Engineering and Computational Tools for Salivary Biomedicine

Thesis by
Yeokyoung (Anne) Kil

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended April 23, 2025

Yeokyoung (Anne) Kil
ORCID: 0000-0002-1235-7379

ACKNOWLEDGEMENTS

This work would not have been possible without the guidance and support of many kindhearted people around me. First and foremost, I am sincerely grateful to my advisor, Lior Pachter, who took a chance by accepting me into his lab and provided unwavering support throughout my academic journey. I'm incredibly lucky to have such a supportive advisor, and it meant a lot to know that he will always have my back.

I would like to thank my committee members, Wei, Joel, and Anne, for their invaluable guidance. Their willingness to share their expertise and constructive feedback have been instrumental in bringing this thesis into the world. A sincere thanks to Anne, who has helped me many times throughout my graduate school journey from all the way across the country, and has greatly motivated my research as the "queen of saliva."

I am deeply grateful to my fellow lab members who made this journey not only possible but also enjoyable. Weekly group meetings full of fascinating research, shared celebrations of big and small successes, commiserations over failures, lab trips all around SoCal, and impromptu boba runs have all become unforgettable memories I will cherish. This work bears the invisible fingerprints of your collective influence and support.

I would like to extend my sincere appreciation to the faculty members who gave me the opportunity to teach the class, Design and Construction of Biodevices. To Justin, Yang, and Richard, thank you for your mentorship, patience, and trust as I taught the class and discovered my passion for teaching. I owe a special thanks to Justin, who entrusted me with building a pulse oximeter from scratch with nothing but an old soldering iron at hand, in the middle of nowhere in Korea during the pandemic lockdown. This experience is something I'll always be proud of, and I couldn't have done it without your guidance.

I am also grateful to the amazing people on Caltech campus for their invaluable support throughout my graduate studies. I thank Laura and Daniel from the ISP, for always having my back and helping me navigate life in America as an international scholar. I thank Christine, the guardian angel of Medical Engineering, for always thinking of us students and creating a space where we can thrive.

I owe special thanks to my dearest friends in the Graduate Student Council who

tirelessly work behind the scenes to improve our community, often without recognition. Thank you for the countless hours we spent together advocating for our fellow students, organizing events, and most importantly, creating spaces where we could all step away from our labs and remember we were not alone in this journey.

Beyond academia, I have been incredibly fortunate to be surrounded by friends who made this journey not just bearable, but truly memorable. While words can never fully express what their presence has meant to me, I would be remiss not to acknowledge how vital they have been to both my work and wellbeing. I thank my camping friends, Saehui, Yoojin, Minkyo, and Si-Hyung, who transformed the isolating first years of graduate school during the pandemic into unexpected adventures. I thank my lab mates again for being amazing friends throughout this whole journey, especially Taleen, Arsani, Gennady, Riley, and Tara, whose companionship and care enriched my life far beyond our shared academic pursuits. I owe immeasurable gratitude to my best friend, Erika, whose parallel journey through her own PhD created a unique understanding that few others could provide. Thank you for being my constant confidant—for patiently listening when I needed to vent, for genuinely celebrating every small victory alongside the major milestones, and for sharing your own experiences in ways that made me feel less alone on this challenging path.

Finally, I reserve my deepest gratitude for my family, whose unwavering support has been the bedrock of this academic journey. To my parents, who instilled in me the curiosity and perseverance that led me here, thank you for your lifelong support and unconditional love. Your early encouragement of my questions and interests laid the groundwork for my academic journey, while your consistent faith in my abilities gave me the courage to pursue this degree. To my siblings, thank you for your encouragement, humor, and perspective throughout this process. Your check-in calls, perfectly timed distractions, and unwavering confidence in me provided welcome reminders of life beyond academia. Even from a distance, your support has been a constant source of strength.

I must also thank Lily, my faithful feline companion, who provided unconditional comfort during long writing sessions and reminded me of the importance of regular breaks. Your purring presence beside my keyboard brought calm to the most stressful days.

Above all, my profound love and gratitude go to my husband, Joseph, who has been my unwavering foundation throughout this journey. Thank you for being my

first listener when ideas needed refining, and for your steadfast belief in me when doubts crept in. You celebrated every small victory as if it were your own and provided perspective when challenges seemed insurmountable. Your sacrifices, encouragement, and partnership have made this achievement as much yours as it is mine. I simply could not have done this without you by my side.

ABSTRACT

Saliva is emerging as a powerful biofluid for noninvasive diagnostics, offering a window into human health through its diverse biomolecular composition. This dissertation advances the field of salivary biomedicine by addressing critical challenges in saliva collection, processing, and analysis. First, a comparative analysis of five saliva collection devices highlighted key usability factors, informing the development of SalivaStraw—a novel device designed to improve collection efficiency and minimize leakage. Next, colosseum, a low-cost, open-source fraction collector, was designed and developed to facilitate scalable saliva processing and improve biomarker isolation. Finally, a computational framework leveraging spline regression was applied to longitudinal salivary transcriptomic data, enabling the identification of temporally regulated genes and underscoring saliva's potential for dynamic health monitoring. Collectively, this work contributes new tools and methodologies that strengthen the foundation of saliva-based diagnostics, broadening its applications in precision medicine and beyond.

PUBLISHED CONTENT AND CONTRIBUTIONS

The materials in this dissertation are largely drawn from the manuscripts listed below. The abbreviations in the list of publications are as follows: LP: Lior Pachter, ASB: A. Sina Booeshaghi, KHM: Kyung Hoi Min, JG: Jase Gehring. Although not indicated in the list below, YK was co-first author on manuscript 3 with ASB.

[1] Y. Kil, A. S. Booeshaghi, and L. S. Pachter, "Comparative survey-based study of noninvasive saliva collection devices," *J. Med. Device.*, vol. 19, pp. 1–19, 2 Jun. 1, 2025. DOI: 10.1115/1.4067232,
YK and LP conceived the study. YK developed the computational pipeline and performed the analyses. YK interpreted the results with input from LP. YK created the figures and wrote the manuscript, with revisions from LP.

[2] Y. Kil and L. Pachter, "Differential analysis reveals isoform switching following pneumococcal vaccination," *bioRxiv*, Mar. 10, 2025. DOI: 10.1101/2025.03.09.642237,
YK and LP conceived the study. YK developed the computational pipeline and performed the analyses. YK interpreted the results with input from LP. YK created the figures and wrote the manuscript, with revisions from LP.

[3] A. S. Booeshaghi, Y. Kil, K. H. Min, J. Gehring, and L. Pachter, "Low-cost, scalable, and automated fluid sampling for fluidics applications," *HardwareX*, vol. 10, e00201, Oct. 1, 2021. DOI: 10.1016/j.ohx.2021.e00201,
ASB, YK, and LP designed the fraction collector. JG helped set instrument specifications. YK assembled and built the fraction collector and performed the experiments. ASB, YK, and KHM designed the GUI. KHM coded the installable GUI and YK, KHM, and ASB coded the web-browser GUI. ASB coded the browser-serial package. ASB, YK, and KHM wrote the documentation. ASB and YK analyzed the data and made figures. ASB, YK, and LP wrote the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Background

Saliva has emerged as a powerful biofluid in biomedical research and diagnostics due to its abundance of diverse biomarkers that reflect local and systemic physiology. Key categories of biomarkers in saliva include but are not limited to nucleic acids, hormones, other proteins and peptides, and metabolites [1]. Nucleic acids, including human and microbial DNA, RNA, and microRNAs, enable genomic and transcriptomic analyses [2]. Hormones such as cortisol and testosterone provide insights into stress and endocrine function, while metabolites like glucose and lactate offer a window into metabolic processes [3]. Proteins and peptides encompass a broad range of biomarkers, including digestive enzymes, inflammatory and immune markers like interleukins and C-reactive protein, antimicrobial peptides such as defensins, and disease-specific antibodies, like those for HIV [4], [5].

Collectively, these biomarkers position saliva as a valuable tool for diagnostics, health monitoring, microbiome profiling, and even environmental exposure assessment [3], [6]. Saliva has been extensively used for oral and systemic disease diagnostics for conditions such as certain cancers [7], [8], autoimmune diseases [9], [10], and neurological disorders [11]. Saliva has played a crucial role in infectious disease detection, with applications ranging from viral infections like HIV [4], [12], SARS-CoV-2 [13], [14], and hepatitis [15], to bacterial and fungal pathogens such as Mycobacterium tuberculosis and Candida [16], [17]. The salivary microbiome has also been extensively studied for its role in oral and systemic diseases [6]. Saliva is also being leveraged for longitudinal health monitoring, particularly in tracking vaccine responses and immune function over time [18]. This growing body of research underscores saliva's vast potential as a versatile and informative biofluid, paving the way for its broader integration into noninvasive disease diagnostics, personalized health monitoring, and public health surveillance.

Aside from the abundance of biomarkers, saliva has several attributes that make it an ideal sample for biomedicine. The greatest advantage of saliva is that it can be self-collected noninvasively [19]. Unlike blood, which requires invasive sampling by a trained professional, saliva can be collected noninvasively, but shares many

analytes with blood, such as metabolites like glucose and lactate [20], antibodies [20], cytokines [21], or hormones [22], as analytes in saliva originate primarily from capillaries that feed the salivary glands [20]. Compared to other noninvasively-collected biofluids such as sweat, interstitial fluid (ISF), tear, and urine, saliva is easier to collect because it is produced more readily [23]. Saliva is stable at room temperature for the detection of hormones like glucocorticoids and androgens [24] and higher temperatures for the detection of viruses such as HIV [4] and SARS-CoV-2 [25], which facilitates the transportation of samples, although saliva samples are usually preserved with protease-inhibiting buffers for protein biomarkers as they contain proteases that can degrade proteins [26].

1.2 Challenges and Gaps in Salivary Biomedicine

Despite its advantages, several challenges hinder saliva's widespread adoption. Some of saliva's inherent characteristics cause difficulties in using saliva as a sample. Saliva composition can vary between individuals [27], posing a challenge for establishing reliable baseline values. Furthermore, the flow rate and composition of saliva can fluctuate significantly even within an individual due to a range of factors like collection method, hydration status, physical activity, or drug use [20], [28], complicating the consistency of biomarker detection. Saliva's viscosity due to mucins, along with enzymes like proteases and RNases, requires pretreatment to ensure accurate biomarker extraction and avoid degradation [29]. The microbial content in saliva also further complicates salivary analysis, as it can interfere with genomic or transcriptomic analyses by introducing non-human sequences and drowning significant biological signals [30].

Saliva's underutilization in research and diagnostics is partly rooted in its inherent complexities, compounded by historical preferences for blood and other biofluids. Blood has been established as the gold standard for diagnostic tests, with a robust infrastructure and well-defined protocols supporting its widespread use [31]. As a result, many tests are specifically designed for blood, leaving saliva relatively unexplored. Despite many studies highlighting the potential of saliva as a diagnostic fluid, particularly for viral detection during the COVID-19 pandemic [13], [14], [32]–[34], nasal and nasopharyngeal swabs remained the preferred method, demonstrating the broader reluctance to adopt saliva-based approaches. The underutilization of saliva is evident in the scarcity of FDA-approved saliva-based tests, highlighting its slow integration into routine clinical practice [35]–[37]. Furthermore, in the scope of salivary genomics and transcriptomics, the scarcity of

DNA-seq or RNA-seq datasets and studies on saliva compared to blood or tissue samples underscores its underutilization as a sample in biological research [38]. In general, the underutilization of saliva is compounded by the lack of standardized saliva collection methods [33], which can amplify the inherent variability in saliva composition and hinder the establishment of baseline levels of salivary biomarkers, further contributing to the reluctance in using saliva as a sample.

1.3 Novel Contributions of This Work

This dissertation addresses challenges in saliva collection, processing, and analysis, by not only advancing methodologies for saliva collection or data analysis but also developing foundational tools that improve and scale salivary research. The contributions of this work encompass novel computational methodologies and devices that can bridge gaps in saliva-based biomedicine and enable more reliable and reproducible studies.

Standardizing saliva collection is essential for ensuring reproducibility and reliability in salivary studies. Variability in collection methods has long been a barrier to the widespread adoption of saliva as a diagnostic biofluid, with differences in device design leading to inconsistencies in sample quality and analyte concentration. These technical inconsistencies are further compounded by usability challenges that affect sample collection in real-world settings. The comparative analysis of saliva collection devices in this dissertation systematically evaluates both collection efficiency and device usability. Unlike most prior studies that primarily focused on biomarker yield, this work emphasizes the role of device usability in shaping the feasibility of saliva-based diagnostics. Building on these insights, a novel saliva collection device, the SalivaStraw, was developed using fluid dynamics principles to optimize sample collection while remaining low-cost and user-friendly. SalivaStraw represents a step toward standardized saliva collection practices by addressing key limitations of existing devices and improving accessibility to saliva-based diagnostic tools.

Enabling scalable and accessible salivary research requires the development of robust, cost-effective technologies that facilitate large-scale studies. The colosseum open-source fraction collector serves as a versatile and affordable liquid handling platform that can automate and enhance the throughput of many types of proteomic workflows, including those used for salivary biomarker research. By making automated fraction collection more accessible, colosseum has the potential to scale proteomic workflows both latitudinally—by enabling high-throughput processing of

multiple experiments simultaneously—and longitudinally—by facilitating the collection and fractionation of repeated samples over time. This scalability strengthens the feasibility of better monitoring of protein biomarkers by allowing researchers to track biological changes with greater temporal resolution and supporting large-scale studies that require extensive sample processing.

Once saliva is efficiently collected using standardized methods and properly processed, the next critical challenge lies in analyzing the resulting data to extract meaningful biological insights. Despite saliva's potential as a sample for longitudinal health monitoring, the lack of robust computational frameworks has limited its utility in capturing dynamic biological changes over time. To bridge this gap, a reproducible computational framework for time-series analysis of longitudinal RNA-seq data was developed using salivary transcriptomic data, enabling the identification of temporally regulated genes across a high number of time points. This method provides a robust framework for analyzing longitudinal RNA-sequencing data in general, but more importantly, demonstrates saliva's potential for dynamic health monitoring like tracking immune responses and other physiological changes. Moreover, by leveraging isoform-level analysis, this work reveals that important biological signals in the transcriptome may be overlooked when analyses are limited to gene-level expression, and lays the fundamental platform for capturing finer-scale dynamics in gene expression. These findings reinforce saliva's viability in real-time health monitoring and precision medicine, while laying the groundwork for future longitudinal studies that may be conducted with saliva.

By addressing key challenges in collection of saliva, scalable proteomic analysis of saliva, and computational analysis of longitudinal saliva samples, this research provides fundamental tools that strengthen the foundation of salivary biomedicine. Through the development of open-source automation technologies, improved saliva collection methods, and computational frameworks for transcriptomic analysis, this work contributes to the growing field of saliva-based diagnostics and broadens the scope of its potential applications.

1.4 Thesis Outline

This thesis is organized into five chapters, each addressing key stages in the process of salivary biomedicine, from collection to processing to analysis, through experimental, computational, and theoretical approaches. Chapter 1 provides the theoretical foundation, introducing the field of salivary biomedicine, outlining its

potential, identifying key challenges, and presenting the methodologies employed in this work. Chapter 2 focuses on the first step of the process: optimizing saliva collection methods through a comparative analysis of existing devices and the development of the SalivaStraw to improve standardization and ease of use. Building on this, Chapter 3 transitions to the next stage—saliva processing—by detailing the design and application of colosseum, an open-source fraction collector, and demonstrating its utility for saliva fractionation in proteomic studies. Building on the foundations of engineered tools for salivary biomedicine laid out in Chapters 2 and 3, Chapter 4 then shifts to the computational analysis of saliva, exploring the use of spline regression models in longitudinal transcriptomic analysis to gain insights into the temporal dynamics of salivary biomarkers. Finally, Chapter 5 integrates these findings into a broader theoretical context, reflecting on the contributions of this work to the field and outlining future directions for salivary diagnostics research.

Chapter 2

OPTIMIZING SALIVA COLLECTION METHODS

[1] Y. Kil, A. S. Booeshaghi, and L. S. Pachter, "Comparative survey-based study of noninvasive saliva collection devices," en, *J. Med. Device.*, vol. 19, pp. 1–19, 2 Jun. 1, 2025. DOI: 10.1115/1.4067232.

2.1 Abstract

While saliva is a valuable sample for noninvasive diagnosis, the usability of saliva collection devices can impact its effectiveness in diagnostics. To address this, we evaluated five saliva collection devices: Salivette (swab), SuperSAL (swab), SalivaBio Passive Drool, Medschenker Saliva Collection Kit (funnel), and a cryovial with a funnel used in SwabSeq COVID-19 (SARS-CoV-2) tests. Saliva collection rates and instruction reading rates for each device were measured for 60 healthy adults. Users then reported the difficulties of instructions, assembly, and saliva collection, as well as whether there was leakage of saliva, through a survey. Overall, Medschenker performed well on most metrics, while SuperSAL did not perform as well. Our results show that no single saliva collection device satisfies all requirements of an ideal device: a device that allows for efficient, easy, and safe saliva collection without leakage. To address these limitations, we developed SalivaStraw, a proof-of-concept collection device designed based on our study findings. SalivaStraw incorporates insights from fluid dynamics and design choices to allow for easy collection of saliva while minimizing saliva leakage, thus improving on existing saliva collection mechanisms.

2.2 Introduction

Saliva is an established biofluid for medical diagnostics [39]. For example, saliva is used in diagnosis of systemic diseases such as Cushing Syndrome [40], in point-of-care testing such as the OraQuick In-Home human immunodeficiency virus (HIV) test [41], in infectious disease monitoring such as the COVID-19 tests [32], [42]. The use of saliva is now expanding beyond medical diagnostics to applications in biological -omics studies, such as genomics, metagenomics, transcriptomics, proteomics, and exosomics [2], [43].

The biggest advantage of using saliva is that whole saliva can be collected nonin-

Method	Advantages	Disadvantages
Swab	Easy use and availability	Retention of analytes by swab
	Suitable for uncooperative users	Possible stimulatory effects
	Homogeneity of samples	Risk of swallowing
Passive Drool	No stimulatory effect	Inhomogeneity of sample
	Represents unstimulated saliva	Long sampling times
		Subject has to be collaborative
Spitting	Little evaporation of saliva	Possible stimulatory effects
	Suitable with low saliva flow	Subject has to be collaborative
		Mucus may interfere with assay
		Increased transmission risk

Table 2.1: Summary of sample methods and representative devices from literature [19], [33]. Devices can be divided into three main categories: swab, passive drool, and spitting. Each device has advantages and disadvantages that may affect device choice for a specific application.

vasively [19], using three main methods: swabs, funnels, and passive drool [19]. Swabs are materials like cotton or synthetic rolls that are placed inside the mouth to passively absorb saliva. Funnels can either be detached or integrated with the collection tube to guide expectorated saliva into the tube. Passive drooling refers to a method in which the user pools saliva passively in their mouths and lets the pooled saliva gravitationally flow into a collection tube. Each mechanism has its own strengths and weaknesses, and one mechanism may be more suitable than others for certain usage cases of the collected saliva, or for specific demographics. For instance, swabs may be used for patients who have difficulty spitting, such as pediatric patients [44]. However, swabs can interfere with salivary contents, such as hormones [45], in which case funnels may be preferred. The characteristics of each saliva collection mechanism are summarized in Table 2.1, and information about the devices that were used in this study are summarized in Table 2.2.

In addition to the characteristics of the saliva collection mechanism, attributes specific to devices should be considered when choosing how to sample saliva. Such attributes may include price, availability, and downstream processes. Selection of a device should also be undertaken in the context of its intended application. For instance, in a mass-testing scenario where many people are collecting saliva for testing purposes, it may be important to minimize saliva leakage to lower the risks of secondary infections, while ensuring that saliva is collected efficiently to increase the throughput of the tests. In this case, testers may be asked to drool into the collection

Device	Method	Mechanism	Price
Salivette	Swab	Absorbent swab, centrifuge to extract saliva	$0.95
SuperSAL	Swab	Absorbent swab, compress swab to extract saliva	$5.00
Passive Drool	Passive Drool	Pooled saliva flows through saliva collection aid	$2.18
Medschenker saliva collection kit	Spitting	12-mm funnel + 10-mL tube	$1.20
SwabSeq saliva collection kit	Spitting	5-mm funnel + 2-mL cryovial	$1.03

Table 2.2: Comparison of the five saliva collection devices used in the study, detailing collection method, extraction mechanism, and cost per unit.

tube rather than spit using a funnel, to minimize risks of secondary infection and to minimize sputum or mucus in the saliva samples for easier downstream processing of the samples. In contrast, at-home tests may emphasize ease-of-use [41].

Despite saliva's diagnostic potential, no standardized method exists for optimized collection across different analytes. This lack of standardization can affect sample quality and test reliability. To address this, we evaluated the performance and usability of five saliva collection devices, aiming to identify the strengths and limitations of each and to demonstrate the critical need for optimized, standardized saliva collection methods. We identified five saliva collection devices: Salivette®, SuperSAL™, Passive Drool, MedSchenker®Saliva Collection Kit, and SwabSeq[1] set-up (Figure 2.1, Table 2.2). We assessed performance, usability, and hygiene using quantitative and qualitative metrics like saliva collection rates and instruction reading times, as well as user-rated difficulties of instructions, assembly, and saliva collection through a survey. Unstimulated saliva flow rates were measured to normalize device performance by individual flow rates. We also measured stimulated saliva flow rates (water and citric acid) to explore future diagnostic applications using stimulated saliva.

[1]We are referring to the funnel and cryovial set-up as SwabSeq, which is the name of a mass-testing platform of COVID-19 testing[32], as this set-up was used in an early implementation of SwabSeq surveillance testing on Caltech Campus. This set-up is not specific to the SwabSeq protocol itself, nor do any of the results from our survey.

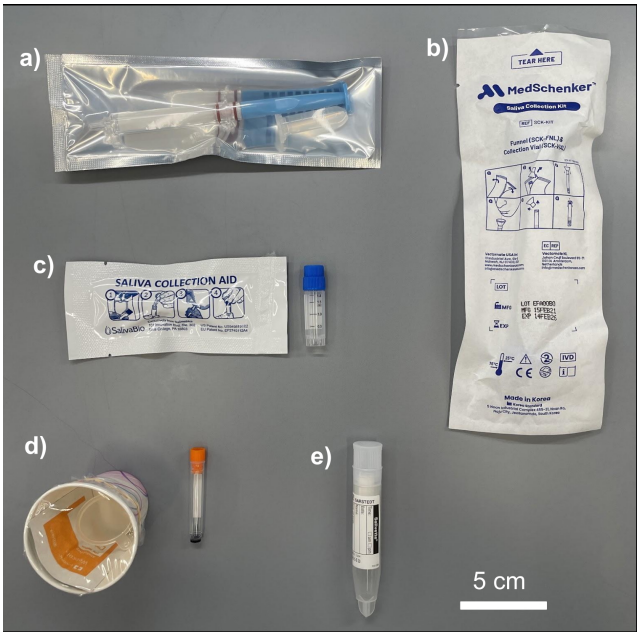


Figure 2.1: The five saliva collection devices selected for the study. (a) SuperSAL, (b) MedSchenker, (c) Passive Drool, (d) SwabSeq, and (e) Salivette. More information on each device can be found in Table 2.2.

2.3 Comparison of Noninvasive Saliva Collection Devices

Results

Out of the 60 subjects recruited for this study, subjects 57-60 could not use the Medschenker device in their sessions and only used the other four devices.

Normalized Saliva Collection Rate

Saliva collection rate was measured for each device by timing how long subjects took to collect volumes of saliva specified by the device's instructions. This number was normalized by each subject's unstimulated (=base) saliva flow rate to account for differences in each person's saliva production. A positive number indicates that the device stimulated saliva production and thus resulted in efficient saliva collection, while a negative number indicates that the device hindered saliva collection. The results are shown in Figure 2.2a and Table 2.3. Salivette (avg. 0.293, standard deviation (SD) 1.885) and Medschenker (avg. 0.178, SD 0.747) ranked highest in average normalized saliva collection rate, both above 0, while the others scored below. SuperSAL the lowest saliva collection rate (avg. 0.673, SD 0.302), with the lowest variance.

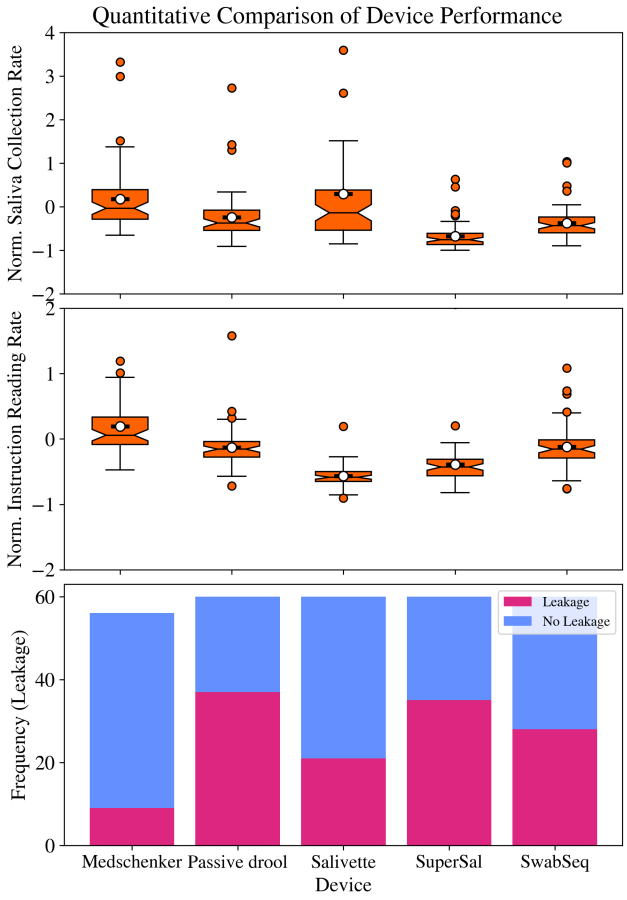


Figure 2.2: Quantitative comparison of saliva collection device performance. The top panel shows normalized saliva collection rates across devices, while the middle panel presents normalized instruction reading rates. The bottom panel illustrates the frequency of leakage events, with leakage indicated in pink and no leakage in blue.

Normalized Instruction Reading Rates

The reading rate for each device's instruction manual was calculated by timing subjects' reading and dividing by the word count, then normalized by their base reading rate measured separately at the beginning of the study. This normalization indicates readability, with a rate around 0 suggesting plain text difficulty. Results are shown in Figure 2.2b and Table 2.3, with manuals available in the study's data repository. The averaged normalized instruction reading rates ranked highest to lowest were: Medschenker (0.194, SD 0.516), SwabSeq (-0.122, SD 0.336), Passive drool (-0.132, SD 0.315), SuperSAL (-0.387, SD 0.374), and Salivette (-0.568, SD 0.162). Medschenker's rate above zero suggests relative ease of reading compared to plain text. However, it also had the widest range and highest variability, indicating

Device	Avg. saliva collection rate (st. dev) [unitless]	Avg. instruction reading rate (st.dev) [unitless]	Percentage of leakage (%)
Salivette	0.294 (1.885)	-0.568 (0.162)	35.0
SuperSAL	-0.673 (0.302)	-0.387 (0.374)	58.3
Passive Drool	-0.243 (0.579)	-0.131 (0.315)	61.7
Medschenker	0.178 (0.747)	0.194 (0.516)	16.1
SwabSeq	-0.372 (0.369)	-0.122 (0.336)	46.7

Table 2.3: Summary statistics of device performance, given by normalized saliva collection rate, normalized instruction reading rate, and percentage of devices with leakage.

some users struggled. Other devices had rates below zero, despite supplementary diagrams in the manuals. Salivette had the lowest average reading rate, indicating that the instructions were difficult to read.

