



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Ph.D. Dissertation

**Spatially-resolved Laser Activated Cell
Sorting for Genomics and Transcriptomics
in Biological Specimens**

레이저 활성 세포 분리 기기를 이용한 조직 내 세포
분리 및 전장 유전체 및 전사체 분석 기술 개발

AUGUST 2020

Graduate School

Seoul National University

INTERDISCIPLINARY PROGRAM IN BIOENGINEERING

CHUNGWON LEE

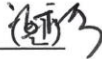
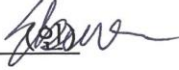



Spatially-resolved Laser Activated Cell Sorting for Genomics and Transcriptomics in Biological Specimens

지도 교수 권 성 훈

이 논문을 공학박사 학위논문으로 제출함
2020 년 6 월

서울대학교 대학원
협동과정 바이오엔지니어링
이 충 원

이충원의 공학박사 학위논문을 인준함
2020 년 6 월

위 원 장	한 원 식	(인) 
부위원장	권 성 훈	
위 원	박 응 양	(인) 
위 원	이 태 걸	(인) 
위 원	정 연 철	(인) 

Abstract

In this dissertation, Spatially-resolved Laser Activated Cell Sorting (SLACS) technique is introduced, and its applications in genomics and transcriptomics are demonstrated. All biological mass is comprised of biological cells, each of which contain its own multi-billion bytes worth of data from genetic molecules, such as DNA or RNA. After the Human Genome Project sequenced one person's genome in ten years, the massively parallel sequencing technologies that are referred to the next generation sequencing (NGS) sprouted innovations in biology, providing further insights into biology and generating revolutions in diagnostics and therapeutics. However, these technologies were only applicable to pools of heterogeneous genetic molecules, hindering thorough explorations of genetic landscapes in the different cells within a biospecimen. Therefore, efforts to separate each and every cell from the pool of cells have generated numerous single cell isolation methodologies, which can be categorized into three: those that separate cell using microfluidics, microarrays, and optics.

Advancement in micro-technologies particularly provided advantages in manipulating single cells because biological cell sizes that usually range from microns to tens of microns. State-of-art cell separation technologies that utilize

microfluidic properties were rapidly commercialized, enabling high throughput single cell analysis that can process hundreds to thousands of single cells at a time. These utilize cell dissociation and compartmentalization in a microfluidic chambers or a pico-liter droplets, in which biomolecular techniques can amplify the desired genetic molecules. The amplified products such as the genomes or the transcriptomes of the single cells are sequenced through NGS, providing insights into how the dissociated cells were functioning in the biospecimen. However, the dissociation process of the cells that are originally adhered to each other can be harsh and requires the surface proteins that interact with another to be degraded. This process has raised many doubts on whether the cell state is the same before it is dissociated within a solvent. Therefore, microarrays of chemically synthesized oligonucleotides that can capture the poly adenosine tail, or poly (A) tail, were developed to capture the messenger RNAs (mRNAs) directly from the biological specimens. These technologies, however, require large resolution of the oligonucleotide spots because of the technical limitations in chemical DNA synthesis technologies and cross-contaminations between the spots.

Optical separation of the cells from biospecimen has been extensively investigated with conventional laser capture microdissection (LCM) devices that utilize laser to transfer target area of interest to the desired receiver.

However, these utilize either ultraviolet (UV) lasers to catapult the desired areas that can be highly damaging to the biomolecules within, or thermoplastics that can be melt down using near-infrared (IR) lasers and transfer the desired region of interest for further biological assays. However the thermoplastic approach often cause cross-contamination and has low throughput because the specimen has to be isolated in a contact manner.

In this dissertation, the development of an optical cell sorter, or spatially-resolved laser activated cell sorter (SLACS) that uses pulsed near-IR laser that can optomechanically isolate the cells with low damage and high throughput is described. The engineering process of this novel device and two softwares and their applications in NGS technologies are described. Furthermore, the applications of SLACS for genomics and transcriptomics are demonstrated. Proof-of-concept studies for future applications of SLACS are also described.

Keywords: laser, cell sorter, spatial omics, spatial genomics, spatial transcriptomics

Student Number: 2014-21548

Table of Contents

ABSTRACT	I
TABLE OF CONTENTS	IV
LIST OF TABLES	VII
LIST OF FIGURES	VIII
CHAPTER 1. INTRODUCTION	1
1.1. Spatially resolved omics for atlasing human cells in the biological circuitry	2
1.1.1. The emergence of single cell sequencing technologies	3
1.1.2. Spatially resolved omics technologies and needs for development	7
1.2. Main Concept: Development of spatially-resolved laser activated cell sorter (SLACS) and compatible omics technologies	1 4
1.3. Outline of the dissertation.....	1 5
CHAPTER 2. BACKGROUND.....	1 6
2.1. Previous spatial omics technologies.....	1 7
2.1.1. <i>In situ</i> spatial omics technologies	1 7
2.1.2. Isolate-and-transfer technologies for spatial omics	2 0

2.2.	Commercialized spatial omics technologies.....	2 3
2.3.	Previous research in the group	2 5
CHAPTER 3. PLATFORM DEVELOPMENT		2 9
3.1.	Development of SLACS and remote selection system	3 0
3.2.	Whole genome sequencing strategies for SLACS	3 3
3.3.	Whole transcriptome sequencing strategies for SLACS	4 2
CHAPTER 4. PLATFORM APPLICATION		4 8
4.1.	Applications of SLACS to spatial genomics	4 9
4.2.	Applications of SLACS to spatial transcriptomics	6 2
4.3.	Applications of OPENchip and future perspectives with SLACS	6 5
CHAPTER 5. CONCLUSION AND DISCUSSION		7 9
5.1.	Summary of dissertation	8 0
5.2.	Comparison with previous technology	8 3
5.3.	Limit of the platform.....	8 4
5.4.	Future work	8 6

BIBLIOGRAPHY.....	8 8
--------------------------	------------

국문 초록	9 5
--------------------	------------

List of Tables

Table 1.1 Comparison of the characteristics of different single cell sequencing technologies.	1 3
Table 3.1 PCR validation primer sets for MDA products [31].	3 8

List of Figures

Figure 1.1 Biological cells comprise biological tissues that comprise the organs.....	4
Figure 1.2 The central dogma describes various parameters for profiling the cells in a biological system.	4
Figure 1.3 Microfluidic devices that are used for single cell sequencing applications [8].	6
Figure 1.4 Conventional workflow of library preparations for spatial omics technologies that uses biochemical isolations. Bluee fragment indicates universal primer' green and red, molecular identifiers; dark grey, poly A tail; Light blue, mRNA; yellow and green circles, transposase Tn5; and purple and yellow Illumina barcodes.....	1 1
Figure 2.1 Comparison chart for major spatial omics platforms.	2 3
Figure 2.2 Prototype built for Sniper Cloning [64].....	2 7
Figure 2.3 Concept of Sniper Cloning [64]	2 8
Figure 3.1 Platforms described in this dissertation.	3 0
Figure 3.2 The design and a picture of the SLACS system.	3 1
Figure 3.3 The closed-up picture of the SLACS system (Left) and the user interface built for the SLACS system (Right).....	3 1
Figure 3.4 The remote system for selecting desired region of the biospecimen.	3 2
Figure 3.5 PHLI-seq uses SLACS to sequence the whole genomes of cells [33]. (a) Serial sections from a tumor is prepared with different staining modalities. (b) The SLACS device isolates the cells in high throughput. (c) The genome within are amplified with whole genome amplification and goes through NGS. (d) Cell isolation images.....	3 5
Figure 3.6 Copy number alteration (CNA) plots of PHLI-seq cell line samples with MDA and MALBAC.	3 6

Figure 3.7 Quality control experiments for different parameters for PHLI-seq [33]. (a,b) Relation between the whole genome amplification start time and copy number ratio. (c,d,e) Correlation between amplification start time and area under the Lorenz curve. (f) CNAs for different samples (g) fraction of bins vs. fraction of reads for different samples.	4 0
Figure 3.8 Platform comparison to laser pulse catapulting (LPC) and laser microdissection (LMD).	4 1
Figure 3.9 Correlation between the barcoded mRNA sequencing of single cell material (~20pg) and bulk mRNA sequencing data.	4 3
Figure 3.10 Gene counts according to different cell types and different cell preparation methods.	4 5
Figure 3.11 Heatmap of the RNA expressions from different samples...	4 6
Figure 3.12 Transcript isoform diversity between the two different samples in gene STX3.	4 7
Figure 4.1 SLACS for spatial genomics applied to HER2+ breast cancer tissue.	5 1
Figure 4.2 3D map of a tumor tissue using spatial genomics platform with SLACS.	5 6
Figure 4.3 Follow-up study for PHLI-seq in breast cancer tissue sections. The CNA plots are accompanied by the corresponding tissue section slide in the left.	5 8
Figure 4.4 Microstructures that capture single CTC are selectively isolated with SLACS [31].	5 9
Figure 4.5 CTCs captured from a breast cancer patient show heterogeneity in the CNA plot drawn in circos format.	6 0
Figure 4.6 Cells from the two distinct layers from a brain organoid section with SLACS went through SmartSeq2.	6 3
Figure 4.7 Highly expressed genes from immune cells isolated from breast cancer tissue section.	6 4
Figure 4.8 Rare CTCs are enriched with a biochip and used for DNA and RNA	

profiling <i>in situ</i> [44].	6 9
Figure 4.9 The OPENchip captures CTCs with high specificity and capture efficiency [44].	7 1
Figure 4.10 The padlock probes for <i>in situ</i> profiling are validated with different cell lines [44].	7 2
Figure 4.11 The cell lines are run through OPENchip for spike-in tests [44].	7 3
Figure 4.12 OPENchip data with cancer patient samples [44].	7 5
Figure 5.1 Compared to other technologies, SLACS is able to produce data from hundreds of single cells at once, producing less artifacts. Also, it is not confined to RNA analysis.	8 3
Figure 5.2 The development timeline for SLACS platform.	8 5

Chapter 1. Introduction

In this chapter, spatially resolved omics for atlasing human cells is described. From the needs for the human cell atlas to the addressing technologies, the contents of this chapter will provide introduction of this dissertation. Specifically, the emergence of the next generation sequencing (NGS) technologies and single cell sequencing technologies that stemmed out from the NGS technologies is described. Then, advanced technologies that can analyze the cells' genetic molecules with spatial context are described. Building on these state-of-art technologies, the dissertation will introduce a novel platform named spatially-resolved laser activated cell sorting (SLACS) technology. With SLACS technology, biological problems will be solved by connecting the spatial assays to the molecular assays.

1.1. Spatially resolved omics for atlasing human cells in the biological circuitry

Human is comprised of 40 trillion biological cells. Each and every cell has different functions that altogether enable biological processes that take place in a human's body. The human genome project funded by the US government through the National Institutes of Health (NIH) was initiated in 1990. It took 13 years to sequence one person's genome, providing clues to how human is encoded [1]. The impact of the human genome project, however, was not a mere scientific discovery, but a gigantic disruption in industry, science, and medicine. The emergence of the next generation sequencing (NGS) technologies is perhaps the best example [2]. NGS was enabled with development and convergence of technologies from diverse fields. Optics, biochemistry, semiconductor technologies, and computer science that were developed during the human genome project became the base for the birth of NGS. Now that the development speed of NGS surpassed the Moore's law, NGS is able to sequence a genome of one person in a day with 100 USD. With this innovation, NGS is used in enormously diverse areas such as diagnostics, medicine, environment monitoring, and even entertainment.

After the human genome project, other projects like the 1,000 genome project sprung out. With the analysis of genomes of people. One of the projects was a researcher-funded project called, the human cell atlas project [3]. The human

cell atlas project aims to profile the different cells in a human's body. Because each and every cell has different functions, understanding what cells exist in a human's body will ultimately lead to the comprehension of what cells comprise which organs. This investigation of the building blocks of the human body has already brought insights into how to cure diseases and how to enhance body functions. With this call for the human cell atlas, single cell sequencing technologies have emerged, further innovating biology and biotechnology.

1.1.1. The emergence of single cell sequencing technologies

Understanding the human body has been one of the greatest enigma that the humankind has sought to solve. Humans have pondered on what they were comprised of and how one is born and deceased in a biological sense is one of the basic and greatest philosophical question. To understand how the biological systems are engineered, we first need to understand how the system is comprised of. Biological cells are defined to be the basic unit that comprises a biological system. With cellular interactions, bigger units such as tissues are able to function in multiple ways to comprise even bigger units such as the organs (Figure 1.1). Therefore, a call has been made to categorize the cells in a biological system according to biologically significant parameters [3]. The human cell atlas that was organized by multi-disciplinary researchers around the globe has produced magnificent number of scientific discoveries, which are

expected to sprout more discoveries when analyzed deeper and with more biospecimens.

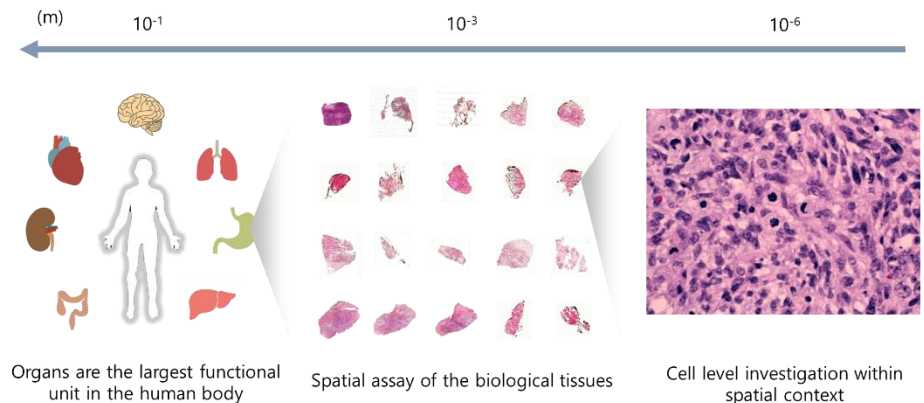


Figure 1.1 Biological cells comprise biological tissues that comprise the organs

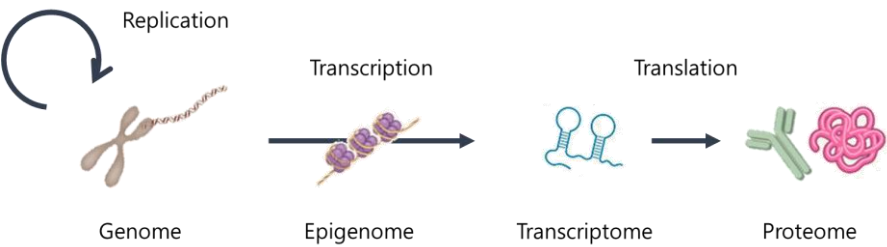


Figure 1.2 The central dogma describes various parameters for profiling the cells in a biological system.

Then the question comes to: what parameters could be the biologically significant parameters? Each and every functioning cell follows the central dogma that dictates the direction of genomic transcription and transcriptomic

translation (Figure 1.2). If the cell is encoded with genome, the cell machinery transcribes the gene according to epigenetic or epitranscriptomic signatures to generate RNAs. Among many kinds of RNAs, messenger RNAs (mRNAs) are able to be translated into amino acid based proteins, that are functionally active inside and outside the cells. Therefore, studying the biomolecules that are related to the central dogma can be crucial in order to discover and categorize different cell types in a biological system. The best example is the fluorescence activated cell sorter (FACS) that uses fluorescent taggers on the desired cells to sort out cells by changing their paths using laser cytometry [4]. Using FACS, researchers have sequenced the genomes and transcriptomes of single cells revealing heterogeneity of these cells in biological masses like tumors or normal and diseased organs [5], [6]. While these provided very powerful methodologies to explore biology in single cell level, the high cost per sample barred extensive single cell studies.

With the milestone paper by Macosko et al., the high-throughput single cell compartmentalization using nanoliter droplets, in which barcoded mRNA capturing molecules are trapped with single cell, opened a vast opportunity in typing thousands of single cells at once with affordable costs [7]. Because mRNAs indicate the functional state of the single cells, typing the barcoded mRNAs enabled principle component analysis (PCA) of the different cells at once. Likewise, using the properties of fluids in micro- or nano-scale, several platforms became available to meet the needs of single cell typing [8] (Figure

1.3). Using the fluidic properties in micro- and nano-scale, the single cell compartmentalization technologies applied microfluidic valve, array of microwells, and droplets to compartmentalize single cells. These platforms were applied to various applications including oncology, microbiology, neurology, development science, genetic health, agrigenomics, and more. These technologies were developed and commercialized by several companies like Fluidigm, 10X genomics, and Wafergen and are widely being used in scientific discoveries and clinical uses. With advances in engineering solutions to overcome the current limitations of these platforms, these platforms provide

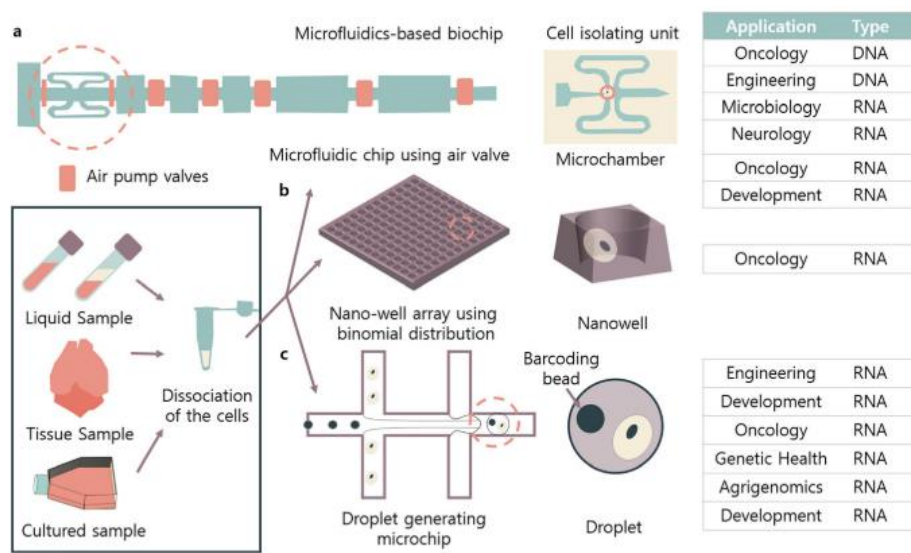


Figure 1.3 Microfluidic devices that are used for single cell sequencing applications [8].

great potential for further impact in the field of single cell biology.

1.1.2. Spatially resolved omics technologies and needs for development

Now that the researchers have the tools to analyze, type, and categorize the single cells in a given biological mass, massive data regarding the genetic biomolecules within thousands of single cells is increasingly accumulated, and these data is being shared by researchers via the human cell atlas [3]. The next step towards building a map for the human body is to analyze the cells within the spatial context. This is because knowing what cells exist is not enough to address where the cells are located. If the same cell type is in different functioning locations within a biological system, understanding the spatial context is critical to upgrade our understanding of the biological mass. Therefore, although they still are in the early stage, spatial omics technologies have been developed to address these needs.

Tagging the desired molecules *in situ* has provided a strong tool for connecting the molecular feature to the spatial information. Since the development of fluorescence *in situ* hybridization (FISH) technologies, various forms of *in situ* molecular analysis technologies have sprung out. The basis of FISH technologies lies in that fluorescence tagged molecules can be hybridized on to the genome or transcriptome that are fixed within a cell. From single molecule FISH (smFISH) [9] to Brainbow [10], the strategy to tag regions of the genome

or the transcriptome have provided tools for the connection between molecular biology and spatial assays. There are various types of *in situ* assays, such as SABER [11], seqFISH [12], seqFISH+ [13], FISSEQ [14], MERFISH [15] and *in situ* sequencing methods [16]. These all utilize *in situ* hybridization methods of oligonucleotide-based barcodes that are complementary to sequence of interest. Although these methods provide a powerful tool to investigate the gene expression or gene mutation within spatial context *in situ*, it often requires high resolution optical instruments and are only able to type a few combinations of gene sets. Therefore, these methods are often supported with NGS-based omics technologies that are more high throughput and less expensive.

While the high throughput single cell sequencing technologies provided a sharp breakthrough in biology and biotechnology, all microfluidic platforms require the cells to be dissociated into a solution before analyzing them. This dissociation step causes the cells that exist in a biological mass to be detached from one to another, with slight treatment of proteases. This step causes the loss of spatial information and requires additional methodologies to reconstruct the spatial information of the analyzed cells [17]. A method to overcome this limitation is spatially-resolved techniques that utilizes immobilized molecular barcodes with poly thymine (poly-T) sequences that can capture the poly adenylated (poly-A) tail of the mRNAs [18]–[20]. These technologies utilize microarray of synthesized single stranded oligonucleotides that work as baits for the mRNA molecules within the cells. If a tissue section is prepared on top

of the microarray, as the mRNA molecules from the cells are released with cell lysis, the molecules are captured on the baits. Spatial Transcriptomics (ST) technology from Lundeberg's group from Science for Life Laboratory (SciLifeLab) invented and commercialized this product, which became the basis for the spatial omics technologies that utilize biochemical capture of the mRNA molecules within spatial context. However, because of the limitations of the spot size of the oligonucleotide synthesis, the spatial resolution is approximately 100 μm in length and width, which will contain approximately 100 cells in unit space. Therefore, high density spatial transcriptomics (HDST) was developed to increase the spatial resolution of this platform [19]. HDST platform utilizes the beads in which the mRNA baits are immobilized, just like the beads that the droplet-based platforms use. After the beads are assembled in a 10 μm by 10 μm area, the spatial barcodes that are on the beads are read by sequencing by synthesis, usually with Illumina NGS sequencers. Here, it is important to note that a highly specific chemistry is used to capture the mRNAs within the cells in spatial context. The procedure is a five-step process: capture of the mRNA molecules, reverse transcription, template switching, polymerase chain reaction (PCR), transposase Tn5 tagmentation and PCR [21] (Figure 1.5). After the mRNA molecules are captured, reverse transcription creates complementary DNA (cDNA) and leaves GGG sequence in the 3' end of the cDNA. By priming to the GGG sequence, the designed universal primer in PCR reaction generates double stranded DNA-RNA hybrid molecules that contains

the universal primer region, cell barcode region, unique molecule identifier, poly T, cDNA region, and universal primer. These molecules are usually 1 to 2 kilobases (kb) long, depending on the length of the original mRNAs. Then these molecules are amplified using PCR, amplifying the whole transcriptome. These amplicons then go through transposase tagmentation. Tn5 transposases are originally seen in bacterial strains that function as gene shufflers. These dimer proteins are able to allocate double stranded nucleic acid fragments into the target regions by cutting the target and inserting the fragments that the proteins are assembled with [22]. Using this property, the long double stranded DNA library can be cut and ligated with Illumina adapters so that the NGS library is ready to be sequenced in a single pool. Because the libraries from different regions of the HDST platform are pooled, only the fragments that contain the cell barcodes are amplified. Tn5 transposases usually tagment approximately 600 bp molecules, the rest sequences are abandoned during the final PCR process. Using the 50 bp sequence, the gene is identified and is applied for gene expression counting. This is very cost-efficient process, but in turn, it loses the full transcriptome data. To illustrate, it is impossible to type the splicing variants of the mRNAs. Also, other genetic molecules like the genome or the proteome cannot be analyzed using this method. In this manner, as Fluidigm puts it, there is a renaissance in cell isolation methodologies that use lasers.

