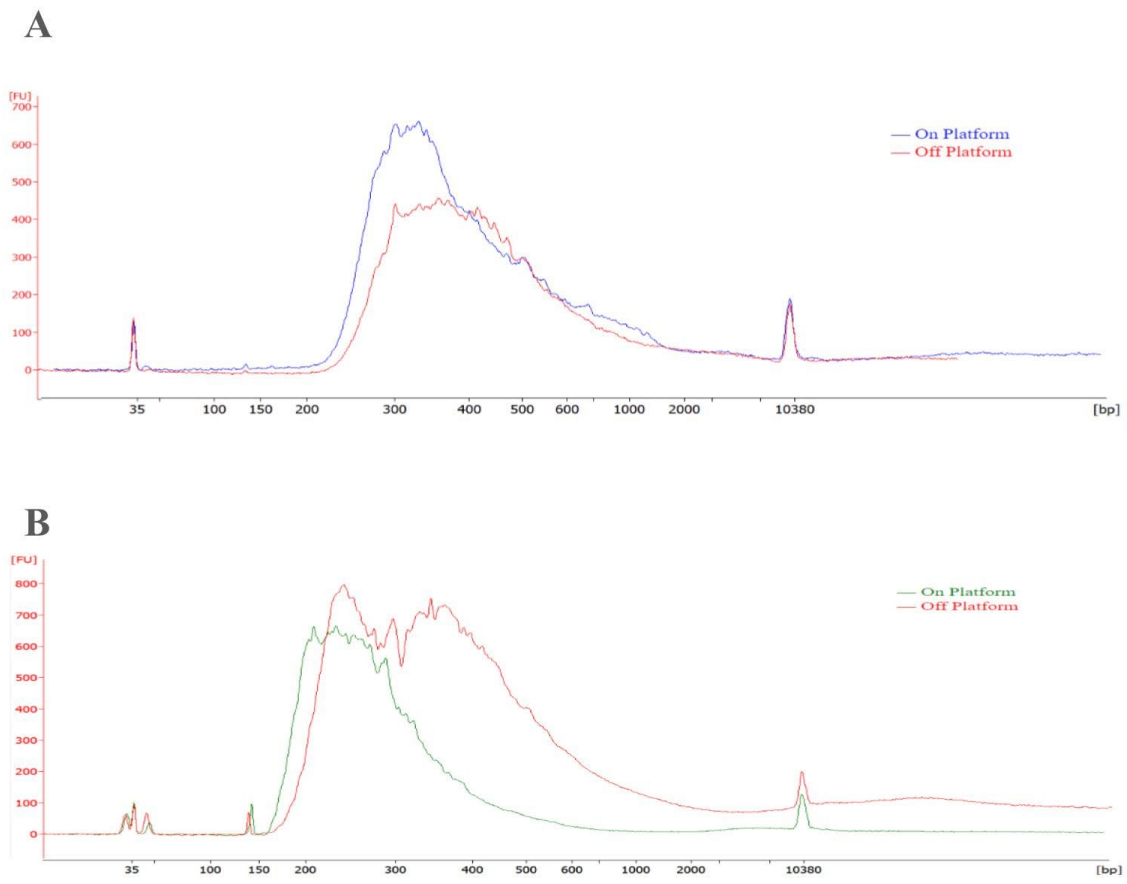


II. Quality Control

To perform quality control on the final libraries generated, the samples were verified by using the Agilent 2100 Bioanalyzer. On Figure 5 we show an overlay between a chosen on-platform sample and off-platform final library for a comparison of the quality between both types of samples. For NextFlex Rapid Directional RNA-Seq 2.0 Kit the on-platform samples (n=8) averaged around 336 bp while the off-platform samples (n=3) averaged

around 296 bp (Fig. 3.5A). Both sets of samples showed a distribution of around 290 - 1,000 bp with a negligible difference from sample to sample. On the other hand, for the Watchmaker RNA Library Prep Kit the on-platform samples (n=8) averaged around 209 bp while the off-platform samples (n=2) averaged 513 bp (Fig. 3.5B). The on-platform samples showed a distribution of around 200 - 550 bp and the distribution of the off-platform samples was around 210 - 525 bp. This slight difference in distribution can be attributed to the different aspects found between manual pipetting and the platform pipetting.

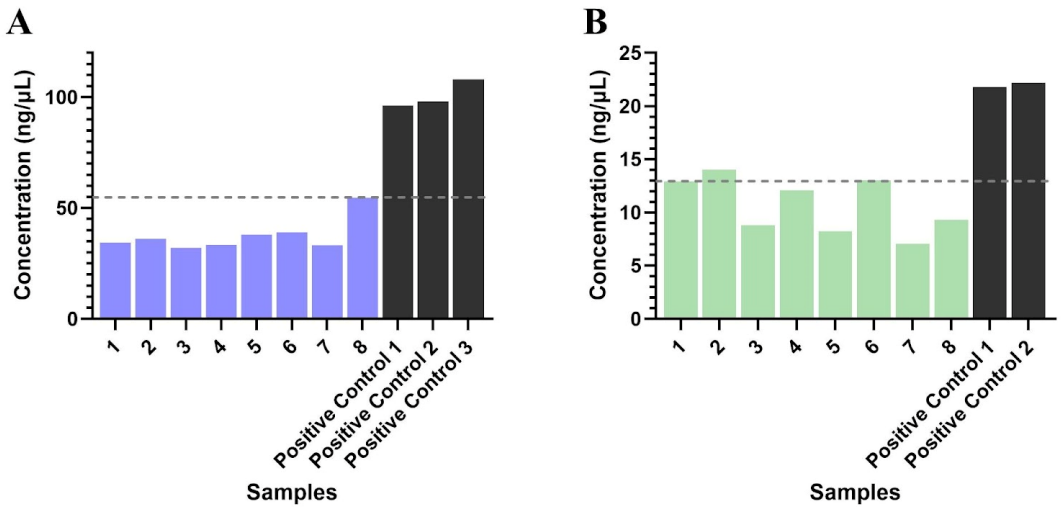
Figure 3.5: Final library quality control. (A) Overlay of on-platform sample (blue) and off-platform sample (red) for NextFlex Rapid Directional RNA-Seq 2.0 Kit (B) Overlay of on-platform sample (green) and off-platform sample (red) for Watchmaker RNA Library Prep Kit.