Leakage of Saliva

Subjects indicated whether saliva leaked onto their hands or clothes, with responses coded as "True" or "False". Results are shown in Figure 2.2b and Table 2.3. Leakage rates were ranked as follows: Passive drool (61.7%), SuperSAL (58.3%), SwabSeq (46.7%), Salivette (35%), and Medschenker (16.1%). Passive drool and SuperSAL had the most leakage, with over half of the devices leaking saliva. Medschenker had significantly fewer instances of leakage compared to other devices.

Survey Results

On the survey, subjects rated the difficulty of instructions, assembly, and saliva collection for each device, with response options ranging from "Very easy" to "Very difficult." Results (Figure 2.3) were categorized into instruction (Figure 2.3a), assembly (Figure 2.3b), and saliva collection (Figure 2.3c). In terms of instruction readability, SwabSeq's instructions were rated as easiest to read, and instructions for Medschenker, Passive drool, and Salivette were also considered easy on average. SuperSAL's instructions were rated as the most difficult. Assembly was rated as easy for all devices, with Medschenker and Salivette being the simplest. SuperSAL had the most "Difficult" ratings. For saliva collection, Medschenker and Salivette were favored, with over 70% rating them as easy to use. SwabSeq was moderately easy, while Passive drool received mixed feedback. SuperSAL was consistently rated the

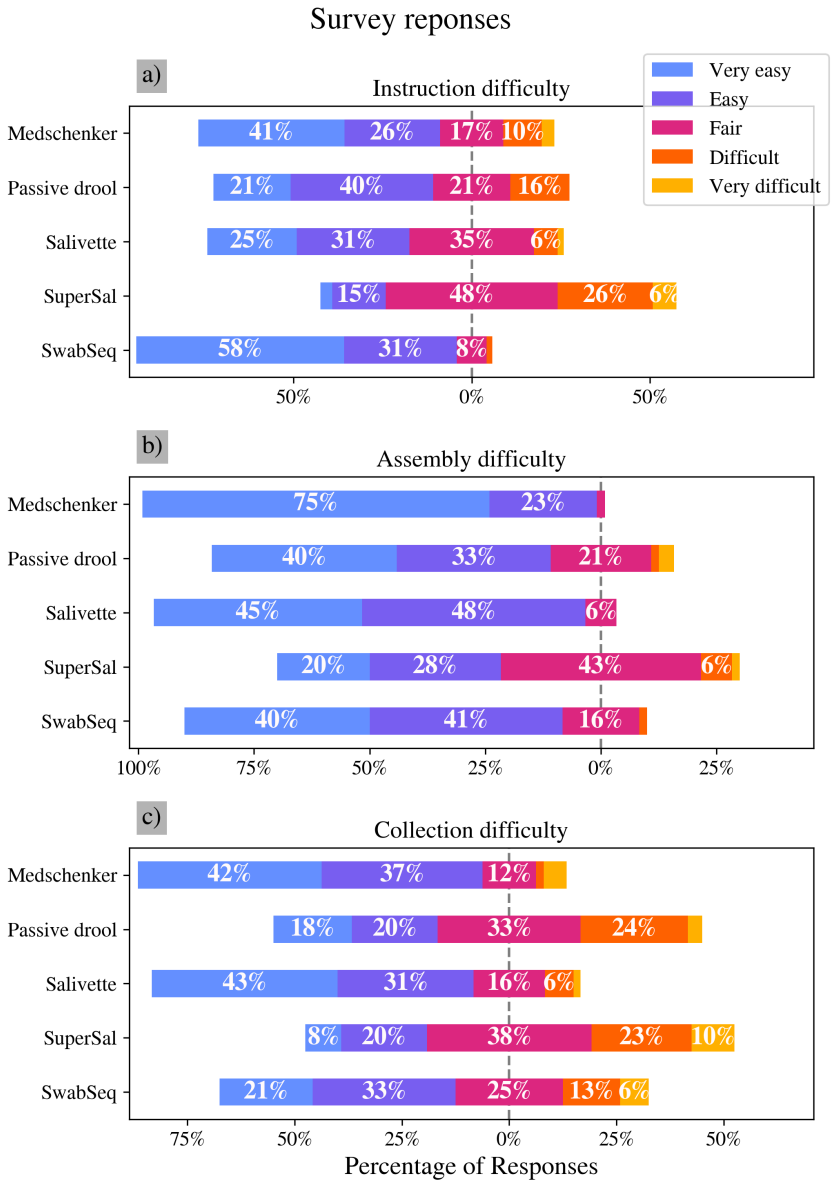


Figure 2.3: Survey responses on (a) instruction difficulty, (b) assembly difficulty, and (c) saliva collection difficulty. Each bar is centered on the midpoint of "fair" responses. (a) SwabSeq's instruction was rated the easiest to read, whereas SuperSAL's was rated the most difficult. (b) Medschenker was rated the easiest to assemble, whereas SuperSAL was rated the most difficult. (c) Medschenker and Salivette were rated the easiest to collect saliva with, whereas SuperSAL and Passive drool were rated the most difficult.

most difficult.

Overall, survey results seemed to align well with the quantitative metrics measured in our study. The saliva collection difficulty survey ranked devices from easiest to most difficult as follows: Medschenker, Salivette, SwabSeq, Passive drool, and SuperSAL. Saliva collection rates followed similar trends, with Salivette and Medschenker having the highest collection rates and the most "Easy" or "Very easy" responses, indicating they were the easiest to use. SuperSAL had the lowest collection rate and the most "Difficult" or "Very difficult" responses, making it the hardest device. Responses for Passive drool and SwabSeq were also consistent with their collection rates.

In the instruction difficulty survey, the ranking from easiest to hardest was SwabSeq, Medschenker, Passive drool, Salivette, and SuperSAL. Instruction reading rates followed similar trends as the survey. Medschenker had the highest reading rate, despite some variance, with 41% of users finding the instructions "Very easy." SwabSeq had the most "Very easy" responses and the highest readability score, though its reading rate was lower than Medschenker's. For Salivette, survey results indicated users didn't find the instructions overly difficult but it had the lowest reading rate, possibly due to the small font and multiple languages that were presented on the same page. SuperSAL had a low reading rate, low readability score, and the most "Difficult" responses, indicating that its instructions were the hardest to follow.

Discussion

While there is prior research focusing more on the usability of saliva collection devices [46] than on the diagnostic quality of the saliva collected, as has been the primary focus of many previous studies [25], [44], [45], [47]–[50], our study is novel in that we use quantitative metrics such as saliva collection rate and instruction reading rate to quantify the usability of common saliva collection devices in tandem with surveys to capture the full image of device usability. Our results underscore the need for an all-encompassing device that can address the functionality and usability-related issues that are present in the myriad of saliva collection devices on the market. An ideal device should not only efficiently collect saliva, but also be user-friendly; clear instructions, intuitive assembly, and comfortable saliva collection are essential. Device design should account for hygiene, by eliminating sources of leakage and providing user instructions on proper sanitization.

Despite the valuable insights gained, this study has several limitations that should be

acknowledged. While this study's focus on the user experience in saliva collection represents a novel approach, further validation in the lab is required to determine the optimal saliva collection device or method for specific biomarkers. Targeted analyses on specific salivary biomarkers should be conducted to ensure that the collection device does not interfere with biomarker levels, regardless of user preferences—for instance, there is evidence that swab-based devices interfere with cortisol in saliva [45], [49], although users generally found it easy to collect saliva with swabs. Moreover, this study focuses on healthy adults, with mean age of 27.27 and standard deviation of 6.99, where most ages are clustered around the 20s and some right skewness from data points in the 40s and 50s. The findings should not be generalized beyond this group, as results may vary across different demographic groups. When choosing saliva collection methods, factors like accessibility and age-related changes in saliva flow should be considered. For example, devices requiring both hands may not suit individuals with limited mobility, and devices with choking hazards should not be used with younger children. Future studies could benefit from surveying a broader demographic or conducting specialized research on specific groups, such as younger children or individuals with lower mobility or dexterity, to validate that our findings may be generalized to a broader demographic.

The study raises several intriguing questions that could be the focus of future research, such as the use of stimulated saliva for diagnostics. In this study, subjects collected unstimulated saliva, which some found challenging. Stimulating salivation could enhance collection efficiency, especially in mass-testing scenarios, as was seen in the results with water-stimulated and citric-acid-stimulated saliva flow rates (Appendix A). Prior literature suggests that water stimulation may yield viable saliva samples for metabolomics [50], as well as saliva stimulated by chewing paraffin for salivary protein analysis [47]. Stimuli like citric acid can greatly boost saliva flow rates, though caution is needed to maintain sample viability. Future research could explore optimal citric acid concentrations for this purpose.

Methods

Aim, design, and registration of the study

The study aimed to assess saliva collection devices and measure individual saliva flow rates under various conditions. The study was deemed exempt by the Caltech Institutional Review Board (IRB) due to the minimal-risk nature of the tasks involved (exemption number IR21-1142). The sample size of 60 subjects was determined

through statistical estimation.

Recruitment of subjects

Subjects above the age of 18 were recruited by fliers, emails, and word-of-mouth advertising. Upon recruitment, subjects were asked for consent to fast for 30 minutes before the session in order to simulate a typical saliva collection protocol. During the 30 minutes, subjects were asked to avoid eating, drinking, smoking, chewing gum, or brushing their teeth. Subjects provided written informed consent at the beginning of the study. Then, subjects were timed while reading the Informed Consent Form in order to calculate the base reading rate in words per minute.

Saliva collection with devices

Subjects used five saliva collection devices in random order. To improve precision, saliva samples were measured by mass rather than volume. To convert Salivette saliva mass to volume, we used a median unstimulated saliva density of 1.007 g/mL from prior literature [48]. Subjects were timed while reading the instruction manual for each device. Next, subjects collected saliva using the device while being timed until the specified amount was collected: 1 mL for MedSchenker and Passive drool, 0.5 mL for SwabSeq, until the indicator turned red for SuperSAL, and for 2 minutes with the Salivette swab. After each device, subjects answered a questionnaire about saliva leakage, the difficulty of assembly, instructions, collection, and provided optional comments. The questionnaire is presented as Pocket Material at the end of this dissertation.

Statistical analysis

The statistical justification for the sample size of the experiment is described in detail in the first subsection of Methods. One-way ANOVA and Tukey's HSD Pairwise Comparison were used to compare saliva collection and instruction reading rates for the five devices. One-way ANOVA and Tukey's HSD Pairwise Comparison were used to compare survey responses between devices by enumerating Likert-scale responses (Very easy, Easy, Fair, Difficult, and Very difficult) to integers 1-5 (respectively, in this order of the responses) where necessary. Alpha (maximum value for p-value) was 0.05 (5%) for all tests to determine statistically significant results.

2.4 Development of SalivaStraw: A Novel Saliva Collection Device

Using the insights learned from our comparative study on existing devices, we developed SalivaStraw, a novel, proof-of-concept saliva collection device.

Design Rationale and Theoretical Considerations

Saliva collection devices face several challenges, including ensuring efficient sample collection, user comfort, ease of use, and proper hygiene. The development of SalivaStraw stemmed from these challenges and the findings of the comparative analysis in Section 2.3, which identified limitations in existing devices. Specifically, issues such as saliva leakage, user variability in collected saliva volume, and confusion in usage instructions were prevalent.

The core design goal for SalivaStraw was to create a device that minimizes saliva leakage during collection, prevents overcollection of saliva, requires no assembly, is compatible with multiple different collection tubes, and can be produced at low-cost by injection molding. The concept was to design a device that would leverage passive drool to collect clear, whole saliva without the leakage problem of passive drool devices. The device would be easy to use, as it is a single-component system that requires no assembly with minimal user operations.

Theoretical considerations for SalivaStraw's design were largely informed by the principles of fluid dynamics, particularly the Poiseuille flow equation, which describes the laminar flow of fluids through cylindrical tubes and is given by

$$\Delta P = \frac{8\mu L Q}{\pi R^4}$$

where μ is the viscosity of fluid, L is the length of the pipe, Q is the volumetric flow rate, and R is the radius of the pipe. This equation was used to determine the lengths and radii of the saliva and air channels while balancing pressure differences needed to drive saliva and air flow. Further details on the theoretical model of SalivaStraw are provided in Appendix B.

Design and Materials

SalivaStraw is a compact, straw-like device designed to be fit into the opening of a standard saliva collection tube. It features two through-holes: a wider saliva inlet and a narrower air outlet (vent). The user pools saliva in their mouths and gently blows or guides the saliva into the saliva inlet. Air inside the tube exits through the vent as saliva fills up inside the tube. The air vent and saliva inlet are vertically offset to optimize collection efficiency and prevent overcollection. This

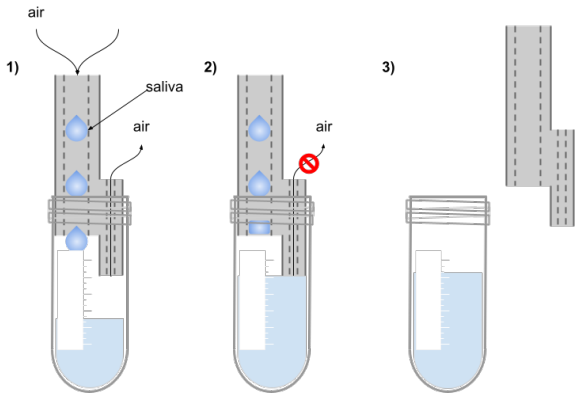


Figure 2.4: Concept of SalivaStraw usage. 1) The user places the SalivaStraw in their mouth, passively allowing saliva to flow through the saliva inlet and into the connected collection tube. 2) Once the saliva fills up and blocks the air outlet (vent), no more air can exit and no more saliva can enter. 3) SalivaStraw is then discarded after a single use.

offset ensures that once the saliva reaches a critical volume, it blocks the air vent, stopping further collection. This mechanism provides a natural stopping point for users and minimizes leakage.

SalivaStraw was initially designed with a tapered form in the first design iteration to fit a variety of collection tube sizes, enhancing modularity. However, the design was modified in Version 2 to incorporate a threaded lid for a more secure fit with a cryovial, after initial testing showed the possibility of saliva leaking through the seal between SalivaStraw and the cryovial. While this modification improved stability, it reduced the device's compatibility with other tubes. Future design iterations could explore hybrid approaches that retain the adaptability of the tapered design while incorporating a more secure sealing mechanism.

The ideal material for SalivaStraw is Delrin or HDPE, both water-resistant and biocompatible plastics commonly used in dental applications. The initial design for manufacturability (DFM) analysis indicated that the tapered version of SalivaStraw could be injection molded in a single mold, with an estimated cost of $0.21 per unit when produced at a scale of 100,000 parts per batch, which is around five times cheaper than the cost of the cheapest device, Salivette, in Table 2.2.

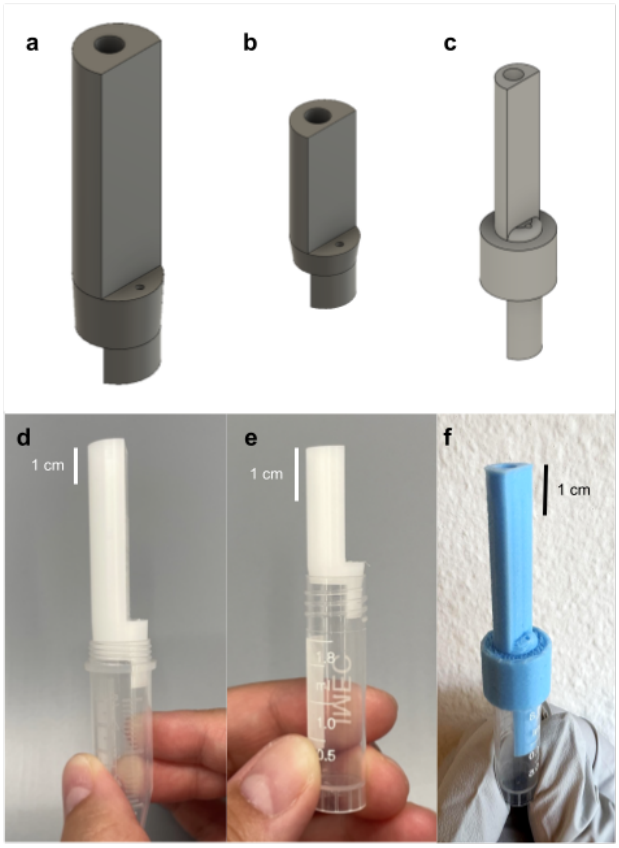


Figure 2.5: CAD models (a, b, c) and photos (d, e, f) of various SalivaStraw versions, showcasing the progression from the initial tapered designs (a, b, d, e) to a lid-integrated design (c, f). a) and d) feature the initial 5-mL prototype, which has the largest size of the three and is press-fit into the mouth of a 5-mL cryovial. b) and e) show the initial version of the 2-mL prototype that was press-fit into a 2-mL cryovial. This model was later changed to the second version shown in c) and f), which has an integrated threaded lid that can be screwed onto the same 2-mL cryovial and improves the airtight seal between the cryovial and SalivaStraw.

Theoretical Performance Estimates

The performance of SalivaStraw was initially estimated using fluid dynamics principles, particularly Poiseuille flow, which informed our design approach for saliva collection. Internal testing was conducted with very early prototypes to establish baseline performance. The initial 5-mL prototype demonstrated promising saliva entry characteristics but exhibited leakage through the air vent during forceful blowing, with bubbly saliva being particularly problematic. Similarly, the first 2-mL prototype encountered challenges with saliva entry due to mouthpiece size constraints and inadequate sealing between components. Subsequent design iterations incorporated refined fluid dynamics modeling that identified critical parameter thresholds

previously overlooked. By optimizing these parameters based on Poiseuille flow principles, later prototypes achieved improved performance in the following round of internal testing.

As testing results showed that saliva could leak from between SalivaStraw and the cryovial, a second version of SalivaStraw was developed with a threaded lid design to enhance sealing with cryovials, with the optimized parameters in place. Further internal testing was conducted to assess the performance of this improved SalivaStraw model. The most significant improvement in this prototype compared to the press-fit prototype was that the sealing issue between SalivaStraw and the cryovial was eliminated. In this round of testing, collection times for approximately 0.75 mL of saliva ranged from 50 to 61 seconds, yielding flow rates between 0.74 and 0.90 mL per minute, with an average rate of 0.84 mL per minute. Under normal usage conditions, the device functioned without observable leakage, which indicated that the refined parameters maintained efficient saliva entry while addressing some of the previous prototypes' limitations. However, excessive pressure—similar to the force required to inflate a balloon—still caused saliva to bubble through the air vent. While this represents an improvement over earlier designs, further refinement of the venting mechanism may be necessary to prevent fluid egress under high-pressure conditions. Furthermore, user instructions for SalivaStraw should indicate clearly that excessive exhalation pressure should not be exerted in normal use cases and that there may be leakage if this is not accounted for. Another key observation was that the saliva inlet remained relatively small, as the improved prototype was built for a 2-mL cryovial, which is small by itself. Once the collection chamber filled to capacity, residual saliva pooled at the inlet instead of entering the tube, as expected. However, due to the narrow inlet diameter, this buildup eventually led to overflow and minor leakage. Increasing the inlet size in future iterations could help mitigate this issue by allowing saliva to enter more efficiently without accumulation at the entry point.

Overall, these refinements mark a significant improvement over early prototypes, with optimized fluid dynamics enhancing saliva entry while reducing leakage. Future iterations should continue refining inlet dimensions and venting mechanisms to ensure robust performance across a range of user conditions. This iterative development process demonstrated the importance of precise fluid dynamics modeling in optimizing saliva collection device design, ultimately leading to more effective prototypes with reduced leakage and enhanced sample collection efficiency.

Discussion

SalivaStraw has the potential to serve as a standardized device for saliva collection, as it enables the passive drool method for clear whole saliva collection while minimizing leakage and reducing the risk of secondary contamination. Its design is informed by usability insights from existing saliva collection devices, prioritizing ease of use and reliable sample collection. Despite these advantages, formal user testing was not conducted due to logistical constraints. Future validation studies will be necessary to assess its usability, efficiency, and performance across diverse user populations.

Beyond individual usability, scalability is an important factor in widespread adoption. The SalivaStraw design lends itself to cost-effective mass production, particularly with the tapered version, which could be injection molded in a single mold at a low per-unit cost. However, future iterations must balance cost, compatibility, and performance trade-offs, particularly when considering modifications such as the threaded lid. Additionally, integration into existing saliva collection workflows remains an open question. Ensuring compatibility with automated processing systems and laboratory pipelines could enhance the device's utility in both clinical and research settings.

Our results with stimulated saliva in the comparative study on existing saliva collection devices revealed differences in collection dynamics, suggesting a potential avenue for future iterations of SalivaStraw. Incorporating a stimulation method—such as citric acid on the mouthpiece—could facilitate faster collection, particularly for individuals with low saliva production. However, further research would be required to assess the diagnostic viability of stimulated saliva before implementing this modification.

More broadly, this work contributes to ongoing efforts to improve and standardize noninvasive sample collection methods. With the growing demand for at-home diagnostics, telemedicine, and field-based sample collection in low-resource settings, innovations like SalivaStraw could help bridge the gap between accessible sample collection and high-quality laboratory analysis. Future iterations, guided by usability studies and further design refinements, have the potential to make saliva-based diagnostics more practical, scalable, and widely available.

Methods

Prototype Development

The initial design of SalivaStraw was developed using Fusion 360[51] for computer-aided design (CAD). Iterative prototyping was carried out using 3D printing, allowing for rapid adjustments to key design parameters such as the saliva inlet geometry and air vent placement. Once a viable prototype was identified, refined versions were fabricated with HDPE through machining, utilizing a lathe and drill press in the shop to create precise features necessary for functional testing.

Internal Testing

Preliminary functional testing was conducted to evaluate saliva entry, minimize leakage through the air vent, and assess overall usability. These assessments informed refinements to SalivaStraw parameters and modifications to the seal between the straw and the collection tube, ensuring reliable sample collection. Further validation studies, including formal user testing, will be necessary to comprehensively assess the device's performance.

2.5 Conclusions

In our comparative study of existing saliva collection devices, we present key findings across the evaluation of five saliva collection devices, measured with metrics such as saliva collection rate, instruction reading rate, and leakage rate, as well as a user survey on the usability of each device. Aggregated results showed that no single device performed well on all of the metrics. We then introduce SalivaStraw, a novel device that was designed to address key challenges in saliva collection by enabling clear whole saliva collection via the passive drool method while minimizing leakage and potential contamination. Its development was informed by insights from the comparative study, prioritizing ease of use, manufacturability, and sample integrity. The prototyping process, including CAD modeling, 3D printing, and machining, allowed for iterative refinement of the design, culminating in a device that is simple to use and scalable for mass production. While preliminary testing confirmed basic functionality, formal user validation is necessary to fully assess SalivaStraw's usability and performance across diverse populations. Future iterations could explore design modifications such as integrated saliva stimulation to enhance collection efficiency, particularly for individuals with low saliva flow. More broadly, this work contributes to efforts to improve and standardize noninvasive sample collection, with

potential implications for expanding access to saliva-based diagnostics in clinical, at-home, and low-resource settings.

An important factor to take into consideration for optimizing saliva collection is the variation in saliva production. Saliva production can be influenced by a wide range of physiological, environmental, and behavioral factors, all of which can not only affect the saliva flow rate but also saliva composition [20], [28]. Major factors that can affect unstimulated saliva production include degree of hydration, body position, or circadian rhythms, while minor factors include gender, age, or psychic effects such as the thought or sight of food [28]. While physiological factors like age or sex are difficult to standardize, environmental or behavioral factors like where or when saliva collection takes place or how the subject is positioned should be controlled for. Moreover, as food intake, introduction of various chemicals, or dental hygiene activities can introduce additional variation in saliva production, saliva collection protocols should explicitly state requirements such as no eating, no brushing teeth or using mouthwash, and no smoking for 30 minutes prior to saliva collection to ensure subject compliance and retrieval of good quality saliva samples. By accounting for these influencing factors, researchers can enhance the reproducibility and reliability of salivary biomarker analyses.

2.6 Appendix A: Salivary Flow Rates

In the comparative study of saliva collections devices, subjects collected saliva under three different conditions: unstimulated, after gargling water (control), and after gargling citric acid (stimulated). The purpose of this part of the study was to assess the feasibility of collecting stimulated saliva with commercial saliva collection devices. Water served as a control to measure the effect of liquid in the mouth. For each condition, three samples were collected at 30-second intervals. In the unstimulated case, subjects pooled saliva for 30 seconds, spat into a tube, and repeated this twice. For the water and citric acid conditions, subjects gargled the liquid for 30 seconds, spat it out, and then collected saliva at 30, 60, and 90 seconds. The weight of all nine samples was recorded, and was then converted to flow rate. Pairwise permutation tests were used to compare saliva flow rates between conditions, between sex, and between time points in each condition.

Water ($p = 0.001$) and citric acid ($p \ll 0.001$) stimulation significantly increased saliva flow rates compared to unstimulated conditions (Figure 2.6a). Citric acid stimulation showed the highest increase in flow rate, around 3.17 times higher

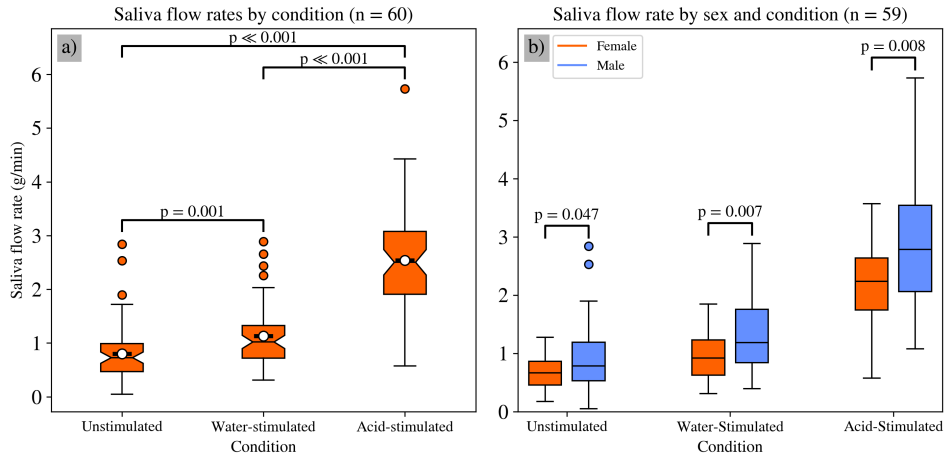


Figure 2.6: Saliva flow rates in unstimulated, water-stimulated, and acid-stimulated conditions with p-values for (a) all 60 subjects and (b) 59 subjects separated by sex at birth, excluding one subject who did not provide information. (a) Both water and citric acid significantly stimulated saliva production, with citric acid increasing saliva flow rates 3.17 times compared to unstimulated saliva production rates. (b) Males had higher flow rates than females in all conditions, with the biggest difference in citric-acid-stimulated saliva.

than no stimulation and 2.25 times higher than water stimulation. Furthermore, we observed significant differences between male and female saliva flow rates for all three conditions (p-values 0.047, 0.007, and 0.008, respectively) (Figure 2.6b). Males had higher saliva flow rates than females in all salivation conditions, with citric acid stimulation showing the most pronounced difference.