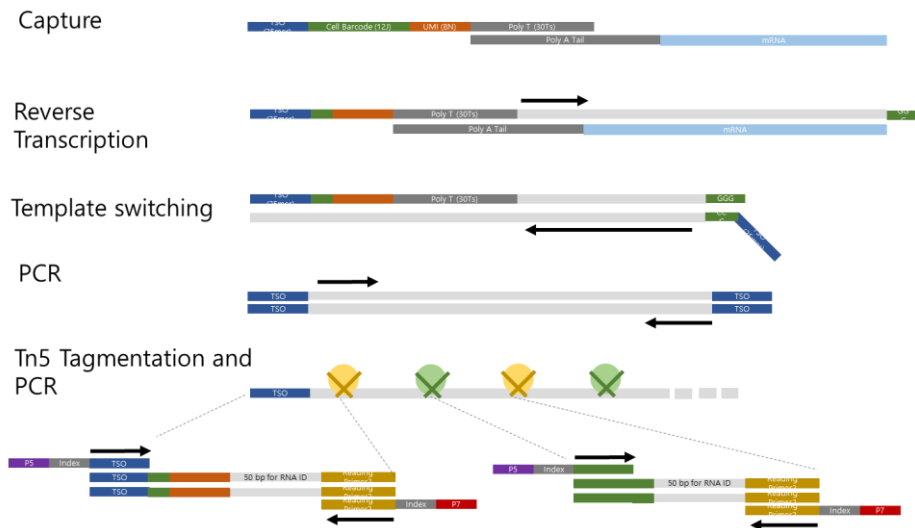


Figure 1.4 Conventional workflow of library preparations for spatial omics technologies that uses biochemical isolations. Blue fragment indicates universal primer; green and red, molecular identifiers; dark grey, poly A tail; Light blue, mRNA; yellow and green circles, transposase Tn5; and purple and yellow Illumina barcodes.

Cell isolation technologies that use optical isolation (i.e. lasers) have been around this field for a long time. Since its invention, laser capture microdissection (LCM) technologies were commercialized by many optics companies such as Leica, Zeiss, and Fluidigm. The basic idea of these technologies is to use laser to dissect out cells from the biospecimen. Utilizing the laser induced forward transfer (LIFT) technology used in semiconductors, the cells instead of microstructures are transferred to a desired retriever. There

are two types of these devices: ones that use ultraviolet (UV) lasers and that use near-infrared (IR) lasers. For the LCM that uses UV, the high energy UV laser burns the periphery of the target area, enabling the microdissection. Then, a weaker UV laser is applied to the center of the target to catapult out the target. The target is retrieved in a PCR tube cap, allowing further downstream analysis. The problem of using UV light in these methods is that the short wavelength covalently links the thymine moieties, generating thymine dimers. This chemical reaction further induces fragmentation and amplification inhibition. Therefore, UV LCMs are very damaging to the biospecimen. Another LCM methodology is the ones that use near-IR lasers. Using the high thermal energy generated by the near-IR lasers, these methodologies require thermoplastics that are melted when the near-IR lasers are applied. If the target region of the thermoplastic is melted and adhered to the target region in the biospecimen, the thermoplastic is soaked into buffers that allow downstream analysis. Compared to the LCM that use UV, the LCM methodologies that use near-IR lasers are less damaging to the biospecimen. However, the contact-based isolation method in near-IR laser LCMs is tremendously low-throughput and is prone to cross-contaminations. To improve this method Fluidigm recently announced its new system named Acculift. The platform uses IR lasers and gecko-feet adhesives to increase the efficiency compared to conventional LCMs. However, the fundamental issue of using IR lasers and thermos-adhesives still exists.

Table 1.1 Comparison of the characteristics of different single cell sequencing technologies.

Cell isolation	Fluidic	Biochemical	Optical
Spatial resolution	Loss of spatial information during cell dissociation process	Preserves the spatial information	Preserves the spatial information
Quality of biomolecules	Dissociation process affects the single cell to be detached from one to another	Diffusion in between different areas with capturing molecules	The ultraviolet (UV) light generates thymine dimers and fragments the DNA. Contact methods produce cross-contamination.
Possible molecular features	High freedom to operate	Highly specific chemistry using poly T sequence to capture mRNAs.	High freedom to operate

Nonetheless, LCMs have widely been used for analyzing the target of interest in terms of genomes and transcriptomes within spatial context [23]–[29]. Therefore, all cell isolation methods used in omics technologies have their pros and cons that are compared in Table 1.1. In regards to spatial omics technologies, there is a need for an improved device that is capable of producing high quality of biomolecules and high freedom to operate. In order to complete the human cell atlas and further dive into human pathology atlas, new spatial omics technologies must be developed.

1.2. Main Concept: Development of spatially-resolved laser activated cell sorter (SLACS) and compatible omics technologies

In this dissertation, I propose a novel technological solution in the spatial omics field. The main concept of the proposed technology that can address the aforementioned technological limitations is to use near-IR lasers and non-contact method using a special sacrificial layer [30]–[33]. In this dissertation, this novel technology, named spatially-resolved laser activated cell sorter (SLACS), will be introduced. I address previous works from the previous dissertations from the research group and describe some of the development processes and applications of SLACS. Using nanosecond pulsed laser with wavelength of 1064 nm, the desired targets are isolated in non-contact manner.

Like the conventional LCMs, SLACS conserves the spatial information as the retriever for the targets is compartmentalized. Also, it has high freedom to operate like other LCMs because the retriever is simply a PCR tube cap. Furthermore, the near-IR laser has low or no damage to the cells and the non-contact method makes the method high-throughput and less prone to cross-contaminations. With the technologically advanced device, I describe its application to DNA sequencing and RNA sequencing. With the molecular demonstrations, I also show *in situ* assay that can be complimentary to SLACS in near future.

1.3. Outline of the dissertation

In this dissertation, a new concept of SLACS is described. In Chapter 2, the previous works in SLACS are described and compared in an aspect of technical background and application. In Chapter 3, development and optimization process of the platform is described. In Chapter 4, as a proof of , several demonstrations of SLACS in DNA sequencing and RNA sequencing and description of *in situ* technology are described. Finally, in Chapter 5, discussions and future directions of the SLACS platform are provided. The summary of the dissertation, the limit of the platform and future work are described.

Chapter 2. Background

In this chapter, previous spatial omics technologies and previous works that became the basis for the development of SLACS are described. Specifically, *in situ* spatial technologies and isolate-and-transfer technologies for spatial omics are described. Then, commercialized spatial omics technologies are compared to illustrate the state-of-art devices for spatial omics. After describing the pros and cons of the conventional spatial omics technologies, laser transfer technologies using pulsed-laser developed previously in the research group are described.

2.1. Previous spatial omics technologies

As described in Chapter 1, spatial omics technologies are being introduced for the post-single cell omics era. To locate the typed cells within the tissue context, several different and unique approaches have been introduced. In this section of the chapter, these technologies are categorized into two: *in situ* spatial omics and isolate-and-transfer technologies. *In situ* spatial omics does not require any transferring steps of the biomolecules, while the isolate-and-transfer technologies utilize one or more methods to transfer the cells or the biomolecules inside the cells.

2.1.1. *In situ* spatial omics technologies

As discussed in Chapter 1, there are several *in situ* spatial omics technologies, such as FISH techniques, SABER [11], seqFISH [12], seqFISH+ [13], FISSEQ [14], MERFISH [15] and *in situ* sequencing methods [16]. All of these methods utilize *in situ* hybridization (ISH) of certain probes onto the genome or transcriptome, allowing further downstream process for multiplex and sensitive methods to profile omics information within spatial context. Since its development in 1969, ISH-based techniques have been widely used for analyzing the nucleic acid contents within the cells [34]. Although they still are very powerful tools to some applications, the early ISH technologies relied on four to five fluorescence channels, enabling only a few probe regions in the

entire genome or transcriptome. Therefore, for extensive analysis of the entire genome or transcriptome, technologies that utilize the programmability of the ISH probes were developed. One of the earliest examples that advanced the ISH technologies is sequential FISH (seqFISH) and seqFISH+ [12], [35], [36]. The seqFISH technologies utilize the barcodes that are attached to the ISH probes that target certain regions in the genome and transcriptome. By repeating hybridization of fluorescent probes to the barcode regions and detachment of them with DNases, the images created in each cycle are merged to produce different colors. SeqFISH+ is an advanced version of seqFISH and utilizes four non-readout “slots” of the barcode regions. These additional slots require less cycles for the same number of combinations of transcripts, and produces indications of 4^n combinations of transcripts, where n equals to the number of cycles. However, the barcoded probes and the fluorescent probes have to be chemically synthesized, making the process more expensive and time consuming compared to other ISH technologies. MERFISH (multiplexed error-robust FISH) is a similar technique that utilizes two flanking regions to the target-hybridized probes [15], [37], [38]. MERFISH adds a computational approach to sort out non-specific hybridizations for error-corrected reads. Also, with integrations to expansion microscopy, MERFISH has increased its sensitivity of reading out overlapping signals inside the cell.

Although *in situ* hybridization methods have been developed to be useful tools with high multiplexity and sensitivity, the high cost of oligonucleotide

synthesis of various hybridizing probes and time consuming procedures barred their extensive uses in conventional laboratories. Therefore, technologies that utilize the advances in sequencing technologies were introduced. *In situ* sequencing methods, which were introduced by Mats Nilsson's group in Stockholm University [16], brought about many applications in diverse groups and commercialization [39]–[46]. *In situ* sequencing methods utilize padlock probes, which are padlock shaped probes that can hybridize onto a specific target region [47]. Using phi29 polymerases that can amplify the DNA with hyperbranching and strand displacement, the circular template formed by the padlock probe is amplified. This process is called rolling circle amplification (RCA), and produces nano- blobs of amplified DNA, or rolling circle amplification products (RCPs). The RCPs are then sequenced by sequencing by ligation (SBL) methodology, which sequences each nucleotide by ligating sequencing probes on the target. These were used in a variety of spatial omics platforms [39]–[42], [45], [46], [48]–[51], including one that is discussed in this dissertation [44]. A similar approach was invented by George Church's group from Harvard University that utilizes sequencing by synthesis (SBS) methodology, instead of SBL [52]. BaristaSeq utilizes the same chemistry of padlock probe hybridization and RCA. The difference comes to the sequencing method applied to the RCPs that sequences each nucleotide with nucleotide synthesis. Although SBS provides highly efficient sequencing methods in terms of high signal-to-noise ratio, BaristaSeq was only demonstrated in cultured

cells. Similarly, there are other sequencing by ligation methods applied to *in situ* sequencing, such as STARmap (Spatially resolved Transcript Amplicon Readout Mapping) [53] and FISSEQ (fluorescent in situ RNA sequencing) [54]. These utilize RCA on the template and use SBL technology to sequence a few bases.

That the shape of the tissue is maintained during and after these assays serves as a great advantage in analyzing the biomolecules in spatial context. Altogether, *in situ* technologies can analyze the genome and the transcriptome in subcellular to tissue level. There are many advanced optics devices that can address some of the difficult issues in imaging biomolecules using the *in situ* technologies. For example, high resolution imaging of molecules within 3 dimensional tissue context can be made possible by integrating the state-of-art molecular techniques and optics techniques.

2.1.2. Isolate-and-transfer technologies for spatial omics

Although *in situ* technologies provide great tools to examine biospecimens within tissue context, throughput, labor-intensiveness, and the depth of the molecular data can be low than the isolate-and-transfer technologies. The isolate-and-transfer technologies use additional method to isolate the molecules from the tissue, instead of typing them *in situ*. The best example is the spatial transcriptomics technology, invented by Joakim Lundeberg's group [20]. The spatial transcriptomics technology utilizes a

microarray of oligonucleotides with spatial barcodes and poly-T sequences. These spots of oligonucleotides are in circular shapes with 100µm diameter, located 200 µm apart. On top of the microarray, the tissue to be examined is prepared. Because the substrate for the microarray is transparent glass, the tissue can be stained and imaged. This provides a great advantage over the ISH technologies because histopathological information can be added onto the spatial context. After the tissue is imaged, the cells are lysed and the mRNAs with poly A tails are hybridized onto the poly T barcodes nearby. Then, similar to droplet-based single cell typing methods, the hybridized molecules go through reverse transcription, template switching, PCR, and fragmentation for NGS library preparation. Similar methods were further reported using beads with 2 µm (HDST) [19] and 10 µm (Slide-seq) [55]. Slide-seq utilizes tissue placed on top of monolayered packed beads on coverslip. The packed beads are attached with the same barcodes that are used in spatial transcriptomics technology. Instead of using the microarray, Slide-seq uses these 10 µm sized beads to type the whole tissue section. However, before the mRNAs can be captured on to the beads, the spatial barcodes on the beads have to be read by *in situ* sequencing technologies, so that the spatial barcodes of the randomly packed beads can be recovered after the captured mRNAs are pooled. HDST is a similar technology with a higher throughput. The beads are assembled on the microwells, on which the tissue section is prepared. After the assembly process, the beads here also have to be sequenced *in situ* for reading the spatial barcodes.

Then, the tissue can be applied for subsequent processes like cell lysis and mRNA capture can take place. These are very powerful tools to type the tissue section by isolating the spatial barcode tagged mRNAs and transferring them into one solution. With these technologies the transcriptome can be sequenced with NGS sequencers. However, the captured gene counts are usually low and the chemistry is highly specific, hindering other omics analysis such as genomics, BCR and TCR sequencing et cetera.

Laser capture microdissection (LCM) [56] technologies have been around for quite a long time compared to other isolate-and-transfer technologies. Despite they either use nucleic acid damaging UV lasers or labor intensive IR lasers, LCM technologies are experiencing a renaissance for their freedom to operate and spatially resolved characteristics. While the cutting-edge isolate-and-transfer spatial transcriptomics technologies addressed above are highly specific to mRNA typing analysis, LCM technologies dissects out region of interest in single cell level that makes the user to choose the downstream assays ranging from mass spectrometry to NGS applications. LCM technologies can be categorized into two types: ones that use UV laser and ones that use IR laser with thermoplastics. The former type uses UV laser to burn the periphery of the region of interest. Then UV laser with lower intensity is applied to the center of the region of interest, catapulting the microdissected area to a PCR tube cap. The latter type uses IR laser to melt down the thermoplastic that is in contact with the tissue. The melted region works as glue to transfer the region of interest

to the PCR tube cap. Geo-seq is an example of a state-of-art work for RNA sequencing using LCM [23], [24], [29]. These groups combined single cell RNA sequencing method (Smart-seq2) [21] with LCM. The group successfully recovered full-length mRNAs from the microdissected areas, showing possibility in connecting imaging assays to the molecular assays. Other groups also applied LCM to genomic studies [25], [26], [28]. Implications that the LCM technologies give are the possibility of complex assays that can connect the spatial assays to the molecular assays, in contrast to other spatial omics technologies which still lack incorporation of different spatial assay modalities.

2.2. Commercialized spatial omics technologies

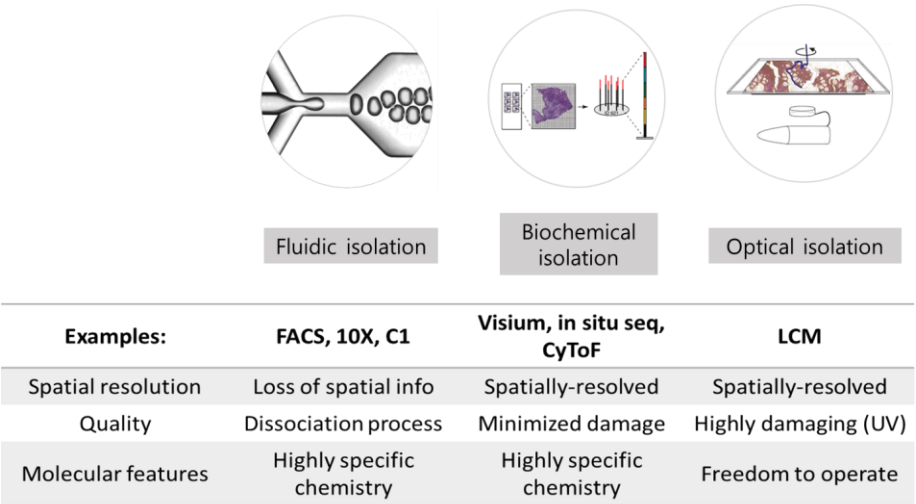


Figure 2.1 Comparison chart for major spatial omics platforms.

Although at its premature stage, the development of spatial omics has

led to many commercializations of different platforms (Figure 2.1). For the *in situ* sequencing technology, CARTANA (Sweden) that was spun off from Mats Nilsson's research group currently provides an integrated assay reagents that include custom padlock probes, RCA reagents, and fluorescent sequencing probes that can be applied to tissue sections applied on normal glass slides. The specific information regarding *in situ* sequencing technology is described in the section 2.1. Readcoor (United States) was spun off from George Church's research group and currently provides assay toolkit for FISSEQ. Through the commercialization process Readcoor developed an *in situ* sequencer with multi-omics properties. Likewise, the detailed explanation for FISSEQ is described in the section 2.1. 10X genomics (United States) was originally founded for commercializing droplet-based single cell sequencing preparation instruments. Its Chromium technology is similar to the drop-seq method and utilizes hydrogel particles with barcoded poly-T baits to capture the transcripts. This commercialization was a disruptive success that bloomed the single cell biology industry. Then, 10X genomics acquired Spatial Transcriptomics (Sweden) that was spun off from Joakim Lundeberg's group. Spatial Transcriptomics is a representative isolate-and-transfer technology that utilizes the barcoded poly-T baits that are in form of a microarray, on which the tissue section is applied. As 10X Genomics acquired Spatial Transcriptomics, 10X Genomics developed a new product, named Visium. In contrast to the original Spatial Transcriptomics technology that used microarray with 100 μm X 100

µm spots, Visium was upgraded with a microarray of 55 µm X 55 µm spots. In other words, the spatial resolution of previous technology was approximately 100 cells, while that of the Visium technology is around 50 cells. Through the recent publications, this technology still has more rooms for improvements such as spatial resolution, chemistry, and labor intensiveness. Lastly, Nanostring (United States) developed a spatially resolved proteomics platform called GeoMx [57]. This platform utilizes the nCounter platform and Nanostring beads that are basically a thread of differently colored nanobeads that represent a specific colorimetric barcode, previously developed for multiplex protein detection. In GeoMx Nanostring used antibodies that are labelled with specific oligonucleotides. The oligonucleotides provide hybridization sites, on which the Nanostring barcodes can be hybridized. In conclusion, although the field is in naïve state, several platforms were commercialized in the recent three years, indicating demand for spatial omics platforms and academic and industrial developments are more to come.

2.3. Previous research in the group

In 2015, the group published a paper regarding a optomechanical isolation device that can print out desired DNA molecules from the NGS sequencing plate [8], [58]. Named Sniper cloning, this method uses laser to the sequenced DNA clones that are formed in NGS sequencing plate. Whether the DNA molecules are sequenced optically (Illumina, Pacific Biosciences, or BGI

Genomics) or chemically (Oxford Nanopore, IonTorrent, or Genapsys), the nature of the NGS technologies are that the DNA molecules to be sequenced are applied to a two-dimensional substrate for further enzymatic or chemical reactions that enable the sequencing process. After the sequencing run is over, the sequencing substrates are discarded and the raw data of the sequencing run contains the coordinates of the DNA clusters inside the two-dimensional substrate and the DNA sequence reads matched to the coordinates. The main idea of this research was to utilize the sequenced substrate instead of discarding it. Because the sequenced oligonucleotides are sequence-verified, if we can isolate out the desired oligonucleotides with desired sequences from the sequencing substrate, the highly pure oligonucleotides can be synthesized with low-cost. This methodology was termed Sniper cloning, as it utilizes nanosecond pulsed laser to “snipe” out the desired oligonucleotides from the sequencing substrate. Using the same technology, Yeom et al. developed a NGS error validation platform that can select out very rarely appearing genetic variants among a pool of oligonucleotide libraries [59]. Also, this synthesis platform was applied to generate guide RNA libraries for Cas9 libraries [60]. While this platform can be a “selective” platform, it also can be a constructive platform where the isolated DNAs can serve as building blocks for building larger nanostructures [61] or the genomes [62]. This pulsed-laser prototype became the basis for SLACS, which basically uses the main idea of applying a pulse of laser to a desired area [31], [33], [63].

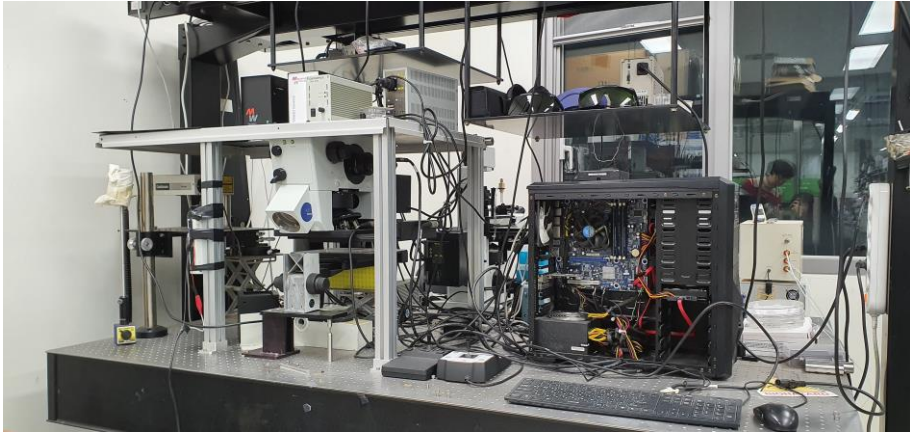


Figure 2.2 Prototype built for Sniper Cloning [64].

