PH.D. DISSERTATION

Single Cell Genome Analysis Platform Development in Blood Sample

채취 혈액의 단일 세포 유전체 분석 플랫폼 개발

BY

YUSHIN JUNG

2017

DEPATRTMENT OF ELECTRICAL AND
COMPUTER ENGINEERING
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Abstract

In this thesis, I introduce the development of single cell genome analyzing platform for blood sample. By using this platform, I observed abnormality of down syndrome by analyzing genome of 5 cells from real patient blood. The number of cells required for genome analysis is important because less number implies high possibility of detecting unique markers. To develop this platform, I improved targeted single cell isolating method and optimized whole genome amplification method.

First, to isolate targeted cell, we need to observe the cell through microscope and isolate them right away. Conventional method burns surrounding of the targeted cells laid on polyethylene naphthalate film. I used indium tin oxide as a sacrificial layer and infrared pulse laser for intact cell isolation. At the same time, since this sacrificial layer has high physical and chemical durability similar to slide glass, modification of general protocol to observe cells are not required.

Optimization of whole genome amplification improved uniformity of amplification of low input genome. Amplifying whole genome with low amplification bias is critical for high quality of genome analysis. Conventional protocol of whole genome amplification using isolated cells were not enough for high quality genome analysis. By optimal cell retrieving, cell lysis and pre-denaturation step improved the amplification efficiency. As the result, we could obtain uniform amplification product.

In order to confirm this platform is applicable for various usage, I demonstrated with 2 cell lines and 2 blood sample from hereditary disorder patients. Since staining and storing decrease quality of genome which result in low uniformity of amplification, we had to find minimum cell number for reliable amplification product. We measured initial amplification rate, PCR success rate and positive rate of target sequencing and directly analyzed genome coverage by low depth whole genome sequencing. We found out 25 cells were minimum cell number required for reliable genome analysis using giemsa stained cells. I believe this novel platform is ready to use for wide usage for discovering biologically significant points, which was unrevealed by high noise from heterogeneous cell population.

keywords : single cell isolation, sacrificial layer, whole genome amplification, genome coverage, amplification bias

Student Number : 2012-20863

Contents

List of Figures

immediately, after freezing and thawing 20 times, storing at room temperature for 8 days, storing at −20 degrees for 9 days. For each cell line, 5, 25 and 100 cells were isolated.

Figure 4.2. 2 Genome coverage of WGA products using blood sample. Graphs of 25 cells isolated after storing for 8 days at −20 degrees, 25 cells isolated after freeze and thaw 20 times, 25 and 100 cells isolated after immediate arrival to laboratory.

Figure 4.2. 3 CNV plot of WGA DS samples.

Figure 4.2. 4 CNV plot of WGA CATCH 22 samples.

Chapter 1

Introduction

The cell is the basic biological unit of all known living organisms. So, to cure major human life threatening diseases such as cancer and hereditary disorder, researchers need to understand functions of cells. Inside the human cell 3 billion of DNA encode the functions of the cell; how the cell to proliferate, interact with their surroundings and even go into apoptosis cycle to end its function. Therefore, analyzing genomes of the cell is closely related to treating the chronic and acute disease.

For example, fluorescent in-situ hybridization (FISH) which is

frequently used method to diagnose cancer is based on observing chromosome abnormalities: translocation, inversion, deletion and duplication. Doctors can identify some types of leukemia, lymphoma and sarcomas based on FISH result[1]. Moreover, rapid accumulation of genomic databases made pharmacogenomics to emerge for disease treatment based on genome analysis[2].

But still, complex diseases, such as cancer, shows lack of clinical success using genomic information. This is because cancer cell proliferates and differentiates rapidly which results in imperfect genome replication. Imperfect genome replication ends up with heterogeneous cell clones having different genomes. Conventional method collects all of these heterogeneous clones and use for genome analysis. This approach average out small number of mutations in minor clones' genome which may be critical to living organisms[3]. Minor clone of the cancer cell which was not identified at the first time is frequently reported as the cause of recurrence of the cancer. Recurrence of the cancer is fatal to survival of the patient[4]. Therefore, diagnosing minor clones from heterogeneous cancer is closely related to patient's survival rate. In addition, appropriate treatment by identifying minor clones helps patient to

save immense medical expense. Inefficient treatment not only slowdowns the recover but also burdens patient financially which make them to give up the cure.

In this thesis, I introduce minor clone targeting genome analyzing platform. This platform is based on pulse laser catapulting system which gently transfers targeting cell from visualizing space to biomolecule analyzing spot. For best genome analyzing quality every step of preparing genome sequencing are optimized such as cell lysis and genome amplification.

Chapter 1 introduces commercialized single cell isolation techniques and genome amplification methods; single cell isolating techniques include fluorescence activated cell sorter (FACS), C1 from Fluidigm and laser capture microdissection (LCM), and genome amplification method includes PCR based genome amplification, multiple displacement amplification (MDA), and multiple annealing and looping based amplification cycles (MALBAC).

Laser catapulting based single cell isolation platform is introduced in chapter 2. Our team tried to isolate cell on bare glass and sacrificial layer coated glass. After we found out indirect laser

isolation is necessary for intact cell isolation, we isolated various types of cell condition and validated my platform.

In chapter 3, optimized genome amplification protocol is introduced. Since genome of the single cell level (~10 pg) is not enough to use for whole genome sequencing, genome amplification is necessary after isolation. However, many information can be lost at this step due to amplification bias. Therefore, we tried to optimize genome amplification step focusing on minimizing amplification bias and at the same time maintaining enough quantity for sequencing. Amplified genome quality was compared based on real-time amplifying efficiency and PCR success rate from random positions of genome.

Single cell genome analyzing platform was validated using cell lines and patients' blood sample in chapter 4. Since we targeted blood sample to use on this platform, cell lines were leukemia cell line (acute promyelocytic leukemia cell (HL60) and chronic myelogenous leukemia (K562)) and patients' sample was blood sample. Random location PCR, target sequencing, and genome wide analysis such as copy number variation (CNV) and genome coverage was processed for validation. Applying on patients' sample is much

tougher than applying on cell line, since quality of the patients' sample varies on their storing condition. Therefore, storing condition was modified to use for validating this platform.

In conclusion, I have summarized my work and suggest further works for various applications.

By using this platform, many researchers can save time and money for increasing signal-to-noise ratio(SNR) to see minor clones' unique genome. Pathologists could directly see the genome of their targeted cells while diagnosing patients' sample. As the result, patients who are suffering from disease such as cancer which is composed of heterogeneous clones will get accurate prescription for their better healthy life. I believe platform depicted here can solve mysteries of diverse cell behavior and accelerate developing medicines and instruments to help people suffering without efficient treatment.

1.1 Single Cell Isolation

Bulk analysis has low sensitivity on rare mutations due to low SNR. Single cell level genome analysis can increase SNR since portion of rare mutation grows[5]. Furthermore, by only collecting the particular cells and use for genome analysis, sensitivity will increase much higher. Therefore, many researchers tried to isolate targeted cells for further analysis.

Fluorescence activated cell sorter(FACS) is the most well-known single cell isolation technique[6]. Firstly, user tags the target cell with fluorescent label from the cell suspension. (Fig. 1.1) After fluorescent label is tagged whole suspension flow through narrow streamline. This streamline can break into droplets with a low probability of more than one cell per droplet. If there is wanted fluorescence signal in the droplet it means targeted cell is in that droplet. Then user can collect the targeted cell with electrostatic deflection system in the droplet. Thousands of targeted single cells can be isolated for further experiments. Recent developments such

as quantum dot allows maximum number of distinguishable fluorescent labels to be 17 or 18[7].

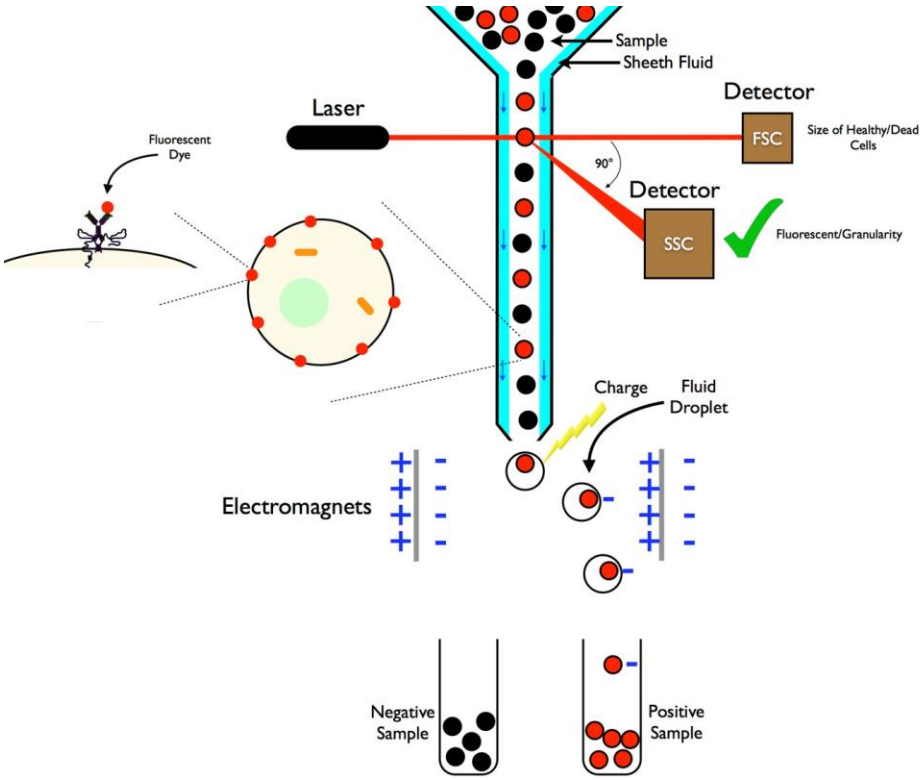


Figure 1. 1 Fluorescence activated cell sorter schematics (https://en.wikipedia.org/wiki/Flow_cytometry)

There are also rising cell isolation technique such as C1 system from Fluidigm[8]. (Fig 1.2) This technique also captures single cell from cell suspension. By flowing suspended cells through microfluidic channel each chamber can catch single cell by negative pressure[9]. After capturing cells user can lysis them and amplify genome or transcriptome in each chamber by controlling

the valve. This technique is user friendly, since all of the process is automatically processed after loading the cell suspension by single pipetting. Therefore, C1 have high potential to be used on many single cell study because of its reliable data quality and user friendly instructions.

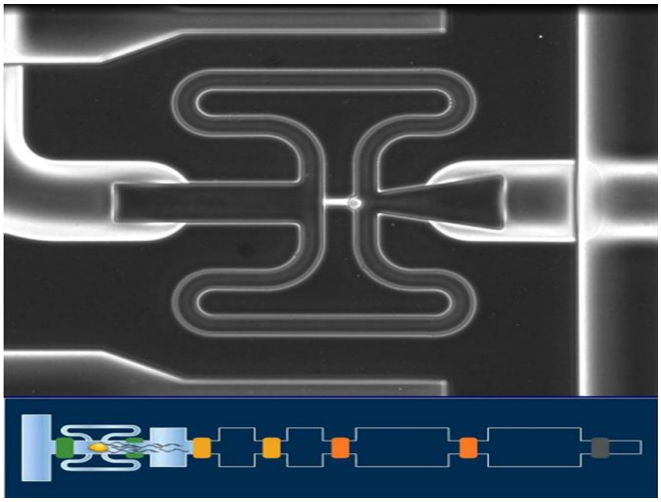


Figure 1. 2 Chamber at C1 system
(http://qb3.berkeley.edu/fgl/single-cell-services/)

Pathologists in the hospital usually adhere patients' blood or tissue on slide glass and stain them to see morphologies of diverse cell types. (Fig 1.3) By observing morphologies of cell, pathologists can quantify abnormal cells or predict how severe the disease is[10]. Moreover, this process is much simpler than isolating single cells for analysis, Therefore, the first medical examination is

based on stained smear samples until nowadays. However, using morphological information only limits accurate diagnosis and research. Conventional single cell isolation techniques such FACS and C1, are popular for many research areas but they are appropriate for isolating suspended cells. If the targeted cell is adhered on the glass, complex resuspension step is required. Bulky resuspension step will mix every cell and will not guarantee that the analyzing cell was the cell which was seen through the microscope.

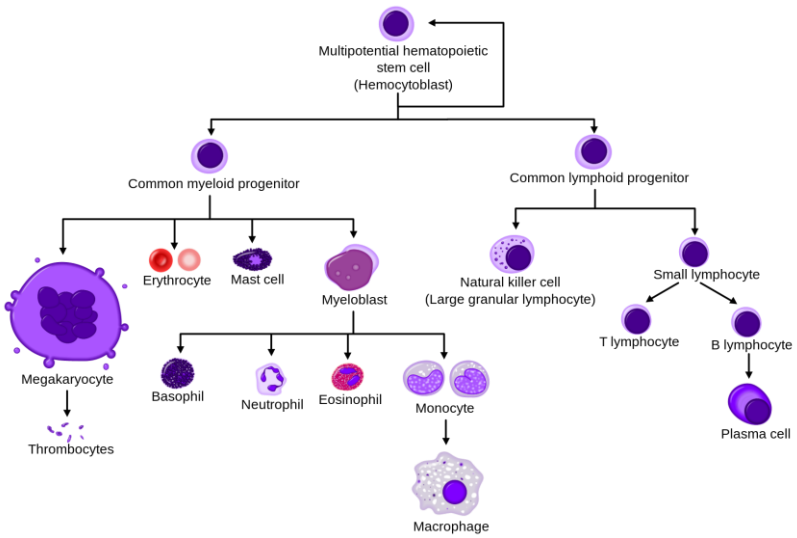


Figure 1. 3 White Blood Cell Types
(https://en.wikipedia.org/wiki/White_blood_cell)

Both morphologic and genomic/transcriptomic information is crucial. Therefore, researchers developed a platform so called

laser microdissection (LMD). LMD can see through microscope and select the cell to isolate[11]. (Fig 1.4) Laser burns out the surroundings of the target area and drop down to collecting chamber, usually small tubes or caps. Many researchers used this technique and found out genomic mutations or transcriptomic difference of minor clone which wasn't revealed while analyzing bulk samples[12].