On-platform and off-platform (positive controls) final library concentrations are represented in Figure 3.6. NextFlex Rapid Directional RNA-Seq Kit 2.0 final libraries averaged 37.58 ± 7.29 ng/µL and 100.73 ± 6.36 ng/µL for on-platform and off-platform respectively (Fig. 3.6A). The off-platform and on-platform final library concentrations averages vary by approximately 63.15 ng/µL. On the other hand, Watchmaker RNA Library Prep Kit final libraries averaged 10.68 ± 2.61 ng/µL and 22.00 ± 0.28 ng/µL for on-platform and off-platform samples respectively (Fig. 3.6B). The final library concentration averages for the on-deck and off-deck samples for this assay vary

approximately 11.32 ng/µL. It is to be expected for on-platform samples to have a lower concentration than off-platform which makes the trend observed normal. The Qubit Flex Fluorometer has an error of 12% reported by the manufacturer. This fact is to be taken into account when comparing the data. Even though the on-platform and off-platform samples vary from each other on both assays, they are still comparable and they fall within our quality metrics. This means that all of the libraries produced both by the automated platform and manually were sequenceable and ready to be analyzed.

Figure 3.6: On-platform and off-platform (positive controls) final library concentrations in ng/µL for (A) 200 ng input Human Total RNA produced by NextFlex Rapid Directional RNA-Seq Kit 2.0 and (B) 100 ng input Human Total RNA produced by Watchmaker RNA Library Prep Kit. Dashed lines show the average for the library preparation samples across on-platform and off-platform samples for each assay.



III.Sequencing Results and Analysis

Following quality control and quantification of the library, sequencing was performed using the Illumina MiniSeq System (Illumina, San Diego, CA). Full run sequencing results from the automated device for both kits were analyzed as well as off platform positive

controls (Watchmaker on platform n=8, NextFlex on platform n=8, Watchmaker positive control n=2, Nextflex positive control n =3). Paired-end reads were aligned to the human genome using GenCode hg38 reference and basic transcript annotations. Both Nextflex and Watchmaker automated protocols yielded properly paired alignment percentages (Fig. 3.7A, 3.7B) and total mapped percentages (Fig. 3.8A, 3.8B) comparable to the positive control. Average percentages for reads properly paired mapped for the automated device for Watchmaker and Nextflex kits were 92.03% and 94.13% respectively, while the average properly paired percentages for the positive controls for Watchmaker and Nextflex were 95.52% and 94.72% respectively. Average percentages for total reads mapped for the automated device for Watchmaker and Nextflex kits were 99.01% and 99.32% respectively, while the average properly paired percentages for the positive controls for Watchmaker and Nextflex were 99.39% and 99.39% respectively .

Figure 3.7: On-platform and off-platform (positive controls) sequenced final libraries properly paired percentage for (A) samples of Human Total RNA produced by NextFlex Rapid Directional RNA-Seq Kit 2.0 and (B) samples of Human Total RNA produced by Watchmaker RNA Library Prep Kit.