Sniper cloning

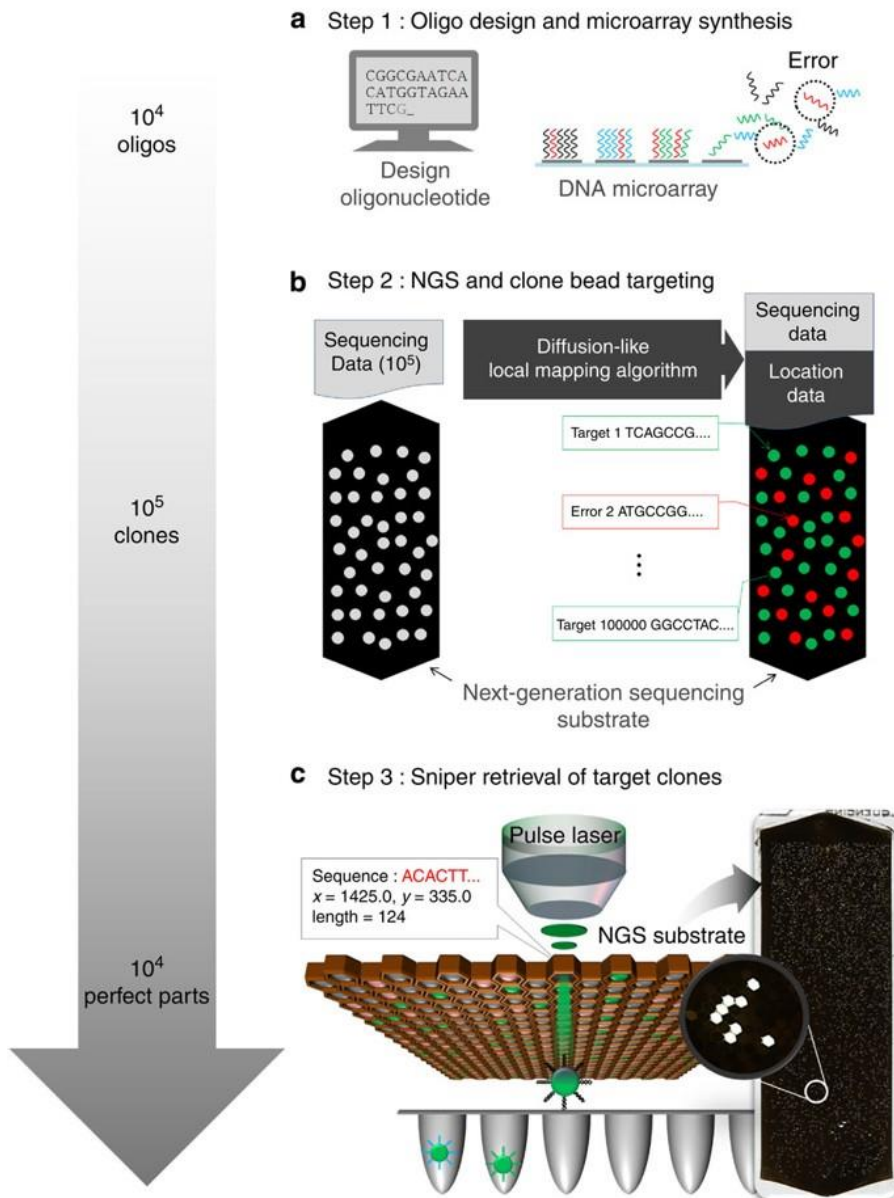


Figure 2.3 Concept of Sniper Cloning [64]

Chapter 3. Platform development

In this chapter, the entire process is described for developing SLACS and *in situ* spatial assay for spatial omics platform (figure 3.1). Developing on the optomechanical transfer of biospecimen using pulsed laser, novel spatially-resolved omics platforms were developed in three different aspects. First, hardware and the software that comes with the hardware were all built in-house. Second, the biochemical methodologies (whole genome amplification, whole transcriptome amplification, and rolling circle amplification) were developed to fit into the hardware platforms. Third, bioinformatics pipeline was developed to analyze the genome and the transcriptome. Because the *in situ* spatial platform used the same approach by Mats Nilsson's group [44], it will be discussed more extensively in Chapter 4.

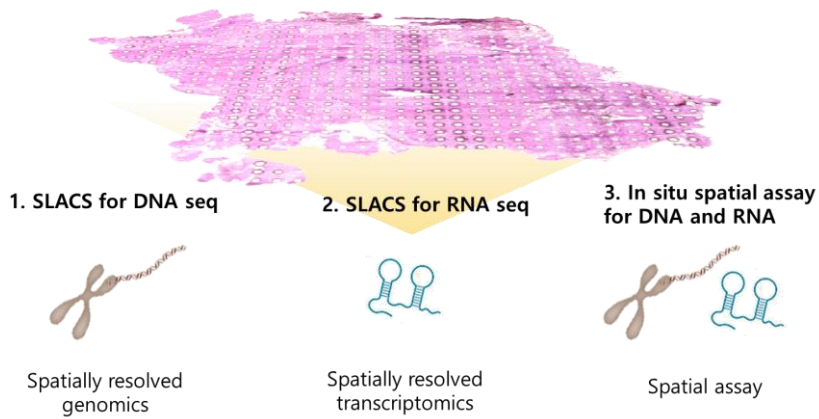


Figure 3.1 Platforms described in this dissertation.

3.1. Development of SLACS and remote selection system

Spatially resolved Laser Activated Cell Sorter (SLACS) was built after the development of the Sniper cloning device [64]. The additive functions are the fluorescent stages, slit that can adjust the isolation spot sizes, and the array of objective lenses that have longer working distances than others (Figure 3.2). Therefore, the whole system is comprised of 1) nanosecond laser (Nd:YAG, $\lambda = 1064$ nm), 2) 1064 nm bandpass filter, 3) reflective bright field that serves as the guide light, 4) laser shape modulating slit (x,y-axes), 5) a CCD camera, 6) fluorescent light source, 7) fluorescent filter array, 8) focus controlling axis and lens array, 9) two motorized stages for sample manipulation and retriever manipulation, and 10) bright field light source. The controller computer was connected to the SLACS device for manual and automatic control.

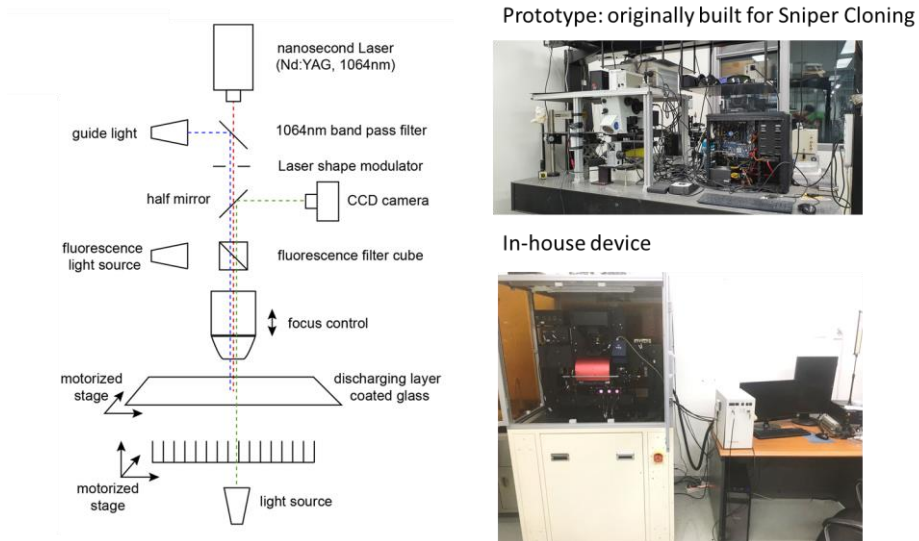


Figure 3.2 The design and a picture of the SLACS system.

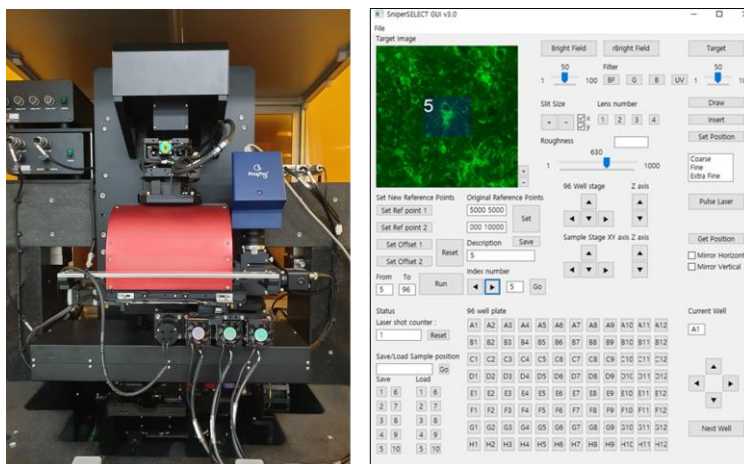


Figure 3.3 The closed-up picture of the SLACS system (Left) and the user interface built for the SLACS system (Right).

The software scripts that can control the in-house device were written

in Python, using the image processing libraries. The software is able to control the X, Y, and Z axes of the each motorized stages, the Z axis of the lens array, the X and Y axes of the light shape modulator, pulsed-laser, and all the light sources integrated into the system. The software can be used for two different modes: the manual mode and automatic mode. In contrast to the manual mode, the automatic mode isolates the regions of interest automatically. This was enabled by designing and building another user interface for selecting regions of interests from the tissue (Figure 3.4).

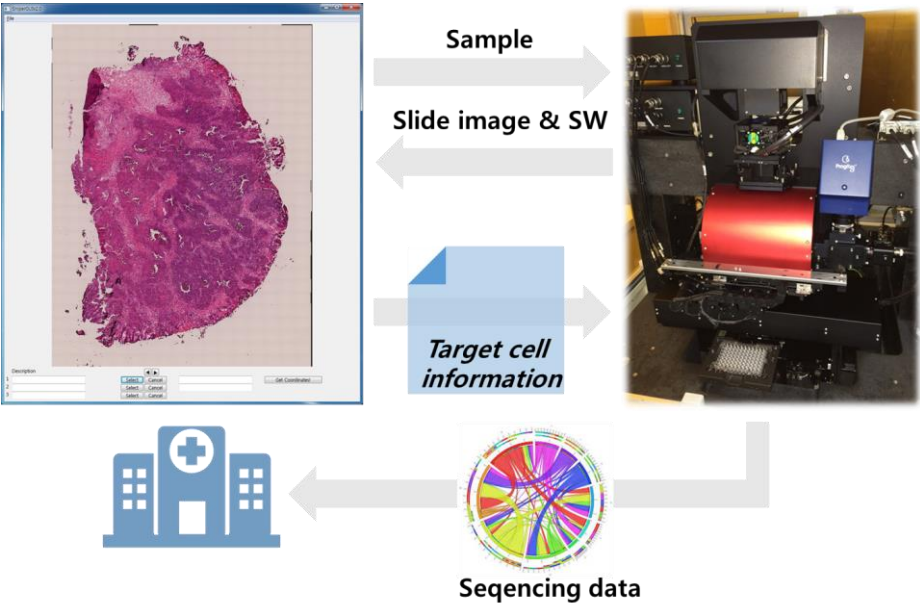


Figure 3.4 The remote system for selecting desired region of the biospecimen.

The remote system was built for the users in hospitals and other institutions for different collaborations. If the biospecimen samples are received

from the hospitals and different collaborators, the whole slide is imaged through an automatic image-stitching fluorescent microscopy. If the tissue section to be analyzed is stained in bright field modality, the tissue was imaged with bright field and if the tissue section is stained in fluorescence modality, the tissue was imaged with corresponding fluorescence channels. Then, using a remote computing server, the sections are sent to the pathologists, medical doctors, life scientists and other collaborators. The graphic user interface that can process the image and read out the coordinates of the regions of interest is used by the collaborators to select the regions to be isolated. Then the output files of the user interface program are transferred to the SLACS controller to isolate the selected regions for downstream analyses.

3.2. Whole genome sequencing strategies for SLACS

The first development of reagents that I and my colleagues applied SLACS was DNA sequencing. It was thought that every cell in one human's body had the same genome. After the human genome atlas that sequenced one human's genome, NGS was developed and it is now a 100 USD genome era in 2020. Now that the sequencing costs are decreasing rapidly, the researchers started to look into the whole genomes in single cell level. The landmark paper by Nicholas Navin revealed that the single cell genomic landscape showed heterogeneity in breast cancer tissue [65]. By sequencing the single cell genome, Navin et al. found that the tumor cells gain variants in a punctuated manner,

forming different subclones within a tumor of a person. This work is significant in single cell genomics because it first applied whole genome amplification method to single cells and demonstrated the evolutionary tree between the cells from one individual. The so-called intra-tumoral heterogeneity has now been extensively studied, but that these cells have to be dissociated into solution and thus lose their spatial context was the limitation that these studies had. SLACS has advantages in analyzing the genomes of these tissues because the cells can be isolated with preserved tissue information and the whole genome amplification methods can be applied for spatial genomics.

Therefore, I and my colleagues introduced phenotype-based high-throughput laser-aided isolation and sequencing (PHLI-seq) using SLACS device and applying whole genome amplification methods to the isolated targets (Figure 3.5). In this section, I will describe how the whole genome amplification methods were developed for SLACS. Human cells normally have approximately 6 pg of DNA that comprises genome. Because conventional NGS technologies require 1 ng of DNA as a minimal input, the single cell genome has to be amplified for NGS analysis. There are several approaches that can amplify the whole genomes. Degenerate oligonucleotide-primed polymerase chain reaction (DOP-PCR) is one of the earliest approaches that can amplify the whole genomes from single cell [66]. DOP-PCR utilizes degenerate oligonucleotide primers to prime nonspecific sites within genomes and amplifies the regions via PCR. However, the method was reported to

produce amplification bias, an amplification tendency of certain regions over other regions.

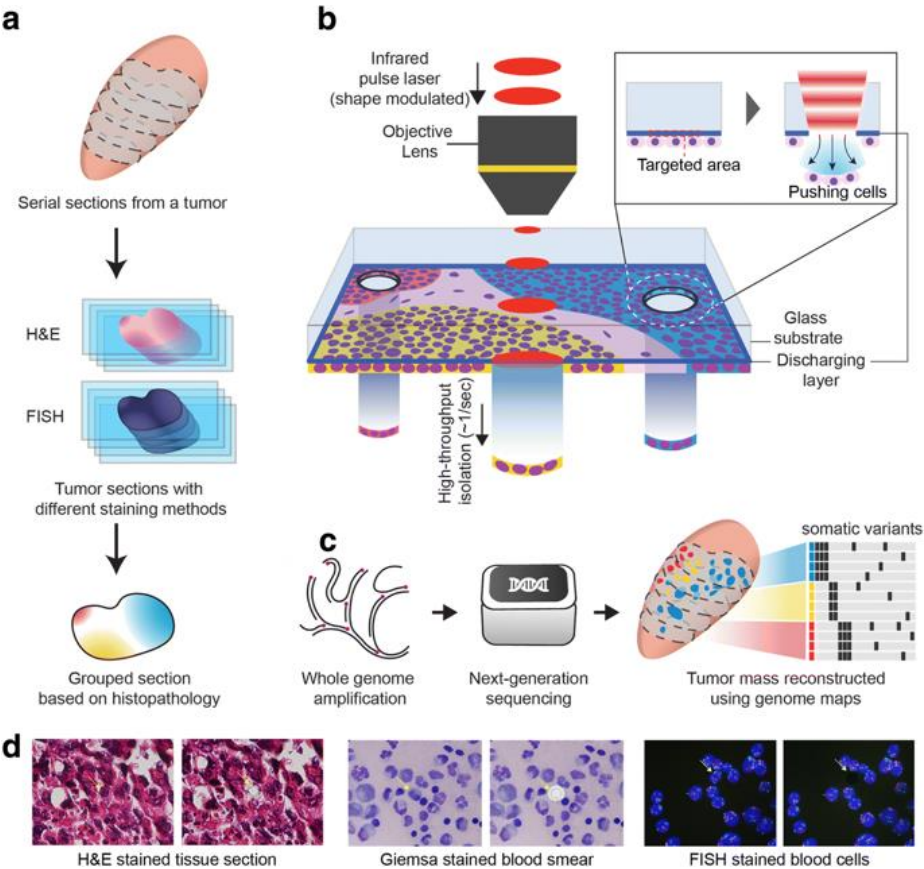


Figure 3.5 PHLI-seq uses SLACS to sequence the whole genomes of cells [33].
 (a) Serial sections from a tumor is prepared with different staining modalities.
 (b) The SLACS device isolates the cells in high throughput. (c) The genome within are amplified with whole genome amplification and goes through NGS.
 (d) Cell isolation images.

Multiple Annealing and Looping Based Amplification Cycles

(MALBAC) is another method that was developed from DOP PCR [67]. The method combines Phi29 polymerase-based hyperbranching amplification method and the DOP PCR method. Despite a step was added, it has been reported to produce similar artifacts to DOP-PCR.

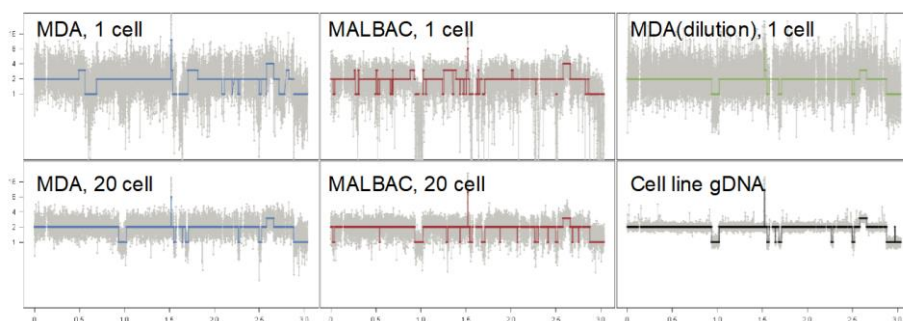


Figure 3.6 Copy number alteration (CNA) plots of PHLI-seq cell line samples with MDA and MALBAC.

Multiple displacement amplification (MDA) uses Phi29 polymerase, which generates large fragments of DNA from the genome template by amplifying with random hexamer primers. If the random hexamers hybridize to the template genome, the Phi29 polymerase elongates the DNA with high strand displacement activity. Using PHLI-seq, MALBAC and MDA were compared to determine the whole genome amplification method better suited for SLACS system. The copy number alterations (CNA) were plotted in Figure 3.6. Some cancer cells are known to have CNAs that represent a genomic signature of how many chromosomes are present in the genome. The experiment was performed

with SK-BR-3 cell line. When compared to the “Cell line gDNA”, which depicts the plot of CNAs of the bulk genomic DNA, the grey noise is significantly low compared to other CNA plots. Both MDA and MALBAC amplified samples show much larger data noise, while the MDA samples show closer CNA profile to that of the bulk sample. Therefore, the research team decided to use MDA in SLACS analysis.

The kit that the PHLI-seq paper utilized was from GE Healthcare’s Genomiphi MDA kit. This kit utilizes chemical lysis using alkaline reagents, but the fixed samples of the tissue renders it difficult to be lysed. Therefore, we modified the kit using a protease called proteinase K. This enzyme is one of the strong proteinases, which can unravel the histones and other proteins that exist within the cell nucleus. This lysis method showed stronger lysis compared to the alkaline lysis method. For the optimization of the whole genome amplification chemistry, a quality control standard was necessary. 11 Primer sets from different chromosomes were designed for PCR validation to see if the whole genome was amplified thoroughly (Table 3.1). The sixteen primer sets are designed for sixteen different genes in sixteen different chromosomes. If more than 15 out of 16 sites were amplified, the whole genome amplification product was considered adequate for NGS.

Table 3.1 PCR validation primer sets for MDA products [31].

Validation primer 1	F	5'-TCTAGACCTGCCACTGGGAA-3'
	R	5'-ATGCAGCAGGTGCTGAGTAA-3'
Validation primer 2	F	5'-ACTGCCCATGCACTTTGACT-3'
	R	5'-CCACACTCCTTCGCCAACTT-3'
Validation primer 3	F	5'-ACACCATGAAGCAGAAGGGG-3'
	R	5'-TGCATGAGCCCATGTACCTC-3'
Validation primer 4	F	5'-GGATGACTGGAGCAGGAAG-3'
	R	5'-TGGGCAGCATCCATTGAGAG-3'
Validation primer 5	F	5'-AAGAGCATTTTATGCTCCATCTG-3'
	R	5'-CACATACAGACCCGCTGGAA-3'
Validation primer 6	F	5'-GCCAACATGGCCAGGAAGTA-3'
	R	5'-TCATGTGCACAAATGTATGTTTCTT-3'
Validation primer 7	F	5'-GGAAGGCTTTGAAGAAGGTGAAT-3'
	R	5'-AGCACCAAAAAGGCACATACC-3'
Validation primer 8	F	5'-AACCTCCCAATCCAGTGC-3'
	R	5'-ACAGTTCTTTTCACTACTGCCG-3'
Validation primer 9	F	5'-GAGCCACATGAGTCTGCCAT-3'
	R	5'-AGAGCCAGGCTTTTGCTGAA-3'
Validation primer 10	F	5'-CTTCTTTGGGGACCATCC-3'
	R	5'-CCCATCGTCTCTGCTGACAA-3'
Validation primer 11	F	5'-GTGTGCGGAAGGTACGGTTA-3'
	R	5'-TTGCTCCTGCTCAGGTCTTG-3'
Validation primer 12	F	5'-TCAATCTCCATGCCAGGGT-3'
	R	5'-TTCAGTCCCAACATTGCACG-3'
Validation primer 13	F	5'-GTGGACAGCTGACACGAGAG-3'
	R	5'-CGAGAGGCCACAGAAGTAGC-3'
Validation primer 14	F	5'-AGGTACCGTACATACCAGGA-3'
	R	5'-TGGTGCTGGCAGGATAACAG-3'
Validation primer 15	F	5'-CTTGCTGGTCTGTCCCTCTG-3'
	R	5'-ATCCTCCCCACCTCCTTTT-3'
Validation primer 16	F	5'-ACCTCAGCAACCTTCAAGAACT-3'
	R	5'-GGGTGTAGAATCAAACAGCG-3'

Another strategy to check for the whole genome amplification quality was the amplification start time (Figure 3.7). By mixing SYBR Green I fluorescent dye, all whole genome amplification was performed in a quantitative PCR machine. The quantitative PCR machine is able to detect the increase in fluorescent signals as the genomic DNA molecules are amplified. We observed a strong correlation between the amplification start time and the whole genome amplification quality. As the amplification started earlier, the whole genome amplification quality was better. The whole genome amplification quality was measured by the area under the Lorenz curve, which shows the frequency of alleles that are aligned with the reference genome. Also, the plots for CNAs were used for checking the quality of the whole genome amplification products. As the bin for counting a DNA fragment unit increases, the CNAs become less strict. Using a bin of 1 kilobase, plots for CNAs for different samples with different start times and different conditions were depicted in Figure 3.7. There was no correlation between the number of laser shots and the whole genome copy number as demonstrated in 3.7 (f). However, formalin fixatives in cell line showed disruptive impact on the whole genome copy number. Therefore, PCR validation, measurement of the amplification start time and low-depth NGS results to determine CNAs were determined to select samples to be sequenced.