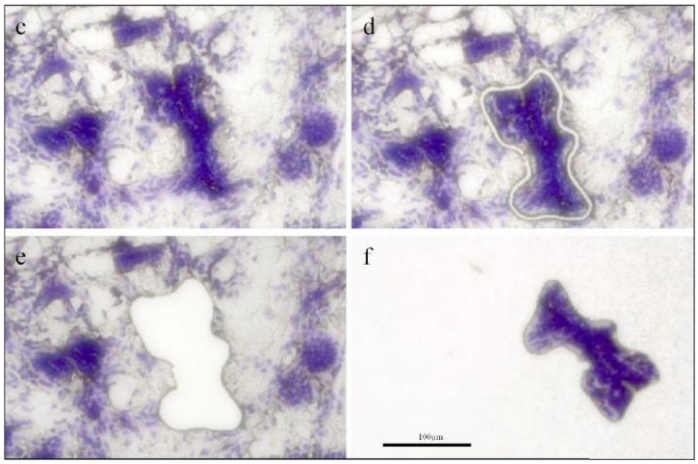


Figure 1. 4 Cell isolation using laser microdissection
(Aaltonen et al., BMC Research Notes, 2011)

Even many break through seemed to be possible using LMD, there are some drawbacks. First of all, burning the surrounding of the target cell limits for single cell application[13]. Isolating lump of cells at once, will have no problem since user can ignore the damage of UV laser. But, in the case of isolating single cell, thermal

damage is serious for further analysis. (Fig 1.5) Moreover, LMD requires to lay the cell sample on polyethylene naphthalate (PEN) membrane slides[14]. Since this polymer film is too thin, handling the slide needs extra care. Surface of the film can be ripped or became rough and bumpy which can disturb obtaining clear morphological information. Most of all, there no settled protocol of genome amplification for single cell level. Amplifying genome with minimum bias is essential for high quality genome analysis[15].

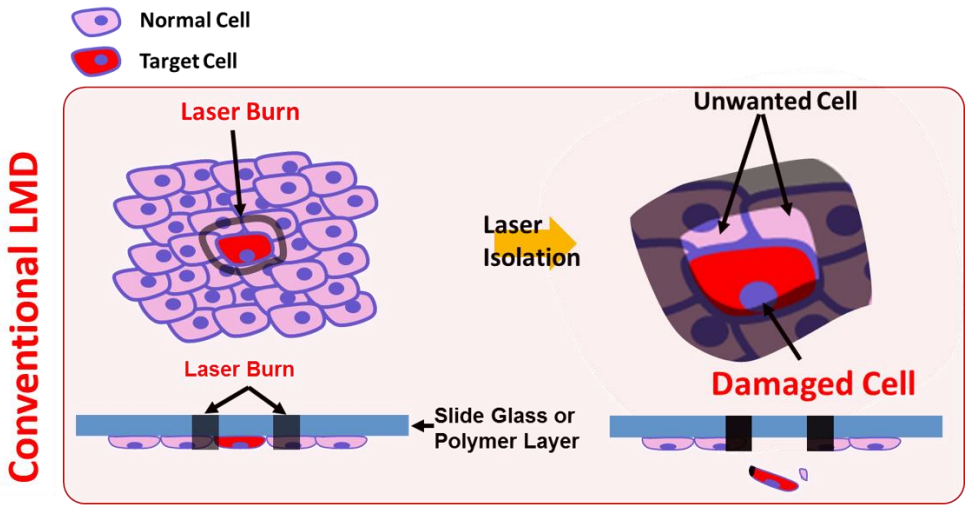


Figure 1. 5 Conventional LMD for single cell isolation

1.2 Genome Amplification

High-throughput sequencing technologies enabled one-thousand-dollar genome sequencing. Commercially available library construction kits such as SureSelect (Agilent) and HyperPrep (KAPA Biosystems) kits provide easier preparation process to use high-throughput sequencing technology. However, this popular kits requires minimum input DNA, more than 200 ng[16]. This minimum amount is hundreds of thousands time bigger than the genome that single cell contains. Recently, optimizing protocol of ligation improved the efficiency and drop down the minimum amount to few tens of nano-grams[17]. But still, this minimum amount is too big compared to single cell level of genome.

Therefore, process for amplifying genome is inevitable when using isolated single cells' genome as input DNA for next generation sequencing (NGS) technology. The most important factor while amplifying genome is reducing amplification bias. Amplification bias could end up with low quality of genome analysis or high depth sequencing to overcome the bias which will be very

expensive. In the other hand, since bulk sample need no amplification step as a DNA input, genome coverage is normally high without less error.

Furthermore, amplifying million times of single cell level genome causes polymerase error. There is plenty of sequencing data to construct a reference sequence when using bulk sample which is not amplified. But in the case of single cell level genome, each polymerase error increase false positive for further analysis. This is because low quantity of input DNA cannot average out the amplification error which happened at initial amplification step.

Conventional method to amplify whole genome with low bias, so called degenerate oligonucleotide primed PCR (DOP-PCR) is based on PCR. Random loci need to be amplified to obtain low amplification bias. For random location amplification, DOP-PCR uses degenerate nucleotides, such as deoxyinosine) with regular base pair (normally 6 bp) on the 3' end as a primer. (Fig 1.6) Since specificity depends on only short regular base pair, primer is much easier to match on genome and stability of binding increases due to degenerate nucleotides. Initial annealing temperature is also low for more stability of binding. DOP-PCR was widely used

technique in many applications such as chromosome painting[18] and microarray hybridization[19] when amount of starting DNA is low. However, whole genome amplification (WGA) using DOP-PCR may give incomplete coverage of loci due to strong amplification bias[20].

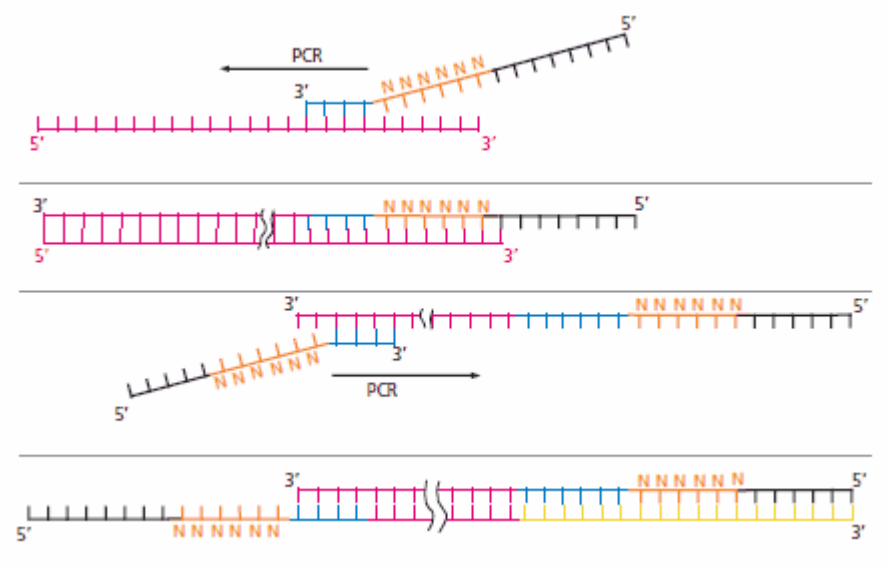


Figure 1. 6 Degenerate oligonucleotide primed PCR mechanism (http://www.sigmaaldrich.com/)

Recently new WGA method such as multiple displacement amplification (MDA) and multiple annealing and looping-based amplification cycles (MALBAC) was introduced. MDA uses phi29 DNA polymerase which can synthesize DNA up to 70 kb in length and random primers[21]. (Fig 1.7) Amplification bias was less than 3-fold in contrast to 4-6 orders of magnitude for PCR-based

WGA. Moreover, since phi29 DNA polymerase displaces strand during amplification, there is no need of temperature changing. MALBAC reduces amplification bias by quasilinear pre-amplification followed by normal PCR cycle[22]. (Fig 1.8) Author claims that single nucleotide variation (SNV) detecting efficiency is higher and allele dropout (ADO) rate is lower than MDA. Also, other group says MALBAC has higher genomic coverage with less amplification bias with slightly higher error rate introduced by Taq polymerase in regular PCR cycle[23]. However, which WGA method produce less bias is still in controversy. Some claims pre-amplification or minor template increments do not alleviate ADO rate[24]. Also since MALBAC had been introduced no longer than 4 years, more research results using MALBAC as WGA needs to be validated.

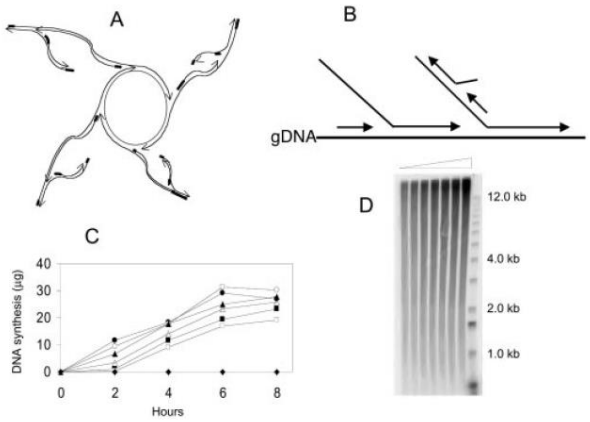


Figure 1. 7 Multiple displacement amplification mechanism and product
(Frank B. Dean et al., PNAS, vol. 99, 2002)

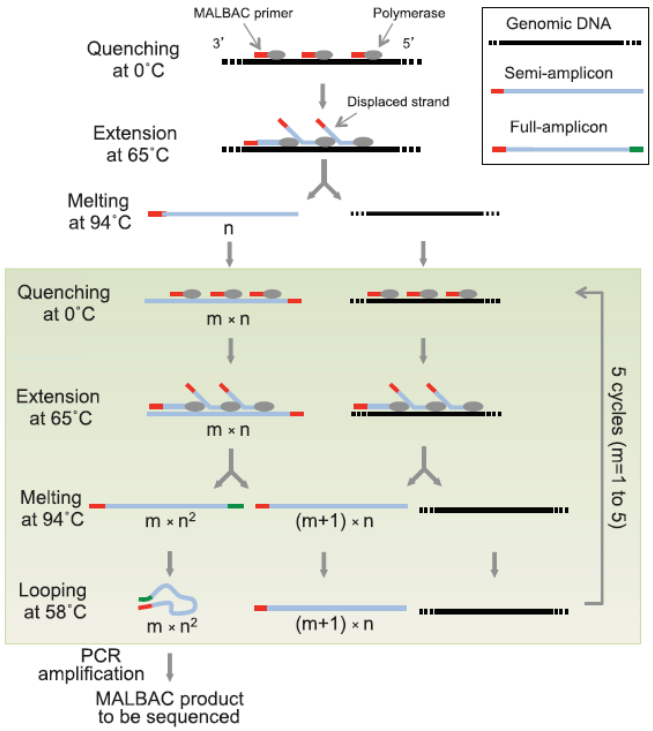


Figure 1. 8 Multiple annealing and looping-based amplification cycles mechanism
(Chenghand Zong, et al., Science, vol. 338, 2012)

MDA has more references of broad applications compared to MALBAC. Therefore, we choose to optimize MDA for WGA at single cell level genome analyzing platform. Main purpose of optimizing MDA is amplifying genome with minimum amplification bias with high coverage of genome. Low amplification bias is close related to amplification efficiency in case of MDA, and this will be explained in chapter 3, before describing laser isolation platform in chapter 2.

Chapter 2

Laser Catapulting Platform

In this chapter, I introduce single cell isolating platform based on laser catapulting. This method enables real time intact single cell level isolation targeting the cell which is adhered on solid glass for researchers to see the morphology. Since laser focuses on sacrificial on the glass and not the cell there is minimal heat damage which guarantee further high quality genome analysis. we demonstrated with various sample such as cell line and blood sample to show this platform can be applied to broad research areas and diagnosis.

2.1 Laser Isolating Platform on bare glass

Most researchers and pathologists observe cells after adhering on slide glass. The reason they use glass is because glass is physically and chemically stable and at the same time transparent. Researchers need to see through the cell without any overlaps of each cell. Therefore, in the case of observing the tissue, pathologists slice the bulk solid sample into one or less thickness of layer could be lay on the glass. In the case of floating cell such as cell lines or blood sample, researchers or pathologists smear them with other slide glass hoping some parts of the cell is not overlapped each other. Glass have perfect solidity to endure the physical stress while spreading the cells, maintaining flat surface for clear observation.

Even the single layer spots were secured to have a clear view of the cell, the cell needs to be stained for better observation. This is because cells are usually too transparent to see. There are many methods of staining around the world; they could use water

base solution or ethanol base solution or even more harsh solutions. Glass provides chemical endurance so that users do not need to consider whether they support broke or dissolve into their solvent.