Over time, unstimulated saliva flow rates remained stable (Figure 2.7). Water stimulation led to an initial increase in flow rate, followed by a significant decrease after 30 seconds ($p = 0.002$). Citric acid stimulation resulted in a more sustained increase, with significant decreases observed between all time points ($p = 0.002$ for 30 to 60 seconds, and $p = 0.010$ for 60 to 90 seconds). These results show that water as a stimulus has an effect that lasts up to around 60 seconds after stimulus introduction, whereas citric acid has a longer-lasting effect that is observed even at 90 seconds after stimulus introduction.

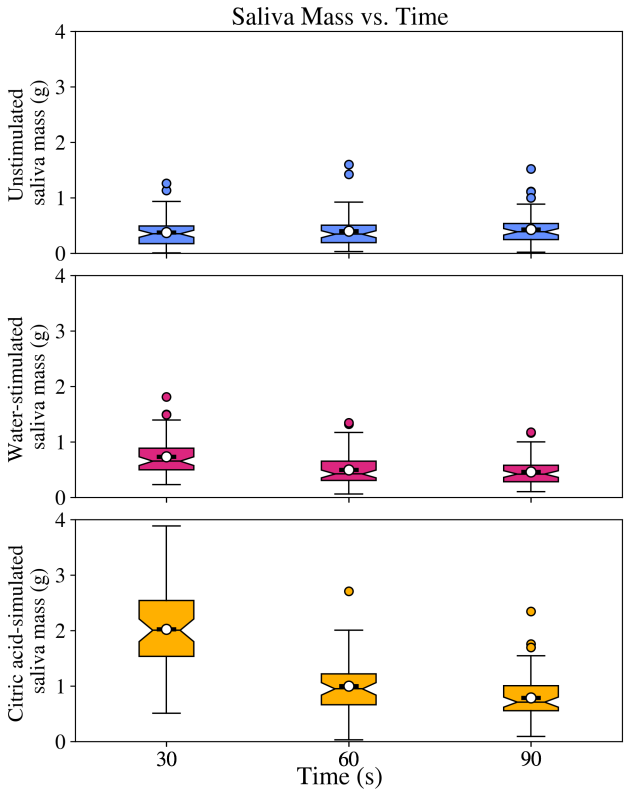


Figure 2.7: Change in saliva mass over time after stimulus introduction at t = 0. Significant decreases in flow rate were observed in water stimulation after 30 s, and between all time points in citric acid stimulation.

2.7 Appendix B: Theoretical Model and Parameters of SalivaStraw

The design of SalivaStraw was derived from the Poiseuille Flow equation, which is given by

$$\Delta P = \frac{8\mu LQ}{\pi R^4}$$

where μ is the viscosity of fluid, L is the length of the pipe, Q is the volumetric flow rate, and R is the radius of the pipe. ΔP denotes the pressure difference between the two ends of the pipe.

In the context of the SalivaStraw design, we use an effective viscosity to approximate saliva's flow behavior, despite its non-Newtonian nature. This approach is valid because, at low shear rates typical of laminar flow, saliva exhibits approximately Newtonian behavior, making it reasonable to apply Poiseuille's law. Saliva is a shear-thinning fluid, meaning its viscosity decreases with increasing shear rate. However, under the moderate shear conditions expected in the device, empirical studies have shown that saliva stabilizes to an effective viscosity that can be treated

as constant [48]. Additionally, Poiseuille's law assumes a viscosity that does not vary radially across the tube, an assumption that holds as long as the flow remains laminar.

To validate our assumption of laminar flow, we calculate the Reynolds number for flow in a pipe with some approximate parameters for SalivaStraw:

$$Re = \frac{\rho v D}{\mu}$$

where ρ is the density of saliva ($\approx$ 1000 kg/m^3), v is velocity, D is the air vent diameter (1-1.5 mm), and μ is the saliva viscosity (1.5 mPa·s). Given the small diameters and low velocities involved, our calculations confirm that $Re < 2000$, indicating laminar flow, justifying the use of Poiseuille's law. Since we ensure that the Reynolds number remains below the critical threshold for turbulence, the effective viscosity provides a sufficiently accurate approximation for modeling saliva flow through the system.

Using the relationship between the characteristics of the pipe and pressure drop across the pipe defined by Poiseuille flow, we were able to determine the lengths and radii for the air outlet and saliva inlet to ensure that a) air can flow through the vent but not saliva, and that b) saliva can flow through the saliva inlet without backflow.

The device consists of two primary components: the air outlet (A) and the saliva inlet (S), modeled as two separate pipes. The pipe lengths and radii are denoted as L_A and L_S, and r_A and r_S for the air outlet and saliva inlet, respectively. The pressure required for saliva to flow through S and A are denoted as P_S^{sal} and P_A^{sal}, respectively. The pressure required for air to flow through S and A are similarly denoted, P_S^{air} and P_A^{air}, respectively. The conditions a) and b) described above can now be rewritten in terms of the pressure notations, such that equation (2.1) and (2.2) correspond to condition a) and equation (2.3) to condition b):

$$P_A^{sal} \gg P_A^{air} \tag{2.1}$$

$$P_A^{sal} \gg P_S^{sal} \tag{2.2}$$

$$P_S^{sal} > P_S^{air} \tag{2.3}$$

Specifically in equation (2.1), the ratio of P_A^{sal} to P_S^{air} must be maximized in order to stop saliva from escaping through A, but the air outlet cannot be so narrow that even air cannot leave through it.

Additionally, because A is narrower than S, it follows that $P_A^{air} > P_S^{air}$. Furthermore, to ensure proper fluid dynamics in the SalivaStraw system, P_S^{sal} must be greater than P_A^{air}. This condition arises due to the significantly higher viscosity of saliva (μ_{sal} = 1.57 mPa·s [48]) compared to air (μ_{air} = 0.02 mPa·s), which increases the resistance to flow in the saliva inlet. Putting the conditions above all together lead to this inequality that describes the system:

$$P_A^{sal} >> P_S^{sal} > P_A^{air} > P_S^{air}$$

From here, we denote the ratio of P_A^{sal} to P_S^{sal} as T, which is the number to maximize in order to stop saliva from leaking through A. While maximizing T, we must ensure that P_S^{sal} is greater than P_A^{air} for proper fluid dynamics, so we denote the ratio of P_S^{sal} to P_A^{air} as T', which has to be greater than 1 at all times. Expanding the expressions for T and $T\prime$ using the Poiseuille Flow equation, we get:

$$T = \frac{P_A^{sal}}{P_S^{sal}} = \frac{L_A/r_A^4}{L_S/r_S^4}$$
$$T' = \frac{P_S^{sal}}{P_A^{air}} = \frac{\mu_{sal}}{\mu_{air}} * \frac{L_S/r_S^4}{L_A/r_A^4} > 1$$

Substituting the terms for T' in the expression for T and plugging in the values for the viscosities of air and saliva, we get:

$$T' = \frac{\mu_{sal}}{\mu_{air}} * \frac{L_S/r_S^4}{L_A/r_A^4} = \frac{\mu_{sal}}{\mu_{air}} * \frac{1}{T} > 1$$

which gives us $T < 78.5$. Iterative testing identified $T < 78.5$ as a critical determinant of SalivaStraw's functionality, with more significance than anticipated in preliminary modeling. The first versions of the prototypes used for internal testing had $T = 58.82$ for the 5-mL prototype and $T = 102.4$ for the 2-mL prototype. Testing results for the 5-mL and the 2-mL prototype showed that saliva could escape through the air outlet if T is too small and saliva could not enter through the saliva inlet if T is too big. Based on these observations, we redesigned the prototypes to achieve T values that approached but remained below the critical threshold of 78.5. These targeted adjustments successfully mitigated both the air outlet leakage problem in the 5-mL prototype and the inlet flow restriction in the 2-mL prototype in the next round of internal testing, demonstrating the practical significance of maintaining T within this optimal range. The values derived for L_A, L_S, r_A, and

r_S using the threshold for T are given in Table 2.4. The parameters are for the fixed 2-mL and 5-mL SalivaStraw initial models (tapered press-fit) and the 2-mL SalivaStraw v2 model (threaded lid). CAD schematics for the three SalivaStraw models are provided as pocket material at the end of the dissertation.

Finally, we validate the pressures associated with pushing saliva through the air vent with physiological data. The maximum exhalatory pressure a male human with large lung capacity can generate is approximately 11,704 Pa [52]. Using the Poiseuille Flow equation to calculate the pressure required for saliva to flow through the air vent at approximately 1 mL/min, we get:

$$\begin{aligned}\Delta P &= \frac{8\mu_{sal} L_A Q}{\pi R_A^4} \\ &= \frac{8 \cdot (1.5mPa \cdot s) \cdot (30mm) \cdot (1mL/min)}{\pi \cdot (0.6mm)^4} \\ &= 147,366Pa\end{aligned}$$

which shows that the pressure required to push saliva through the air vent is 147,366 Pa. Since this is far beyond the physiological capability of a human, saliva leakage through the air vent is not feasible under normal usage conditions, demonstrating that SalivaStraw will be effective in minimizing leakage of saliva through the air vent.

Prototype	L_A (mm)	r_A (mm)	L_S (mm)	r_S (mm)	T
2-mL v1 and v2	30	0.6	48	2	77.16
5-mL v2	40	0.8	51	2.5	74.80

Table 2.4: Parameters for the various models of SalivaStraw, determined from calculations with the Poiseuille Flow equation. Both the press-fit version and the threaded-lid version of the 2-mL model have same the same parameters, but threaded-lid version features a more airtight seal with the collection tube to prevent potential leakage problems.

Chapter 3

FRACTION COLLECTION FOR SALIVARY PROTEOMICS WITH COLOSSEUM

[1] A. S. Booeshaghi, Y. Kil, K. H. Min, J. Gehring, and L. Pachter, "Low-cost, scalable, and automated fluid sampling for fluidics applications," *HardwareX*, vol. 10, e00201, Oct. 1, 2021. DOI: 10.1016/j.ohx.2021.e00201.

3.1 Abstract

We present colosseum, a low-cost, modular, and automated fluid sampling device for scalable fluidic applications. The colosseum fraction collector uses a single motor, can be built for less than $100 using off-the-shelf and 3D-printed components, and can be assembled in less than an hour. Build instructions and source files are available at https://doi.org/10.5281/zenodo.4677604. In addition to its broad utility in fluid handling, colosseum's ability to fractionate biological samples makes it well-suited for applications in salivary proteomics, where isolating specific molecular fractions can enhance biomarker detection and assay reproducibility.

3.2 Introduction

Fraction collectors are liquid handling machines used in analytical and preparative workflows that require sample separation or longitudinal collection [53]. Typical fraction collectors consist of a rotating rack loaded with containers and a distributing arm for collecting fixed volumes of fluid [54]. Fraction collectors are preferable to manual collection that can be tedious and introduce human error [55]. They are integral to various scientific fields, particularly in chromatography-based methods such as high-performance liquid chromatography (HPLC) and fast protein liquid chromatography (FPLC), where they facilitate the separation and collection of complex mixtures for downstream analysis [56]. The ability to split and collect samples is especially valuable in proteomics and metabolomics, where metabolic and proteomic analytes are often bound together—by enabling the isolation of molecules of specific size prior to mass spectrometry analysis, fraction collectors can separate unbound metabolite biomarkers from bound ones [57]. Especially relevant to salivary biomedicine is the use of fraction collectors in isolating and purifying extracellular vesicles (EVs) from biological samples [58]. Extracellular vesicles (EVs)

are membrane-bound particles secreted by various cell types that play an important role in intercellular communication through the exchange of proteins, nucleic acids, and lipids [59]. Salivary EVs contain various protein biomarkers and are highly enriched in microRNAs (miRNAs), making them potentially valuable for noninvasive monitoring of various diseases [2], [59].

Saliva is increasingly recognized as a viable biofluid in fields such as proteomics, metabolomics, and extracellular vesicle (EV) research [59]–[61]. Its diverse composition of analytes makes it well-suited for chromatography-based separations [62], where fraction collectors facilitate the isolation of specific biomolecular components. In a typical salivary proteomics pipeline such as liquid chromatography tandem mass spectrometry (LC-MS/MS), samples progress through several key stages: initial collection and preparation, chromatographic separation, mass spectrometry analysis, and data processing. After saliva samples have been collected and preprocessed, they undergo chromatographic separation. A fraction collector interfaces between chromatographic separation and mass spectrometry analysis—downstream of the chromatography column, a fraction collector is stationed to automatically collect specific aliquots of fluid exiting the chromatography column, preserving separated biomolecules before they proceed to mass spectrometry for identification and quantification. This fractionation step is essential for detecting low-abundance proteins that might otherwise be masked by highly abundant components such as amylase and mucins, thereby enhancing the depth and reliability of salivary proteome analysis [63] and improving protein biomarker discovery in saliva for diseases ranging from cancer to autoimmune disorders [64]–[69]. In another application of proteomics, EV isolation from saliva benefits from fraction collection when it is used in conjunction with size exclusion chromatography (SEC), as it can separate EVs from free proteins and RNA, enhancing their diagnostic utility [70]. Fractionation in saliva may offer additional advantages by separating mucins and enzymes from biomarkers of interest and thereby mitigating interference [71] and reducing sample degradation [72]. These advantages make automated fraction collectors an invaluable tool for advancing saliva-based diagnostics and biomarker discovery, especially in proteomics.

Most laboratories currently rely on commercial fraction collectors, which are expensive and difficult to customize (Table 3.1). To reduce cost and facilitate custom applications, a number of open-source fraction collectors have been developed. These devices, while less expensive, continue to rely on complex engineering de-

Model	Capacity (# tubes)	Price (USD)
Cytiva Frac30	30	2,089.00
BioFrac Fraction Collector	60-252	7,556.00
Gilson Fraction Collector	105-240	4,424.55
GE AKTA Frac-900 Series	95, 392	3,935.00
Buchi C-660	12, 30, 60	12,888.97
Open-Source	Customizable	<100

Table 3.1: Costs and capacity of commercial fraction collectors. The costs are based on new, unused models. The capacity of each fraction collector is given by how many tubes the device can hold.

signs and parts that may be difficult to source and manufacture, thus driving costs higher, lengthening the assembly process, and complicating operation.

We have designed and built a simple, low-cost, and modular fraction collector that is easy to assemble and use. This open-source fraction collector, which we call colosseum, is based on design principles for modular, robust, open-source hardware [73], and offers advantages to commercial systems by virtue of being significantly less expensive and easily customizable.

The colosseum fraction collector can be assembled in less than an hour and costs $96.25. Unlike the micrIO [74], which is built from parts of a salvaged Illumina Genome Analyzer that costs $1,500, the colosseum fraction collector uses off-the-shelf and 3D-printed parts. The LEGO MINDSTORM fraction collector [75] costs $500, and while it uses more commonly available components, it still requires cutting and bending of steel C-channel. Furthermore, most fraction collectors require the use of multiple axes to position a dispenser head over a reservoir. Control of such a system can require communicating with and driving up to three separate motors in tandem. The colosseum fraction collector is based on a simpler design where a mechanical coupling between the motor, the tube rack, and the dispenser arm enables rotation of the rack and position of the arm with only one motor. Designing around a single motor simplifies operation, and reduces cost, complexity, and assembly time.

3.3 Results

The colosseum fraction collector consists of four 3D-printed components, two rotary shafts, five rubber feet, one stepper motor, an Arduino, and a motor controller (Figure 3.1a,b). We chose the spiral tube layout (Figure 3.1c) instead of the rectangular tube layout of previously published fraction collectors as it enables serial fraction

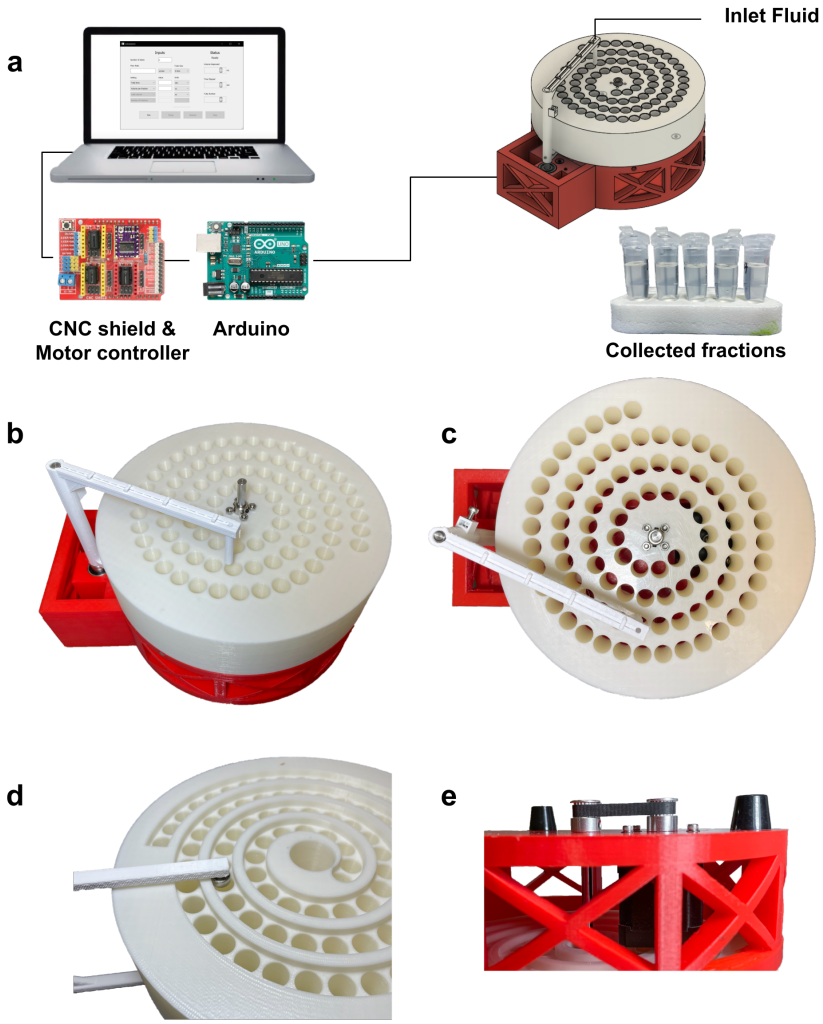


Figure 3.1: a) The colosseum fraction collector (left) is controlled by a single motor. A motor controller shield (red) is connected to an Arduino Uno (blue) and drives the motor. The computer's Graphical User Interface (right) and Python backend sends motor movement instructions to the Arduino. The Arduino-motor controller then sends those instructions to the motor. A motor located in the base turns the shaft of the tube rack. Grooves in the bottom of the fraction collector constrain the dispenser arm to rotate in tandem. (b) Angled view, (c) top view, (d) bottom view of mechanical coupling between the dispenser arm and tube rack, (e) side view of mechanical coupling of motor and tube rack.

collection with only one motor. By coupling the dispenser arm to the tube rack with a slot-cam mechanical coupling (Figure 3.1d) we constrained the rotation of the tube rack and movement of the dispenser arm to rotation of a single stepper motor located in the base of the fraction collector (Figure 3.1e).

The device is modular: each component can be developed, tested, and fabricated separately using mutually compatible interfaces. The tube rack fits 1.5 mL Eppendorf tubes and can easily be modified to accept tubes of varying sizes under the constraint that they follow the spiral pattern (Figure 3.2a). The tube rack fits 88 tubes with a packing efficiency of 60.2% relative to the optimal packing of 146 tubes on a circular disk of the same size as the tube rack [76]. In addition, the dispenser arm can be modified to accept connectors and tubing of various sizes to enable parallel dispensing.

The device is controlled by a graphical user interface (GUI) that communicates with an Arduino, CNC motor shield, stepper motor driver, and software adapted from the poseidon syringe pump [73]. Experiment parameters such as flow rate, total volume, total time, volume per fraction, and number of fractions are input by the user in the GUI and the Python or JavaScript back-end structures and sends Arduino-interpretable commands to the Arduino for execution. The GUI can be installed with the pip package-management tool and run with a single command on Mac, Linux, or Windows, or it can be run directly in a web browser at `https://pachterlab.github.io/colosseum/`. The web-browser implementation takes advantage of the web serial specification [77] and the browser-serial API [78] to read and write from the serial port within a web browser environment.

To ensure that commands set by the GUI correctly align the dispenser arm with each collection tube, we measured and converted the angle between pairs of tubes to motor steps, and programmed this list of angular displacements into the control software (Methods). We also used a simple iterative scheme to approximate the position of equally spaced points along an Archimedean spiral and compared it to our measurements. We found high concordance in the angular displacements (Figure 3.2b,c). This allows us to programmatically generate arbitrary spiral motor displacements based on the distances between successive tubes and distances between successive arms of the spiral.

In order to characterize collection errors across a range of flow rates commonly used in microfluidics and FPLC [79], [80], we sampled 180 fractions over six flow rates ranging from 22.5 mL/hr to 720 mL/hr. We found that the collection errors

were within ±6.5%, with one sample having -10.6% error due it being the first fraction collected. These data suggest that the use of the colosseum system with the poseidon syringe pump results in accurately collected fractions (Figure 3.2d). Next, we sought to assess the fraction collecting performance over an increasing amount of sample volume, as is commonly performed in gradient elution series [54]. For a fixed flow rate, we collected 20 fractions with 12-second increments in collection time per tube over the course of 42 minutes in three replicates (Figure 3.2e). We found that the collected fractions closely followed the expected fraction amount with a Spearman correlation of 0.997, showing that the colosseum fraction collector can be used to accurately collect gradient elution series.

3.4 Discussion

We have demonstrated a low-cost, modular, and automated fraction collector that uses 3D-printed parts and off-the-shelf components, can be built in an hour, and is simple to run. We show how colosseum samples fluid accurately over a wide range of flow rates making it useful for microfluidics experiments and proteomics protocols. The low cost of our device could enable several instruments to run in parallel. For example, a single control board can in principle run multiple fraction collectors and syringe pumps thus facilitating large-scale experiments (Figure 3.2f). We have also thoroughly documented the build process with instructional README's and videos, and we have made all of the results described in this paper reproducible on Google Colab.

Although colosseum has not yet been demonstrated in salivary experiments, its ability to facilitate fraction collection in other fluidic systems suggests strong potential for advancing salivary biomarker research, particularly in proteomics. When used in tandem with chromatography and mass spectrometry, colosseum could enhance salivary assay sensitivity and reproducibility by isolating specific molecular fractions, improving the detection of low-abundance protein analytes while reducing interference from mucins and enzymes [71], [72]. Additionally, its affordability and scalability make it well-suited for high-throughput studies, such as longitudinal biomarker monitoring and assessing individual responses to therapeutic interventions. As an automated liquid sampler, colosseum could streamline time-series experiments by automatically collecting aliquots of saliva over time following a perturbation, ensuring reproducible and efficient analysis. Further exploration of colosseum's specific applications in salivary biomedicine, including potential adaptations for handling saliva's unique properties, is discussed in Chapter 5 of this

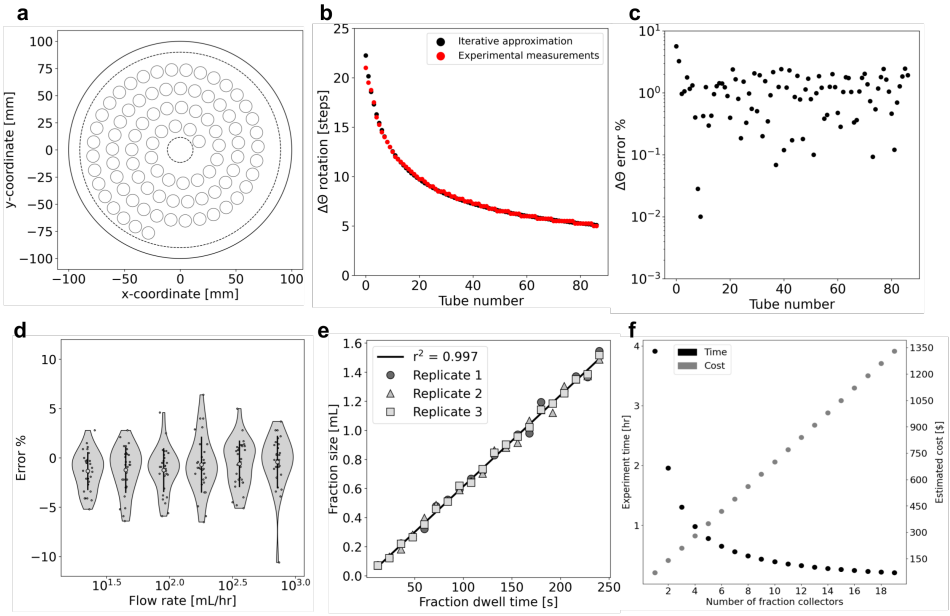


Figure 3.2: a) Tube placement on the colosseum is defined by an Archimedean spiral with tubes distributed 13 mm apart along the spiral and with 17.39 mm distance between subsequent arms of the spiral. b) The tubes are placed uniformly along the spiral where the arc length between any two tubes is constant, but the rotational displacement between any two tubes is nonconstant. c) Iterative approximation to the tube locations is similar to the measured tube locations. d) The error in the fraction size for 88 samples across a range of flow rates. e) The fraction size increases with increasing dwell time for a constant flow rate and the Spearman correlation of the means is 0.997. f) Multiple fraction collectors enable parallel collection which drastically decreases experimental time at a marginal increase in cost. The demonstrated accuracy across various flow rates (22.5-720 mL/hr) encompasses the range typically used for chromatography-based applications in salivary biomedicine (60-120 mL/hr) [81]. This validation ensures that the colosseum can reliably handle the flow rates commonly employed in salivary proteomics, where stepwise separation of proteins based on size or charge can significantly improve detection sensitivity for low-abundance biomarkers [63].

dissertation.

3.5 Methods

We designed the colosseum fraction collector by following basic principles of open-source hardware design [73].

Part Design and Fabrication

The fraction collector consists of four 3D-printed parts: a base, base plate, dispenser arm, and tube rack. The base holds the base plate, dispenser arm, and tube rack in place with additional hardware. The base plate acts as a horizontal support for the main rotary shaft, with rotational bearings that support the shaft in two places. The dispenser arm consists of two connected parts: the top part of the arm holds the fluid tubing and the bottom part acts as a cam follower that follows the spiral track on the bottom of the tube rack. Collection tubes are placed in the tube rack and are organized in a spiral pattern that mirrors the pattern the dispenser arm follows during rotation. The tube rack is constrained to the shaft with a flange coupling set screw and is mechanically coupled to the motor with a timing belt so that rotation of the motor results in rotation of the tube rack and the dispenser arm. After numerous sketch iterations, we used Fusion 360 [51] to generate a 3D CAD model of the device and added dimensional tolerances of +3-5% to all parts to account for variance in 3D printing. To prepare the appropriate files for 3D printing, Simplify3D [82] was used to slice the STL model and generate GCode with 10% infill and 0.2 mm layer height. Parts were printed on a Prusa i3 Mk3 3D printer [83] with 1.75 mm diameter PLA filament, at 215 °C nozzle temperature and 60 °C bed temperature, with 10% infill. STL files for all parts can be found in the GitHub repository `https://doi.org/10.5281/zenodo.4677604`. A 3D interactive viewer through Autodesk Fusion 360 is available at this link: `https://a360.co/4iOSAf5`. The time to print all parts separately was approximately 73 hours, but may vary depending on the printer model used and the print settings. All parts required to assemble the colosseum fraction collector can be found in the bill of materials.

