Novel High-throughput Technologies for Applications in Microbiology

by

Sijie Chen

B.S., B.Econ., Peking University (2017)
S.M., Massachusetts Institute of Technology (2019)

Submitted to the Department of Mechanical Engineering and
the Institute for Data, Systems, and Society
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering and Statistics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2023

Authored by: Sijie Chen
Department of Mechanical Engineering and
the Institute for Data, Systems, and Society
July 31, 2023

Certified by: Cullen R. Buie
Associate Professor of Mechanical Engineering
Thesis Supervisor

Accepted by: Nicolas Hadjiconstantinou
Professor of Mechanical Engineering
Chairman, Department Committee on Graduate Theses

Accepted by: Themistoklis Sapsis
Professor of Mechanical and Ocean Engineering
Chairman, Interdisciplinary Doctoral Program in Statistics

Novel High-throughput Technologies for Applications in Microbiology

by

Sijie Chen

Submitted to the Department of Mechanical Engineering and
the Institute for Data, Systems, and Society
on July 31, 2023, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering and Statistics

Abstract

Electroporation, a widely used gene delivery technique for bacterial transformation, involves applying an electric field to increase cell membrane permeability. However, the traditional cuvette-based protocol is laborious, time-consuming, low throughput, and operator dependent, hindering the progress of bacterial genetic engineering. To address these challenges, this study focuses on developing high-throughput solutions to streamline the process, improve efficiency, and ensure consistent results.

The primary contribution of this research is the development of—to the best of our knowledge—the first automated system for high-throughput bacterial genetic engineering. At the core of this system is a microfluidic device designed to integrate seamlessly with liquid handling robots. By performing electroporation column-by-column, this automated system achieves a 96-well electroporation in just 5 minutes, 30 times faster than performing 96 independent genetic transformations using cuvettes. Through experiments with *Escherichia coli* NEB10β, it demonstrates high efficiency and consistency, enabling rapid and reliable genetic manipulation.

The introduction of high throughput electroporation results in other bottlenecks in the process, particularly in the assessment of transformation efficiency through colony counting. Existing automated colony counting solutions often struggle with merged colonies, a common issue in high-throughput plating. To address the need for a high-throughput evaluation solution to assess the efficiency of bacterial electroporation, we introduce MCount. MCount outperforms existing solutions by optimizing contour and regional pairing, resulting in low error rates and minimal reliance on hyperparameters. Moreover, we propose statistical methods that require few labeled or even unlabeled datapoints, ensuring consistently low error rates and facilitating deployment in scenarios with limited labeled data.

Finally, as conventional electroporation methods are limited to relatively small sample volumes ($<$ 200 L), we present M-tube, a disposable and user-friendly microfluidic device for large volume ($>$ 10 mL) bacterial gene delivery. With minimal fabrication requirements and straightforward operation, M-tube surpasses cuvettes in transformation

efficiency and facilitates the creation of transposon mutant libraries.

Collectively, the high-throughput solutions for microbiology overcome the limitations of cuvette-based electroporation, significantly improving efficiency and consistency in bacterial genetic engineering. These advancements pave the way for accelerated research and development in the field, fostering breakthroughs in genetic manipulation and biotechnology applications.

Thesis Supervisor: Cullen R. Buie
Title: Associate Professor of Mechanical Engineering

Acknowledgments

I wish to express my profound gratitude to my advisor, Professor Cullen R. Buie, for his unwavering support and guidance over these past six years. Far beyond his role as an academic supervisor, Professor Buie has served as a life advisor. During a period when I felt lost, he didn't simply provide answers, but shared his personal experiences, teaching me how to find and navigate the right path. He exemplified the kindness and decency a person could embody and taught me to balance work and family life effectively. He made MIT feel like home. My respect for him extends well beyond our academic relationship.

I wish to express my sincere thanks to my committee members, Professor Rohit N. Karnik and Professor Jeff Gore, for their insightful feedback and advice across all my research undertakings. Professor Karnik, with his empathetic approach, ensured the continuation of individual, in-person discussions even amidst a pandemic-induced virtual landscape. Despite physics being his main domain, Professor Gore consistently demonstrated a deep understanding of my work, providing valuable insights, whether related to mechanical engineering, microbiology, or statistics. Their contributions were instrumental in shaping my PhD thesis.

My profound thanks go to my colleagues at the Laboratory for Energy and Microsystems Innovation, whose assistance was indispensable in completing my PhD journey. I am particularly thankful to Dr. Qianru Wang for her comprehensive guidance on mechanical engineering in my early days in the lab. My collaborations with Dr. Po-Hsun Huang over four years and the shared growth with Dr. Hyungseok Kim over six years were significant parts of my journey. Dr. Chris Vaiana, Dr. Chelsea Catania, and Dr. Jonathan Cottet deserve special mention for their contributions to our discussions on bioengineering. Additional thanks go to Dr. Paulo Garcia, Rameech McCormack, Dr. Zhifei Ge, Dr. Kameron Conforti, and our administrator Maral Banosian.

I extend my heartfelt gratitude to all the friends who have enriched my PhD journey. Their support, friendship, and camaraderie have greatly contributed to my growth and success.

Lastly, I wish to express my deepest gratitude to my family. Their steadfast understanding, encouragement, and love have been my greatest source of strength, propelling me through this challenging journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Electroporation Overview

Gene delivery is a crucial step in genetic engineering, involving the transfer of foreign genetic materials like DNA and RNA into host cells to modify the characteristics of organisms [1]. This process plays a fundamental role in understanding gene function and protein expression, as well as enabling the development of genetically modified organisms. The applications of gene delivery span a wide range of fields, including medicine, agriculture, gene therapy, and industrial biotechnology. These advancements have led to significant breakthroughs such as the production of genetically modified animals [2], the synthesis of human insulin [3, 4] and hormones [5], the cultivation of genetically modified crops with improved traits [6, 7], and the development of industrial enzymes with enhanced efficiency [8].

Various techniques are employed for gene delivery, each with its advantages and limitations. Viral vectors, such as retroviruses and adenoviruses, are highly efficient at delivering genes into cells but suffer from problems including frame-shift mutations by random insertion of target genes into the host genome, toxic side effects caused by viral components like inflammatory responses, nonspecific targeting issues, and high cost of vector production [9–11]. Non-viral vectors, such as liposomes and nanoparticles, offer

safer alternatives but often exhibit lower transfection efficiency [12]. Conjugation, another biological approach, is a natural mechanism of gene transfer observed primarily in bacterial systems, where genetic material is directly transferred between bacterial cells through physical contact [13]. Physical techniques, including sonoporation, photoporation, magnetofection, hydroporation, and electroporation, create temporary pores in the cell membrane, enabling the entry of DNA or other molecules, which are often combined with microfluidic techniques for a more precise and controllable gene delivery [14–41].

Electroporation, a gene delivery method that increases membrane permeability with the assistance of an electric field, has garnered particular interest for microbiology transformation since the 1950s [42]. By the early 1970s, researchers recognized the phenomenon of dielectric breakdown in the cell membrane at critical electric field strengths [43]. In 1972, electroporation was proposed as a technique for non-viral gene delivery for the first time [44]. Subsequent research showed that small molecules could pass through these electric field-induced membrane pores in various cell types [45].

Over the years, electroporation has emerged as the preferred gene delivery method for bacterial transformation. Its simple setup, ease of control, and high efficiency set it apart from other techniques. Unlike chemical or biological methods that often require specific chemicals or vectors tailored to particular bacterial strains, electroporation offers a versatile and general approach. Other physical methods, like sonoporation, microinjection, and photoporation, create pores that are usually larger than or comparable to the size of bacteria, limiting their suitability for bacterial transformation. Additionally, magnetofection and hydroporation involve more complex systems compared to electroporation. In contrast, electroporation's simplicity, controllability, and high efficiency make it the optimal choice for bacterial transformation [1, 3–12, 14–42, 44, 46–74].

1.2 Limitations of Existing Electroporation Protocols

The traditional electroporation protocol typically involves the use of electroporation cuvettes to facilitate the process. Initially, genetic material is prepared and combined with host cells in a suitable buffer solution. The resulting mixture is then transferred into an electroporation cuvette equipped with electrodes. Subsequently, the cuvette is inserted into an electroporation device, where an electrical pulse is administered to induce electroporation. Following the pulse, the cuvette is removed, and the cells undergo a recovery period in a growth medium for subsequent analysis. While this protocol has demonstrated effectiveness in numerous applications, it is important to consider certain limitations associated with this approach.

One significant limitation of the cuvette-based electroporation protocol is its inherent batch nature, resulting in low throughput, processing less than 30 samples per hour. This constraint becomes especially problematic when establishing electroporation protocols for unfamiliar strains, requiring researchers to conduct thousands of experiments with various factors such as plasmids, buffer medium, and electric conditions, often spanning months or even years. To address this challenge, commercial products like the Bio-Rad Gene Pulser MXcell™ Electroporation System, the Lonza 4D-Nucleofector® 96-Well Unit, and the BTX Electroporation Plate were introduced for high-throughput cuvette electroporation. However, these solutions have their drawbacks. Bio-Rad has discontinued their product, the Lonza device is tailored for mammalian cells, and the BTX plate faces issues like frequent arcing, significant residual volume in cuvettes, and reduced transformation efficiency.

The cuvette-based electroporation protocol also suffers from the accumulation of joule heat within the cuvette, which can result in a high cell mortality rate and electric breakdown of the cell sample. To mitigate this concern, microfluidic techniques have been introduced to enable flow-through electroporation, where a continuous flow of media through the system effectively dissipates heat and minimizes cell death. Flow-through electroporation also eliminates the dead volume that can remain in the cu-

vettes, thereby improving overall process efficiency. Furthermore, microfluidics offers a compatible platform for integrating additional techniques aimed at enhancing transformation efficiency. For example, the utilization of polyelectrolytic gel electrodes can help suppress bubble formation [14]. Semicontinuous flow and hydrodynamics enable precise control of the flow dynamics [15, 20–24]. Additionally, the incorporation of microfluidic valves and DC voltage sources allows for the generation of faster electric pulses compared to conventional pulse generators [17–19]. Despite these advancements in microfluidic integration contributing to the overall enhancement of transformation efficiency during the electroporation process, such approaches have not been explored in high-throughput electroporation settings. There remains a promising opportunity to leverage microfluidic technology to further improve the efficiency and scalability of electroporation in high-throughput applications.

A final issue with existing electroporation protocols, whether microfluidic-based or commercial products, is the requirement for extensive manual operation. This not only makes the process time-consuming and labor-intensive but also introduces the potential for inconsistent results due to human factors. To address this issue, it is crucial to incorporate automation into the electroporation process, similar to the advancements seen in other fields where automation has brought numerous benefits. For instance, automation has revolutionized processes such as high-throughput screening in drug discovery [75] and laboratory automation using robotic systems [76]. Despite these advancements in other areas, the field of electroporation has yet to witness significant efforts towards automation, highlighting the potential for further development in this critical area.

1.3 Study Objectives and Thesis Organization

This study aims to develop high-throughput solutions that overcome the limitations of the current cuvette-based bacterial electroporation approach, with a focus on enhancing automation and improving consistency. The thesis is divided into several chapters, each

addressing different facets of this objective.

Chapter 2 proposes and constructs—to the best of our knowledge—the first automated system for high-throughput bacterial genetic engineering. At the core of this system is our innovative microfluidic, flow-through 96-well electroporation device, designed to be compatible with commercial liquid handling robots. By performing electroporation column-by-column, our system achieves a 96-well electroporation in approximately 5 minutes, which is 30 times faster than performing 96 independent cuvette electroporations. Through experiments with *Escherichia coli* NEB10β, we demonstrated the system's effectiveness in screening electroporation conditions and achieving transformation efficiencies comparable to cuvettes. Notably, the system exhibited remarkable efficiency and consistency across channels and columns, underscoring its reliability and reproducibility. Overall, our automated system significantly expands the capabilities of bacterial electroporation, enabling rapid and efficient genetic manipulation in a scalable format.

Chapter 3 addresses the demand for a high-throughput evaluation solution to assess the efficiency of bacterial electroporation. Existing automated counting solutions struggle with the issue of merged colonies, a common occurrence in high throughput plating. To overcome this limitation, we propose MCount, the only known solution that incorporates both contour information and regional algorithms for colony counting. By optimizing the pairing of contours with regional candidate circles, MCount can accurately infer the number of merged colonies. We evaluate MCount on a precisely labeled *Escherichia coli* dataset of 960 images (15,847 segments) and achieve an averaged error rate of 3.99%, significantly outperforming existing published solutions such as NICE (16.54%). MCount also demonstrates strong generalization with only two hyperparameters. To further facilitate deployment in scenarios with limited labeled data, we propose statistical methods that require few labeled or even unlabeled datapoints, all of which guarantee consistently low error rates. MCount presents a promising solution for accurate and efficient colony counting in application workflows requiring high throughput.

Chapter 4 deals with the limitation of conventional electroporation methods, including cuvette-based and microfluidics-based approaches, as well as the automated system, which are typically designed for small sample volumes. This constraint significantly hampers their effectiveness in applications demanding large volumes, such as the generation of mutant libraries. To address this challenge, we introduce M-tube, a disposable and user-friendly microfluidic electroporation device providing a scalable and efficient bacterial gene delivery approach. This device requires minimal fabrication and offers straightforward operation. It can surpass traditional cuvettes in a variety of situations, including across different *Escherichia coli* strains with various electroporation efficiencies. Moreover, we utilized its large-volume bacterial electroporation capability to produce a transposon mutant library in the anaerobic gut commensal Bifidobacterium longum.

Finally, Chapter 5 discusses the major conclusions and a broader impact of this study, as well as opportunities for further improvement and future work.

Chapter 2

Automated Bacterial Electroporation System

This chapter is reproduced in part from:

Chen, S.*, Huang, P. H.*, & Buie, C. R.. A microfluidic 96-well electroporation device for automated, high-throughput bacterial genetic engineering. *Submitted*. *Equal contribution.

2.1 Introduction

In Chapter 1, we identified the limitations of the traditional cuvette-based electroporation protocol for bacterial genetic engineering and discussed the need for high-throughput solutions to overcome these challenges. In this chapter, we present our research on the development of an automated system specifically designed for high-throughput bacterial genetic engineering, as depicted in Figure 2-1. We describe the design and construction of a microfluidic device as the core component of the system, the overall system architecture, and the experiment with *Escherichia coli* for evaluating the performance and reliability of the system.

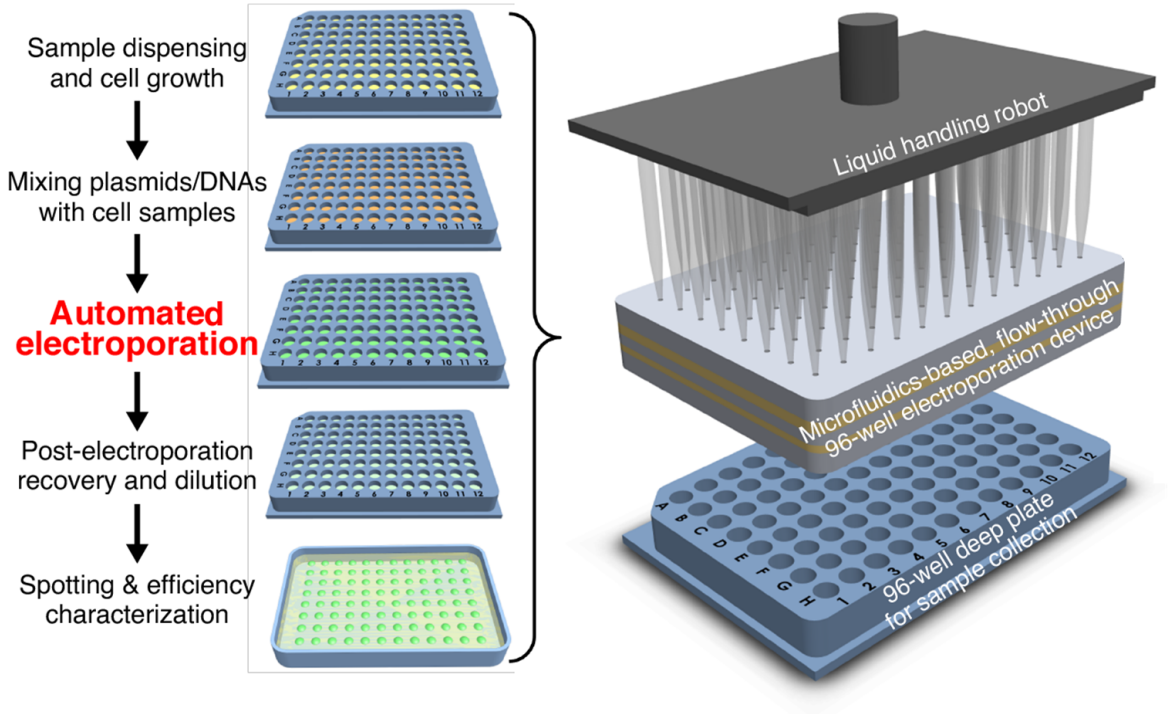


Figure 2-1: **Schematic of the automated system for high-throughput bacterial genetic engineering.** The diagram depicts an ideal, fully automated electrotransformation workflow facilitated by liquid handling robots. The system's primary focus and the most challenging task is the "Automated electroporation" step in a 96-well format, which is the main emphasis of this study.

2.2 Design of the Microfluidic Device

This section focuses on the design of the microfluidic device, which serves as the central element of the automated system. Our discussion encompasses key considerations and design principles that were adopted during its development.

The microfluidic device, as depicted in Figure 2-2a, consists of five main layers: two sealing layers, two electrode layers, and an acrylic channel layer. Mechanical integrity is ensured by bonding each layer together using a pressure-sensitive adhesive layer. To establish an electric field within the channel layer, an electric pulse is applied across the two electrode layers. For enhanced electrical safety, the sealing layers are constructed using electrically insulating materials, effectively enclosing the electrode layers. Seam-

less contact with pipette tips is guaranteed by fabricating the top sealing layer with silicon material, which also features a converging shape at the fluid entrance. This design facilitates optimal interaction with pipette tips while minimizing residual volume. Additionally, a nozzle is incorporated into the bottom sealing layer to prevent liquid splattering at the fluid exit.

To ensure compatibility with commercially available liquid handling robots capable of automated liquid transfer, we specifically designed the microfluidic device in a 96-format, allowing column-by-column operations (Figure 2-2b). To minimize inter-channel and inter-device variation, we adopt injection molding as the fabrication process, which allows for standardized manufacturing of the microfluidic device. Moreover, the use of cost-effective PCB boards for the electrode layers contributes to the overall affordability of the microfluidic device.

In achieving high transformation efficiency through electroporation, the configuration of the channel layer plays a pivotal role. Multiple design considerations and trade-offs are meticulously evaluated. While a thicker layer is desirable for more consistent flow and electric field within the channel, the upper bound of the voltage provided by the power system must be considered. It is crucial to strike a balance and ensure that the layer is thick enough to establish a sufficient electric field within the channel layer while adhering to the limitations imposed by the power system's voltage capacity.

Additionally, the shape and diameter of the channel significantly impact the efficiency of the electroporation process. Given the challenges associated with precise fabrication of straight channels using injection molding, a practical alternative is to adopt a converging channel shape. This design choice has been proven to contribute to higher transformation efficiency [77]. The channel diameter also warrants careful consideration as it directly influences the current during electroporation. Selecting the optimal diameter is crucial to prevent exceeding the power system's limits or causing electric arcing. Meanwhile, the resolution of the injection molding process imposes constraints on the lower bound of the channel diameter. Furthermore, the residence time, reflecting the sample's flow duration through the channel, is a critical factor in

achieving high transformation efficiency and is influenced by the channel shape.

To determine the appropriate channel configuration, a combination of COMSOL simulations and trial-and-error experiments was conducted. Based on these evaluations, a channel thickness of 3 mm and a converging channel with a minimum (maximum) diameter of 0.5 (1) mm, as illustrated in Figure 2-2a, were determined to be the optimal parameters for the microfluidic device. These findings strike a balance between efficient electroporation and the practical limitations of the power system and injection molding process.

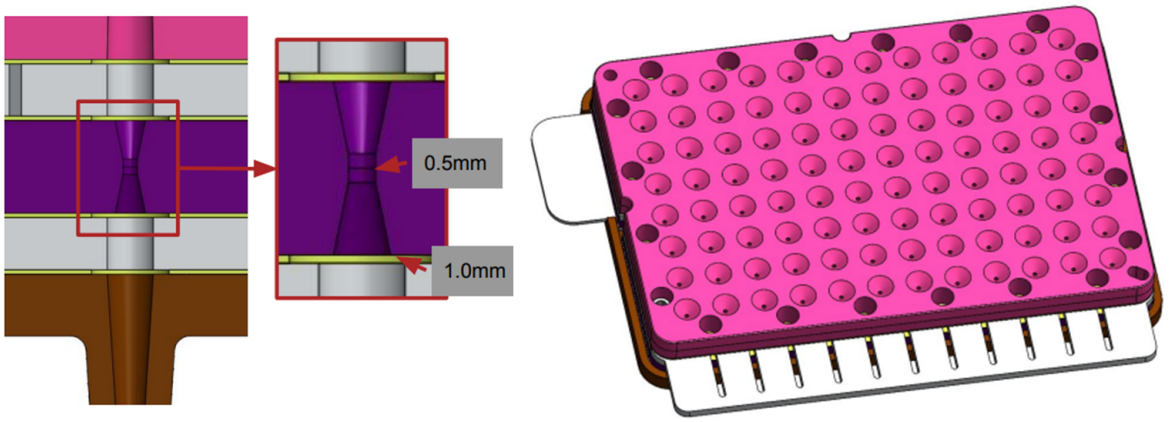


Figure 2-2: **Microfluidic device design for high-throughput bacterial genetic engineering.** (a) Illustration of the microfluidic device comprising five main layers, including sealing layers, electrode layers, and an acrylic channel layer. (b) Close-up view of the microfluidic device in the 96-well format for compatibility with liquid handling robots, showcasing its standardized design for injection molding fabrication. This figure is reproduced from the collaboration with manufacturing consultant RE:Build Fikst.

2.3 Design of the Automated System

In this section, we present the design and architecture of the automated system specifically developed for high-throughput bacterial genetic engineering. The overall system consists of several vital subsystems that work together seamlessly to enhance efficiency and productivity.

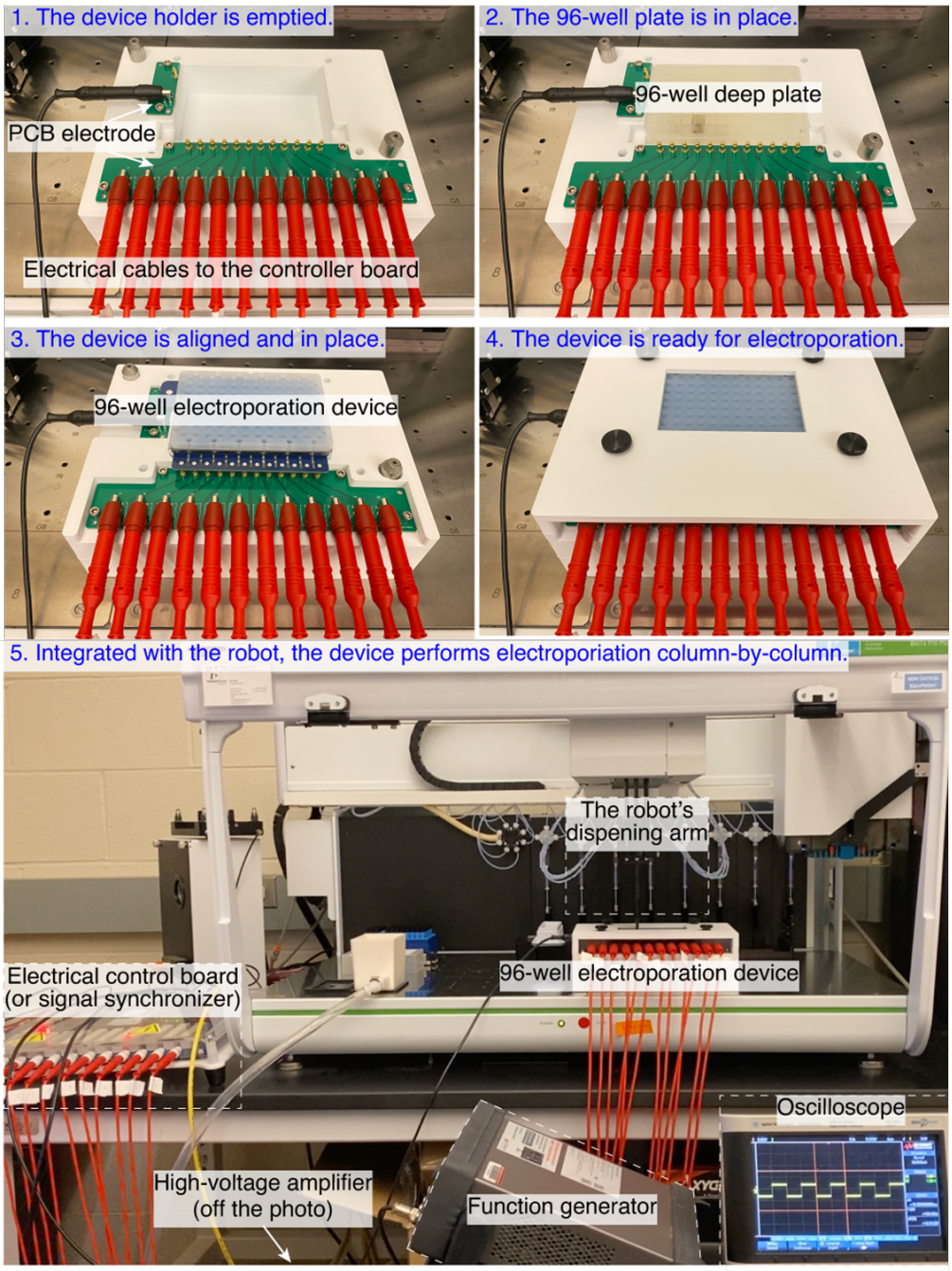


Figure 2-3: **Images of the experimental setup and procedure for the 96-well electroporation device.** With the device holder, the device and the 96-well deep plate can securely sit on the platform of the robot, ensuring stable operation for the entire electroporation process. The control board serves to communicate between the robot, function generator, and high-voltage amplifier, ensuring the synchronization between sample dispensing and electric pulse delivery. Our automated system can perform electroporation on 8 channels at a time (i.e., column-by-column), and the electroporation of an entire device can be completed in roughly 5 mins.

As depicted in Figure 2-3, a device holder with embedded PCB-based connectors is designed to encapsulate both the 96-well deep plate for sample collection and the microfluidic device. This holder ensures precise alignment and secure accommodation of the device's electrodes, enabling the activation of arbitrary columns for electroporation.

The power supply subsystem is another essential component, comprising a waveform generator and a high-voltage amplifier. The waveform generator offers flexibility in delivering electric pulses with user-defined parameters such as amplitude, frequency, and duty cycle, while the high-voltage amplifier guarantees the provision of sufficient electric amplitude. Additionally, an oscilloscope is connected to monitor the applied electric field in real-time.

Bubble formation is a common occurrence at the beginning of sample fulfillment in the channel, which can potentially lead to electric arcing if an electric field is present. Similarly, residual sample after dispensation also poses a risk of electric arcing. To mitigate these challenges and minimize the occurrence of electric arcing, it is essential to synchronize the pulse delivery and sample dispensation. For this purpose, a customized electrical control board is developed. This control board receives commands from the liquid handling robot and delivers amplified electric signals from the power supply to the selected columns of the device holder. This synchronization ensures accurate timing and precise delivery of pulses to each column of the microfluidic device, minimizing the risk of electric arcing and ensuring reliable electroporation results.

By integrating these subsystems, the automated system enables electroporation to be performed column-by-column, significantly reducing processing time. The entire electroporation process for a 96-well device can be completed in approximately 5 minutes, which is 30 times faster compared to performing 96 independent genetic transformations using cuvettes. This remarkable improvement in throughput underscores the efficiency and productivity benefits of the automated system for high-throughput bacterial genetic engineering.