For these reasons, pathologists and researchers used glass to observe spread cells. As a result, plenty of rare patients' smeared cells on the glass are stored in hospitals waiting for novel technology to analysis their genome information. Therefore, we aim to isolate the cell on the bare glass. LMD cut the surroundings and then push the cut spot with out-focused pulse laser. For higher throughput and less heat damage while cutting the surroundings, we decided to isolate the cell by single pulse laser shot. Platform followed the DNA isolating platform introduced at Nature Communications, in 2015[25]. (Fig 2.1.1)

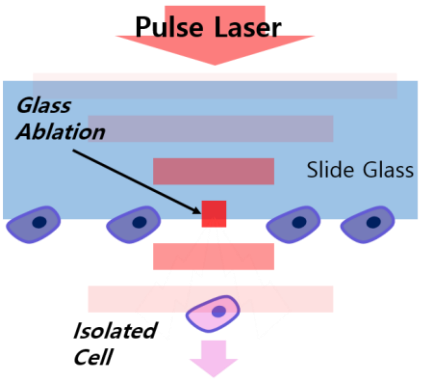


Figure 2.1. 1 Schematic of cell isolation on bare glass by pulse laser catapulting

Laser was focused directly on the cell. Pulse width was 7~10 nanoseconds with two wavelengths, green, 532 nm and infrared, 1064 nm. Firstly, green laser was used. With low power of laser shot cells' shape and transparency changed. (Fig 2.1.2) Cell disappeared when higher power of pulse laser was shot leaving the trace of the existence of the cell and dark dust around the laser shot spots on the glass.

Since the cell disappeared from the glass we decided to see whether the cell was safely isolated. Clean retrieving glass (RG) was closely placed on the isolating glass (IG) facing the cell side. A number of cells were shot by green laser with high power. After isolation of cells, RG was placed on microscope to see the isolated cells. There was no shape of cell but only black dust claiming there was something fell apart from the IG. (Fig 2.1.3) Summing this results, there was no evidence that the cell was isolated safely when shooting with green pulse laser. On the contrary to intact isolation the cell seems to burn when directly focusing with low or high power laser. Retrieved black dust seems to be fragments of burnt cell pieces. IR laser didn't show any difference with maximum power (100 mJ) of laser shot, which means there is no

interaction in this wavelength.

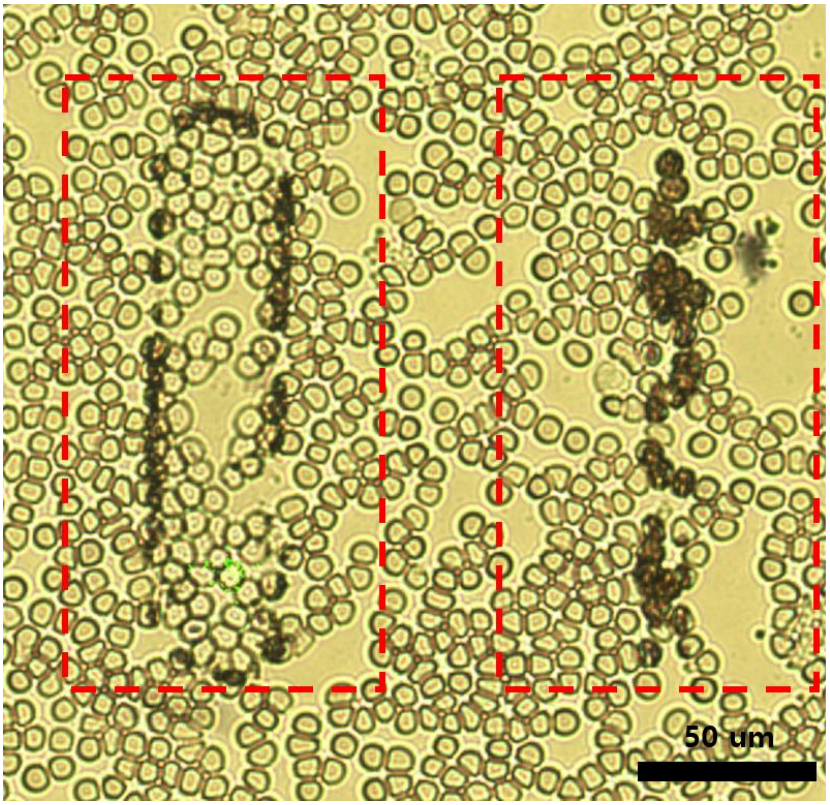


Figure 2.1. 2 Cell isolation with green laser on bare glass

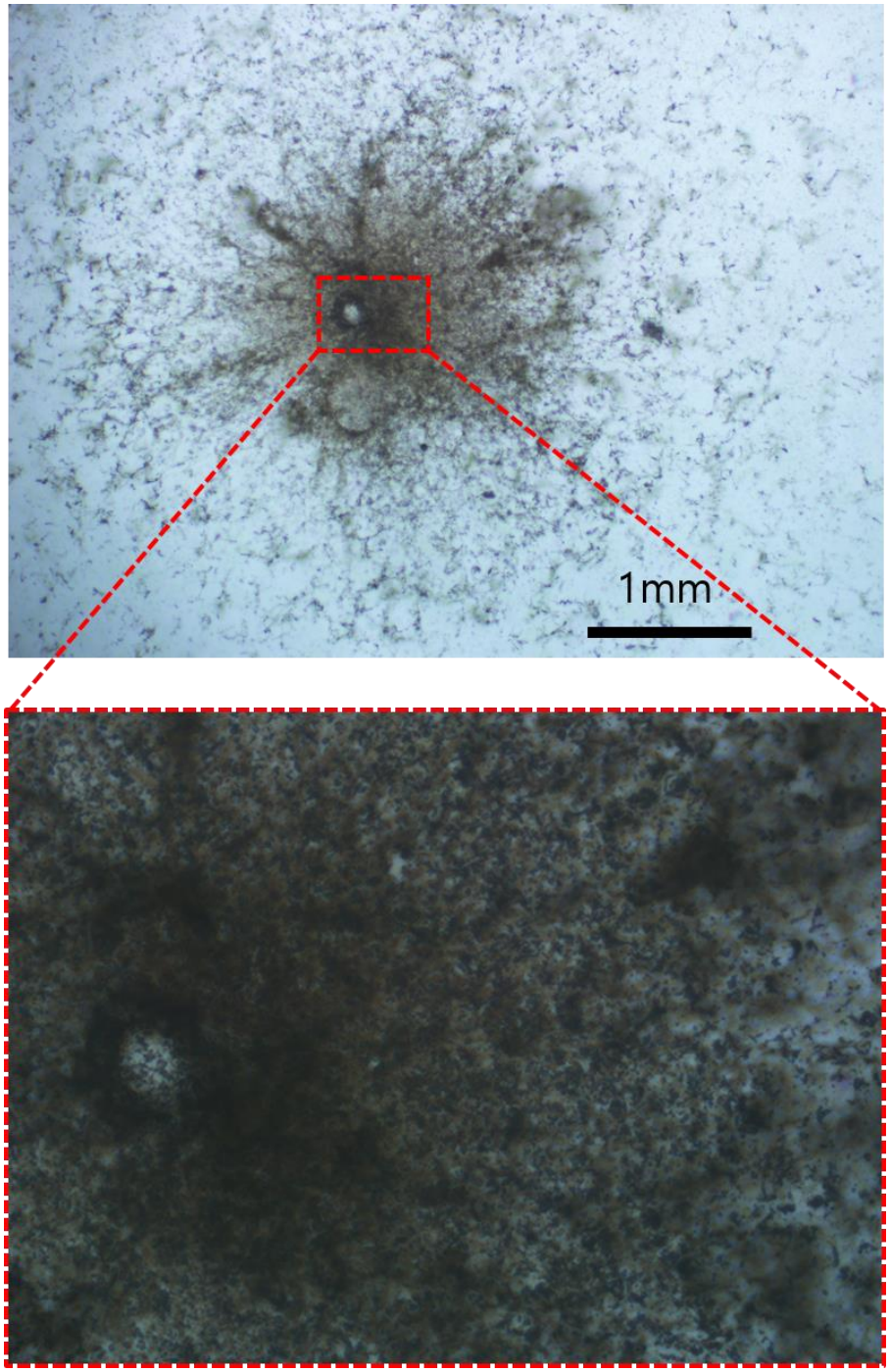


Figure 2.1. 3 Retrieved burnt cell fragments isolated from bare glass

Cell was isolated by focusing laser below the targeting cell. Focusing the laser underneath the cell, which is glass, will result in weaker interaction between laser and the cell. The deeper out-focusing laser spot the weaker interaction is expected. (Fig 2.1.4, a) Therefore, there will be certain out-focusing depth which cell is not damaged while glass ablate to push the cell to fall. We tested coloring slide glass with blue paint. (Fig 2.1.4, b) Blue paint was a substitute for cell. There was almost no thermal damage of pulse laser to blue paint when laser was out focused around 200 micrometers. This result implies only if we can break the glass beneath the target cell, cell isolation would be properly done, ignoring how large the disruption of the surface would be.

In-focus shot, 20, 40 and 60 um out focused shot was tried using maximum power of the laser. The sound of breaking the glass was clear. But surface of the glass didn't fall apart when laser was out-focused over 40 um. (Fig 2.1.5) This means glass is too solid to be shattered by laser ablation beneath the surface over 40 um. Besides, the power of the laser was enough to burn the cell, even the focus was out. (Fig 2.1.6)