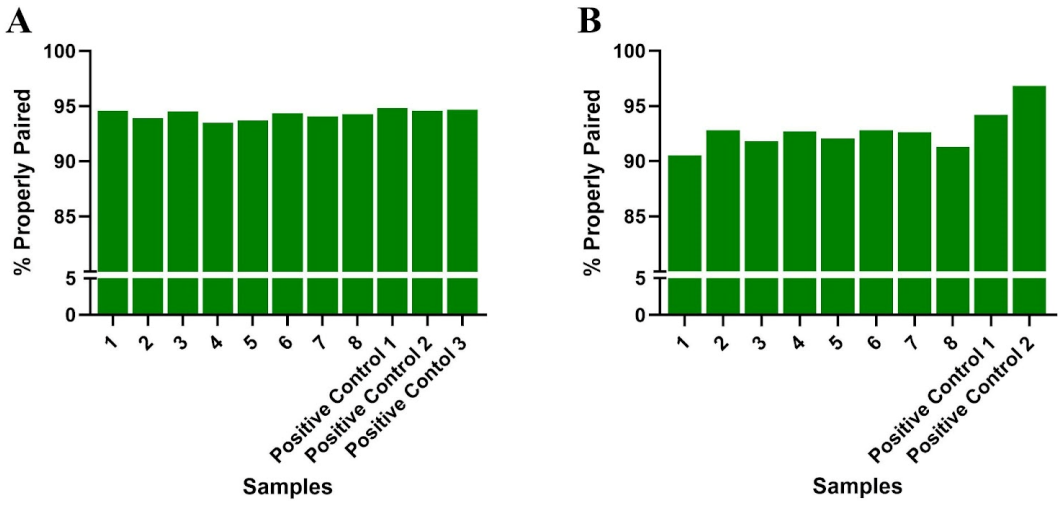
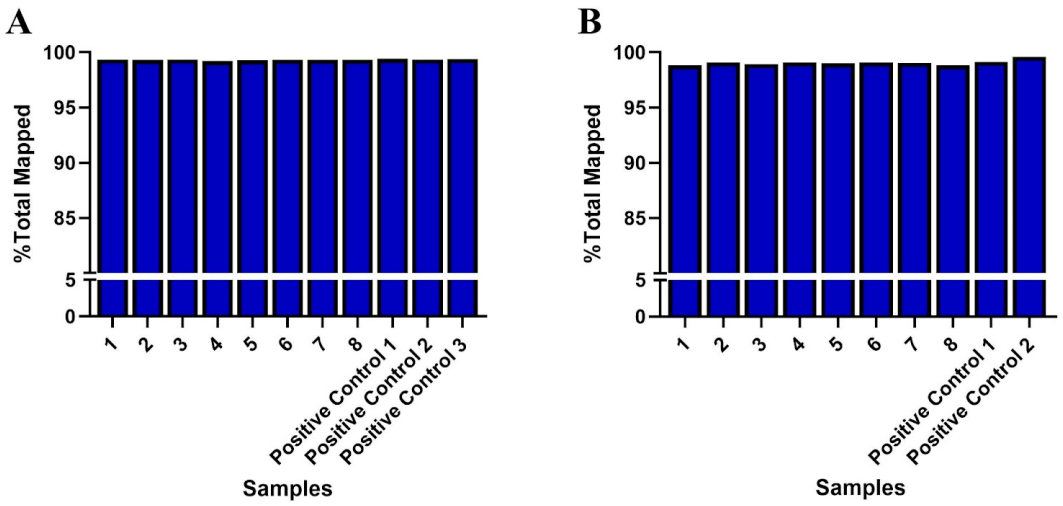
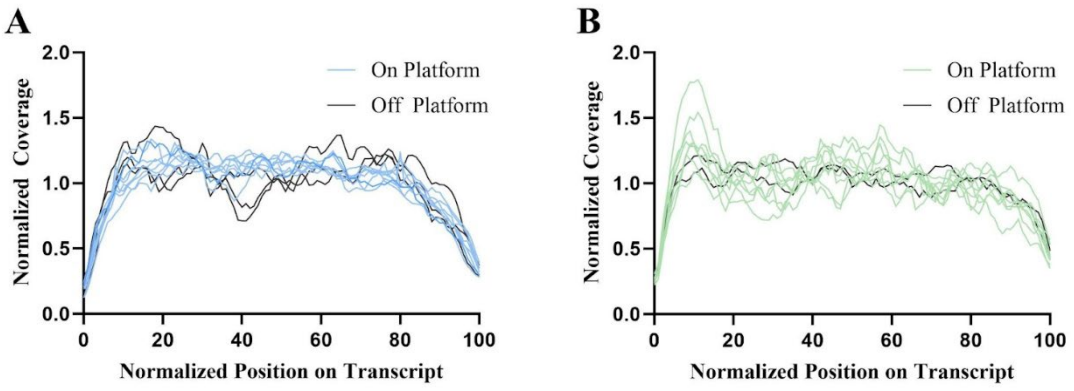


Figure 3.8: On-platform and off-platform (positive controls) sequenced final libraries total mapped percentages for (A) samples of Human Total RNA produced by NextFlex Rapid Directional RNA-Seq Kit 2.0 and (B) samples of Human Total RNA produced by Watchmaker RNA Library Prep Kit.



Many applications of RNA sequencing revolve around new and known variant identification for diagnostics and clinical research. Therefore, there is a need for consistent coverage across the transcript. Normalized coverage across normalized position on the transcript was compared between positive control and automated device for both protocols (Fig. 3.9A, 3.9B). Both automated protocols generated a library with coverage across the transcript comparable to, and in NextFlex more consistent than, the positive control. The coverage distribution and mapping metrics indicate that the library generated from the automated procedure is effectively sequenceable and allows for a broad range of applications typical of RNA sequencing.

Figure 3.9: On-platform and off-platform (positive controls) sequenced final libraries normalized coverage for (A) samples of Human Total RNA produced by NextFlex Rapid Directional RNA-Seq Kit 2.0 and (B) samples of Human Total RNA produced by Watchmaker RNA Library Prep Kit.



3.4 CONCLUSIONS

As the use of NGS for RNA sequencing becomes more rampant, generating automated workflows to decrease hands-on time and increase the efficiency of library generation is of great importance. However, because applications of RNA sequencing are primarily related to clinical diagnostics and research surrounding variant identification, these automated workflows must produce reproducible and sequenceable libraries. With this work we show different automated workflows for two commonly used RNA-Seq NGS kits. The results show that the automated workflows produce samples with comparable library quality and sequencing metrics to those generated off-platform for each of the kits. One of the common issues encountered in automated handling of RNA is the degradation and loss of samples. Therefore, several iterations of the automated workflows were performed to optimize sample preservation and maximum library output. The finalized workflows serve to standardize and streamline the library preparation process, reducing human error and variability in samples. This is supported by the consistently high mapping rates and comprehensive sample coverage across the transcript. For the NextFlex kit workflow, there is an improved uniformity of coverage, potentially indicating increased viability and

reliability of samples for downstream analysis such as variant calling and differential transcriptomics.