Device Assembly

A complete guide on how to assemble the colosseum fraction collector can be found on YouTube. A step-by-step assembly guide is also available on protocols.io [84].

The assembly of the device starts with the base. Five rubber feet are screwed onto the bottom of the base to stabilize the device and to ensure that the timing pulleys on

the motor and the tube rack shaft are elevated and free of obstruction. A timing belt pulley is secured to the shaft of a NEMA 17 motor and the motor is then screwed onto the floor of the base. The tube-rack shaft is also inserted into the floor of the base along with a bearing that acts to stabilize the shaft. A timing belt pulley is secured to the shaft and couples its rotation with that of the motor. The motor and the shaft are connected by a timing belt of length 120 mm. The mounting holes in the base for the motor are designed so that the user can adjust the distance between the two timing pulleys in order to prevent slippage of the timing belt. Additionally, washers are inserted in between the base floor and the screws holding the motors so that the plastic of the base does not get worn out over time. The base plate is then screwed onto the floor of the base using M5 screws and nuts.

The dispenser arm, which is secured to a shaft with an M3 set screw, is placed into the base plate along with a bearing. A torsion spring is placed on the shaft, between the dispenser arm and the base plate to lessen slack between the dispenser cam follower and the tube rack spiral groove. The tube rack shaft is then inserted into the tube rack and secured in place with a flange coupling set screw.

The motor cables are routed through the side of the base and connected to the Arduino. The Arduino is connected to a CNC shield and DRV8825 Pololu motor controller [85]. The Arduino is also connected to a computer. This allows the user to send and receive signals to the motor via serial commands. Power is supplied to the stepper motor driver via terminals on the CNC motor shield. We supply the stepper motor driver with 12 V DC at 3.0 A. The DRV8825 has a maximum current rating of 2.2 A per phase and the bipolar NEMA 17 stepper motor has a rated current of 2.0 A per phase. The stepper motor driver limits the amount of current that can be delivered to the stepper motor via a potentiometer.

User Interface

The GUI translates the parameters set by the user into motor commands sent to the Arduino. The Arduino runs our custom firmware, pegasus [86], which sends command strings to the motor controller which in turn sends pulse-width-modulated signals to the motor. The GUI is written in Python using Qt, an open-source, cross-platform GUI framework. All packages related to the GUI are pip-installable and the GUI can be launched with a single command from the command line. The web-browser GUI is written in JavaScript, requires no installation, and can be run by simply navigating to a website. The GUI consists of two parts: parameter inputs

and a status monitor, the latter of which displays the total volume dispensed, time elapsed, and current tube location. Upon opening the GUI, users are prompted to connect to an Arduino. To run the colosseum fraction collector, users must specify three parameters: the flow rate, total time or total volume, and volume per fraction or number of fractions. The remaining parameters are calculated using the ones provided. In addition to these parameters, users must also specify the tube size to ensure that the fraction size will not be greater than the capacity of the tube. Users can operate the colosseum by pressing the run, pause, resume, and stop buttons in the GUI. All software required to run the colosseum fraction collector is freely available on Github under an open source BSD-2-Clause License.

Python 3.6 and JavaScript code is used on the back end to interpret user input from the GUI and send custom commands to the Arduino, accordingly. The Python implementation uses the pyserial package [87] to interface with the serial port and the web-browser implementation uses the browser-serial package [78]. Parameters from the GUI are translated into dwell time per tube and number of tubes to fill. The angle between each tube in the spiral was measured on Fusion 360 using the Inspect tool, saved as a csv file, and specified in the Python backend. These angles are then converted into the number of steps the motor must rotate. The motor stops rotating at each tube location for a specified amount of time in order to dispense the fluid into the tube. The motor then moves a set number of steps to reach the next tube. The status monitor displays the amount of total volume dispensed, how much time has elapsed since the start of the experiment and which tube the fraction is being dispensed into.

Testing and Validation

We tested the functionality of the device with numerous experiments where tap water is flown in at a set flow rate, or varying flow rate. We used the poseidon syringe pump, a 60 mL syringe, microfluidic tygon tubing and 1.5 mL Eppendorf tubes to pump fluid to the colosseum. The poseidon syringe pump was controlled with the pegasus software [86]. For a varying number of flow rates and a set dwell time per tube for each flow rate, we collected 30 fractions and compared the fraction sizes to the predicted fraction size of 1 mL by weighing each tube before and after collection (Figure 3.2d). We used a 200x 1 mg analytical scale manufactured by Yae First Trading Co., ltd part number TEK-AB-0392 to measure the amount of collected fluid. In order to properly fit the Eppendorf tubes on the tube rack, we cut off the caps from the Eppendorf tubes before collecting fractions in them and put

them back on for the final measurement, making sure that the cap corresponded to the tube from which it was removed.

In follow up experiments we fixed the flow rate and linearly increased the collection time. For a fixed flow rate of 22.5 mL/hr and 20 fractions with 12-second increments in collection time per tube, we collected fractions and compared the observed fraction sizes to the predicted fraction sizes (Figure 3.2e). We used pegasus to run the colosseum with varying dwell times per tube.

We estimated the cost and time for using k fraction collectors to show that these devices, when used in parallel, can reduce the experimentation time. For example, if we collect n fractions on each of k fraction collectors with a volume per fraction v and a constant flow rate f per collector then the time it takes to run this collection is t = n/k*v/f.

To test the accuracy of the measured angles between two successive tubes we used an iterative scheme to estimate the radius and angular position based of the polar form of Archimedean spiral of $r = b * \theta$ for a constant b. The radius and the arc length are used to update the angular position and then the angular position is used to update the radius. Optimal packing was calculated with the "best known packings of equal circles in a circle" [76] online tool with the outermost disk corresponding to the diameter of the area available for tube placement and the packing disks corresponding to the distance between tubes along the arc.

Data Analysis

All data analysis was performed with Python 3.7. Jupyter notebooks that run in Google Colab and all experimental data to reproduce Figure 3.2 can be found on our static GitHub repository containing all files relevant to the published manuscript for colosseum: `https://github.com/pachterlab/BKMGP_2021`. The project can also be found in the project's main GitHub repository, which reflects all updates to the project: `https://github.com/pachterlab/colosseum`.

3.6 Appendix A: Optimizing colosseum for Salivary Proteomics

In salivary proteomics, fractionation is a critical step that follows chromatography to further separate and isolate proteins of interest before mass spectrometry analysis. The flow rate during this step, particularly the rate at which sample fractions are collected, is a key factor that influences both the resolution of the separation and the recovery of proteins. This section discusses the relationship between the flow rate

through the chromatography column and the flow rate feeding the fraction collector, along with considerations for optimizing fractionation protocols.

Liquid chromatography separates biomolecules based on their differential interactions with two key components: the stationary phase and the mobile phase. The stationary phase is a fixed material within the chromatography column that retains analytes to varying degrees based on their physicochemical properties, while the mobile phase is the solvent that flows through the column, carrying the sample components. The interaction between these phases dictates the separation of proteins and other biomolecules.

Flow Rates

The flow rate into the chromatography column is typically controlled by a pump feeding into the liquid chromatography (LC) set-up and is typically dependent on the chromatography method being used. For example, fast protein liquid chromatography (FPLC) can tolerate much higher flow rates up to 5 mL/min[88], whereas the typical range for high-performance liquid chromatography (HPLC) is kept below 2 mL/min[81]. The flow rate feeding the fraction collector is generally controlled to match the chromatography system's flow, allowing fractions to be collected continuously as the eluate exits the chromatography column. However, the flow rate entering the fraction collector can be different from the flow rate used in the chromatography column due to the use of flow splitters and the specific settings of the fraction collector.

Flow rates across many types of LC set-ups typically range from 0.3 mL/min to 2.0 mL/min, although with 4.6-mm-ID columns, the typical range is limited to 1.0 to 2.0 mL/min [81]. Lower flow rates are often used for more detailed separation or for smaller columns, while higher flow rates (1.5 to 2.0 mL/min) may be employed to increase the speed of processing, though this may reduce the resolution of separation. All of these flow rates are within the range that can reliably be used with colosseum. It is important to match the time intervals for fraction collection to the flow rate feeding into the fraction collector, as a fraction collector is typically set to collect fractions based on time intervals or volume increments, if not used in tandem with a UV detector that triggers fraction collectors at certain conditions. The user interface in colosseum allows the user to choose between time-based collection and volume-based collection, and automatically times collections based on the system's flow rate, thus facilitating the experiment.

Fraction Size

When performing liquid chromatography for salivary proteomics, one of the key considerations is the size of the fractions collected during the separation process. The size of the fractions collected during chromatography is influenced by several factors, including the flow rate, the resolution of the chromatography system, and the complexity of the sample. Fractions are typically collected in sub-milliliter volumes, depending on the separation method and the desired resolution. In LC-MS workflows with saliva, flow rates commonly range from 200-250 μL per minute, with fractions often collected at one-minute intervals, resulting in fraction volumes that balance resolution and analyte recovery [89]–[91].

The size of the fraction affects the purity and yield of proteins in the sample. Smaller fractions lead to higher purity as they allow for more precise separation of compounds eluting from the chromatography column, but they may result in lower overall yield since the total amount of protein collected in each fraction is reduced [92]. On the other hand, larger fractions can increase yield by collecting more of the eluate, capturing a greater quantity of the target protein, but may be lower in purity as broader collection windows can include multiple overlapping peaks, leading to co-elution of impurities mixed in with the protein of interest [92]. Furthermore, the capacity of the fraction collector, such as the volume of a single fraction or the number of total fractions, could be a factor in choosing the fraction size. The fraction size in colosseum is 1.5 mL per fraction, for a total of 88 fractions usable in one run, which makes it suitable for salivary proteomics pipelines.

Gradient Elutions

An elution gradient is commonly used to improve separation efficiency. In a gradient elution system, the composition of the mobile phase is gradually changed over time to elute analytes based on their affinity for the stationary phase [93]. For example, in reverse-phase chromatography, the organic solvent concentration is increased to progressively release hydrophobic proteins, whereas in ion-exchange chromatography, a salt gradient disrupts ionic interactions, causing proteins to elute sequentially. The elution gradient significantly affects the resolution and timing of protein elution, with different proteins or peptides eluting at different points in the gradient based on their physicochemical properties. As the gradient progresses, proteins are separated based on their affinity for the stationary phase, allowing for the collection of fractions at various time points that contain distinct sets of proteins. In one salivary experiment, whole saliva samples with 100 μg proteins in each were

separated using an HP1100 LC system with a Vydac C4 reversed-phase column at 250 μL/min, employing a gradient elution from 5% to 85% mobile phase B over 40 minutes, followed by 25 minutes at 85% B, resulting in 35 one-minute LC fractions per sample [90]. The colosseum fraction collector is well-suited for elution gradient workflows. In one of its validation experiments, each consecutive fraction had a linearly increasing flow volume, indicating that the fractionation process effectively captured the progressive elution profile of proteins.

Other Factors for Optimization: Removal of α-Amylase

Salivary α-amylase is the most abundant protein in saliva, and along with albumin and immunoglobulins, comprises up to 75% of the total saliva proteome. These high-abundance proteins cause significant ion suppression for low-abundance biomarkers in mass spectrometric analysis, obscuring potential biomarkers of interest. Studies have shown that starch-based depletion methods can achieve up to a 97% reduction in amylase activity, allowing for enhanced detection of several potential biomarkers including desmoplakin, short palate lung and nasal epithelium carcinoma-associated protein 2, mucin-7, and several immunoglobulin isoforms that would otherwise remain undetected [94].

Amylase depletion from saliva samples can be achieved using a simple starch affinity-based approach. A potato starch solution is prepared by washing with water, centrifuging in between washes. Saliva is then mixed with the washed starch and incubated in a rotating shaker in room temperature, for the amylase to react with the starch. After incubation, the mixture is centrifuged and the supernatant, which now contains saliva with reduced amylase levels, is collected for protein analysis. [91]

Chapter 4

LONGITUDINAL TRANSCRIPTOMICS WITH SALIVA

[1] Y. Kil and L. Pachter, "Differential analysis reveals isoform switching following pneumococcal vaccination," *bioRxiv*, Mar. 10, 2025. DOI: 10.1101/2025.03.09.642237.

4.1 Abstract

Advances in RNA-sequencing (RNA-seq) technology have enabled scalable and accessible transcriptomics studies, including those using noninvasive biosamples like saliva. Longitudinal RNA sequencing studies have been used to track gene expression over time, revealing biological pathways and expression patterns. Saliva's noninvasiveness allows for frequent sampling, opening new possibilities in health monitoring. Traditional approaches for such longitudinal studies rely on pairwise comparisons or linear regression models, but these methods face challenges when dealing with many time points. Spline regression offers a robust alternative by efficiently capturing temporal patterns. In this study, we apply spline regression to analyze longitudinal RNA-seq data and demonstrate its advantages in isoform-level differential expression analysis. Our findings highlight the importance of isoform switching, which may be missed in gene-level analyses.

4.2 Introduction

Advances in RNA-sequencing (RNA-seq) technology have made RNA-seq scalable and accessible [95], allowing researchers to conduct longitudinal transcriptomics studies that measure the expression of genes over time [96]–[103]. Longitudinal -omics studies, which can span multiple types of omics data such as genomics, transcriptomics, and proteomics, have become increasingly important in biomedicine, offering powerful tools for monitoring health and disease over time [104]–[106]. Longitudinal -omics leverages the ability to observe molecular changes dynamically to uncover insights into disease mechanisms, often on a personalized level, enabling more precise and effective healthcare strategies.

Saliva is an increasingly recognized biosample for longitudinal -omics studies due to its noninvasive, easy-to-collect nature, and the wealth of biological information it contains. Unlike other biological fluids, such as blood or urine, saliva can be

collected frequently over extended periods without discomfort, making it particularly suited for longitudinal studies that require repeated sampling. Saliva harbors a wide array of biomolecules, including RNA, DNA, proteins, and small metabolites, which reflect systemic changes in the body and can provide insights into various physiological and pathological processes [107]. Moreover, salivary biomarkers have been implicated in a variety of chronic disease groups such as cancer, autoimmune diseases, and metabolic disorders, thus establishing saliva as a potential medium for health monitoring [8], [9], [108], [109]. The longitudinal analysis of salivaomics, including transcriptomics, proteomics, and metabolomics, offers a unique opportunity to track changes over time, providing valuable data on the progression of diseases or responses to treatments. In this study, bulk RNA-seq of saliva [18] is used to explore longitudinal transcriptomic changes, further demonstrating saliva's potential in noninvasive health monitoring and personalized medicine.

When analyzing RNA-seq data from saliva or other biosamples, an important consideration is the concept of isoforms—different RNA transcript variants produced from the same gene. These variants result from processes like alternative splicing, where different exons can be included or excluded from the final transcript [110]. Isoforms from the same gene can encode mRNA or proteins with different functions, localizations, or regulatory properties [110]. A particularly significant phenomenon is isoform switching, where the relative abundance of these transcript variants changes between conditions or over time [111]. This switching can reveal subtle but biologically important signals that would be completely masked in conventional gene-level analyses where transcript counts are simply summed together [112]. Detecting these isoform-level changes is especially valuable in longitudinal studies of complex biological processes, such as immune responses, where the specific mRNA isoform being expressed can dramatically affect biological outcomes.

The most commonly used approaches for biological discovery from longitudinal RNA-seq studies are specific to the number of time points present in the study. Most published longitudinal transcriptomics studies, such as the aforementioned ones, have only 3-5 time points [96]–[100], [102], [103], making it difficult or impossible to apply traditional time series methods. In these cases where the number of time points is small, conducting pairwise comparisons between time points is the most intuitive and effective way to look for differentially-expressed genes. When there are more time points, pairwise comparisons can still be useful for examining changes in gene expression at specific times, but the high-dimensionality

of the analysis resulting from tens of thousands of genes and many time points can make implementation tedious, and aggregating results from the many pairs can be challenging and suffer from multiple testing difficulties.

One way to identify differentially expressed (DE) genes in time-course RNA sequences with many time points is to model the data with splines. A spline is a smooth function that consists of piecewise polynomials connected at points called "knots"—in simpler terms, splines can be thought of as flexible curves that adapt to follow the shape of data over time, similar to how a draftsman's spline tool bends to create smooth curves through fixed points. Fitting splines to time course data is a straightforward approach to reduce dimensionality of the data, while preserving qualitative features. Splines can be fit to longitudinal data with uniformly spaced, or uneven time points in cases where sampling is not uniform. Moreover, spline regression can be used to analyze longitudinal data without biological replicates, whereas replicates are necessary for classic DE methods such as pairwise comparisons [113].

The shape of the polynomials that make up a spline and the number and placement of knots, define the many types of splines. The degree of the spline refers to the degree of the piecewise polynomials forming the spline: a cubic spline, for instance, is constructed from polynomials of degree 3. To ensure that a spline is continuous, there are two conditions that need to be met: 1) all polynomials of degree d in the spline must be (d-1) times differentiable, and 2) all derivatives up to degree (d-1) must be continuous at each knot. In addition to these conditions, a natural spline is constrained to be linear at the end points. Natural splines are widely used in spline regression because their edge constraints make them mimic natural phenomena more than other splines.

The number of knots in a spline can also be expressed in terms of "degrees of freedom," which is a statistical term denoting the number of independently variable parameters in a system. The number of knots, k, and the degrees of freedom, df, are related to each other, but the relationship depends on the type of the spline. For instance, B-splines, which are splines additionally defined by control points for enhanced shape control, have df = k + d degrees of freedom, where d denotes the degree of the polynomials [114]. Natural splines have df = k + 1 degrees of freedom, due to the constraints at the end points [114]. The number and placement of knots can determine how well the spline fits the data, so it is important to choose an appropriate number of knots for the spline. With more knots the spline fit may be better, but additional parameters increase the risk of overfitting [113]. As for

the locations of the knots, they can be placed uniformly throughout the range of the data with equal numbers of points between each knot, or be optimized to minimize variance [115].

Figure 4.1: Time series data and analysis workflow. Left: Hourly samples with 20 pre- and 20 post-vaccination time points, matched by hour. a) Broad DE analysis between pre- and post-vaccination without time consideration. b) DE analysis within each condition to detect temporal expression changes. c) DE analysis with both time and condition, comparing splines fit to each. Right: Daily samples with 30 post-vaccination time points. Splines are fit to each isoform, and DE isoforms with spline-like patterns are identified.

In this work, we analyze a dataset published in Mias et. al. [18], which aimed to demonstrate the potential of longitudinal saliva multiomics for noninvasive health monitoring by tracking immune responses to a vaccination [18]. Researchers monitored one healthy individual across three distinct time periods: (1) hourly sampling for 24 hours under normal conditions, (2) hourly sampling for 24 hours after administering the pneumococcal polysaccharide vaccine (PPSV23), and (3) daily sampling for 33 days following the vaccination. The original study utilized a pipeline named MathIOmica, which employs spectral methods to analyze time series data based on temporal trends through autocorrelation analysis, where each time series is compared with a delayed version of itself and categorized into classes based on autocorrelations and signal spikes [18]. From the original data, we sub-sample and construct two time series, hourly and daily, from the bulk RNA-seq data generated from this study and show that spline regression coupled applied to accurate quantifications can be used to identify isoforms that may be biologically relevant, but that are missed

in a naïve gene-level analysis.

4.3 Results

Description of the Time Series

The hourly series has 20 time points prior to vaccination and 20 time points after. The pre-vaccination hourly timepoints are matched with the post-vaccination time points for the hour of the day, and time points are evenly distributed. The principal components analysis (PCA) of the hourly samples shows some separation between early and late time points, but not much separation between conditions (Figure 4.2, left). The daily series has 30 time points, all of which are post-vaccination, sampled at the same time every day. All 30 time points in the daily series were after vaccination, so the only covariate for the daily series was time. Therefore, we used sleuth to look for genes that varied in expression over time without changes in condition. The PCA of the daily samples did not show a conclusive pattern on the time-course behavior of the samples (Figure 4.2, right).

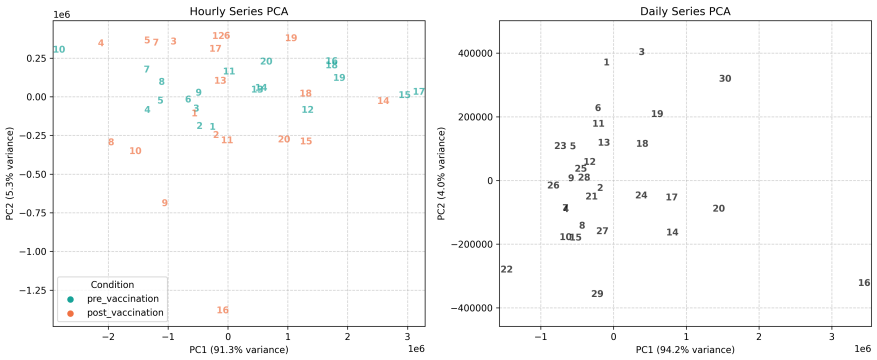


Figure 4.2: [PCA of hourly (left) and daily (right) samples, colored by condition and labeled by time. Left) For the hourly series, there is some separation between the early and late time points but not much separation between pre- and post-vaccination samples. Right) Overall, the samples are more clustered than spread out, and the PCA does not show a conclusive pattern on how samples change over time.

Results of Immune Perturbation

Hourly Series, Pre- and Post-Vaccination

From the hourly series with 20 pre- and 20 post-vaccination samples, a total of 1208 genes were identified as differentially expressed over time and between conditions (Table 4.1, Appendix A). 178 of the 1208 DE genes from the hourly series were also found in the original study. Among the top DE genes are TYROBP, STXBP2,

PCBP2, and ADGRE5, which are genes associated with immune-system-related pathways, such as the Immune System pathway (R-HSA-168256) and Innate Immune System pathway (R-HSA-168249) from the Reactome 2024 database [116]. TMPRSS11E, the top DE gene, is known to be highly expressed in the squamous epithelium and to be related to the innate immune response [117]. TYROBP, another DE gene from the hourly series, is a key immune signaling adaptor that regulates the activation of natural killer cells and myeloid cells, playing a crucial role in innate immunity and neuroinflammation [118]. The identification of these immune-related genes and Reactome pathways in salivary transcriptomics demonstrates the potential of saliva as a diagnostic medium for monitoring immune responses. The ability to detect genes like TYROBP and TMPRSS11E in saliva is particularly significant as it shows that this noninvasive biofluid can effectively capture systemic immune signatures that traditionally might require blood sampling. This reinforces the utility of saliva in biomedicine as a window into systemic physiological processes.

Gene Set Enrichment Analysis (GSEA) for Reactome pathways using all 1208 DE genes showed numerous pathways expected of an immune response to a vaccination, such as "Immune System," "Innate Immune System," and "Neutrophil Degranulation," but there were pathways not specific to immune response, such as "Cellular Responses to Stimuli/Stress" (Table 4.2, Appendix A). In spite of the differences in Reactome databases used for GSEA between the original study and our results, terms from the original study (Reactome 2022) and from sleuth (Reactome 2024) showed significant overlap in immune-system-related terms, suggesting that our spline-based analysis was able to capture the immune response as well.

Daily Series, Post-Vaccination

In the daily series with 30 post-vaccination samples, sleuth identified 241 genes with expression patterns that varied over time more than noise as differentially expressed genes (Table 4.3, Appendix). 123 out of the 241 DE genes were also marked as DE in the original study. The top 10 DE genes for the daily samples are presented in Table 4.3 (Appendix A), ranked by statistical significance, as well as the splines fit on the isoforms of the top DE gene, PLAT. PLAT is known to modulate both innate and adaptive immunity by influencing immune cell activation, migration, and cytokine production [119]. NAMPT has also been found to regulate immune responses by maintaining NAD+ levels in immune cells [120].

GSEA for Reactome pathways with DE genes from the daily series revealed many

significant pathways related to immune responses, such as “Immune System,” “Neutrophil Degranulation,” and “Cytokine Signaling in Immune System” (Table 4.4, Appendix A). However, similar to the hourly series, the top pathways contained Reactome pathways that are not specific to immune response, such as “Metabolism of Proteins.” As with the hourly series, GSEA terms from the original study (Reactome 2022) and from sleuth (Reactome 2024) showed overlap in immune-system-related terms, such as 'Adaptive Immune System', 'Immune System', 'Innate Immune System', 'Interferon Gamma Signaling', 'Interleukin-1 Signaling', and 'Neutrophil Degranulation,' showing that our methods were able to capture pathways similar to the original study that are relevant to the immune response.

Differential Expression Analysis in the Isoform Resolution

Using kallisto and sleuth in tandem enables DE analysis on the isoform resolution. With spline regression on each isoform, we can capture isoform-level temporal dynamics that could be hidden if isoform counts are aggregated into gene-level counts prior to analysis. For instance, if two transcripts have opposing temporal expression patterns, adding them together for a gene-level analysis may cancel out their expression levels. We show that performing differential expression analysis in the isoform level captures genes more specific to the immune response. Furthermore, we show that for many of the DE genes in both the hourly and daily series, isoform-level results show isoform switching, which is defined as a significant change in the relative contribution of isoforms to the parent gene expression [111].

Hourly Series, Pre- and Post-Vaccination

We ran sleuth in transcript mode with p-value aggregation enabled on abundance estimates from spline regression applied to each isoform. The gene-level DE results were compiled by aggregating p-values of the isoforms with the Lancaster method. As sleuth allows for gene-level analysis, we analyzed the hourly series on both isoform-level and gene-level and compared the results. 615 genes were identified as differentially expressed from gene-level analysis, as opposed to the 1208 genes from the isoform-level analysis. The top DE gene for both gene- and isoform-level analyses was TMPRSS11E. In gene-mode, sleuth aggregates all isoforms for a given gene prior to filtering and analysis, which may lead to isoforms with highly contrasting patterns getting added to form a lower contrasting pattern and getting filtered out as a result. GSEA with the 615 DE genes from gene-level analysis identified Reactome pathways not particularly relevant to the immune response,

whereas GSEA with DE genes from isoform-level analysis showed immune-related pathways, as expected from a longitudinal study with an immune perturbation.

When running sleuth directly on isoform-level abundance estimates, we were able to capture genes with isoforms that show varying expression patterns over time. One of the top genes, VASP, showed evidence of isoform switching, as shown in Figure 4.3. The post vaccination splines for all 8 isoforms of VASP are plotted in the top panel, as well as a breakdown of each isoform's expression (bottom left) compared to the aggregate gene-level expression (bottom right). The aggregated gene-level expression over time looks different from most of the splines, suggesting that dynamics of temporal expression are not the same for all isoforms and that summing the isoform counts may hide individual isoform expression patterns. We can see from the isoform fractions that the isoform that dominates the gene's expression switches from ENST00000588273 to ENST00000705987, and later to ENST00000245932. This isoform switching phenomenon in VASP is particularly interesting in this time series with immune perturbation, as VASP is known to play a crucial role in T-cell activation and expansion [121]. The detection of such nuanced expression changes in saliva demonstrates the depth of molecular information that can be captured through salivary transcriptomics when appropriate computational methods are applied. Such granular insights would be entirely missed in conventional gene-level analyses, highlighting the value of our isoform-level approach for salivary biomedicine applications.