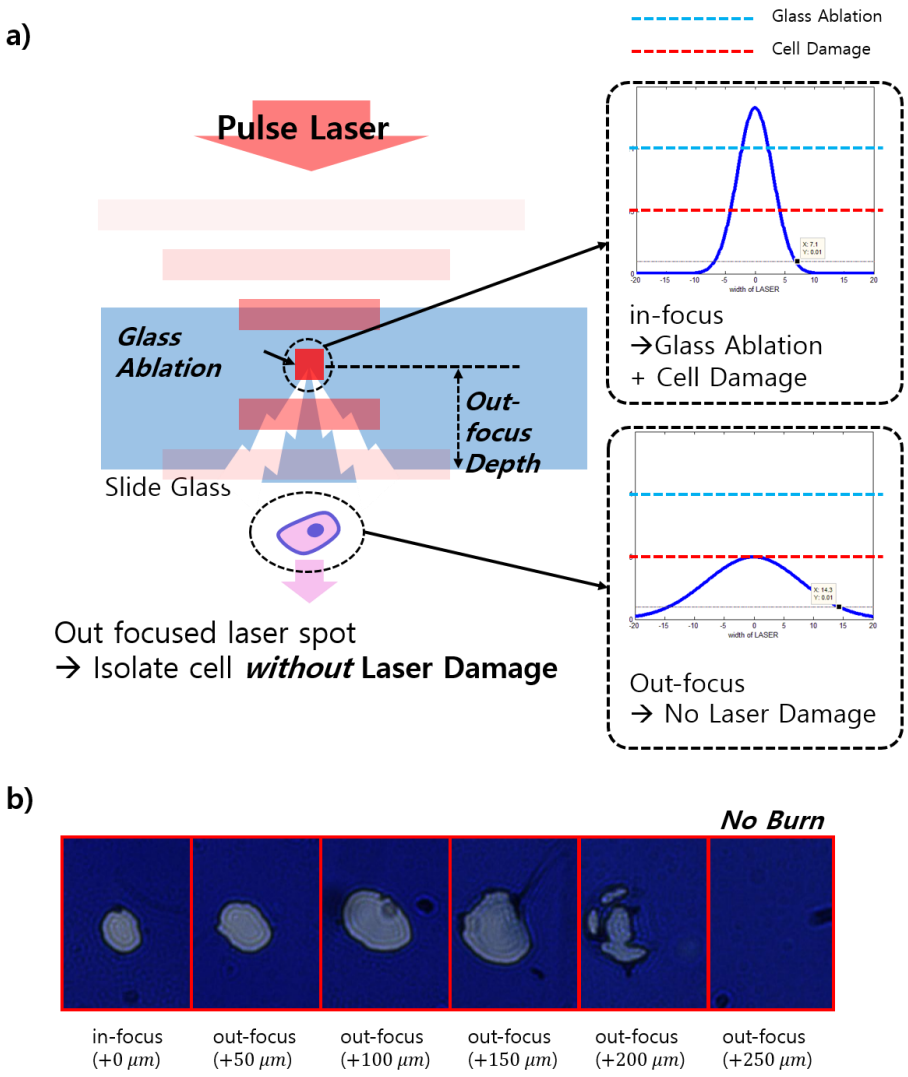


Figure 2.1. 4 Schematics of out-focusing pulse laser for isolating cell

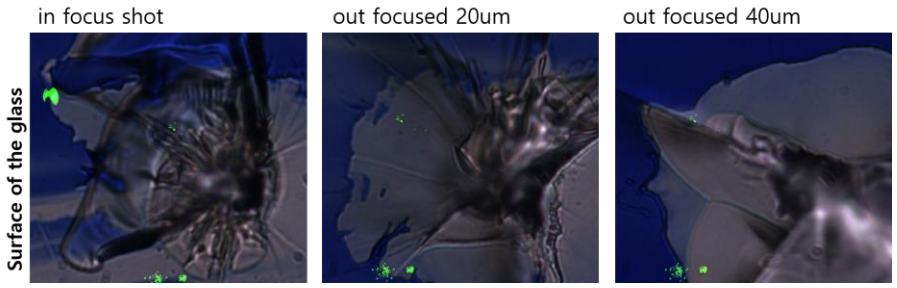


Figure 2.1. 5 Surface of the underneath ablated glass

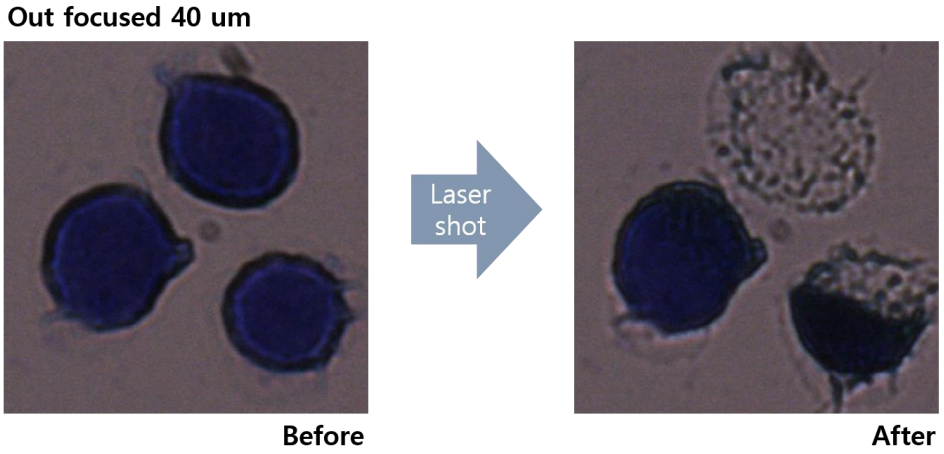
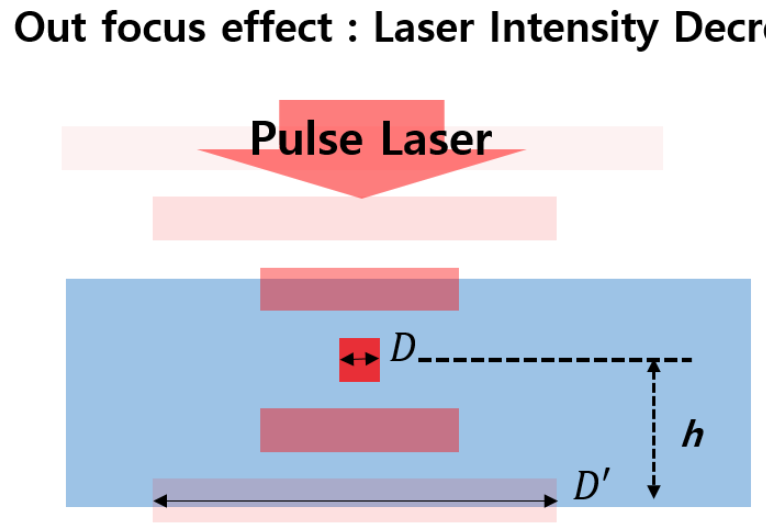


Figure 2.1. 6 Burnt cell when laser shot was out of focus

Decreasing interaction between cell and laser while maintaining falling pressure with nanosecond pulse laser did not work by out focusing laser. Simple calculation helped what factor needs to be improved for intact isolation of cell. In figure 2.1.7 effect of out-focusing laser is modeled and calculated.



$$\frac{P'}{P} = \frac{A}{A'} = \frac{D^2}{D'^2} = \frac{1}{(1 + h\frac{8NA^2}{\lambda\sqrt{n^2 - NA^2}})^2} = \frac{1}{(1 + hc(NA))^2}$$

$$(where\ NA = \ numerical\ aperture,$$
$$c(NA) = \frac{8NA^2}{\lambda\sqrt{n^2 - NA^2}}\ )$$

Figure 2.1. 7 Model of out focusing laser to isolate cell safely

Since glass should ablated while cells on the surface should not be damaged we can define two inequalities as below.

Ablate glass

$$I\, a_{glass} > K_{glass}$$

No damage cell

$$I \frac{1}{(1+hc(NA))^2} a_{cell} < K_{cell}$$

$(a_{cell,glass} = absorbance\ factor,$
$K_{cell,glass} = ablation\ point)$

Laser power is same while both event happens. To satisfy both inequalities at the same time laser power should satisfy the following inequality.

$$\boldsymbol{K_{glass} \frac{1}{a_{glass}} < I < K_{cell} \frac{\left(1+hc(NA)\right)^2}{a_{cell}}}$$

$$\boldsymbol{\therefore \frac{a_{cell}}{a_{glass}} < \frac{K_{cell}}{K_{glass}} (1+hc(NA))^2}$$

Absorption of the cell and ablation point of the glass and cell is constant. Therefore, we can change the distance of the out focusing depth (h), choose higher NA lens or increase absorbance factor of the glass. Unfortunately, as I mentioned before, there is limit of out-focusing depth to under 60 um. So I assume the model of isolating single cell should follow plot at figure 2.1.8.

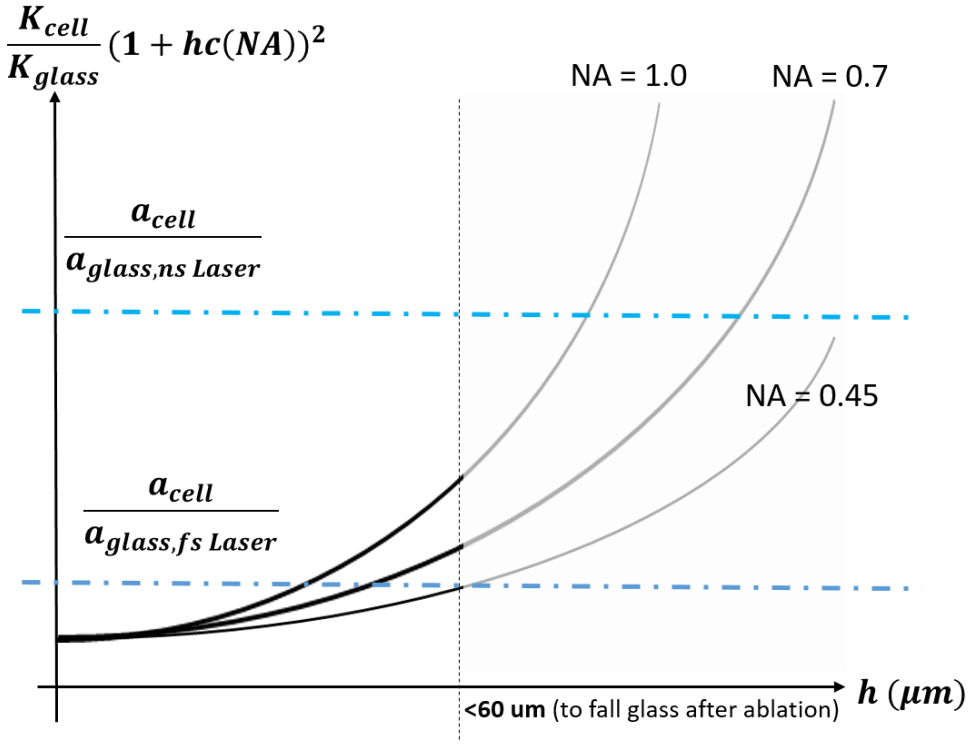


Figure 2.1. 8 Effect of out-focusing laser for intact cell isolation

According to this plot, there is no way to isolate cell safely on bare glass using nanosecond laser. Higher NA lens may approach the proper isolating condition. But the lens would be too expensive maintaining longer working distance to pass through the slide glass thickness (~1 mm) with higher NA than 1.0. Changing laser from nanosecond laser to femtosecond laser might increase absorbance of glass. This is because nanosecond time scale is long enough to occur extra radiation, ionization, vaporization and

convection which decreases absorbance constant of the glass while femtosecond timescale is short enough for absorption and excitation of the atoms can happen[26]. However, it is uncertain that surface of the glass where cell is adhered will break apart for appropriate isolation. Also, femtosecond laser is too expensive and hard to maintain the quality of it for many researchers or hospitals to use.

Increasing absorption constant by changing the laser is not appropriate solution for cell isolation. Instead, using sacrificial layer would have same effect of increasing the absorption constant easily. At the same time, not only increasing the absorption but also maintaining high physical and chemical endurance with high transparency is important. However, this implies new platform cannot be applied to stored patients' sample which is rare and precious for further research. I believe my model of effect of out-focusing will inspire future researchers and develop better platform for appropriate solution which I couldn't make out.

2.2 Laser Isolating Platform on sacrificial layer

Isolating single cell adhered on bare glass by pulse laser catapulting requires further research. Adhering cells on layer which has higher absorption rate of laser instead of bare glass can overcome the burning cell problem. Since this layer absorbs laser and burn instead of cell above, the layer is so-called sacrificial layer. Sacrificial layer should be transparent while maintaining high physical and chemical durability such as glass for clear observation of cells.

My research team leader introduced a transparent metal, indium tin oxide (ITO), which is widely used as electrical path patterning layer. This metal is popular for its transparency of visible light while producing high conductivity[27]. Interestingly, this metal's absorbance rate of infrared is higher than visible light[28]. Mentioned at chapter 2.1, cell didn't interact with IR laser. This means, by catapulting with IR laser which doesn't interact with cell but is absorbed at ITO sacrificial layer can expect intact cell isolation.

ITO coated glass was ordered from “Fine Chemical Industry”, with different thickness of coating layer: 150 nm and 300 nm. They were both coated with sputtering method. The view seeing through a ITO coated glass became little dark but the quality difference of a view compared to seeing through a bare glass was almost negligible. (Fig 2.21) Cell was stained with Giemsa staining protocol and staining protocol was from prof. Lee at Seoul National University Hospital (SNUH). Giemsa staining protocol is as followed. First, smear the cell on the ITO coated glass. Before the smeared cell dries, glass is dipped into 100 % methanol for 30 seconds for fixation. Dip the glass to giemsa staining solution right away for 7 minutes. The time of staining can be varied depending on thickness of stained color. After staining process is done, rinse the staining solution with phosphate-buffer saline(PBS) followed by rinsing in deionized water. After drying in air, the sample is prepared for observation and isolation.

There was slight difference while cell adhering step due to different hydrophilicity between glass and ITO. Glass interacts with water more than ITO resulting bigger contact angle while adhering cell on the glass. But this can simply overcome by treating the

surface of ITO coated glass by extra treatment for hydrophilic surface, such as coating the surface 2-propanol or plasma treatment. However, even without any extra treatment user can easily overcome this hydrophobic situation by simply slowly smearing the cell sample on the ITO coated glass. After we found image quality of the stained cell and staining protocol is same as imaging and staining on the bare glass we isolated the cell using IR laser.

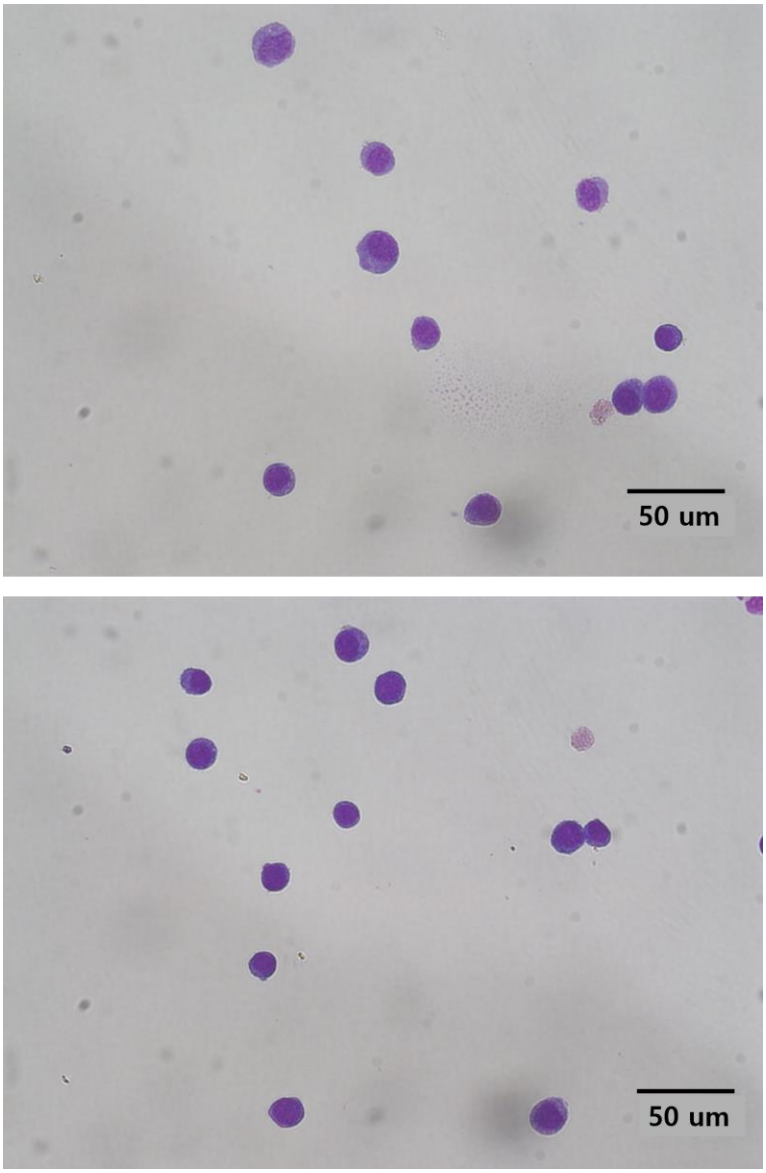


Figure 2.2. 1 Image quality of cell staining.
Cell image on ITO glass (up) is little darker than cell image on bare glass (down).

First of all, thickness of ITO layer for intact isolation was tested. (Fig 2.2.2) Cell was stained with Giemsa protocol and then

isolated with IR laser. Cell disappeared on IG with low power of laser in both thickness of ITO layer. Retrieved cell was also observed after the isolation from both thickness of ITO layer at RG. Some cell did break into half but no burnt damage was observed. (Fig 2.2.3) This result implies isolating efficiency is not critically depends on thickness of ITO layer between 150 nm and 300 nm. Therefore, we choose to use 150 nm thickness of ITO layer coated slide glass from now on since it is cheaper than 300 nm thickness of ITO layer.