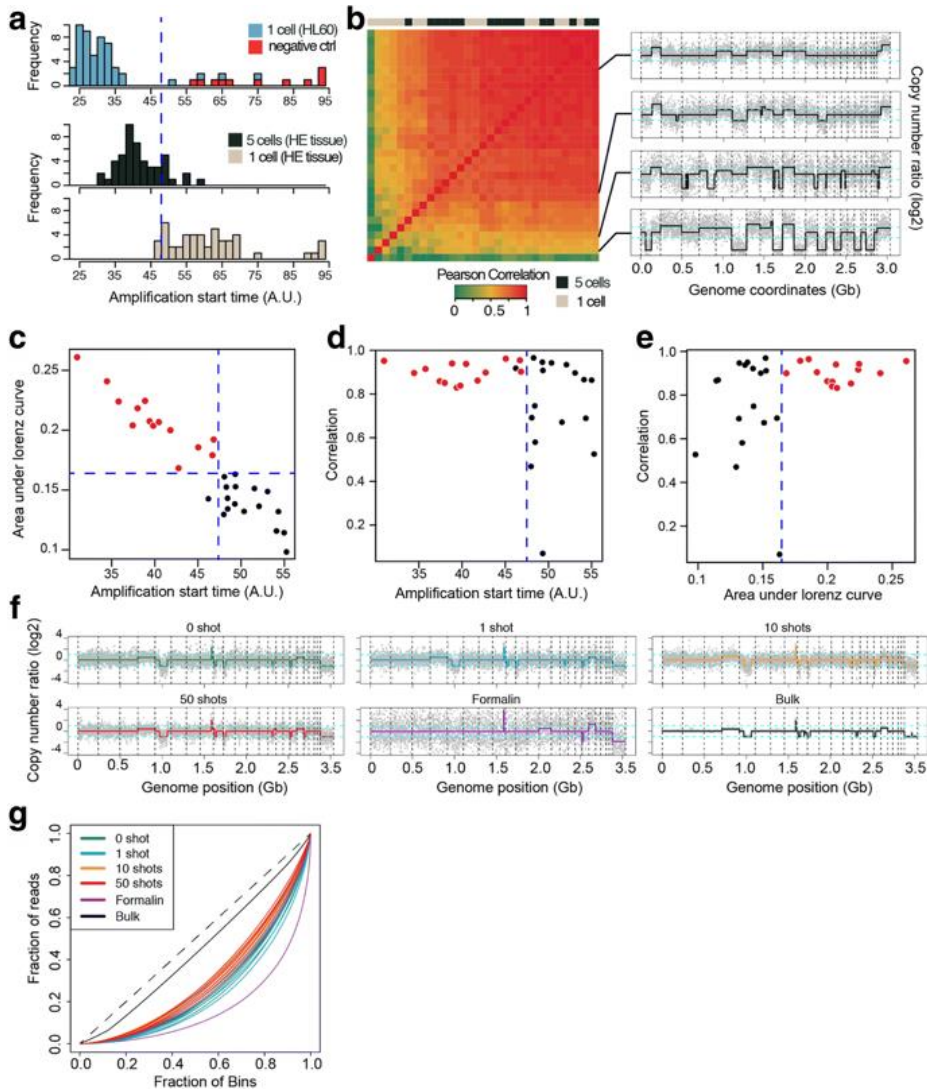


Figure 3.7 Quality control experiments for different parameters for PHLI-seq [33]. (a,b) Relation between the whole genome amplification start time and copy number ratio. (c,d,e) Correlation between amplification start time and area under the Lorenz curve. (f) CNAs for different samples (g) fraction of bins vs.

fraction of reads for different samples.

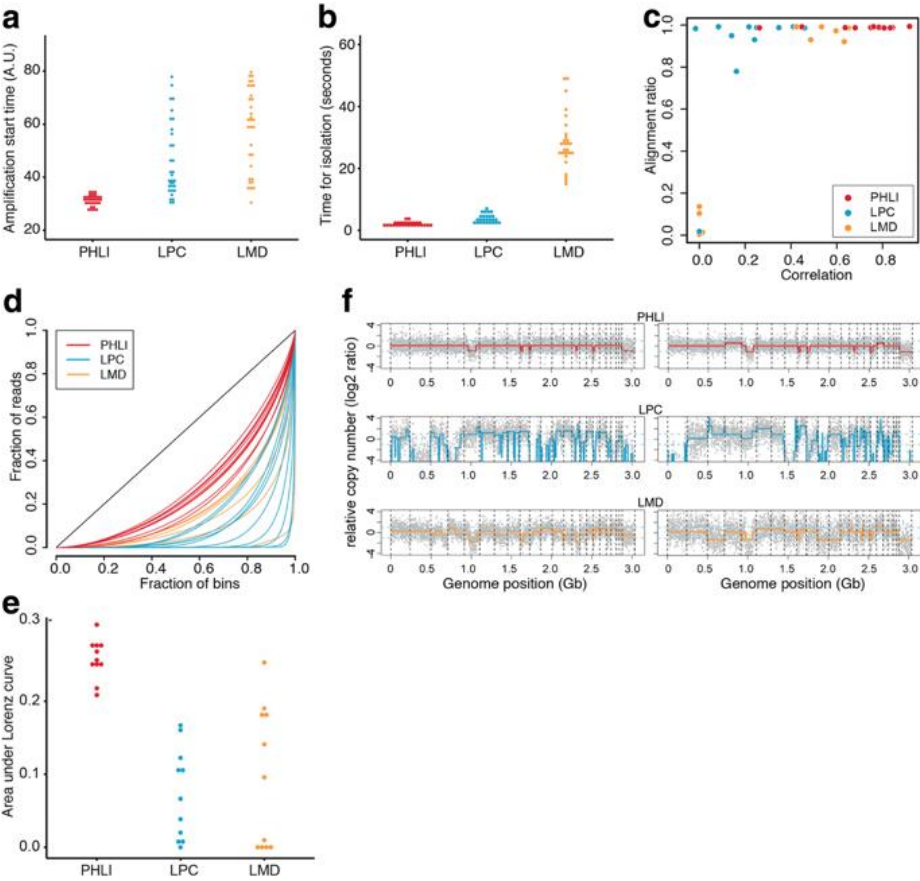


Figure 3.8 Platform comparison to laser pulse catapulting (LPC) and laser microdissection (LMD).

The sequenced genome from cells isolated with SLACS showed higher quality compared to that with LPC and LMD, which are commercialized competitive technologies that utilize laser to isolate cells (Figure 3.8). Through

analyzing the amplification start time (Figure 3.8 (a)), time for isolation (Figure 3.8 (b)), alignment ratio to genome (Figure 3.8 (c,)), Lorenz curve (Figure 3.8 (d,e)) and relative CNA plots were compared. In conclusion, PHLI-seq showed superior whole genome quality compared to other laser-based cell isolators.

3.3. Whole transcriptome sequencing strategies for SLACS

Similar to the genomic contents, the transcriptomic contents within single cell range from 30 pg to 100 pg. To type the gene expressions, these messenger RNAs (mRNAs) must be amplified before sequencing. There are several strategies for this and they are described in Chapter 2 of this dissertation. All methods described in Chapter 2 acquire approximately hundreds to thousands of gene counts from single cell or small number of cells. Smart-seq2 [21] and Smart-seq3 [68], however, are current state-of-art transcriptome amplification strategies that can acquire tens of thousands gene counts. The specific details for Smart-seq 2 are described in Figure 1.4. There has been several spatially resolved transcriptomics approaches using cryosections. The so-called RNA tomography, or tomo-seq, lysed the whole cryosection that contains tens of thousands of cells and performed Smart-seq2 [69]. An upgraded versions are studies using LCMs and these are discussed in detail in Chapter 2. Similar to tomo-seq, these technologies utilized the high sensitivity and high recovery rate of Smart-seq2. I and my colleagues decided to use Smart-seq2 and a modified version of Smart-seq2. The step is as follows: (1)

hybridization of barcoded poly-T sequence; (2) reverse transcription to synthesize complementary DNA (cDNA); (3) template switching to generate double stranded DNA/RNA hybrid; (4) PCR to amplify the hybrids; (5) Tn5 tagmentation of the hybrids to generate a library of NGS adapters; and (6) PCR to amplify the NGS library. We compared the data to the bulk mRNA sequencing data to generate R squared value of 0.81 as shown in Figure 3.8.

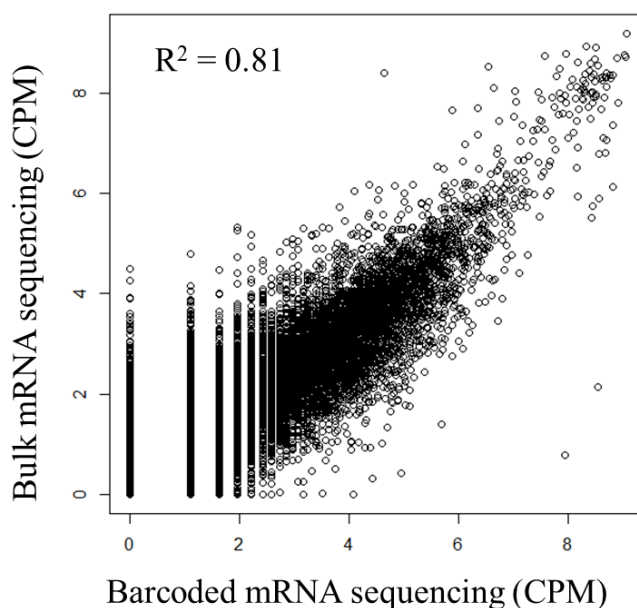


Figure 3.9 Correlation between the barcoded mRNA sequencing of single cell material (~20pg) and bulk mRNA sequencing data.

To further characterize the amplified transcriptome quality, we performed gene counting from different samples. Using SLACS, 10 cells from

three different cell line were isolated (Figure 3.9). 3T3 is a mouse fibroblast cell line, while HEK293T and HL60 cell lines are human embryonic kidney cells and human leukemia cells, respectively. Two different cell line preparation types, named “fixed” and “live” represent methanol-fixed cells and live cells, respectively. The red dotted line in Figure 3.9 represents approximate gene count number from research that uses Fluidigm’s C1 platform, while the blue dotted line represents approximate gene count number from research that uses 10X genomics’ Chromium platform. The dotted line is different for 3T3 and other two lines because the gene count from mouse samples is usually lower than that from human samples. As the box plots in figure 3.9 demonstrates, the gene counts from cells that were isolated with SLACS were higher than the counterparts of C1 and Chromium. The box plot that demonstrates the gene count from HL60 sample shows little lower gene count than others, but it was still higher than the average gene count generated with Chromium. This is due to the naturally high RNase contents in immune cells. Because of the high contents of RNases, it is difficult to recover RNAs from the immune cells. Nevertheless, the gene count from cells isolated with SLACS were high compared to that from cells treated with different scRNA seq platform, showing high quality of the SLACS method. The average gene counts from other spatial transcriptomics platform such as ST and Visium show little less gene count number compared to Chromium. In that manner, we were able to provide evidence that the RNA quality from cells isolated with SLACS is high.

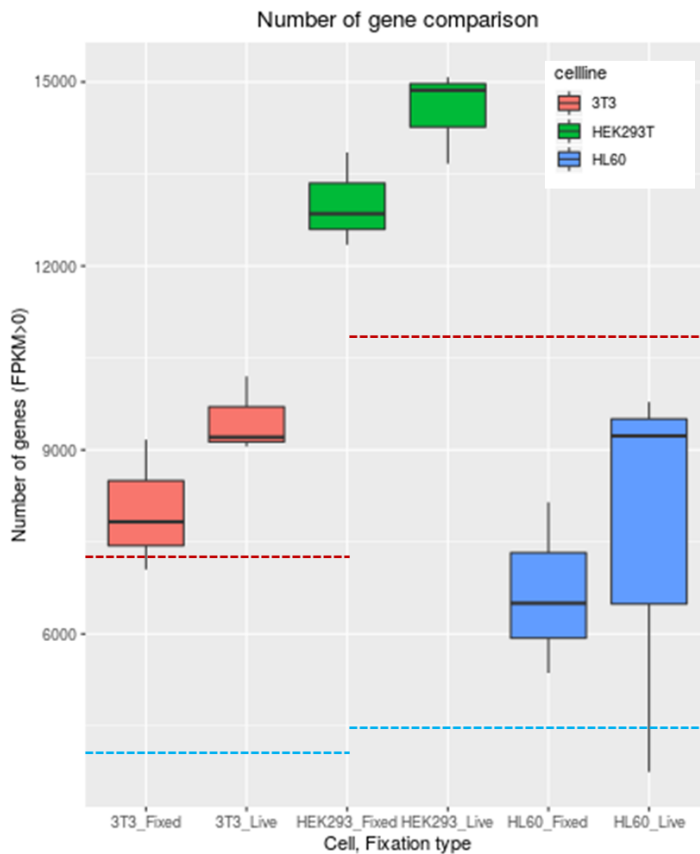


Figure 3.10 Gene counts according to different cell types and different cell preparation methods.

Also, using the RNA seq data, by counting the reads aligned to different genes, gene expression levels can be discovered in different samples analyzed above (Figure 3.10). The top panel of the Figure 3.10 displays the gene expression level in common from different cell lines. The middle panel displays the differential gene expression level from different cell lines.

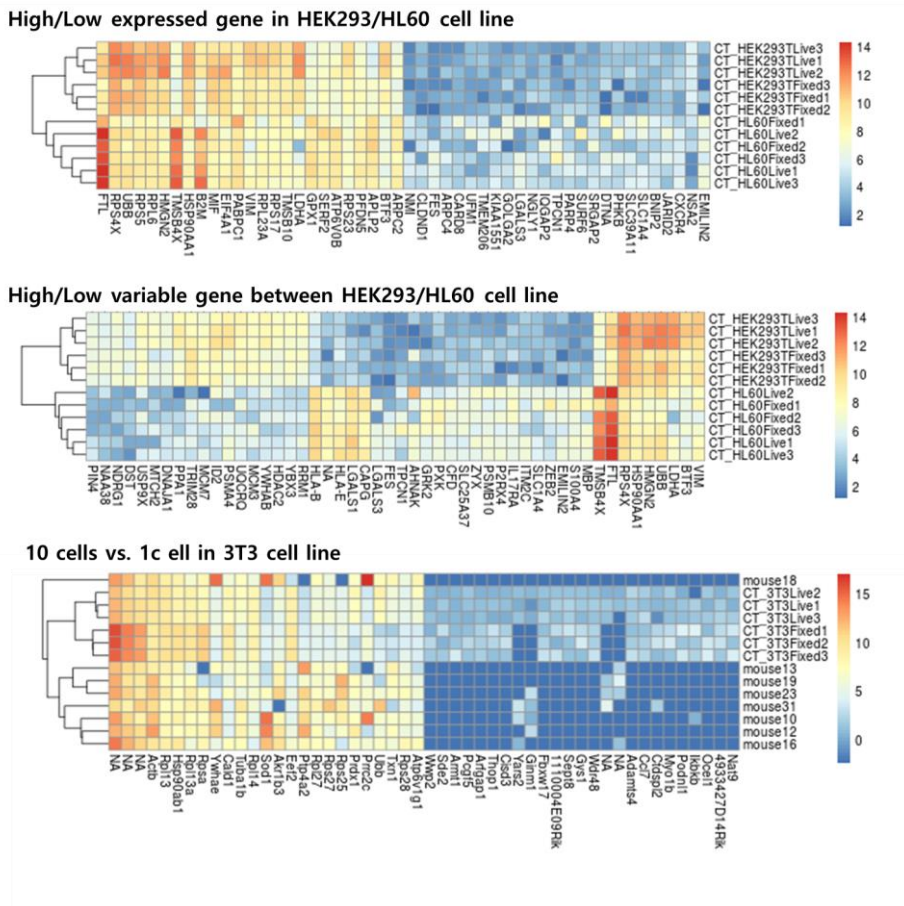


Figure 3.11 Heatmap of the RNA expressions from different samples.

As illustrated in this panel, two different clusters of the differential gene expression show that the data was successfully divided into two different sample types. The bottommost panel shows the correlation between the gene expression level of 10 cells and 1 cell from the same cell line. The heatmap shows there is a high correlation between the two different groups.

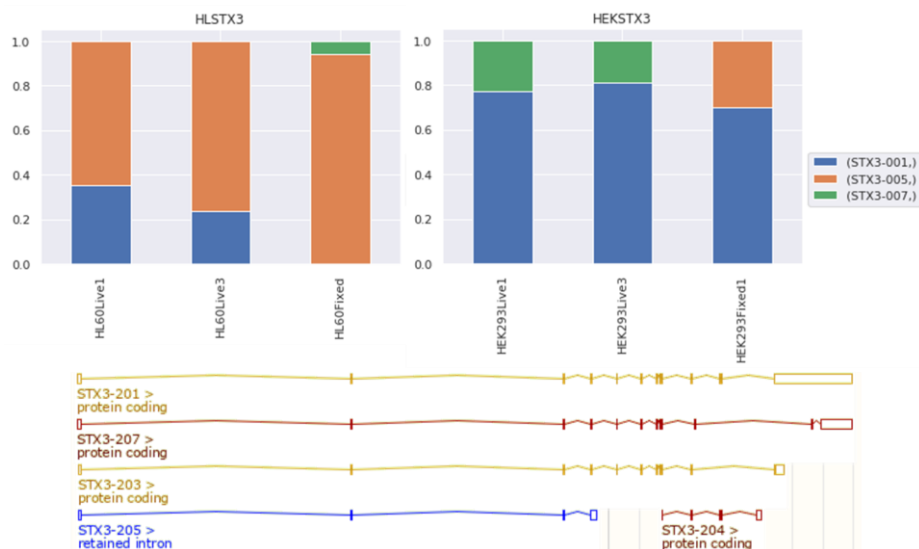


Figure 3.12 Transcript isoform diversity between the two different samples in gene STX3.

Furthermore, because we used Smart-seq2, the amplified cDNA library is not pooled. For high throughput isolate-and-transfer technologies such as Visium, the molecules must be pooled. Therefore fragments that do not contain barcodes must be discarded in the pooling process. However, for SLACS, the amplicons are confined to a microwell or a microtube, enabling full transcriptome sequencing of the samples. To demonstrate, we found a known splicing variant of STX3 gene that are differentially expressed in HEK 293T cells and HL60 cells. Through assessing the full transcriptome data, the splicing variant diversity was illustrated in a bar graph (Figure 3.11).

Chapter 4. Platform application

In this chapter, I describe the applications of the proposed platforms in oncology and neuroscience. SLACS for spatial genomics was applied to breast cancer and circulating tumor cell applications. SLACS for spatial transcriptomics platform was applied to breast cancer tissue, mouse brain tissue, and brain organoids. Finally, a spatial assay platform using *in situ* molecular profiling was applied to CTCs from breast and pancreatic cancer patients, for the platform's future application to SLACS device.

4.1. Applications of SLACS to spatial genomics

SLACS was applied to spatial genomics using MDA method as described in Chapter 3. In this section of the Chapter 4, I describe two different studies in tumor cells and circulating tumor cells, both in breast cancer. Through these applications, the powerful potential of SLACS in spatial genomics is demonstrated.

We first applied the SLACS for spatial genomics to the tumor sections from breast cancer patients [33]. This was a collaboration with Seoul National University Hospital that utilized the SLACS device, the remote selection software, and the bioinformatics pipeline to infer tumor evolution from the different subclones within a single tumor. Cancer cells display intra- and inter-tumor heterogeneity, which hinders the extensive investigation of the cancer subclones within a tumor. Therefore, a spatially resolved analysis of the heterogeneous cancer genome, in which the data are connected to the three-dimensional space of a tumour, is crucial to understand cancer biology and the clinical impact of cancer heterogeneity on patients. However, despite recent progress in spatially resolved transcriptomics, spatial mapping of genomic data in a high-throughput and high-resolution manner has been challenging due to current technical limitations. By applying PHLI-seq, the heterogeneity of breast cancer tissues at a high resolution is revealed and the genomic landscape of the cells and phenotypes in the tumour mass is mapped to their corresponding

spatial locations. Additionally, with different staining modalities, the genotypes of the cells can be connected to corresponding phenotypic information of the tissue. Together with the spatially resolved genomic analysis, the histories of heterogeneous cancer cells in two or three dimensions are inferred, providing significant insight into cancer biology and precision medicine.

First, a tumor section from human epidermal growth factor (HER2)-positive breast cancer patient for initial demonstration of the spatial genomics platform. I applied PHLI-seq to a HER2-positive invasive breast cancer to demonstrate that the genomes of the cell clusters from a stained tissue section can be effectively analyzed. A tissue section of a preserved fresh frozen sample was prepared and stained by hematoxylin & eosin (H&E). The prepared section was scanned by automated microscope to generate a high-resolution whole-section image (Figure 4.1). The image was segmented into cell clusters to generate a binary image, and spatial information and morphological phenotypes were extracted from each cell cluster. Then, the grouping of the cell clusters was performed using weighted hierarchical clustering to generate six groups. Based on an additional inspection by pathologists, 53 cell clusters were selected from the groups to be analyzed by PHLI-seq. With an average of 20-30 cells in each cell cluster, whole genome amplification and quality filtering were performed. The filter-passed samples were analyzed by low-depth whole genome (n=53), whole exome (n=12), and targeted sequencing (n=53).

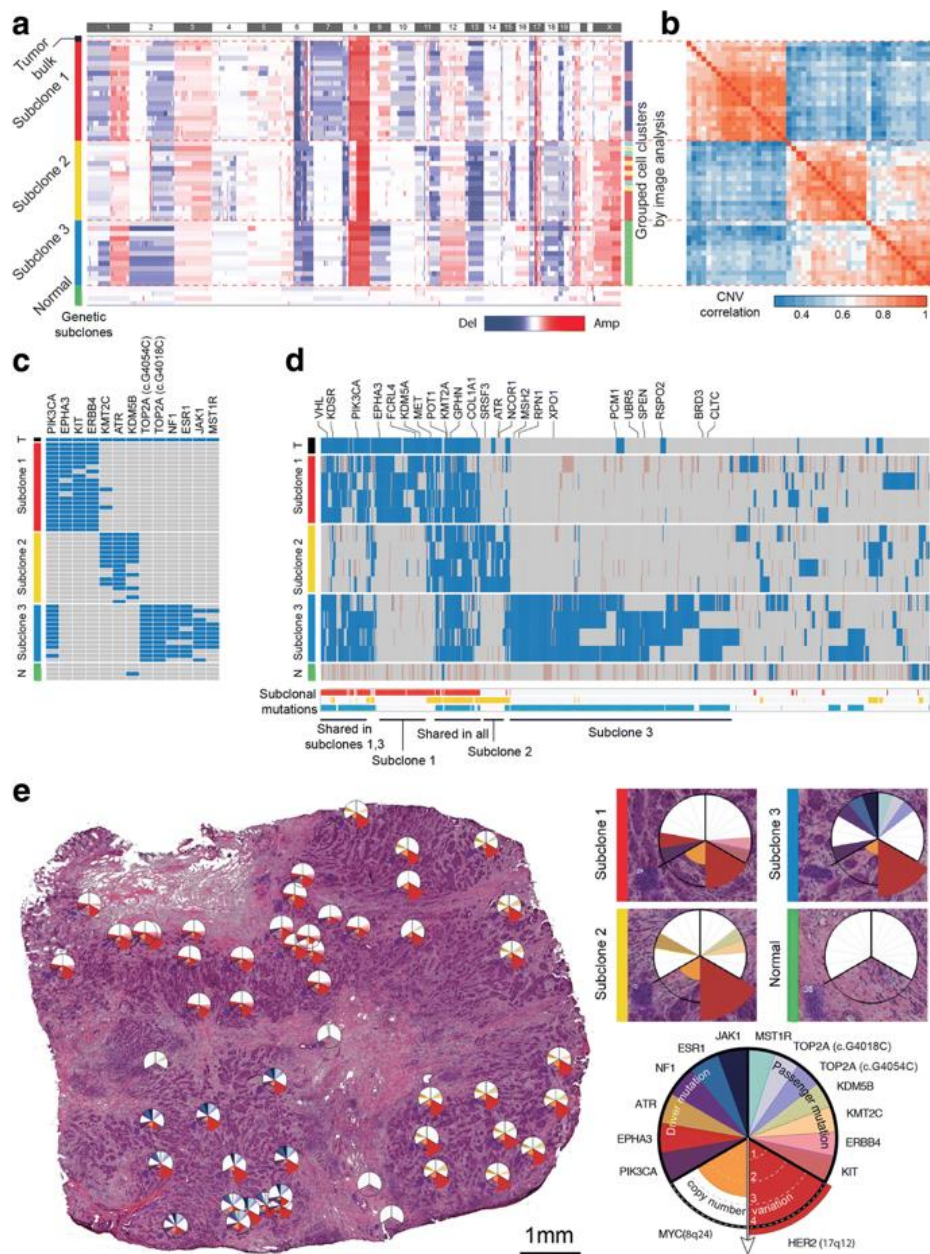


Figure 4.1 SLACS for spatial genomics applied to HER2+ breast cancer tissue.