Daily Series, Post-Vaccination

The difference between gene-level and isoform-level analysis results were also seen in the daily series. While 241 genes were marked as DE in the isoform-level analysis with p-value aggregation, only 106 genes were marked as DE in the gene-level analysis. The top 10 DE genes from isoform-level analysis, except for RPS27L, were not identified as DE in gene-level analysis. As with the hourly series, GSEA with the 106 genes from gene-level analysis identified Reactome pathways that were not specific to an immune response that is expected of a vaccination, suggesting that many of the genes that were relevant to the immune response were lost in gene-level analysis.

Similar to the hourly series, DE analysis on the isoform level identified genes with varying isoform expression patterns, with some showing signs of isoform switching. Figure 4.4 shows the temporal expression of isoforms for gene PLAT,

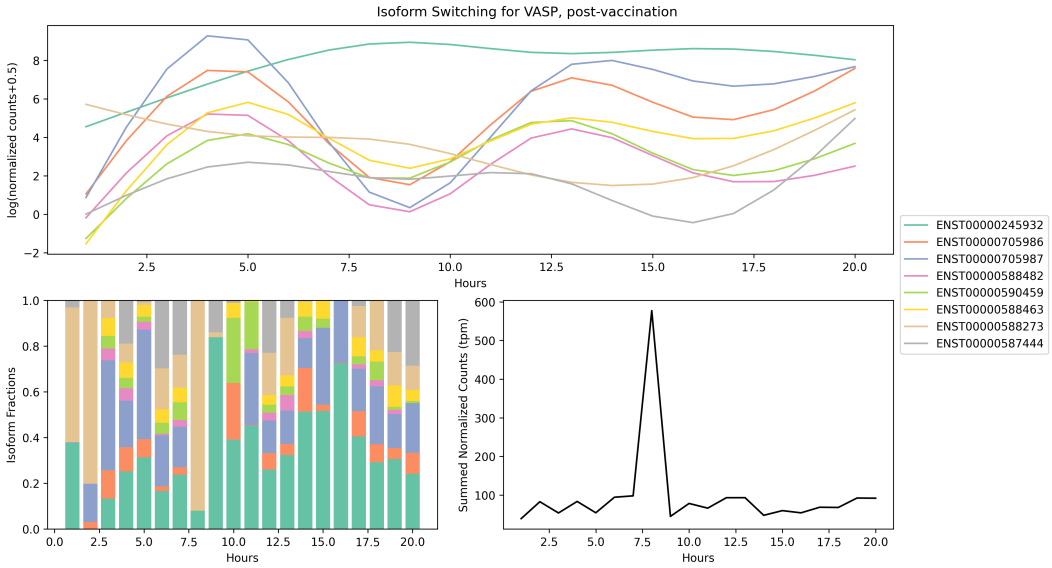


Figure 4.3: Isoform switching seen in gene VASP, in post-vaccination series. Top: Overlayed isoform expression patterns show different temporal trends. Bottom left: Fraction of isoform expression of total gene expression over time, showing change in which isoform dominates the expression of the gene. Bottom right: Total gene expression over time, which can deviate from the constituent isoform expression patterns, implying that isoform-level analysis may highlight temporal dynamics of a gene that might be missed on a gene-level analysis.

the top DE gene from the daily series, as an example. For the five isoforms that passed the filter for low expression and were aggregated into the PLAT gene, there seem to be distinct groups of splines that resemble each other (Fig. 4.4, top panel). This is further corroborated by the isoform fractions (Fig. 4.4, lower left) as it shows ENST00000352041 and ENST00000677722 taking up most of the expression of PLAT in the first few days of the series and then later being replaced by ENST00000429089 and ENST0000067867 later in the series. Interestingly, for the daily series, this switch in isoform fractions can also be observed in other DE genes from the isoform-level results; it does not seem to be universal among all DE genes, especially in those where all isoforms are consistently expressed over time. However, the observed isoform switching in PLAT, a gene involved in modulating both innate and adaptive immunity by encoding tissue-type plasminogen activator (tPA) [119], further illustrates how salivary transcriptomics can reveal dynamic molecular changes during immune responses. This level of detail in longitudinal monitoring would be difficult to achieve with more invasive sampling methods due to the lower frequency of sampling, further supporting saliva's value for precision medicine applications.

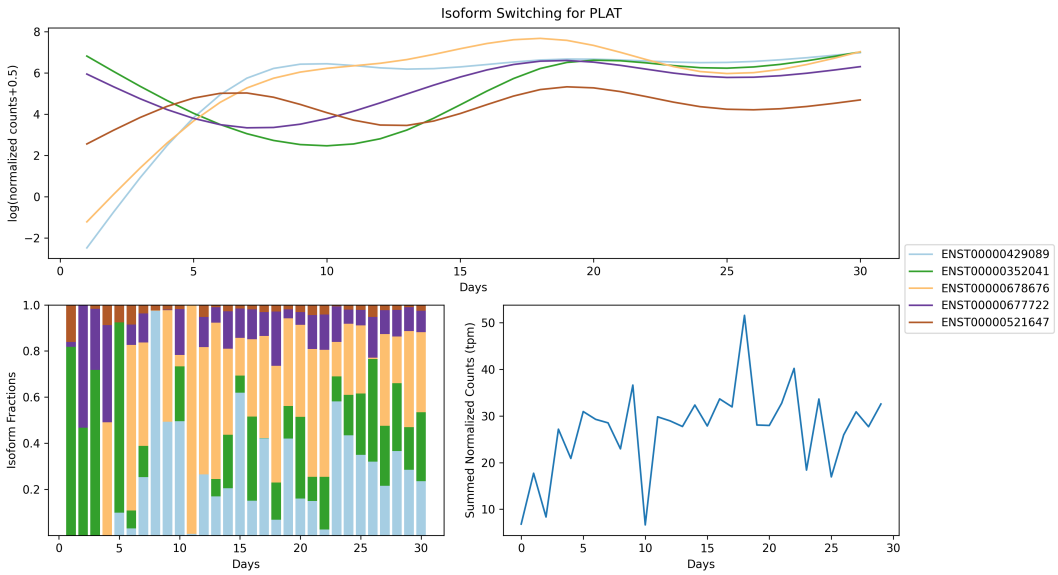


Figure 4.4: Isoform switching seen in gene PLAT from the daily time series. Top: Overlayed isoform expression patterns show different temporal dynamics of isoforms for PLAT. Bottom left: Fraction of isoform expression to total gene expression shows the change in which isoform dominates the expression for the gene. Bottom right: Total expression for gene PLAT, which does reflect the complexity of isoform expression patterns as evident in the hourly series.

4.4 Conclusions and Discussion

We present a method of analyzing longitudinal transcriptomic data that leverages sleuth for time-course DE analysis at the isoform-level using spline regression. While this pipeline can be applied to longitudinal RNA-seq data generated from any sample, applying this to salivary transcriptomics is especially valuable as saliva, with its noninvasive nature and readily production, is uniquely suited for longitudinal health monitoring applications. The computational approach developed in this chapter establishes a foundation for more advanced applications in salivary biomedicine, as explored in Chapter 5. Our isoform-level analysis method enhances the value of salivary transcriptomics in several ways: it enables the detection of nuanced expression patterns that could serve as more precise biomarkers, it supports personalized medicine applications through noninvasive longitudinal monitoring, and it provides a framework that can be integrated with other omics data types from saliva. By addressing the computational challenges in analyzing time-series salivary RNA-seq data, particularly at the isoform level, this work contributes to the broader goal of establishing saliva as a valuable biofluid for precision diagnostics and health monitoring. Chapter 5 builds upon these methodological advances to explore practical applications and future directions for salivary biomedicine.

In this chapter, we also show that spline regression is an effective method of identifying temporally DE genes when compared to other methods of analyzing longitudinal data. Furthermore, we show that sleuth's ability to perform isoform-level DE analysis enables the identification of genes with isoform switching that may be missed in a classic, gene-level DE analysis. The findings from this study demonstrate the pipeline's potential for more precise and comprehensive analysis of time-dependent transcriptomic data in similar longitudinal studies with multiple experimental conditions and many time points. However, the concept of analyzing time-course RNA-seq with splines is not new. The analysis has been introduced as an approach that can be utilized with different DE analysis tools such as edgeR [122] and limma [123] (Appendix C). The novelty of our approach is to apply spline regression on the isoform level and to aggregate them into gene-level conclusions via sleuth. Our results show that the immune perturbations in this study are better captured with isoform-level analysis than with gene-level analysis.

Splines have many advantages over traditional pairwise comparisons or linear regression approaches that are suitable for longitudinal data analysis with many time points. Compared to the longitudinal transcriptomic studies from other publications with 3-5 time points, the study used in this pipeline contains dozens of time points, and comparing timepoints pairwise is a tedious task that does not utilize the time-series structure fully. Spline regression offers a flexible approach for modeling complex, non-linear expression patterns, whereas linear regression may be too restrictive to accurately capture such variability. Moreover, it is possible to use splines to model time-course RNA-seq data with missing time points, as well as time points that are unevenly spaced. Although there are some benefits of using splines to interpolate transcriptomic time-series data, there is a need for caution, especially when estimating parameters. When using splines to model time-course transcriptomic data, it is important to choose an appropriate number of knots, or degrees of freedom. In the user manual for limma and in Svensson's walkthrough for sleuth on modeling time-course experiments with splines, the recommended degrees of freedom is given as 3 to 5 [123], [124]. Though the reasons for this seemingly arbitrary range are not presented anywhere, this range of degrees of freedom is also recommended and seems to work well for modeling biological phenomena other than transcriptomics, such as serum glucose concentration over 30 days [113]. The degrees of freedom chosen for the current pipeline was 5, to account for the fact that there are 20 and 30 data points in the two time series of interest.

The fit of the splines was evaluated using R^2, with average values of 0.366 for pre-vaccination hourly samples, 0.489 for post-vaccination hourly samples, and 0.476 for daily samples. While these R^2 values are not particularly high, several factors should be considered. Increasing the degrees of freedom generally improves the R^2 value; however, AICc indicates that a model with five degrees of freedom is preferable to one with seven. Additionally, many time points have zero counts, even with a stricter filter than the default in sleuth, which likely contributes to the lower fit. Alternative spline approaches, such as manually placing knots, may enhance the fit, although our attempts with B-splines resulted in poor-quality genes (primarily mitochondrial), suggesting that natural splines, which are constrained to better model biological patterns, would be more suitable for this type of analysis.

4.5 Methods

Here we outline our computational pipeline for analyzing longitudinal salivary RNA-seq data, which integrates several specialized tools to extract biologically meaningful patterns from the complex time-series data.

Subsampling Time Points

We perform spline regression on this data using sleuth [125], a program to analyze RNA-Seq experiments for which transcript abundances have been quantified with kallisto [126]. Natural cubic splines are fit to the hourly series, with 20 pre-vaccination and 20 post-vaccination samples, and daily series, with 30 post-vaccination time points. Regarding sample selection for each time series, the original study selected 24 hourly samples, whereas the current analysis used only 20. The first four samples were omitted in the current pipeline because the vaccine was administered between hours 4 and 5 of the 24-hour post-vaccination series. This omission ensures that the remaining 20 samples and their matched pre-vaccination counterparts more accurately represent the change in condition.

Pseudo-Alignment with kallisto

All samples in the time series data were pseudo-aligned to Ensembl v. 112 annotation for Homo Sapiens (GRCh38.p14) with kallisto. The k-mer length was set to the default value of 31. The -b argument indicates that the number of bootstraps was set to 100.

Differential Expression Analysis with sleuth

All analyses were undertaken with R version 4.4.2 [127] and sleuth version 0.30.1. The following R packages were loaded into RStudio: sleuth, splines, and biomaRt. The sample-to-covariates matrix, which is a required argument for sleuth, was compiled outside R for each of the time series. A transcript-to-gene mapping was created using biomaRt's useEnsembl() function, in order to enable p-value aggregation on the results [112]. The annotation used for the transcript-to-gene mapping was the same annotation used for pseudo-alignment in kallisto. When importing kallisto results into sleuth, the default filter for sleuth (minimum of 5 reads observed in 47% of reads) was modified to be stricter (minimum of 5 reads observed in 67% of reads). The two main R scripts for analyzing the hourly and daily series are provided in Appendix B.

The objective of our study was to find genes that are highly variable in expression over time, either within one experimental condition (daily) or between two different experimental conditions (hourly) such as treatment and control, in which time and condition can interact and introduce complexities in the data. The models for the two datasets were set up as follows. For the hourly series with pre- and post-vaccination, the expression for the full model is given by:

$$Expr \sim \beta_0 + \sum_{c \in \{pre, post\}} \left(\beta_c \cdot c + \sum_{k=1}^{d.f.} \gamma_{c,k} \cdot c \cdot S_k(t) \right) + \epsilon.$$

For the reduced model, the expression is given by:

$$Expr \sim \beta_0 + \sum_{c \in \{pre, post\}} \beta_c \cdot c + \epsilon.$$

The expression weight, β_0 represents the intercept term, β_c represents coefficients for the main effects of each condition (pre or post), c indicates condition (pre or post), $\gamma_{c,k}$ represent coefficients for the interaction of condition c (pre or post) with the k-th spline basis function, $S_k(t)$ represents spline basis functions from $k = 1$ to n degrees of freedom, which is equal to 5 in this case, and lastly, ϵ is a noise term.

In the syntax of the R language, the full model for the hourly series is given by `(~0 + group + group:X)` and the reduced model `(~0 + group)`, where `group` is the condition covariate releveled by the pre-vaccination levels and `X` is a natural cubic spline with 5 degrees of freedom. The `(group:X)` denotes that a spline was fit to each condition, pre- and post-vaccination. The reduced model includes `group`,

such that the difference between the full and reduced models represents the interaction terms, which is given as `(group:X)`. By comparing the full and reduced model in this manner, sleuth identifies isoforms that have different expression patterns for each condition. In contrast, the original study treated pre-vaccination samples as baseline levels and subtracted them from post-vaccination samples to create a 'delta hourly' time series for the main analysis [18]. For the daily series where all time points are post-vaccination, the full model for the daily series includes time, modeled as a natural cubic spline with 5 degrees of freedom and expressed as `(~X)` in R, and the reduced model includes noise, expressed as `(~1)` in R. The expressions for the full and reduced model are given by:

$$Expr \sim \beta_0 + \sum_{k=1}^{d.f.} \gamma_k \cdot S_k(t) + \epsilon,$$

$$Expr \sim \beta_0 + \epsilon.$$

Similar to the hourly series, 0 represents the intercept term, γ_k represent coefficients for the k-th spline basis function, $S_k(t)$ represents spline basis functions from $k = 1$ to n degrees of freedom, which is equal to 5 in this case, and ϵ represents residual error.

Sleuth was run on transcript (isoform) mode such that splines are fit to each isoform, after which the results are compiled into gene-level differential expression results through p-value aggregation with the Lancaster method [112]. There are two testing methods in sleuth: the likelihood ratio test and the Wald test. For the likelihood ratio test, which is used in the current study, sleuth compares two models: a full model with all parameters of the experiment, and a reduced model with the parameters to control for. For each type of analysis, full and reduced models were fit to the data, with specific details about the model used for each analysis provided in the Results section. To implement splines in R, the ns() package in the splines library was used [127]. The DE analysis results were filtered for q-values < 0.05, and were exported to a Python notebook for plotting. All R and Python notebooks as well as the full list of DE genes and related pathways may be found on a designated GitHub repository: `https://github.com/pachterlab/KP_2025`.

Gene Set Enrichment Analysis (GSEA)

The DE analysis results from sleuth were exported as CSV files, and cleaned up in a Python notebook. The list of DE genes were analyzed on Enrichr [128] for Reactome

2024 pathways. To match the original study where the data was published, GSEA results were filtered for adjusted p-values < 0.003.

Visualization

To visualize the spline regression results, normalized transcript counts and full models containing spline coefficients were exported as CSV files from sleuth into a Python notebook. Splines were visualized for each isoform of a DE gene, as sleuth was operated on transcript mode and the spline coefficients were fit to each isoform instead of the gene. Transcript TPMs were aggregated to plot gene-level expression. All Python notebooks and CSV files may be found on a designated GitHub repository: `https://github.com/pachterlab/KP_2025`.

Data Availability

The data used for this study is publicly available at NCBI Gene Expression Omnibus database (accession GSE108664) and has an accompanying publication [18].

4.6 Appendix A: Differential Gene Expression Analysis and Gene Set Enrichment Analysis Results

Presented in the following tables are differential gene expression analysis results and gene set enrichment analysis results. The full list of DE genes and GSEA results cn abe found on the GitHub repository: `https://github.com/pachterlab/KP_2025`.

Gene ID	Gene Name	# Aggregated Transcripts	P-value	Q-value
ENSG00000274058	TMPRSS11E	1	1.014E-08	8.215E-05
ENSG00000011600	TYROBP	6	2.409E-08	9.754E-05
ENSG00000197111	PCBP2	8	3.672E-08	9.912E-05
ENSG00000076944	STXBP2	10	1.198E-07	2.426E-04
ENSG00000117984	CTSD	3	1.861E-07	2.502E-04
ENSG00000162244	RPL29	1	2.065E-07	2.502E-04
ENSG00000125753	VASP	8	2.163E-07	2.502E-04
ENSG00000134333	LDHA	4	2.784E-07	2.505E-04
ENSG00000170296	GABARAP	4	4.086E-07	3.008E-04
ENSG00000196352	CD55	8	3.991E-07	3.008E-04

Table 4.1: Top 10 Differentially expressed genes from the hourly series. A total of 1208 genes were identified as differentially expressed over time and between conditions. Among the top DE genes are TYROBP, STXBP2, PCBP2, and ADGRE5, which are genes associated with immune-system-related pathways, such as "Immune System" and "Innate Immune System" from the Reactome 2024 database.

Term	# Genes	P-value	Adj-P-value
Immune System	247	2.56E-28	2.15E-25
Cellular Responses to Stimuli	139	2.60E-28	2.15E-25
Cellular Responses to Stress	125	5.09E-26	2.81E-23
SRP-dependent Cotranslational Protein Targeting to Membrane	44	1.22E-24	5.06E-22
Innate Immune System	149	9.82E-22	3.25E-19
Nonsense Mediated Decay (NMD) Enhanced by the Exon Junction Complex (EJC)	41	1.54E-21	3.63E-19
Nonsense-Mediated Decay (NMD)	41	1.54E-21	3.63E-19
Peptide Chain Elongation	36	5.26E-21	1.09E-18
Response of EIF2AK4 (GCN2) to Amino Acid Deficiency	38	6.60E-21	1.21E-18
Neutrophil Degranulation	84	1.50E-20	2.48E-18

Table 4.2: Top 10 Reactome Pathways identified from DE genes in the hourly series. GSEA with all 1208 DE genes from the hourly series shows some immune-response-related pathways such as “Immune System,” “Innate Immune System,” and “Neutrophil Degranulation” but also many non-immune-specific pathways.

Gene ID	Gene Name	# Aggregated Transcripts	P-value	Q-value
ENSG00000104368	PLAT	5	5.10E-09	3.30E-05
ENSG00000105835	NAMPT	22	5.10E-08	1.63E-04
ENSG00000130066	SAT1	7	3.89E-07	6.22E-04
ENSG00000176788	BASP1	2	3.75E-07	6.22E-04
ENSG00000185201	IFITM2	5	1.34E-06	1.71E-03
ENSG00000167996	FTH1	9	2.55E-06	2.71E-03
ENSG00000136167	LCP1	2	4.10E-06	3.15E-03
ENSG00000163902	RPN1	1	3.58E-06	3.15E-03
ENSG00000185088	RPS27L	2	4.63E-06	3.15E-03
ENSG00000119535	CSF3R	9	4.93E-06	3.15E-03

Table 4.3: Top 10 Differentially expressed genes from the daily series. A total of 241 genes were identified as differentially expressed in the daily series, for which all samples were post-vaccination. Many of the top DE genes are relevant to the immune system, including PLAT and NAMPT, which are known to modulate immune cells.

Term	# Genes	P-value	Adj-P-value
Immune System	62	2.17E-11	2.00E-08
Neutrophil Degranulation	22	7.00E-08	3.23E-05
Cytokine Signaling in Immune System	27	6.32E-07	1.95E-04
Innate Immune System	34	8.63E-07	1.99E-04
Metabolism of Proteins	47	9.88E-06	1.83E-03
Signaling by Interleukins	17	3.16E-05	4.87E-03
Class I MHC Mediated Antigen Processing & Presentation	15	4.37E-05	5.77E-03
Signaling by Rho GTPases	21	5.74E-05	6.64E-03
Signaling by Rho GTPases, Miro GTPases and RHOBTB3	21	8.01E-05	7.50E-03
Apoptotic Cleavage of Cellular Proteins	5	8.34E-05	7.50E-03

Table 4.4: Top 10 Reactome Pathways identified from DE genes in the daily series. GSEA with all 241 genes from the daily series identified many Reactome pathways related to the immune response, with the top 10 pathways including terms such as “Immune System,” “Neutrophil Degranulation,” and “Cytokine Signaling in Immune System.”

4.7 Appendix B: Code Implementation for Longitudinal RNA-seq Analysis in sleuth

This appendix provides the complete R code implementation used to perform the differential expression analysis described in Chapter 4. The code demonstrates the practical application of sleuth for analyzing longitudinal RNA-seq data at the isoform level, including data import, normalization, time-course modeling using natural splines, and differential expression testing. This implementation served as the computational foundation for all results presented in the main chapter and is included here to ensure reproducibility and to serve as a reference for researchers seeking to apply similar methodologies to their own longitudinal transcriptomic datasets. While the two main R scripts are included in this Appendix, all R and Python notebooks may be found on a designated GitHub repository: `https://github.com/pachterlab/KP_2025`.

Below is the script used to analyze the hourly series, which contains 20 pre-vaccination and 20 post-vaccination time points.

```r
# Import dependencies
library(sleuth, splines)
library(biomaRt)

# Set up annotations for genes and transcripts
ensembl <- biomaRt::useEnsembl(biomart = "genes",
                               dataset = "hsapiens_gene_ensembl",
                               mirror='www')
t2g <- biomaRt::getBM(attributes = c("ensembl_transcript_id",
                                     "ensembl_gene_id",
                                     "external_gene_name"),
                      mart = ensembl)
t2g <- dplyr::rename(t2g, target_id = ensembl_transcript_id,
                     ens_gene = ensembl_gene_id,
                     ext_gene = external_gene_name)

# Set working directory and import metadata
setwd("~/longsaliva")
s2c <- read.delim("s2c_hourly.txt", sep=" ", header=TRUE)
time <- rep(seq(from=1, to=length(s2c$sample)/2, by=1), times=2)
s2c <- dplyr::mutate(s2c, time=time)
sample <- paste0(rep(c('pre', 'post'), each=20),'_',s2c$time)
s2c$sample <- sample

colnames(s2c) <- c("path", "sample", "condition", "time")
```

```r

# Setting the pre-vaccination points as reference
group <- relevel(factor(s2c$condition), ref='pre_vaccination')

# Changing filter on low-count reads to be stricter than default
new_filter <- function(row, min_reads = 5, min_prop = 0.67) {
  mean(row >= min_reads) >= min_prop
}

# Create sleuth object for analysis
so <- sleuth_prep(s2c, target_mapping = t2g,
                  aggregation_column = "ens_gene",
                  extra_bootstrap_summary = TRUE,
                  filter_fun = new_filter)

# Fit "full" and "reduced" model, compare with LRT
X <- splines::ns(s2c$time, df=5)
full_design <- model.matrix(formula(~0 + group + group:X))
colnames(full_design)
so <- sleuth_fit(so, full_design, "full")
so <- sleuth_fit(so, ~0+group, "reduced")
so <- sleuth_lrt(so, "reduced", "full")

# View DE analysis results
sleuth_table <- sleuth_results(so, 'reduced:full', 'lrt', show_all
    = FALSE, pval_aggregate = TRUE)
sleuth_de <- dplyr::filter(sleuth_table, qval <= 0.05)
head(sleuth_de)

# Optionally output results as CSV
# write.csv(sleuth_de, 'hourly_results.csv')
```

Below is the script used to analyze the daily series, which contains 30 post-vaccination time points, which results in a slightly different analysis process to that used for the hourly series.

```r
# Import dependencies
library(sleuth, splines)
library(biomaRt)

# Set up annotations for genes and transcripts
ensembl <- biomaRt::useEnsembl(biomart = "genes",
                               dataset = "hsapiens_gene_ensembl",
                               mirror = "useast")
```

```
t2g <- biomaRt::getBM(attributes = c("ensembl_transcript_id",
                                     "ensembl_gene_id",
                                     "external_gene_name"),
                      mart = ensembl)
t2g <- dplyr::rename(t2g, target_id = ensembl_transcript_id,
                     ens_gene = ensembl_gene_id,
                     ext_gene = external_gene_name)

# Set working directory and import metadata
setwd("~/longsaliva")
s2c <- read.delim("s2c_daily.txt", sep=" ", header=TRUE)
time <- seq(from=1, to=length(s2c$sample), by=1)
s2c <- dplyr::mutate(s2c, time=time)
colnames(s2c) <- c("path", "sample", "time")

# Changing filter on low-count reads to be stricter than default
new_filter <- function(row, min_reads = 5, min_prop = 0.67) {
  mean(row >= min_reads) >= min_prop
}

# Create sleuth object for analysis
so <- sleuth_prep(s2c, target_mapping = t2g,
                  aggregation_column = "ens_gene",
                  extra_bootstrap_summary = TRUE,
                  filter_fun=new_filter)

# Fit "full" and "reduced" model, compare with LRT
X <- splines::ns(s2c$time, df=5)
full_design <- model.matrix(formula(~ X))
so <- sleuth_fit(so, formula = full_design, fit_name = "full")
so <- sleuth_fit(so, formula = ~ 1, fit_name = "reduced")
so <- sleuth_lrt(so, "reduced", "full")

# View DE analysis results
sleuth_table <- sleuth_results(so, 'reduced:full', 'lrt', show_all
    = FALSE, pval_aggregate = TRUE)
sleuth_de <- dplyr::filter(sleuth_table, qval <= 0.05)

# Optionally output results as CSV
# write.csv(sleuth_de, 'daily_results.csv')
```

4.8 Appendix C: Comparison of Pipeline Performance for Longitudinal RNA-seq Data

The development of computational pipelines for RNA-seq analysis has significantly advanced our understanding of transcriptomic dynamics in various biological contexts. The pipeline introduced in this chapter using sleuth[125] is designed to perform differential expression analysis on RNA-seq data on the isoform level[112], which can uncover isoform-level signals that may be hidden in a traditional gene-level analysis. We applied sleuth on longitudinal, time-course RNA-seq data on saliva to see whether modeling data with splines could uncover temporal dynamics in the isoform-level, and through the results in Chapter 4, we show that sleuth indeed was able to detect temporal signals in gene expression.

This appendix presents a focused comparison between the pipeline described in Chapter 4, which utilizes sleuth, and a leading alternative approach combining edgeR for import, organization, filtering, and normalization with limma[123] for linear modeling and empirical Bayes moderation in differential expression analysis. The edgeR[122]-limma pipeline represents the most competitive alternative to our sleuth-based methodology for several key reasons. First, edgeR provides a direct method for importing kallisto pseudo-alignment outputs through tximport, allowing us to use identical quantification data as input to both pipelines. This enables a more objective comparison by eliminating variability that might arise from different alignment strategies. Second, limma explicitly supports spline regression—the same statistical approach employed in sleuth—making it theoretically capable of modeling similar expression patterns across time. The comparison is focused on the hourly time series, which has 20 pre-vaccination and 20 post-vaccination time points, to highlight the viability of using these computational pipelines on time series data that have matched time points in different conditions.