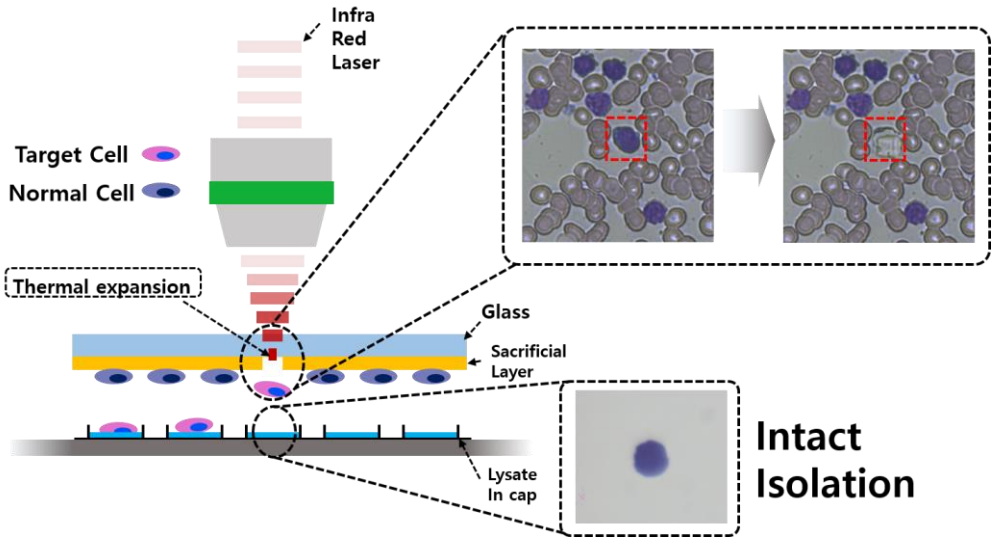


Figure 2.2. 2 Schematic of single cell isolation on sacrificial layer (ITO layer)

ITO coated glass is useful material for its characteristics. At this single cell isolation platform, light absorption characteristic (high transparency of visible light and high absorbance of infrared) and

high physical and chemical durability were used. The typical disadvantage of this material is the price, since it's hundred time expensive than normal bare glass. Therefore, some researchers or pathologists in hospital may hesitate using this ITO coated glass for storing a number of samples which are produced every day. But the price of this expensive glass is negligible for post process for further analysis such as genome analysis which will be explained in next chapter. Potential of genome analysis is sufficient for storing every sample on this expensive glass, without wasting precious sample which may require long period of time for recollection. Fortunately, there are some trial to use sample on bare glass by stamping prepared cells to on newly designed glass[29]. Moreover, conductivity of ITO layer leaves a potential that this glass surface may be used as an electrical circuit for automatic cell detecting, labeling or biomolecule transportation. Further researches may eventually come up with solutions.

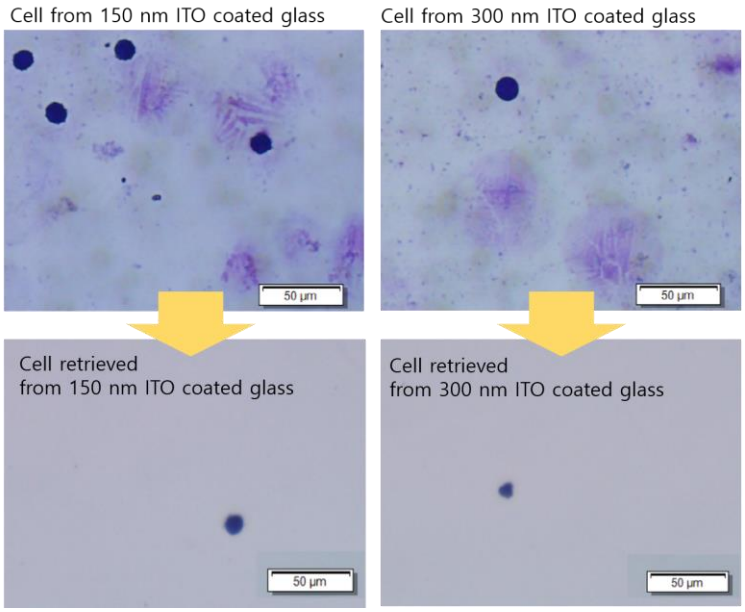


Figure 2.2. 3 Retrieved cell isolated from ITO coated glass

2.3 Laser Catapulting Platform validation

Cell isolation platform based on pulse laser catapulting was implemented by adhering cells on ITO coated glass. For broad application performance of isolating required to be validated. Broad applicability was number on priority and dispersity of isolated cell was the second for availability of further analysis.

We tried various types of sample for isolation. Cell from cell line, blood sample (peripheral blood and bone marrow), FISH stained sample and tissue stamped sample was tried. (Fig 2.3.1) Blood samples and FISH stained sample was from prof Lee at SNUH and tissue stamped sample was from prof Han at SNUH. HL60 cell line was used for cell line sample. A number of isolating experience ensured intact cell isolation succeeds if only cell is disappeared from IG.

Samples from cell line did not require any modification for isolation via isolating this platform. In the case of blood sample, cell to cell distance was not far so that fine aiming required to prevent

isolation of unwanted cells. Since, red blood cell doesn't contain genome inside the cell, isolating cell contains genome from peripheral blood did not require change of laser power. Besides, sample from bone marrow contains high portion of white blood cell resulting the distance between cells were not far enough. This issue was solved by decreasing the laser spot size while increasing laser power for isolation. was required. FISH stained cell was covered with sticky oil whose role is maintaining the fluorescence signal by preventing bleaching and providing clear view. This sample required higher power of pulse laser while isolation to prevent moving from spot to spot without falling from IG. As a result, cell was isolated clearly while trying on a number of FISH stained cells. Tissue stamped samples did not have clear physical edge of the cell. After isolating this kind of sample leaved torn part of the cell. This means even whole part of the cell was isolated, user cannot sure whether the genome was isolated safely. To apply this platform for isolating to tissue sample or tissue stamped sample, further research is required for practical use.

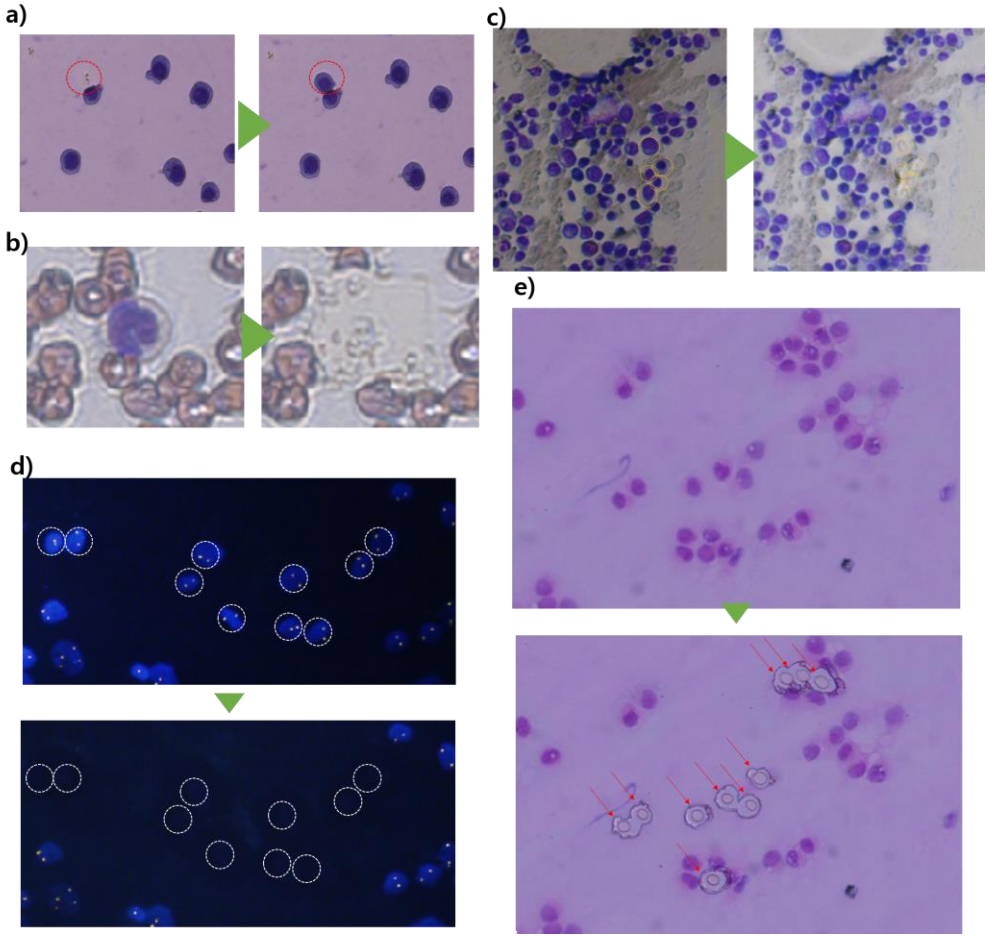


Figure 2.3. 1 Various sample for laser isolation

Dispersion of isolated cells was also studied. After isolating hundred cells on IG, I observed the pattern of retrieved cell from RG. (Fig 2.3.2) Cell sample was peripheral blood sample stained with typical hematoxylin and eosin. Dispersity of isolated cells was calculated by measuring the distance between isolated point and retrieved cell. We assumed isolated point is the center point (mean

or median of x and y position) of the retrieved cells on RG, since isolated cells will disperse around the center point. Distance between IG and RG was 1 mm. 91 cells were found on the RG. Ignoring data points above six-sigma, which is only 2 cells, every cell was retrieved inside 1 mm radius circle. Also 99 % of the cells which were found was inside less than 1.25 mm radius circle. Unfound 9 cells might have been retrieved outside of the observed region which is 4 x 4 mm rectangle.

Validation of cell isolation platform was demonstrated on various sample types typically used on many research fields and hospital to ensure this platform is ready for practical use. For single cell isolation for tissue samples need further research. Dispersity of retrieved cells was also studied to see efficiency of isolation. Since more than 90 % of isolated cells were retrieved inside less than 2 mm radius circle, this platform ensures single cell isolation would be sufficiently done while retrieving with popularly used 96-well PCR tubes or tube caps, which diameter is 9 mm. This platform showed clear improvement of isolating cells by potential of higher throughput and safer isolation compared to conventional laser microdissection.

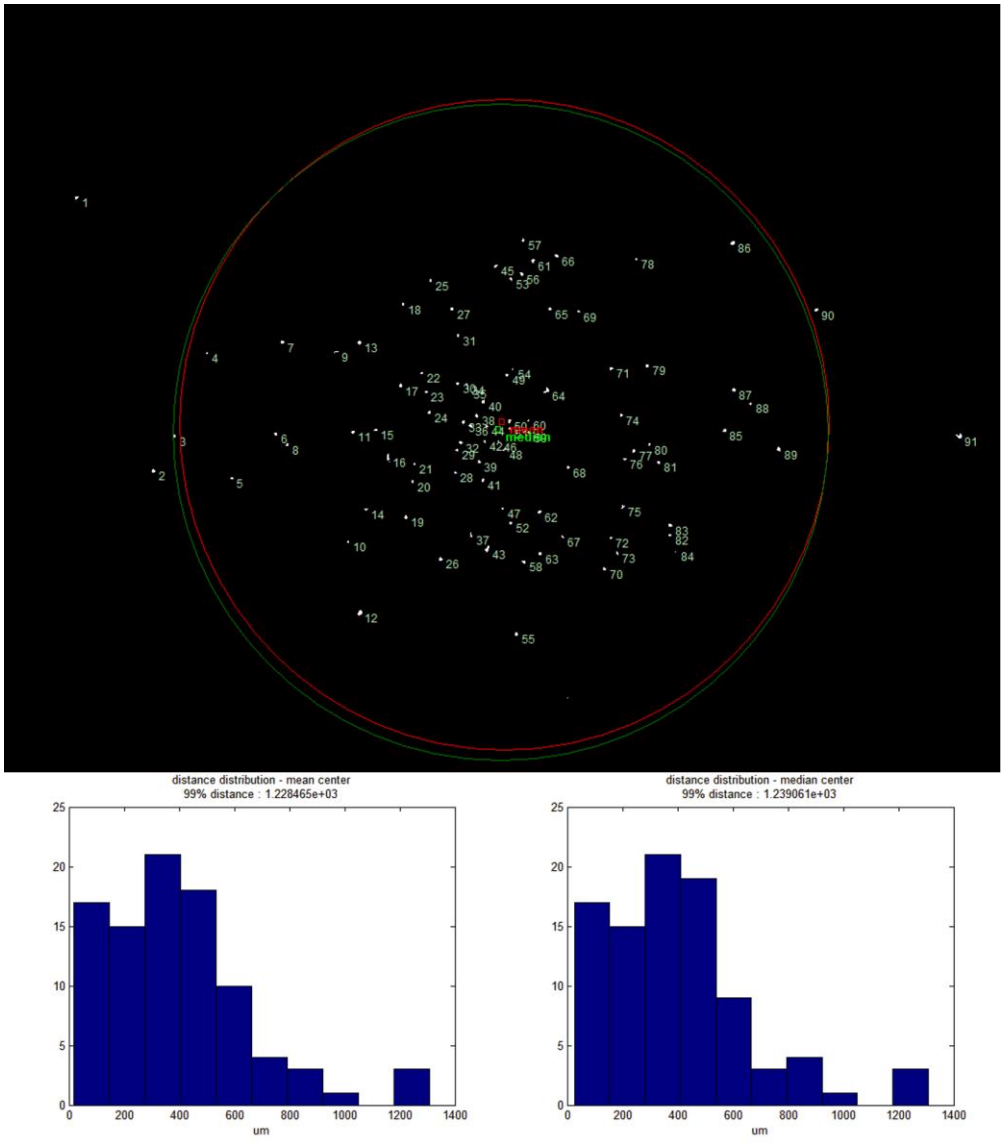


Figure 2.3. 2 Dispersion of retrieved cells.