Genome-wide copy number alteration analysis by low-depth whole

genome sequencing showed that there are three major subclonal populations in the tumor sample (approximately unbiased p-value > 0.99, multiscale bootstrap resampling with 10,000 iterations). The three subclonal populations had both shared and unique alteration profiles. The shared alterations include 1q gain, 8q gain, 8p loss and HER2 amplifications, all of which had been previously reported as frequent copy number alterations in human breast cancer and other types of cancer. In this sample, the causative fitness gains underlying breast carcinogenesis appeared to take place in the MYC (8q24), AKT3 (1q44), or ERBB2 (HER2) (17q12) regions, all of which were gained or amplified in 12.4%~21.3% of cases in The Cancer Genome Atlas (TCGA) project. Also, the up-regulation of these genes have been known to be associated with the causation of breast cancer. In contrast, subclones 1, 2 and 3 harbor unique gains or focal amplifications at 6q22.32-q23.2, 2q14.1-q14.2 and 8p21.3-p22, respectively. Among them, the amplification of CUG2 (Cancer-Upregulated Gene 2, also called CENPW, 6q22.32), known to be associated with tumor progression, might contribute to the high proliferation ability in subclone 1. One interesting observation is that the copy number alteration status was clearly divided into three distinct populations with no intermediate subclones. Since intermediate subclones might be excluded from the sampling process, we isolated a few tens of additional cell clusters at the boundaries between subclones. The isolated samples were analyzed by low depth whole genome sequencing. Then, the clustering analysis was performed based on inferred copy

number data of both previously isolated samples (n=53) and additionally isolated ones (n=27). The result shows that the 80 cell clusters from the HER2 positive tissue section are classified into one of the three previously defined cancer subclones. This result reinforces the evidence for the punctuated copy number change followed by a period of stasis, as demonstrated in previous studies.

To investigate the genomic diversity at the single-nucleotide level, I and my colleagues performed targeted sequencing of 121 genes associated with breast cancer. The results revealed unique mutational profiles in each subclone, consistent with those determined by whole-genome sequencing. In our targeted sequencing analysis of 53 cell cluster samples (20 (subclone 1), 16 (subclone 2), 13 (subclone 3), and 4 (no classification)), I found that mutations in PIK3CA, EPHA3, KIT, ERBB4 and KMT2C occurred in subclone 1, mutations in KMT2C, ATR and KDM5B in subclone 2, and mutations in TOP2A, NF1, ESR1, JAK1 and MST1R in subclone 3. Notably, PIK3CA p.M1043I, previously known to be an oncogenic mutation causing an increased PI3K lipid kinase activity, a constitutive AKT activation and the transformation of NIH3T3 cells and chick embryo fibroblasts, was shared between the subclones 1 and 3. In addition, PIK3CA p.M1043I and EPHA3 p.E794K in subclone 1 and ATR p.R2363X in subclone 2 were reported to be recurrently observed in breast and other types of cancer^{28,31}. Intriguingly, the stop-gain mutation ATR p.R2363X completely destroys the PI3_PI4_kinase domain, a key functional

structure, in tumor suppressor ATR that plays a role in DNA mismatch repair pathway. Another validated mutation that could cooperate in driving the subclonal heterogeneity is ESR1 p.E380Q in subclone 3, which has been known to be associated with endocrine-therapy resistance in breast cancer xenograft model [reference xx]. Even though the nonsynonymous EPHA3 p.E794K is a novel mutation that has not been reported so far in dbSNP, we confirmed that this mutation with the highest deleterious risk scores (SIFT_score, 0 and Polyphen2_HDIV_score, 1) occurred in Protein Tyrosine Kinase domain (PF07714) in EPHA3, sufficiently deserving further functional study in the future. Stop-gain mutation KMT2C p.Q3417X in subclone 1 completely truncated SET domain (PF00856), a key functional structure, implying a consistency with a recent report that low KMT2C expression has been associated with a poor outcome in ER-positive breast cancer patients [reference x3]. Cooperating with the oncogenic copy number amplifications mentioned above, those deleterious mutations in the three subclones might make significant contributions to the independent oncogenic evolution of each subclone.

For further analysis, I performed whole-exome sequencing for four samples selected from each subclone. I found that 75 mutations were shared in the three subclones and that 99, 75, and 382 mutations in VHL, KDSR, PIK3CA, EPHA3, FCRL4, KDM5A, MET, POT1, KMT2A, GPHN, COL1A1, SRSF3, ATR, NCOR1, MSH2, RPN1, and XPO1, occurred exclusively in subclones 1,

2, and 3, respectively. In contrast to the whole-exome mutation profiles in the three subclones by PHLI-seq, we could not find such representative mutation profiles in the sequencing data from the tumor bulk. The mutations detected in the tumor bulk mainly covered the mutational profile in subclone 1 (78.8% of exclusive mutations), whereas only approximately 6.81% of the mutations in subclone 3 were detected in the tumor bulk (Supplementary Fig. 9). This implies that PHLI-seq could provide rich information about subclonality and variants with low-level allele fraction in heterogeneous tumors, even those with subclones that are too minor to be detected by conventional methods.

I further analyzed consecutive sections of another breast cancer tissue to discover how heterogeneous tumor subclones exist in the three-dimensional space of the tissue, and to demonstrate how PHLI-seq can be an empowering tool to bridge genomics to histopathology (Figure 4.2). Each tissue section was about 7 mm x 6 mm in size and prepared to have an interval of 700 um between the consecutive sections. Seven tissue sections were prepared from the breast cancer tissue for PHLI-seq to discover genetic heterogeneity in 7 mm x 6 mm x 4.9 mm mass. Total 177 cell clusters were isolated and sequenced by PHLI-seq as described. First, I performed low depth whole genome sequencing to grasp a picture of heterogeneity of copy number alterations (Fig. 5b). I discovered that three genetic subclones existed, and they were denoted as in situ clone 1, in situ clone 2, and invasive clone. In situ clone 1 and in situ clone 2 had the same copy number profiles, except for the deletion of a q arm of

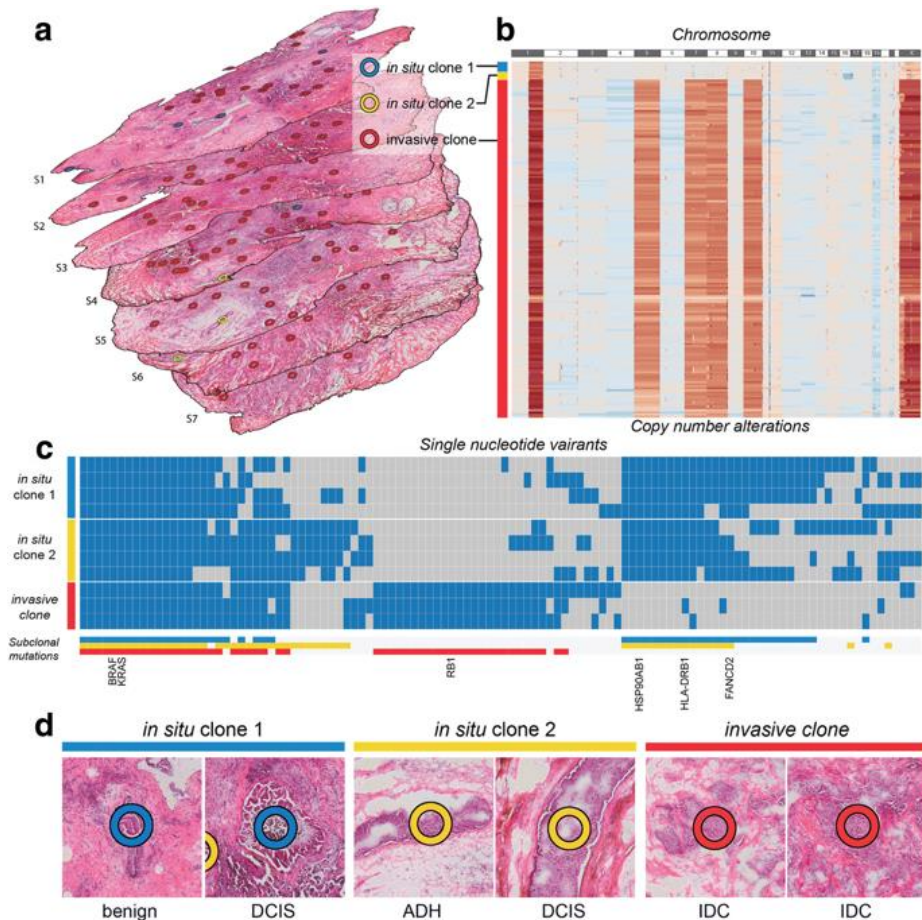


Figure 4.2 3D map of a tumor tissue using spatial genomics platform with SLACS

chromosome 16 and a p arm of chromosome 17 in *in situ* clone 2, suggesting *in situ* clone 2 was derived from the ancestral cells of *in situ* clone 1 by additional chromosomal deletion. Despite the difference in large chromosomal deletions, cancer cells in the two subclones were not classified into the same groups defined by genomic profiles by histological classification. Ductal

carcinoma in situ (DCIS) and benign cancer cells determined by histology were included in in situ clone 1, whereas DCIS and atypical ductal hyperplasia (ADH) were included in in situ clone 2 (Fig. 5c). As benign cancer cells progress to ADH, and then to DCIS, the histological classification of cancer cells in in situ clone 1 and in situ clone 2 are not exactly matches to the implication on lineages derived from the whole genome copy number alterations. This suggests that combined analysis of histopathology and spatially resolved genomics enabled by PHLI-seq has potential to contribute to clinical diagnostics.

On the other hand, invasive clone showed additional chromosomal amplification in chromosome 1 (q arm), 6, 7, 8, 10, and X. To investigate more genomic differences in single nucleotide level in the three subclones, I performed whole exome sequencing in each group. The result shows that the cancer cells in the invasive clone had mutations which do not exist in in situ clones. Surprisingly, even in situ clones also had mutations where invasive clone did not. This may suggest that invasive ductal carcinoma (IDC) is derived from early ancestry of DCIS, ADH, and benign cancer cells, not directly from DCIS in sequential manner. Moreover, in situ clone 2 showed mutations exclusive to in situ clone 1, which supports the previous explanation that, in situ clone 2 is derived from the ancestral cells of in situ clone 1, based on CNA profiles.

For future work, we are planning to apply PHLI-seq to a cohort of triple negative breast cancer (TNBC) patients who had relapses of breast cancer

after neoadjuvant chemotherapy (Figure 4.3). The cohort (n=24) comprises patients with no response, complete response, and partial response. Interestingly, a section was found to develop HER2+ breast cancer, which implies clonal expansion of a minor subclone in TNBC patients.

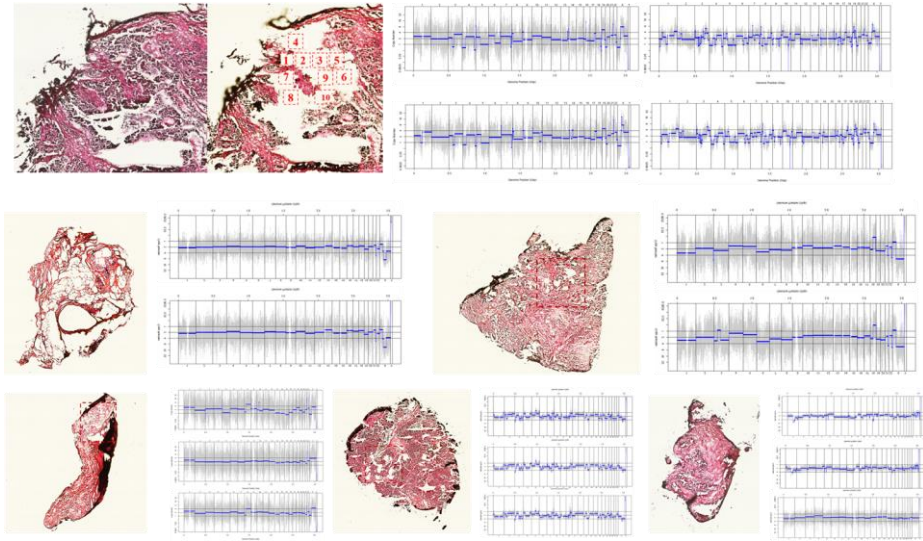


Figure 4.3 Follow-up study for PHLI-seq in breast cancer tissue sections. The CNA plots are accompanied by the corresponding tissue section slide in the left.

Another application of spatial genomics to oncology is to apply SLACS to circulating tumor cells (CTCs) [70]. CTCs are cells that are shed off the primary tumor and circulate around a cancer patient's body. Their importance in prognosis, diagnosis, and cancer monitoring is increasing, as their possible roles in metastasis have been revealed and they can represent the primary tumor that they were originated from [8], [71]–[73]. Because of their

clinical potential, CTCs have been greatly studied in various cancers. However, because these are rare cells, there are still many technical hurdles for their extensive studies. Especially, cancer heterogeneity also applies to CTCs, rendering them difficult to study. Therefore, platforms that can enrich the rare CTCs from whole blood and can isolate single CTC for downstream process like whole genome sequencing need to be developed.

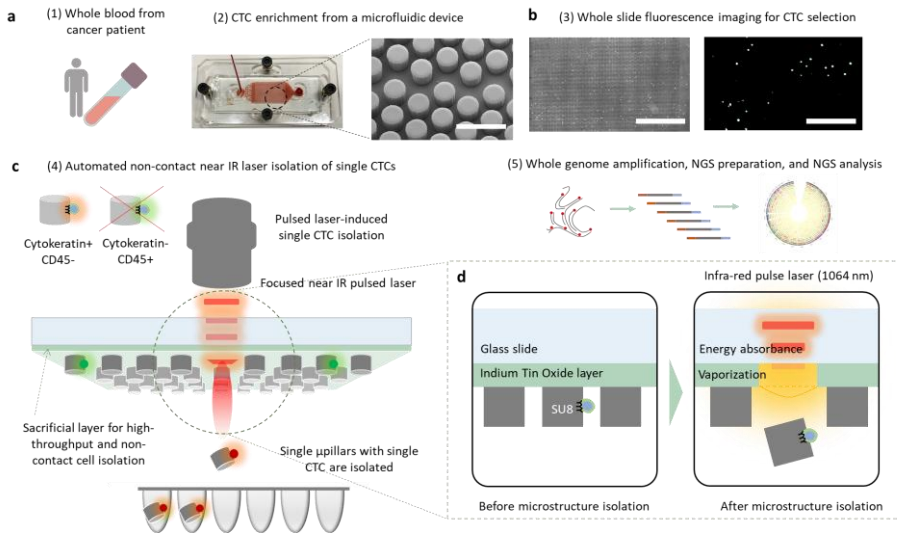


Figure 4.4 Microstructures that capture single CTC are selectively isolated with SLACS [31].

SLACS is applied to a biochip that isolates circulating tumor cells via immune-affinity capture method using anti-epithelial cell adhesion molecule (EpCAM) antibody (Figure 4.4). Then the isolated cell from the biochip goes through whole genome amplification. This platform was named LIMO-seq,

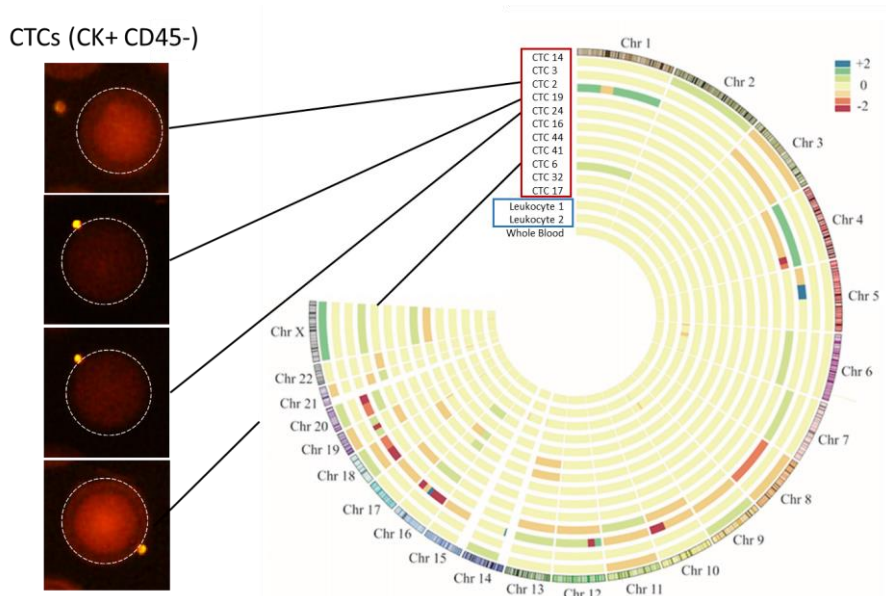


Figure 4.5 CTCs captured from a breast cancer patient show heterogeneity in the CNA plot drawn in circos format.

(an abbreviation for laser-induced isolation of microstructure on optomechanically-transferrable-chip and sequencing) to differentiate its value from PHLI-seq. Adopting the microfluidic chip developed by Nagrath et al. [74], CTCs were captured by the micropillars that are coated with anti-EpCAM antibodies. On top of the glass slide coated with the sacrificial layer, conventional photolithography was applied to generate an array of micro pillars with 100 μm diameter and 100 μm height. In microscale, the whole blood will flow in laminar manner. However, the micro pillar array would disturb the laminar flow, generating more chances for the CTCs to be captured onto the

microstructures. Then, the whole chip is imaged for CTC selection. All cells captured on the chip were stained with Hoechst dye, FITC labelled anti-CD45 antibody, and TRITC labelled anti-cytokeratin (CK) antibody, which stain nucleus of cells, lymphocytes, and cancer cells, respectively. The CK-positive and CD45-negative cells were selected with SLACS device and genomic DNA within was amplified with MDA for whole genome sequencing.

To demonstrate the clinical application of LIMO-seq, loaded 6 mL whole blood from a breast cancer patient who was diagnosed to have invasive lobular carcinoma and was aged 53. A pathology expert in The Seoul National University Hospital evaluated the patient to have negative expressions for ER, PR, and HER2 (Figure 4.5). Thirteen CTCs (CK+CD45-) were selected for analysis. As shown in Fig. 4b, the pathological results showing negative expressions for ER, PR, and HER2 have been well explained by the low copy number values on the corresponding gene loci in 12 of the analyzed single CTCs, except for one CTC (cell number 14) with relatively high copy number values on those loci. Such discrepancy in the one CTC might be attributed to a tumor heterogeneity in the patient, which could be too finely minor to accurately pinpoint by present pathology diagnosis method. This result strongly corroborates the accuracy and efficiency of LIMO-seq in capturing and analyzing various single CTCs in the genome level, which could represent the patient's tumor heterogeneity that could not be uncovered by conventional pathological test.

4.2. Applications of SLACS to spatial transcriptomics

Spatial transcriptomics, in contrast to spatial genomics, can reveal the functional landscape of the cells within the spatial context in gene expression level. For the tissue specimen, Fixed and Recovered Intact Single-cell RNA (FRISCR) was applied, which was developed by Thomsen et al. [75]. Because tissues are more complicated than the homogeneous pool of cells from one cell line, this strategy that utilizes a strong protease was applied to the tissues samples that went through SLACS.

Perhaps the most complex biological specimen with complexly functioning cells is the brain. For a proof-of-concept study, SLACS with Smart-seq2 was applied to a brain organoid sample. This study was a collaboration with Kyungsun Kang's group at Seoul National University. A brain organoid sample with two distinct layers was sectioned after OCT embedding. The targets were isolated from a densely populated area of cells and a less-densely populated area of cells (Figure 4.6). As demonstrated in a fluorescence image in Figure 4.6, the brain organoid section went through two different modalities of immunofluorescence staining. The purple immunofluorescence indicates a stem cell marker, while the green immunofluorescence indicates a neural cell marker. After cell isolation of SLACS from the section, the gene expression levels were acquired through SmartSeq2, and differentially expressed genes shown in the heatmap revealed two different clusters, which matched the

differentially stained regions.

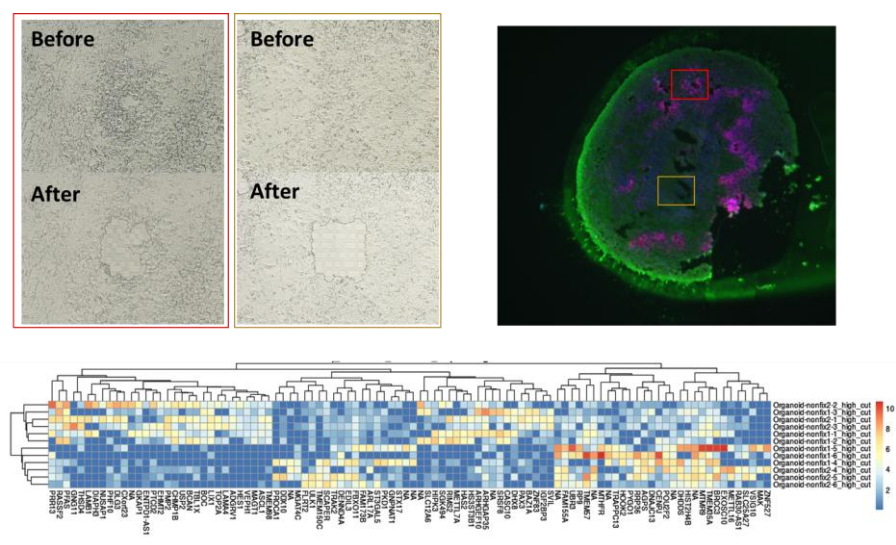


Figure 4.6 Cells from the two distinct layers from a brain organoid section with SLACS went through SmartSeq2.