Despite these methodological similarities, our preliminary analyses revealed substantial differences in the ability of these pipelines to detect biologically meaningful signals in longitudinal RNA-seq data. While both approaches utilize comparable statistical frameworks, our sleuth-based pipeline demonstrated enhanced sensitivity in identifying time-dependent expression patterns with biological relevance. By benchmarking against the edgeR-limma approach—which represents current best practices in the field—we provide evidence for the specific advantages of our sleuth-based pipeline in the context of longitudinal RNA-seq analysis. The results presented in this appendix serve to validate the technical advancements of our approach and

provide practical guidance for researchers analyzing longitudinal transcriptomic data who must choose between these competing methodological frameworks.

```
# Import dependencies
library(tximport, tidyr)
library(limma)
library(splines)

# Set working directory and import metadata
setwd("~../../longsaliva")
s2c <- read.delim("s2c_hourly.txt", sep=" ",
                  header=TRUE)
df <- dir(file.path("."))
time <- rep(seq(from=1, to=length(s2c$sample)/2, by=1),
                times=2)
s2c <- dplyr::mutate(s2c, time=time)
colnames(s2c) <- c("path", "sample", "condition",
                   "time")

# Set up annotations for genes and transcripts
ensembl <- biomaRt::useEnsembl(biomart = "genes",
                       dataset = "hsapiens_gene_ensembl",
                       mirror = "useast")
t2g <- biomaRt::getBM(attributes = c("ensembl_transcript_id",
                                     "ensembl_gene_id"),
                      mart = ensembl)
t2g <- dplyr::rename(t2g, TXNAME = ensembl_transcript_id,
                     GENEID = ensembl_gene_id)
files <- file.path(s2c$path, "abundance.h5")
names(files) <- s2c$sample
all(file.exists(files))

# Import kallisto outputs
txi <- tximport(files, type = "kallisto",
                tx2gene = t2g,
                ignoreTxVersion = TRUE,
                countsFromAbundance = "lengthScaledTPM")
names(txi)

# Filter low-expression counts with edgeR
y <- edgeR::DGEList(txi$counts)
keep <- edgeR::filterByExpr(y)
y <- y[keep, ]

```

```
# Modeling time as natural spline with 5 d.f. as with sleuth
X <- ns(s2c$time, df=5)

#Then fit separate curves for the control and treatment groups:
Group <- factor(s2c$condition)
Group <- factor(s2c$condition,
                levels = c("pre_vaccination", "post_vaccination"))
design <- model.matrix(~0 + Group*X, data=y)
colnames(design)

# Normalize and run voom transformation on limma
y <- edgeR::calcNormFactors(y)
v <- voom(y, design)
fit <- lmFit(v, design)
fit <- eBayes(fit)

# View DE results
tt <- topTable(fit, coef=8:12)
```

The edgeR-limma pipeline above yielded only 10 DE genes, which was not a sufficient number of genes to yield any significant gene set enrichment analysis (GSEA) results. The genes marked as differentially expressed over time are shown below:

Gene ID	Gene Name	Avg. Expr.	F-statistic	P.Value	Adj.P.Val
ENSG00000274058	None	4.783	11.286	6.20E-07	4.43E-03
ENSG00000135390	ATP5MC2	3.067	10.923	8.98E-07	4.43E-03
ENSG00000101146	RAE1	1.053	9.718	3.23E-06	1.06E-02
ENSG00000006534	ALDH3B1	2.288	9.375	4.70E-06	1.16E-02
ENSG00000149573	MPZL2	2.284	9.027	6.96E-06	1.21E-02
ENSG00000168421	RHOH	5.018	8.879	8.22E-06	1.21E-02
ENSG00000083844	ZNF264	2.680	8.845	8.55E-06	1.21E-02
ENSG00000141665	FBXO15	2.431	8.566	1.18E-05	1.36E-02
ENSG00000189266	PNRC2	3.531	8.520	1.24E-05	1.36E-02
ENSG00000151366	NDUFC2	2.267	8.263	1.68E-05	1.49E-02

Table 4.5: Genes identified as differentially-expressed over time by the edgeR-limma pipeline on the hourly time series. Only 10 genes were identified, as opposed to the 1208 identified by sleuth on the same data. Gene Set Enrichment Analysis was not performed due to the insufficient number of genes.

The comparative analysis between our sleuth-based pipeline and the edgeR-limma

approach revealed striking differences in sensitivity for detecting differential expression in longitudinal RNA-seq data. The edgeR-limma pipeline yielded only 10 differentially expressed genes, despite using identical kallisto pseudoalignment outputs and employing comparable statistical frameworks for modeling temporal patterns. Our sleuth-based methodology identified 1,208 differentially expressed genes from the same dataset, demonstrating significantly higher detection power.

The substantial disparity in detection sensitivity had significant downstream implications for biological interpretation. The limited set of 10 genes identified by edgeR-limma proved insufficient to yield any statistically significant results in gene set enrichment analysis (GSEA), effectively preventing meaningful pathway-level insights. In contrast, the much larger gene set identified by sleuth enabled robust enrichment analysis, revealing biologically coherent temporal patterns as detailed in Chapter 4.

These results underscore the critical importance of pipeline selection in longitudinal RNA-seq studies, particularly for isoform-level analyses. While both methods theoretically support spline-based modeling of time-course data, sleuth demonstrated dramatically enhanced capacity to detect biologically relevant expression changes. This performance difference may be attributed to sleuth's specialized design for transcript-level analysis and its sophisticated handling of technical variance in RNA-seq data.

For researchers conducting longitudinal transcriptomic studies, these findings suggest that sleuth offers substantial advantages over the edgeR-limma approach, particularly when temporal patterns and isoform-specific dynamics are of interest. The practical implications extend beyond mere statistical power to the fundamental ability to extract meaningful biological narratives from complex time-series RNA-seq data.

Chapter 5

FUTURE DIRECTIONS IN SALIVARY BIOMEDICINE

5.1 Introduction

The work presented in this dissertation has contributed to advancing salivary biomedicine by addressing key challenges in saliva collection, fractionation, and longitudinal analysis. From optimizing collection methods to developing open-source tools for fluid handling, each chapter has laid the foundation for improving the accessibility and reliability of salivary biomedicine by leveraging techniques and methods from biomedical engineering. The comparative analysis of saliva collection devices and the development of SalivaStraw offer insights into enhancing sample quality and usability across diverse populations. The colosseum fraction collector presents a versatile platform for saliva fractionation, expanding opportunities for biomarker discovery and downstream analyses such as proteomics and exosome isolation. Finally, the exploration of longitudinal transcriptomics with saliva underscores the potential of high-frequency sampling to capture dynamic biological changes.

While these contributions advance the infrastructure for salivary research, this work does not encompass large-scale studies linking these tools directly to biomarker discovery or clinical outcomes. Bridging this gap requires comprehensive validation of collection methods across diverse populations through extensive clinical trials. Such endeavors would demand significant time and resources, as well as coordinated efforts across multiple research sites to capture diverse biological and technical variables. The following sections propose future projects aimed at addressing these gaps, leveraging the tools and methodologies developed in this dissertation to pave the way for more standardized and scalable approaches in salivary diagnostics.

5.2 Advancing Saliva Collection Technologies

Validation of SalivaStraw

A primary limitation of the current SalivaStraw development is the lack of formal user testing. While theoretical fluid dynamics principles and internal testing suggest promising performance characteristics, a systematic validation study is necessary to establish SalivaStraw's effectiveness in real-world settings. Future work should include a comparative user study similar to that conducted for existing de-

vices in Section 2.3, using metrics such as collection efficiency, leakage rates, and user-rated usability scores. The study design should incorporate diverse participant demographics, including pediatric subjects, elderly individuals, and those with limited dexterity, to ensure the device's universality. Additionally, analytical validation should assess whether SalivaStraw affects salivary biomarker concentrations compared to established collection methods. This would involve collecting paired samples using SalivaStraw and reference devices, followed by analysis of key biomarkers including proteins, metabolites, and nucleic acids. Such a comprehensive validation approach would address both the usability and analytical reliability aspects of SalivaStraw, providing a solid foundation for its broader implementation in research and clinical settings.

Stimulated Saliva Collection with SalivaStraw

A key area for future development is the exploration of stimulation methods to enhance saliva yield and consistency for diagnostic applications. Our comparative study on saliva collection devices included an investigation into the effects of citric acid as a gustatory stimulus for salivation [129]. Findings demonstrated that citric acid stimulation increased saliva production more than three-fold compared to unstimulated conditions [84]. While further research is necessary to assess how citric acid stimulation affects specific biomarkers [130], this approach could particularly benefit elderly patients, individuals with xerostomia, or others with low resting saliva flow [131]. SalivaStraw, our proof-of-concept saliva collection device, is well-suited to investigate this method further, as citric acid could be applied directly to the mouthpiece without requiring additional equipment. Additionally, since citric acid stimulation reduces saliva viscosity [130], a device that minimizes leakage and overcollection, such as SalivaStraw, would be essential for ensuring a sterile and efficient saliva collection process.

Devices for Non- or Less-Cooperative Subjects

Usability is a critical factor in saliva collection, particularly for individuals who may have difficulty complying with collection protocols. While most saliva collection devices are accessible for healthy adults, they may not be as user-friendly for younger or elderly populations or individuals with disabilities. As identified in our comparative analysis of existing saliva collection devices [129], no single device is universally suitable across all demographics.

During later iterations of SalivaStraw's design, we developed a version with a bent

saliva inlet, paired with a squeezable collection tube, to facilitate saliva collection in non-upright subjects. This design was particularly motivated by the need for single-handed saliva collection in pediatric subjects, allowing caregivers to hold the child while simultaneously collecting saliva. The prototype included an air outlet with a narrow, blunt syringe needle inserted into the collection tube lid and a saliva inlet made from PTFE tubing to enable flexibility. Collection is initiated by inserting the bent straw into the mouth, compressing the low-density polyethylene (LDPE) collection tube, and blocking the air vent. Once saliva enters the straw, the air vent is uncovered, allowing continuous flow until the collected saliva reaches the vent, at which point suction ceases—following the same fluidic principles as the standard SalivaStraw design. While initial informal testing confirmed the feasibility of this approach, further iterations and formal testing are necessary to refine the design and improve its applicability across diverse populations.

Another potential advancement in saliva collection accessibility involves integrating a candy-like segment onto the mouthpiece to encourage compliance, particularly in pediatric and non-cooperative subjects. A similar concept has been implemented in an oral bacterial sampling device [132], but future work could tailor this approach for general saliva collection. Ideally, this concept could be incorporated into SalivaStraw by adding a flavored coating to the mouthpiece. However, further investigation is needed to determine the impact of this modification on the fluid dynamics of SalivaStraw—as the user may aspirate instead of blowing saliva—and the potential effects of the coating on saliva sample integrity.

5.3 Expanding Applications of Saliva-Based Analysis

Exosome Extraction with colosseum

Having established tools for efficient collection and fractionation, we can now explore broader applications of these technologies. A particularly promising direction is exosome extraction using the colosseum, which is a generic fraction collector that can be applied to saliva fractionation. Exosomes, a type of extracellular vesicles that carry RNA and proteins, are abundant in saliva and are valuable for biomarker discovery and disease monitoring [133]. Future work could integrate the colosseum fraction collector with size exclusion chromatography (SEC) [134] to isolate salivary exosomes from saliva. This would require optimizing specific parameters: flow rates, filtration sizes, and buffer compositions. The ability to extract exosomes efficiently would enable downstream proteomic and RNA analyses, expanding the utility of salivary diagnostics in personalized medicine and noninvasive disease

monitoring.

Applications of Longitudinal Health Monitoring with Saliva

The longitudinal monitoring capabilities of salivary diagnostics have significant implications for clinical applications, particularly in cancer surveillance and infectious disease management. Saliva is increasingly recognized as a valuable medium for "liquid biopsy"—a noninvasive alternative to traditional tissue sampling that can provide real-time insights into disease progression [135]. In oral squamous cell carcinoma (OSCC) and gastric cancer, several salivary mRNA biomarkers have demonstrated diagnostic potential [136], [137], with longitudinal monitoring of these markers potentially enabling early detection of recurrence without requiring invasive procedures. The COVID-19 pandemic further highlighted saliva's utility in infectious disease monitoring, where salivary mRNA biomarkers could be tracked over time to understand disease severity and progression patterns [138], [139]. The noninvasive nature of saliva collection makes it particularly suited for these applications, as it enables high-frequency sampling with minimal patient discomfort, allowing clinicians to capture disease dynamics with unprecedented temporal resolution. This approach to frequent, minimally invasive monitoring of transcriptomic biomarkers represents a significant advancement in disease surveillance that could transform patient care across multiple conditions.

Beyond its applications in precision and personalized medicine with transcriptomic biomarkers, saliva holds significant potential for advancing female and mental health research, specifically. Saliva presents a particularly compelling sample for female health due to the hormonal complexity and fluctuations inherent to female physiology. Across various life stages in female physiology such as menstrual cycles, pregnancy, menopause, and hormonal therapies, key hormones like estrogen, progesterone, and luteinizing hormone undergo dynamic changes that influence reproductive health, mood, metabolism, and overall well-being [140]–[144]. Tracking these fluctuations can also offer crucial insights into conditions like polycystic ovary syndrome (PCOS), menstrual irregularities, or diseases that disproportionately affect more women than men, while enabling personalized approaches to treatment and management of these conditions [145]–[148]. Saliva's noninvasive nature enables stress-free, at-home sampling, making it uniquely suited for longitudinal studies that capture these subtle, time-sensitive shifts. Furthermore, the ability to measure free, bioavailable hormone levels in saliva offers direct insights into reproductive health, fertility, and broader hormonal patterns [149]–[154]. Leveraging saliva for female

health monitoring not only addresses these biological needs but also contributes to closing the research gap created by the historical underrepresentation of women in biomedical studies [155], [156], paving the way for more inclusive and personalized healthcare.

Saliva also holds great promise for mental health research and monitoring, particularly because its collection method inherently reduces stress compared to other biosamples like blood or urine [157]. There is already a substantial body of research exploring salivary biomarkers of mental health and stress such as cortisol or alpha-amylase, reinforcing its utility in this field [158]–[162]. Mental health studies often aim to measure stress-related biomarkers, such as cortisol, where the collection process itself can influence results: traditional sampling methods, like venipuncture, may induce anxiety or discomfort, potentially skewing these measurements [163]. In contrast, saliva collection is quick, painless, and can be done in familiar environments, making it ideal for capturing more accurate baseline stress levels, while reflecting accurate levels of biomarkers in the body [164]. This low-stress sampling method not only improves data quality but also makes mental health monitoring more accessible, encouraging frequent sampling and long-term tracking without adding burden to the participant.

Beyond its advantages for specific applications like stress monitoring or female health monitoring, saliva offers significant practical benefits for at-home, routine monitoring of specific biomarkers. This is because saliva collection does not require supervision from a healthcare professional, unlike blood sampling. There have been many studies developing point-of-care salivary hormone biosensors, mostly centered around cortisol, using various mechanisms, such as electrochemical sensors [165], [166], and lateral flow assays [167]–[171], which could allow users to obtain immediate results without needing to send samples to a lab. Despite these advancements, there are currently no FDA-approved, at-home saliva-based hormone tests available on the market. Instead, most available salivary hormone test kits require users to collect saliva in a tube with preservation buffers and then send the samples to a laboratory for analysis [3], [24]. This lack of FDA-approved saliva-based hormone tests may be due to challenges related to standardization and the variability of hormone concentrations in saliva, which can be influenced by factors such as time of day, hydration levels, and individual differences [28]. Additionally, interpreting hormone data can be complex, often requiring medical expertise to contextualize results within a person's broader health profile. Given these challenges, a more realistic and

impactful future direction would be the development of a system for at-home saliva collection and analysis, where results could be logged and shared with healthcare professionals for remote monitoring and interpretation. A telemedicine-based approach could empower individuals to track hormone levels over time while ensuring that medical guidance is integrated into the process. Through the development of standardized collection methods and investigation of key salivary biomarkers, my research contributes to laying the groundwork for such a system, enhancing the feasibility and reliability of saliva-based hormone monitoring and bringing it one step closer to clinical implementation.

Understanding Biological Timescales with Saliva

Longitudinal salivaomics also presents an opportunity to explore fundamental biological questions about the temporal dynamics of biomarkers. Different biological processes operate on distinct timescales, with some biomarkers fluctuating rapidly in the order of minutes to hours, while others follow longer trends in the order of days to weeks [172]–[174]. Saliva's readily production and noninvasiveness enable frequent sampling, making it an ideal biofluid for capturing these dynamic changes with high temporal resolution. This raises the possibility of detecting biological signals that might otherwise be overlooked due to limited sampling resolution. As demonstrated in Chapter 4, immune response signals in saliva exhibited measurable changes on an hourly scale. This suggests that increasing the frequency of sampling could provide new insights into biomarker fluctuations and their biological relevance. Future research should systematically investigate optimal sampling intervals for different biomarkers, integrating high-frequency saliva sampling with advanced statistical models to refine health monitoring and disease progression studies.

5.4 Scalable Fluidics and High-Throughput Analysis

Fluidic systems are essential to a wide range of scientific applications, but their broader adoption depends on improvements in scalability, automation, and cost-effectiveness. Increasing the scalability and accessibility of fluidic systems could enable new discoveries by making it easier to conduct large-scale, high-throughput experiments. The colosseum open-source fraction collector, a key component of this dissertation, plays an important role in enabling high-throughput fluid collection for various research applications as described in Chapter 3. Although the use case of colosseum was presented in the context of salivary biomedicine in Chapter 3, the usefulness of colosseum is further demonstrated by its wide adoption across disci-

Studies using colosseum			
Title / Usage	Institution	Area	Year
Effects of Core Size and Surfactant Choice on Fluid Saturation Development in Surfactant/Polymer Corefloods [175]	ICL	Chemical & Environmental Science	2024
Multimodality Imaging of Fluid Saturation and Chemical Transport for Two-Phase Surfactant/Polymer Floods in Porous Rocks [176]	ICL	Chemistry & Geochemistry	2025
Water Treatment Sampling with colosseum	Yale	Chemical & Environmental Engineering	2024
Continuous culture for malaria parasite, Plasmodium falciparum	UC Berkeley	Infectious Diseases and Vaccinology	2024
Studies benchmarking colosseum			
Title	Area	Price (USD)	Year
Low-cost modular chromatography column rack and vial holders [177]	Column chromatography	334	2023
REVOLVER: A low-cost automated protein purifier based on parallel preparative gravity column workflows [178]	Protein purification	250	2022
Customizable large-scale HPLC fraction collection using low-cost 3D printing [179]	HPLC	280	2025
Turning a 3D Printer into a HPLC Fraction Collector: A Tool for Compound-Specific Stable Isotope Measurements [180]	HPLC	N/A (uses existing 3D printer)	2023
Open source fraction collector/MALDI spotter for proteomics [181]	Proteomics	1795	2022

Table 5.1: Applications of the colosseum fraction collector across scientific disciplines. The upper section showcases its diverse applications across research institutions and scientific disciplines, from chemical engineering to infectious disease research. The lower section highlights how colosseum has influenced the scientific community to develop and benchmark similar open-source fraction collectors, spanning various applications from column chromatography to proteomics. These benchmarking represent a growing ecosystem of affordable alternatives to commercial equipment. Together, these applications and derivative works illustrate how colosseum has not only lowered barriers to complex fluidic experiments but also inspired other researchers to continue advancing the democratization of scientific tools, amplifying the impact of the original innovation.

plines (Table 5.1), where its low cost and customizability have enabled researchers to conduct complex fluidics experiments without being constrained by the availability of specialized commercial equipment or advanced technology. Likewise, the poseidon syringe pump [73], designed by our lab in previous years, has contributed to the advancement of more efficient fluidic systems. Together, these devices create more opportunities for conducting complex fluid-based experiments at scale. Future research should focus on the continued development of open-source devices like colosseum, which have the potential to democratize scientific discovery by lowering accessibility barriers and enabling more laboratories to perform complex fluidic experiments and advance knowledge across disciplines.

Building on the concept of open-source and low-cost devices, the development of an affordable open-source sequencing platform would be transformative when integrated with fluidic systems like colosseum and poseidon. This integration would make large-scale studies, such as those in salivary transcriptomics, significantly more accessible and cost-effective. While existing platforms like Oxford Nanopore and open-source solutions such as DropSeq [182] have advanced the field, a purpose-built low-cost sequencing solution used in tandem with colosseum and poseidon could dramatically lower the barriers to conducting complex RNA sequencing experiments. This combination would create a fully open-source pipeline that enhances both accessibility and scalability for researchers with limited resources. In the context of salivary transcriptomics, this would enable more efficient RNA sequencing pipelines and facilitate broader applications in population health studies, precision medicine, and biomarker discovery, further unlocking the potential of salivary diagnostics as a noninvasive and scalable tool for health monitoring.

5.5 Prospective

The future of saliva-based diagnostics lies in improving accessibility, reliability, and scalability across collection methods, biomarker analyses, and data interpretation. Innovations such as optimized collection devices, tailored to diverse populations, and open-source fluidic systems like colosseum will pave the way for broader adoption of salivary biomedicine in both research and clinical settings. Furthermore, integrating at-home saliva collection with telemedicine-based analysis could empower individuals to monitor dynamic changes in their health easily and noninvasively while ensuring medical oversight. As salivary biomedicine evolves, the technologies and approaches developed in this dissertation provide a foundation for transformative advances in healthcare. By lowering the barrier to noninvasive biological monitoring,

these tools offer not just a window into human biology, but a gateway to previously inaccessible insights. The future of personalized medicine, population health, and patient empowerment will be significantly shaped by our ability to harness the full diagnostic potential of this remarkable biofluid.

BIBLIOGRAPHY

[1] C. Dawes and D. T. W. Wong, "Role of saliva and salivary diagnostics in the advancement of oral health," en, *J. Dent. Res.*, vol. 98, pp. 133–141, 2 Feb. 2019. DOI: 10.1177/0022034518816961.

[2] T. Nonaka and D. T. W. Wong, "Saliva diagnostics," en, *Annu. Rev. Anal. Chem. (Palo Alto Calif.)*, vol. 15, pp. 107–121, 1 Jun. 13, 2022. DOI: 10.1146/annurev-anchem-061020-123959.

[3] D. Malamud, "Saliva as a diagnostic fluid," en, *Dent. Clin. North Am.*, vol. 55, pp. 159–178, 1 Jan. 2011. DOI: 10.1016/j.cden.2010.08.004.

[4] K. Stark, C. Warnecke, V. Brinkmann, H. R. Gelderblom, U. Bienzle, and G. Pauli, "Sensitivity of HIV antibody detection in saliva," en, *Med. Microbiol. Immunol.*, vol. 182, pp. 147–151, 3 Jul. 1993. DOI: 10.1007/BF00190267.

[5] L. Zhang, H. Xiao, and D. T. Wong, "Salivary biomarkers for clinical applications," en, *Mol. Diagn. Ther.*, vol. 13, pp. 245–259, 4 2009. DOI: 10.1007/BF03256330.

[6] D. Belstrøm, "The salivary microbiota in health and disease," en, *J. Oral Microbiol.*, vol. 12, p. 1 723 975, 1 Feb. 4, 2020. DOI: 10.1080/20002297.2020.1723975.

[7] Y. Li, M. A. R. St John, X. Zhou, *et al.*, "Salivary transcriptome diagnostics for oral cancer detection," en, *Clin. Cancer Res.*, vol. 10, pp. 8442–8450, 24 Dec. 15, 2004. DOI: 10.1158/1078-0432.CCR-04-1167.

[8] X. Wang, K. E. Kaczor-Urbanowicz, and D. T. W. Wong, "Salivary biomarkers in cancer detection," en, *Med. Oncol.*, vol. 34, p. 7, 1 Jan. 2017. DOI: 10.1007/s12032-016-0863-4.

[9] S. G. Foddai, M. Radin, A. Barinotti, *et al.*, "New frontiers in autoimmune diagnostics: A systematic review on saliva testing," en, *Int. J. Environ. Res. Public Health*, vol. 20, p. 5782, 10 May 10, 2023. DOI: 10.3390/ijerph20105782.

[10] K. Ohyama, M. Moriyama, J.-N. Hayashida, *et al.*, "Saliva as a potential tool for diagnosis of dry mouth including sjögren's syndrome," en, *Oral Dis.*, vol. 21, pp. 224–231, 2 Mar. 2015. DOI: 10.1111/odi.12252.

[11] R. Farah, H. Haraty, Z. Salame, Y. Fares, D. M. Ojcius, and N. Said Sadier, "Salivary biomarkers for the diagnosis and monitoring of neurological diseases," en, *Biomed. J.*, vol. 41, pp. 63–87, 2 Apr. 2018. DOI: 10.1016/j.bj.2018.03.004.

[12] P. P. Mortimer and J. V. Parry, "Detection of antibody to HIV in saliva: A brief review," en, *Clin. Diagn. Virol.*, vol. 2, pp. 231–243, 4-5 Aug. 1994. DOI: 10.1016/0928-0197(94)90048-5.

[13] K. K.-W. To, O. T.-Y. Tsang, C. C.-Y. Yip, *et al.*, "Consistent detection of 2019 novel coronavirus in saliva," en, *Clin. Infect. Dis.*, vol. 71, pp. 841–843, 15 Jul. 28, 2020. DOI: 10.1093/cid/ciaa149.

[14] A. L. Wyllie, J. Fournier, A. Casanovas-Massana, *et al.*, "Saliva or nasopharyngeal swab specimens for detection of SARS-CoV-2," en, *N. Engl. J. Med.*, vol. 383, pp. 1283–1286, 13 Sep. 24, 2020. DOI: 10.1056/NEJMc2016359.

[15] J. V. Parry, K. R. Perry, S. Panday, and P. P. Mortimer, "Diagnosis of hepatitis a and B by testing saliva," en, *J. Med. Virol.*, vol. 28, pp. 255–260, 4 Aug. 1989. DOI: 10.1002/jmv.1890280410.

[16] A. R. Namuganga, N. N. Chegou, P. Mubiri, G. Walzl, and H. Mayanja-Kizza, "Suitability of saliva for tuberculosis diagnosis: Comparing with serum," en, *BMC Infect. Dis.*, vol. 17, p. 600, 1 Aug. 31, 2017. DOI: 10.1186/s12879-017-2687-z.

[17] P. R. Zhou, H. Hua, and X. S. Liu, "Quantity of candida colonies in saliva: A diagnostic evaluation for oral candidiasis," en, *Chin. J. Dent. Res.*, vol. 20, pp. 27–32, 1 2017. DOI: 10.3290/j.cjdr.a37739.

[18] G. I. Mias, V. V. Singh, L. R. K. Rogers, *et al.*, "Longitudinal saliva omics responses to immune perturbation: A case study," en, *Sci. Rep.*, vol. 11, p. 710, 1 Jan. 12, 2021. DOI: 10.1038/s41598-020-80605-6.

[19] F. G. Bellagambi, T. Lomonaco, P. Salvo, *et al.*, "Saliva sampling: Methods and devices. an overview," *Trends Analyt. Chem.*, vol. 124, p. 115 781, Mar. 1, 2020. DOI: 10.1016/j.trac.2019.115781.