2.4 Experimental Evaluation

This section focuses on the experimental evaluations conducted to assess the performance and reliability of the automated system. We describe two sets of experiments: screening electroporation conditions and evaluating electroporation consistency. For the screening experiments, we investigated parameters including electric field strength and flow rate for *Escherichia coli* NEB10β. In the consistency evaluation experiments, we analyze the reproducibility of the automated system by performing multiple electroporations using the same conditions.

2.4.1 Cultivation and Preparation of Competent Cells

To prepare *Escherichia coli* NEB10β, a glycerol stock was used to inoculate 5 mL of LB medium in two 14-mL culture tubes. The cultures were incubated overnight at 37°C and 250 rpm. The following day, 5 mL of each culture was transferred to 245 mL of LB medium and grown at 37°C and 200 rpm until reaching the exponential phase, as determined by measuring the OD_{600} using a UV spectrophotometer (UV-1800, Shimadzu). The cell cultures were then subjected to centrifugation at 3500 rpm and 4°C using an Allegra 64R Benchtop Centrifuge (F0650 rotor, Beckman Coulter), and the supernatant was carefully removed. The resulting cell pellets were washed with 2 mL of pre-chilled 10% glycerol at 4°C. This washing process was repeated after three more centrifugations at 8000 rpm and 4°C using an Allegra 64R Benchtop Centrifuge (F1202 rotor, Beckman Coulter). After the final centrifugation, the cell pellets were resuspended in 10% glycerol, adjusting the volume to achieve a final cell concentration of approximately OD_{600} of 10. To introduce the desired genetic material, ampicillin resistance and GFP encoding DNA plasmids (Parts Registry K176011) were mixed with the cell sample, resulting in a final concentration of 0.1 ng/μL. The sample was then subjected to electroporation using the automated system. For cuvette-based electroporation, the sample was loaded into 2 mm VWR electroporation cuvettes and electroporated using a MicroPulser™ (Bio-Rad) set at 2.5 kV with a 6 ms time constant. Immediately after

electroporation, the sample was suspended in 900 μL of pre-warmed LB medium at 37°C and 250 rpm for one hour. Finally, 100 μL of the cell suspension, either undiluted or diluted by 10-fold, 100-fold, or 1000-fold, was carefully pipetted onto LB agar plates containing ampicillin (50 μg/mL). The plates were then incubated overnight at 37°C and 250 rpm to facilitate colony growth and enable subsequent analysis of transformed colonies.

2.4.2 Screening Electroporation Conditions

To investigate the system's ability for screening electroporation conditions, e.g., electric field, we performed experiments using *E. coli* NEB10β. With a fixed flow rate of 400 μL/sec, we applied varying electric fields, ranging from 1.67 to 18.3 kV/cm, to different columns of the device. This allowed for 8 technical replicates at each electric field. As a positive control, we conducted electroporation in cuvettes with 2-mm gaps using different voltages, while a negative control involved the cell sample flowing through the device without the presence of an electric field. After electroporation, the cell suspensions were immediately transferred to pre-warmed LB medium and incubated at 37°C and 250 rpm for one hour to facilitate recovery and gene expression. These experiments were repeated for three biological replicates.

The results of the experiment regarding colony formation under 1000-fold dilution are illustrated in Figure 2-4, and the corresponding transformation efficiency for each condition is presented in Figure 2-5. It is evident that our system can provide a wider range of electric fields compared to cuvettes. Notably, when the electric field exceeds 15 kV/cm, high current in the channels leads to electric arcing, resulting in only a few channels within the same column yielding colonies. It is anticipated that by reducing the conductivity of the cell sample through buffer medium modifications or lowering the sample concentration, the upper bound of the working window, representing the electric field range where no electric arcing occurs, could be increased.

We also conducted experiments with similar setup to showcase the system's capa-

bility of screening flow rates under a fixed electric field of 10 kV/cm as presented in Figure 2-6. While our system can provide a wide range of flow rates, when the flow rate is low, e.g., 25 μL/sec, arcing will occur due to the formation of bubbles induced by Joule heat accumulation in the channel.

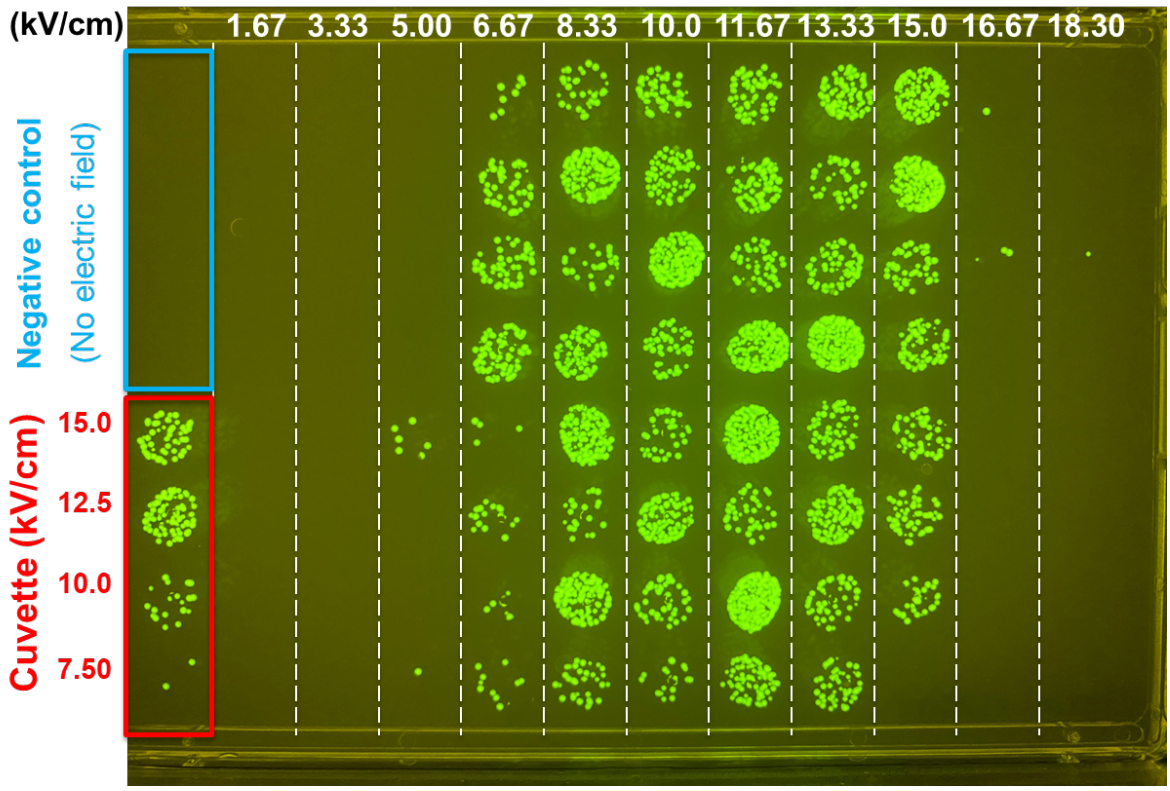


Figure 2-4: **Screening electroporation conditions for *Escherichia coli* NEB10β.** Colony formation on LB agar plate under 1000-fold dilution transformation efficiency, showcasing the impact different electric field strength under 400 μL/sec.

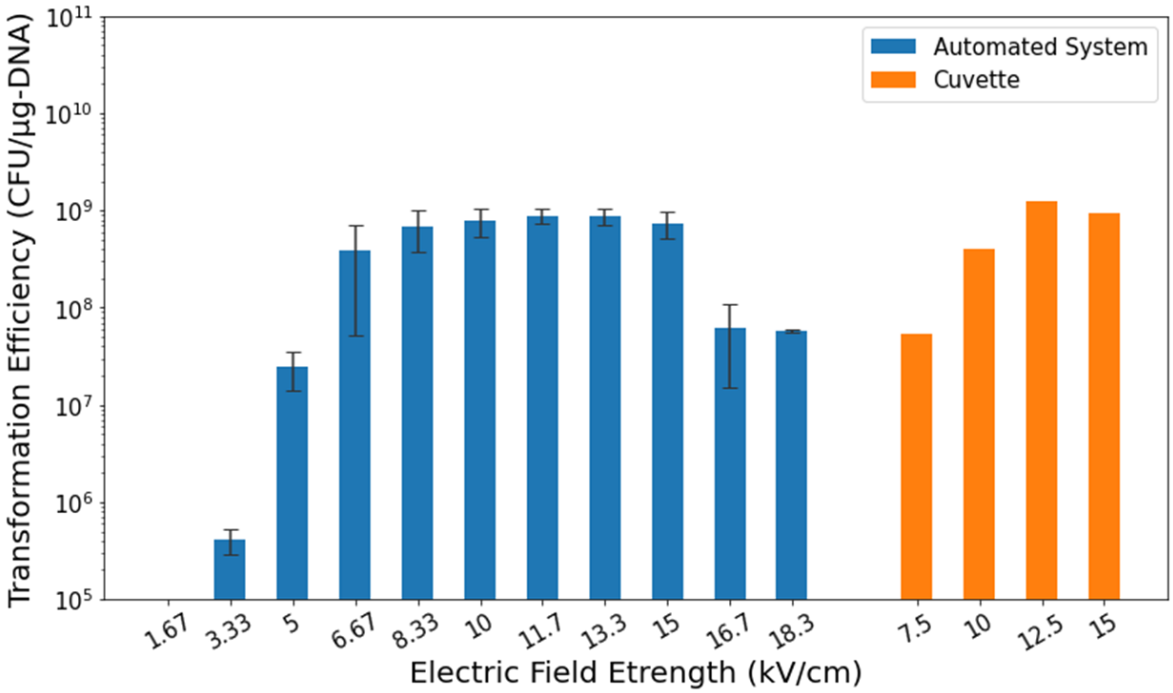


Figure 2-5: **Transformation efficiency under different electric field.** Transformation efficiencies for each condition, indicating the performance and effectiveness of the system in varying electric field range. Data represent the average of eight technical replicates in a column (n = 8) and error bars represent 1 standard deviation.

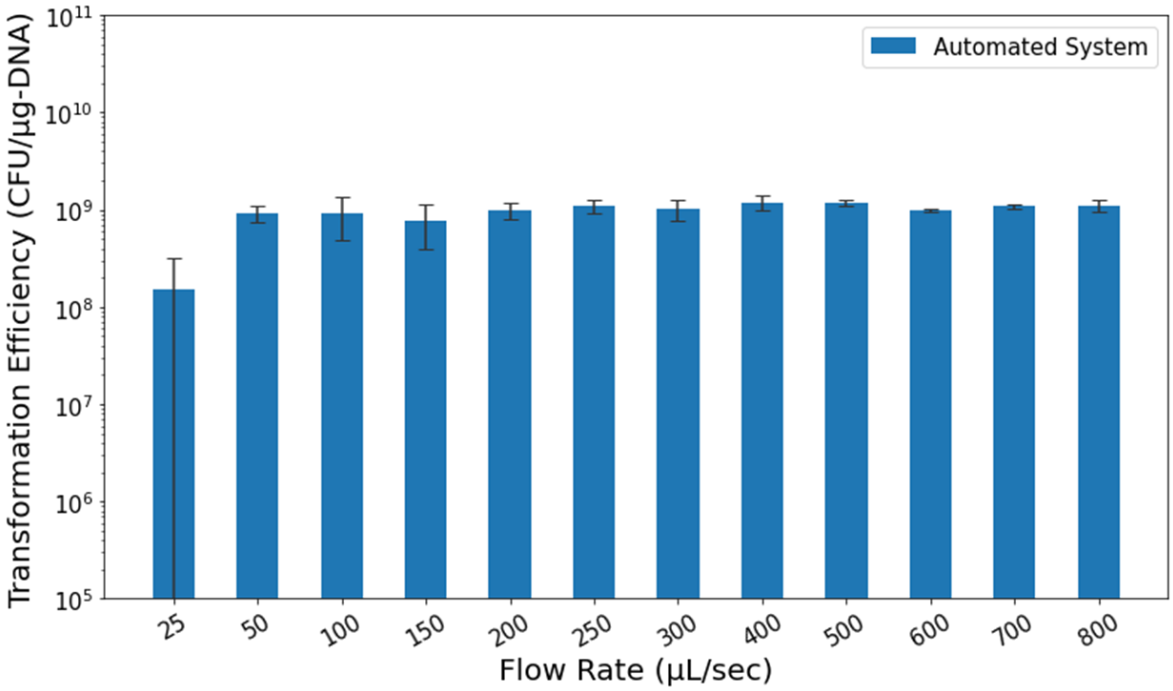


Figure 2-6: **Transformation efficiency under different flow rates.** Transformation efficiencies for each condition, indicating the performance and effectiveness of the system in varying flow rate range. Data represent the average of eight technical replicates in a column (n = 8) and error bars represent 1 standard deviation.

2.4.3 Evaluating Electroporation Consistency

To assess the electroporation consistency of the automated system, we conducted multiple electroporations using three different combinations of electroporation conditions: 8.33 kV/cm & 200 μL/sec, 8.33 kV/cm & 400 μL/sec, and 5 kV/cm & 400 μL/sec. Each combination was tested in a single experiment.

Figure 2-7 presents the transformation efficiencies for all three combinations. Despite occasional channel clogging or malfunctions (generally affecting 1-2 out of 96 channels), the overall efficiency and consistency across channels and columns are noteworthy, indicating the reliability and reproducibility of our device. Furthermore, the device exhibits sensitivity to varying conditions, resulting in different levels of transformation efficiencies with uniformly distributed performance. This feature enhances the adaptability and versatility of our automated system for a wide range of applications.

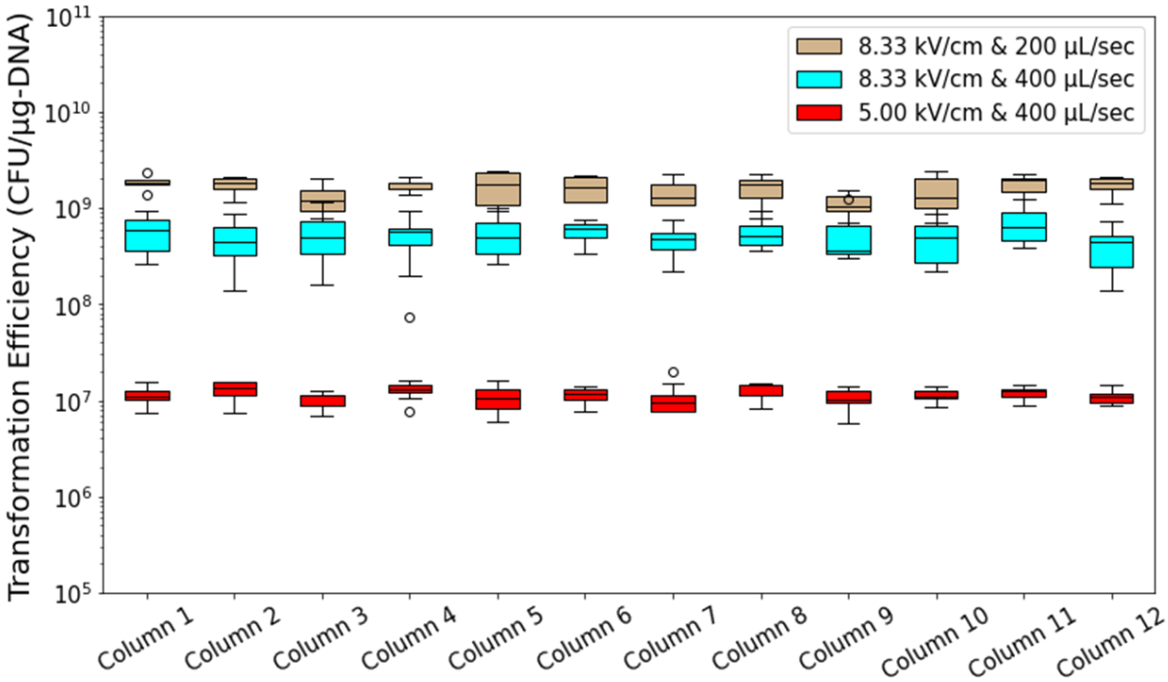


Figure 2-7: **Evaluating electroporation consistency for *Escherichia coli* NEB10β.** Transformation efficiencies for different combinations of electroporation conditions (8.33 kV/cm & 200 μL/sec, 8.33 kV/cm & 400 μL/sec, and 5 kV/cm & 400 μL/sec), demonstrating the system's reliability and reproducibility in producing consistent results. Data represent the average of eight technical replicates in a column (n = 8) and dots represent the outliers of boxplot.

2.5 Conclusions

In this chapter, we have presented an automated system specifically designed for high-throughput bacterial genetic engineering. Our microfluidic device, serving as the core component, offers notable advantages over traditional cuvettes, including enhanced efficiency, improved throughput, and consistent results. The overall design and architecture of the automated system were discussed, emphasizing its ability to streamline the electroporation process and revolutionize bacterial genetic engineering workflows.

Through extensive experimental evaluations, we have demonstrated the exceptional performance and reliability of our automated system. The results obtained validate its high efficiency, transformation efficiency, and reproducibility, affirming the effectiveness

of our approach. By addressing the limitations of cuvette-based methods and introducing automation and parallelization, our system provides researchers with a reliable, scalable, and high-throughput solution for bacterial genetic engineering.

In conclusion, our automated system marks a significant advancement in the field of high-throughput bacterial genetic engineering. Its innovative design, combined with its robust performance and scalability, empowers researchers to overcome traditional limitations and expedite genetic manipulation in bacteria. With the ability to achieve reliable and reproducible results, our system opens new avenues for strain engineering, functional genomics, and synthetic biology applications.

Chapter 3

MCount: An automated colony counting tool for high throughput microbiology

3.1 Introduction

Quantitative assessment of microorganisms is a critical procedure in the field of microbiology, and various methods have been developed to estimate microorganism levels, including quantitative PCR [78], flow cytometry [79], spectrophotometry [80, 81], and counting colony forming units (CFU) [82, 83]. Among these methods, CFU counting is the oldest and most widely used method, and its efficacy and reliability have been examined since the 1880s [83, 84]. In addition, its simple operational protocol and minimal consumables cost have made it a long-established gold standard.

However, traditional manual CFU counting is a time-consuming and labor-intensive process when dealing with many images (e.g. more than one hundred). Thus, researchers have devoted considerable effort to developing numerous solutions over the decades to realize an easier, faster, more accurate, and reliable counting method. Currently, popular solutions include commercial products such as SphereFlash (IUL Instru-

ments) and ProtoCOL 3 (Synbiosis) and open-sourced tools based on various counting algorithms [85–92]. While commercial products are usually easy to operate, they are expensive and highly specialized for specific counting scenarios, and their proprietary programming nature makes them difficult to modify or share. In contrast, open-source tools are gaining more interest, especially those capable of batch processin11[86–88]. NIST's Integrated Colony Enumerator (NICE) is a tool that has been popular since 2009, which is based on the combination of extended minima function and thresholding algorithms [86]. It has a relatively short image processing time ($<$ 5 seconds per image) and a friendly user interface. OpenCFU, publicized in 2013, is a tool claimed to work faster, more accurate, and more robust than NICE [87]. It adopts the watershed algorithm along with a series of pre- and post-processing filters. AutoCellSeg was developed in 2017 [88] to reduce the hyperparameter selection effort during counting, based on a feedback-based watershed algorithm, and it has an interactive graphical user interface that is appealing to people not familiar with programming.

The existing colony counting solutions work well on single petri dish plates of 60 cm^2, where 25 $\sim$ 250 colonies are randomly distributed, making colony merging issues rare [84, 93, 94]. Most colony plating protocols require careful selection of the serial dilution ratio to minimize the chance of colony merging. However, modern colony counting requires higher throughput, and a typical example is plating 96 samples on a single-well rectangular plate of 109 cm^2 (Figure 3-1a). In this case, more than 10 colonies are plated on an area less than 1 cm^2 for each sample, resulting in $>$10 times higher density, which often leads to colony merging. Furthermore, in this high-throughput workflows, the smaller area means fewer image pixels per sample, resulting in lower image quality. These difficulties lead to significant counting underestimation of current solutions (Figure 3-1c). Although NICE can recognize most single colonies, it tends to count merged colonies as one. AutoCellSeg always counts each continuous region as one, regardless of the region's shape, so it cannot handle scenarios involving colony merging. As for OpenCFU, it fails to recognize most colonies, including single colonies, due to the lower image quality for each sample of the rectangular plate. The failure of

these solutions results from the fact that they rely on region-based algorithms and do not take contour information into account.

Colony counting is a sub-domain of object counting that includes cell counting [95], pollen [96, 97], and bubble [98, 99], and object counting approaches can be categorized into contour-based or region-based methods. Contour-based methods aim to recognize the shape only from the contour pixels and use methods such as the Hough-transform [100, 101] and least-squared circle [98, 100, 102] to fit the contour. They also involve algorithms such as concave point detection methods to split the whole contour into segments [99, 102–105]. Contour-based methods can provide precise results given a high-definition image, but they typically require more computational resources. Region-based methods take all shape pixels into account, and classical methods include the extended minima function method [86, 106], morphological operations [107], distance transformation [103, 108], and the watershed algorithm [87, 88, 96, 106]. Regional-based methods can tolerate more noise and have faster processing speed, but their recognition accuracy is usually lower. To combine the advantages of both methods, researchers have made efforts to pair the extended minima with contour segments [98, 99, 103, 104]. In recent studies, machine learning approaches have been adopted [103, 105], eliminating the computationally expensive exhausted pairing [98, 99]. Surprisingly, although algorithms are advancing rapidly in other sub-domains of object counting, colony counting is still lagging behind with reliance on region-based algorithms, leading to poor accuracy in high-throughput workflows.

To ensure user-friendliness and operational simplicity, counting tools typically emphasize a few intuitive hyperparameters. However, while these tools focus more on the physical meaning of different hyperparameters, they often lack a discussion on properly tuning these hyperparameters, i.e., the hyperparameter optimization problem, which is an essential topic in the field of machine learning to ensure high quality performance with minimal human effort [109, 110]. Consequently, deploying such solutions to various counting tasks, even with minor differences, can be challenging. Therefore, there is a need for not only a better colony counting algorithm but also a consistent method

for hyperparameter optimization.

In this chapter, we propose a new solution, MCount (Merged-colony Counting), that can precisely infer the number of merged colonies from an image of a relatively small size, meeting the demands of high-throughput colony counting. MCount employs both region-based and contour-based algorithms, which leads to much higher accuracy than existing tools. Since there is no standard benchmark for high-throughput colony counting tasks, we construct a dataset for performance evaluation. In addition, we address the hyperparameter optimization problem. Given that the number of labeled data is often insufficient, we propose novel statistical methods for hyperparameter optimization, ensuring a low error rate even with a low number of labeled data or even unlabeled data. Finally, we examine the statistical robustness of the proposed hyperparameter-tuning methods.

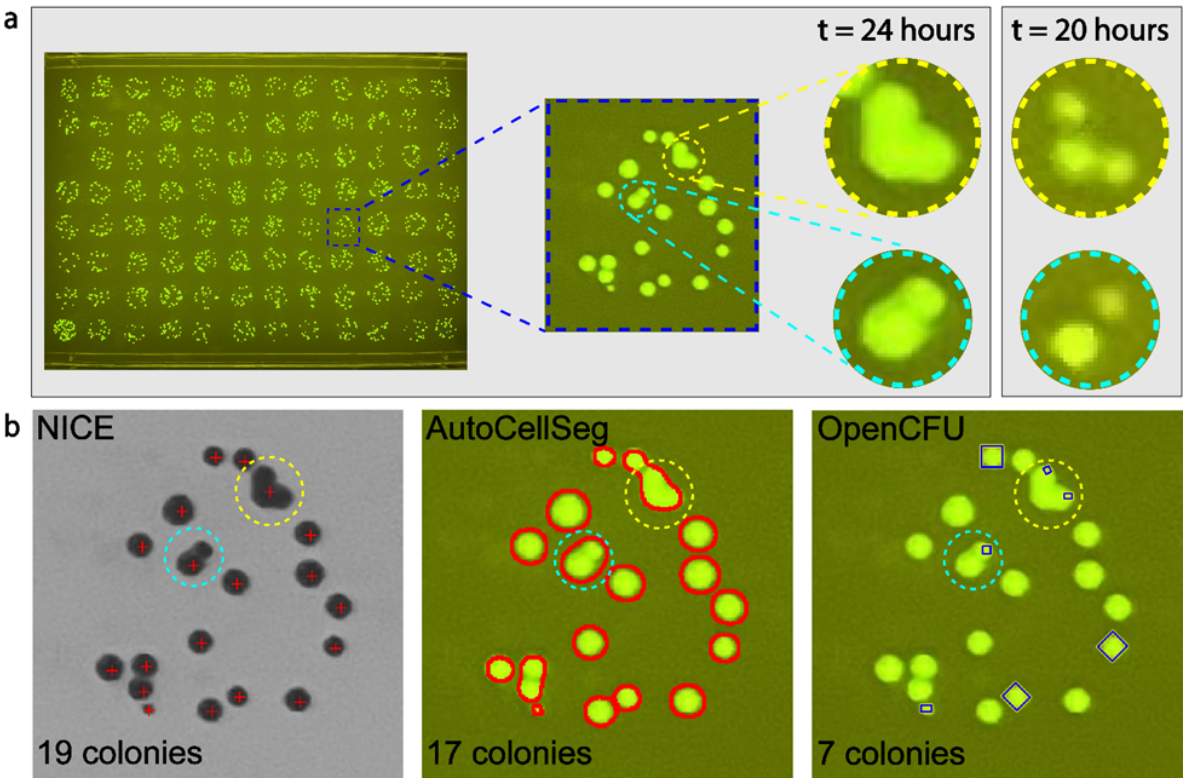


Figure 3-1: **The need for a more powerful colony counting algorithm.** (a) Colonies of fluorescent *E. coli* NEB10β are plated on agar in an 8×12 array format. As the area of each well is small, colony merging occurs frequently and the sub-image in the blue rectangle is an example. While it is challenging to determine the exact number of merged colonies in the yellow and cyan circles, a skilled person can count 3 and 2 colonies, respectively, which can be verified by the photograph taken a few hours earlier. (b) NICE correctly counts all single colonies but counts all merged colonies as one [86]; AutoCellSeg uses the watershed algorithm to count every connected region as one colony, regardless of the region's shape [88]; OpenCFU applies sophisticated rules to count only perfect circle-shaped colonies but fails to recognize most single colonies [87]. All three algorithms underestimate the number of merged colonies, which is 22.

3.2 Methods

Figure 3-2 presents a detailed flowchart of MCount, depicting the process of foreground extraction, contour extraction, regional circle fitting, and optimization.