To further demonstrate the spatial transcriptomics platform enabled with SLACS, I performed immune cell isolation from a immunofluorescence-stained breast cancer tissue section (Figure 4.7). The tissue was stained with anti-CD45 antibodies with FITC fluophores. CD45 is a surface protein that most immune cells express on their membranes and therefore serves as a pan-immune cell marker. Tumor microenvironment (TME) is a naïvely explored field of study for the technical difficulties to analyze. However, it is considered to be important and has potential to provide therapeutic implications. This is because the tumor microenvironment is the source for which the tumor cells

can grow on, and it is the battlefield for the immune cells to fight the abnormally growing cells.

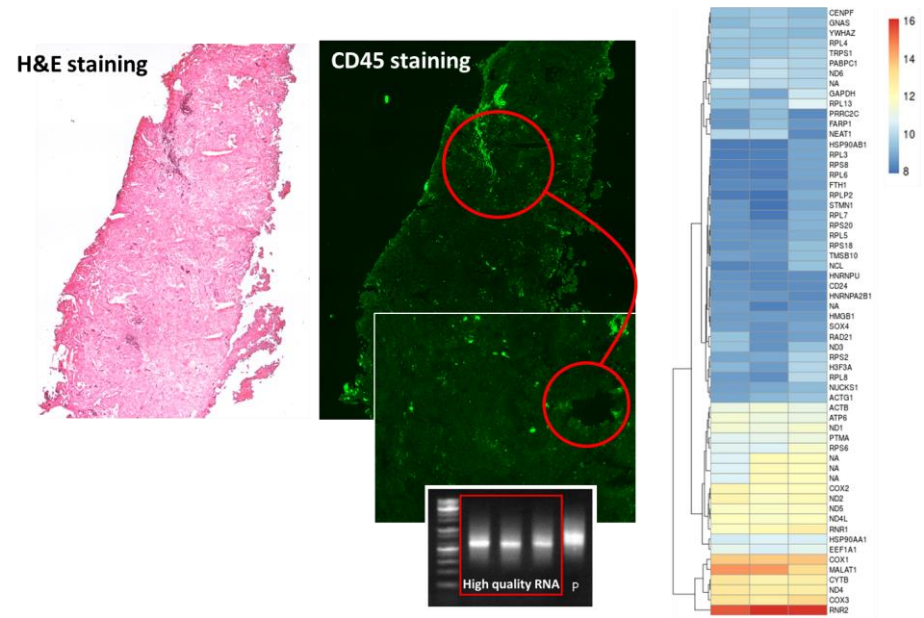


Figure 4.7 Highly expressed genes from immune cells isolated from breast cancer tissue section.

To analyze the TME of the breast cancer tissue section, we isolated areas stained with CD45, and attained gene expression profile in these cells that were embedded in the breast cancer tissue section. Some of the genes that were detected in this experiment were: CD44, CD74, CD46, CD59, CD164, CD9, CD63, CD276, CD302, CD55, CD151, CD83, CD82, CD47, CD81, CSF1, and MIF. These genes are some of the signature genes that are expressed in immune cells, providing further evidence that SLACS is able to isolate immune cells

with high specificity. Interestingly, CD9 is expressed in platelets, pre-B cells, eosinophils, basophils, activated T cells, endothelial cells, and epithelial cells. It has been described that CD9 is a unique marker signature gene for marginal zone B cells [76]. During the development of B cells, they migrate from the bone marrow, where they are generated, to the peripheral microenvironments. Although it is difficult to postulate that the isolated cells are B cells, all other gene expressions that were detected in the three immune cells indicate that they can be identified as functional activated immune cells. Because the adaptive immune system of the human body is able to develop B cells and antibodies that can respond to the activities of the cancer cells, it is important to study how and what are the naturally retained B cells are produced in cancer patients. For future study, SLACS will overcome the sampling issues in selecting the immune cells of interest within tissue context.

4.3. Applications of OPENchip and future perspectives with SLACS

Because the main purpose of development of SLACS was to bridge the gap between the spatial assay and omics assays, I sought to develop a spatial assay applicable to SLACS. By modifying the chip developed for whole genome sequencing of CTCs [31], we applied *in situ* molecular profiling technique, enabled by padlock probes [44]. We named this chip that enables the

in situ molecular profiling of the captured CTCs on chip as OPENchip (On-chip Post-processing Enabling chip). The scientific and clinical potential of RNA analysis of the rare and heterogeneous circulating tumour cell (CTC) analysis in single cell level has motivated the development of specialized rare-cell-enriching and analysing technologies[25], [77]–[81]. To overcome the rarity and heterogeneity, these techniques have integrated single cell analysis methods that can enrich single CTCs to molecular biotechnologies that can analyse the CTCs. These integrated single cell analysis methods can be categorized into two: methods that require single cell separation[77], [80], [81] and those that detect RNA expressions *in situ*[82]–[84].

When single cell separation methods are used, molecular biotechnologies such as the quantitative PCR or the next generation sequencing are followed so that RNA analysis can be performed in massively parallel and high-throughput. Few examples include fluorescence activated cell sorting (FACS) methods and microfluidic-based methods[80], [85]. After whole blood from a patient is centrifuged, heavier cells are labelled with appropriate molecules (i.e. antibodies with fluorescent molecules), FACS sorts out single CTCs, which are readily analysed through PCR or NGS[4]. However, the inherent problem with FACS lies in the poor yield, because not only the technique itself has low recovery rate, but also CTCs are already rare (approximately 10 to 100 CTCs per 25 million cells, or 5 mL in average patients with metastasis)[70], [74], [86], [87]. Although microfluidic-based CTC

enriching methods can be integrated to recover the yield, the two-step integration makes the procedure difficult-to-perform and leaves more rooms for the loss of the rare cells. Therefore, microfluidic-based and FACS-less CTC enriching platforms were developed utilizing immunoaffinity[74], density gradient centrifugation[88], inertial focusing[89], and much more. However, when used by themselves, these technologies only enable bulk sequencing of the captured CTCs, averaging out the RNA expression levels of the individual and heterogeneous CTCs[78], [82], [90]. Therefore, they were integrated to manual picking or other microfluidic devices to analyse single cells. When integrated, the single CTCs have to be processed to be analysed through quantitative PCR machines or NGS platforms, making the procedure complex. Moreover, while quantitative PCR produces too little information (i.e. usually single gene expression), the NGS technologies often produce too much information than what is necessary for patients in general clinics. Because the other non-extensively-studied profiles are unnecessary for many medical implementations, trading off undesired sequencing information provides advantages in simplicity, expense in cost, and accessibility. Also, sampling bias might produce inaccurate results and lose the nature of the true cell compartment targeted for expression profiling. Therefore, although the state-of-art biomolecular technologies enable extensive investigation of the CTCs, a simpler, yet high throughput platform that can assess RNA profiles in single CTCs is necessary for practical implementation of CTCs as a tool for medical

purposes.

Imaging-based single CTC analysis has been developed because the methods can address the simplicity and throughput issues. The Epic platform (Epic Sciences, San Diego) uses digital pathology and therefore provides solution of image-based CTC enumeration, but requires micromanipulations to analyse genetic profiles of single CTCs[91]. Fluorescence in situ hybridization (FISH) techniques that can be applied to mRNA expression profiling are representative examples that overcame this issue[82]–[84]. By hybridizing nucleic acid probes with fluorescence, cells on a glass slides are imaged in different fluorescence wavelengths. Although these platforms successfully demonstrated downstream CTC molecular analyses using FISH, the nature of hybridizing fluorophore-labelled DNA probes requires high-resolution fluorescence microscope and can only measure mRNA strand enumeration. However, for most cancers, along with the RNA expression variations, profiles of a few but critical single nucleotide variations (SNV) are directly related to diagnosis, prognosis, and medications. Thus, to implement lab-on-a-chip devices for clinical purposes, an integrated platform for the rare cell enrichment and appropriate RNA profiling is essential for high-throughput RNA screening in circulating tumour cells.

In this part of the chapter I demonstrate in situ RNA profiling of single circulating tumour cells on chip and applied the method to SK-BR3 cell line

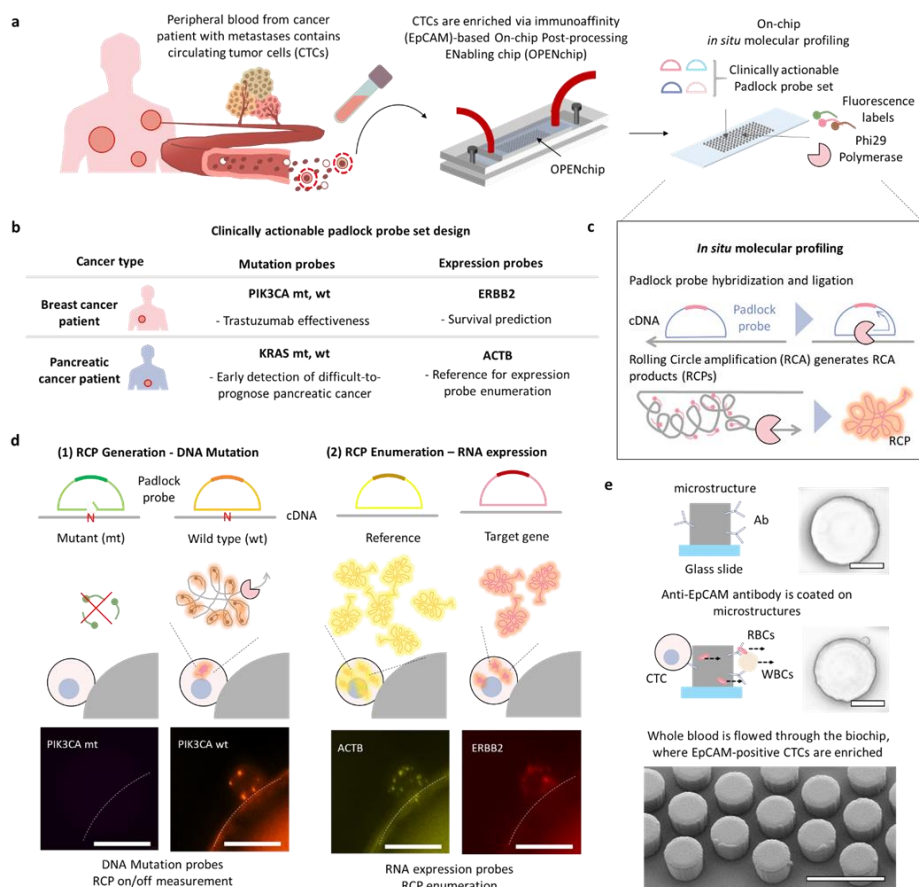


Figure 4.8 Rare CTCs are enriched with a biochip and used for DNA and RNA profiling *in situ* [44].

and whole blood samples from epithelial cancer patients (breast cancer patient and pancreatic cancer patient) with metastases. By applying padlock probe-based *in situ* RNA profiling technology [16], [47] to the CTCs captured on microstructures [74], we analysed HER2 and ACTB RNA expressions, and PIK3CA mutations on chip (Figure 4.8). The microstructures were fabricated

on normal glass slides with simple photolithography, while the *in situ* RNA profiling technology concerns of isothermal reaction, creating a blob of fluorophores as a result. We applied the proposed technique to whole blood samples from a breast cancer patient and a pancreatic cancer patient to demonstrate the applicability and practicality of the technique. Through the simple process by converging the two technologies, we envision that practical implementation of CTC molecular analysis will be made possible for cancer patients.

Because CTCs are rare, we first sought to enrich the CTCs before they can be analysed through *in situ* RNA profiling method. Among various methods designed to capture CTCs, we selected immunoaffinity-based capture of CTCs to demonstrate our platform to the most widely used method. Using our previously developed CTC enrichment chip [31] (Figure 4.9), with CTC capture efficiency of 98.7%, we first enriched cell line that can represent breast cancer, one of the major epithelial cancers. To investigate if the capture efficiency matches with that from the previous work, we infused SK-BR-3 cells-spiked whole blood samples into the chip. The chip was comprised of micro pillars that have diameters of 100 μm and heights of 100 μm . By criss-crossing the microstructures in an array, the laminar flow was perturbed to increase the probability of capture. After the photocurable polymer (SU-8) was spin-coated, the microstructures were directly fabricated on the normal glass slides using

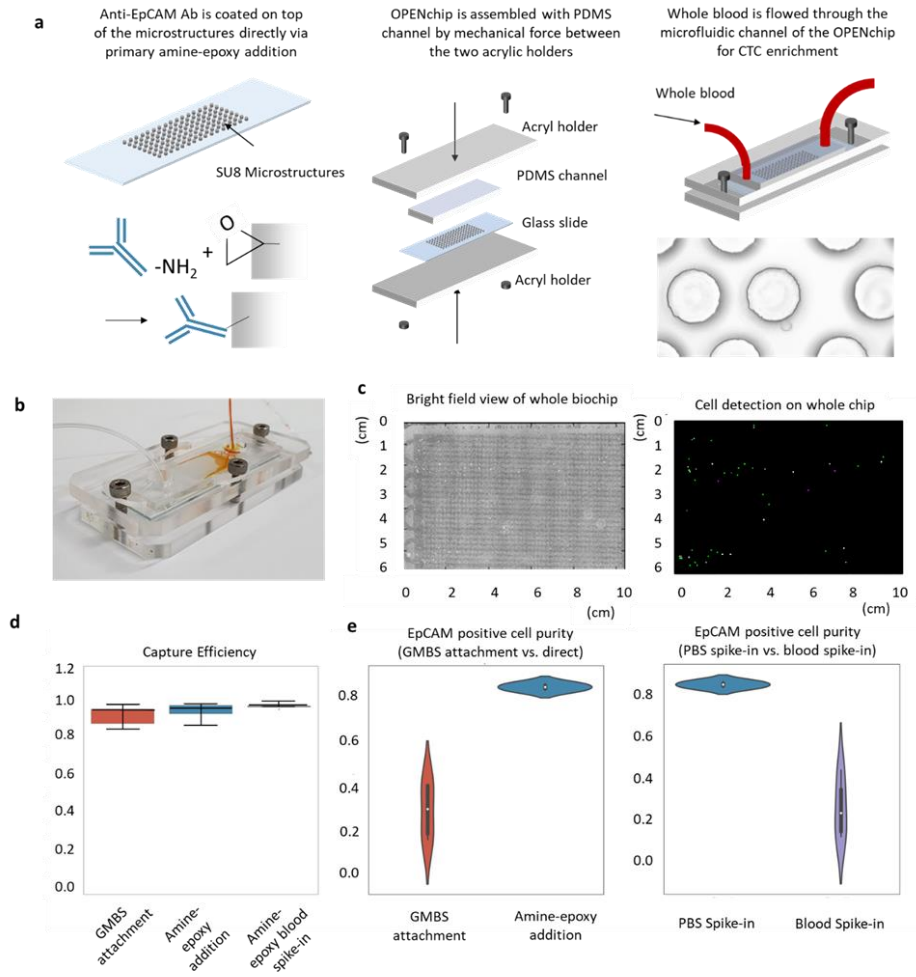


Figure 4.9 The OPENchip captures CTCs with high specificity and capture efficiency [44].

conventional photolithography system. After the spiked-in whole blood was infused in the chip, the whole blood residues were washed off using PBS. The cells then were fixed and stained with DAPI, cytokeratin antibody (CK), and

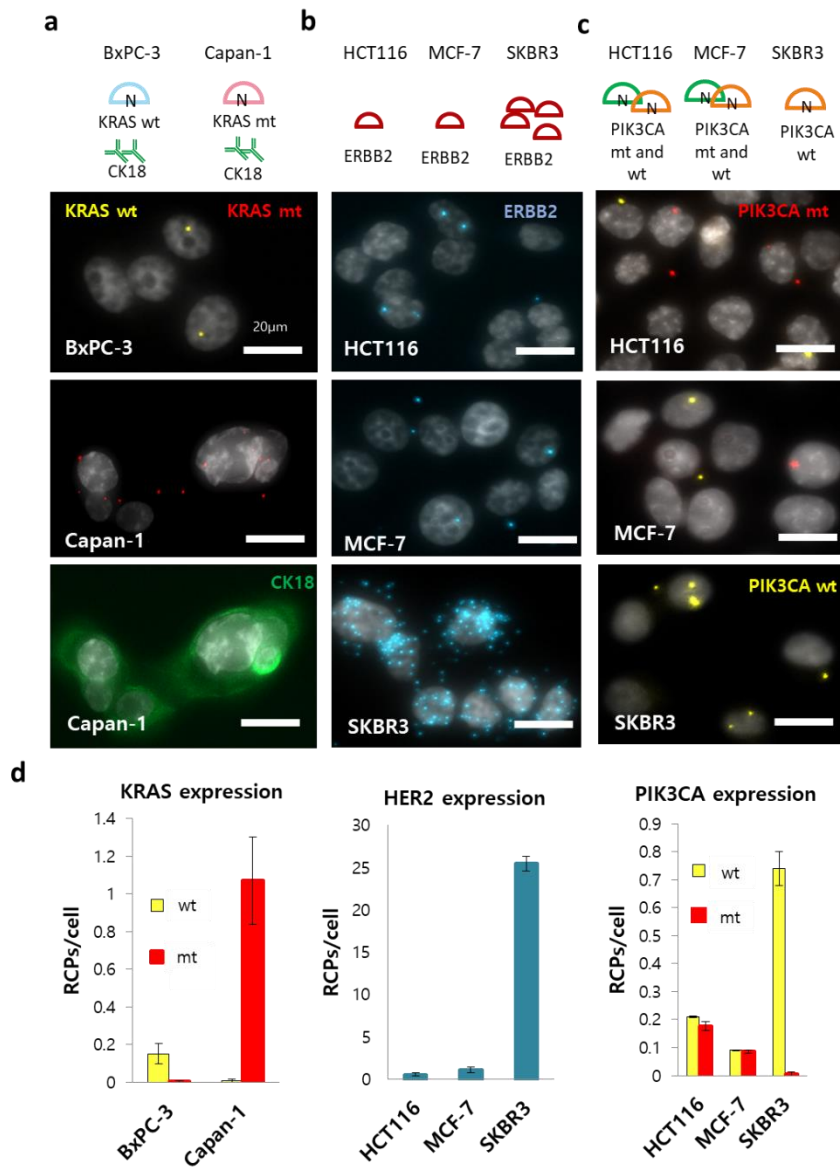


Figure 4.10 The padlock probes for *in situ* profiling are validated with different cell lines [44].

CD 45 antibody to confirm that the chip adequately captures the cells (Figure

4.10, Figure 4.11). DAPI staining was used to stain cell nucleus, while CK and CD45 antibodies were used to confirm if the cells were from epithelial origin or leukocyte origin, respectively.

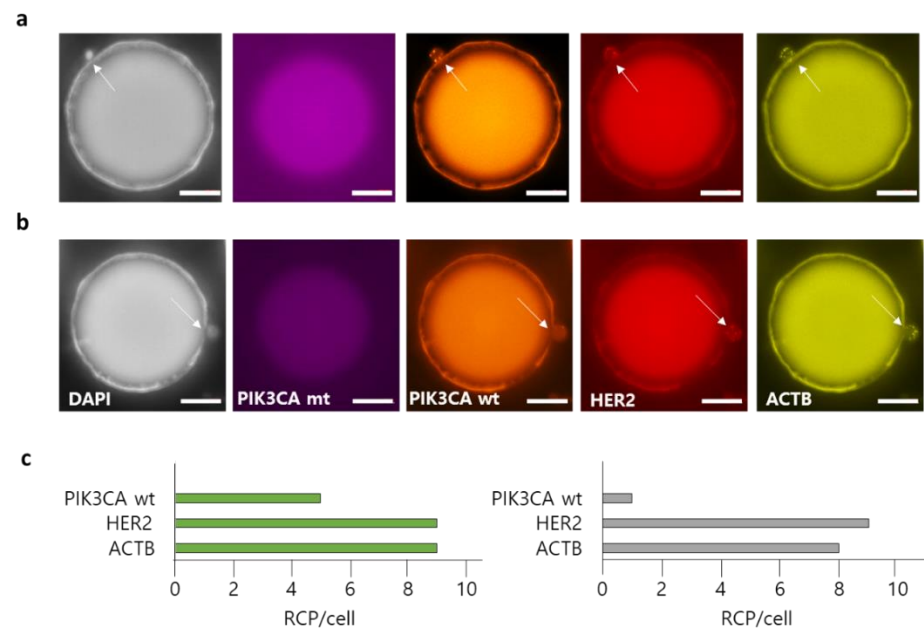


Figure 4.11 The cell lines are run through OPENchip for spike-in tests [44].

In order to validate the specificity and sensitivity of the PLPs used for the *in situ* RCA assay, we selected five cancer cell lines, spanning two pancreatic, two breast and one colon cancer, to perform molecular profiling. The chosen pancreatic cancer cell lines, BxPC-3 and Capan-1, represent the wild type *KRAS* and homozygous *KRAS G12V* mutation, respectively; the breast cancer cell line MCF-7 is HER2 non-amplified and heterozygous

PIK3CA E545K, and SKBR3 is HER2 amplified and *PIK3CA* wild type; the colon cancer cell line HCT116 is HER2 non-amplified and heterozygous *PIK3CA H1047R*. The expression profiles of the three genes were tested off-chip on glass slides and the results are seen in Fig. 2. The expression of the individual genes correlates with the characteristic features of the respective cell lines. The results of the performed RCA assay are given as the number of RCPs, each RCP representing individual RNA molecule per cell. The slide containing the Capan-1 cells were also subjected to CK18 antibody staining, thus reflecting on the combinatorial performance of PLP and immunofluorescence in the same assay. As an internal control, detection of ACTB was also performed in all cell lines (Figure 4.10), showing high expression.

A representative cell line from the above panel, namely SKBR3, was used for spiking the blood for infusion into the microfluidic chip. Using this, the performance of the breast cancer PLP assay was further tested on-chip as shown in Figure 4.11. The expression patterns correlate both with the characteristics of the chosen cell line as well as to that of the results from the assay performed off-chip. The reproducibility of the assay was tested with triplicates, showing the robustness of the method.