[20] J. Heikenfeld, A. Jajack, B. Feldman, *et al.*, "Accessing analytes in biofluids for peripheral biochemical monitoring," en, *Nat. Biotechnol.*, vol. 37, pp. 407–419, 4 Apr. 2019. DOI: 10.1038/s41587-019-0040-3.

[21] S. Williamson, C. Munro, R. Pickler, M. J. Grap, and R. K. Elswick Jr, "Comparison of biomarkers in blood and saliva in healthy adults," en, *Nurs. Res. Pract.*, vol. 2012, p. 246 178, Apr. 30, 2012. DOI: 10.1155/2012/246178.

[22] J. Kells and C. M. Dollbaum, "Saliva tests, part 1: Clinical use, elements of testing, and guidelines for posttreatment interpretation," *Int. J. Pharm. Compd.*, vol. 13, 4 Jul. 2009. (visited on 09/06/2024).

[23] M. Villiger, R. Stoop, T. Vetsch, *et al.*, "Evaluation and review of body fluids saliva, sweat and tear compared to biochemical hydration assessment markers within blood and urine," en, *Eur. J. Clin. Nutr.*, vol. 72, pp. 69–76, 1 Jan. 2018. DOI: 10.1038/ejcn.2017.136.

[24] M. Gröschl, "Current status of salivary hormone analysis," en, *Clin. Chem.*, vol. 54, pp. 1759–1769, 11 Nov. 2008. DOI: 10.1373/clinchem.2008.108910.

[25] I. M. Ott, M. S. Strine, A. E. Watkins, *et al.*, "Stability of SARS-CoV-2 RNA in nonsupplemented saliva," en, *Emerg. Infect. Dis.*, vol. 27, pp. 1146–1150, 4 Apr. 2021. DOI: 10.3201/eid2704.204199.

[26] K. Thomadaki, E. J. Helmerhorst, N. Tian, *et al.*, "Whole-saliva proteolysis and its impact on salivary diagnostics," en, *J. Dent. Res.*, vol. 90, pp. 1325–1330, 11 Nov. 2011. DOI: 10.1177/0022034511420721.

[27] M. Quintana, O. Palicki, G. Lucchi, *et al.*, "Inter-individual variability of protein patterns in saliva of healthy adults," en, *J. Proteomics*, vol. 72, pp. 822–830, 5 Jul. 21, 2009. DOI: 10.1016/j.jprot.2009.05.004.

[28] C. Dawes, "Physiological factors affecting salivary flow rate, oral sugar clearance, and the sensation of dry mouth in man," en, *J. Dent. Res.*, vol. 66 Spec No, pp. 648–653, Feb. 1987. DOI: 10.1177/00220345870660S107.

[29] L. d'Amone, G. Matzeu, and F. G. Omenetto, "Stabilization of salivary biomarkers," en, *ACS Biomater. Sci. Eng.*, vol. 7, pp. 5451–5473, 12 Dec. 13, 2021. DOI: 10.1021/acsbiomaterials.1c01138.

[30] K. E. Kaczor-Urbanowicz, Y. Kim, F. Li, *et al.*, "Novel approaches for bioinformatic analysis of salivary RNA sequencing data for development," en, *Bioinformatics*, vol. 34, pp. 1–8, 1 Jan. 1, 2018. DOI: 10.1093/bioinformatics/btx504.

[31] D. T. Wong, "Salivary diagnostics: Amazing as it might seem, doctors can detect and monitor diseases using molecules found in a sample of spit," en, *Am. Sci.*, vol. 96, pp. 37–43, 1 Jan. 1, 2008. DOI: 10.1511/2008.69.3669.

[32] J. S. Bloom, L. Sathe, C. Munugala, *et al.*, "Massively scaled-up testing for SARS-CoV-2 RNA via next-generation sequencing of pooled and barcoded nasal and saliva samples," en, *Nat Biomed Eng*, vol. 5, pp. 657–665, 7 Jul. 2021. DOI: 10.1038/s41551-021-00754-5.

[33] S. H. Tan, O. Allicock, M. Armstrong-Hough, and A. L. Wyllie, "Saliva as a gold-standard sample for SARS-CoV-2 detection," en, *Lancet Respir Med*, vol. 9, pp. 562–564, 6 Jun. 2021. DOI: 10.1016/S2213-2600(21)00178-8.

[34] C. B. F. Vogels, A. E. Watkins, C. A. Harden, *et al.*, "SalivaDirect: A simplified and flexible platform to enhance SARS-CoV-2 testing capacity," en, *Med*, vol. 2, 263–280.e6, 3 Mar. 12, 2021. DOI: 10.1016/j.medj.2020.12.010.

[35] 23andMe, Inc. "23andMe and the FDA." (Feb. 15, 2025), [Online]. Available: https://customercare.23andme.com/hc/en-us/articles/211831908-23andMe-and-the-FDA (visited on 03/14/2025).

[36] U.S. Food and Drug Administration, Center for Drug Evaluation and Research., *OraQuick® HIV self-test 180-day supplement approval order*, Dec. 16, 2024. [Online]. Available: https://www.fda.gov/media/184589/download?attachment (visited on 03/14/2025).

[37] U.S. Food and Drug Administration, Center for Drug Evaluation and Research., *SalivaDirect letter of authorization*, Jul. 1, 2024. [Online]. Available: https://www.fda.gov/media/141194/download (visited on 03/14/2025).

[38] P. Ostheim, A. Tichý, I. Sirak, *et al.*, "Overcoming challenges in human saliva gene expression measurements," en, *Sci. Rep.*, vol. 10, p. 11 147, 1 Jul. 7, 2020. DOI: 10.1038/s41598-020-67825-6.

[39] K. E. Kaczor-Urbanowicz, C. Martin Carreras-Presas, K. Aro, M. Tu, F. Garcia-Godoy, and D. T. Wong, "Saliva diagnostics - current views and directions," en, *Exp. Biol. Med.*, vol. 242, pp. 459–472, 5 Mar. 2017. DOI: 10.1177/1535370216681550.

[40] National Cancer Institute. "NCI dictionary of cancer terms." (Feb. 2, 2011), [Online]. Available: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/cushing-syndrome (visited on 07/11/2023).

[41] Center for Biologics Evaluation and Research. "Information regarding the OraQuick in-home HIV test." (Feb. 25, 2022), [Online]. Available: https://www.fda.gov/vaccines-blood-biologics/approved-blood-products/information-regarding-oraquick-home-hiv-test (visited on 07/11/2023).

[42] C.-C. Chen, K.-Y. Hsiao, C.-H. Bai, and Y.-H. Wang, "Investigation of the diagnostic performance of the SARS-CoV-2 saliva antigen test: A meta-analysis," en, *J. Microbiol. Immunol. Infect.*, vol. 55, pp. 1084–1093, 6 Pt 1 Dec. 2022. DOI: 10.1016/j.jmii.2022.07.003.

[43] N. Grassl, N. A. Kulak, G. Pichler, *et al.*, "Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome," en, *Genome Med.*, vol. 8, p. 44, 1 Apr. 21, 2016. DOI: 10.1186/s13073-016-0293-0.

[44] G. Hiremath, A. Olive, S. Shah, C. M. Davis, R. J. Shulman, and S. Devaraj, "Comparing methods to collect saliva from children to analyze cytokines related to allergic inflammation," en, *Ann. Allergy Asthma Immunol.*, vol. 114, pp. 63–64, 1 Jan. 2015. DOI: 10.1016/j.anai.2014.09.012.

[45] S. Kidd, P. Midgley, N. Lone, *et al.*, "A re-investigation of saliva collection procedures that highlights the risk of potential positive interference in cortisol immunoassay," en, *Steroids*, vol. 74, pp. 666–668, 8 Aug. 2009. DOI: 10.1016/j.steroids.2009.02.009.

[46] O. M. Allicock, M. E. Petrone, D. Yolda-Carr, *et al.*, "Evaluation of saliva self-collection devices for SARS-CoV-2 diagnostics," en, *BMC Infect. Dis.*, vol. 22, pp. 1–9, 1 Mar. 25, 2022. DOI: 10.1186/s12879-022-07285-7.

[47] J. Figueira, S. Gouveia-Figueira, C. Öhman, P. Lif Holgerson, M. L. Nording, and A. Öhman, "Metabolite quantification by NMR and LC-MS/MS reveals differences between unstimulated, stimulated, and pure parotid saliva," en, *J. Pharm. Biomed. Anal.*, vol. 140, pp. 295–300, Jun. 5, 2017. DOI: 10.1016/j.jpba.2017.03.037.

[48] S. Gittings, N. Turnbull, B. Henry, C. J. Roberts, and P. Gershkovich, "Characterisation of human saliva as a platform for oral dissolution medium development," en, *Eur. J. Pharm. Biopharm.*, vol. 91, pp. 16–24, Apr. 2015. DOI: 10.1016/j.ejpb.2015.01.007.

[49] T. Kozaki, N. Hashiguchi, Y. Kaji, A. Yasukouchi, and Y. Tochihara, "Effects of saliva collection using cotton swab on cortisol enzyme immunoassay," en, *Eur. J. Appl. Physiol.*, vol. 107, pp. 743–746, 6 Dec. 2009. DOI: 10.1007/s00421-009-1178-3.

[50] Y. Maruyama, Y. Nishimoto, K. Umezawa, *et al.*, "Comparison of oral metabolome profiles of stimulated saliva, unstimulated saliva, and mouth-rinsed water," en, *Sci. Rep.*, vol. 12, p. 689, 1 Jan. 13, 2022. DOI: 10.1038/s41598-021-04612-x.

[51] P. P. Song, Y. Qi, and D. Cai, "Research and application of autodesk Fusion360 in industrial design," *IOP Publishing*, vol. 359, pp. 12 037–12 037, 2018. DOI: 10.1088/1757-899x/359/1/012037.

[52] C. G. Lausted, A. T. Johnson, W. H. Scott, M. M. Johnson, K. M. Coyne, and D. C. Coursey, "Maximum static inspiratory and expiratory pressures with different lung volumes," *Biomed. Eng. Online*, vol. 5, p. 29, May 5, 2006. DOI: 10.1186/1475-925X-5-29. [Online]. Available: http://dx.doi.org/10.1186/1475-925X-5-29.

[53] J. E. Varner and W. A. Bulen, "An automatic constant-volume fraction collector," *American Chemical Society*, vol. 29, pp. 625–625, 12 1952. DOI: 10.1021/ed029p625.

[54] A. Polson, "A simplified fraction collector for gradient elution chromatography," *Elsevier BV*, vol. 5, pp. 116–120, 1961. DOI: 10.1016/s0021-9673(01)92828-6.

[55] M. M. Jessop-Fabre and N. Sonnenschein, "Improving reproducibility in synthetic biology," *Frontiers Media*, vol. 7, 2019. DOI: 10.3389/fbioe.2019.00018.

[56] A. Madadlou, S. O'Sullivan, and D. Sheehan, "Fast protein liquid chromatography," *Springer Science+Business Media*, pp. 439–447, 2010. DOI: 10.1007/978-1-60761-913-0_25.

[57] Y. Li, M. Kuhn, J. Zukowska, *et al.*, "Coupling proteomics and metabolomics for the unsupervised identification of protein–metabolite interactions in chaetomium thermophilum," *Public Library of Science*, vol. 16, 7 2021. DOI: 10.1371/journal.pone.0254429.

[58] Izon Science Limited. "Isolate & measure your extracellular vesicle samples." (2023), [Online]. Available: https://www.izon.com/application/extracellular-vesicles (visited on 02/26/2025).

[59] L. Reseco, Á. Molina-Crespo, M. Atienza, E. González, J. M. Falcón-Pérez, and J. L. Cantero, "Characterization of extracellular vesicles from human saliva: Effects of age and isolation techniques," *Multidisciplinary Digital Publishing Institute*, vol. 13, pp. 95–95, 1 2024. DOI: 10.3390/cells13010095.

[60] T. J. Griffin, "The human saliva proteome: Overview and emerging methods for characterization," *Spiegel*, vol. 9490, pp. 949 003–949 003, 2015. DOI: 10.1117/12.2183316.

[61] I. Takeda, C. Stretch, P. Barnaby, *et al.*, "Understanding the human salivary metabolome," *Wiley Interdiscip. Rev. RNA*, vol. 22, pp. 577–584, 6 2009. DOI: 10.1002/nbm.1369.

[62] T. Guo, P. A. Rudnick, W. Wang, C. S. Lee, D. L. DeVoe, and B. M. Balgley, "Characterization of the human salivary proteome by capillary isoelectric focusing/nanoreversed-phase liquid chromatography coupled with ESI-tandem MS," *American Chemical Society*, vol. 5, pp. 1469–1478, 6 2006. DOI: 10.1021/pr060065m.

[63] Y. Sun, U. Braga-Neto, and E. R. Dougherty, "A systematic model of the LC-MS proteomics pipeline," en, *BMC Genomics*, vol. 13 Suppl 6, S2, S6 Oct. 26, 2012. DOI: 10.1186/1471-2164-13-S6-S2.

[64] S. AlTarawneh, M. B. Border, C. F. Dibble, and S. Bencharit, "Defining salivary biomarkers using mass spectrometry-based proteomics: A systematic review," *OMICS*, vol. 15, pp. 353–361, 6 2011. DOI: 10.1089/omi.2010.0134. [Online]. Available: http://dx.doi.org/10.1089/omi.2010.0134.

[65] J. M. Amenábar, B. M. D. Silva, and C. Punyadeera, "Salivary protein biomarkers for head and neck cancer," *Expert Rev. Mol. Diagn.*, vol. 20, pp. 305–313, 3 2020. DOI: 10.1080/14737159.2020.1722102. [Online]. Available: http://dx.doi.org/10.1080/14737159.2020.1722102.

[66] S. Hu, J. Wang, J. Meijer, *et al.*, "Salivary proteomic and genomic biomarkers for primary sjögren's syndrome," *Wiley Interdiscip. Rev. RNA*, vol. 56, pp. 3588–3600, 11 2007. DOI: 10.1002/art.22954.

[67] S. Hu, M. Arellano, P. Boontheung, *et al.*, "Salivary proteomics for oral cancer biomarker discovery," *American Association for Cancer Research*, vol. 14, pp. 6246–6252, 19 2008. DOI: 10.1158/1078-0432.ccr-07-5037.

[68] A. Zhang, H. Sun, P. Wang, and X. Wang, "Salivary proteomics in biomedical research," *Clin. Chim. Acta*, vol. 415, pp. 261–265, 2012. DOI: 10.1016/j.cca.2012.11.001. [Online]. Available: http://dx.doi.org/10.1016/j.cca.2012.11.001.

[69] X. Zhang, Z. Wu, Y. Xu, *et al.*, "Saliva proteomic analysis reveals possible biomarkers of renal cell carcinoma," *De Gruyter Open*, vol. 18, pp. 918–926, 1 2020. DOI: 10.1515/chem-2020-0048.

[70] S. Williams, M. Fernández-Rhodes, A. Law, B. Peacock, M. P. Lewis, and O. G. Davies, "Comparison of extracellular vesicle isolation processes for therapeutic applications," *SAGE Publishing*, vol. 14, 2023. DOI: 10.1177/20417314231174609.

[71] K. Helton, K. E. Nelson, E. Fu, and P. Yager, "Conditioning saliva for use in a microfluidic biosensor," *Royal Society of Chemistry*, vol. 8, pp. 1847–1847, 11 2008. DOI: 10.1039/b811150b.

[72] D. Esser, G. Álvarez-Llamas, M. P. de Vries, D. Weening, R. J. Vonk, and H. Roelofsen, "Sample stability and protein composition of saliva: Implications for its use as a diagnostic fluid," *SAGE Publishing*, vol. 3, 2008. DOI: 10.4137/bmi.s607.

[73] A. S. Booeshaghi, E. da Veiga Beltrame, D. Bannon, J. Gehring, and L. Pachter, "Principles of open source bioinstrumentation applied to the poseidon syringe pump system," *Nature Portfolio*, vol. 9, 1 2019. DOI: 10.1038/s41598-019-48815-9. [Online]. Available: https://doi.org/10.1038/s41598-019-48815-9.

[74] S. A. Longwell and P. M. Fordyce, "micrIO: An open-source autosampler and fraction collector for automated microfluidic input–output," *Royal Society of Chemistry*, vol. 20, pp. 93–106, 1 2019. DOI: 10.1039/c9lc00512a.

[75] M. Caputo, J. T. Lyles, M. S. Salazar, and C. L. Quave, "LEGO MINDSTORMS fraction collector: A low-cost tool for a preparative high-performance liquid chromatography system," *American Chemical Society*, vol. 92, pp. 1687–1690, 2 2019. DOI: 10.1021/acs.analchem.9b04299.

[76] E. Specht. "The best known packings of equal circles in a 1x0.30000 rectangle (complete up to N = 300)." (Aug. 11, 2010), [Online]. Available: http://hydra.nat.uni-magdeburg.de/packing/crc_300/crc.html (visited on 03/24/2025).

[77] M. Cáceres, R. Grant, R. Waldron, *et al.* "Web serial API." (2013), [Online]. Available: https://wicg.github.io/serial/ (visited on 03/24/2025).

[78] A. S. Booeshaghi, *Browser-serial: First release*, comp. software, version 1.0.3, Mar. 18, 2021. [Online]. Available: https://github.com/sbooeshaghi/browser-serial.

[79] M. S. Cheri, H. Latifi, J. Sadeghi, M. S. Moghaddam, H. Shahraki, and H. Hajghassem, "Real-time measurement of flow rate in microfluidic devices using a cantilever-based optofluidic sensor," *Royal Society of Chemistry*, vol. 139, pp. 431–438, 2 2013. DOI: 10.1039/c3an01588b.

[80] E. Westman, S. Eriksson, T. Låås, P.-Å. Pernemalm, and S.-e. Sköld, "Separation of DNA restriction fragments by ion-exchange chromatography on FPLC columns mono P and mono Q," *Elsevier BV*, vol. 166, pp. 158–171, 1 1987. DOI: 10.1016/0003-2697(87)90558-6.

[81] A. Labs. "Chromatography flow rates explained: HPLC, LC-MS & GC." (Jan. 27, 2025), [Online]. Available: https://axionlabs.com/chromatography-training/chromatography-flow-rates-explained-hplc-lc-ms-gc/ (visited on 03/17/2025).

[82] Simplify3D. "Simplify3D." (Jan. 2013), [Online]. Available: https://www.simplify3d.com/ (visited on 03/24/2025).

[83] Prusa Research. "Original prusa 3D printers directly from josef prusa." (2017), [Online]. Available: https://www.prusa3d.com/category/original-prusa-i3-mk3/ (visited on 03/24/2025).

[84] Y. Kil. "Assembly instructions for colosseum protocol guidelines." (2021), [Online]. Available: http://dx.doi.org/10.17504/protocols.io.btz3np8n (visited on 03/24/2025).

[85] Pololu Corporation. "DRV8825 stepper motor driver carrier, high current." (2010), [Online]. Available: https://www.pololu.com/product/2133/specs (visited on 03/24/2025).

[86] A. S. Booeshaghi, *Pegasus: Modular stepper motor control with arduino, CNC motor sheild, and pololu stepper driver. also the workhorse of poseidon and colosseum*, comp. software, version 32ca075b38a72a7955209657a8326ac749f658a3, Aug. 8, 2019. [Online]. Available: https://github.com/pachterlab/pegasus (visited on 02/27/2025).

[87] C. Liechti, *Pyserial*, comp. software, version 3.5, Nov. 22, 2020. [Online]. Available: https://pypi.org/project/pyserial/#description.

[88] H. G. Pontis, "Protein and carbohydrate separation and purification," in *Methods for Analysis of Carbohydrate Metabolism in Photosynthetic Organisms*. Elsevier, 2017, pp. 45–63. DOI: 10.1016/b978-0-12-803396-8.00003-x. [Online]. Available: http://dx.doi.org/10.1016/b978-0-12-803396-8.00003-x.

[89] H. Xiao, Y. Zhang, Y. Kim, *et al.*, "Differential proteomic analysis of human saliva using tandem mass tags quantification for gastric cancer detection," *Sci. Rep.*, vol. 6, p. 22 165, 1 Feb. 25, 2016. DOI: 10.1038/srep22165. [Online]. Available: https://www.nature.com/articles/srep22165 (visited on 03/31/2025).

[90] S. Hu, J. Jiang, and D. T. Wong, "Proteomic analysis of saliva: 2D gel electrophoresis, LC-MS/MS, and western blotting," *Methods Mol. Biol.*, vol. 666, pp. 31–41, 2010. DOI: 10.1007/978-1-60761-820-1_3. [Online]. Available: http://dx.doi.org/10.1007/978-1-60761-820-1_3.

[91] P. Sivadasan, M. Kumar Gupta, G. J. Sathe, *et al.*, "Data from human salivary proteome - a resource of potential biomarkers for oral cancer," *Data Brief*, vol. 4, pp. 374–378, Sep. 2015. DOI: 10.1016/j.dib.2015.06.014. [Online]. Available: http://dx.doi.org/10.1016/j.dib.2015.06.014.

[92] Gilson Incorporated. "Maximize purity, yield, and throughput with improved fraction collection." (Aug. 31, 2023), [Online]. Available: https://www.gilson.com/default/learninghub/post/maximize-purity-yield-and-throughput-with-improved-fraction-collection.html (visited on 03/31/2025).

[93] L. R. Snyder, "Principles of gradient elution," *Chromatogr. Rev.*, vol. 7, pp. 1–51, 1965. DOI: 10.1016/0009-5907(65)80002-3. [Online]. Available: http://dx.doi.org/10.1016/0009-5907(65)80002-3.

[94] Z. LaCasse, P. Chivte, K. Kress, *et al.*, "Enhancing saliva diagnostics: The impact of amylase depletion on MALDI-ToF MS profiles as applied to COVID-19," *J. Mass Spectrom. Adv. Clin. Lab*, vol. 31, pp. 59–71, Jan. 2024. DOI: 10.1016/j.jmsacl.2024.01.003. [Online]. Available: http://dx.doi.org/10.1016/j.jmsacl.2024.01.003.

[95] A. Gondane and H. M. Itkonen, "Revealing the history and mystery of RNA-seq," en, *Curr. Issues Mol. Biol.*, vol. 45, pp. 1860–1874, 3 Feb. 24, 2023. DOI: 10.3390/cimb45030120.

[96] M. H. Askenase, B. A. Goods, H. E. Beatty, *et al.*, "Longitudinal transcriptomics define the stages of myeloid activation in the living human brain after intracerebral hemorrhage," en, *Sci Immunol*, vol. 6, 56 Feb. 19, 2021. DOI: 10.1126/sciimmunol.abd6279.

[97] J. Bouquet, M. J. Soloski, A. Swei, *et al.*, "Longitudinal transcriptome analysis reveals a sustained differential gene expression signature in patients treated for acute lyme disease," en, *MBio*, vol. 7, e00100–16, 1 Feb. 12, 2016. DOI: 10.1128/mBio.00100-16.

[98] M. Eltobgy, F. Johns, D. Farkas, *et al.*, "Longitudinal transcriptomic analysis reveals persistent enrichment of iron homeostasis and erythrocyte function pathways in severe COVID-19 ARDS," en, *Front. Immunol.*, vol. 15, p. 1 397 629, Aug. 5, 2024. DOI: 10.3389/fimmu.2024.1397629.

[99] J. C. Foo, N. Trautmann, C. Sticht, *et al.*, "Longitudinal transcriptome-wide gene expression analysis of sleep deprivation treatment shows involvement

of circadian genes and immune pathways," en, *Transl. Psychiatry*, vol. 9, p. 343, 1 Dec. 18, 2019. DOI: 10.1038/s41398-019-0671-7.

[100] Y. Lefol, T. Korfage, R. Mjelle, *et al.*, "TiSA: TimeSeriesAnalysis-a pipeline for the analysis of longitudinal transcriptomics data," en, *NAR Genom Bioinform*, vol. 5, lqad020, 1 Mar. 2023. DOI: 10.1093/nargab/lqad020.

[101] A. Michna, H. Braselmann, M. Selmansberger, *et al.*, "Natural cubic spline regression modeling followed by dynamic network reconstruction for the identification of radiation-sensitivity gene association networks from time-course transcriptome data," en, *PLoS One*, vol. 11, e0160791, 8 Aug. 9, 2016. DOI: 10.1371/journal.pone.0160791.

[102] H. Mulenga, M. Musvosvi, S. C. Mendelsohn, *et al.*, "Longitudinal dynamics of a blood transcriptomic signature of tuberculosis," en, *Am. J. Respir. Crit. Care Med.*, vol. 204, pp. 1463–1472, 12 Dec. 15, 2021. DOI: 10.1164/rccm.202103-0548OC.

[103] R. Su, M. M. Li, N. R. Bhakta, *et al.*, "Longitudinal analysis of sarcoidosis blood transcriptomic signatures and disease outcomes," en, *Eur. Respir. J.*, vol. 44, pp. 985–993, 4 Oct. 2014. DOI: 10.1183/09031936.00039714.

[104] P. Gao, X. Shen, X. Zhang, *et al.*, "Precision environmental health monitoring by longitudinal exposome and multi-omics profiling," en, *Genome Res.*, vol. 32, pp. 1199–1214, 6 Jun. 2022. DOI: 10.1101/gr.276521.121.

[105] V.-K. S. Oh and R. W. Li, "Wise roles and future visionary endeavors of current emperor: Advancing dynamic methods for longitudinal microbiome meta-omics data in personalized and precision medicine," en, *Adv. Sci. (Weinh.)*, vol. 11, e2400458, 47 Dec. 2024. DOI: 10.1002/advs.202400458.

[106] S. M. Schüssler-Fiorenza Rose, K. Contrepois, K. J. Moneghetti, *et al.*, "A longitudinal big data approach for precision health," en, *Nat. Med.*, vol. 25, pp. 792–804, 5 May 2019. DOI: 10.1038/s41591-019-0414-6.

[107] N. J. Bonne and D. T. Wong, "Salivary biomarker development using genomic, proteomic and metabolomic approaches," en, *Genome Med.*, vol. 4, p. 82, 10 Oct. 30, 2012. DOI: 10.1186/gm383.

[108] M. Soukup, I. Biesiada, A. Henderson, *et al.*, "Salivary uric acid as a non-invasive biomarker of metabolic syndrome," en, *Diabetol. Metab. Syndr.*, vol. 4, p. 14, 1 Apr. 19, 2012. DOI: 10.1186/1758-5996-4-14.

[109] B. Zyśk, L. Ostrowska, and J. Smarkusz-Zarzecka, "Salivary adipokine and cytokine levels as potential markers for the development of obesity and metabolic disorders," en, *Int. J. Mol. Sci.*, vol. 22, p. 11 703, 21 Oct. 28, 2021. DOI: 10.3390/ijms222111703.

[110] E. T. Wang, R. Sandberg, S. Luo, *et al.*, "Alternative isoform regulation in human tissue transcriptomes," en, *Nature*, vol. 456, pp. 470–476, 7221 Nov. 27, 2008. DOI: 10.1038/nature07509.

[111] K. Vitting-Seerup and A. Sandelin, "The landscape of isoform switches in human cancers," en, *Mol. Cancer Res.*, vol. 15, pp. 1206–1220, 9 Sep. 2017. DOI: 10.1158/1541-7786.MCR-16-0459.

[112] L. Yi, H. Pimentel, N. L. Bray, and L. Pachter, "Gene-level differential analysis at transcript-level resolution," en, *Genome Biol.*, vol. 19, p. 53, 1 Apr. 12, 2018. DOI: 10.1186/s13059-018-1419-z.