3.5 ACKNOWLEDGEMENTS

The authors thank Revitty for the grant given to Brown University to be able to carry out this research and development project.

Competing Interests:

The authors declare no conflict of interest with this project.

Anubhav Tripathi is a paid scientific advisor and consultant for Revvity.

3.6 METHODS AND MATERIALS

I.Library Preparation Methods

An automated workflow was developed and tested on two common Total RNA library preparation kits, Watchmaker RNA Library Preparation Kit (Watchmaker Genomics, Boulder, CO) and Nextflex Rapid Directional RNA-Seq Kit 2.0 (Revvity, Waltham, MA).

Nextflex Total RNA: Positive Controls and automated libraries were assembled with the Nextflex Rapid Directional RNA-Seq Kit 2.0 (Revvity, Waltham, MA) to do Total RNA Library Preparation. The manufacturer protocol was utilized to assemble the positive controls. Water, Total RNA, and fragmentation buffer were mixed together. The samples were incubated at 94 ℃ for 12 minutes. Directional First Strand Synthesis Buffer Mix and Rapid Reverse Transcriptase were added to the mix which was then incubated at 25 ℃ for 10 minutes, 50 ℃ for 15 minutes, ending with 70 ℃ for 10 minutes. After the incubation, the Directional Second Strand Synthesis mix was added, the mixture was incubated at 16

℃ for 60 minutes, and purified with Clean Up Beads and 70% IPA. The samples were resuspended with nuclease-free water and the resulting supernatant was transferred to a new well, mixed with adenylation buffer and enzyme, and incubated at 65 ℃ for 30 minutes. This is followed by a purification with Clean Up beads and 70% IPA, the samples were resuspended with nuclease-free water and the resultant supernatant was transferred to a new well. Lastly, the PCR Master Mix and primer mix were added to the mixture followed by an incubation and final purification wash. The incubation consisted of an initial denaturation at 98 ℃ for 30 seconds; 13 PCR cycles of denaturation at 98 ℃ for 15 seconds, annealing at 65 ℃ for 30 seconds, and extension at 72 ℃ for 30 seconds; and a final extension at 72 ℃ for 2 minutes. The final purification was done with Clean Up Beads and 70% IPA.

Automated Adaptation: The automated version of the protocol ended up being divided into two different plates because of the amount of reagents needed to carry out this workflow. The first plate does the cDNA Synthesis portion of the protocol. Once this is finished the samples generated were transferred into the second plate that performs the library preparation portion of the kit.

Watchmaker Total RNA: Positive controls and automated libraries were assembled using the Watchmaker RNA Library Preparation Kit (Watchmaker Genomics, Boulder, CO) for Total RNA preparation. The standard manufacturer protocol was used for positive controls. 100 ng solution of input samples and fragmentation and priming buffer were mixed and incubated at 85 C for 3 minutes. Following incubation, first strand buffer and enzymes were added to the fragmented RNA and the mixture was incubated at 25 C for 10 minutes, 42 C for 15 minutes, and 70 C for 15 minutes. Next, a second strand buffer and enzyme

were added to the product of first strand synthesis, and the mixture was incubated at 42 C for 5 minutes and 62 C for 10 minutes. Post second strand synthesis, the product was mixed with adapters, ligation buffer, and an ligation enzyme and incubated at 20 C for 15 minutes. Purification then occurred using SPRI beads and 70% IPA. The adapter ligated samples were resuspended in Tris-Cl (pH 8.0), and the resulting mixture was transferred to a different well. This was mixed with Equinox Amplification Master Mix, and incubated at 98 C for 15 seconds, thermocycled for 8 cycles at 60 C for 30 seconds and 72 C for 45 seconds, and incubated at 72 C for 60 seconds. Finally, a post PCR purification step occurred with SPRI beads and 70% IPA. The final product was resuspended in Tris-Cl and transferred to a clean well.

Automated Adaptation: Several adaptations were made to the manufacturer protocol to make it compatible with the automation platform and workflow. To eliminate the need for a centrifuge or vortex within the device or additional user interaction, pipette mixing was used for all steps that required mixing, centrifugation, or vortexing. It was found that liquid transfer would periodically result in bubbles within the liquids. Therefore specific functions were developed for mixing to prevent bubbles within liquids in wells and for simple pipette mixing. The thermal system has a minimum temperature of 20 C, so thermal steps requiring holding at 4 C were altered to hold at room temperature. It was found that volumes less than 10 uL held the risk of evaporation, so adapter volumes were increased to 10 uL instead of 5 (maintaining the given concentration). A similar issue and solution was presented for the use of primer mix during PCR; 5 uL p5/p7 Primer mix and 5 uL nucleus free water used instead of just 5 uL primer. During bead based purification, the standard protocol indicates that pellets should be allowed to dry for 3-5 minutes between

ethanol washes. However, matching this time resulted in overdrying and decreased performance. Therefore, dry times were decreased to two minutes across all purifications. Finally, resuspension elution volumes, both for post ligation purification and post PCR purification were increased (to 25 and 40 uL respectively). This was to again prevent evaporation observed during the automated runs. All changes were individually tested to validate (i.e. one change was made at a time).

II. Nucleic Acid Input Sample

200 ng (concentration 14.29 ng/ul) of Human Total RNA (Male, BioChain, Newark, CA) was used as input for the Nextflex Rapid Directional RNA-Seq Kit 2.0 for Illumina Platforms. On the other hand, 100 ng (concentration 6.67 ng/ul) of Human Total RNA (Male, BioChain, Newark, CA) was used as the input for the Watchmaker RNA Library Preparation Kit.