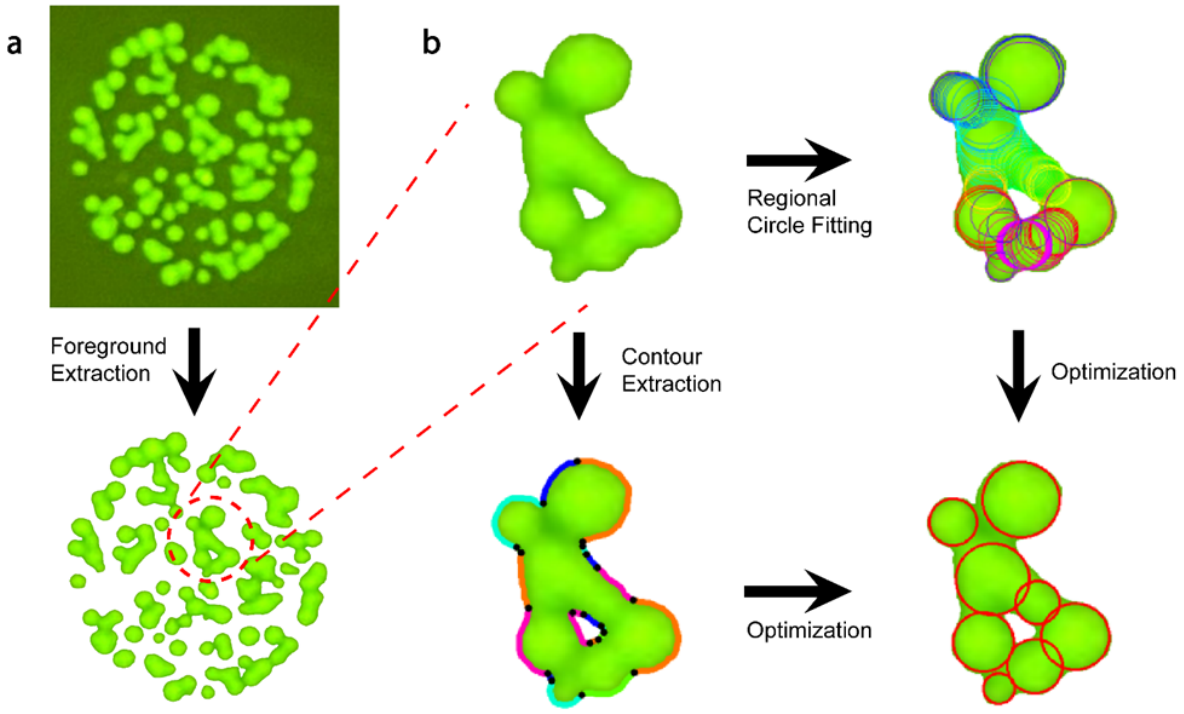


Figure 3-2: **MCount uses a combination of contour-based and region-based information to accurately count merged colonies.** (a) The foreground and background are separated using Otsu thresholding [111], and colonies are segmented into disconnected segments, which are then smoothed to remove noise. (b) Contour-based algorithms are applied to each colony segment to identify turning points (colored in black), which divide the inner and outer contours into contour pieces represented by different colors. Meanwhile, region-based algorithms generate candidate circles denoted in different colors that can fit the boundary well. Finally, proper circles are selected from candidate circles based on the shape of all contour pieces. This selection process is converted into an optimization problem that is solved to correctly recognize merged colonies.

3.2.1 Algorithms adopted by MCount

3.2.1.1 Foreground extraction

The first step is to obtain connected (overlapped) colony segments by using a series of filters. The original image is binarized using Otsu thresholding, which separates image pixels into two categories, i.e., foreground and background, by maximizing inter-class

variance [111]. Next, the connected component labeling algorithm is used to separate disconnected segments. Finally, morphological operations (erosion operation followed by a dilation operation) are applied to each colony segment twice to remove white noise in the image.

3.2.1.2 Contour extraction

The next step is to extract contours from each colony segment using a classic border following algorithm [112]. As each contour may correspond to the edges of several colonies, we need to split each contour into several pieces, so that each piece only corresponds to one colony. We assume that any overlap between multiple colonies results in a shape with concave edge points that correspond to the intersections of the colony boundaries, which allows us to use the concave points to split the contour. To identify the concave points, we use an algorithm based on the polygon approximation algorithm [99] and the algorithm proposed by Zafari *et al.* [113], and the pseudo-code is described in Figure 3-3.

Algorithm 1: Extracting concave points of a contour

Input: C: list of contour pixel points; n points in total
d: adjustable parameter that defines the fineness of contour

Output: T: list of turning pixel points; m points in total
$Conc$: list of concave pixel points

1 $T.append(C_0)$
2 $i \leftarrow 0, \quad k \leftarrow 2$
3 **while** $k < n$ **do**
4 $\quad d_{max} \leftarrow -inf$
5 $\quad$ **for** $j \leftarrow i+1$ **to** k **do**
6 $\quad\quad dist \leftarrow$ the distance from point C_j to line $\overline{C_iC_k}$
7 $\quad\quad$ **if** $dist > d_{max}$ **then**
8 $\quad\quad\quad d_{max} \leftarrow dist$
9 $\quad\quad\quad ind \leftarrow j$
10 $\quad\quad$ **end**
11 $\quad$ **end**
12 $\quad$ **if** $dist > d$ **then**
13 $\quad\quad T.append(C_{ind})$
14 $\quad\quad i \leftarrow ind, \quad k \leftarrow ind + 2$
15 $\quad$ **else**
16 $\quad\quad k \leftarrow k + 1$
17 $\quad$ **end**
18 **end**
19 $Conc.append(T_0)$
20 **for** $i \leftarrow 1$ **to** $m - 1$ **do**
21 $\quad$ **if** $Angle(\overrightarrow{C_{i-1}C_i}, \overrightarrow{C_iC_{i+1}}) < 0$ **then**
22 $\quad\quad Conc.append(T_i)$
23 $\quad$ **end**
24 **end**

Figure 3-3: **Contour segmentation based on concave point detection.** Given a contour C with a hyperparameter d that determines contour fineness, this algorithm first approximates C using polynomial fitting and selects turning points T from the polygon approximation. Then, it identifies concave points $Conc$ from the turning points T to split the contour into several pieces.

In brief, as shown in lines 1-18, the contour is approximated using polynomial fitting, where all the vertices of the polygon are on the contour. Such vertices are called turning points, denoted as T, and hyperparameter d controls how densely the turning points are selected from the contour, as shown in Figure 3-4. For example, when d increases, fewer turning points are selected. Next, all concave vertices of the polygon are further

selected from the turning points by implementing code from lines 19-24. These concave points divide the contour into several pieces, and each piece is denoted by a different color in Figure 3-2b.

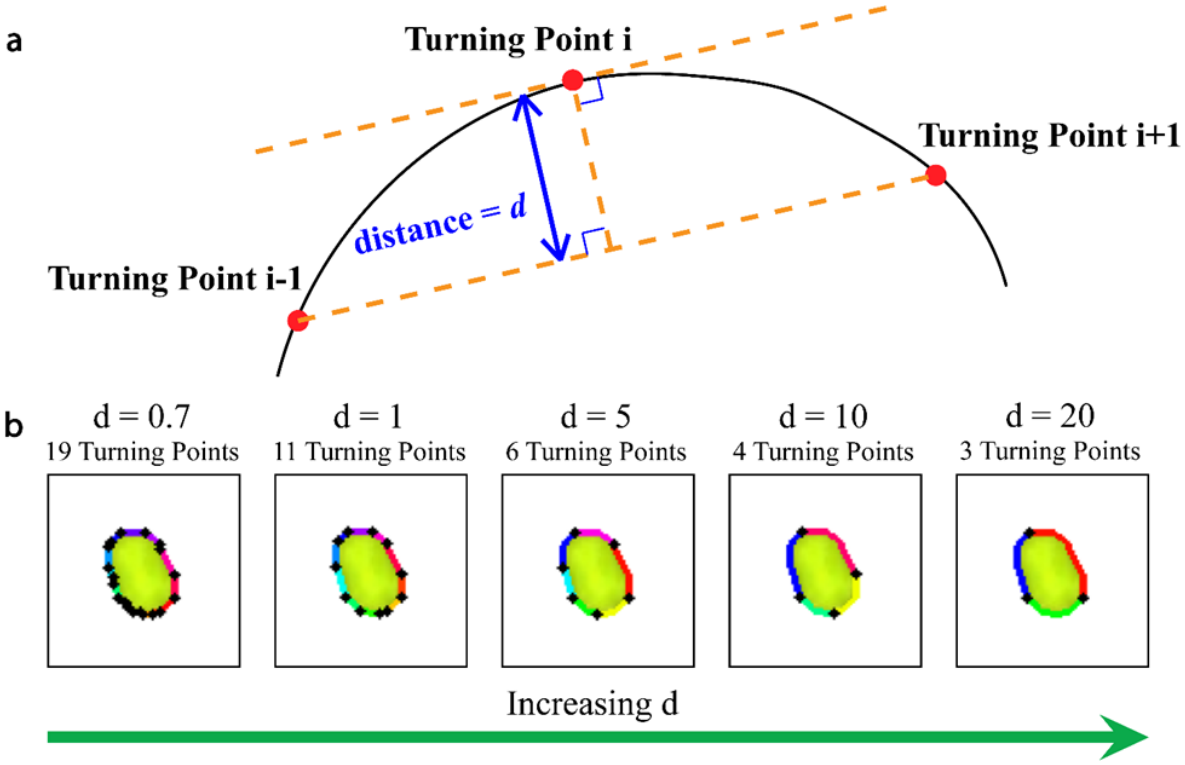


Figure 3-4: **The Polygon Approximation Algorithm uses turning points on the contour to represent an inscribed polygon, where d controls how many turning points are generated.** (a) For consecutive three turning points T_{i-1}, T_i, and T_{i+1}, d represents the distance from T_i to the line $\overline{T_{i-1}T_{i+1}}$. (b) The larger the value of d, the more turning points are generated to represent the contour. The turning points are represented as black dots, while contour pieces divided by turning points are shown in different colors.

3.2.1.3 Regional circle fitting

In this step, a set of candidate circles is generated from each colony segment. Two types of circles are considered. The first type of circle is identified using a region-based algorithm, where distance transformation is applied to the binary foreground image so that the center of circles is the local maxima of distance [114]. The other type of circle is obtained by a least square fit of each split contour piece [99]. In Figure 3-2b, all the

candidate circles are denoted in different colors.

3.2.1.4 Optimization

The final step of the algorithm is to pair the split contour pieces with candidate circles, so that each contour will only match one circle. Denoting C_i as the ith contour piece and O_j as the jth circle (the edge of the circle rather than the center), the pairing problem that we are solving is:

$$X_{ij} = \begin{cases} 1 & \textit{if allocating } C_i \textit{ to } O_j \\ 0 & \textit{otherwise} \end{cases} \tag{3.1}$$

$$Y_j = \begin{cases} 1 & \textit{if any contour is allocated to } O_j \\ 0 & \textit{otherwise} \end{cases} \tag{3.2}$$

O_j is a recognized as a colony only when $Y_j = 1$. To formulate this pairing problem, we define the distance D_{ij} of allocating O_j to C_i as:

$$D_{ij} = \sum_p \min_q \overline{pq}, \quad p \in C_i, q \in O_j \tag{3.3}$$

where p and q are the pixels of contour piece and circle, respectively, and $\overline{pq}$ represents the distance between pixel p and q.

The smaller D_{ij}, the better the pairing of circle O_j with contour piece C_j. Thus, an intuitive way to pair all contour pieces with proper circles is to match every contour piece to the closest circle, i.e., we are looking for a pairing that could lead to a minimized total distance, which could be converted into an optimization problem. However, simply minimizing the total distance without any constraints will lead to overfitting because different contour pieces from the same colony might be fitted to multiple adjacent circles. Therefore, some constraints must be considered, and an example is to control the total number of recognized circles, $\sum_j Y_j$, which can be converted to a zero-one

integer programming problem with objective function $L(X_{ij}, Y_i)$:

$$
\begin{aligned}
(X_{ij}, Y_i) = \underset{X_{ij}, Y_j}{argmin}\, L(X_{ij}, Y_i) = \underset{X_{ij}, Y_j}{argmin} \sum_{i,j} D_{ij} X_{ij} + \lambda \sum_{j} Y_j \\
s.t. \quad \textstyle\sum_j X_{ij} = 1 \\
0 \leq X_{ij} \leq 1 \\
Y_j \leq \textstyle\sum_i X_{ij} \leq Y_j \cdot M \\
0 \leq Y_j \leq 1
\end{aligned}
\tag{3.4}
$$

where λ denotes the strength of the constraint and M denotes the total number of contour pieces. By solving Equation (3.4), we can obtain all the recognized colonies O_j with $Y_j = 1$.

3.2.1.5 Reducing computational complexity of the optimization step

We recognize that the most time-consuming part during the implementation of MCount is the optimization step because the zero-one integer programming problem is an NP-complete problem whose worst-case runtime grows exponentially. To mitigate this issue, the segment separation procedure is introduced in the foreground extraction step, which can dramatically reduce the computational complexity of the optimization step by reducing the total number of contour pieces to a few hundred.

In addition, as there are a bunch of approximation algorithms available [115, 116], which are already integrated into widely adopted linear programming modelers, e.g., PuLP in Python, we choose one (PuLP solver "COIN_CMD") that yields a high recognition accuracy with relatively short computation time. The actual runtime of MCount is discussed in the Results and Discussion section.

3.2.1.6 Code availability

The source code is included in Appendix B.

3.2.2 Benchmark for performance evaluation

Next we sought to evaluate the performance of MCount by using a benchmark dataset. As there is no standard benchmark available, existing colony counting tasks have relied on their own datasets [85–90]. However, we find that the datasets exhibited at least one of the following concerns when applied to MCount: (a) insufficient number of labeled images (less than 100 or ∼20); (b) too many colonies in each labeled image (more than 50), so that counting number may no longer be the best criteria to evaluate recognition performance; (c) incorrect ground-truth labels, especially on highly merged colonies.

To address these issues, we have created a benchmark dataset colonies of GFP fluorescent *E. coli* DH10-beta, a model organism designed for creating a mutant library suitable for high-throughput purpose (Figure 3-5). The dataset includes 960 labeled images that are large enough to address the problem of an insufficiently large data set. Additionally, we increased the size of our training set by decomposing these images into 15,847 colony segments using foreground extraction. As most colony segments have less than 5 colonies, which also helps addressing problem (b). To address problem (c), we carefully labeled historical images in which the merged colonies were separated. The cell preparation procedure and more detailed discussions on problem (b) and (c) are included in the following section.

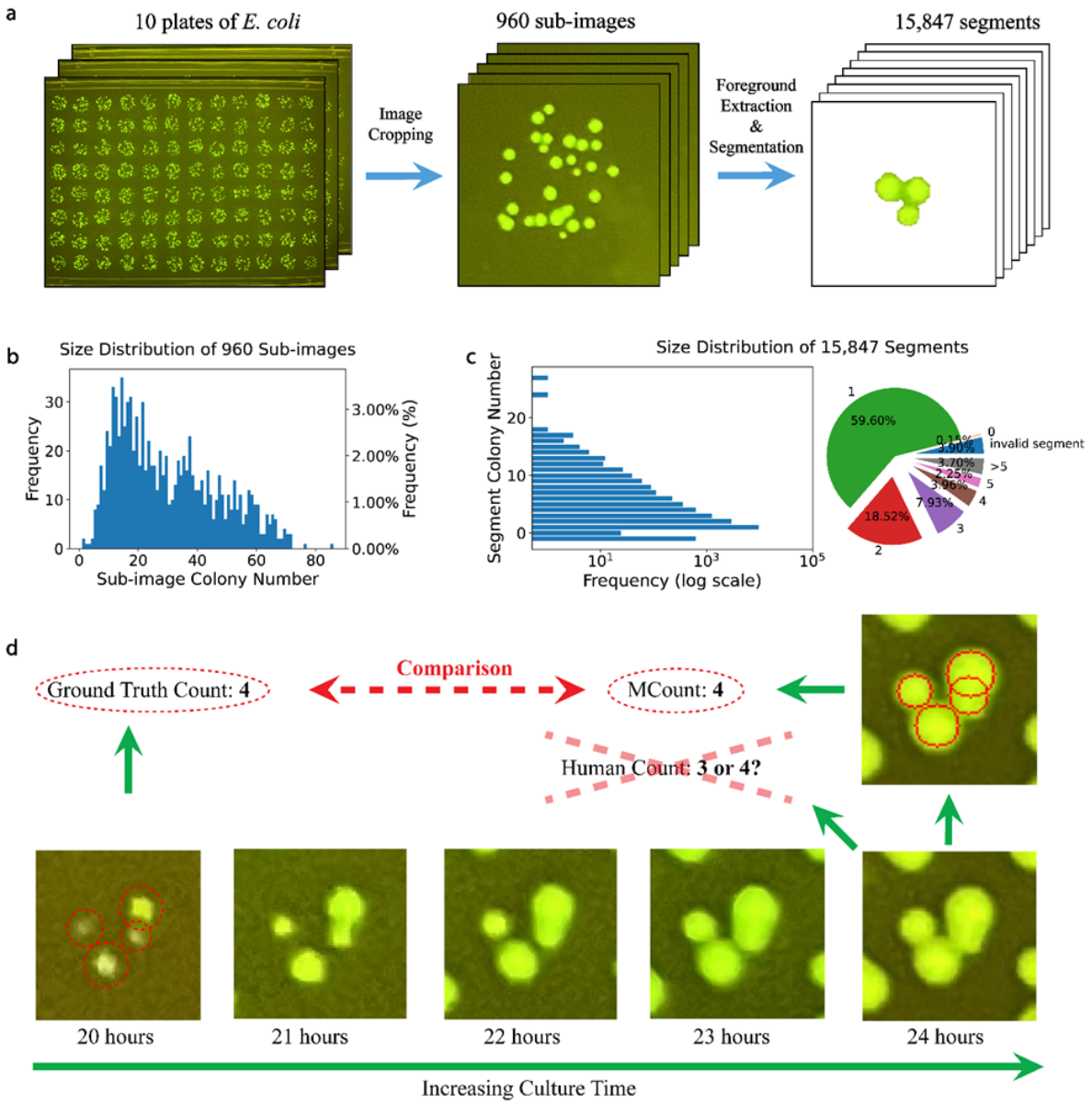


Figure 3-5: **Contour segmentation based on concave point detection.** Given a contour C with a hyperparameter d that determines contour fineness, this algorithm first approximates C using polynomial fitting and selects turning points T from the polygon approximation. Then, it identifies concave points $Conc$ from the turning points T to split the contour into several pieces.

3.2.2.1 Cultivation and preparation of *E. coli* dataset

The first step involved transforming *E. coli* NEB10β with a GFP fluorescent plasmid via electroporation. A glycerol stock was used to inoculate 5 mL of LB medium in a

14-mL culture tube, which was incubated overnight at 37°C and 250 rpm. The following morning, 800 μL of the cell culture was transferred to 80 mL of LB medium and grown at 37°C and 250 rpm until reaching the exponential phase, which was determined by measuring the OD600 using a UV spectrophotometer (UV-1800, Shimadzu). The cell culture was then centrifuged at 3500 rpm at 4°C (F0650 rotor, Allegra 64R Benchtop Centrifuge, Beckman Coulter), and the supernatant was discarded. The cell pellets were washed with 2 mL of pre-chilled 10% glycerol at 4°C, and the process was repeated after two more centrifugations at 8000 rpm at 4°C (F1202 rotor, Allegra 64R Benchtop Centrifuge). After the final centrifugation, the cell pellets were suspended in 10% glycerol, and the volume was adjusted so that the final cell concentration was about OD600 of 10. Ampicillin resistance and GFP encoding DNA plasmids (Parts Registry K176011) were mixed with the cell sample to obtain a final concentration of 0.1 ng/μL. The sample was loaded into 2 mm VWR electroporation cuvettes, and electroporated using a MicroPulser™ (Bio-Rad) at 2.5 kV with a 6 ms time constant. The electroporated sample was immediately suspended in 900 μL of pre-warmed LB medium at 37°C and 250 rpm for one hour. Finally, 100 μL of the cell suspension was pipetted onto LB agar plates containing ampicillin (50 μg/mL), which were then incubated overnight at 37°C and 250 rpm.

In the next step, the GFP fluorescent E. coli was plated in a 96-well format. An individual colony was picked from the plate made in the previous step and inoculated in 5 mL of LB medium in a 14-mL culture tube. The culture was incubated overnight at 37°C and 250 rpm, and 200 μL of the resulting cell culture was transferred to 20 mL of LB medium and grown at 37°C and 250 rpm until the OD600 reached 0.5. The cell sample was then serially diluted using LB medium to obtain a dilution of about $1 \times 10^7 \sim 8 \times 10^7$. A liquid handling robot (Perkin Elmer JANUS G3 BioTx Pro Plus) was used to dispense 5 μL of the sample into each well of an agar plate in an 8×12 format. The dilution ratio could be adjusted to obtain different colony plating densities. The plate was then incubated at 37°C for 12 hours. Photos were taken once an hour using an iPhone 12 Pro until the colonies were overgrown, which was about 25 hours.

3.2.2.2 Discussion on *E. coli* dataset

In this study, we used GFP fluorescent *E. coli* strain NEB10β (New England Biolabs) to create the dataset. We chose to use the GFP fluorescent strain because accurately labeling nonfluorescent plates using the human eye is challenging, even though MCount can accurately recognize nonfluorescent colonies. We also want to address problems (b) and (c) in this section.

In cases where a labeled image of a dataset has too many colonies (more than 150), an algorithm may struggle to accurately identify all colonies even if it provides the same counting number as the label. This can occur due to under-counting issues such as colony merging or over-counting issues like incorrectly identifying image flaws or agar defects as colonies. To address this issue, each sub-image was further decomposed into individual colony segments using foreground extraction, resulting in a dataset of 15,847 segments, as shown in Figure 3-5a. Figure 3-5b and Figure 3-5c shows the colony number distribution of sub-images and segments, respectively, where most sub-images have 10 60 colonies, and most segments only have few colonies that are less than five. While high accuracy on sub-images does not necessarily guarantee high accuracy on segments, a good performance on segments usually implies an accurately counting on sub-images. Furthermore, as 59.6% segments are single-colony and about 40% segments have more than two colonies, a good algorithm should be capable of accurately recognizing both single colonies and overlapped colonies. Thus, evaluating MCount on segments rather than sub-images is a more reasonable approach.

Problem (c) arises from the difficulty of correctly labeling certain images, such as those with highly overlapped colonies, even for humans. While some datasets simply discard these difficult images, others may take the average of the labeling numbers provided by multiple individuals. However, the former approach leads to a dataset with selection bias, and the latter approach cannot guarantee that the average number is the ground truth. To address this issue, we took pictures of the colonies every hour and labeled every colony segment according to its historical image, as shown in Figure

3-5d. This method enables the evaluation of MCount's ability to infer the number of colonies, even when the task is challenging for humans. Furthermore, this approach allows the evaluation of MCount's performance on images with a higher degree of colony merging. An example is shown in Figure 3-5d, where the ground truth colony count is 4, based on the image taken when the colonies had been growing for 20 hours. Looking at the image taken when the colonies had been growing for 24 hours, it is difficult to determine whether there are 3 or 4 colonies, but MCount can correctly recognize the count as 4

3.3 Results and Discussion

3.3.1 MCount only has two adjustable hyperparameters

Although providing more hyperparameters can increase the flexibility of algorithm tuning, which often improves recognition accuracy, however, to prevent overfitting as well as to ensure the user-friendliness, we aim to offer minimized number of adjustable hyperparameters by choosing the most important ones. In the case of MCount, we have identified two hyperparameters that significantly influence its accuracy: d and λ, which control the contour extraction and optimization steps, respectively.

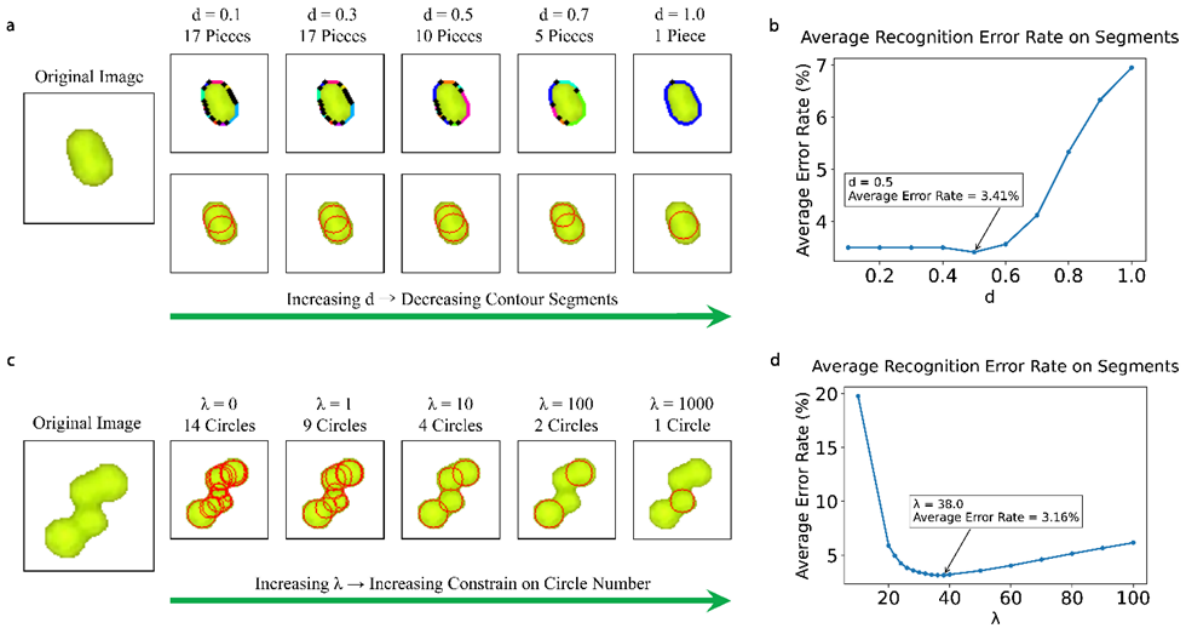


Figure 3-6: **MCount has two adjustable hyperparameters, d and λ, which control the contour fineness and constrain the circle number, respectively.** (a) The number of contour pieces decreases as d increases, but excessively large d may lead to a failure to recognize merged colonies. The original image has two colonies. (b) The average error rate on the segment dataset versus d at $\lambda = 26$, where the minimum average error rate of 3.41% is achieved when $d = 0.5$. (c) A larger λ imposes a stronger constraint on the circle number. By tuning the λ value within the proper range, e.g., $\lambda = 10$, MCount can correctly recognize merged colonies. The original image has four colonies. (d) The average error rate on the segment dataset versus λ when $d = 0.5$, where the minimum average error rate of 3.16% is achieved at $\lambda = 38$.

Hyperparameter d represents the fineness of the polygon approximation fitting of the contour. When d is small, the contour will have more turning points, resulting in more contour pieces being generated, as shown in Figure 3-6a. However, setting too small d should be avoided, because it will result in an unnecessarily large number of contour pieces and candidate circles (each contour piece will generate one least-squared circle), leading to a longer runtime in the optimization step. It also reduces the algorithm's robustness, because each contour piece will be too short to accurately calculate the curvature. Conversely, if d is too large, very few turning points will be used to fit the contour, leading to an underestimation of the colony count. To minimize the average recognition error on segments, we tried with different values of d on 15,847

colony segments, as shown in Figure 3-6b, and found that the best value for d is 0.5 (when $\lambda = 26$). The recognition error on a segment is defined as the difference between the counted number and the label number divided by the label number. Another way to define recognition error is to set 100% for a segment whenever the counted number differs from the label number. The plot for such a recognition error on segments is shown in Figure 3-7a.

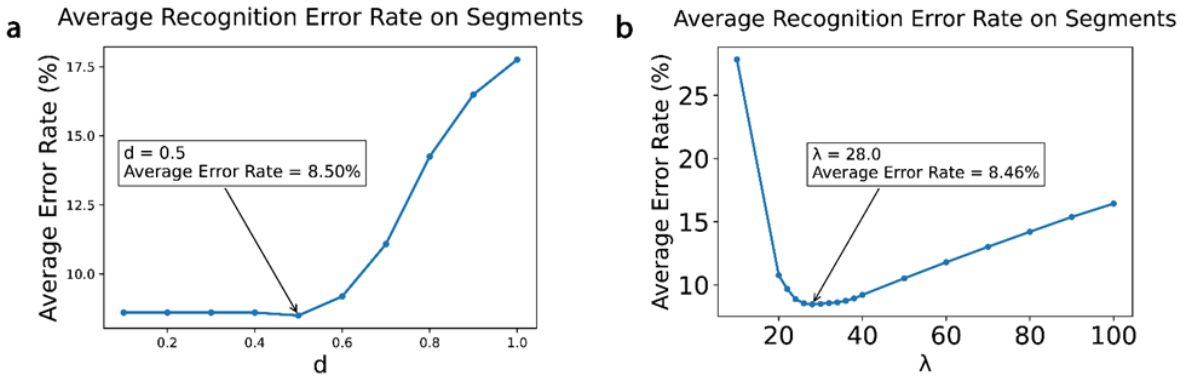


Figure 3-7: **The optimal values of d and λ are determined using a different definition of recognition error, where the error of each segment is defined as $\mathbb{1}(OLCount \neq Label)$.** (a) The average error rate on the segment dataset is plotted against d when $\lambda = 26$. The minimum average error rate of 8.50% is achieved at $d = 0.5$. (b) The average error rate on the segment dataset is plotted against λ when $d = 0.5$. The minimum average error rate of 8.46% is achieved at $\lambda = 28$.