Chapter 3

Genome Amplification

In this chapter, optimization of genome amplification is described. Genome amplification is essential while handling low input of genomes, which is the case of preparing isolated single level of cells for further analysis. Starting from retrieving step, lysing cells, preparing lysate for WGA and WGA steps are optimized. Even using same kit, quality of amplified product dramatically changes with and without optimization.

3.1 Cell Lysing Protocol Optimization

Quantity of WGA product is important but uniformity of amplification is more critical for further analysis. Best way of measuring uniformity of amplification is analyzing WGA product with whole genome sequencing (WGS) and see the genome coverage. WGS cost dramatically decreased nowadays but still too expensive to use for routine measurement such as sanger sequencing.

Instead, we assumed uniformity of genome amplification can also be measured by initial amplification rate. (Fig 3.1.1) Polymerase of MDA amplifies genome from random hexamer and hexamer can bind to original genome DNA or elongated DNA. Uniformity of amplification depends on the number of hexamer bind to original genome DNA. This is because if there is few hexamer bind to original genome DNA and many hexamers bind to elongated DNA starting from single site of original genome, the quantity of WGA product can be high but highly biased to single site of the

genome. Therefore, elongation starting from multiple site of genome is essential for uniform amplification. The number of elongation starting point will directly affect initial amplification rate. To see the initial amplification rate, we required to measure DNA quantity in real time. By measuring the intensity of WGA mixture added with 'SYBR Green I', which stains nucleic acid, by real time PCR machine we could achieve real time DNA quantity. We assumed uniformity of amplification can predicted by measuring time of reaching certain level of intensity.

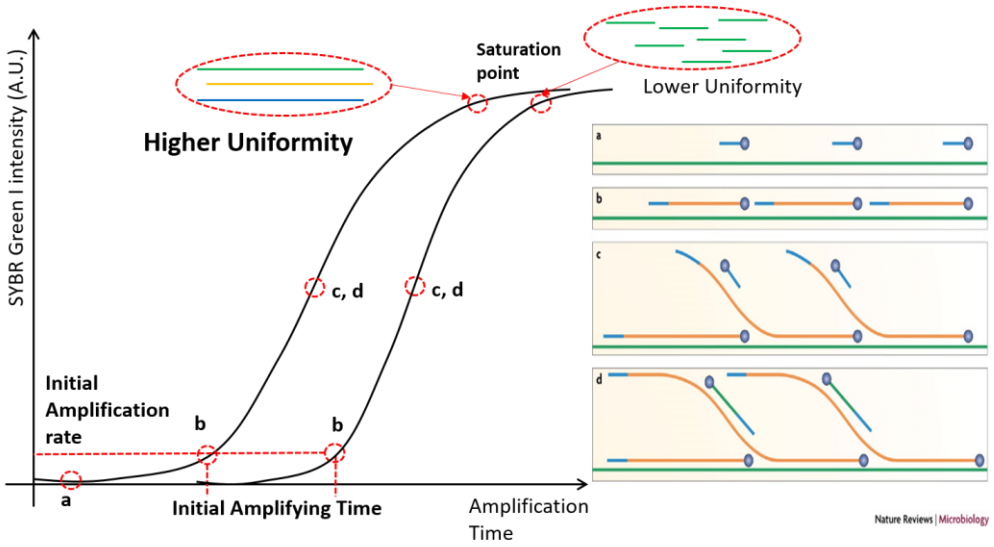


Figure 3.1. 1 Real time amplified DNA quantity deciding uniformity of WGA (Alan Walker and Julian Parkhill, Nature Reviews Microbiology, 6, 2008)

After isolating the cell from IG, genome needs to be isolated from retrieved cell for amplification. Genome in the cell is surrounded by various structure. Also, genome is highly packed with histone protein, which role is well known for gene regulating. Therefore, these various barriers hinder polymerase and random hexamer primer to bind at genome for uniform amplification. Popular WGA kit from 'GE Healthcare' and 'Qiagen' recommends alkaline lysis or proteinase based lysis for cell lysing step.

Protocol of cell lysing from GE Healthcare (Illustra GenomiPhi) is as followed. Prepare cell lysis solution (KOH 400 mM, EDTA 10 mM, DTT 100 mM) and neutralization buffer (HCl 400 mM, Tris-HCl, pH 7.5 600 mM). Mix diluted cell in PBS with cell lysis solution and incubate 10 min on ice followed by neutralization buffer. Caution is not use vortex for lysing cell, seems concerning fragmentize of genome. This protocol, based on alkaline lysis, is widely used for blood cell which handles plenty number of cells. Therefore, this protocol was not proper for amplifying single level of cells. Our team used this kit for WGA ten isolated cells several times following the protocol but only obtained high variance of amplified genome product.

On the other hand, lysing protocol from Qiagen kit (REPLI-g SC) is performed at high temperature (65 degree) compared to GE Healthcare protocol which is performed at ice. There is no description of components for given lysing solution, but due to the step which adds 'Stop Solution' just before the amplification we could guess Qiagen kit is also based on alkaline lysing. WGA ten isolated cells was also done using Qiagen kit which was 3 times more expensive than GE Healthcare. Quantity of amplified genome was ~10 ug which was extremely greater than GE Healthcare (~3 ug), but still uniformity of amplification was questionable.

Protocol of amplifying low input of genome requires optimization for further single cell applications. There are two factors which many make alkaline lysis is not proper for single level cell lysis. Firstly, while alkaline lysing, long length of DNA can be fragmentized due to harsh condition of solutions. Fragmentized genome is hard to amplify without bias by MDA method. Secondly, lysing cell without any mixing step may not be enough to eliminate amplification disturbing factors especially DNA adhered protein, histone. If histone protein is not lysed before WGA, this site may not be amplified.

To reduce harshness while lysing we used proteinase K for cell lysing. Proteinase K is also widely used enzyme for cell lysing in genome extraction. Since this enzyme operates well in wide range of pH and temperature, there is no need to put fragile genome into harsh condition which was required while alkaline lysis. To apply proteinase K lysing to widely used WGA kit (GE Healthcare, Illustra GenomiPhi V2), we mixed 0.5 ul of proteinase K (Sigma-Aldrich) to 9 ul of 'Sample Buffer, (SB)' provided in the kit. This mixture was preloaded on retrieving cap to increase retrieving efficiency. After isolating cells, caps were connected to PCR tube followed by gentle cell spin down. Cell lysing was done for 1 hour at 50 degrees in thermocycler. After cell lysing with proteinase K, this enzyme required to be fully inactivated for WGA amplification. Inactivation was guaranteed by adding inhibitors (DIFP or PMSF) or increasing pH. But, adding new materials or changing solution to harsh condition was not appropriate for our purpose of using proteinase K. Therefore, we tried to inactivate proteinase K by increasing temperature. Incubating the mixture at 70 degrees for 10 minutes appeared to achieve proper inactivation of proteinase K. As a result, our new protocol of lysing cells

followed by mixing normal WGA mixture appeared to have improved efficiency of amplification.

50 cells were isolated into retrieving cap followed by new cell lysing method using proteinase K lysis. (Fig 3.1.2) The human promyelocytic leukemia cells (HL-60) cell line was used after fixation with MeOH on IG. After inactivation of proteinase K, WGA amplification buffer with strand displacement polymerase, phi29 was added with SYBR Green I. DNA quantity was measured in real time while WGA. For control experiment group, 50 isolated cells were lysed by alkaline lysing method followed by WGA. Lysing cell with proteinase K appeared to show higher initial amplification rate compared to general alkaline lysing group. High reproducibility output of WGA product was shown. Three replicated isolated cells were amplified at almost same time inferring our new adjusted lysing protocol is reliable for cell lysing. NTC group is WGA without template which varies by contamination rate.

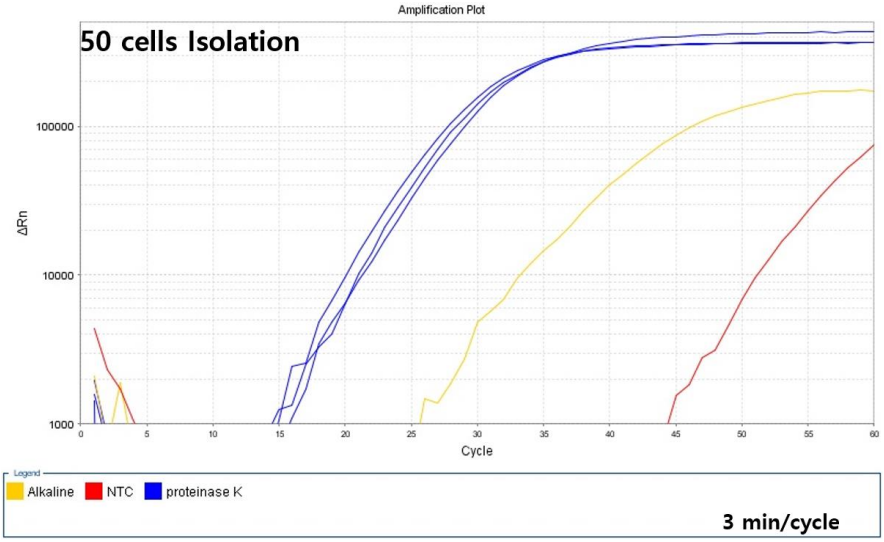


Figure 3.1. 2 DNA quantity measurement while WGA with Proteinase K lysing method

This protocol was still not enough to WGA single cell. (Fig 3.1.3) Initial amplification rate showed high variance of initial amplification rate using same protocol when amplifying single isolated cell. Since, our isolating platform ensures retrieving rate more than 90 %, high variance could not be explained. We assumed variance of initial amplification rate was due to incomplete lysing of proteins, interfering WGA. Higher lysing was required. Therefore, we add gentle tapping while cell lysing, ignoring the caution that vortex might fragmentize genome. Intensity of tapping was minimum but strong enough to mix whole lysate. As a result, mixing lysate while cell lysing step improved initial amplification

rate and at the same reduced variance. This implies protocol of additional mixing while cell lysing step is reliable.

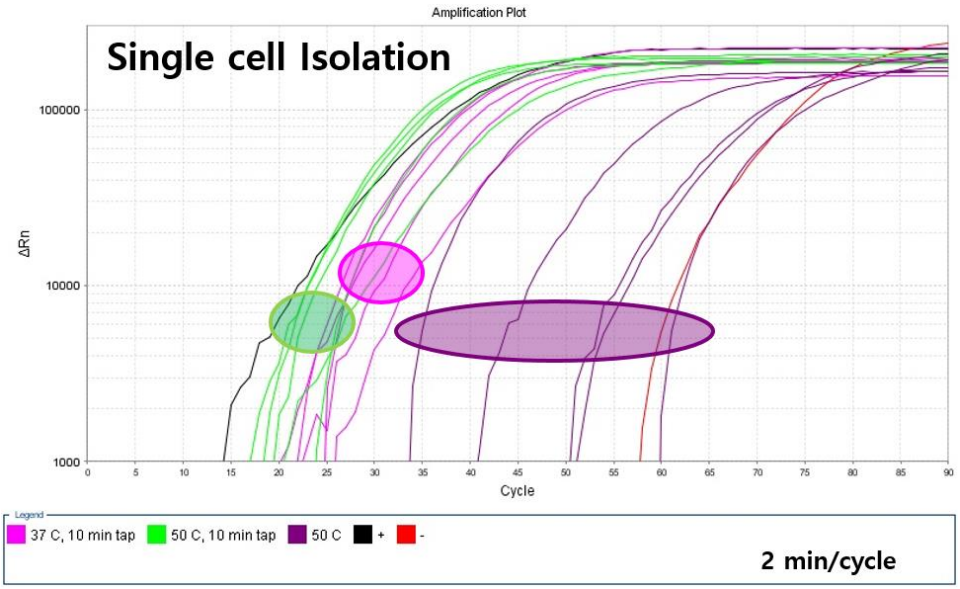


Figure 3.1. 3 Higher efficiency of WGA with additional mixing

Interestingly, lysing cell at 50 degrees showed higher initial amplification rate compared to lysing cell at 37 degrees which is optimal temperature for proteinase K and more stable temperature for long DNA. This result implies higher temperature is required for protein lysing which is more critical for uniform WGA than stability of long DNA. Positive control group is WGA using template with single level of genomic DNA (6 pg) diluted from extracted genomic DNA with high purity. Negative control group is

same as above.

Since new lysing protocol appeared to show higher initial amplification rate we decided to use the optimized lysing protocol from now on. On the other hand, even data show mixing lysate with tapping is reliable enough, we decided to mix with temperature controllable vortexing machine. This is because for higher throughput and reduce human dependency of mixing for wider applicability.

3.2 Genome Amplification Optimization

Even cell lysing step prepared pure genomic DNA for WGA, amplification step itself needs to be optimized for further analysis. According to protocol written at GE Healthcare, amplification step was followed right after lysing and mixing additional buffer and polymerase. But in some paper using this first version of this kit says that mixture was heat denatured and cool down before mixing additional buffer and polymerase[30]. This step was so called, pre-denaturation which seems to fully denature whole double stranded genome to single strand DNA for random hexamer to bind. (Fig 3.2.1) This step may be erased at protocol due to high temperature of pre-denaturation which might fragmentize of genomic DNA.