III. Quality Control

The input Total RNA Quality Score and Integrity Number, and concentration were all obtained by using a Nanodrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). The final libraries were all quantified by using an 8 channel Qubit Flex Fluorometer with the dsDNA High Sensitivity (HS) assay (Invitrogen, Waltham, MA). As for quality control, the final libraries were verified by using an Agilent 2100 Bioanalyzer instrument with their High Sensitivity DNA assay for dsDNA (Agilent Technologies, Santa Clara, CA). The final libraries were diluted to the appropriate concentrations before quantification to fall within the range of the HS Qubit assay and Bioanalyzer High Sensitivity DNA assay before loading the chip.

IV. Automated Platform

Device: The automated liquid handler used for this work was the BioQule ™ NGS System (Revvity, Waltham, MA).

Consumables: The NextFlex Rapid Directional RNA-Seq Kit 2.0 automated workflow used two Thermo Scientific™ Nunc™ DeepWell™ 384-Well Plates (Thermo Fisher Scientific, Waltham, MA). It also used a new clear polymer tubing, double sided, 8-cannula cartridge for pipetting in each run (Revitty, Waltham, MA). The Watchmaker RNA Library Prep Kit automated workflow used one Thermo Scientific™ Nunc™ DeepWell™ 384-Well Plate (Thermo Fisher Scientific, Waltham, MA). This workflow also used a new black offset, double-sided, 8-cannula cartridge for each run (Revitty, Waltham, MA).

V. Scripting Resources

Assay development, as well as software and firmware interaction, were facilitated through Python. A single master script facilitates the control of the device by setting parameters and functions for mechanical feedback and linear stage and liquid movement as defined by pump rotations. This includes functions for conversion of pump rotation to linear motion, 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 specific script development is conducted in supplementary scripts that call upon functions from the master script. Functions for liquid displacement rely on assay developer input parameters of the 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. Functions for thermal heating require developer input parameters for cannula being used, desired

temperature, number of cycles, and desired heating time. Mixing functions are defined locally, and are specific to the intensity, quantity and purpose of the mixing. Each step of the automated workflows use the global and local functions to facilitate the reactions required to build the entire manual library preparation protocols. All supplementary scripts are stored in an assay development repository, hosted and managed via Bitbucket (Atlassian, Sydney, Australia).

VI. Sequencing Analysis

All samples were sequenced using the MiniSeq System (Illumina, San Diego, CA). Sequencing was paired end with read length 51 bp. Data analysis was performed using FastQC, Trimmomatic, STAR, Samtools, picard and custom R scripts. R functions used to generate figures were ggplot2, reshape2, tidyverse, scales, pals, RColorBrewer, and gridExtra.

Chapter 4

Conclusion and Future Directions

4.1 CONCLUSION

This thesis shows different automated approaches for an array of NGS library preparation kits. Even though Chapter 2 went in-depth into the device used and DNA-Seq NGS library preparation kits, this work was set to have an added focus on RNA-Seq. Since the trend of using RNA-Seq techniques to discover different ways to diagnose and treat already known diseases and discover new diseases, it is of key importance to find automated solutions to these workflows. The work presented shows optimized automated workflows that take into consideration the challenges explained in Chapter 1 to try and make them null. This was managed to be done by creating automated workflows that are cost-effective and easy to use, and by choosing to do this work in a robotic liquid handler that is small enough to fit in a laboratory bench top. Because of the facts explained above this makes our workflows accessible which in turn has the potential of making healthcare more accessible depending of the results found when using the methods with clinical samples.

4.2 FUTURE DIRECTIONS

This work only scratches the surface of the NGS library preparation kits that can be automated with this device. As a continuation of the RNA-Seq work, future work includes the automation of an rRNA depletion and a Poly(A) Tail enrichment workflow to complement the library preparation workflow. With this, it will be possible to get more specific data from mRNA components. Future work for this project outside of the scope of

RNA-Seq includes the automation of different protocols to analyze components such as tissues, proteins, cells, and much more. This will open the way for many opportunities in the research and clinical setting which in turn could influence making healthcare more accessible and affordable.

REFERENCES

1. Caprara, M. G., & Nilsen, T. W. (2000). RNA: Versatility in form and function. *Nature Structural Biology*, *7*, 831–833. https://doi.org/10.1038/82816
2. Butz, H., & Patócs, A. (2019). Brief Summary of the Most Important Molecular Genetic Methods (PCR, qPCR, Microarray, Next-Generation Sequencing, etc.). *Experientia Supplementum*, *111*, 33–52. https://doi.org/10.1007/978-3-030-25905-1_4
3. Page, G. P., Zakharkin, S. O., Kim, K., Mehta, T., Chen, L., & Zhang, K. (2007). Microarray analysis. *Methods in Molecular Biology (Clifton, N.J.)*, *404*, 409–430. https://doi.org/10.1007/978-1-59745-530-5_20
4. Bachman, J. (2013). Chapter Two - Reverse-Transcription PCR (RT-PCR). *Methods in Enzymology* , *530*, 67–74. https://doi.org/10.1016/B978-0-12-420037-1.00002-6.
5. Muzzey, D., Evans, E. A., & Lieber, C. (2015). Understanding the Basics of NGS: From Mechanism to Variant Calling. *Current Genetic Medicine Reports*, *3*(4), 158–165. https://doi.org/10.1007/s40142-015-0076-8
6. Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry*, *55*(4), 641–658. https://doi.org/10.1373/clinchem.2008.112789
7. Han, Y., Gao, S., Muegge, K., Zhang, W., & Zhou, B. (2015). Advanced Applications of RNA Sequencing and Challenges. *Bioinformatics and Biology Insights*, *9s1*. https://doi.org/10.4137/bbi.s28991