The other important hyperparameter is λ, which is defined in Equation (4) and reflects the trade-off between colony number constraints and pairing of contour pieces to candidate circles. As shown in Figure 3-6c, when $\lambda = 0$, there is no constraint on colony number, and almost every contour piece matches to a unique candidate circle, even if some pieces come from the same colony. When λ increases, the counted number monotonically increases, and $\lambda = 10$ gives a reasonable recognition result. When λ is very large, such as $\lambda = 1000$, the constraint is so strong that only one circle is drawn. Through testing on colony segments, we identified that the best value for λ to minimize the average recognition error on segments is 38 (when $d = 0.5$). Figure 3-7b shows the optimal values for λ using the alternative definition for recognition error.

Both hyperparameters d and λ are critical for achieving accurate colony counting results with MCount. However, their optimum values are dependent on various factors such as image resolution and colony size, and there is no universal set of values that will work for all scenarios. The standard approach for determining the optimal hyperparameter values is to optimize metrics such as minimizing the average recognition error through grid search and k-fold cross-validation, as demonstrated in the following section.

3.3.2 The performance of MCount

To determine the appropriate values for the hyperparameters (d, λ) in a systematic manner, we utilized grid search and 10-fold cross-validation to minimize the average error rate on 15,847 colony segments and 960 sub-images. The optimal hyperparameter values were found to be $(d, \lambda) = (0.5, 38)$ and $(d, \lambda) = (0.5, 26)$, respectively. In this section, as we focus on the performance on sub-images, we will assess the performance of MCount using $(d, \lambda) = (0.5, 26)$.

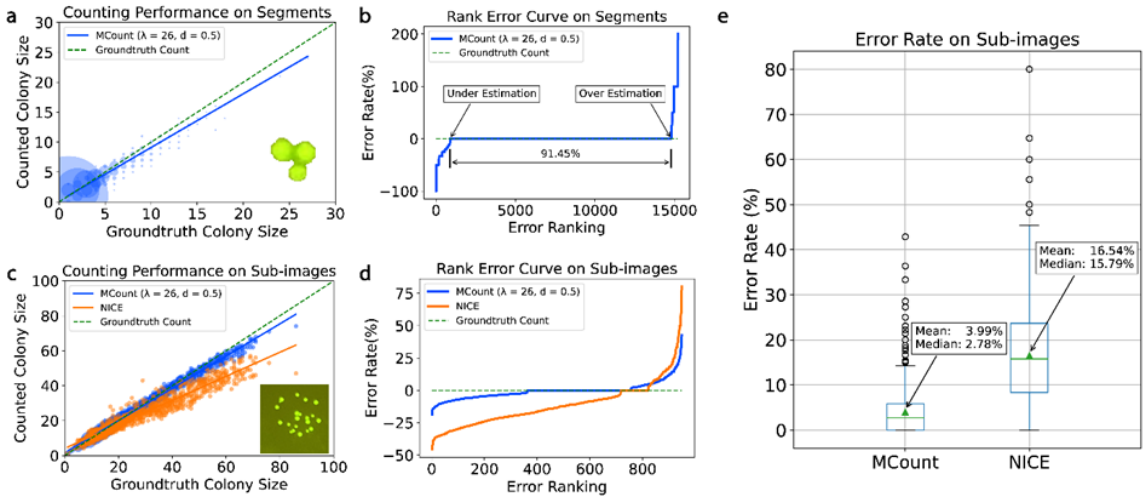


Figure 3-8: **Comparison of MCount** ($\lambda = 26, d = 0.5$) **and NICE performance in recognizing colony segments and sub-images.** (a) Bubble plot of MCount counting result versus ground truth label on 15,847 colony segments. The size of each bubble at location (x, y) is proportional to the number of segments that have x merged colonies but are recognized as y colonies by MCount. The blue line represents the regression line for MCount, while the green dashed line represents the regression line for a perfect algorithm that always gives the result as the label. MCount exhibits slight underestimation as the blue line is lower than the dashed green line. (b) Rank error curve of error rate values for 15,847 segments. MCount recognizes 91.45% of segments with zero error rate. (c) Bubble plot of MCount and NICE counting results versus ground truth count label on 960 colony sub-images. The blue line is much closer to the dashed green line compared to the orange line, indicating that MCount is more accurate in recognizing sub-images. (d) Comparison of MCount and NICE performance in underestimation or overestimation of colony numbers. MCount has fewer underestimated results and less severe underestimation and overestimation compared to NICE. (e) Boxplot of error rate on 960 sub-images for MCount and NICE. MCount has a much lower average error rate and a smaller interquartile range, indicating more consistent better performance.

Figure 8a shows a bubble plot of MCount's counting result versus the ground-truth label on colony segments. The regression line of MCount is close to the ground-truth line, indicating that MCount is a good estimator for colony recognition on the segment dataset. It is worth noting that the bubbles representing coordinates (1,1), (2,2), and (3,3) are large compared to other bubbles because segments with 1-, 2-, and 3-colonies account for a large portion (59.6%, 18.52%, and 7.93%, respectively) of the segment dataset, and most of them are correctly counted by MCount. Therefore,

the regression line of MCount largely depends on segments with few colonies, even though the segment size in the dataset ranges from 1 to 27. MCount tends to slightly underestimate the number of colonies, but it is still accurate for most segments, as 92.26% of segments have no more than 5 colonies. In this case, the actual biases resulting from underestimation are zero because the counted number can only be natural numbers. The counted number monotonically decreases as λ increases, as shown in Figure 3-9a, and this underestimation issue can be addressed by choosing an appropriate value for λ. To better visualize the accuracy of MCount on colony segments, we used a rank error curve in Figure 3-8b, where all 15,847 segments are ranked increasingly according to their error rate. The x-axis represents the ranking of each segment, and the y-axis represents the corresponding error rate. When the error rate is larger than zero, counting tends to overestimate, while counting underestimates when the error rate is smaller than zero. A zero error rate was achieved on 91.45% of the segments, demonstrating good accuracy of MCount on colony segments.

The goal of the colony counting algorithm is to give an accurate count on sub-images rather than segments, therefore we further evaluated the accuracy of MCount on sub-images (Figure 3-8c-d) and compared it with NICE (Figure 3-8e). Figure 3-8c is a bubble plot showing both MCount and NICE versus the ground-truth label on sub-images. Both algorithms have relatively uniform distributions indicated by the uniform bubble sizes. While NICE largely underestimates the number of colonies, MCount slightly underestimates the number of colonies. One can tune the λ value to address the underestimation issue as shown in Figure 3-9b. Figure 3-8d shows the rank error curve of both algorithms, with MCount having better accuracy than NICE. Figure 3-8e shows the boxplot of both algorithms' recognition error on each of the 960 sub-images, with MCount having a 3.99% average error rate, substantially lower than NICE's average error rate of 16.54%.

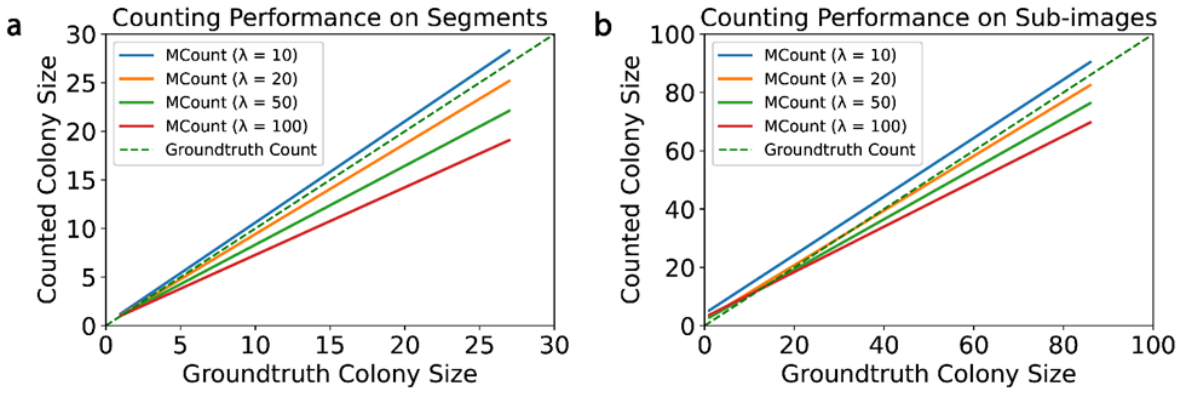


Figure 3-9: **Increasing λ results in monotonously less counting leading to underestimation of colony number.** The plot shows the MCount counting result with different λ values versus ground truth label on (a) 15,847 colony segments and (b) 960 sub-images, respectively. The green dashed line represents the regression line for a perfect colony estimator that always gives the result as the label, while the colored line represents the regression line for MCount. By tuning λ, it is possible to address the overestimation/underestimation issue. Increasing λ results in a monotonous decrease in the number of colonies counted, leading to an underestimation of colony number.

The processing time of MCount was evaluated on the dataset by randomly selecting 100 sub-images. The average time per sub-image was found to be less than 1.7 seconds, indicating that MCount can meet the demand for high-throughput processing. However, the processing time can be further optimized by using a better optimizer solver. A comparison of the processing time for MCount and other solutions is presented in Table 3.1.

Algorithms / Software	**Total Processing Time (Second)**
MCount	169.8
NICE	15.9
OpenCFU	13.2
AutoCellSeg	79.4

Table 3.1: **Processing time of different colony counting solutions for 100 randomly selected sub-images.** The optimization solver for MCount is 'COIN_CMD' in PuLP, and the solver can be changed to achieve faster processing times at the cost of sacrificing accuracy.

3.3.3 Hyperparameter optimization

In previous sections, we utilized a standard method of grid search and cross-validation with the goal of minimizing the average error rate on the entire dataset to tune the hyperparameter set (d, λ), which can be further improved by using techniques such as random search [110]. However, it can be computationally intensive and time-consuming when exploring the hyperparameter space in practice. Furthermore, this method require a pre-labeled dataset with a large quantity of data points, e.g., 960 (as shown earlier), which is impractical when deploying MCount to other counting tasks. In this section, we explore alternative methods for determining the hyperparameter set (d, λ) that only require a few labeled or even unlabeled data points. Table 3.2 summarizes the hyperparameter optimization methods used in this work.

	Image Type	**Number of Images**	d **Optimization**	λ **Optimization**
1	Labeled	960	Grid Search and Cross Validation	
2	Labeled	$10 \sim 20$	Empirical	Average of Samples
3	Unlabeled	$40 \sim 50$	Empirical	Equidispersion Assumption

Table 3.2: **Summary of hyperparameter optimization methods used in MCount.** The table includes the method number, image type (labeled or unlabeled), number of images used for optimization, hyperparameter d optimization method, and hyperparameter λ optimization method. Method 1 uses grid search with cross-validation on 960 labeled images. Method 2 and method 3 decoupled the optimization of d and λ and use the same empirical method for determining d. Method 2 uses the average λ of $10 \sim 20$ labeled images, while method 3 chooses the value of λ that leads to equidispersion on $40 \sim 50$ unlabeled images.

Although d and λ are not completely independent, they can be considered separately since they control different aspects of the algorithm. Specifically, d determines the level of detail in the contour representation, while λ controls the number of circles used in the optimization step. In this way, we can decouple these two hyperparameters and investigate them separately.

To determine the appropriate value of d, we propose an empirical method. MCount

can recognize elliptical segments that are usually comprised of two colonies by dividing the contour of the elliptical colonies into multiple contour pieces, as shown in Figure 3-4a. We began with $d = 100$ and selected a few elliptical segments that appeared to have two colonies. We then visualized the contour of each selected elliptical segment and gradually decreased the value of d until the contour was divided into 3 10 pieces. By applying this approach to the benchmark, we determined that $d = 0.5$ is a suitable value.

In the following sections, we focus on statistical methods for tuning λ and examine their consistency using statistical procedures.

3.3.3.1 The selection of λ by averaging a small number of labeled images

Although labeling hundreds of colony images to calibrate MCount requires an extensive effort, for other counting tasks, labeling only a few dozen images may be feasible. Therefore, we are interested in statistical methods that require only a few labeled images for tuning λ. Here we propose such a method:

1. Obtain a small number of labeled images (e.g., $n = 10$).

2. For each labeled image, find the value of λ_i such that MCount gives the correct colony count. Note that the count monotonically decreases as λ increases, making it easy to find λ_i.

3. Set the hyperparameter λ to the average value, $\lambda = \overline{\lambda} = \frac{1}{n}\sum_i \lambda_i$.

In order to examine whether the proposed method can consistently lead to an appropriate λ value and determine how many images are needed in step 1, we carried out a simulation using the benchmark dataset of 960 labeled images. The simulation procedure is as follows:

1. Randomly sample n images from the 960 labeled images.

2. For each labeled image, find the value of λ_i such that MCount gives the same colony count as the label. Note that the count monotonically decreases as λ increases, making it easy to find λ_i.

3. Calculate the average recognition error rate on the 960 images with $\lambda = \overline{\lambda} = \frac{1}{n}\sum_i \lambda_i$.

4. Repeat steps 1 to 3 1000 times.

By implementing the above simulation procedure, we can investigate the distribution of the average recognition error rates from 1000 trials to see whether this hyperparameter-tuning method can consistently lead to low error rates. Figure 3-10 shows the boxplot of simulation results under different values of n. Not surprisingly, a larger n leads to a narrower distribution and a lower mean/median. When $n = 10(20)$, 985 (998) out of 1000 trials have a λ value leading to a low error rate ($<8\%$) on the 960 images, which indicates consistently low recognition error rates. In practice, labeling ten to twenty colony images is a common practice in bio-labs, making it easy to calibrate MCount when applied to different counting tasks.

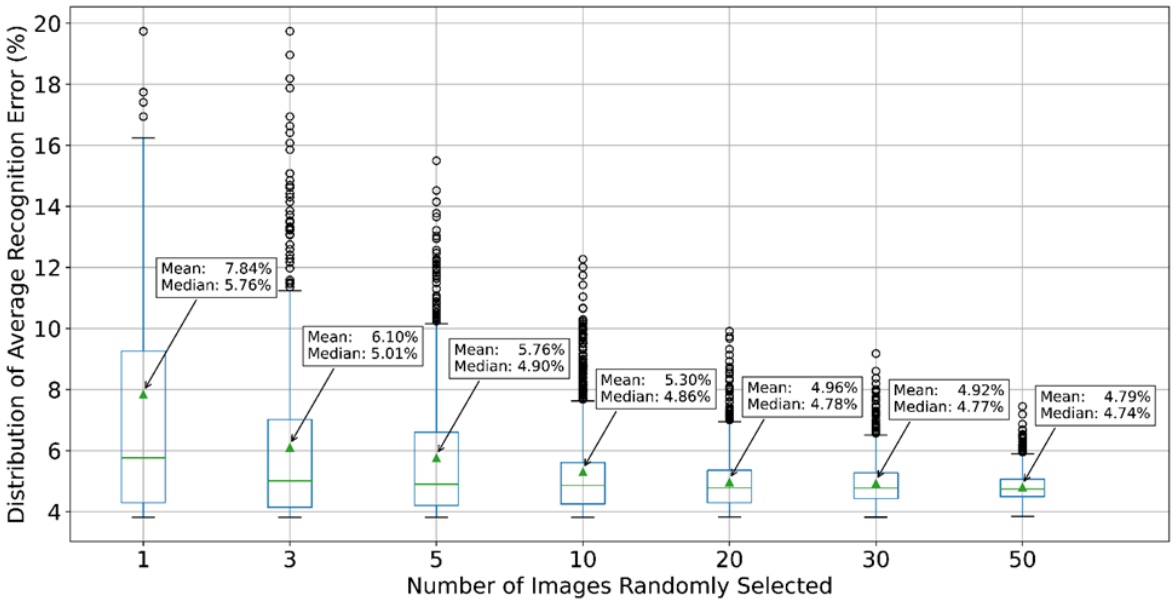


Figure 3-10: **Distribution of average recognition error rates on the benchmark when repeatedly implementing the hyperparameter-tuning method using n labeled images 1000 times.** In each replicate, a few (1, 3, 5, 10, 20, 30, or 50) images are randomly selected from the benchmark, and λ is tuned for each image to a proper value so that MCount gives the same counting number as the label. Then, the average of λ is chosen for this replicate, and the average recognition error rate is calculated on the benchmark using this λ. By simulating this procedure for 1000 replicates, we can plot the distribution of average recognition error rates. As expected, increasing n results in a narrower distribution of average recognition error rates, leading to more consistent performance. Note that when $n = 10$, all recognition errors fall in the range of 3.5% to 14% with a mean of 5.30% (median of 4.86%), much lower than the recognition error rate of NICE at 16.54% (15.79%).

3.3.3.2 The selection of λ based on equidispersion assumption on unlabeled images

In many labs, there are often cases where acquiring a large number of images is feasible, but labeling even a single image is challenging. Therefore, we need statistical methods that can tune λ using unlabeled images. Introducing some prior information could help make good use of the unlabeled data. One such prior is the assumption that the distribution of colony numbers has equidispersion, meaning that the mean is equal to the variance, a key property of the Poisson distribution. We discuss this prior information

in more detail before introducing and examining our method in the following section. By utilizing this prior information, we hypothesize that we can obtain a reasonably accurate λ value using unlabeled images. By analyzing the simulation results, we can determine the minimum number of unlabeled images required to obtain an accurate λ value for a specific task.

3.3.3.2.1 Hypothesis testing for assessing equidispersion

It is predicted that many biological observations follow a Poisson distribution, and colony numbers are an example, where a fixed volume of liquid is independently and randomly sampled from the same source for every plating [117, 118]. In this case, the equidispersion property is naturally satisfied.

However, it is important to note that the prior knowledge chosen in this method is the equidispersion property, not specifically the Poisson distribution. Equidispersion is a weaker requirement than the Poisson distribution, and it is possible that other distributions can also exhibit equidispersion. In the Appendix A, we further discuss this scenario and show that our benchmark is closer to a normal distribution than a Poisson distribution, although our equidispersion assumption still leads to a low recognition error.

Another reason why we prefer equidispersion to the Poisson distribution is that it can be challenging to use statistical methods to confirm that a distribution follows Poisson, while equidispersion can be tested more easily. Here we propose a simple hypothesis testing method to assess equidispersion, known as the Poisson Dispersion test, which also serves as a likelihood test for the Poisson distribution

H_0 : the data has equidispersion, *i.e., Mean=Variance*

H_1 : $Mean \neq Variance$

The test statistic is:

$$D = \sum_{i=1}^{N} \frac{\left(X_i - \overline{X}\right)^2}{\overline{X}} \tag{3.5}$$

Then we can calculate the p-value associated with the chi-square distribution ($N-1$ degrees of freedom). If the p-value is greater than the chosen level of significance, we fail to reject the null hypothesis H_0 and can assume that equidispersion is satisfied.

3.3.3.2.2 Proposed method and its examination

Assuming that equidispersion is satisfied, we propose a statistical method that utilizes unlabeled images to tune λ as follows:

1. Obtain a set of unlabeled images (e.g., $n = 40$).
2. Apply MCount to the images and obtain the counting results under different λ values.
3. Calculate the mean and variance of the counting results under different λ values and find a value $\hat{\lambda}$ where the mean equals the variance.
4. Set the hyperparameter $\lambda = \hat{\lambda}$.

To evaluate the effectiveness of the proposed method using unlabeled images, we conducted a simulation procedure as follows:

1. Randomly select n images from a dataset comprising 96 labeled images with equidispersion. Erase the label of the selected images.
2. Apply MCount to the n images and obtain counting results under different λ values.
3. Calculate the mean and variance of the counting results under different λ values and find a value $\hat{\lambda}$ where the mean equals the variance.
4. Calculate the average recognition error rate on 96 labeled images with $\lambda = \hat{\lambda}$.
5. Repeat steps 1 to 4 1000 times.

The boxplot in Figure 3-11 shows the distribution of the 1000 average recognition error rates for different values of n. Like the case for labeled images, larger n values lead to a narrower distribution and smaller mean. Interestingly, the median remains constant regardless of n. When $n \geq 40$, the maximum line of the boxplot is lower than 8%, indicating consistently low recognition error rates. However, the required n value is much larger than that for labeled images, as unlabeled images contain less information, requiring more training data for hyperparameter optimization.

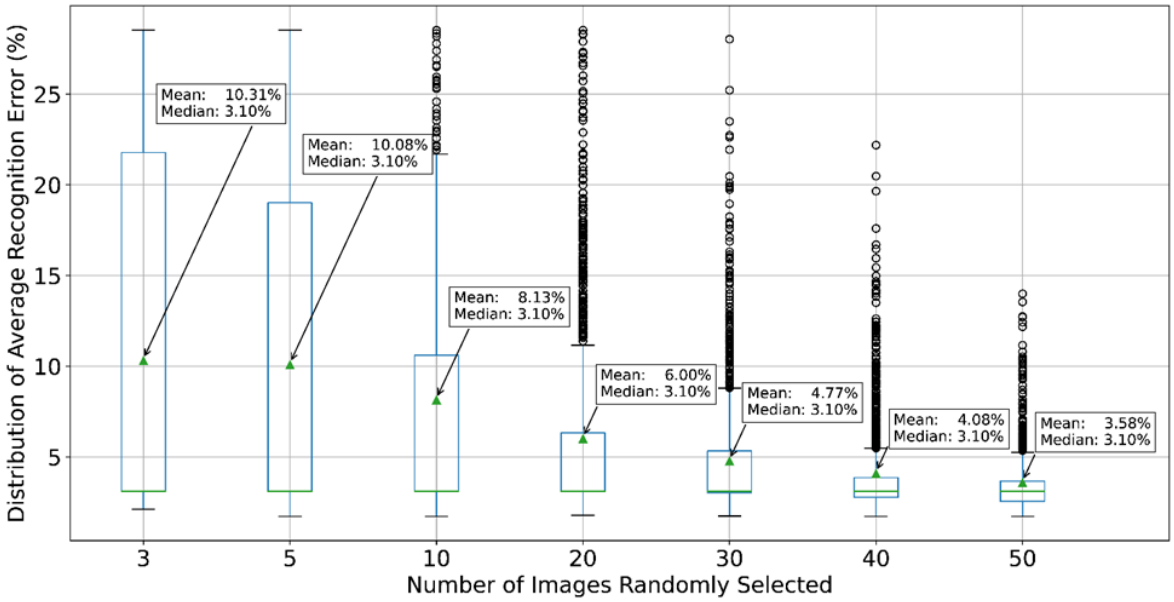


Figure 3-11: **Distribution of average recognition error rates on 96 images with equidispersion when repeatedly implementing the hyperparameter-tuning method using n unlabeled images 1000 times.** In each replicate, a few (3, 5, 10, 20, 30, 40, or 50) images with the label erased are randomly selected from 96 images, and λ is tuned so that the mean of the counting equals the variance. Then, the average recognition error rate is calculated on the 96 labeled images using this λ. By simulating this procedure for 1000 replicates, we can plot the distribution of average recognition error rates. As n increases, the distribution of average recognition error rates becomes narrower, leading to a more consistent distribution. Note that only when $n > 40$, the maximum line of the box plot is lower than 8%, and the mean is lower than 4.08%. The required n is much larger than the case for labeled images because unlabeled images have less information, requiring more training data for hyperparameter optimization. These results demonstrate that the proposed method is effective in achieving consistently low error rates using only a small number of unlabeled images for hyperparameter tuning.

3.4 Conclusion

In this work, we propose a colony counting solution, MCount, that can recognize merged colonies that frequently occur in high-throughput workflows, which is beyond the capabilities of current solutions that only adopt regional algorithms. MCount extracts contour information and combines it with regional information using an optimization algorithm. To evaluate the performance of MCount, we prepared a GFP fluorescent E. coli. DH10-beta based colony dataset, which is comprised of both sub-images and segments labeled according to historical photos. MCount maintains an average recognition error of 3.99% on the sub-image dataset (using grid search and 10-fold cross-validation to minimize the average error rate), which is much lower than current solutions like NICE of 16.54%.

While MCount only has two hyperparameters, it is easy to deploy MCount to other counting tasks. In addition to the standard hyperparameter optimization method which requires several labeled images, we further proposed two methods that require a small number of labeled images or unlabeled images, respectively. To examine the statistics of proposed methods, we conducted simulations and found that they all guarantee consistently low error rates compared to existing methods. The simulations showed that the method requiring labeled images achieved low error rates with as few as 10 labeled images, while the method requiring unlabeled images required at least 40 images to achieve consistently low error rates.

Overall, the statistical evaluation of our proposed methods provides a strong basis for their potential deployment in various counting tasks. Future applications of this solution include colony classification for multiple strains on the same plate, which could be achieved by classification algorithms such as K-nearest neighbor based on the color and size of colonies.

Chapter 4

M-TUBE: a microfluidic device for large-volume bacterial electroporation

This chapter is reproduced in part from:

Huang, P. H., Chen, S., Shiver, A. L., Culver, R. N., Huang, K. C., & Buie, C. R. (2022). M-TUBE enables large-volume bacterial gene delivery using a high-throughput microfluidic electroporation platform. *PLoS Biology, 20*(9), e3001727.

4.1 Introduction

One of the key steps in bacterial genetic engineering is the delivery of DNA into cells, which can be realized by mechanical, chemical, or electrical methods [119–121]. Among these methods, electroporation has been the gold standard because it is not cell-type-specific [120], can deliver molecules of various sizes [122], and can exhibit relatively high efficiency under optimized conditions [120, 123]. For optimal electric field conditions, genetic material enters cells through reversible pores formed in the cell membrane [124, 125]. Electroporation is typically performed using cuvettes, in an operator-dependent manner that is limited to small batches of volume 1 mL or less. Even with high efficiency, creation of a comprehensive mutant library with hundreds of thousands of mutants [126–128] for functional-genomics studies can require electroporation of large volumes

(tens of milliliters) of saturated bacterial culture, which corresponds to hundreds of cuvette-based electroporation reactions. Performing serial electroporation with manual pipetting is a labor-intensive, time-consuming, and costly process. Moreover, cuvette-based electroporation suffers from issues such as residual volume and joule heating [129, 130], which affect electroporation efficiency, cell viability, and overall yield.

Performing electroporation in a microfluidic format [129–132] can remove the need for manual pipetting and improve heat dissipation [129, 132], thereby increasing electroporation efficiency and cell viability. However, most microfluidic devices involve complicated fabrication processes using PDMS [77, 133–136], which is an obstacle to widespread adoption, particularly within the microbiology community that would most benefit. Microfluidics-based electroporation devices are also typically limited by the sample volume they can handle. These devices are commonly used for mammalian cells [136, 137], with just a few examples of applications to bacteria [77, 138]. Several commercial products [139–143] have demonstrated the potential for scaling up electroporation to throughput of up to ∼100 mL at 8 mL/min [143], but most have been applied only to mammalian cells and still rely on batch-wise operation [139–143]. Moreover, existing commercial systems require sophisticated electroporation chambers that limit the volume that they can process. Thus, the capabilities of these systems for large-volume bacterial electroporation are yet unproved.

The ideal genetic transformation system would allow for a wide range of sample volumes to accommodate different applications, especially involving the creation of mutant libraries given the low electroporation efficiency of many understudied yet health-relevant bacterial species [128, 144, 145]. A scalable, high-volume electroporation device should be easily assembled by a microbiologist without sophisticated fabrication, compatible with commercially available and common laboratory equipment, and able to process relevant sample volumes in minutes to minimize biological variability. To this end, here we introduce a simple yet powerful Microfluidic TUbing-based Bacterial Electroporation device (M-TUBE) that enables flexible electroporation of large-volume bacterial samples. M-TUBE facilitates scalable, continuous flow, large-volume bacte-

rial electroporation without the need for micro/nanofabrication, PDMS casting, or 3D printing of microfluidic channels and electrodes.