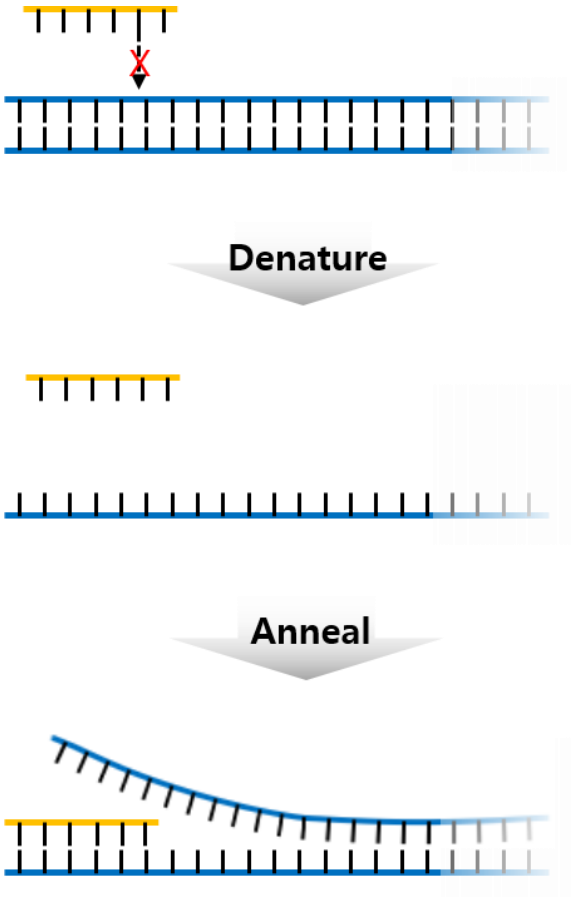


Figure 3.2. 1 Hexamer binding at genomic DNA

We thought this step is essential for random hexamer to bind on genomic DNA which will result in uniform WGA. Following the same protocol, lysate was put into 95 degrees for 3 minutes before mixing with remaining buffer and polymerase. Pre-denatured lysate was cooled to 4 degrees fast as possible for random hexamer to bind at genomic DNA. 10 cells were isolated and same condition of

experiments were replicated 3 times to see the reliability. As a result, we achieved higher initial amplification rate while maintaining low variance by adding pre-denaturation step compared to WGA without pre-denaturation step. (Fig 3.2.1)

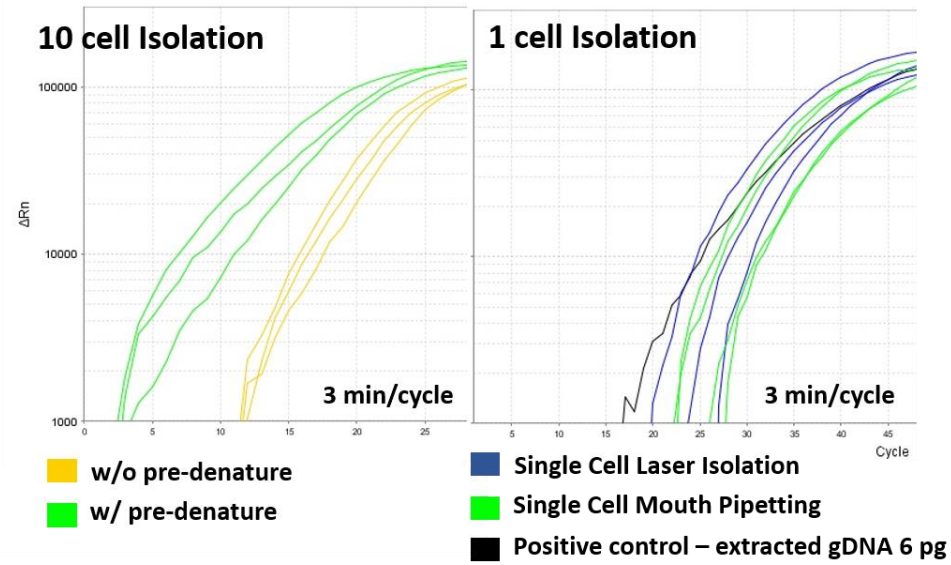


Figure 3.2. 2 Higher initial amplification rate by adding pre-denaturation step

As result, we focused on optimizing preparing steps of WGA and success on improving initial amplification rate which expects to be related to uniformity of amplification. Retrieving cells at liquid filled cap show higher retrieving rate when isolating with laser catapulting system. Lysing cell with proteinase K with gentle vortexing made better state of genomic DNA appropriate for WGA. Adding pre-denaturation step also improved initial amplification rate.

We also tried to optimize amplification reagent itself. But since components of commercialized kit is not published, even copying same efficiency of WGA was hard. To achieve uniform single cell level WGA, more development of reagent itself is required.

Chapter 4

Platform Validation

I have introduced isolating single cell by laser catapulting and optimized protocol for whole genome amplifying single cell level of genomic DNA. In this chapter, we demonstrated the introduced platform with general cell line and patients' blood sample. Technology have their meaning when it is used at real life. In real life, quality of the input genomic DNA is not guaranteed. To see morphological information, one of the important information for pathologist and researcher to diagnose and distinguish cell types,

the cell needs to be stained. Giemsa staining is general method to see morphology of adhered cells, and this staining method turn out to lower the quality of genomic DNA. Unfortunately, we failed amplifying whole genome from isolated single cell which is stained with giemsa staining method compared to quality of genomic DNA extracted from bulk sample. Instead, we found out increased number of isolated cells restore the quality of amplified genomic DNA and isolating 25 cells were enough for PCR analysis. I believe this platform is applicable for many research fields and hospitals right away.

4.1 Platform Validation on Cell Line

Two cell lines were used for platform validation: HL60 and K562. They were both purchased from ATCC and passage number was less than 10 for validation. For accurate measurement of uniformity of amplification four types of data was collected: 1) initial amplification rate, 2) PCR success number when targeting eight random locations on human genome, 3) target sequencing data with leukemia panel and 4) whole genome sequencing profile. Since general depth of whole genome sequencing is too expensive, we followed low depth coverage analyzing method[31]. Leukemia target panel was designed by prof. Lee from SNUH. Staining protocol was also from prof. Lee.

We firstly tested weather staining the cell affects WGA efficiency. (Fig 4.1.1) HL60 cell line was used and 5 cells were isolated. Staining protocol was giemsa staining method. For positive control, cell was isolated without staining step. Initial amplification rate dramatically fell when cell was stained. By this

result, we could say components inside giemsa staining adheres to nucleic acid and interfere WGA.

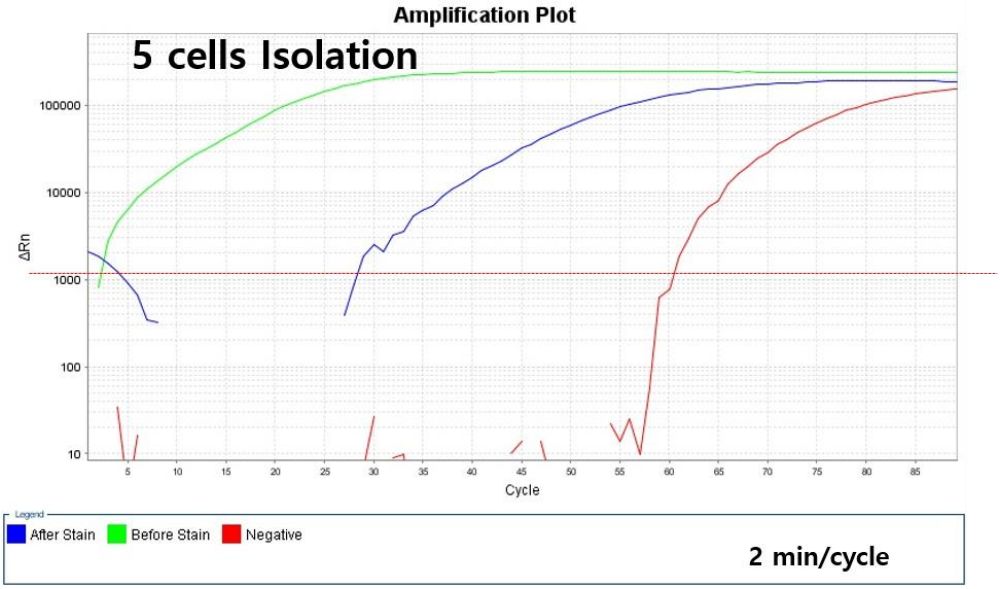


Figure 4.1. 1 Giemsa staining affect initial amplification rate

We tried four different staining methods. (Fig 4.1.2) Two different giemsa staining method, one from SNUH and other protocol was from Sigma-Aldrich, 'Hemacolor' staining method which is commercialized product for fast staining and 'Trypan Blue' staining method which binds to proteins of cell membrane and not adhere to the DNA inside nuclei. HL60 cell line was used and 5 cells were isolated from each staining method with three replications. For positive control, 30 pg of extracted genomic DNA,

same quantity of 5 cells' genomic DNA, was used as template. After WGA, each sample was purified and purified product was used as template for PCR targeting 8 different locations of genome. The reason I amplified 8 different locations by PCR was to measure uniformity of WGA. Uniform WGA will show high number of succeed amplified locations. As a result, retrieved cell groups which were stained with trypan blue showed highest initial amplification rate and highest success number PCR. All of the targeted location succeeded in amplification. Trypan blue doesn't stain the nucleic acid but the cell membrane. This will be the reason of low interference of WGA after staining with trypan blue. On the other hand, this staining method cannot be used for distinguishing various cell types, because trypan blue binds to protein not the DNA. To distinguish cell types, shape of the cell membrane is not the only requirement. By observing nuclei formation researchers or pathologists can obtain more morphological information. Therefore, some research fields cannot use trypan blue as their staining method.

Cell image stained with hemacolor and two different giemsa staining protocol was appropriate for distinguishing cell types,

since they stained the nuclei and the cell membrane. Cell groups which were stained with two different giemsa staining protocol showed similar decrease of initial amplification rate and low PCR success rate. This means protocol difference affect little of WGA efficiency when using same reagent. Protocol of staining cell with hemacolor was handy and quickest but efficiency of WGA was poorest. Interestingly, air smeared cell which wasn't even go through fixing step show low initial amplification rate compared to trypan blue staining group. Since, staining will not improve amplification efficiency, missing fixation step would might be the key for maintaining high quality of genomic DNA. Maintaining quality of genomic DNA will discussed in chapter 4.2 again.

Distinguishing various cell types and maintaining WGA efficiency for single cell level are both important. Since trypan blue staining method is limited for distinguishing cell types and staining with hemacolor dramatically lowers WGA efficiency, we decided to use giemsa staining for further demonstration. Between protocol from Sigma-Aldrich and SNUH, we decided to use protocol from SNUH since patients' blood sample is prepared with the same giemsa staining protocol.

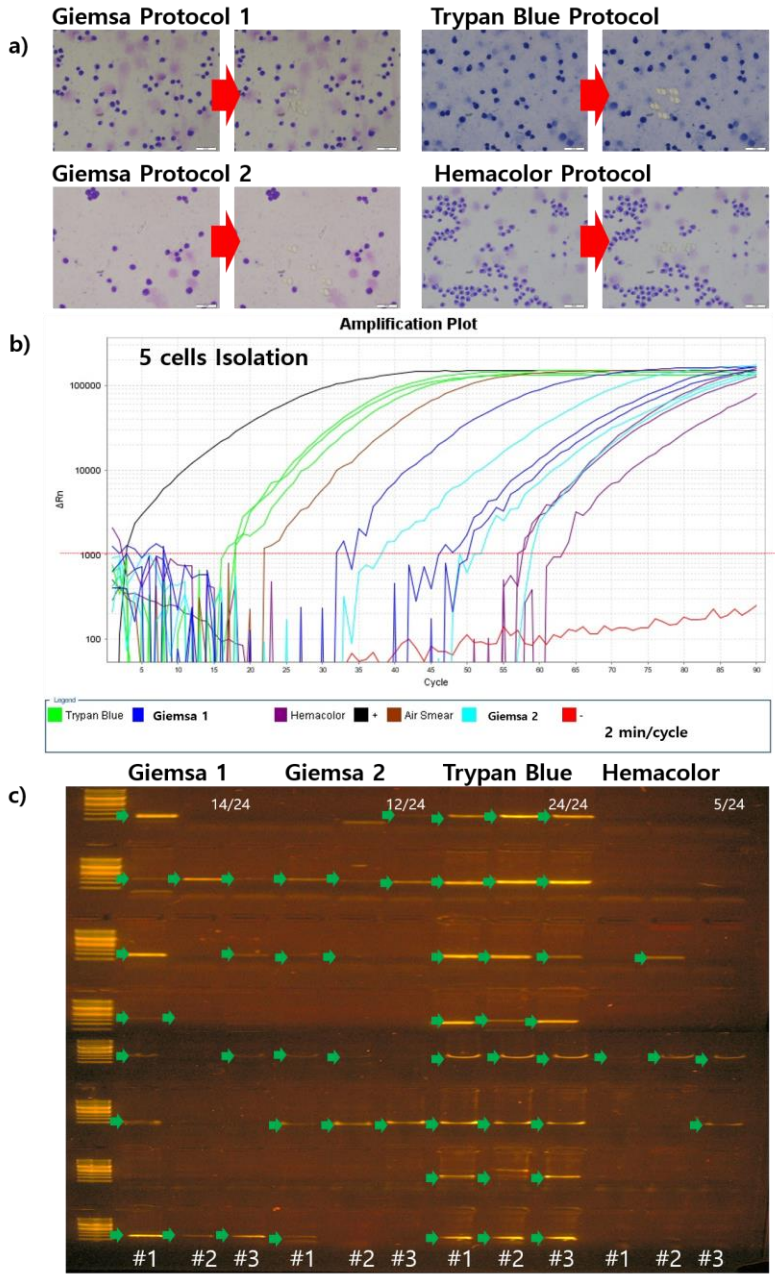


Figure 4.1. 2 Four different staining methods and WGA efficiency. a) Isolating

Minimum number of isolating cells needed to be found for

further research, since staining process lowers WGA efficiency. Therefore, we measured WGA efficiency by differing retrieving cell numbers. Two cell lines, HL60 and K562, were adhered on different ITO glass and fixed with MeOH for 30 seconds. After giemsa staining protocol from SNUH, cells were isolated to single cap. 1, 5, 25, 100 cells were isolated in single cap with 5 replications for each retrieving number of cells. Each sample was lysed and amplified with optimized protocol which I introduced above, and final product was purified for further assays to measure the uniformity of amplification. Further assays include processing PCR with 8 different primers targeting genome, target sequencing with leukemia panel from SNUH and whole genome sequencing with low depth.