[113] J. Gauthier, Q. V. Wu, and T. A. Gooley, "Cubic splines to model relationships between continuous variables and outcomes: A guide for clinicians," en, *Bone Marrow Transplant.*, vol. 55, pp. 675–680, 4 Apr. 2020. DOI: 10.1038/s41409-019-0679-x.

[114] A. Perperoglou, W. Sauerbrei, M. Abrahamowicz, and M. Schmid, "A review of spline function procedures in R," *BMC Med. Res. Methodol.*, vol. 19, p. 46, 1 Mar. 6, 2019. DOI: 10.1186/s12874-019-0666-3.

[115] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, en. Springer Nature, Jul. 29, 2021, 607 pp.

[116] B. Jassal, L. Matthews, G. Viteri, *et al.*, "The reactome pathway knowledgebase," en, *Nucleic Acids Res.*, vol. 48, pp. D498–D503, D1 Jan. 8, 2020. DOI: 10.1093/nar/gkz1031.

[117] Executive Management Group. "TMPRSS11E protein expression summary." (), [Online]. Available: https://www.proteinatlas.org/ENSG00000087128-TMPRSS11E (visited on 02/07/2025).

[118] National Center for Biotechnology Information. "TYROBP transmembrane immune signaling adaptor TYROBP [homo sapiens (human)] - gene." (Mar. 8, 2025), [Online]. Available: https://www.ncbi.nlm.nih.gov/gene/7305 (visited on 03/24/2025).

[119] C. Seillier, P. Hélie, G. Petit, *et al.*, "Roles of the tissue-type plasminogen activator in immune response," en, *Cell. Immunol.*, vol. 371, p. 104 451, 104451 Jan. 2022. DOI: 10.1016/j.cellimm.2021.104451.

[120] V. Audrito, V. G. Messana, and S. Deaglio, "NAMPT and NAPRT: Two metabolic enzymes with key roles in inflammation," en, *Front. Oncol.*, vol. 10, p. 358, Mar. 19, 2020. DOI: 10.3389/fonc.2020.00358.

[121] M. M. Waldman, J. T. Rahkola, A. L. Sigler, *et al.*, "Ena/VASP protein-mediated actin polymerization contributes to naïve CD8+ T cell activation and expansion by promoting T cell-APC interactions in vivo," *Front. Immunol.*, vol. 13, p. 856 977, Jun. 9, 2022. DOI: 10.3389/fimmu.2022.

856977. [Online]. Available: http://dx.doi.org/10.3389/fimmu.2022.856977.

[122] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: A bioconductor package for differential expression analysis of digital gene expression data," en, *Bioinformatics*, vol. 26, pp. 139–140, 1 Jan. 1, 2010. DOI: 10.1093/bioinformatics/btp616.

[123] M. E. Ritchie, B. Phipson, D. Wu, *et al.*, "Limma powers differential expression analyses for RNA-sequencing and microarray studies," en, *Nucleic Acids Res.*, vol. 43, e47, 7 Apr. 20, 2015. DOI: 10.1093/nar/gkv007.

[124] V. Svensson. "Timecourse analysis with sleuth — what do you mean "heterogeneity"?" en. (Nov. 29, 2015), (visited on 08/14/2024).

[125] H. Pimentel, N. L. Bray, S. Puente, P. Melsted, and L. Pachter, "Differential analysis of RNA-seq incorporating quantification uncertainty," en, *Nat. Methods*, vol. 14, pp. 687–690, 7 Jul. 2017. DOI: 10.1038/nmeth.4324.

[126] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter, "Near-optimal probabilistic RNA-seq quantification," en, *Nat. Biotechnol.*, vol. 34, pp. 525–527, 5 May 2016. DOI: 10.1038/nbt.3519.

[127] R. R Foundation for Statistical Computing, "R: A language and environment for statistical computing," *RA Lang Environ Stat Comput*, 2018.

[128] M. V. Kuleshov, M. R. Jones, A. D. Rouillard, *et al.*, "Enrichr: A comprehensive gene set enrichment analysis web server 2016 update," en, *Nucleic Acids Res.*, vol. 44, W90–7, W1 Jul. 8, 2016. DOI: 10.1093/nar/gkw377.

[129] Y. Kil, A. S. Booeshaghi, and L. S. Pachter, "Comparative survey-based study of noninvasive saliva collection devices," en, *J. Med. Device.*, vol. 19, pp. 1–19, 2 Jun. 1, 2025. DOI: 10.1115/1.4067232.

[130] F. Amado, M. J. Calheiros-Lobo, R. Ferreira, and R. Vitorino, "Sample treatment for saliva proteomics," en, *Adv. Exp. Med. Biol.*, vol. 1073, pp. 23–56, 2019. DOI: 10.1007/978-3-030-12298-0_2.

[131] J. A. Saavedra, R. P. Muñoz, G. R. Alcayaga, *et al.*, "Salivary flow rate response to stimulation with 2% citric acid in patients with xerostomia," *J. Oral Diagn.*, vol. 3, pp. 1–6, Jan. 20, 2018. DOI: 10.5935/2525-5711.20180005.

[132] U. N. Lee, X. Su, D. L. Hieber, *et al.*, "CandyCollect: At-home saliva sampling for capture of respiratory pathogens," en, *Lab Chip*, vol. 22, pp. 3555–3564, 18 Sep. 13, 2022. DOI: 10.1039/d1lc01132d.

[133] Y. Han, L. Jia, Y. Zheng, and W. Li, "Salivary exosomes: Emerging roles in systemic disease," en, *Int. J. Biol. Sci.*, vol. 14, pp. 633–643, 6 Apr. 30, 2018. DOI: 10.7150/ijbs.25018.

[134] K. Sidhom, P. O. Obi, and A. Saleem, "A review of exosomal isolation methods: Is size exclusion chromatography the best option?" en, *Int. J. Mol. Sci.*, vol. 21, p. 6466, 18 Sep. 4, 2020. DOI: 10.3390/ijms21186466.

[135] K. Aro, F. Wei, D. T. Wong, and M. Tu, "Saliva liquid biopsy for point-of-care applications," *Front. Public Health*, vol. 5, p. 77, Apr. 11, 2017. DOI: 10.3389/fpubh.2017.00077. [Online]. Available: http://dx.doi.org/10.3389/fpubh.2017.00077.

[136] S. Y. Oh, S.-M. Kang, S. H. Kang, *et al.*, "Potential salivary mRNA biomarkers for early detection of oral cancer," en, *J. Clin. Med.*, vol. 9, p. 243, 1 Jan. 16, 2020. DOI: 10.3390/jcm9010243.

[137] F. Xu and M. Jiang, "Evaluation of predictive role of carcinoembryonic antigen and salivary mRNA biomarkers in gastric cancer detection," en, *Medicine (Baltimore)*, vol. 99, e20419, 22 May 29, 2020. DOI: 10.1097/MD.0000000000020419.

[138] S. D. Tyrkalska, F. Pérez-Sanz, L. Franco-Martínez, *et al.*, "Salivary biomarkers as pioneering indicators for diagnosis and severity stratification of pediatric long COVID," en, *Front. Cell. Infect. Microbiol.*, vol. 14, p. 1 396 263, May 31, 2024. DOI: 10.3389/fcimb.2024.1396263.

[139] L. Verdiguel-Fernández, R. Arredondo-Hernández, J. A. Mejía-Estrada, *et al.*, "Differential expression of biomarkers in saliva related to SARS-CoV-2 infection in patients with mild, moderate and severe COVID-19," en, *BMC Infect. Dis.*, vol. 23, p. 602, 1 Sep. 15, 2023. DOI: 10.1186/s12879-023-08573-6.

[140] S. Abo, D. Smith, M. Stadt, and A. Layton, "Modelling female physiology from head to toe: Impact of sex hormones, menstrual cycle, and pregnancy," en, *J. Theor. Biol.*, vol. 540, p. 111 074, 111074 May 7, 2022. DOI: 10.1016/j.jtbi.2022.111074.

[141] M. M. Costantine, "Physiologic and pharmacokinetic changes in pregnancy," en, *Front. Pharmacol.*, vol. 5, p. 65, Apr. 3, 2014. DOI: 10.3389/fphar.2014.00065.

[142] G. E. Hale and H. G. Burger, "Hormonal changes and biomarkers in late reproductive age, menopausal transition and menopause," en, *Best Pract. Res. Clin. Obstet. Gynaecol.*, vol. 23, pp. 7–23, 1 Feb. 2009. DOI: 10.1016/j.bpobgyn.2008.10.001.

[143] V. Hendrick, L. L. Altshuler, and R. Suri, "Hormonal changes in the postpartum and implications for postpartum depression," en, *Psychosomatics*, vol. 39, pp. 93–101, 2 Mar. 1998. DOI: 10.1016/s0033-3182(98)71355-6.

[144] H. A. Zacur, "Hormonal changes throughout life in women," en, *Headache*, vol. 46 Suppl 2, S49–54, s2 Oct. 2006. DOI: 10.1111/j.1526-4610.2006.00554.x.

[145] M. A. Farage, S. Neill, and A. B. MacLean, "Physiological changes associated with the menstrual cycle: A review," en, *Obstet. Gynecol. Surv.*, vol. 64, pp. 58–72, 1 Jan. 2009. DOI: 10.1097/OGX.0b013e3181932a37.

[146] O. L. Quintero, M. J. Amador-Patarroyo, G. Montoya-Ortiz, A. Rojas-Villarraga, and J.-M. Anaya, "Autoimmune disease and gender: Plausible mechanisms for the female predominance of autoimmunity," en, *J. Autoimmun.*, vol. 38, J109–19, 2-3 May 2012. DOI: 10.1016/j.jaut.2011.10.003.

[147] M. Saei Ghare Naz, M. Rostami Dovom, and F. Ramezani Tehrani, "The menstrual disturbances in endocrine disorders: A narrative review," en, *Int. J. Endocrinol. Metab.*, vol. 18, e106694, 4 Oct. 2020. DOI: 10.5812/ijem.106694.

[148] J. Yang and C. Chen, "Hormonal changes in PCOS," en, *J. Endocrinol.*, vol. 261, 1 Apr. 1, 2024. DOI: 10.1530/JOE-23-0342.

[149] P. H. Gann, S. Giovanazzi, L. Van Horn, A. Branning, and R. T. Chatterton Jr, "Saliva as a medium for investigating intra- and interindividual differences in sex hormone levels in premenopausal women," en, *Cancer Epidemiol. Biomarkers Prev.*, vol. 10, pp. 59–64, 1 Jan. 2001.

[150] N. Gavrilova and S. T. Lindau, "Salivary sex hormone measurement in a national, population-based study of older adults," en, *J. Gerontol. B Psychol. Sci. Soc. Sci.*, vol. 64 Suppl 1, pp. i94–105, Supplement 1 Nov. 2009. DOI: 10.1093/geronb/gbn028.

[151] E. Harmon-Jones and J. S. Beer, Eds., *Methods in Social Neuroscience*. New York, NY: Guilford Publications, Feb. 9, 2009, ch. Assessment of Salivary Hormones, 353 pp.

[152] Y.-H. Lee, Y.-Y. Kim, J.-Y. Chang, and H.-S. Kho, "Changes in oral mucosal MUC1 expression and salivary hormones throughout the menstrual cycle," en, *Oral Dis.*, vol. 21, pp. 962–968, 8 Nov. 2015. DOI: 10.1111/odi.12367.

[153] J. K. Walker, C. C. Dillard, D. E. Gonzalez, H. S. Waldman, and M. J. McAllister, "Impact of the menstrual cycle phases and time of day on markers of stress: Salivary α-amylase and secretory immunoglobulin a," en, *Stress*, vol. 28, p. 2 449 098, 1 Dec. 2025. DOI: 10.1080/10253890.2024.2449098.

[154] N. Wendland, J. Opydo-Szymaczek, D. Formanowicz, A. Blacha, G. Jarząbek-Bielecka, and M. Mizgier, "Association between metabolic and hormonal profile, proinflammatory cytokines in saliva and gingival health in adolescent females with polycystic ovary syndrome," en, *BMC Oral Health*, vol. 21, p. 193, 1 Apr. 13, 2021. DOI: 10.1186/s12903-021-01553-9.

[155] Committee on Increasing the Number of Women in Science, Technology, Engineering, Mathematics, and Medicine (STEMM), Committee on Women in Science, Engineering, and Medicine, Policy and Global Affairs, and National Academies of Sciences, Engineering, and Medicine, *Promising practices for addressing the underrepresentation of women in science, engineering, and medicine: Opening doors*, J. Saunders, Ed. Washington, D.C.: National Academies Press, Apr. 30, 2020. DOI: 10.17226/25785.

[156] C. M. Mazure and D. P. Jones, "Twenty years and still counting: Including women as participants and studying sex and gender in biomedical research," en, *BMC Womens. Health*, vol. 15, p. 94, 1 Oct. 26, 2015. DOI: 10.1186/s12905-015-0251-9.

[157] D. Soo-Quee Koh and G. Choon-Huat Koh, "The use of salivary biomarkers in occupational and environmental medicine," en, *Occup. Environ. Med.*, vol. 64, pp. 202–210, 3 Mar. 2007. DOI: 10.1136/oem.2006.026567.

[158] Y. J. Bae, J. Reinelt, J. Netto, *et al.*, "Salivary cortisone, as a biomarker for psychosocial stress, is associated with state anxiety and heart rate," en, *Psychoneuroendocrinology*, vol. 101, pp. 35–41, Mar. 2019. DOI: 10.1016/j.psyneuen.2018.10.015.

[159] C. S. Carter, H. Pournajafi-Nazarloo, K. M. Kramer, *et al.*, "Oxytocin: Behavioral associations and potential as a salivary biomarker," en, *Ann. N. Y. Acad. Sci.*, vol. 1098, pp. 312–322, 1 Mar. 2007. DOI: 10.1196/annals.1384.006.

[160] S. Chojnowska, I. Ptaszyńska-Sarosiek, A. Kępka, M. Knaś, and N. Waszkiewicz, "Salivary biomarkers of stress, anxiety and depression," en, *J. Clin. Med.*, vol. 10, p. 517, 3 Feb. 1, 2021. DOI: 10.3390/jcm10030517.

[161] Y. Noto, T. Sato, M. Kudo, K. Kurata, and K. Hirota, "The relationship between salivary biomarkers and state-trait anxiety inventory score under mental arithmetic stress: A pilot study," en, *Anesth. Analg.*, vol. 101, pp. 1873–1876, 6 Dec. 2005. DOI: 10.1213/01.ANE.0000184196.60838.8D.

[162] M. Pundir, S. Papagerakis, M. C. De Rosa, *et al.*, "Emerging biotechnologies for evaluating disruption of stress, sleep, and circadian rhythm mechanism using aptamer-based detection of salivary biomarkers," en, *Biotechnol. Adv.*, vol. 59, p. 107 961, 107961 Oct. 2022. DOI: 10.1016/j.biotechadv.2022.107961.

[163] T. Iqbal, A. Elahi, W. Wijns, and A. Shahzad, "Cortisol detection methods for stress monitoring in connected health," *Health Sciences Review*, vol. 6, p. 100 079, Mar. 1, 2023. DOI: 10.1016/j.hsr.2023.100079.

[164] J. Blair, J. Adaway, B. Keevil, and R. Ross, "Salivary cortisol and cortisone in the clinical setting," en, *Curr. Opin. Endocrinol. Diabetes Obes.*, vol. 24, pp. 161–168, 3 Jun. 2017. DOI: 10.1097/MED.0000000000000328.

[165] B. Arévalo, V. Serafín, J. F. Beltrán-Sánchez, *et al.*, "Simultaneous determination of four fertility-related hormones in saliva using disposable multiplexed immunoplatforms coupled to a custom-designed and field-portable potentiostat," en, *Anal. Methods*, vol. 13, pp. 3471–3478, 31 Aug. 12, 2021. DOI: 10.1039/d1ay01074c.

[166] A. Gevaerd, E. Y. Watanabe, C. Belli, L. H. Marcolino-Junior, and M. F. Bergamini, "A complete lab-made point of care device for non-immunological electrochemical determination of cortisol levels in salivary samples," en, *Sens. Actuators B Chem.*, vol. 332, p. 129 532, 129532 Apr. 2021. DOI: 10.1016/j.snb.2021.129532.

[167] A. Apilux, S. Rengpipat, W. Suwanjang, and O. Chailapakul, "Development of competitive lateral flow immunoassay coupled with silver enhancement for simple and sensitive salivary cortisol detection," en, *EXCLI J.*, vol. 17, pp. 1198–1209, Dec. 21, 2018. DOI: 10.17179/excli2018-1824.

[168] S. Dalirirad, D. Han, and A. J. Steckl, "Aptamer-based lateral flow biosensor for rapid detection of salivary cortisol," en, *ACS Omega*, vol. 5, pp. 32 890–32 898, 51 Dec. 29, 2020. DOI: 10.1021/acsomega.0c03223.

[169] H.-K. Oh, J.-W. Kim, J.-M. Kim, and M.-G. Kim, "High sensitive and broad-range detection of cortisol in human saliva using a trap lateral flow immunoassay (trapLFI) sensor," *Analyst*, 2018. DOI: 10.1039/c8an00719e.

[170] M. Yamaguchi, Y. Matsuda, S. Sasaki, *et al.*, "Immunosensor with fluid control mechanism for salivary cortisol analysis," en, *Biosens. Bioelectron.*, vol. 41, pp. 186–191, Mar. 15, 2013. DOI: 10.1016/j.bios.2012.08.016.

[171] M. Zangheri, L. Cevenini, L. Anfossi, *et al.*, "A simple and compact smartphone accessory for quantitative chemiluminescence-based lateral flow immunoassay for salivary cortisol detection," en, *Biosens. Bioelectron.*, vol. 64, pp. 63–68, Feb. 15, 2015. DOI: 10.1016/j.bios.2014.08.048.

[172] M. Andraud, O. Lejeune, J. Z. Musoro, B. Ogunjimi, P. Beutels, and N. Hens, "Living on three time scales: The dynamics of plasma cell and antibody populations illustrated for hepatitis a virus," en, *PLoS Comput. Biol.*, vol. 8, e1002418, 3 Mar. 1, 2012. DOI: 10.1371/journal.pcbi.1002418.

[173] D. V. Buonomano, "The biology of time across different scales," en, *Nat. Chem. Biol.*, vol. 3, pp. 594–597, 10 Oct. 2007. DOI: 10.1038/nchembio1007-594.

[174] M. Shamir, Y. Bar-On, R. Phillips, and R. Milo, "SnapShot: Timescales in cell biology," en, *Cell*, vol. 164, 1302–1302.e1, 6 Mar. 10, 2016. DOI: 10.1016/j.cell.2016.02.058.

[175] A. Rovelli, J. Brodie, B. Rashid, W. J. Tay, and R. Pini, "Effects of core size and surfactant choice on fluid saturation development in surfactant/polymer corefloods," *Energy Fuels*, vol. 38, pp. 2844–2854, 4 Feb. 15, 2024. DOI:

10.1021/acs.energyfuels.3c04313. [Online]. Available: http://dx.doi.org/10.1021/acs.energyfuels.3c04313.

[176] A. Rovelli, J. Brodie, B. Rashid, W. J. Tay, and R. Pini, "Multimodality imaging of fluid saturation and chemical transport for two-phase surfactant/polymer floods in porous rocks," *Transp. Porous Media*, vol. 152, 1 Jan. 2025. DOI: 10.1007/s11242-024-02146-0. [Online]. Available: http://dx.doi.org/10.1007/s11242-024-02146-0.

[177] R. L. Schaufler and N. C. Slowey, "Low-cost modular chromatography column rack and vial holders," *HardwareX*, vol. 13, e00388, e00388 Mar. 2023. DOI: 10.1016/j.ohx.2022.e00388. [Online]. Available: http://dx.doi.org/10.1016/j.ohx.2022.e00388.

[178] P. Diep, J. L. Cadavid, A. F. Yakunin, A. P. McGuigan, and R. Mahadevan, "REVOLVER: A low-cost automated protein purifier based on parallel preparative gravity column workflows," *HardwareX*, vol. 11, e00291, e00291 Apr. 2022. DOI: 10.1016/j.ohx.2022.e00291. [Online]. Available: http://dx.doi.org/10.1016/j.ohx.2022.e00291.

[179] W. J. Crandall, M. Caputo, L. Marquez, Z. R. Jarrell, and C. L. Quave, "Customizable large-scale HPLC fraction collection using low-cost 3D printing," *HardwareX*, vol. 21, e00612, e00612 Mar. 2025. DOI: 10.1016/j.ohx.2024.e00612. [Online]. Available: http://dx.doi.org/10.1016/j.ohx.2024.e00612.

[180] M. C. Carvalho and J. M. Oakes, "Turning a 3D printer into a HPLC fraction collector: A tool for compound-specific stable isotope measurements," *Hardware*, vol. 1, pp. 29–53, 1 Dec. 7, 2023. DOI: 10.3390/hardware1010004. [Online]. Available: http://dx.doi.org/10.3390/hardware1010004.

[181] S. B. Ficarro, W. Max Alexander, I. Tavares, and J. A. Marto, "Open source fraction collector/MALDI spotter for proteomics," *HardwareX*, vol. 11, e00305, e00305 Apr. 2022. DOI: 10.1016/j.ohx.2022.e00305. [Online]. Available: http://dx.doi.org/10.1016/j.ohx.2022.e00305.

[182] E. Z. Macosko, A. Basu, R. Satija, *et al.*, "Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets," en, *Cell*, vol. 161, pp. 1202–1214, 5 May 21, 2015. DOI: 10.1016/j.cell.2015.05.002.

INDEX

POCKET MATERIAL: SURVEY FOR COMPARATIVE ANALYSIS OF SALIVA COLLECTION DEVICES

Below is the survey used in the comparative analysis of saliva collection devices, as presented in Chapter 2.

1

Questionnaire

Answer all of the questions below for each device that you use.

1. How many times have you used this device before?

⬜ More than 3 times ⬜ 1-3 times ⬜ Never

2. How difficult were the user instructions for this device?

Very difficult Difficult Fair Easy Very easy

Opinions/comments (optional):

_____
_____
_____

3. How difficult was it to assemble the device components?

Very difficult Difficult Fair Easy Very easy

Opinions/comments (optional):

_____
_____
_____

4. How difficult was it to collect saliva with this device?

Very difficult Difficult Fair Easy Very easy

Opinions/comments (optional):

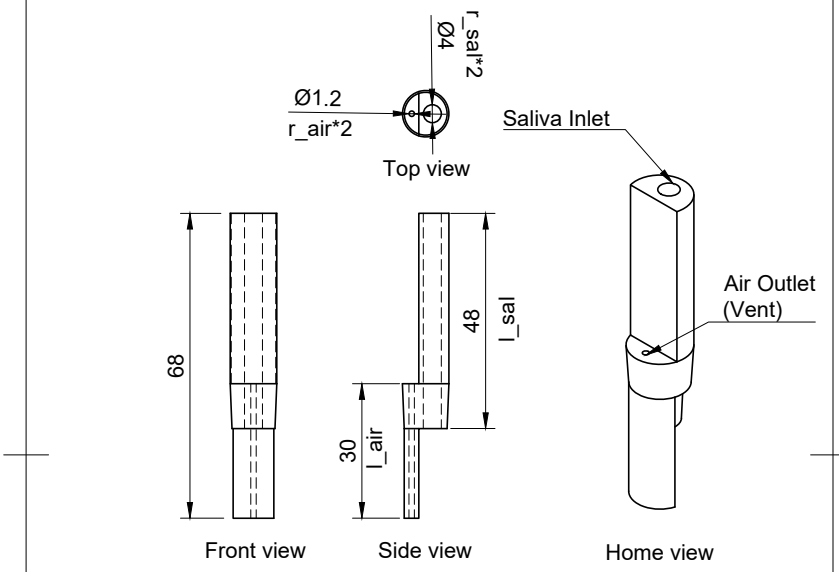
_____
_____
_____

5. Did saliva leak out of the device or get on your hands or clothes?

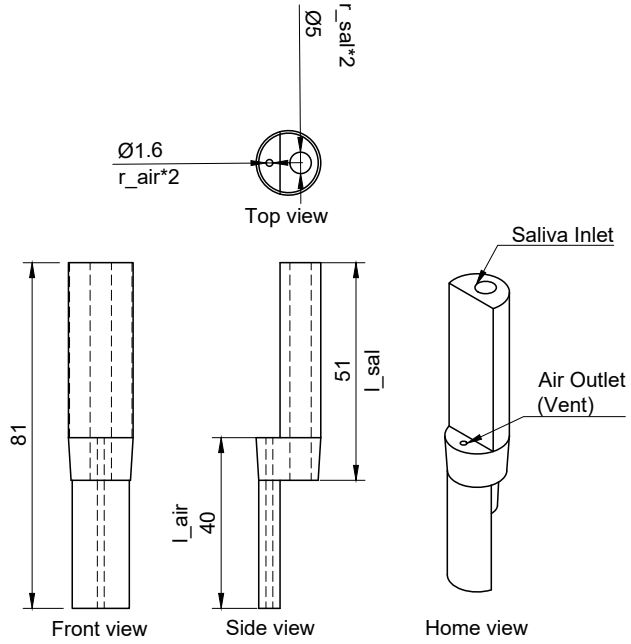
⬜ Yes ⬜ No

POCKET MATERIAL: CAD SCHEMATICS OF SALIVASTRAW

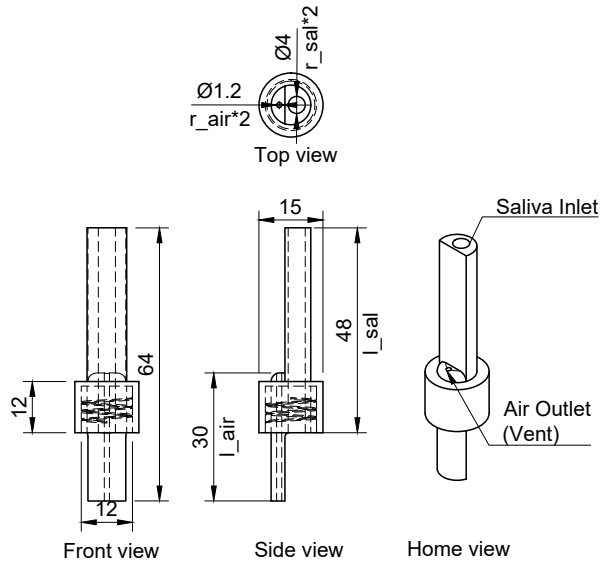
Below are the CAD schematics of all three models of SalivaStraw, as presented in Chapter 2 of this dissertation.



Dept.	Technical reference	Created by Yeokyoung Kil 3/25/2025	Approved by
		Document type	Document status
		Title: SalivaStraw 2mL model Tapered, press-fit	DWG No.
			Rev. / Date of issue / Sheet 1/1



Dept.	Technical reference	Created by Yeokyoung Kil 3/25/2025	Approved by	
		Document type	Document status	
		Title SalivaStraw 5-mL model Tapered, press-fit	DWG No.	
			Rev. / Date of issue	Sheet 1/1



Dept.	Technical reference	Created by Yeokyoung Kil 3/25/2025	Approved by	
		Document type	Document status	
		Title SalivaStraw 2-mL model Threaded Lid	DWG No.	
			Rev. / Date of issue	Sheet 1/1