8. Holland, I., & Davis, J. (2020). Automation in the Life Science Research Laboratory. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.571777
9. Kong, F., Yuan, L., Zheng, Y. F., & Chen, W. (2012). Automatic Liquid Handling for Life Science. *Journal of Laboratory Automation*, *17*(3), 169–185. https://doi.org/10.1177/2211068211435302
10. Esteve-Codina, A. (2018). Chapter Four - RNA-Seq Data Analysis, Applications and Challenges. *Comprehensive Analytical Chemistry*, *82*, 71–106. Elsevier. https://doi.org/10.1016/bs.coac.2018.06.001.
11. Zhong, Y., Xu, F., Wu, J., Schubert, J., & Li, M. M. (2021). Application of Next Generation Sequencing in Laboratory Medicine. *Annals of Laboratory Medicine*, *41*(1), 25–43. https://doi.org/10.3343/alm.2021.41.1.25
12. Genzen, J. R., Burnham, C.-A. D., Felder, R. A., Hawker, C. D., Lippi, G., & Peck Palmer, O. M. (2018). Challenges and Opportunities in Implementing Total Laboratory Automation. *Clinical Chemistry*, *64*(2), 259–264. https://doi.org/10.1373/clinchem.2017.274068
13. Christofyllakis, K.; Bittenbring, J. T.; Thurner, L.; Ahlgrimm, M.; Stilgenbauer, S.; Bewarder, M.; Kaddu-Mulindwa, D. 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* **2022**, *16* (1), 21. DOI: 10.3892/mco.2021.2453.

14. Onoyama, T.; Ishikawa, S.; Isomoto, H. Gastric cancer and genomics: review of literature. *J Gastroenterol* **2022**, *57* (8), 505-516. DOI: 10.1007/s00535-022-01879-3.
15. 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* **2011**, *30* (2), 199-210. DOI: 10.1007/s10555-011-9278-z.
16. Servetto, A.; Napolitano, F.; De Angelis, C.; De Placido, P.; Giuliano, M.; Arpino, G.; De Placido, S.; Bianco, R.; Formisano, L. 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* **2021**, *157*, 103191. DOI: 10.1016/j.critrevonc.2020.103191.
17. Cinque, L.; Angeletti, C.; Orrico, A.; Castellana, S.; Ferrito, L.; Ciuoli, C.; Mazza, T.; Castori, M.; Guarnieri, V. 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* **2020**, *8* (12). DOI: 10.3390/biomedicines8120631.
18. Schnitzler, L. J.; Schreckenbach, T.; Nadaj-Pakleza, A.; Stenzel, W.; Rushing, E. J.; Van Damme, P.; Ferbert, A.; Petri, S.; Hartmann, C.; Bornemann, A.; et al. Sporadic late-onset nemaline myopathy: clinico-pathological characteristics and review of 76 cases. *Orphanet J Rare Dis* **2017**, *12* (1), 86. DOI: 10.1186/s13023-017-0640-2.

19. Suhaimi, S. A.; Zulkipli, I. N.; Ghani, H.; Abdul-Hamid, M. R. W. Applications of next generation sequencing in the screening and diagnosis of thalassemia: A mini-review. *Front Pediatr* **2022**, *10*, 1015769. DOI: 10.3389/fped.2022.1015769.
20. 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* **2021**, *258*, 424-429. DOI: 10.1016/j.ejogrb.2021.01.008.
21. Craig, D. J.; Morrison, T.; Khuder, S. A.; Crawford, E. L.; Wu, L.; Xu, J.; Blomquist, T. M.; Willey, J. C. 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* **2019**, *19* (1), 1081. DOI: 10.1186/s12885-019-6313-x.
22. Ebner, S.; Winkelmann, R.; Martin, S.; Kollermann, J.; Wild, P. J.; Demes, M. Sequencing of BRCA1/2-alterations using NGS-based technology: annotation as a challenge. *Oncotarget* **2022**, *13*, 464-475. DOI: 10.18632/oncotarget.28213.
23. Endris, V.; Stenzinger, A.; Pfarr, N.; Penzel, R.; Mobs, M.; Lenze, D.; Darb-Esfahani, S.; Hummel, M.; Sabine Merkelbach, B.; Jung, A.; 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 Arch* **2016**, *468* (6), 697-705. DOI: 10.1007/s00428-016-1919-8.
24. Petrackova, A.; Vasinek, M.; Sedlarikova, L.; Dyskova, T.; Schneiderova, P.; Novosad, T.; Papajik, T.; Kriegova, E. Standardization of Sequencing Coverage Depth in NGS: Recommendation for Detection of Clonal and Subclonal

Mutations in Cancer Diagnostics. *Front Oncol* **2019**, *9*, 851. DOI: 10.3389/fonc.2019.00851.