4.2 Methods

4.2.1 Cultivation and Preparation of Cell Strain

4.2.1.1 Culturing and preparation of *E. coli* strains

Three *E. coli* strains, including NEB10β (New England Biolabs), K-12 MG1655 (Coli Genetic Stock Center, Yale University) and Nissle 1917 (Mutaflor®, Canada), were employed in this study to test the M-TUBE device. The strains, unless otherwise specified, were cultured, harvested, and made electrocompetent using the same conditions. In brief, glycerol stocks were inoculated into two 14-mL cultures tubes containing 6 mL of LB medium and incubated at 37 °C and 250 rpm. The next morning, 5 mL from each overnight culture was inoculated into 245 mL of LB and grown at 37 °C and 200 rpm to an OD_{600} of 0.5-0.7. Note that each set of *E. coli* experiments involved 15-20 mL of electrocompetent cells at $OD_{600} = 10$, which required two 250-mL cultures. Each 250 mL culture was divided equally into six 50-mL centrifuge tubes and spun down at 4 °C and 3500 rpm for 10 min using an Allegra 64R centrifuge (Beckman Coulter). The supernatant was discarded and 6 mL of ice-cold 10% glycerol was used to wash and combine the six cell pellets into one suspension. Each 6-mL cell suspension was equally divided into four 2.0-mL microcentrifuge tubes. The eight microcentrifuge tubes generated from the two 250-mL cultures were centrifuged at 4 °C and 8000 rpm for 5 min, the supernatants were discarded, and 1 mL of ice-cold 10% glycerol was used to wash and resuspend the pellet in each of the 8 tube. These washing steps were repeated twice more. Next, all cell pellets were combined into a concentrated suspension using 8 mL of ice-cold 10% glycerol and the cell concentration (typically $OD_{600} = 20 - 30$) was measured using a UV spectrophotometer (UV-1800, Shimadzu). Depending on the measured concentration, a final sample with $OD_{600} = 10$ was prepared by adding

an appropriate volume of ice-cold 10% glycerol. This sample was placed on ice prior to electroporation. DNA plasmids (Parts Registry K176011) [77] encoding ampicillin resistance and green fluorescent protein (GFP) were added to this sample at a final concentration of 0.1 ng/μL for NEB10β and MG1655 cultures; for Nissle 1917, a final concentration of 1 ng/ μL was employed so that the number of colony forming units (CFUs) was above the limit of detection. For electroporation, the sample was loaded into a 30-mL plastic syringe (see section on M-TUBE operation).

4.2.1.2 *B. longum* culturing and preparation for M-TUBE electroporation with plasmid DNA

A 5-mL *B. longum* culture was maintained in an anaerobic chamber (Coy) via daily dilution into fresh medium to prepare for electroporation. Briefly, 1 mL of a *B. longum* culture was inoculated into 9 mL of MRS medium in a culture tube, and five additional serially diluted (at 1:10 ratio) cultures were prepared; these six cultures were incubated at 37 °C overnight. The next morning, the optical density of each culture was measured using a spectrometer, and the culture with $OD_{600} = 3 - 4$ was used for subsequent outgrowth. The selected culture was diluted to $OD_{600} = 0.54$ in 60-70 mL and grown to $OD_{600} = 1.5 - 2$, after which cells were harvested and made electrocompetent following the same steps described above for *E. coli*. The 60-70 mL were then divided equally into two 50-mL centrifuge tubes and spun down outside the anaerobic chamber at 4 °C and 3500 rpm for 10 min using an Allegra 64R ultracentrifuge (Beckman Coulter). Next, the two 50-mL centrifuge tubes were returned to the anaerobic chamber, the supernatant was discarded, and 5 mL of ice-cold 10% glycerol were used to wash and combine the two cell pellets into one suspension. The 5-mL cell suspension was divided equally into four 2-mL microcentrifuge tubes. The four tubes were centrifuged inside the chamber at room temperature and 10,000 rpm for 2 min using an Eppendorf 5418 microcentrifuge, the supernatants were discarded, and 1 mL of ice-cold 10% glycerol was used to wash and resuspend the pellet in each of the 4 tubes. These washing steps were repeated two more times. Next, all pellets were combined into a concentrated

suspension using 5 mL of ice-cold 10% glycerol. Depending on the concentration, the final sample at OD_{600} = 10 was prepared by adding the appropriate volume of ice-cold 10% glycerol and then placed on ice prior to electroporation. The pAM5 plasmid encoding tetracycline resistance was added to the sample at a final concentration of 2 ng/μL. The mixture of the plasmid DNA with the cells was loaded into a 10-mL plastic syringe for electroporation.

4.2.1.3 Transposon mutagenesis of *B. longum* NCIMB8809

Previous transformation protocols [146–148] were combined with minor modifications to prepare electrocompetent cells of *B. longum* NCIMB8809. Briefly, a glycerol stock of *B. longum* NCIMB8809 was recovered for 24 h in 5 mL of MRS broth (MRS media, Difco) at 37 °C and passaged overnight (16 h) in 10 mL of MRS in a 10-fold dilution series. The next morning, the incubator temperature was raised to 40 °C and one of the overnight cultures in the dilution series was used inoculate 50 mL of MRS (MRS media, HIMEDIA) in a 250-mL Erlenmeyer flask at an initial OD600OD_{600} (optical density at λ=600 nm) of 0.18, as measured by a 96-well plate reader (Epoch2, BioTek) in a 96-well flat bottom microplate (Grenier Bio-One, Cat. #655161) with 200 µL of culture per well. In the dilution series, the overnight culture with the lowest optical density that still provided enough cells to proceed was used to inoculate the next culture. The 50 mL of culture in HIMEDIA-brand MRS was grown to an OD_{600} of 1.0 and used to inoculate MRS broth reconstituted from individual components, modified with 1% lactose as the sole carbon source and an additional 133 mM NaCl, at an initial OD_{600} of 0.18. This culture was harvested at an OD_{600} of 0.5, pelleted, washed three times with 15% glycerol, and resuspended at an OD_{600} of 6.7 in 15% (v/v) glycerol. To harvest the cells, the culture was moved to a pre-reduced 50 mL conical tube (Fisher Scientific, Cat. #06-443-19) on ice, brought out of the anaerobic chamber, centrifuged for 10 min at 3,428g (Centrifuge 5920R, Eppendorf), and transferred back into the anaerobic chamber. After cells were harvested, the incubator temperature was lowered back down to 37 °C. Subsequent washes were performed at a volume of 5 mL in 5-mL Eppendorf

tubes (Cat. #0030122321, Eppendorf) and pelleted with a compatible microcentrifuge (MC-24™ Touch, Benchmark Scientific) that had been brought into the chamber, using 2-min 10,000g centrifugation steps. Transposomes were assembled *in vitro* by mixing an erythromycin resistance cassette with commercially available EZ-Tn5 transposase according to anufacturer's instructions. Transposomes were mixed with competent cells at a concentration of 2U transposase/mL competent cells and electroporated using the M-TUBE device. Electroporated cells were recovered for 2 h at 37 °C, concentrated 10-fold through centrifugation and resuspension in MRS, and plated on RCM-agar plates with 5 μg/mL erythromycin. Colonies were harvested for sequencing after ∼36 h of growth at 37 °C.

4.2.1.4 Preparing a Tn-seq library for *B. longum* NCIMB8809

Erythromycin-resistant colonies from the Tn5 transposome electroporation were scraped from the selective plates into 500 μL of MRS broth (MRS media, Difco) for each Petri dish. Samples from this suspension were taken, glycerol (Fisher Bioreagents, Cat. #BP229-1) was added to a final concentration of 15% (v/v), and the cryostocks were stored in 11-mm crimp vials (Thermo Scientific™, Cat. #C4011-11) with sealed aluminum crimp caps (Thermo Scientific™, Cat. #11-03-400) at -80 °C. Simultaneously, most of the suspension was stored directly at -20 °C for subsequent DNA isolation. Genomic DNA (gDNA) was isolated from the colony suspension using a DNeasy Blood and Tissue Kit (QIAGEN, Cat. #69506) according to the manufacturer's protocol for Gram-positive organisms.

Isolated gDNA was first sheared in a Covaris S220 Sonicator with microTUBE AFA fiber pre-slit snap-cap tubes (Covaris, Cat. #520045) according to the manufacturer's instructions for 300-bp fragments. A KAPA HyperPrep Kit (Roche, 07962312001) with custom oligos was then used to prepare the library. Briefly, sonicated gDNA was subjected to a dual bead-based size selection using AMPure XP magnetic beads (Beckman Coulter, Cat. #A63881) according to the manufacturer's instructions for 300-bp sized fragments. An end-repair and A-tailing reaction was performed followed by an adap-

tor ligation by following the KAPA HyperPrep protocol and using a custom adaptor. After a one-sided bead cleanup, the entire sample of adaptor-ligated gDNA fragments was used as input for a PCR reaction that simultaneously amplified transposon-gDNA junctions and added Illumina TruSeq adaptors. An Ultra II Q5 PCR mix (New England Biolabs, Cat. #E7649A) was used for all PCR reaction components except the template DNA and custom primers. The PCR reaction involved an initial denaturation step of 98 °C for 2 min, followed by 25 cycles of three steps: 98°C for 30 s, 65 °C for 20 s, and 72 °C for 30 s. After a final extension at 72 °C for 10 min, the sample was cleaned up using a NucleoSpin® Gel and PCR Clean-up kit (Machery-Nagel, Cat. #740609.250). The Tn-seq library was run on a MiSeq (Illumina, Cat. #SY-410-1003), with the 150-cycle MiSeq Reagent Kit V3 (MS-102-3001), 150-bp read 1 length, and no indexing reads.

4.2.2 Device Fabrication and Simulation

4.2.2.1 Materials

Syringe needles of various gauges (16, 20, or 23) with blunt tips were purchased from CML Supply LLC. Plastic tubing of various diameters were purchased from Cole-Parmer: 0.5-mm inner diameter (ID) (PB-0641901), 0.8-mm ID (EW-07407-70), and 1.6-mm ID (EW-07407-71). Plastic syringes of various volumes with Luer-Lok tips were purchased from Thomas Scientific: 30 mL (BD302832), 20 mL (BD302830), and 10 mL (BD302995). Luria broth (LB) (BD244620) and dehydrated agar (BD214010) were purchased from VWR. MRS broth (BD288130) and Reinforced Clostridial Medium (RCM) (CM0149B) were purchased from Fisher Scientific. Carbenicillin disodium salt (C3416), tetracycline (T7660), L-Cysteine (C7352), α-Lactose monohydrate (L2643), and sucrose (S7903) were purchased from Millipore Sigma. Oligonucleotides were purchased from Integrated DNA Technologies.

4.2.2.2 Modeling electric field strength and temperature distribution in an M-TUBE microchannel

To simulate the electric field when using plastic tubings of different diameters to assemble M-TUBE devices and the temperature distribution under different combinations of electroporation conditions, we built a numerical model in COMSOL Multiphysics v6.0 (Burlington, MA). The model was based around the multiphysics module of electromagnetic heating, which couples the physics of electric currents, laminar flow, and heat transfer in solids and fluids. To reduce the complexity of the model and computation, we used a simplified channel geometry that only includes the tips of the two needle electrodes and the microchannel formed between two needle electrodes; in short, the geometry of M-TUBE used for simulation is 500 μm in diameter and 3 mm in length. Equations, boundary conditions, assumptions and numerical techniques used to compute electric fields, flow fields and temperatures can be found in previous literatures [77, 149, 150]. It is worthwhile pointing out here that to conservatively model the temperature distribution inside M-TUBE, an electric conductivity of 0.01 s/m was assumed for 10% (v/v) glycerol to take the potential increase in conductivity during electroporation [151–153] into account.

4.2.2.3 Protocol for preparation of an M-TUBE device

An M-TUBE device is assembled from two syringe needles and one piece of plastic tubing with a pre-defined length (Figure 4-1d). Here, we describe the details of assembly of an M-TUBE device with a microchannel length of 3 mm and a tubing ID of 0.5 mm. First, we cut plastic tubing (50 feet per roll) into 20-mm segments on a cutting mat with metric dimensions. Second, we take two syringe needles of 23 gauge with a tip length of 0.5 in, which has an outer diameter of 0.63 mm that ensures tight fitting between the tubing inner surface and the outer surface of the syringe needle. Next, we insert one of the syringe needles into the tubing and repeatedly rotate back and forth the tubing and/or syringe needle, until the tip of the syringe needle is close to

the middle of the tubing and there is also a small portion of the needle for electrical connection that is not inserted into the tubing. We then insert the other syringe needle and rotate back and forth the tubing/syringe needle or the 2nd syringe needle until a gap (i.e., the microchannel length) of a 2-4 mm between the tips of the two syringe needles is established. The gap size can be checked by placing the entire assembly close to a tape measure. After assembling the three components, we remove the plastic hub from either of the syringe needles. Upon removal of the plastic hub, the gap size should then be carefully re-checked with a tape measure, and slight adjustments can be made to establish a gap of 3 mm by gently twisting either needle inward or outward. After this final adjustment, the M-TUBE device is completely assembled.

As discussed above, assembly of one M-TUBE device requires only 60-90 s, hence we typically prepare 50 M-TUBE devices at a time, in ∼1 h. The M-TUBE devices are placed in a Petri dish, which is sterilized in a biosafety cabinet with UV irradiation overnight. After UV sterilization, M-TUBE devices are stored in a -20 °C freezer or refrigerator until just before conducting electroporation experiments, a step similar to the pre-chilling of electroporation cuvettes.

To prepare M-TUBE devices with other tubing sizes, all steps remain unchanged, but it is necessary to ensure that the plastic tubing is assembled with syringe needles that have complementary outer diameters in their tips.

4.2.3 Experiment Setup and Protocol

4.2.3.1 The external high-voltage power supply system

The external high-voltage power supply simply consists of a function generator (Agilent Technologies, 33220A), a high-voltage amplifier (Trek Inc., 623B), and an oscilloscope (Agilent Technologies, DSO-X 2022A). The function generator supplies any electric signals that we pre-program/pre-define (AC or DC, sine or square waves, frequency, voltages, etc) to the high-voltage amplifier, and then the high-voltage amplifier amplifies those signals by 1,000 folds, while the oscilloscope is employed to monitor if the amplified

signals are being output correctly or not. The function generator provides original, non-amplified signals to the high-voltage amplifier through a BNC cable, and then the high-voltage amplifier outputs the 1000-fold amplified signals through a pair of high-voltage cables, which are then customized with alligator clips (or test clips) and connected to the two electrodes of the M-TUBE device. The switching ON/OFF of the high-voltage signals is primarily controlled by the function generator; by simply engaging and disengaging a "trigger" button on the function generator, we can, respectively, switch on and off the output of signals. Note also that the function generator, the high-voltage amplifier and the oscilloscope are of standard, common electronics equipment that can be accessed in many research laboratories/facilities or they are also readily available at a reasonable cost.

4.2.3.2 Electroporation of *E. coli* strains using M-TUBE

The final sample of cells mixed with plasmid DNA was loaded into a plastic syringe, which was mounted on a syringe pump (Legato 210P, KD Scientific) that could be operated horizontally or vertically. To prevent bending of the plastic tubing of the M-TUBE device and to enable convenient collection of the electroporated sample directly into tubes, we typically operate the syringe pump as shown in Figure 4-1c. After arranging the pump to operate vertically, an M-TUBE device was attached to the sample-loaded syringe via Luer-Lok connection, and the two syringe-needle electrodes were connected to the external high-voltage power supply system. Upon confirming a tight connection between the M-TUBE device and the power supply, we pre-filled the M-TUBE microchannel by infusing the cell sample at a relatively low flow rate (typically 250-500 μL/min), to prevent air bubbles and thereby arcing/sparking in M-TUBE, until we visually confirmed that the microchannel was filled with the liquid cell sample. Next, a collection tube (reservoir) was placed underneath the M-TUBE device (Figure 4-1c) so that the electroporated sample could be directly and automatically collected. We programmed the pumping parameters including target pumping volume and pumping flow rate, and started flow using the syringe pump at the pre-set flow

rate; immediately after starting flow, we started the application of electric signals to the M-TUBE device to initiate electroporation.

As a positive control, the same batch of electrocompetent cells was also electroporated at various field strengths using 0.2-cm electroporation cuvettes (VWR, 89047-208). One hundred microliters were pipetted into a pre-chilled electroporation cuvette. Each cuvette was pulsed with an electroporator (MicroPulser™, Bio-Rad) at field strengths including 8.33 kV/cm, 10.0 kV/cm, 12.5 kV/cm, and 15 kV/cm with time constants between 5.0-5.5 ms. Immediately after the application of electric pulses to each cuvette, 900 μL of pre-warmed (37 °C) LB recovery medium were added to each cuvette, and the 100-μL electroporated cells was mixed with the 900-μL recovery medium via pipetting. We then aspirated as much electroporate sample volume as possible from the cuvette and dispensed it into designated wells on a 96-well deep plate, along with the electroporated samples from M-TUBE for subsequent recovery at 37 °C for 1 h.

4.2.3.3 Electroporation of *B. longum* via M-TUBE

Most steps for *B. longum* were the same as for *E. coli* described above; the differences are described here. After pre-filling an M-TUBE device with the *B. longum* sample, a 50-mL conical tube (reservoir) containing MRS recovery medium was placed underneath the M-TUBE device, so that electroporated B. longum cells could be directly and automatically flowed into the recovery medium. For *B. longum* electroporation with M-TUBE, one flow rate (7.2 mL/min, or 592 mm/s for the 0.5-mm M-TUBE device) was tested at three field strengths (3.33, 5.00, and 8.33 kV/cm).

As a positive control, the same batch of electrocompetent cells was electroporated at the same three field strengths using 0.2-cm electroporation cuvettes. One hundred microliters of the final cell sample were pipetted into a pre-chilled electroporation cuvette. Each cuvette was pulsed by the electroporator with time constants ranging between 5.4-5.8 ms. Immediately after the application of an electric pulse, 1000 μL of pre-warmed (37 °C) LB recovery medium were added to each cuvette and mixed with the cells via

pipetting. We then aspirated as much electroporated sample volume as possible from the cuvette and dispensed it into a 1.5-mL microcentrifuge tube.

4.2.3.4 Collection, recovery, and evaluation of electroporated E. coli samples

In each set of *E. coli* experiments, a range of flow rates and electric field strengths were tested; for each combination of testing conditions, 1 mL of electroporated sample was collected in a microcentrifuge tube. One hundred microliters of the electroporated sample was aspirated and dispensed into each of four wells of a 96 deep-well plate containing LB recovery medium. In each 96-well plate, we were able to test 20 combinations of electroporation conditions. After filling all designated wells of the 96-well plate, the plate was incubated in a shaking incubator at 37 °C and 250 rpm for 1 h. After 1 h of recovery, the 96-well sample plate was placed in a designated position on a liquid handling robot (Janus BioTx Pro Plus, PerkinElmer) for automated serial dilution: 10X, 100X and 1000X dilution for *E. coli* NEB10β; 10X and 100X dilution for *E. coli* K12 MG1655 or Nissle 1917. Following serial dilution, 5 μL from each well were dispensed onto LB-agar plates containing 50 μg/mL carbenicillin, and the selective plates were incubated overnight at 37 °C. The next morning, each plate was photographed for CFU counting.

4.2.3.5 Collection, recovery, and evaluation of electroporated *B. longum* samples

After electroporating *B. longum* using M-TUBE, 1 mL of cells was flowed directly into 10 mL of MRS recovery medium. *B. longum* samples electroporated by M-TUBE or in cuvettes were incubated at 37 °C for 3 h. Following recovery, 1.1 mL from each M-TUBE or cuvette sample were aspirated and pipetted into separate 1.5-mL microcentrifuge tubes and spun down at 10,000 rpm for 2 min. The supernatants were discarded and 200 μL of MRS medium were added into each 1.5-mL tubes to resuspend the cell

pellets. Next, the 200-μL suspension was plated onto RCM-agar plates with 10 μg/mL tetracycline, and the selective plates were incubated at 37 °C for at least 48 h. Following the 48-h incubation, each plate was photographed for CFU counting.

4.2.3.6 CFU quantification

Photos of selective plates for electroporation with plasmids were captured using an iPhone 11 (Apple) on a tripod with a remote shutter. The photos were imported to ImageJ (NIH) and CFU.Ai v. 1.1 for enumerating the CFUs. The transformation efficiency was defined as the number of CFUs on selective plates per μg of DNA.

4.3 Results

4.3.1 Assembly and characterization of the M-TUBE device

The M-TUBE device consists of 2 syringe needles and 1 plastic tube of a defined length (Figure 4-1a). The plastic tubing serves as the microfluidic channel, and the syringe needles serve as the two electrodes, which, when connected to an external high-voltage power supply, establish an electric field across the tubing microchannel. Upon establishing an electric field in the channel, bacterial cells flowing through the channel can be electrotransformed and uptake surrounding genetic material. The syringe needles and plastic tubing used to assemble M-TUBE are commercially and readily available at low cost (<\$0.21 per device), and the overall size of an M-TUBE device is similar to that of a conventional cuvette (Figure 4-1b). Because syringe needles of standard common formats can be used, M-TUBE can be attached to any commercially available syringe with complementary connectors and can be conveniently interfaced with any syringe pump for sample delivery (Figure 4-1c).

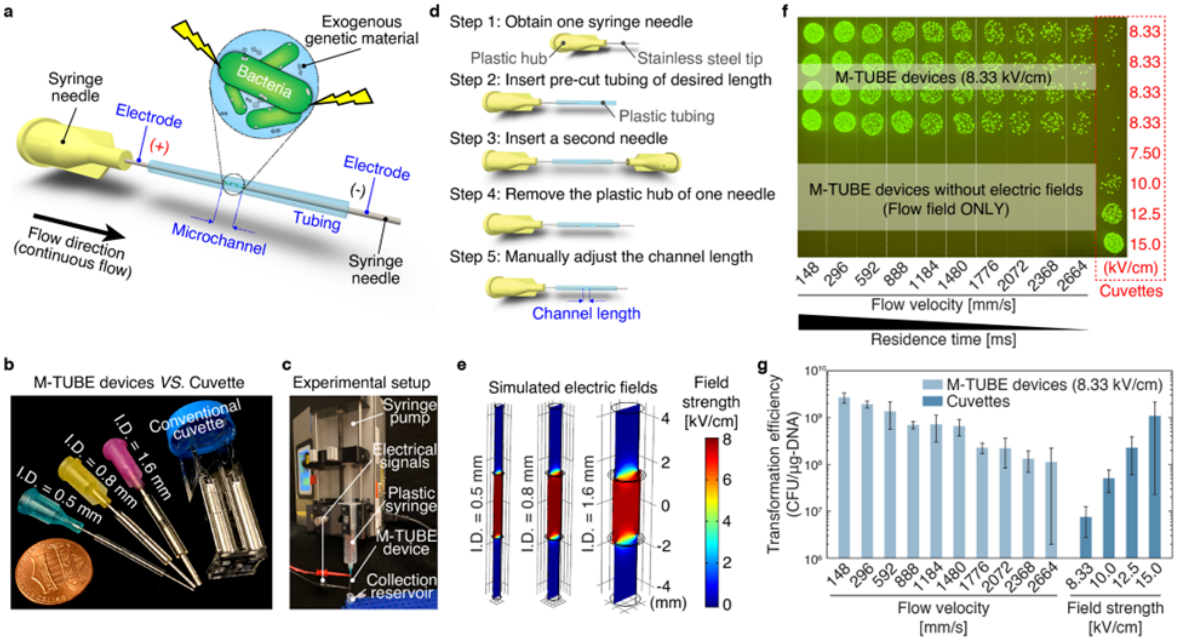


Figure 4-1: **M-TUBE is a fabrication-free, microfluidics tubing-based bacterial electroporation device that is simple to assemble and exhibits higher electroporation efficiency than cuvettes.** (a) Schematic of the M-TUBE device. The device is composed of two syringe needles and one piece of plastic tubing of predefined length. The two syringe needles and plastic tubing serve as the two electrodes and microchannel, respectively. When the two electrodes are connected to an external power supply (or electrical signal generator), an electric field is established within the microchannel, where bacterial electroporation can take place. (b) M-TUBE devices with three inner diameters (I.D.) are all similar in size to a conventional cuvette. (c) Photograph of the experiment setup when using the M-TUBE device. (d) Detailed breakdown of the protocol for M-TUBE assembly. One device can be completely assembled in 90–120 s. The total cost of parts is currently less than \$0.22 and this price could be lowered if parts are bought in bulk. (e) Simulations of the electric field established in M-TUBE devices using COMSOL Multiphysics 5.5 predict similar field strengths irrespective of I.D. (f) Spot-dilution assay to quantify viability on selective plates when *E. coli* NEB10β cells were flowed through the device with a plasmid encoding ampicillin resistance and GFP in the presence or absence of an electric field. Transformation was dependent on the electric field. For M-TUBE devices, a voltage of $\pm$2.50 kV (AC field) was applied, which results in an electric field of 8.33 kV/cm. The same batch of cells was used to conduct cuvette-based electroporation as a comparison. (g) Comparison of transformation efficiency (colony forming units (CFUs) per μg of DNA) corresponding to the plates in (f). The electroporation efficiency of M-TUBE decreased as the fluid velocity was increased, as expected due to the shorter duration of exposure to the electric field. Regardless of the fluid velocity, the efficiency of M-TUBE was at least one order of magnitude higher than that of cuvettes with the same field strength (8.33 kV/cm). Data represent the average ($n \geq 3$) and error bars represent 1 standard deviation.

The M-TUBE device can be easily assembled in five steps (Figure 4-1d). In brief, device assembly is accomplished by inserting one syringe needle into the plastic tubing cut to a particular length, and a second syringe needle is inserted into the other end of the tubing. Once both needles are inserted, the length of the channel is manually adjusted to a pre-defined value by modifying the gap between the facing ends of the two syringe needles. Assembling a single M-TUBE device requires only 90-120 s, far more convenient than typical fabrication processes for microfluidic devices (usually require several days).

Simulations of the electric field established in the tubing microchannel of M-TUBE (Figure 4-1e) indicate that the electric field strength is unaffected by the size of the microchannel (i.e., the tubing inner diameter), assuming that the applied voltage (e.g., 2.50 kV) and distance between the two electrodes (gap, or microchannel length) are held constant. This characteristic enables M-TUBE devices to cover a wider range of sample flow rates without having to adjust the applied voltage to maintain the same field strength. The gap of M-TUBE devices can be easily adjusted without additional assembly, unlike devices that rely on microfabrication, CNC machining, or 3D printing [154], providing a simple method for adjusting electric field strength of a device. Another beneficial feature is that the residence time within M-TUBEs can be adjusted to control cell exposure to the electric field. Since M-TUBE electroporates bacterial cells in a continuous flow manner, the residence time is dictated by the fluid velocity (or flow rate), such that residence time decreases with an increase in fluid velocity if the gap is fixed. These two features, gap length and flow rate, offer users more flexibility in tuning important electroporation parameters such as the electric field strength and the residence time, respectively, which are not always readily tunable in conventional electroporators.