Blood samples from two breast cancer and two pancreatic cancer patients, with metastatic disease, were obtained for infusion into the chip. CTCs characterised by positive CK18 staining and/or expression of any of the genes from the chosen panel were enumerated for each patient sample. The breast

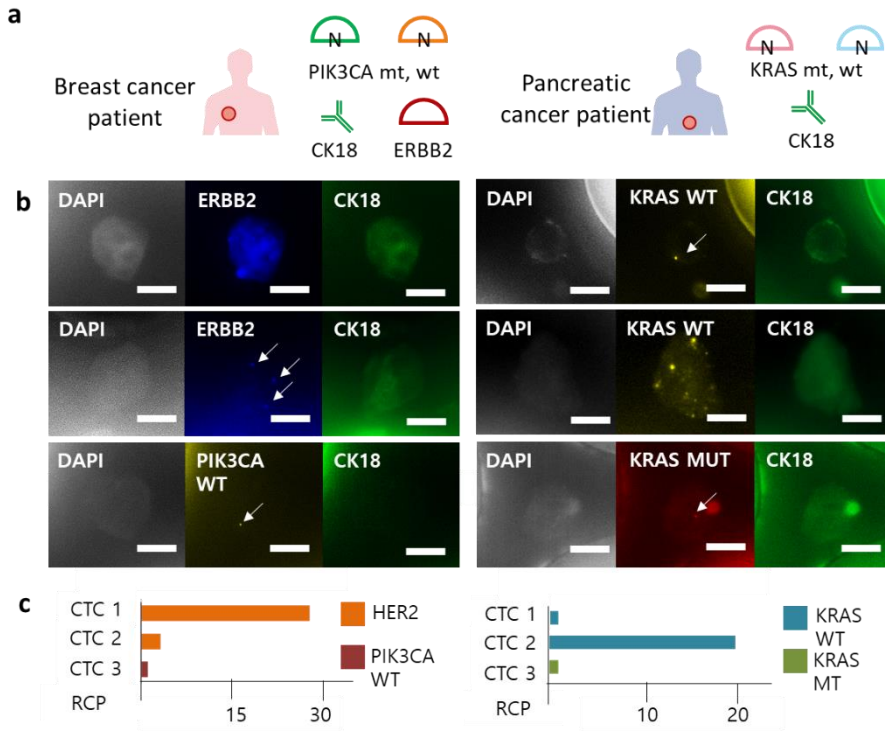


Figure 4.12 OPENchip data with cancer patient samples [44].

cancer patients showed a count of 15 and 36 CTCs, while the pancreatic cancer patients showed a count of 36 and 16. Images of *in situ* expression profiles of captured CTCs from one breast cancer and one pancreatic cancer patients are shown in Figure 4.12. RCP counts per cell range from 1 to 10 for HER2, and 1 to 8 for KRAS wild type; PIK3A is less abundantly expressed reflected by the single RCP count per cell. It should be noted that there is usually a significant loss of the capture efficiency during each step of the PLP assay, attributed to the intermittent washing steps of the method. This corresponds to almost half

the original cell number loaded on-chip. CK18 antibody staining showed both positive and negative patterns, however both being considered as CTCs, as also previously observed [43]. Heterogeneity was observed concerning the size of the captured CTCs, of which the diameter ranged from 20 to 40 μm . Such heterogeneity both in physical and molecular aspects become interesting for precise characterisation of single cells, in general and CTC, in specific.

The expression of the relevant markers targeted correlate with their biological significance and the disease state analysed. To briefly summarize the medical status of the breast cancer patient (female, age of 49), in 2013 July, the patient was diagnosed as invasive ductal carcinoma, clinical stage two (cT2N1), nuclear grade 3, histologic grade III (ER/PR/HER2, +95%/+80%/-). The patient was treated with neoadjuvant chemotherapy (4 cycles of Adriamycin and cyclophosphamide and then 4 cycles of docetaxel). In January of 2014, the patient went through right total mastectomy with sentinel lymph node biopsy and treated with tamoxifen. In March of 2016, the patient went through axillary lymph node dissection because of recurrence in the right axillary lymph node. At this point the patient was diagnosed ER/PR/HER2, 80%/5%/3+ and Ki67 15% through immunohistochemistry (IHC) testing. Because the patient was diagnosed as HER2+, she was treated with Taxol, cyclophosphamide, and Herceptin for two cycles. In June of 2016, liver metastasis was found and went through tissue biopsy. At this point the patient was diagnosed ER/PR/HER2 95%/1%/1+ and Ki67 3% through IHC testing. Therefore, the patient was

treated with Letrozole, but showed disease progress. Then, the patient was treated Aromasine and Everolimus, but showed disease progress. Finally, the patient was treated with Fulvestrant, but showed disease progress. Therefore, the cancer was resistant to several endocrine treatments. The blood was acquired in February of 2018, and the expression of HER2 and the mutation status of PIK3CA were analysed. Since HER2 is a prominent therapeutic target, determining its expression can be crucial for assessing treatment response and for strategizing on new HER2 directed therapies [92] . Our method aims to overcome the obstacles other RNA-based methods face because of the HER2 expression heterogeneity that is presented in the CTC population. PIK3CA predominant mutations in breast cancer have also been proven to be as heterogeneous and as highly relevant to therapy resistance therefore as important to profile in CTCs for future personalized therapeutic advances. CTCs with mutant PIK3CA have been reported to be enriched after chemotherapy [93].

The two pancreatic cancer patients analysed were diagnosed through TNM staging system, from which pancreatic cancer patient 1 (male, age of 76) was diagnosed as T2N1M0 and pancreatic cancer patient 2 (female, age of 80) was diagnosed to have tumour in pancreas (>2 cm in diameter) and regional lymph node metastasis (T2N2M0). For pancreatic cancer patients we analysed frequently mutated codon 12/13 of the KRAS gene, because of its 95% prevalence in pancreatic cancer [94], its big potential as a prognostic factor [95]

and the successful employment of the same target in our previous studies [43] that gave promising molecular profiling of pancreatic CTCs using other technologies. Future studies can implement *in situ* analysis of the primary and metastatic tumors. Antimetastatic therapies could be based on the individual evaluation of the CTCs' molecular profiles and their correlation with the primary or the metastatic tumour respective profiles [92], [96]

In the future, SLACS will be merged with *in situ* techniques similar to the OPENchip platform. Because *in situ* techniques have highly confined biochemistry that can analyze multiplex RNA. After profiling clinically actionable genetic aberrations from DNA and RNA simultaneously, SLACS can be used to isolate the analyzed CTC from the biochip to sequence them with NGS for massively parallel whole genome or whole transcriptome sequencing. Because SLACS can connect the spatial assays to the molecular assays, the developed *in situ* assay will provide useful technologies for the developed spatial genomics and transcriptomics technologies with SLACS, if combined.

Chapter 5. Conclusion and Discussion

In this Chapter, the proposed platform will be summarized. Then, the platform will be compared with the conventional technique in the aspect of throughput, spatial resolution, and degree of freedom. In addition, the limit of this platform will be described. Finally, future works that can make the platform will be presented.

5.1. Summary of dissertation

With the advancement of NGS and optics and semiconductor technologies, tools for spatial omics have been developed in many ways [8], [97]. In this dissertation, I described the importance and need for integrated analysis of spatial assays and biomolecular assays. Because each and every cell in a human body has different functions, which can be represented by different genetic molecules expressed, it is important to have tools that can enable life scientists and healthcare experts to look deeply inside these cells. To address such needs, an optical system was built and developed for its extensive uses in analyzing DNA and RNA within tissue context. Spatially-resolved Laser Activated Sorter (SLACS) was developed from previously developed Sniper Cloning technology, by adopting two automated X, Y, and Z axes and software to communicate between the machine and the user interface.

To demonstrate the uses of SLACS, I described its use for phenotype-based high-throughput laser-aided isolation and sequencing (PHLI-seq). SLACS was integrated with multiple displacement amplification (MDA) to address intra-tumoral heterogeneity within breast cancer tissue section from HER2+ breast cancer patients [33]. MDA utilizes Phi29 polymerase that has high strand displacement activity with replication properties. Specifically, using PHLI-seq, molecular subclones in a tumor were revealed and their evolutionary history was predicted. Using the same principle, the study using tumors from

24 triple negative breast cancer (TNBC) patients is being conducted to reveal the genomic landscape of the residual tumor from the patients. Also, when integrated with a biochip, SLACS was able to select out the targets based on the molecular features of the rare cells captured on the biochip [31]. Specifically, circulating tumor cells (CTCs) were captured on a biochip coated with anti-epithelial-cell-adhesion-molecule (EpCAM) antibodies. The anti-EpCAM antibodies bind to the EpCAM proteins that are expressed on the epithelial cells and not on the leukocytes that are present in the peripheral blood of the cancer patients. This technology was named LIMO-seq (Laser-induced Isolation of Microstructure on Optomechanically-transferrable-chip and sequencing). Processing a whole blood sample from a breast cancer patient, LIMO-seq was able to produce single CTC genomes from the cytokeratin positive and CD45 negative cancer cells. To further demonstrate the clinical and scientific potential of the LIMO-seq, experiments with different biochips and applying SLACS to them are being conducted.

Next, SLACS was applied to RNA sequencing technologies. Spatially resolved transcriptomics is important to understand how the cells are functioning and where they are located within the tissue. By combining SmartSeq2 with SLACS, full mRNA sequencing was conducted from the isolated targets. SmartSeq2 utilizes the poly-A tail capture using poly-T hybridization to the mRNAs and generates double stranded cDNA-RNA hybrid molecule that can readily be amplified by PCR with universal primers. The

universal primers can amplify the full transcriptome including the 3' end to the 5' end. Then, transposases are applied to the cDNA-RNA hybrid molecules to fragment and ligate the sequencing primers and barcodes to the fragmented molecules. SmartSeq2 was applied to a breast cancer tissue section to reveal the immune landscape of the breast cancer. Deciphering immune landscape within tissue context is important because the adaptive immune system interaction to the cancer microenvironment is a rigorous and complex process that is yet to be studied. For example, extremely diverse subtypes of B cells that have or will interact with the cancer cells exist within the tissue and in the peripheral blood. Therefore, the breast cancer tissue section was stained with anti-CD45 antibodies, and the fluorescent cells were isolated using SLACS. Then, SmartSeq2 was applied to the isolated cells.

Then, for spatial assay development, *in situ* molecular profiling was applied to the CTC-capturing biochip [44]. The OPENchip (On-chip post-processing enabling chip) utilizes the same CTC-capturing chip used for LIMO-seq. After the cells are captured, the cells were fixed and permeabilized for *in situ* molecular profiling of clinically actionable genes. To demonstrate the clinical potential for this lab-on-a-chip device, whole blood samples from breast cancer patients and pancreatic cancer patients were run through the chip, capturing CTCs for *in situ* profiling analysis. In the future this assay will be integrated with SLACS device for in-depth molecule analysis after the *in situ* profiling assay.

5.2. Comparison with previous technology

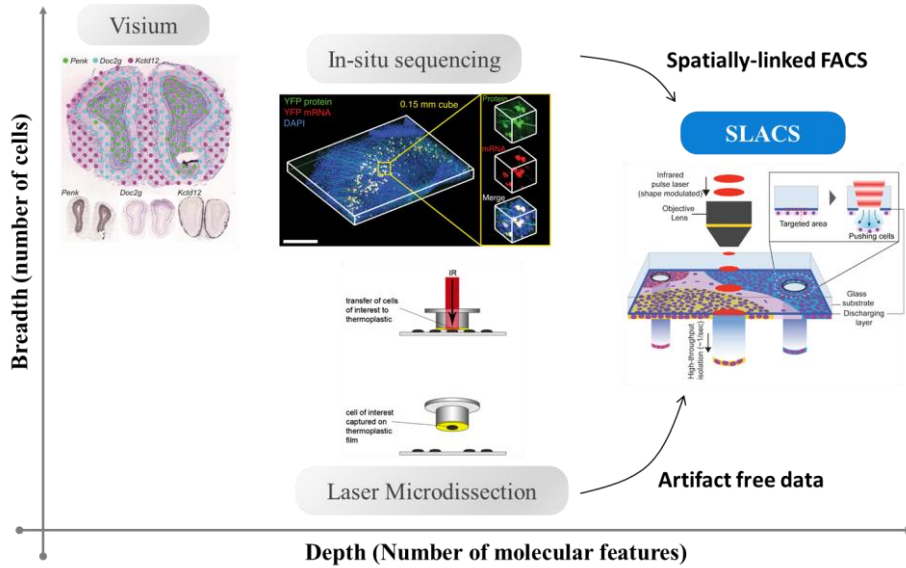


Figure 5.1 Compared to other technologies, SLACS is able to produce data from hundreds of single cells at once, producing less artifacts. Also, it is not confined to RNA analysis.

There are three parameters that were considered for the promotion of SLACS compared to other spatial omics technology: throughput, artifact, and freedom to operate. Visium from 10X Genomics, *in situ* sequencing from CARTANA, and Laser microdissection from Zeiss, Leica, and Fluidigm are the state-of-art commercialized technologies for spatial omics (Figure 5.1). Because the details of these technologies are addressed in the previous chapters, this section will only discuss the comparison between these technologies to SLACS. The throughput in terms of cell number analyzed in single run is the

highest in Visium technology and *in situ* sequencing technology (approximately thousand cells at once), and lowest for laser microdissection technologies (less than 10 cells at once). SLACS is situated in between, addressing approximately 100 cells at once. However, in contrast to the Visium technology and *in situ* sequencing technology, laser microdissection technology and SLACS has much higher freedom to operate. To elaborate, Visium technology has to work with poly-A capture of mRNAs from the tissue. *In situ* sequencing technology is also based on nucleic acid amplification procedures, allowing extensive analysis of nucleic acids with a few number of base reads. Laser microdissection and SLACS, however, isolates cells into normal PCR tubes and the contents within, such as proteins and nucleic acids, can be extensively analyzed for multiplex and simultaneous analysis of the genetic molecules that are inside the cells. However, conventional LCM technologies utilize UV light to push out the cells that are inside the microdissected target. The UV light renders the cells to be damaged and it is widely-known that the UV beam causes thymine dimer, nucleic acid fragmentation, and protein denaturation. However, because SLACS uses IR light, the damage is minimized [33].

5.3. Limit of the platform

The limitations in these platforms lie in their throughput as the previous section suggests. Although the throughput of the platform is much higher than conventional laser microdissection technologies, that of SLACS is

lower than that of the Visium technology and *in situ* sequencing technology. The basic throughput for the laser pulse is one target per second. However, all platforms addressed herein used 96 wells to retrieve the cells. Therefore, the retrieval capacity is the main bottleneck for the throughput of the platform. The research group is putting efforts to increase the retrieval capacity by upgrading the retrieval stage of the device. Also, with plans for commercialization, the research group plans to productize the SLACS device with upgraded features including the retrieval capacity and other minor issues.

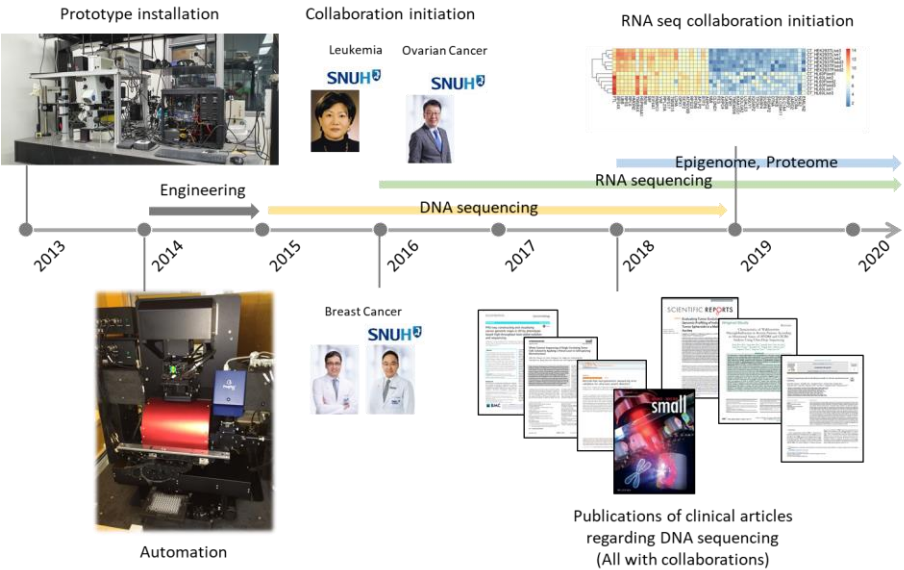


Figure 5.2 The development timeline for SLACS platform.

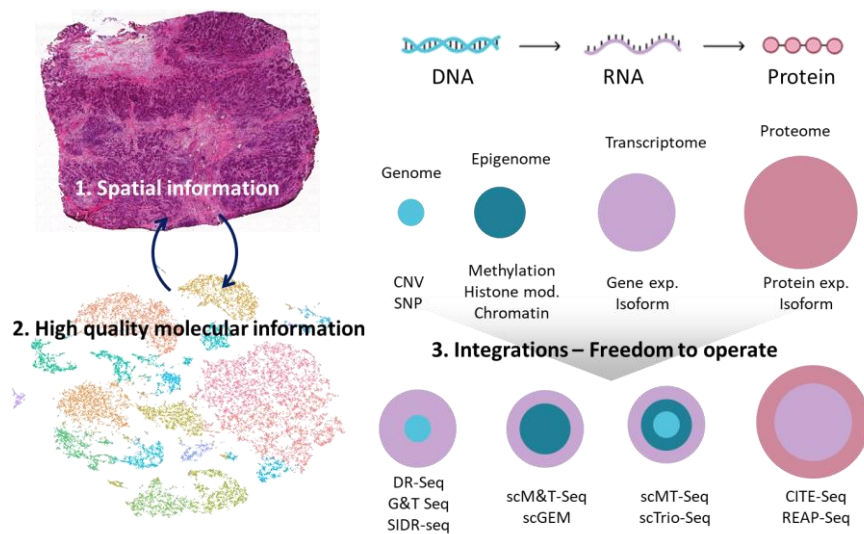


Figure 5.3 Integrations of SLACS to other technologies.

5.4. Future work

There are far still more validations to be performed for the utilization of SLACS with other molecular biology techniques (Figure 5.2 and Figure 5.3). NGS-based technologies such as epigenetic sequencing technologies and integrated technologies. For epigenome sequencing technologies, ATAC-seq (Assay for Transposase-Accessible Chormatin using sequencing) [98], Methylation sequencing, or bisulfite sequencing technologies can be applied to SLACS. Also, mass spectrometry technologies can provide useful for spatial proteomics when combined with SLACS. With optimizations and development, SLACS will prove useful for its high degree of freedom to operate. Furthermore, the integrated platforms are blooming out with advancement of single cell

isolation technologies. Technologies that integrate genomics and transcriptomics such as G&T seq [99] or SIDR-seq [100]; those that integrate epigenomics and transcriptomics such as scM&T-seq [101]; those that integrate genomics, epigenomics, and transcriptomics such as scMT-seq [101] or scTrio-seq [102]; and those that integrate proteomics and transcriptomics such as CITE-seq [103] or REAP-seq [104].

In terms of spatial assays, integrations of SLACS to platforms such as OPENchip assay can provide useful information. Such spatial assays include various types of ISH technologies such as FISH, FICTION, and other fluorescently labelled *in situ* hybridization technologies, *in situ* sequencing technologies, and classic staining methods that can imply pathological and histological information of the given tissue.

Together with the future integrations and development, I believe that SLACS will prove its potential for various spatial omics technology. With development and commercialization, SLACS will further will be developed as a powerful tool in addressing many scientific and clinical uses.

Bibliography

- [1] J. Craig Venter *et al.*, “The sequence of the human genome,” *Science* (80-.), vol. 291, no. 5507, pp. 1304–1351, Feb. 2001.
- [2] M. L. Metzker, “Sequencing technologies — the next generation,” *Nat. Rev. Genet.*, vol. 11, 2009.
- [3] “The human body at cellular resolution: the NIH Human Biomolecular Atlas Program,” *Nature*, vol. 574, no. 7777. Nature Publishing Group, pp. 187–192, 10-Oct-2019.
- [4] L. A. Herzenberg, R. G. Sweet, and L. A. Herzenberg, “Fluorescence-activated Cell Sorting,” *Scientific American*, vol. 234. Scientific American, a division of Nature America, Inc., pp. 108–118, 1976.
- [5] N. Navin *et al.*, “Tumour evolution inferred by single-cell sequencing,” *Nature*, vol. 472, no. 7341, pp. 90–94, 2011.
- [6] J. M. Segal *et al.*, “Single cell analysis of human foetal liver captures the transcriptional profile of hepatobiliary hybrid progenitors,” *Nat. Commun.*, vol. 10, no. 1, p. 3350, Dec. 2019.
- [7] E. Z. Macosko *et al.*, “Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets Resource Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets,” pp. 1202–1214, 2015.
- [8] A. C. Lee, Y. Lee, D. Lee, and S. Kwon, “Divide and conquer: A perspective on biochips for single-cell and rare-molecule analysis by next-generation sequencing,” *APL Bioeng.*, vol. 3, no. 2, p. 020901, Jun. 2019.
- [9] K. L. Frieda *et al.*, “Synthetic recording and in situ readout of lineage information in single cells,” *Nature*, vol. 541, no. 7635, pp. 107–111, Jan. 2017.
- [10] S. Hampel, P. Chung, C. E. McKellar, D. Hall, L. L. Looger, and J. H. Simpson, “Drosophila Brainbow: A recombinase-based fluorescence labeling technique to subdivide neural expression patterns,” *Nat. Methods*, vol. 8, no. 3, pp. 253–259, Mar. 2011.
- [11] J. Y. Kishi *et al.*, “SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues,” *Nat. Methods*, vol. 16, no. 6, pp. 533–544, Jun. 2019.
- [12] S. Shah, E. Lubeck, W. Zhou, and L. Cai, “seqFISH Accurately Detects Transcripts in Single Cells and Reveals Robust Spatial Organization in the Hippocampus,” *Neuron*, vol. 94, no. 4, pp. 752-758.e1, May 2017.
- [13] C.-H. L. Eng *et al.*, “Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+,” *Nature*, vol. 568, no. 7751, pp. 235–239, Apr. 2019.
- [14] J. H. Lee *et al.*, “Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues,” *Nat. Protoc.*, vol. 10, no. 3, pp. 442–458, Mar. 2015.
- [15] C. Xia, J. Fan, G. Emanuel, J. Hao, and X. Zhuang, “Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression,” *Proc. Natl. Acad. Sci. U. S. A.*, vol.