25. Schneider, L.; Cui, F.; Tripathi, A. Isolation of target DNA using synergistic magnetic bead transport and electrokinetic flow. *Biomicrofluidics* **2021**, *15* (2), 024104. DOI: 10.1063/5.0045307.
26. 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* **2021**, *11* (24), 14459-14474. DOI: 10.1039/d1ra01732b.
27. 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* **2022**, *42* (2), 240-253. DOI: 10.1002/pd.6092.
28. Snider, A.; Nilsson, M.; Dupal, M.; Toloue, M.; Tripathi, A. A Microfluidics Workflow for Sample Preparation for Next-Generation DNA Sequencing. *SLAS Technol* **2019**, *24* (2), 196-208. DOI: 10.1177/2472630318796133.
29. Gordon, L. G.; White, N. M.; Elliott, T. M.; Nones, K.; Beckhouse, A. G.; Rodriguez-Acevedo, A. J.; Webb, P. M.; Lee, X. J.; Graves, N.; Schofield, D. J. Estimating the costs of genomic sequencing in cancer control. *BMC Health Serv Res* **2020**, *20* (1), 492. DOI: 10.1186/s12913-020-05318-y.
30. Lee, K.; Tripathi, A. Parallel DNA Extraction From Whole Blood for Rapid Sample Generation in Genetic Epidemiological Studies. *Front Genet* **2020**, *11*, 374. DOI: 10.3389/fgene.2020.00374.

31. Bronner, I. F.; Quail, M. A. Best Practices for Illumina Library Preparation. *Curr Protoc Hum Genet* **2019**, *102* (1), e86. DOI: 10.1002/cphg.86.
32. Ribarska, T.; Bjornstad, 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* **2022**, *23* (1), 92. DOI: 10.1186/s12864-022-08316-y.
33. Chiniquy, J.; Garber, M. E.; Mukhopadhyay, A.; Hillson, N. J. Fluorescent amplification for next generation sequencing (FA-NGS) library preparation. *BMC Genomics* **2020**, *21* (1), 85. DOI: 10.1186/s12864-020-6481-8.
34. Dallavilla, T.; Marceddu, G.; Casadei, A.; De Antoni, L.; Bertelli, M. A fast, reliable and easy method to detect within-species DNA contamination. *Acta Biomed* **2020**, *91* (13-S), e2020019. DOI: 10.23750/abm.v91i13-S.10531.
35. Hess, J. F.; Kohl, T. A.; Kotrova, M.; Ronsch, K.; Paprotka, T.; Mohr, V.; Hutzenlaub, T.; Bruggemann, M.; Zengerle, R.; Niemann, S.; et al. Library preparation for next generation sequencing: A review of automation strategies. *Biotechnol Adv* **2020**, *41*, 107537. DOI: 10.1016/j.biotechadv.2020.107537.
36. Huang, J.; Li, C. A modified protocol with less clean-up steps increased efficiency and product yield of sequencing library preparation. *3 Biotech* **2022**, *12* (5), 111. DOI: 10.1007/s13205-022-03168-5.
37. Muscarella, L. A.; Fabrizio, F. P.; De Bonis, M.; Mancini, M. T.; Balsamo, T.; Graziano, P.; Centra, F.; Sparaneo, A.; Trombetta, D.; Bonfitto, A.; et al. Automated Workflow for Somatic and Germline Next Generation Sequencing

Analysis in Routine Clinical Cancer Diagnostics. *Cancers (Basel)* **2019**, *11* (11). DOI: 10.3390/cancers11111691.

38. Tegally, H.; San, J. E.; Giandhari, J.; de Oliveira, T. Unlocking the efficiency of genomics laboratories with robotic liquid-handling. *BMC Genomics* **2020**, *21* (1), 729. DOI: 10.1186/s12864-020-07137-1.
39. Hess, J. F.; Hess, M. E.; Zengerle, R.; Paust, N.; Boerries, M.; Hutzenlaub, T. Automated library preparation for whole genome sequencing by centrifugal microfluidics. *Anal Chim Acta* **2021**, *1182*, 338954. DOI: 10.1016/j.aca.2021.338954.
40. Kim, H.; Jebrail, M. J.; Sinha, A.; Bent, Z. W.; Solberg, O. D.; Williams, K. P.; Langevin, S. A.; Renzi, R. F.; Van De Vreugde, J. L.; Meagher, R. J.; et al. A microfluidic DNA library preparation platform for next-generation sequencing. *PLoS One* **2013**, *8* (7), e68988. DOI: 10.1371/journal.pone.0068988.
41. Tan, S. J.; Phan, H.; Gerry, B. M.; Kuhn, A.; Hong, L. Z.; Min Ong, Y.; Poon, P. S.; Unger, M. A.; Jones, R. C.; Quake, S. R.; 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* **2013**, *8* (7), e64084. DOI: 10.1371/journal.pone.0064084.
42. 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* **2020**, *92* (3), 2519-2526. DOI: 10.1021/acs.analchem.9b04086.
43. *Plastics - Thermal Conductivity Coefficients*. https://www.engineeringtoolbox.com/thermal-conductivity-plastics-d_1786.html

44. *Material Properties*. https://www.engineeringtoolbox.com/material-properties-t_24.html

45. Aird, D.; Ross, M. G.; Chen, W. S.; Danielsson, M.; Fennell, T.; Russ, C.; Jaffe, D. B.; Nusbaum, C.; Gnirke, A. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* **2011**, *12* (2), R18. DOI: 10.1186/gb-2011-12-2-r18.

46. Avila-Rios, S.; Parkin, N.; Swanstrom, R.; Paredes, R.; Shafer, R.; Ji, H.; Kantor, R. Next-Generation Sequencing for HIV Drug Resistance Testing: Laboratory, Clinical, and Implementation Considerations. *Viruses* **2020**, *12* (6). DOI: 10.3390/v12060617.