4.3.2 Optimization of bacterial electroporation with M-TUBE

To establish the utility of M-TUBE, optimize its design, and showcase its ability to electrotransform bacterial cells, we used a strain of *E. coli* (NEB10β) with high transformation efficiency. The M-TUBE devices employed for most experiments conducted in this study were comprised of a 500-μm diameter tube and 3-mm gap, and were supplied with a voltage of $\pm$2.50 kV or 5.00 kV_{pp} (peak-to-peak AC signal, square wave), which leads to a field strength of 8.33 kV/cm within the microchannel. Cuvettes with 2-mm gaps were used to perform electroporation at different voltages for as a control. We first confirmed that the flow field (or flow shear stress) along the tube does not by itself lead to genetic transformation. In the absence of an electric field, simply flowing cells through M-TUBE at fluid velocities ranging from 148 mm/s (1.8 mL/min) to 2664 mm/s (32.6 mL/min) did not result in any transformation events (Figure 4-1f, bottom). By contrast, once a sufficient electric field was established within M-TUBE, colonies were obtained across the entire range of flow rates tested (Figure 4-1f, top), with transformation efficiencies ranging from 108-1010 CFUs/μg of DNA (Figure 4-1g). A reduction in electroporation efficiency was observed as the fluid velocity was increased. This trend was expected because the residence time decreases as the flow rate increases, hence cells are exposed to the electric field for a shorter duration at higher flow rates. Despite the lower efficiency at higher flow rates, the overall efficiency obtained using the M-TUBE device was at least one order of magnitude higher than that obtained using cuvettes with the same field strength (8.33 kV/cm). We also note that, compared to cuvettes (typically used at 10-15 kV/cm), M-TUBE was able to produce a comparable efficiency using a lower electric field. The finding that M-TUBE outperforms cuvettes in terms of transformation efficiency may be due to a synergistic effect of the flow field and the electric field [155].

Given the strong dependence of transformation efficiency on field strength in cuvette-based electroporation, we next evaluated how M-TUBE performs across field strengths. Compared to cuvette-based electroporation at 8.33 kV/cm, regardless of the supplied

field strength, M-TUBE exhibited higher transformation efficiencies across the range of flow rates tested (Figure 4-2a, left). This finding indicates that M-TUBE can either achieve the same efficiency with lower field strengths or higher efficiency with the same field strength. Moreover, electroporation efficiencies with M-TUBE had a smaller standard deviation than those obtained with cuvette-based electroporation. Thus, M-TUBE provides several benefits compared with cuvettes in addition to its high-volume capability.

Most of our M-TUBE electroporation experiments were carried out using an electric field generated with alternating current (AC) rather than direct current (DC). With DC fields, M-TUBE also exhibited higher electroporation efficiency than cuvettes using the same field strength or comparable efficiency using a lower field strength, although efficiency and reproducibility with DC fields were overall lower than when using AC fields (Figure 4-3). To determine whether M-TUBE transformation efficiency depends on AC field frequency, we conducted electroporations across five fluid velocities in the range 148-1184 mm/s with a distinct frequency (50, 100, 200, 300, 400 Hz) for each fluid velocity so that cells flowing through the microchannel were exposed to only a single pulse (Figure 4-4). For a comparison, an electroporation was also carried out with at a common frequency (400 Hz) for all fluid velocities tested. Electroporation efficiency was largely independent of AC field frequency (Figure 4-4). This result contrasted with a previous study that observed frequency dependence [156], potentially due to differences in channel geometry. Regardless, our findings highlight the flexibility of M-TUBE.

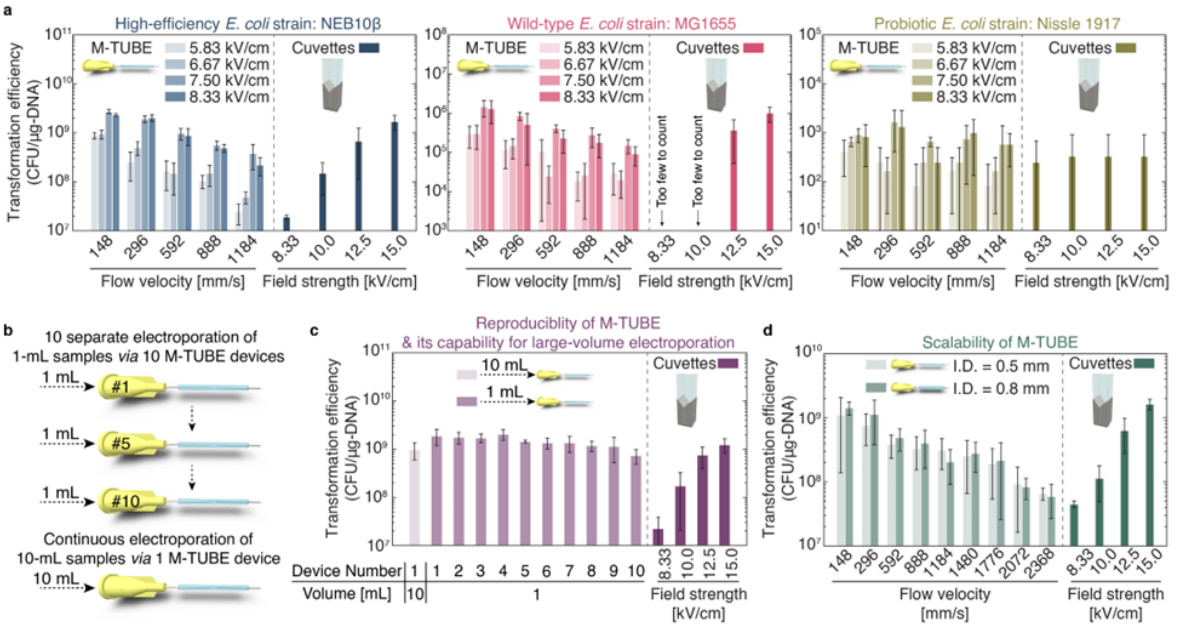


Figure 4-2: **The M-TUBE device exhibits higher efficiency than cuvettes across *E. coli* strains, is reproducible, and maintains high efficiency across tubing sizes.** (a) Comparison of M-TUBE device performance when transforming the high-efficiency strain NEB10β, the wild-type strain MG1655, and the probiotic strain Nissle 1917 across voltages and fluid velocities. M-TUBE outperformed cuvettes at an equivalent electric field strength for all strains. Data represent the average ($n \geq 3$) and error bars represent 1 standard deviation. (b) Schematic of the experiment comparing 10 separate 1 mL electroporations and 1 continuous electroporation of a 10 mL sample (c) Transformation efficiency for the experiments in (b) demonstrates that sample volume can be increased without compromising efficiency. The same batch of cells was used to conduct cuvette-based electroporation as a comparison. Data represent the average ($n \geq 3$) and error bars represent 1 standard deviation. (d) Transformation efficiency was similar across 0.5-mm and 0.8-mm diameter M-TUBE devices. For M-TUBE devices, a voltage of ± 2.50 kV (AC field) was applied, which results in an electric field of 8.33 kV/cm. Data represent the average ($n \geq 3$) and error bars represent 1 standard deviation.

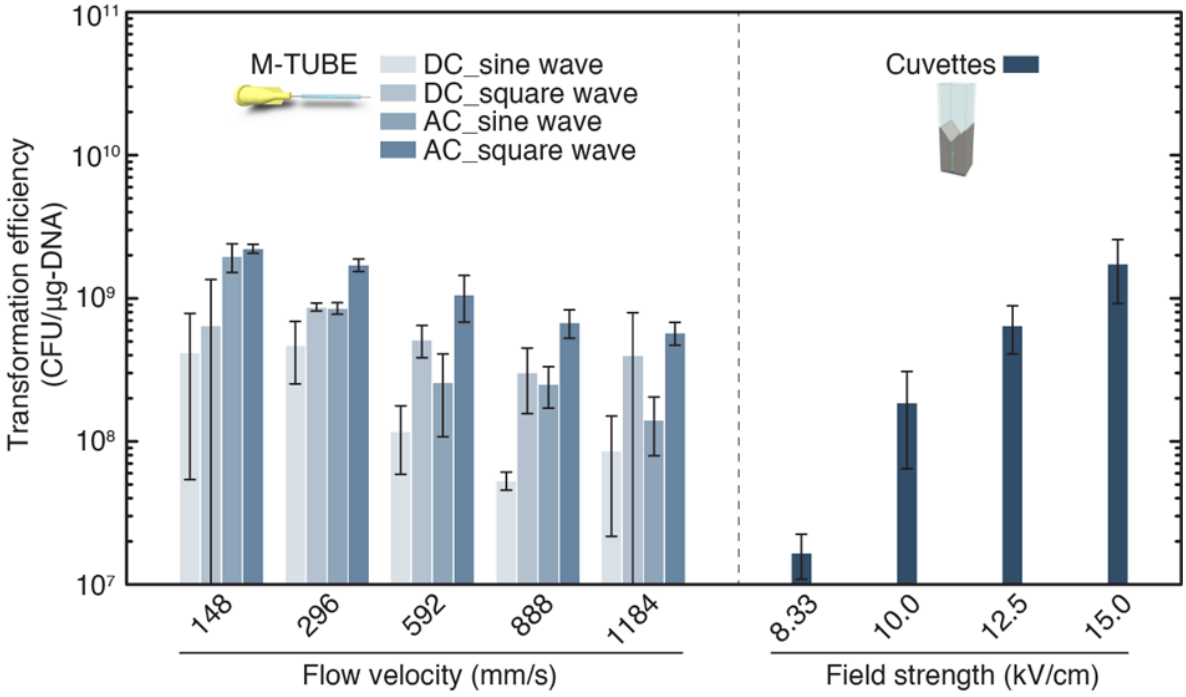


Figure 4-3: **M-TUBE device performance is higher using AC fields compared with DC fields.** Using DC fields, M-TUBE devices achieved higher transformation efficiency than cuvettes using the same field strength or comparable efficiency using a lower field strength. Overall, electroporation efficiency and reproducibility were lower using DC fields compared with AC fields. For M-TUBE devices, a voltage of ± 2.50 kV (AC field) or 2.50 kV (DC field with a duty cycle of 95%) was applied, which results in an electric field of 8.33 kV/cm. Data represent the average ($n \geq 3$) and error bars represent 1 standard deviation.

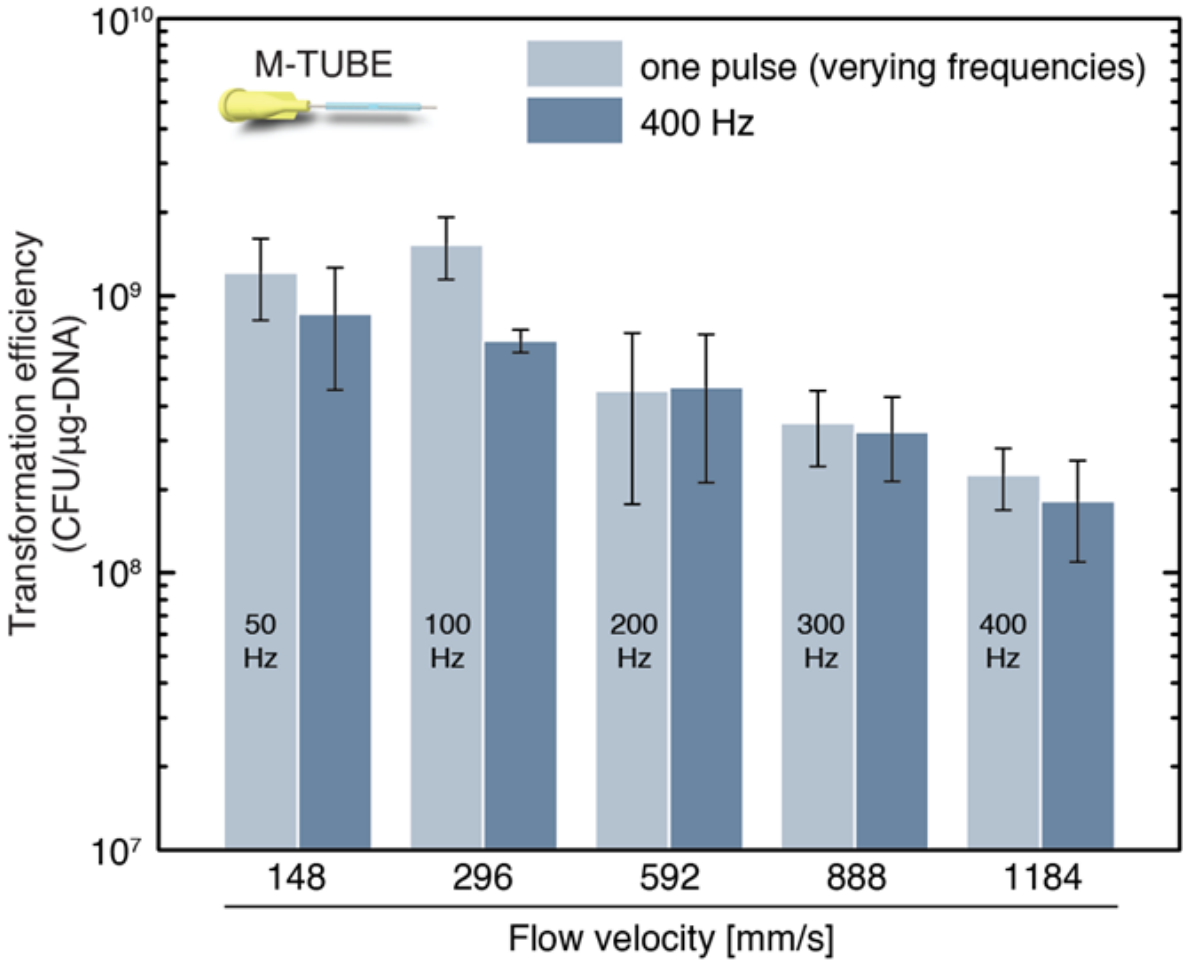


Figure 4-4: **Dependence of M-TUBE transformation efficiency on the frequency of the applied AC field.** With M-TUBE devices, electroporation efficiency was largely independent of the applied AC field frequency. For M-TUBE devices, a voltage of ±2.50 kV (AC field) was applied, which results in an electric field of 8.33 kV/cm. Data represent the average (n $\geq$ 3) and error bars represent 1 standard deviation.

4.3.2.1 Assembly has negligible effect on reproducibility of M-TUBE

Since M-TUBE is hand-assembled, small fluctuations in the microchannel length are inevitable across independently assembled M-TUBE devices (even assembled by the same user). Given that the field strength is defined as the ratio of the applied voltage to the microchannel length, we sought to evaluate if the field strength differs significantly

across identical but separately assembled M-TUBE devices, thereby causing variation in electroporation performance for NEB10β cells (Figure 4-2b, top). We concurrently carried out electroporation of a large-volume sample (10 mL) to demonstrate the capacity of M-TUBE for high-volume electroporation (Figure 4-2b, bottom), from which we were able to determine if there is a substantial difference in transformation efficiency between multiple small volume electroporation experiments and continuous flow large volume electroporation. The variation across 10 M-TUBE devices was insignificant and negligible, and each of the tested devices outperformed cuvettes regardless of the field strength (Figure 4-2c), confirming that assembly has negligible impact on the reproducibility of the M-TUBE.

Furthermore, M-TUBE was able to electroporate the entire 10-mL sample at a flow rate of 3.6 mL/min with efficiency higher than or comparable to cuvettes (Figure 4-2c) and the transformation efficiency for 10 mL of continuous electroporation was not significantly different from that of 10 separate 1-mL experiments. Continuous electroporation of 10 mL is equivalent to 100 individual 0.1-mL cuvette-based electroporations, for which the configuration of M-TUBE that we tested would shorten the entire electroporation time by two to three orders of magnitude (depending on the flow rate). Put in other terms, M-TUBE can process two to three orders of magnitude more volume of sample in a given period of time compared with cuvettes. In terms of cost, M-TUBE is at least 10-fold cheaper than cuvettes. Moreover, using M-TUBE for large-volume bacterial electroporation can also circumvent the need for manual pipetting by flowing the electroporated sample directly into recovery medium, thereby decreasing total processing time and potentially improving cell viability and transformation efficiency. Taken together, these features make M-TUBE an ideal candidate for large-volume bacterial electroporation.

4.3.3 M-TUBE exhibits comparable or better efficiency compared with cuvettes across *E. coli* strains

Motivated by the successful transformation of *E. coli* NEB10β, M-TUBE was then tested on the wild-type strain *E. coli* MG1655, which typically has lower transformation efficiency than NEB10β. The results show that M-TUBE maintained higher efficiency than cuvettes for MG1655 (Figure 4-2, middle). With a field strength of 8.33 kV/cm, M-TUBE yielded efficiencies at least two orders of magnitude higher than cuvettes; even though cuvettes were supplied with a field strength of 10 kV/cm, the number of successfully transformed colonies was too low to reliably enumerate. To further test M-TUBE performance on *E. coli* strains, we used M-TUBE to electroporate the probiotic strain Nissle 1917 [144, 145]. While both M-TUBE and cuvettes exhibited much lower electroporation efficiencies for Nissle 1917 compared with MG1655, M-TUBE was comparably efficient to cuvettes and showed slightly better reproducibility (Figure 4-2a, right). Moreover, the ability of M-TUBE to process arbitrarily large sample volumes in a continuous fashion means that a desired number of transformed cells of a low-efficiency strain such as Nissle can be obtained with M-TUBE simply by processing a sufficiently large volume. Conversely, using cuvettes for the same goal would be expensive and technically challenging. Overall, M-TUBE showed robust performance across *E. coli* strains with a wide range of electroporation efficiencies, with performance and reproducibility higher than or comparable to cuvette-based electroporation.

Our next goal was to evaluate the ability to scale up the M-TUBE to process even larger volume samples. To this end, the performance of the M-TUBE device with three different inner diameters was compared (500, 800, and 1600 μm, with the size of syringe needles altered accordingly) (Figure 4-2d, 4-5). As long as the gap and the fluid velocity were held fixed, M-TUBE devices with different diameters maintained a high electroporation efficiency for NEB10β cells and outperformed cuvettes. With the same fluid velocity, an M-TUBE device with larger diameter would enable processing larger volumes: with a diameter of 1600 /mum, an average fluid velocity of 592 mm/s

allows for electroporation of ∼70 mL/min, several orders of magnitude more than what is possible with cuvettes. These results again demonstrate the capabilities of M-TUBE for large-volume bacterial electroporation, and confirm that M-TUBE can be readily scaled up without compromising efficiency simply by changing the tubing and syringe needles sizes while maintaining fluid velocity.

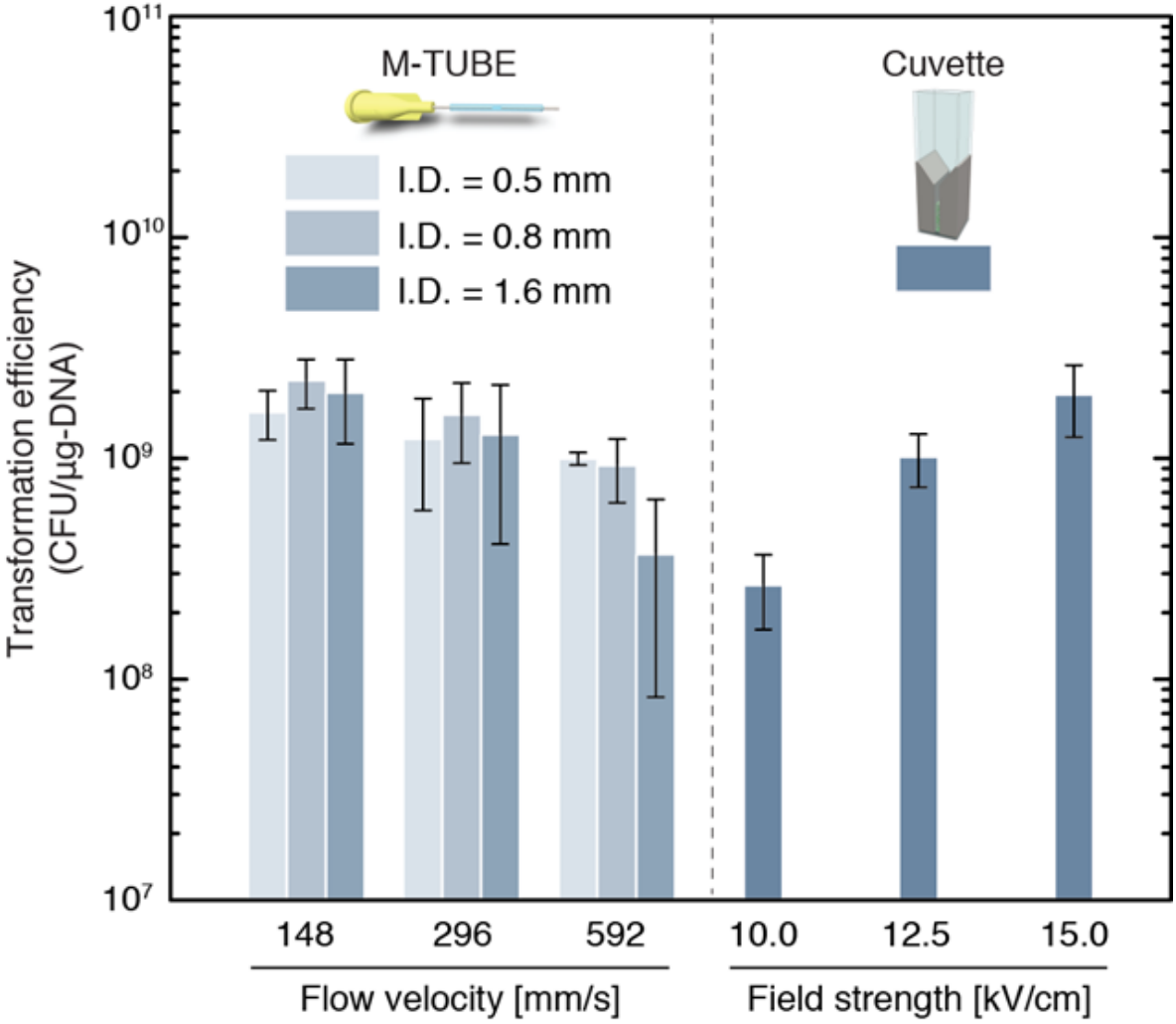


Figure 4-5: **Transformation efficiency is maintained across M-TUBE devices with different diameters.** To further evaluate the scalability of M-TUBE, M-TUBE devices made using plastic tubing with 0.5-mm, 0.8-mm, and 1.6-mm inner diameters and compared to conventional cuvettes. A voltage of $/pm2.50$ kV (AC field) was applied to each M-TUBE device, resulting in an electric field of 8.33 kV/cm. The same batch of cells was used to conduct electroporation with 0.2-mm cuvettes and various voltages as a comparison. Data represent the average ($n \geq 3$) and error bars represent 1 standard deviation.

4.3.4 Generation of a transposon mutant library in an anaerobic gut commensal with M-TUBE

As a demonstration of the utility of M-TUBE in other organisms, we sought to use the system to generate a set of transposon insertion mutants in a human gut commensal. Many of these organisms are obligate anaerobes and hence require more complex handling during growth, washing, and electroporation. We assembled the M-TUBE electroporation platform inside an anerobic chamber and ran an experiment to generate a small-scale transposon insertion pool in *Bifidobacterium longum* subsp. *longum* NCIMB8809. *B. longum* species are used as probiotics and are actively investigated for their health-promoting effects [157]. To identify optimal electroporation conditions for maximizing transposome delivery, we first electroporated *B. longum* NCIMB8809 cells with the pAM5 plasmid (Figure 4-6a). As with *E. coli*, M-TUBE plasmid transformation efficiency was comparable to or higher than that of cuvettes for *B. longum* (Figure 4-6a). With the optimal electroporation conditions, *B. longum* cells were successfully transformed with *in vitro*-assembled EZ-Tn5 transposomes, demonstrating its utility both in an anaerobic chamber and for high-throughput transposon mutagenesis (Figure 4-6b, 4-6c). Like plasmids, M-TUBE transposome electroporation efficiency was comparable to or higher than that of cuvettes. Transposon sequencing of the transformants revealed >2,000 unique transposition events spread across the genome (Figure 4-6c). Given these encouraging results, we expect that a scaled-up transformation protocol will produce a transposon pool of sufficient diversity for future chemical-genomic investigation using barcode sequencing [126, 158, 159]. Furthermore, we expect M-TUBE should have wide applicability for generation of libraries of thousands of transposon mutants, even in bacterial species with complex growth requirements.

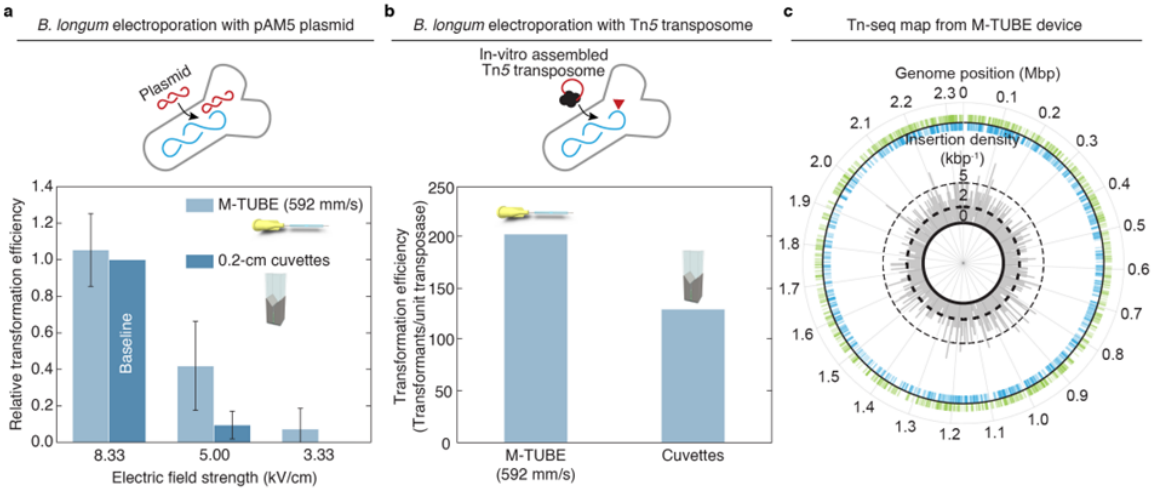


Figure 4-6: **M-TUBE efficiently transforms anaerobic bacteria and enables transposon insertion mutagenesis.** (a) Comparison of M-TUBE performance during electrotransformation of *B. longum* NCIMB8809 with the plasmid pAM5 at various electric field strengths. For M-TUBE devices, voltages of ±2.50, ±1.50, and ±1.00 kV (AC field) were applied to produce electric fields of 8.33, 5.00, and 3.33 kV/cm, respectively. A fluid velocity of 592 mm/s was used for the M-TUBE device because the ∼5 ms residence time with an M-TUBE inner diameter of 0.5 mm is similar to the time constant observed in cuvette electroporation (5.2-5.6 ms). Data represent the average ($n \geq 3$) and error bars represent 1 standard deviation. (b) Comparison of M-TUBE performance during electrotransformation of *B. longum* NCIMB8809 with Tn5 transposome. For the M-TUBE device, a field strength of 8.33 kV/cm and fluid velocity of 592 mm/s were used, motivated by the results in (a). (c) The transposon insertions recovered from Tn5 transposome electroporation are spread approximately uniformly across the *B. longum* NCIMB8809 genome. The locations of individual mapped insertions are recorded on the outer circle. Green ticks on the outside indicate insertions in the positive (+) orientation, blue ticks on the inside indicate insertions in the negative (-) orientation. The insertion density (kbp-1) (both positive and negative orientation) is plotted in 1-kbp bins on the inner circle. Transposon insertions are distributed throughout the genome in both the positive and negative orientations, indicating that *B. longum* NCIMB8809 can be transformed by Tn5 transposomes using M-TUBE without major insertional bias.

4.4 Conclusions

Taken together, the disposable, fabrication-free M-TUBE device can process large volumes of bacterial cells with dramatically reduced processing time and effort, without

compromising transformation efficiency and cell viability. Due to the simplicity of its fabrication and the wide availability of its components, M-TUBE presents an electroporation strategy that can be immediately implemented in the microbiology community. The flexibility that M-TUBE offers in tuning electroporation conditions such as field strength and residence time make the device a powerful tool for working with hard-to-transform strains. Given the relatively high transformation efficiency compared with cuvettes and its ability to deal with both small and large volumes, M-TUBE has the potential to be a viable alternative to cuvettes and an indispensable tool for applications requiring large volumes such as the creation of mutant libraries.