- 116, no. 39, pp. 19490–19499, Sep. 2019.
- [16] R. Ke *et al.*, “In situ sequencing for RNA analysis in preserved tissue and cells,” *Nat. Methods*, vol. 10, no. 9, pp. 857–60, Sep. 2013.
 - [17] R. Satija, J. a Farrell, D. Gennert, A. F. Schier, and A. Regev, “Spatial reconstruction of single-cell gene expression data,” *Nat. Biotechnol.*, vol. 33, no. 5, pp. 495–502, 2015.
 - [18] S. G. Rodriques *et al.*, “Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution,” *Science*, vol. 363, no. 6434, pp. 1463–1467, Mar. 2019.
 - [19] S. Vickovic *et al.*, “High-density spatial transcriptomics arrays for in situ tissue profiling,” *bioRxiv*, p. 563338, Feb. 2019.
 - [20] P. L. Ståhl *et al.*, “Visualization and analysis of gene expression in tissue sections by spatial transcriptomics,” *Science*, vol. 353, no. 6294, pp. 78–82, Jul. 2016.
 - [21] S. Picelli, O. R. Faridani, Å. K. Björklund, G. Winberg, S. Sagasser, and R. Sandberg, “Full-length RNA-seq from single cells using Smart-seq2,” *Nat. Protoc.*, vol. 9, no. 1, pp. 171–181, Jan. 2014.
 - [22] S. Picelli, Å. K. Björklund, B. Reinis, S. Sagasser, G. Winberg, and R. Sandberg, “Tn5 transposase and tagmentation procedures for massively scaled sequencing projects,” *Genome Res.*, vol. 24, no. 12, pp. 2033–2040, Dec. 2014.
 - [23] J. Chen, S. Suo, P. P. Tam, J.-D. J. Han, G. Peng, and N. Jing, “Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq,” *Nat. Protoc.*, vol. 12, no. 3, pp. 566–580, 2017.
 - [24] S. Nichterwitz *et al.*, “Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling,” *Nat. Commun.*, vol. 7, no. 1, p. 12139, Nov. 2016.
 - [25] L. Zhao *et al.*, “High-purity prostate circulating tumor cell isolation by a polymer nanofiber-embedded microchip for whole exome sequencing,” *Adv. Mater.*, vol. 25, no. 21, pp. 2897–2902, 2013.
 - [26] R. Jiang *et al.*, “A comparison of isolated circulating tumor cells and tissue biopsies using whole-genome sequencing in prostate cancer,” *Oncotarget*, 2015.
 - [27] S. M. Rothman *et al.*, “Human Alzheimer’s disease gene expression signatures and immune profile in APP mouse models: a discrete transcriptomic view of A β plaque pathology,” *J. Neuroinflammation*, vol. 15, no. 1, p. 256, Sep. 2018.
 - [28] D. A. Drew *et al.*, “Proximal Aberrant Crypt Foci Associate with Synchronous Neoplasia and Are Primed for Neoplastic Progression,” *Mol. Cancer Res.*, vol. 16, no. 3, pp. 486–495, Mar. 2018.
 - [29] S. Nichterwitz, J. A. Benitez, R. Hoogstraaten, Q. Deng, and E. Hedlund, “LCM-seq: A method for spatial transcriptomic profiling using laser capture microdissection coupled with PolyA-based RNA sequencing,” in *Methods in Molecular Biology*, vol. 1649, Humana Press Inc., 2018, pp. 95–110.
 - [30] S. Kim, A. C. Lee, H. Lee, J. Kim, and Y. Jung, “Constructing and Visualizing Cancer Genomic Maps in 3D Spatial Context by Phenotype-based High-throughput Laser-aided Isolation and Sequencing (PHLI-seq),” vol. 11,

- 2018.
- [31] O. Kim *et al.*, “Whole Genome Sequencing of Single Circulating Tumor Cells Isolated by Applying a Pulsed Laser to Cell-Capturing Microstructures,” *Small*, p. 1902607, Jun. 2019.
 - [32] J. Noh *et al.*, “High-throughput retrieval of physical DNA for NGS-identifiable clones in phage display library,” *bioRxiv*, p. 370809, Jul. 2018.
 - [33] S. Kim *et al.*, “PHLI-seq: constructing and visualizing cancer genomic maps in 3D by phenotype-based high-throughput laser-aided isolation and sequencing,” *Genome Biol.*, vol. 19, no. 1, p. 158, Dec. 2018.
 - [34] J. G. Gall and M. L. Pardue, “Formation and detection of RNA-DNA hybrid molecules in cytological preparations,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 63, no. 2, pp. 378–383, Jun. 1969.
 - [35] S. Shah, E. Lubeck, W. Zhou, and L. Cai, “seqFISH Accurately Detects Transcripts in Single Cells and Reveals Robust Spatial Organization in the Hippocampus,” *Neuron*, vol. 94, no. 4, pp. 752–758.e1, May 2017.
 - [36] C. H. L. Eng *et al.*, “Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+,” *Nature*, vol. 568, no. 7751, pp. 235–239, Apr. 2019.
 - [37] K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang, “Spatially resolved, highly multiplexed RNA profiling in single cells,” *Science* (80-.), vol. 348, no. 6233, 2015.
 - [38] C. Xia, H. P. Babcock, J. R. Moffitt, and X. Zhuang, “Multiplexed detection of RNA using MERFISH and branched DNA amplification,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–13, Dec. 2019.
 - [39] M. Kühnemund *et al.*, “Targeted DNA sequencing and in situ mutation analysis using mobile phone microscopy,” *Nat. Commun.*, vol. 8, p. 13913, Jan. 2017.
 - [40] B. Koos *et al.*, “Next-generation pathology - Surveillance of tumor microecology,” *J. Mol. Biol.*, vol. 427, no. 11, pp. 2013–2022, 2015.
 - [41] S. Ciftci *et al.*, “A novel mutation tolerant padlock probe design for multiplexed detection of hypervariable RNA viruses,” *Sci. Rep.*, vol. 9, no. 1, p. 2872, Dec. 2019.
 - [42] J. Svedlund *et al.*, “Generation of in situ sequencing based OncoMaps to spatially resolve gene expression profiles of diagnostic and prognostic markers in breast cancer,” *EBioMedicine*, Sep. 2019.
 - [43] A. El-Heliebi *et al.*, “In situ detection and quantification of AR-V7, AR-FL, PSA, and KRAS point mutations in circulating tumor cells,” *Clin. Chem.*, vol. 64, no. 3, pp. 536–546, 2018.
 - [44] A. C. Lee *et al.*, “OPENchip: An on-chip: In situ molecular profiling platform for gene expression analysis and oncogenic mutation detection in single circulating tumour cells,” *Lab Chip*, vol. 20, no. 5, pp. 912–922, Mar. 2020.
 - [45] X. Qian *et al.*, “Probabilistic cell typing enables fine mapping of closely related cell types in situ,” *Nat. Methods*, Nov. 2019.
 - [46] M. Asp *et al.*, “A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart,” *Cell*, 2019.
 - [47] C. Larsson *et al.*, “In situ genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes,” *Nat. Methods*, vol. 1, no. 3, pp. 227–232, 2004.

- [48] S. Ciftci *et al.*, “The sweet detection of rolling circle amplification: Glucose-based electrochemical genosensor for the detection of viral nucleic acid,” *Biosens. Bioelectron.*, p. 112002, Dec. 2019.
- [49] D. Wu *et al.*, “Profiling surface proteins on individual exosomes using a proximity barcoding assay,” *Nat. Commun.*, vol. 10, no. 1, p. 3854, Dec. 2019.
- [50] S. Ciftci *et al.*, “Digital rolling circle amplification-based detection of Ebola and other tropical viruses,” *J. Mol. Diagnostics*, Dec. 2019.
- [51] C. Strell *et al.*, “Placing <scp>RNA</scp> in context and space – methods for spatially resolved transcriptomics,” *FEBS J.*, vol. 286, no. 8, pp. 1468–1481, Apr. 2019.
- [52] X. Chen, Y.-C. Sun, G. M. Church, J. H. Lee, and A. M. Zador, “Efficient in situ barcode sequencing using padlock probe-based BaristaSeq,” *Nucleic Acids Res.*, vol. 46, no. 4, p. e22, Feb. 2018.
- [53] X. Wang *et al.*, “Three-dimensional intact-tissue sequencing of single-cell transcriptional states,” *Science (80-.)*, vol. 361, no. 6400, Jul. 2018.
- [54] J. H. Lee *et al.*, “Highly multiplexed subcellular RNA sequencing in situ,” *Science (80-.)*, vol. 343, no. 6177, pp. 1360–1363, Mar. 2014.
- [55] S. G. Rodrigues *et al.*, “Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution,” *Science*, vol. 363, no. 6434, pp. 1463–1467, Mar. 2019.
- [56] V. Espina *et al.*, “Laser-capture microdissection,” *Nat. Protoc.*, vol. 1, no. 2, pp. 586–603, 2006.
- [57] C. R. Merritt *et al.*, “Multiplex digital spatial profiling of proteins and RNA in fixed tissue,” *Nat. Biotechnol.*, vol. 38, no. 5, pp. 586–599, May 2020.
- [58] H. Lee *et al.*, “A high-throughput optomechanical retrieval method for sequence-verified clonal DNA from the NGS platform,” *Nat. Commun.*, vol. 6, p. 6073, Feb. 2015.
- [59] H. Yeom *et al.*, “Barcode-free next-generation sequencing error validation for ultra-rare variant detection,” *Nat. Commun.*, vol. 10, no. 1, p. 977, Dec. 2019.
- [60] N. Cho *et al.*, “High-throughput construction of multiple cas9 gene variants via assembly of high-depth tiled and sequence-verified oligonucleotides,” *Nucleic Acids Res.*, vol. 46, no. 9, May 2018.
- [61] Y. Choi, H. Choi, A. C. Lee, H. Lee, and S. Kwon, “Reconfigurable DNA Accordion Rack,” *Angew. Chemie Int. Ed.*, Jan. 2018.
- [62] H. Yeom *et al.*, “Cell-free bacteriophage genome synthesis using low cost sequence-verified array-synthesized oligonucleotides,” *ACS Synth. Biol.*, vol. 0, no. ja, May 2020.
- [63] J. Noh *et al.*, “High-throughput retrieval of physical DNA for NGS-identifiable clones in phage display library,” *MAbs*, vol. 11, no. 3, pp. 532–545, Apr. 2019.
- [64] H. Lee *et al.*, “A high-throughput optomechanical retrieval method for sequence-verified clonal DNA from the NGS platform,” *Nat. Commun.*, vol. 6, no. 1, p. 6073, Dec. 2015.
- [65] N. Navin *et al.*, “Tumour evolution inferred by single-cell sequencing,” *Nature*, vol. 472, no. 7341, pp. 90–94, 2011.
- [66] H. Telenius, N. P. Carter, C. E. Bebb, M. Nordenskjöld, B. A. J. Ponder, and

- A. Tunnacliffe, "Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer," *Genomics*, vol. 13, no. 3, pp. 718–725, Jul. 1992.
- [67] C. Zong, S. Lu, A. R. Chapman, and X. S. Xie, "Genome-wide detection of single-nucleotide and copy-number variations of a single human cell," *Science*, vol. 338, no. 6114, pp. 1622–6, 2012.
- [68] M. Hagemann-Jensen *et al.*, "Single-cell RNA counting at allele and isoform resolution using Smart-seq3," *Nat. Biotechnol.*, pp. 1–7, May 2020.
- [69] F. Kruse, J. P. Junker, A. van Oudenaarden, and J. Bakkers, "Tomo-seq: A method to obtain genome-wide expression data with spatial resolution," *Methods Cell Biol.*, vol. 135, pp. 299–307, Jan. 2016.
- [70] ASHWORTH and TR., "A case of cancer in which cells similar to those in the tumours were seen in the blood after death," *Aust Med J.*, vol. 14, p. 146, 1869.
- [71] G. Rossi and M. Ignatiadis, "Promises and pitfalls of using liquid biopsy for precision medicine," *Cancer Research*, vol. 79, no. 11. American Association for Cancer Research Inc., pp. 2798–2804, 01-Jun-2019.
- [72] K. C. Andree, G. van Dalum, and L. W. Terstappen, "Challenges in circulating tumor cell detection by the CellSearch system," *Mol. Oncol.*, no. December 2015, pp. 1–13, 2015.
- [73] C. Alix - Panabieres, H. Schwarzenbach, and K. Pantel, "Circulating tumor cells and circulating tumor DNA," *Annu Rev Med*, vol. 63, pp. 199–215, 2012.
- [74] S. Nagrath *et al.*, "Isolation of rare circulating tumour cells in cancer patients by microchip technology," *Nature*, vol. 450, no. 7173, pp. 1235–1239, Dec. 2007.
- [75] E. R. Thomsen *et al.*, "Fixed single-cell transcriptomic characterization of human radial glial diversity," *Nat. Methods*, vol. 13, no. 1, 2015.
- [76] W.-J. Won and J. F. Kearney, "CD9 Is a Unique Marker for Marginal Zone B Cells, B1 Cells, and Plasma Cells in Mice," *J. Immunol.*, vol. 168, no. 11, pp. 5605–5611, Jun. 2002.
- [77] D. T. Miyamoto *et al.*, "RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance," *Science (80-.)*, vol. 349, no. 6254, 2015.
- [78] Y. Zhang *et al.*, "Single-Cell Codetection of Metabolic Activity, Intracellular Functional Proteins, and Genetic Mutations from Rare Circulating Tumor Cells," *Anal. Chem.*, vol. 87, no. 19, pp. 9761–9768, Oct. 2015.
- [79] D. T. Ting *et al.*, "Single-Cell RNA Sequencing Identifies Extracellular Matrix Gene Expression by Pancreatic Circulating Tumor Cells," *Cell Rep.*, vol. 8, no. 6, pp. 1905–1918, 2014.
- [80] J. G. Lohr *et al.*, "Genetic interrogation of circulating multiple myeloma cells at single-cell resolution," vol. 147, 2016.
- [81] a F. Sarioglu *et al.*, "A microfluidic device for label-free, physical capture of circulating tumor cell clusters.," *Nat. Methods*, vol. 12, no. April, pp. 1–10, 2015.
- [82] M. Dhar *et al.*, "Label-free enumeration, collection and downstream cytological and cytogenetic analysis of circulating tumor cells," 2016.

- [83] M. A. Leversha *et al.*, “Fluorescence in situ hybridization analysis of circulating tumor cells in metastatic prostate cancer,” *Clin. Cancer Res.*, vol. 15, no. 6, pp. 2091–7, Mar. 2009.
- [84] J. F. Swennenhuis, A. G. J. Tibbe, R. Levink, R. C. J. Sipkema, and L. W. M. M. Terstappen, “Characterization of circulating tumor cells by fluorescence in situ hybridization,” *Cytom. Part A*, vol. 75A, no. 6, pp. 520–527, Jun. 2009.
- [85] X. Zhang, S. L. Marjani, Z. Hu, S. M. Weissman, X. Pan, and S. Wu, “Single-Cell Sequencing for Precise Cancer Research: Progress and Prospects,” *Cancer Res.*, pp. 1–9, 2016.
- [86] N. Aceto, “Circulating tumor cell clusters are precursors of breast cancer metastasis,” *Aacr*, vol. 158, no. 5, p. Abstract #LB-192, 2014.
- [87] K. Pantel and M. R. Speicher, “The biology of circulating tumor cells,” *Oncogene*, vol. 35, no. 10, pp. 1216–1224, Mar. 2016.
- [88] C. E. Yoo *et al.*, “Vertical Magnetic Separation of Circulating Tumor Cells for Somatic Genomic-Alteration Analysis in Lung Cancer Patients,” *Nat. Publ. Gr.*, 2016.
- [89] E. Sollier *et al.*, “Size-selective collection of circulating tumor cells using Vortex technology,” *Lab Chip*, vol. 14, no. 1, pp. 63–77, 2014.
- [90] B. Polzer *et al.*, “Molecular profiling of single circulating tumor cells with diagnostic intention,” *EMBO Mol Med*, vol. 6, pp. 1371–1386, 2014.
- [91] A. E. Dago *et al.*, “Rapid Phenotypic and Genomic Change in Response to Therapeutic Pressure in Prostate Cancer Inferred by High Content Analysis of Single Circulating Tumor Cells,” *PLoS One*, vol. 9, no. 8, p. e101777, Aug. 2014.
- [92] S. Riethdorf *et al.*, “Detection and HER2 expression of circulating tumor cells: Prospective monitoring in breast cancer patients treated in the neoadjuvant GeparQuattro trial,” *Clin. Cancer Res.*, vol. 16, no. 9, pp. 2634–2645, 2010.
- [93] M. Janiszewska *et al.*, “In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer,” *Nat. Genet.*, vol. 47, no. 10, pp. 1212–1219, 2015.
- [94] K. L. Bryant, J. D. Mancias, A. C. Kimmelman, and C. J. Der, “KRAS: feeding pancreatic cancer proliferation,” *Trends Biochem. Sci.*, vol. 39, no. 2, pp. 91–100, Feb. 2014.
- [95] L. Tao, L. Zhang, D. Xiu, C. Yuan, Z. Ma, and B. Jiang, “Prognostic significance of K-ras mutations in pancreatic cancer: a meta-analysis,” *World J. Surg. Oncol.*, vol. 14, no. 1, p. 146, 2016.
- [96] A. Lyberopoulou *et al.*, “Mutational analysis of circulating tumor cells from colorectal cancer patients and correlation with primary tumor tissue,” *PLoS One*, vol. 10, no. 4, pp. 1–12, 2015.
- [97] M. Asp, J. Bergenstr hle, and J. Lundeberg, “Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration,” *BioEssays*, p. 1900221, May 2020.
- [98] J. D. Buenrostro, B. Wu, H. Y. Chang, and W. J. Greenleaf, “ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide,” *Curr. Protoc. Mol. Biol.*, vol. 109, no. 1, pp. 21.29.1–21.29.9, Jan. 2015.
- [99] S. S. Dey, L. Kester, B. Spanjaard, M. Bienko, and A. Van Oudenaarden,

- “Integrated genome and transcriptome sequencing of the same cell,” *Nat. Biotechnol.*, vol. 33, no. 3, pp. 285–289, Jan. 2015.
- [100] K. Y. Han *et al.*, “SIDR: simultaneous isolation and parallel sequencing of genomic DNA and total RNA from single cells,” *Genome Res.*, vol. 28, no. 1, pp. 75–87, Jan. 2018.
 - [101] Y. Hu *et al.*, “Simultaneous profiling of transcriptome and DNA methylome from a single cell,” *Genome Biol.*, vol. 17, no. 1, p. 88, May 2016.
 - [102] Y. Hou *et al.*, “Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas,” *Cell Res.*, vol. 26, no. 3, pp. 304–319, Mar. 2016.
 - [103] M. Stoeckius *et al.*, “Simultaneous epitope and transcriptome measurement in single cells,” *Nat. Methods*, vol. 14, no. 9, pp. 865–868, Sep. 2017.
 - [104] V. M. Peterson *et al.*, “Multiplexed quantification of proteins and transcripts in single cells,” *Nat. Biotechnol.*, vol. 35, no. 10, pp. 936–939, Oct. 2017.

국문 초록

본 학위 논문에서는 이 논문에서는 SLACS (Spatially-resolved Laser Activated Cell Sorting) 기술이 도입되었으며 유전체학 및 전사체학에 대한 응용이 시연되었다. 모든 생물학적 덩어리는 생물학적 세포로 구성되며, 각각의 세포는 DNA 또는 RNA와 같은 유전자 분자로부터 얻은 수십억 바이트의 데이터를 포함한다. 휴먼 게놈 프로젝트가 10 년 안에 한 사람의 게놈을 시퀀싱 한 후, 차세대 시퀀싱 (NGS)과 관련되는 대규모 병렬 시퀀싱 기술은 생물학의 혁신을 불러 일으켜 생물학에 대한 통찰력을 제공하고 진단 및 치료에서 혁명을 일으켰다. 그러나 이 기술들은 이중 유전자 분자의 풀에만 적용 할 수 있으며, 생물 표본 내의 다른 세포에서 유전자 지형의 철저한 탐색을 방해하였다. 따라서, 세포 풀에서 각각의 모든 세포를 따로 분리하려는 노력은 수많은 단일 세포 분리 방법론을 생성하였으며, 이는 미세유체학, 마이크로 어레이 및 광학을 사용하여 세포를 분리하는 것의 세 가지로 분류 될 수 있다.

일반적으로 수 마이크로미터에서 수십 마이크로미터에 이르는 생물학적 세포 크기 때문에 단일 기술의 진보는 단일 세포

조작에 이점을 제공 하였다. 미세 유체 특성을 이용하는 최첨단 세포 분리 기술이 빠르게 상용화되어 한 번에 수백에서 수천 개의 단일 세포를 처리 할 수 있는 고 처리량 단일 세포 분석이 가능해졌습니다. 이들은 미세 분자 챔버 또는 피코 리터 액적에서 세포 해리 및 구획화를 이용하며, 여기서 생체 분자 기술은 원하는 유전자 분자를 증폭시킬 수 있다. 단일 세포의 게놈 또는 전 사체와 같은 증폭 된 생성물은 NGS를 통해 시퀀싱되어, 해리 된 세포가 생체 시편에서 어떻게 기능하는지에 대한 통찰력을 제공한다. 그러나, 원래 서로 부착 된 세포의 해리 과정은 가속할 수 있으며, 서로 상호 작용하는 표면 단백질이 분해 될 것을 요구한다. 이 공정은 전지 상태가 용매 내에서 해리되기 전에 동일한 지에 대해 많은 의문을 제기했다. 따라서, 폴리아데노신 꼬리 또는 폴리 (A) 꼬리를 포획 할 수 있는 화학적으로 합성 된 올리고 뉴클레오티드의 마이크로 어레이는 생물학적 표본으로부터 메신저 RNA (mRNA)를 직접 포획하기 위해 개발되었다. 그러나, 이들 기술은 화학적 DNA 합성 기술의 기술적 한계 및 스폿 간의 교차 오염으로 인해 올리고 뉴클레오티드 스폿의 큰 해상도를 요구한다. 생체 시료로부터 세포의 광학적 분리는 관심 대상 영역을 원하는 수신기로 전달하기 위해 레이저를 이용하는 종래의 레이저 캡처

미세 해부 (LCM) 장치로 광범위하게 조사되었다. 그러나, 이들은 자외선 (UV) 레이저를 사용하여 생체 내 분자에 크게 손상을 줄 수 있는 원하는 영역을 만들거나 근적외선 (IR) 레이저를 사용하여 녹일 수 있고 추가 생물학적 물질을 위해 원하는 관심 영역을 전달할 수 있는 열가소성 수지를 사용한다. 분석. 그러나 열가소성 방식은 종종 교차 오염을 유발하고 시편을 접촉 방식으로 분리해야하기 때문에 처리량이 낮다.

이 논문에서는 광학 셀 분류기 또는 낮은 손상과 높은 처리량으로 셀을 광학적으로 분리 할 수 있는 펄스 형 근적외선 레이저를 사용하는 SLACS (공간적으로 해결 된 레이저 활성화 셀 분류기)의 개발에 대해 설명하였다. 이 새로운 장치의 엔지니어링 프로세스와 NGS 기술의 두 소프트웨어 및 응용 프로그램에 대해 설명하였다. 또한, 게놈 및 전 사체에 대한 SLACS의 적용이 입증되었다. SLACS의 향후 응용에 대한 개념 증명 연구도 설명하였다.

주요어: 레이저, 세포 분류기, 공간 오믹스, 공간 유전체학, 공간 전사체학

학번: 2014-21548