47. Zhang, L.; Wang, J.; Coetzer, M.; Angione, S.; Kantor, R.; Tripathi, A. One-Step Ligation on RNA Amplification for the Detection of Point Mutations. *J Mol Diagn* **2015**, *17* (6), 679-688. DOI: 10.1016/j.jmoldx.2015.07.001.

48. 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* **2012**, *40* (7), e54. DOI: 10.1093/nar/gkr1263.

49. Li, T. W.; Weeks, K. M. Structure-independent and quantitative ligation of single-stranded DNA. *Anal Biochem* **2006**, *349* (2), 242-246. DOI: 10.1016/j.ab.2005.11.002.

50. Bauer, R. J.; Zhelkovsky, A.; Bilotti, K.; Crowell, L. E.; Evans, T. C., Jr.; McReynolds, L. A.; Lohman, G. J. S. Comparative analysis of the end-joining activity of several DNA ligases. *PLoS One* **2017**, *12* (12), e0190062. DOI: 10.1371/journal.pone.0190062.

51. *Comparison of fluorescence-based quantitation with UV absorbance measurements*. 2022. https://www.k-state.edu/igenomics/services/documents/comparisson%20between%20qubit.pdf

52. *Best practices for manually normalizing library concentrations*. https://support.illumina.com/bulletins/2017/03/best-practices-for-manually-normalizing-library-concentrations.html

53. *Quality Scores for Next-Generation Sequencing* https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf

54. Piovesan, A.; Pelleri, M. C.; Antonaros, F.; Strippoli, P.; Caracausi, M.; Vitale, L. On the length, weight and GC content of the human genome. *BMC Res Notes* **2019**, *12* (1), 106. DOI: 10.1186/s13104-019-4137-z.

55. 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* **2018**, *6* (25). DOI: 10.1128/genomeA.00515-18.

56. Tanaka, N.; Takahara, A.; Hagio, T.; Nishiko, R.; Kanayama, J.; Gotoh, O.; Mori, S. Sequencing artifacts derived from a library preparation method using enzymatic fragmentation. *PLoS One* **2020**, *15* (1), e0227427. DOI: 10.1371/journal.pone.0227427.

57. Nutman, A.; Marchaim, D. How to: molecular investigation of a hospital outbreak. *Clin Microbiol Infect* **2019**, *25* (6), 688-695. DOI: 10.1016/j.cmi.2018.09.017.

58. Quainoo, S.; Coolen, J. P. M.; van Hijum, S.; Huynen, M. A.; Melchers, W. J. G.; van Schaik, W.; Wertheim, H. F. L. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. *Clin Microbiol Rev* **2017**, *30* (4), 1015-1063. DOI: 10.1128/CMR.00016-17

59. Qin, D. (2019). Next-generation Sequencing and Its Clinical Application. *Cancer Biology & Medicine*, *16*(1), 4–10. https://doi.org/10.20892/j.issn.2095-3941.2018.0055

60. Buermans, H. P. J., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1842*(10), 1932–1941. https://doi.org/10.1016/j.bbadis.2014.06.015

61. Hong, M., Tao, S., Zhang, L., Diao, L.-T., Huang, X., Huang, S., Xie, S.-J., Xiao, Z.-D., & Zhang, H. (2020). RNA sequencing: new technologies and applications in cancer research. *Journal of Hematology & Oncology*, *13*(1). https://doi.org/10.1186/s13045-020-01005-x

62. Guo, P., Haque, F., Hallahan, B., Reif, R., & Li, H. (2012). Uniqueness, Advantages, Challenges, Solutions, and Perspectives in Therapeutics Applying RNA Nanotechnology. *Nucleic Acid Therapeutics*, *22*(4), 226–245. https://doi.org/10.1089/nat.2012.0350

63. Riscado, M., Baptista, B., & Sousa, F. (2021). New RNA-Based Breakthroughs in Alzheimer's Disease Diagnosis and Therapeutics. *Pharmaceutics*, *13*(9), 1397. https://doi.org/10.3390/pharmaceutics13091397

64. Rački, N., Morisset, D., Gutierrez-Aguirre, I., & Ravnikar, M. (2013). One-step RT-droplet digital PCR: a breakthrough in the quantification of waterborne RNA viruses. *Analytical and Bioanalytical Chemistry*, *406*(3), 661–667. https://doi.org/10.1007/s00216-013-7476-y

65. Chheda, U., Selvi Pradeepan, Esposito, E. A., Strezsak, S. R., Fernandez-Delgado, O., & Kranz, J. K. (2023). Factors affecting stability of RNA – temperature, length, concentration, pH, and buffering species. *Journal of Pharmaceutical Sciences*, *113*(2). https://doi.org/10.1016/j.xphs.2023.11.023

66. CD Genomics. (n.d.). *mRNA Sequencing vs Total RNA Sequencing - CD Genomics*. Www.cd-Genomics.com. Retrieved April 18, 2024, from https://www.cd-genomics.com/mrna-sequencing-vs-total-rna-sequencing.html#:~:text=Total%20RNA%2DSeq%2C%20however%2C

67. Illumina. (n.d.). *Total RNA Sequencing | Whole-transcriptome sequencing solutions*. Www.illumina.com. https://www.illumina.com/techniques/sequencing/rna-sequencing/total-rna-seq.html

68. Pereira, M. A., Imada, E. L., & Muniz Guedes, R. L. (2017). RNA-seq: Applications and Best Practices. *Applications of RNA-Seq and Omics Strategies - from Microorganisms to Human Health*. https://doi.org/10.5772/intechopen.69250