Chapter 5

Conclusions and Future Work

In this thesis, we have addressed the limitations of traditional cuvette-based electroporation methods in bacterial genetic engineering by developing high-throughput solutions that enhance efficiency and consistency. These solutions include an automated system for high-throughput bacterial genetic engineering, a colony counting tool (MCount), and a large-volume electroporation system (M-TUBE).

The automated system presented in this study offers a significant advancement in bacterial genetic engineering. It features a microfluidic device that enables column-by-column electroporation, resulting in a 96-well electroporation process completed in just 5 minutes. Through experimental evaluations, we have demonstrated the system's efficiency, reproducibility, and compatibility with liquid handling robots. This automated system provides a reliable and scalable approach for rapid and efficient genetic manipulation.

While the automated system shows great potential, there are several areas that could be further investigated and improved. One important aspect is gaining a better understanding of the mechanism of electroporation to enhance the system. It is crucial to explore the factors that contribute to arcing and develop strategies to prevent it, as this issue can hinder the efficiency and reliability of the system. Additionally, leveraging the flow dynamics, such as the use of inertia force, can be explored to further optimize

the transformation efficiency.

Another avenue for improvement lies in refining the design of the microfluidic device. By optimizing and improving the device's design, it can become even more reliable, repeatable, user-friendly, and cost-effective for broader utilization. This includes refining the dimensions, materials, and fabrication processes to enhance the performance and compatibility of the device.

Further optimization of the power system and pulse delivery system is also essential. Fine-tuning these components and achieving better synchronization can lead to improved efficiency and accuracy in delivering electric pulses to the cells. Additionally, efforts can be made to enhance the compatibility of the automated system with other brands of liquid handling robots, such as Tecan and Hamilton, broadening its accessibility and integration with existing laboratory infrastructure.

The development of a user-friendly operating interface is another important future direction. This could involve the creation of a graphical user interface (GUI) or a Python-based interface that simplifies the operation of the automated system and improves communication with the controller with minimal delay. Such improvements will enhance the overall user experience and increase the system's usability and adoption in different laboratory settings.

To broaden the applicability of the automated system, it would be valuable to develop specific protocols for commonly studied microbes. Additionally, the system can be adapted to quickly screen unknown microbes, enabling rapid and efficient characterization and manipulation of diverse microbial strains.

Furthermore, exploring the deployment of the automated system in anaerobic environments for anaerobic microbes presents an interesting future prospect. Adapting the system to work under anaerobic conditions would open new possibilities for studying and manipulating anaerobic bacteria, expanding the scope of applications and research opportunities.

By addressing these areas, the automated system can continue to evolve and make significant contributions to the field of bacterial genetic engineering and synthetic biol-

ogy. With its potential applications in the generation of mutant libraries, DNA library screening, and high-throughput gene expression analysis, the automated system has the capacity to revolutionize the field and drive advancements in diverse areas of research and biotechnology.

The counting tool, MCount, developed in this study, offers an efficient solution for high-throughput colony counting in bacterial electroporation. By utilizing contour information and employing advanced algorithms, MCount achieves accurate and reliable colony quantification, outperforming existing solutions. Its ability to handle merged colonies makes it particularly valuable for precise assessment of transformation efficiency in genetic engineering experiments.

Although we have introduced various hyperparameter optimization methods that require minimal labeled or even unlabeled datapoints, further improvements are needed to enhance the user-friendliness of the counting tool. Developing a user interface that is intuitive and streamlined will simplify the operation and enable biologists to easily incorporate MCount into their research workflow. By making the tool more user-friendly, it will be more accessible and widely adopted by researchers who prefer simple and straightforward tools for their experiments.

In addition to colony counting, the counting tool has the potential to be utilized for colony classification tasks. Leveraging its robust contour-based counting algorithms, MCount can be extended to classify colonies based on different morphological characteristics, such as size, shape, and color. This expanded functionality opens up new possibilities for automated analysis and characterization of bacterial colonies, providing valuable insights for various genetic engineering and microbiology applications.

The large volume system, M-TUBE, represents a significant advancement in bacterial gene delivery by providing a scalable and efficient solution for large volume transformations. With its disposable and user-friendly microfluidic device, M-TUBE overcomes the limitations of traditional cuvettes and enables high-efficiency gene delivery in volumes exceeding 10 mL. By optimizing the device design and operation, M-TUBE offers a reliable and cost-effective approach for large volume bacterial transformation.

The applications of the M-tube system are diverse and impactful. Once the optimal electroporation conditions have been determined using the automated system, M-tube can be employed for large volume transformations in various genetic engineering applications. It enables the rapid and efficient delivery of genetic material into bacterial cells at a scale that was previously challenging to achieve. This opens up new possibilities for generating mutant libraries, screening DNA libraries, and performing high-throughput gene expression analysis, among other applications. The ability to perform large volume transformations using M-tube enhances the efficiency and scalability of these genetic engineering workflows, accelerating research and development in fields such as synthetic biology, biotechnology, and drug discovery.

In conclusion, this thesis has successfully addressed the limitations of traditional cuvette-based electroporation methods in bacterial genetic engineering by introducing high-throughput solutions that enhance efficiency and consistency. The developed automated system, colony counting tool, and large-volume electroporation system offer promising avenues for rapid and reliable genetic manipulation. The advancements made in this research contribute to the field of bacterial genetic engineering and open doors for further improvements and future applications in various domains of biotechnology and synthetic biology.

Appendix A

Discussion on the Distribution of Colony Numbers

Although it is theoretically predicted that colony numbers should follow a Poisson distribution, the distribution may deviate from Poisson and become closer to a Normal distribution due to noise induced by liquid handling operations. We illustrate this phenomenon using our benchmark data.

Our benchmark consists of ten *E. coli* plates, and we investigated the distribution of all ten plates using hypothesis testing. We used the Kolmogorov–Smirnov (KS) test to test the null hypothesis (H_0) that the distribution of 96 colony numbers follows a Poisson or Normal distribution with a significant level of 0.05. The p-values of the KS test (Table S2) show that we cannot reject the null hypothesis that the distribution follows either a Poisson or Normal distribution for any of the datasets. However, we cannot conclude that they follow Poisson or Normal distribution, either. As the p-values of KS test (Poisson) are smaller than those of KS test (Normal) for almost all datasets, we conclude that our benchmark is closer to a Normal distribution than a Poisson distribution.

We also conducted Poisson Dispersion tests (Table S2) and found that some of the datasets still have the equidispersion property, despite being closer to a Normal

distribution.

Dataset	Mean	Variance	KS (Normal)	KS (Poisson)	Poisson Dispersion
1	58.79	85.24	0.6604	0.6539	0.0747
2	20.64	53.21	0.7431	0.3207	2.11e-07*
3	48.01	42.07	0.8749	0.8492	0.6450
4	15.51	28.59	0.6963	0.4516	0.0021*
5	9.38	9.89	0.8258	0.7202	0.4385
6	34.60	43.73	0.3595	0.0523	0.0658
7	22.33	21.97	0.6214	0.0947	0.5507
8	56.69	122.76	0.2082	0.0974	7.95e-06*
9	38.28	57.64	0.8063	0.3398	0.0087*
10	15.19	23.26	0.3048	0.3699	0.0017*

Table A.1: **Mean, variance, and p-values of tests for 10 datasets.** The mean and variance values were calculated from the 96 colony counts of each dataset. The first column represents the dataset index. The second column represents the mean value of the 96 colony counts, and the third column represents the variance of the 96 colony counts. The 4th and 5th columns show the p-values for Kolmogorov–Smirnov (KS) tests for Normal and Poisson distributions, respectively, under the significant level of 0.05, where no rejection is made for any datasets. The 6th column shows the p-values for Poisson Dispersion tests, and datasets whose null hypotheses are rejected under the significant level of 0.05 are denoted with a star symbol *.

Appendix B

Source code of MCount

B.1 Imports

```
%matplotlib inline
import time
from nd2reader import ND2Reader
import matplotlib.pyplot as plt
import cv2
import os
import numpy as np
import math
import pulp
import numpy as np
from pprint import pprint
import pandas as pd
import datetime
from scipy import optimize
from scipy.ndimage.filters import maximum_filter
```

B.2 Auxiliary Function Definitions

```
filename = 'Example/single image example.jpg'
destination_filename = 'Example/single image example_result.jpg'
lam = 30           # Hyperparameter: Lambda
d_t = 0.5          # Hyperparameter: d

# Find the first type of candidate circle via local maxima
def find_maxima(img):
    kernel = cv2.getStructuringElement(cv2.MORPH_RECT, (3,3))
    sub_opening = cv2.morphologyEx(img, cv2.MORPH_OPEN, kernel, iterations=2)
    dist = cv2.distanceTransform(sub_opening, cv2.DIST_L2, 3)

    maxima = (dist > 0) * (dist == maximum_filter(dist, 3, mode='constant', cval=math.inf))
    res = np.where(1 == maxima)

    centroids = list(map(list, zip(list(res[1]), list(res[0]), list(dist[res]))))
    count = len(centroids)
    m = maxima
    return count, m, centroids, dist

# Define the distance from Point j to Line jk
def distance(cnt_i, cnt_j, cnt_k):
    x1, y1 = cnt_i[0][0], cnt_i[0][1]
    line_start_x, line_start_y = cnt_j[0][0], cnt_j[0][1]
    x2, y2 = cnt_k[0][0], cnt_k[0][1]
    array_longi = np.array([x2-x1, y2-y1])
    array_trans = np.array([x2-line_start_x, y2-line_start_y])
    array_temp = (float(array_trans.dot(array_longi)) / array_longi.dot(array_longi))
    array_temp = array_longi.dot(array_temp)
    return np.sqrt((array_trans - array_temp).dot(array_trans - array_temp))

def clockwise_angle(v1, v2):
    x1, y1 = v1
    x2, y2 = v2
    dot = x1*x2+y1*y2
    det = x1*y2-x2*y1
    theta = np.arctan2(det, dot)
    return theta

# Calculate the distance of each 2D points from the center (xc, yc)
def calc_R(x,y, xc, yc):
    return np.sqrt((x-xc)**2 + (y-yc)**2)

# Calculate the algebraic distance between the data points and the mean circle centered at c=(xc, yc)
def f(c, x, y):
    Ri = calc_R(x, y, *c)
    return Ri - Ri.mean()

# Find the second type of candidate circle via least square fitting
def leastsq_circle(x,y):
    x_m = np.mean(x)
    y_m = np.mean(y)
    center_estimate = x_m, y_m
    center, ier = optimize.leastsq(f, center_estimate, args=(x,y))
    xc, yc = center
    Ri       = calc_R(x, y, *center)
    R        = Ri.mean()
    residu   = np.sum((Ri - R)**2)
    return int(xc), int(yc), R, residu
```

B.3 Main Function Definition

```
def seg_counting(labels, xxxxxx, img, lam):
    img_c = np.uint8(255*(labels==xxxxxx))
    count, m, centroids, dist = find_maxima(np.uint8(img_c))
    contours, h = cv2.findContours(img_c,cv2.RETR_TREE, cv2.CHAIN_APPROX_NONE)
    contours_split = []
    for cnt in contours:
        turning_points = [cnt[0]]
        i, k = 0, 2
        while k < len(cnt):
            d_max, ind = 0, 0
            for j in range(i+1, k):
                d = distance(cnt[i], cnt[j], cnt[k])
                if d > d_max:
                    d_max = d
                    ind = j
            if d_max > d_t:
                turning_points = np.append(turning_points, [cnt[ind]], axis=0)
                i, k = ind, ind+2
            else:
                k = k + 1

        angles = []
        concave_points = [turning_points[0]]

        for i in range(len(turning_points)-2):
            angles.append(clockwise_angle([turning_points[i+1][0][0]-turning_points[i][0][0],
                                           turning_points[i+1][0][1]-turning_points[i][0][1]],
                                          [turning_points[i+2][0][0]-turning_points[i+1][0][0],
                                           turning_points[i+2][0][1]-turning_points[i+1][0][1]])/math.pi*180)
            if angles[-1] > 0:
                concave_points = np.append(concave_points, [turning_points[i+1]], axis=0)

        index = []
        k = 0

        for i in range(len(cnt)):
            x, y = concave_points[k][0][0], concave_points[k][0][1]
            x_now, y_now = cnt[i][0][0], cnt[i][0][1]
            if x==x_now and y==y_now:
                index.append(i)
                k = k+1
            if k == len(concave_points):
                break

        if contours_split == []:
            contours_split = [[cnt[index[x]:index[x+1]]] for x in range(len(index)-1)]
        else:
            contours_split = contours_split + [[cnt[index[x]:index[x+1]]] for x in range(len(index)-1)]
        contours_split.append([cnt[index[-1]:]])

    for ss in range(len(contours_split)):
        x = [i[0][0] for i in contours_split[ss][0]]
        y = [i[0][1] for i in contours_split[ss][0]]
        if len(x) <= 3:
            continue
        xc, yc, _, _ = leastsq_circle(x,y)
        y_max, x_max = m.shape
        if yc < y_max and yc >= 0 and xc < x_max and xc>= 0:
            if img_c[yc][xc] > 0 and dist[yc][xc] > 0:
                count = count + 1 - int(m[yc][xc])
                m[yc][xc] = 1
                centroids.append((xc, yc, dist[yc][xc]))

    # d[i][j] defines the distance from Contour i to Candidate circle j
    d = np.zeros((len(contours_split), count))
    dd = [math.inf for i in contours_split]
    index_d = [-1 for i in contours_split]
    for i in range(len(contours_split)):
        for j in range(count):
            for z in range(len(contours_split[i][0])):
                d[i][j] = d[i][j] + abs(math.dist([contours_split[i][0][z][0][0], contours_split[i][0][z][0][1]],
                                                  [centroids[j][0], centroids[j][1]]) - centroids[j][2])
            if dd[i] > d[i][j]:
                dd[i] = d[i][j]
                index_d[i] = j
```

```
    # Solve the optimization problem
    d = np.array(d)
    prob = pulp.LpProblem('contour_mapping', sense=pulp.LpMinimize)
    var_x = [[pulp.LpVariable('x{:5d}{:5d}'.format(i, j), cat=pulp.LpBinary) for j in range(count)]
             for i in range(len(contours_split))]
    flatten = lambda x: [y for l in x for y in flatten(l)] if type(x) is list else [x]
    var_z = [pulp.LpVariable('z{:5d}'.format(j), cat=pulp.LpBinary) for j in range(count)]
    prob += pulp.lpDot(flatten(var_x)+var_z, np.append(d.flatten(), [lam for j in range(count)]))
    for i in range(len(contours_split)):
        prob += (pulp.lpSum(var_x[i]) == 1)
    for j in range(count):
        prob += (pulp.lpSum([var_x[i][j] for i in range(len(contours_split))] + [-var_z[j]]) >= 0)
        prob += (pulp.lpSum([var_x[i][j] for i in range(len(contours_split))] + [-len(contours_split)*var_z[j]]) <= 0)

    # Solver is COIN_CMD
    prob.solve(solver = pulp.getSolver('COIN_CMD'))

    res = {'objective':pulp.value(prob.objective),
           'var_x': [[pulp.value(var_x[i][j]) for j in range(count)] for i in range(len(contours_split))],
           'var_z': [pulp.value(var_z[j]) for j in range(count)]}

    final_img = np.uint8(img)
    img_seg = np.uint8(img*np.array([(labels==xxxxxx),(labels==xxxxxx),(labels==xxxxxx)]).transpose(1, 2, 0))
    for i in range(count):
        if res['var_z'][i]:
            img = cv2.circle(final_img, (centroids[i][0], centroids[i][1]), int(centroids[i][2]), (0, 0, 255), 1)
            img_seg = cv2.circle(img_seg, (centroids[i][0], centroids[i][1]), int(centroids[i][2]), (0, 0, 255), 1)

    return int(sum(res['var_z'])), final_img, img_seg
```

B.4 Execution and Results

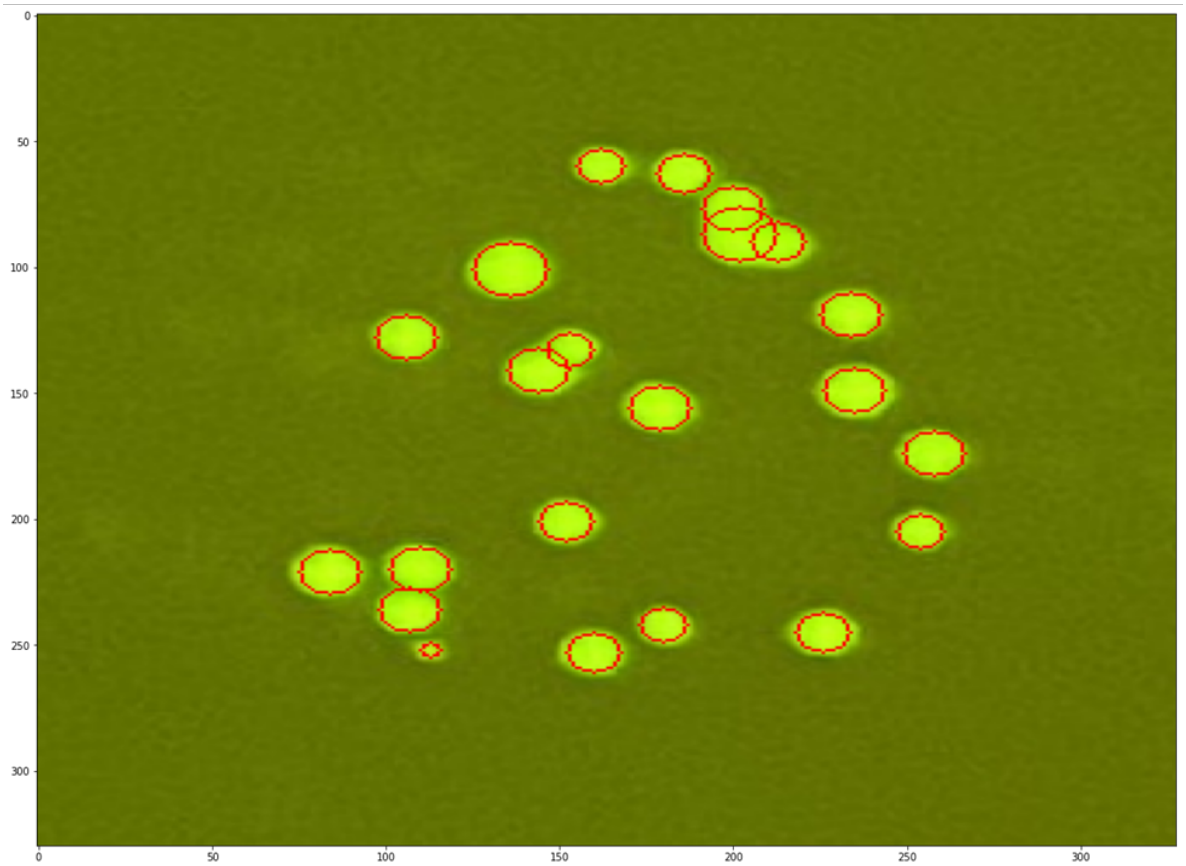
```
img = cv2.imread(filename)
img = img.copy()
gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
ret, thresh = cv2.threshold(gray, 0, 255, cv2.THRESH_OTSU)
num_labels, labels, stats, centroids = cv2.connectedComponentsWithStats(thresh)

counting_result = 0

for xxxxxx in range(1, num_labels):
    counted_cell_number, img, img_seg = seg_counting(labels, xxxxxx, img, lam)
    counting_result = counting_result + counted_cell_number

print("Saved as " + destination_filename)
cv2.imwrite(destination_filename, img)
print("Counting Result (" + chr(955) + "=" + str(lam) + "): ", counting_result)
plt.figure(figsize = (20,15))
plt.imshow(img[...,::-1], aspect = 'auto')
plt.show()
```

```
Saved as Example/single image example_result.jpg
Counting Result (λ=30):  22
```



Bibliography

[1] Terms and acronyms. *U.S. Environmental Protection Agency online*.

[2] Liangxue Lai, Donna Kolber-Simonds, Kwang-Wook Park, Hee-Tae Cheong, Julia L Greenstein, Gi-Sun Im, Melissa Samuel, Aaron Bonk, August Rieke, Billy N Day, et al. Production of α-1, 3-galactosyltransferase knockout pigs by nuclear transfer cloning. *Science*, 295(5557):1089–1092, 2002.

[3] H Boyer et al. First successful laboratory production of human insulin announced. news release. Technical report, Genentech 1978-09-06, 1978.

[4] Ilanit Tof. Recombinant dna technology in the synthesis of human insulin. *Little Tree Publishing, retrieved*, 19:2010, 1994.

[5] Jens Nielsen. Production of biopharmaceutical proteins by yeast: advances through metabolic engineering. *Bioengineered*, 4(4):207–211, 2013.

[6] A. S. Bawa and K. R. Anilakumar. Genetically modified foods: safety, risks and public concerns-a review. *Journal of Food Science and Technology-Mysore*, 50(6):1035–1046, 2013.

[7] C James and AF Krattiger. Global review of the field testing and commercialization of transgenic plants, 1986 to 1995: The first decade of crop biotechnology. isaaa brief no. 1. *Ithaca, NY, International Service for the Acquisition of Agri-Biotech Applications, 31 p.(isaaa-brief-01-1996. pdf)*, 1996.

[8] K. H. Maurer. Detergent proteases. *Current Opinion in Biotechnology*, 15(4):330–334, 2004.

[9] Niels-Bjarne Woods, Virginie Bottero, Manfred Schmidt, Christof Von Kalle, and Inder M Verma. Gene therapy: therapeutic gene causing lymphoma. *Nature*, 440(7088):1123, 2006.

[10] Clare E Thomas, Anja Ehrhardt, and Mark A Kay. Progress and problems with the use of viral vectors for gene therapy. *Nature Reviews Genetics*, 4(5):346, 2003.

[11] R. Waehler, S. J. Russell, and D. T. Curiel. Engineering targeted viral vectors for gene therapy. *Nature Reviews Genetics*, 8(8):573–587, 2007.

[12] Conchita Tros de Ilarduya, Yan Sun, and Nejat Düzgüneş. Gene delivery by lipoplexes and polyplexes. *European journal of pharmaceutical sciences*, 40(3):159–170, 2010.

[13] Matxalen Llosa, F Xavier Gomis-Rüth, Miquel Coll, and Fernando de la Cruz. Bacterial conjugation: a two-step mechanism for dna transport. *Molecular microbiology*, 45(1):1–8, 2002.

[14] Sang Kyung Kim, Jae Hyun Kim, Kwang Pyo Kim, and Taek Dong Chung. Continuous low-voltage dc electroporation on a microfluidic chip with polyelectrolytic salt bridges. *Analytical chemistry*, 79(20):7761–7766, 2007.

[15] Shengnian Wang, Xulang Zhang, Weixiong Wang, and L James Lee. Semicontinuous flow electroporation chip for high-throughput transfection on mammalian cells. *Analytical chemistry*, 81(11):4414–4421, 2009.

[16] Shengnian Wang, Xulang Zhang, Bo Yu, Robert J Lee, and L James Lee. Targeted nanoparticles enhanced flow electroporation of antisense oligonucleotides in leukemia cells. *Biosensors and Bioelectronics*, 26(2):778–783, 2010.

[17] Jun Wang, M Jane Stine, and Chang Lu. Microfluidic cell electroporation using a mechanical valve. *Analytical chemistry*, 79(24):9584–9587, 2007.

[18] Hsiang-Yu Wang and Chang Lu. Microfluidic electroporation for delivery of small molecules and genes into cells using a common dc power supply. *Biotechnology and bioengineering*, 100(3):579–586, 2008.

[19] Tao Geng, Yihong Zhan, Hsiang-Yu Wang, Scott R Witting, Kenneth G Cornetta, and Chang Lu. Flow-through electroporation based on constant voltage for large-volume transfection of cells. *Journal of Controlled Release*, 144(1):91–100, 2010.

[20] Tao Zhu, Chunxiong Luo, Jianyong Huang, Chunyang Xiong, Qi Ouyang, and Jing Fang. Electroporation based on hydrodynamic focusing of microfluidics with low dc voltage. *Biomedical microdevices*, 12(1):35–40, 2010.

[21] Zewen Wei, Deyao Zhao, Xueming Li, Mengxi Wu, Wei Wang, Huang Huang, Xiaoxia Wang, Quan Du, Zicai Liang, and Zhihong Li. A laminar flow electroporation system for efficient dna and sirna delivery. *Analytical chemistry*, 83(15):5881–5887, 2011.

[22] Arash Noori, P Ravi Selvaganapathy, and Joanna Wilson. Microinjection in a microfluidic format using flexible and compliant channels and electroosmotic dosage control. *Lab on a Chip*, 9(22):3202–3211, 2009.

[23] Andrea Adamo and Klavs F Jensen. Microfluidic based single cell microinjection. *Lab on a Chip*, 8(8):1258–1261, 2008.

[24] Jun Wang, Yihong Zhan, Victor M Ugaz, and Chang Lu. Vortex-assisted dna delivery. *Lab on a Chip*, 10(16):2057–2061, 2010.

[25] J Baumgart, W Bintig, A Ngezahayo, S Willenbrock, H Murua Escobar, W Ertmer, H Lubatschowski, and A Heisterkamp. Quantified femtosecond laser based opto-perforation of living gfshr-17 and mth53a cells. *Optics express*, 16(5):3021–3031, 2008.

[26] Ning Bao, Jun Wang, and Chang Lu. Microfluidic electroporation for selective release of intracellular molecules at the single-cell level. *Electrophoresis*, 29(14):2939–2944, 2008.

[27] Ning Bao, Thuc T Le, Ji-Xin Cheng, and Chang Lu. Microfluidic electroporation of tumor and blood cells: observation of nucleus expansion and implications on selective analysis and purging of circulating tumor cells. *Integrative biology*, 2(2-3):113–120, 2010.

[28] Justin K Valley, Steven Neale, Hsan-Yin Hsu, Aaron T Ohta, Arash Jamshidi, and Ming C Wu. Parallel single-cell light-induced electroporation and dielectrophoretic manipulation. *Lab on a Chip*, 9(12):1714–1720, 2009.

[29] Andries Dirk van der Meer, MMJ Kamphuis, AA Poot, J Feijen, and I Vermes. A microfluidic device for monitoring sirna delivery under fluid flow. *Journal of Controlled Release*, 132(3):e42–e44, 2008.

[30] David J Stevenson, Frank J Gunn-Moore, Paul Campbell, and Kishan Dholakia. Single cell optical transfection. *Journal of the Royal Society Interface*, 7(47):863–871, 2010.

[31] Ting-Hsiang Wu, Tara Teslaa, Sheraz Kalim, Christopher T French, Shahriar Moghadam, Randolph Wall, Jeffery F Miller, Owen N Witte, Michael A Teitell, and Pei-Yu Chiou. Photothermal nanoblade for large cargo delivery into mammalian cells. *Analytical chemistry*, 83(4):1321–1327, 2011.

[32] Robert F Marchington, Yoshihiko Arita, Xanthi Tsampoula, Frank J Gunn-Moore, and Kishan Dholakia. Optical injection of mammalian cells using a microfluidic platform. *Biomedical optics express*, 1(2):527–536, 2010.

[33] Albert Tsung-Hsi Hsieh, Nicole Hori, Rustin Massoudi, Patrick Jen-Hao Pan, Hirotaka Sasaki, Yuh Adam Lin, and Abraham P Lee. Nonviral gene vector formation in monodispersed picolitre incubator for consistent gene delivery. *Lab on a Chip*, 9(18):2638–2643, 2009.

[34] Andreas Grodrian, Josef Metze, Thomas Henkel, Karin Martin, Martin Roth, and J Michael Köhler. Segmented flow generation by chip reactors for highly parallelized cell cultivation. *Biosensors and bioelectronics*, 19(11):1421–1428, 2004.

[35] Jon F Edd, Dino Di Carlo, Katherine J Humphry, Sarah Köster, Daniel Irimia, David A Weitz, and Mehmet Toner. Controlled encapsulation of single-cells into monodisperse picolitre drops. *Lab on a Chip*, 8(8):1262–1264, 2008.

[36] Max Chabert and Jean-Louis Viovy. Microfluidic high-throughput encapsulation and hydrodynamic self-sorting of single cells. *Proceedings of the National Academy of Sciences*, 105(9):3191–3196, 2008.

[37] Alfred Vogel, Joachim Noack, G Hüttman, and G Paltauf. Mechanisms of femtosecond laser nanosurgery of cells and tissues. *Applied Physics B*, 81(8):1015–1047, 2005.

[38] Judith Baumgart, Willem Bintig, Anaclet Ngezahayo, Saskia Willenbrock, and Hugo MURUA. Quantified femtosecond laser based opto-perforation of living cells. In *Proceedings of LPM2008-the 9th International Symposium on Laser Precision Microfabrication*, 2009.

[39] Ming Lei, Hanpeng Xu, Hao Yang, and Baoli Yao. Femtosecond laser-assisted microinjection into living neurons. *Journal of neuroscience methods*, 174(2):215–218, 2008.

[40] Motowo Tsukakoshi, S Kurata, Y Nomiya, Y Ikawa, and T Kasuya. A novel method of dna transfection by laser microbeam cell surgery. *Applied Physics B*, 35(3):135–140, 1984.

[41] Xanthi Tsampoula, Veneranda Garcés-Chávez, Muriel Comrie, David James Stevenson, Ben Agate, CTA Brown, Frank Gunn-Moore, and Kishan Dholakia. Femtosecond cellular transfection using a nondiffracting light beam. *Applied Physics Letters*, 91(5):053902, 2007.

[42] Donald Chang. *Guide to electroporation and electrofusion*. Academic Press, 1991.

[43] Joseph M Crowley. Electrical breakdown of bimolecular lipid membranes as an electromechanical instability. *Biophysical journal*, 13(7):711–724, 1973.

[44] Eberhard Neumann and Kurt Rosenheck. Permeability changes induced by electric impulses in vesicular membranes. *The Journal of membrane biology*, 10(1):279–290, 1972.

[45] Stéphane Orlowski, Jean Belehradek Jr, Claude Paoletti, and Lluis M Mir. Transient electropermeabilization of cells in culture: increase of the cytotoxicity of anticancer drugs. *Biochemical pharmacology*, 37(24):4727–4733, 1988.

[46] Charles H Jones, Chih-Kuang Chen, Anitha Ravikrishnan, Snehal Rane, and Blaine A Pfeifer. Overcoming nonviral gene delivery barriers: perspective and future. *Molecular pharmaceutics*, 10(11):4082–4098, 2013.

[47] Kenya Kamimura, Takeshi Suda, Guisheng Zhang, and Dexi Liu. Advances in gene delivery systems. *Pharmaceutical medicine*, 25(5):293–306, 2011.

[48] B Alberts. Johnson-a. lewis j. raff m. roberts k. and p. walter (editor): Molecular biology of the cell. garland science, 2002.

[49] Inês Chen and David Dubnau. Dna uptake during bacterial transformation. *Nature Reviews Microbiology*, 2(3):241, 2004.

[50] Daniel W Pack, Allan S Hoffman, Suzie Pun, and Patrick S Stayton. Design and development of polymers for gene delivery. *Nature reviews Drug discovery*, 4(7):581, 2005.

[51] Zeger Debyser. A short course on virology/vectorology/gene therapy. *Current gene therapy*, 3(6):495–499, 2003.

[52] Tilak Jain, Ryan McBride, Steven Head, and Enrique Saez. Highly parallel introduction of nucleic acids into mammalian cells grown in microwell arrays. *Lab on a Chip*, 9(24):3557–3566, 2009.

[53] Theodore L Roth, Cristina Puig-Saus, Ruby Yu, Eric Shifrut, Julia Carnevale, P Jonathan Li, Joseph Hiatt, Justin Saco, Paige Krystofinski, Han Li, et al. Reprogramming human t cell function and specificity with non-viral genome targeting. *Nature*, 559(7714):405, 2018.

[54] Anders Laustsen and Rasmus O Bak. Electroporation-based crispr/cas9 gene editing using cas9 protein and chemically modified sgrnas. In *CRISPR Gene Editing*, pages 127–134. Springer, 2019.

[55] Yihong Zhan, Jun Wang, Ning Bao, and Chang Lu. Electroporation of cells in microfluidic droplets. *Analytical chemistry*, 81(5):2027–2031, 2009.

[56] Ana Valero, Janine Nicole Post, Jan William van Nieuwkasteele, Paulus Martinus ter Braak, W Kruijer, and Albert van den Berg. Gene transfer and protein dynamics in stem cells using single cell electroporation in a microfluidic device. *Lab on a Chip*, 8(1):62–67, 2008.

[57] Zhengzheng Fei, Shengnian Wang, Yubing Xie, Brian E Henslee, Chee Guan Koh, and L James Lee. Gene transfection of mammalian cells using membrane sandwich electroporation. *Analytical chemistry*, 79(15):5719–5722, 2007.

[58] Robert J Turnbull, E Neumann, and K Rosenheck. An alternate explanation for the permeability changes induced by electrical impulses in vesicular membranes. *Journal of Membrane Biology*, 14(1):193–196, 1973.

[59] Tadej Kotnik, Lea Rems, Mounir Tarek, and Damijan Miklavčič. Membrane electroporation and electropermeabilization: mechanisms and models. *Annual review of biophysics*, 48, 2019.

[60] James C Weaver and Yu A Chizmadzhev. Theory of electroporation: a review. *Bioelectrochemistry and bioenergetics*, 41(2):135–160, 1996.

[61] Min Ju Park, Myeong Soo Park, and Geun Eog Ji. Improvement of electroporation-mediated transformation efficiency for a bifidobacterium strain to a reproducibly high level. *Journal of microbiological methods*, 159:112–119, 2019.

[62] Mun Su Rhee, Jin-woo Kim, Yilei Qian, LO Ingram, and KT Shanmugam. Development of plasmid vector and electroporation condition for gene transfer in sporogenic lactic acid bacterium, bacillus coagulans. *Plasmid*, 58(1):13–22, 2007.

[63] Rachel Binet and Anthony T Maurelli. Transformation and isolation of allelic exchange mutants of chlamydia psittaci using recombinant dna introduced by electroporation. *Proceedings of the National Academy of Sciences*, 106(1):292–297, 2009.

[64] YH Kim, KS Han, S Oh, Seungkwon You, and Sae Hun Kim. Optimization of technical conditions for the transformation of lactobacillus acidophilus strains by electroporation. *Journal of applied microbiology*, 99(1):167–174, 2005.

[65] Ding-Qiang Chen, Shao-Song Huang, and Yong-Jun Lu. Efficient transformation of legionella pneumophila by high-voltage electroporation. *Microbiological research*, 161(3):246–251, 2006.

[66] Haihong Zhang, Yuanting Li, Ximing Chen, Hongmei Sheng, and Lizhe An. Optimization of electroporation conditions for arthrobacter with plasmid part2. *Journal of microbiological methods*, 84(1):114–120, 2011.

[67] Paulo A Garcia, Zhifei Ge, Jeffrey L Moran, and Cullen R Buie. Microfluidic screening of electric fields for electroporation. *Scientific reports*, 6:21238, 2016.

[68] Bruce L Roth, Martin Poot, Stephen T Yue, and Paul J Millard. Bacterial viability and antibiotic susceptibility testing with sytox green nucleic acid stain. *Appl. Environ. Microbiol.*, 63(6):2421–2431, 1997.

[69] Q. R. Wang, A. A. D. Jones, J. A. Gralnick, L. W. Lin, and C. R. Buie. Microfluidic dielectrophoresis illuminates the relationship between microbial cell envelope polarizability and electrochemical activity. *Science Advances*, 5(1), 2019.

[70] Mario R Capecchi. Generating mice with targeted mutations. *Nature medicine*, 7(10):1086, 2001.

[71] R. Jaenisch and B. Mintz. Simian virus 40 dna sequences in dna of healthy adult mice derived from preimplantation blastocysts injected with viral dna. *Proceedings of the National Academy of Sciences of the United States of America*, 71(4):1250–1254, 1974.

[72] M. Ashraf and N. A. Akram. Improving salinity tolerance of plants through conventional breeding and genetic engineering: An analytical comparison. *Biotechnology Advances*, 27(6):744–752, 2009.

[73] T. Kotnik, P. Kramar, G. Pucihar, D. Miklavcic, and M. Tarek. Cell membrane electroporation-part 1: The phenomenon. *Ieee Electrical Insulation Magazine*, 28(5):14–23, 2012.

[74] Frank J Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.

[75] Gisbert Schneider. Automating drug discovery. *Nature reviews drug discovery*, 17(2):97–113, 2018.

[76] G Richard Cathcart, Thomas Brennan-Marquez, John A Bridgham, George S Golda, Harry A Guiremand, Marianne Hane, Louis B Hoff, Eric Lachenmeier, Melvyn N Kronick, Douglas H Keith, et al. Automated molecular biology laboratory, August 22 1995. US Patent 5,443,791.

[77] Paulo A Garcia, Zhifei Ge, Laura E Kelley, Steven J Holcomb, and Cullen R Buie. High efficiency hydrodynamic bacterial electrotransformation. *Lab on a Chip*, 17(3):490–500, 2017.

[78] Cindy J Smith and A Mark Osborn. Advantages and limitations of quantitative pcr (q-pcr)-based approaches in microbial ecology. *FEMS microbiology ecology*, 67(1):6–20, 2009.

[79] Hazel M Davey and Douglas B Kell. Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. *Microbiological reviews*, 60(4):641–696, 1996.

[80] Jacob Beal, Natalie G Farny, Traci Haddock-Angelli, Vinoo Selvarajah, Geoff S Baldwin, Russell Buckley-Taylor, Markus Gershater, Daisuke Kiga, John Marken, Vishal Sanchania, et al. the igem interlab study contributors. *Robust estimation of bacterial cell count from optical density. bioRxiv*, 2019.

[81] Ronen Hazan, Yok-Ai Que, Damien Maura, and Laurence G Rahme. A method for high throughput determination of viable bacteria cell counts in 96-well plates. *BMC microbiology*, 12(1):1–7, 2012.

[82] Sander Sieuwerts, Frank AM De Bok, Erik Mols, Willem M De Vos, and JET van Hylckama Vlieg. A simple and fast method for determining colony forming units. *Letters in applied microbiology*, 47(4):275–278, 2008.

[83] Joanne M Willey, Linda M Sherwood, Christopher J Woolverton, and Mark Schneegurt. *Study Guide to Accompany Prescott, Harley, and Klein's Microbiology*. McGraw-Hill Higher Education, 2008.

[84] Robert S Breed and WD Dotterrer. The number of colonies allowable on satisfactory agar plates. *Journal of bacteriology*, 1(3):321–331, 1916.

[85] Silvio D Brugger, Christian Baumberger, Marcel Jost, Werner Jenni, Urs Brugger, and Kathrin Mühlemann. Automated counting of bacterial colony forming units on agar plates. *PloS one*, 7(3):e33695, 2012.

[86] Matthew L Clarke, Robert L Burton, A Nayo Hill, Maritoni Litorja, Moon H Nahm, and Jeeseong Hwang. Low-cost, high-throughput, automated counting of bacterial colonies. *Cytometry Part A*, 77(8):790–797, 2010.

[87] Quentin Geissmann. Opencfu, a new free and open-source software to count cell colonies and other circular objects. *PloS one*, 8(2):e54072, 2013.

[88] Arif Ul Maula Khan, Angelo Torelli, Ivo Wolf, and Norbert Gretz. Autocellseg: robust automatic colony forming unit (cfu)/cell analysis using adaptive image segmentation and easy-to-use post-editing techniques. *Scientific reports*, 8(1):7302, 2018.

[89] Priya Choudhry. High-throughput method for automated colony and cell counting by digital image analysis based on edge detection. *PloS one*, 11(2):e0148469, 2016.

[90] Zhongli Cai, Niladri Chattopadhyay, Wenchao Jessica Liu, Conrad Chan, Jean-Philippe Pignol, and Raymond M Reilly. Optimized digital counting colonies of clonogenic assays using imagej software and customized macros: comparison with manual counting. *International journal of radiation biology*, 87(11):1135–1146, 2011.

[91] Sander Sieuwerts, Frank AM De Bok, Erik Mols, Willem M De Vos, and JET van Hylckama Vlieg. A simple and fast method for determining colony forming units. *Letters in applied microbiology*, 47(4):275–278, 2008.

[92] Mark-Anthony Bray, Martha S Vokes, and Anne E Carpenter. Using cellprofiler for automatic identification and measurement of biological objects in images. *Current Protocols in Molecular Biology*, 109(1):14–17, 2015.

[93] Scott Sutton. Counting colonies. *Pharmaceutical Microbiology Forum Newsletter*, 12(9):2–5, 2006.

[94] Diane M Tomasiewicz, Donald K Hotchkiss, George W Reinbold, Ralston B Read Jr, and Paul A Hartman. The most suitable number of colonies on plates for counting. *Journal of Food Protection*, 43(4):282–286, 1980.

[95] Diane M Tomasiewicz, Donald K Hotchkiss, George W Reinbold, Ralston B Read Jr, and Paul A Hartman. The most suitable number of colonies on plates for counting. *Journal of Food Protection*, 43(4):282–286, 1980.

[96] Clayton M Costa and Suann Yang. Counting pollen grains using readily available, free image processing and analysis software. *Annals of Botany*, 104(5):1005–1010, 2009.

[97] Fernando C Monteiro. Pollen grain recognition through deep learning convolutional neural networks. In *AIP Conference Proceedings*, volume 2425. AIP Publishing, 2022.

[98] Mathieu De Langlard, Hania Al-Saddik, Sophie Charton, Johan Debayle, and Fabrice Lamadie. An efficiency improved recognition algorithm for highly overlapping ellipses: Application to dense bubbly flows. *Pattern Recognition Letters*, 101:88–95, 2018.

[99] Wen-Hui Zhang, Xiaoya Jiang, and Yin-Mingzi Liu. A method for recognizing overlapping elliptical bubbles in bubble image. *Pattern Recognition Letters*, 33(12):1543–1548, 2012.

[100] Dilip K Prasad, Maylor KH Leung, and Siu-Yeung Cho. Edge curvature and convexity based ellipse detection method. *Pattern Recognition*, 45(9):3204–3221, 2012.

[101] Michael Thompson. Pattern classification and scene analysis by richard o. duda and peter e. hart. *Leonardo*, 7(4):370–370, 1974.

[102] Sahar Zafari, Tuomas Eerola, Jouni Sampo, Heikki Kälviäinen, and Heikki Haario. Segmentation of partially overlapping nanoparticles using concave points. In *Advances in Visual Computing: 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part I 11*, pages 187–197. Springer, 2015.

[103] Sahar Zafari, Tuomas Eerola, Jouni Sampo, Heikki Kälviäinen, and Heikki Haario. Segmentation of partially overlapping nanoparticles using concave points. In *Advances in Visual Computing: 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part I 11*, pages 187–197. Springer, 2015.

[104] Xiaoyuan Guo, Hanyi Yu, Blair Rossetti, George Teodoro, Daniel Brat, and Jun Kong. Clumped nuclei segmentation with adjacent point match and local shape-based intensity analysis in fluorescence microscopy images. In *2018 40th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3410–3413. IEEE, 2018.

[105] Sahar Zafari, Mariia Murashkina, Tuomas Eerola, Jouni Sampo, Heikki Kälviäinen, and Heikki Haario. Resolving overlapping convex objects in silhouette images by concavity analysis and gaussian process. *Journal of Visual Communication and Image Representation*, 73:102962, 2020.

[106] Chanho Jung and Changick Kim. Segmenting clustered nuclei using h-minima transform-based marker extraction and contour parameterization. *IEEE transactions on biomedical engineering*, 57(10):2600–2604, 2010.

[107] Chiwoo Park, Jianhua Z Huang, Jim X Ji, and Yu Ding. Segmentation, inference and classification of partially overlapping nanoparticles. *IEEE transactions on pattern analysis and machine intelligence*, 35(3):1–1, 2012.

[108] David W Paglieroni. Distance transforms: Properties and machine vision applications. *CVGIP: Graphical models and image processing*, 54(1):56–74, 1992.

[109] Li Yang and Abdallah Shami. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415:295–316, 2020.

[110] Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren. *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019.

[111] Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren. *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019.

[112] Satoshi Suzuki et al. Topological structural analysis of digitized binary images by border following. *Computer vision, graphics, and image processing*, 30(1):32–46, 1985.

[113] Sahar Zafari, Tuomas Eerola, Jouni Sampo, Heikki Kälviäinen, and Heikki Haario. Comparison of concave point detection methods for overlapping convex objects segmentation. In *Image Analysis: 20th Scandinavian Conference, SCIA 2017, Tromsø, Norway, June 12–14, 2017, Proceedings, Part II 20*, pages 245–256. Springer, 2017.

[114] Gunilla Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Transactions on pattern analysis and machine intelligence*, 10(6):849–865, 1988.

[115] Egon Balas. An additive algorithm for solving linear programs with zero-one variables. *Operations Research*, 13(4):517–546, 1965.

[116] Fred Glover. A multiphase-dual algorithm for the zero-one integer programming problem. *Operations Research*, 13(6):879–919, 1965.

[117] Hugh M Pettigrew and William C Mohler. A rapid test for the poisson distribution using the range. *Biometrics*, pages 685–692, 1967.

[118] RA Fisher et al. The accuracy of the plating method of estimating the density of bacterial populations, with particular reference to the use of thornton's agar medium with soil samples. *Annals of Applied Biology*, 9(3–4):325–359, 1922.

[119] Philippe Mazodier and Julian Davies. Gene transfer between distantly related bacteria. *Annual review of genetics*, 25(1):147–171, 1991.

[120] Naoto Yoshida and Misa Sato. Plasmid uptake by bacteria: a comparison of methods and efficiencies. *Applied microbiology and biotechnology*, 83:791–798, 2009.

[121] Ana Leonor Rivera, Denis Magana-Ortiz, Miguel Gomez-Lim, Francisco Fernandez, and Achim M Loske. Physical methods for genetic transformation of fungi and yeast. *Physics of life reviews*, 11(2):184–203, 2014.

[122] YuLing Sheng, Valeria Mancino, and Bruce Birren. Transformation of escherichia coli with large dna molecules by electroporation. *Nucleic acids research*, 23(11):1990–1996, 1995.

[123] Bruce M Chassy, Annick Mercenier, and Jeannette Flickinger. Transformation of bacteria by electroporation. *Trends in Biotechnology*, 6(12):303–309, 1988.

[124] Tadej Kotnik, Wolfgang Frey, Martin Sack, Saša Haberl Meglič, Matjaž Peterka, and Damijan Miklavčič. Electroporation-based applications in biotechnology. *Trends in biotechnology*, 33(8):480–488, 2015.

[125] Eberhard Neumann. Membrane electroporation and direct gene transfer. *Journal of Electroanalytical Chemistry*, 343(1-2):247–267, 1992.

[126] Anthony L Shiver, Rebecca Culver, Adam M Deutschbauer, and Kerwyn Casey Huang. Rapid ordering of barcoded transposon insertion libraries of anaerobic bacteria. *Nature protocols*, 16(6):3049–3071, 2021.

[127] Gaelen T Hess, Laure Frésard, Kyuho Han, Cameron H Lee, Amy Li, Karlene A Cimprich, Stephen B Montgomery, and Michael C Bassik. Directed evolution using dcas9-targeted somatic hypermutation in mammalian cells. *Nature methods*, 13(12):1036–1042, 2016.

[128] Lorena Ruiz, Francesca Bottacini, Christine J Boinett, Amy K Cain, Mary O'Connell-Motherway, Trevor D Lawley, and Douwe van Sinderen. The essential genomic landscape of the commensal bifidobacterium breve ucc2003. *Scientific Reports*, 7(1):5648, 2017.

[129] Saeid Movahed and Dongqing Li. Microfluidics cell electroporation. *Microfluidics and Nanofluidics*, 10:703–734, 2011.

[130] Tao Geng and Chang Lu. Microfluidic electroporation for cellular analysis and delivery. *Lab on a Chip*, 13(19):3803–3821, 2013.

[131] Iris van Uitert Séverine Le Gac. *Electroporation in Microfluidic Devices*. 2006.

[132] MB Fox, DC Esveld, A Valero, R Luttge, HC Mastwijk, PV Bartels, A Van Den Berg, and RM Boom. Electroporation of cells in microfluidic devices: a review. *Analytical and bioanalytical chemistry*, 385:474–485, 2006.

[133] Hsiang-Yu Wang and Chang Lu. Electroporation of mammalian cells in a microfluidic channel with geometric variation. *Analytical chemistry*, 78(14):5158–5164, 2006.

[134] Hsiang-Yu Wang and Chang Lu. Microfluidic electroporation for delivery of small molecules and genes into cells using a common dc power supply. *Biotechnology and bioengineering*, 100(3):579–586, 2008.

[135] Roee Ziv, Yair Steinhardt, Gadi Pelled, Dan Gazit, and Boris Rubinsky. Micro-electroporation of mesenchymal stem cells with alternating electrical current pulses. *Biomedical microdevices*, 11:95–101, 2009.

[136] Tao Geng, Yihong Zhan, Jun Wang, and Chang Lu. Transfection of cells using flow-through electroporation based on constant voltage. *nature protocols*, 6(8):1192–1208, 2011.

[137] Emmanuel G Guignet and Tobias Meyer. Suspended-drop electroporation for high-throughput delivery of biomolecules into cells. *Nature methods*, 5(5):393–395, 2008.

[138] Andrew C Madison, Matthew W Royal, Frederic Vigneault, Liji Chen, Peter B Griffin, Mark Horowitz, George M Church, and Richard B Fair. Scalable device for automated microbial electroporation in a digital microfluidic platform. *ACS Synthetic Biology*, 6(9):1701–1709, 2017.

[139] Mohamad Hamieh, Anton Dobrin, Annalisa Cabriolu, Sjoukje JC van der Stegen, Theodoros Giavridis, Jorge Mansilla-Soto, Justin Eyquem, Zeguo Zhao, Benjamin M Whitlock, Matthew M Miele, et al. Car t cell trogocytosis and cooperative killing regulate tumour antigen escape. *Nature*, 568(7750):112–116, 2019.

[140] Judith Feucht, Jie Sun, Justin Eyquem, Yu-Jui Ho, Zeguo Zhao, Josef Leibold, Anton Dobrin, Annalisa Cabriolu, Mohamad Hamieh, and Michel Sadelain. Calibration of car activation potential directs alternative t cell fates and therapeutic potency. *Nature medicine*, 25(1):82–88, 2019.

[141] Justin Eyquem, Jorge Mansilla-Soto, Theodoros Giavridis, Sjoukje JC van der Stegen, Mohamad Hamieh, Kristen M Cunanan, Ashlesha Odak, Mithat Gönen, and Michel Sadelain. Targeting a car to the trac locus with crispr/cas9 enhances tumour rejection. *Nature*, 543(7643):113–117, 2017.

[142] Daniel P Dever, Rasmus O Bak, Andreas Reinisch, Joab Camarena, Gabriel Washington, Carmencita E Nicolas, Mara Pavel-Dinu, Nivi Saxena, Alec B

Wilkens, Sruthi Mantri, et al. Crispr/cas9 β-globin gene targeting in human haematopoietic stem cells. *Nature*, 539(7629):384–389, 2016.

[143] Matthias Bozza, Alice De Roia, Margareta P Correia, Aileen Berger, Alexandra Tuch, Andreas Schmidt, Inka Zörnig, Dirk Jäger, Patrick Schmidt, and Richard P Harbottle. A nonviral, nonintegrating dna nanovector platform for the safe, rapid, and persistent manufacture of recombinant t cells. *Science advances*, 7(16):eabf1333, 2021.

[144] Ulrich Sonnenborn. Escherichia coli strain nissle 1917—from bench to bedside and back: history of a special escherichia coli strain with probiotic properties. *FEMS microbiology letters*, 363(19):fnw212, 2016.

[145] Pichet Praveschotinunt, Anna M Duraj-Thatte, Ilia Gelfat, Franziska Bahl, David B Chou, and Neel S Joshi. Engineered e. coli nissle 1917 for the delivery of matrix-tethered therapeutic domains to the gut. *Nature communications*, 10(1):5580, 2019.

[146] A O'callaghan, F Bottacini, M O'Connell Motherway, and D Van Sinderen. Pangenome analysis of bifidobacterium longum and site-directed mutagenesis through by-pass of restriction-modification systems. *BMC genomics*, 16:1–19, 2015.

[147] Min Ju Park, Myeong Soo Park, and Geun Eog Ji. Improvement of electroporation-mediated transformation efficiency for a bifidobacterium strain to a reproducibly high level. *Journal of microbiological methods*, 159:112–119, 2019.

[148] Lorena Ruiz, Mary O'Connell Motherway, Noreen Lanigan, and Douwe van Sinderen. Transposon mutagenesis in bifidobacterium breve: construction and characterization of a tn5 transposon mutant library for bifidobacterium breve ucc2003. *PLoS One*, 8(5):e64699, 2013.

[149] Blanca Del Rosal, Chen Sun, Despina Nelie Loufakis, Chang Lu, and Daniel Jaque. Thermal loading in flow-through electroporation microfluidic devices. *Lab on a Chip*, 13(15):3119–3127, 2013.

[150] Roberto C Gallo-Villanueva, Michael B Sano, Blanca H Lapizco-Encinas, and Rafael V Davalos. Joule heating effects on particle immobilization in insulator-based dielectrophoretic devices. *Electrophoresis*, 35(2-3):352–361, 2014.

[151] Aude Silve, Isabelle Leray, Clair Poignard, and Lluis M Mir. Impact of external medium conductivity on cell membrane electropermeabilization by microsecond and nanosecond electric pulses. *Scientific reports*, 6(1):19957, 2016.

[152] Paulo A Garcia, John H Rossmeisl, Robert E Neal, Thomas L Ellis, John D Olson, Natalia Henao-Guerrero, John Robertson, and Rafael V Davalos. Intracranial nonthermal irreversible electroporation: in vivo analysis. *The Journal of membrane biology*, 236:127–136, 2010.

[153] Davorka Sel, David Cukjati, Danute Batiuskaite, Tomaz Slivnik, Lluis M Mir, and Damijan Miklavcic. Sequential finite element model of tissue electropermeabilization. *IEEE Transactions on Biomedical Engineering*, 52(5):816–827, 2005.

[154] Qingfu Zhu, Megan Hamilton, Bryan Vasquez, and Mei He. 3d-printing enabled micro-assembly of a microfluidic electroporation system for 3d tissue engineering. *Lab on a Chip*, 19(14):2362–2372, 2019.

[155] Johannes Akinlaja and Frederick Sachs. The breakdown of cell membranes by electrical and mechanical stress. *Biophysical journal*, 75(1):247–254, 1998.

[156] Yihong Zhan, Zhenning Cao, Ning Bao, Jianbo Li, Jun Wang, Tao Geng, Hao Lin, and Chang Lu. Low-frequency ac electroporation shows strong frequency dependence and yields comparable transfection results to dc electroporation. *Journal of controlled release*, 160(3):570–576, 2012.

[157] Hirosuke Sugahara, Toshitaka Odamaki, Shinji Fukuda, Tamotsu Kato, Jin-zhong Xiao, Fumiaki Abe, Jun Kikuchi, and Hiroshi Ohno. Probiotic bifidobacterium longum alters gut luminal metabolism through modification of the gut microbial community. *Scientific reports*, 5(1):13548, 2015.

[158] Kelly M Wetmore, Morgan N Price, Robert J Waters, Jacob S Lamson, Jennifer He, Cindi A Hoover, Matthew J Blow, James Bristow, Gareth Butland, Adam P Arkin, et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio*, 6(3):10–1128, 2015.

[159] Hualan Liu, Anthony L Shiver, Morgan N Price, Hans K Carlson, Valentine V Trotter, Yan Chen, Veronica Escalante, Jayashree Ray, Kelsey E Hern, Christopher J Petzold, et al. Functional genetics of human gut commensal bacteroides thetaiotaomicron reveals metabolic requirements for growth across environments. *Cell reports*, 34(9), 2021.