We first saw initial amplification rate after staining cells. (Fig 4.1.3) Two positive control was used. One was amplifying 30 pg of extracted genomic DNA, quantity same as 5 cells of genomic DNA, and other was amplifying 5 isolated cells which were fixed only. As we expected, more retrieved cells in single cap showed higher initial amplification rate. Both positive control showed similar initial amplification rate which means there are almost no difference of

genomic DNA quality between extraction from bulk samples and laser isolation after fixation. Since initial amplification rate of amplifying 25 stained cells was similar as positive control, we expected that 25 cells will show high quality of amplified product.

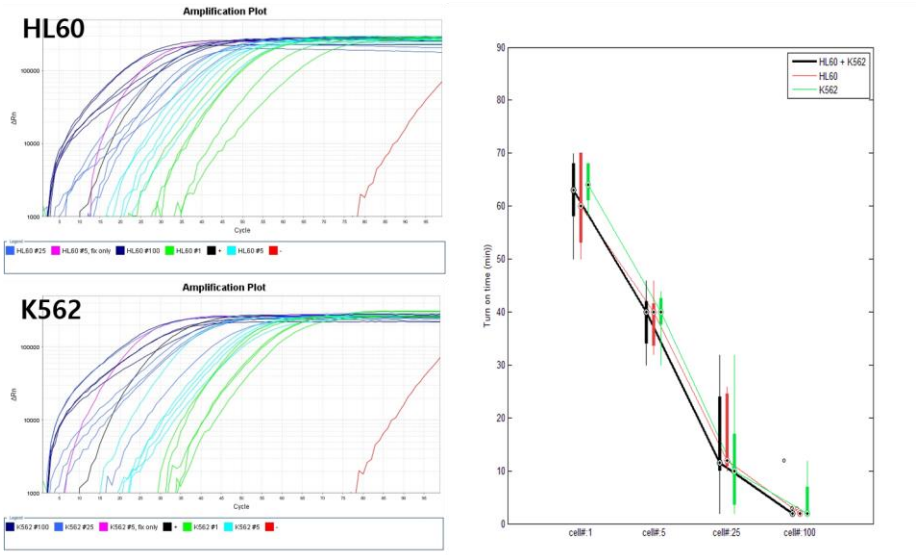


Figure 4.1. 3 WGA of two cell lines, HL60 and K562

After purification of WGA products, they were PCR targeting 8 different locations. Among 8 different locations 6 targets were single nucleotide polymorphism (SNP)marker of HL60 with high allele frequency, 1 site for SNP with 50 % of allele frequency in HL60 and 1 site for reference. Whether primers were targeting proper locations and whether cell lines contain known SNP sites were pre-tested and validated with genomic DNA extracted from

bulk HL60 cell lines. Position of SNP, primer sequences used for PCR and lengths of PCR products are listed in table 4.1.

ChrID	Pos	Ref	Alt	Length	AF %	Forward / Reverse Primer
2	228243905	G	A	303	>99	AGGCCCTTCTTCACATCAAGT
						AACCAGAAGATTGGGATAAAAGGCA
4	2175733	A	G	271	~50	TCTAGACCTGCCACTGGGAA
						ATGCAGCAGGTGCTGAGTAA
5	205565	G	A	249	<1	GAGCCACATGAGTCTGCCAT
						AGAGCCAGGCTTTTGCTGAA
7	25266400	G	A	333	>99	ACTGCCCATGCACTTTGACT
						CCACACTCCTTCGCCAACTT
10	5435918	G	A	163	>99	CTTCCTTGGGGACCACATCC
						CCCATCGTCTCTGCTGACAA
15	4243837	T	C	189	>99	GTGTGCGGAAGGTACGGTTA
						TTGCTCCTGCTCAGGTCTTG
16	20648702	G	A	309	>99	GGATGACTGGAGCAGGGAAG
						TGGGCAGCATCCATTGAGAG
17	7214824	C	T	332	>99	TCCCTGGGCCTACTACCTTC
						TTTGCCATGGCCTTGATTGC

Table 4. 1 SNP locations for PCR validation of uniformity of WGA

PCR success rate was closely related to initial amplification rate which depends on number of retrieving cells. (Fig 4.1.4) Retrieving more than 25 cells were enough to obtain high quality of amplified whole genome which ensures >90 % of PCR success rate. PCR success is closely related to applicability on diagnosis, since until now many tests are based on sequencing PCR products of patients' sample.

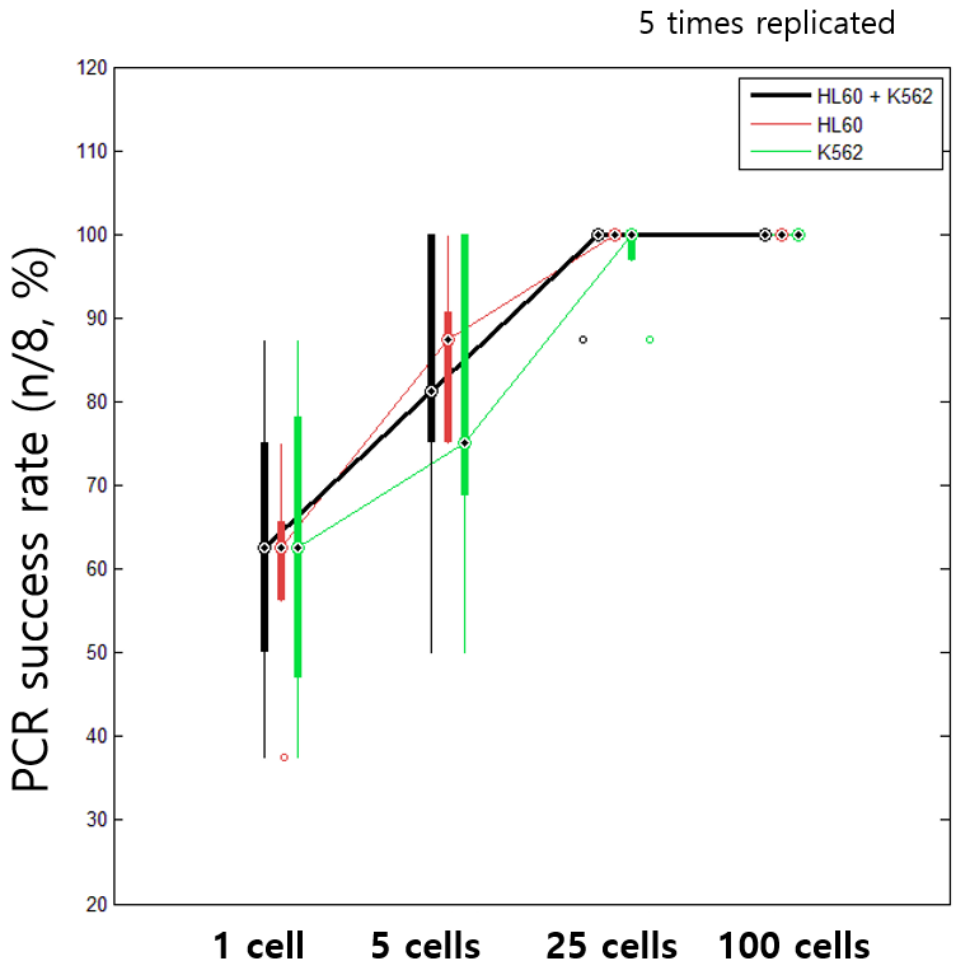
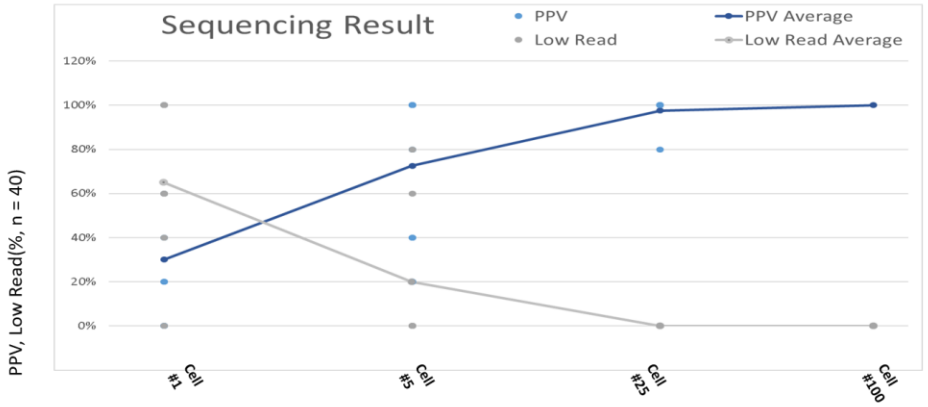


Figure 4.1. 4 PCR success rate of cell line WGA

We purified every PCR product and send for sequencing to see whether designed SNP has been amplified. (Fig 4.1.5) Sequencing data from PCR products using Genomic DNA extracted from bulk sample as a template show our designed primer amplifies the correct region and alternative and reference sequence of our

cell line follows the known allele frequency. Mixed sequencing result means reference sequence and alternative sequence exists in similar portion. Same as PCR success rate, true positive rate increases as retrieving cell number increases. Isolating more than 25 cells ensures >90 % of true positive SNP sequencing result. False rate includes false negative data points and not estimated data points due to low signal of sequencing data. Latter type of data points is major portion of false rate, which means this platform's PPV is high enough to be compared to conventional sequencing platform even isolating single cell. Improving low signal rate for higher sensitivity is remaining issue. Catching heterogeneous SNP was not suitable for this platform. WGA starting from low input of genomic DNA has high probability of ADO. This stochastic phenomenon will remain until better WGA method develops. Fortunately, this platform can select the cell to analysis. Therefore, user do not need to worry about missing low portion of sequence, which is the case when analyzing bulk sample. User can choose targeted cells which expects to contain same genomic DNA and select homogeneous SNP to analysis from targeted cells. Isolated 5 cells without staining step show high

quality of sequencing data compared to bulk sequencing data. There was no low signal sequencing data nor false data. This result also implies staining the cell lowers the quality of genomic DNA when using MDA for WGA.



	>n/2 is True Positive
	>n/2 is False Positive or Not counted

	BULK	MtOH (5 cell)	1 cell	5 cell	**25 cell**	**100 cell**
HL60_chr4_ALT_G_REF_A	Mixed	Mixed	0/1/1/3	2/2/1/0	0/4/1/0	0/5/0/0
HL60_chr5_ALT_A_REF_G	REF	REF	0/0/2/3	0/0/5/0	0/0/5/0	0/0/5/0
HL60_chr2_ALT_A_REF_G	ALT	ALT	5/0/0/0	5/0/0/0	5/0/0/0	5/0/0/0
HL60_chr7_ALT_A_REF_G	ALT	ALT	1/0/1/3	5/0/0/0	5/0/0/0	5/0/0/0
HL60_chr10_ALT_A_REF_G	ALT	ALT	0/0/0/5	2/0/0/3	5/0/0/0	5/0/0/0
HL60_chr15_ALT_C_REF_T	ALT	ALT	0/0/0/5	1/0/0/4	5/0/0/0	5/0/0/0
HL60_chr16_ALT_A_REF_G	ALT	ALT	3/0/0/2	5/0/0/0	5/0/0/0	5/0/0/0
HL60_chr17_ALT_T_REF_C	ALT	ALT	0/0/0/5	4/0/0/1	5/0/0/0	5/0/0/0

(ALT/Mixed/REF/Not count due to low read)

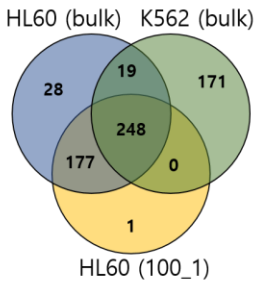
Figure 4.1. 5 SNP sequencing of cell line WGA

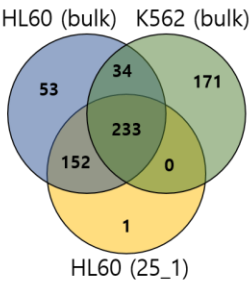
WGA product was also send for target sequencing using leukemia panel from SNUH. 2 samples were selected to send for target sequencing from each group among 5 replications. Single cell isolated result was not send for target sequencing due to high

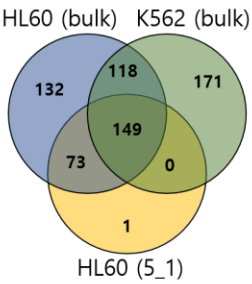
false rate during PCR sequencing. Target sequencing panel targets 477 SNPs of HL60 and 441 SNPs of K562. 267 SNPs are common between HL60 and K562 and 205 SNPs and 171 SNPs are unique to each cell lines. Bulk genomic DNA without any amplification was screened using this panel. Our HL60 cell line contained 472 SNPs and K562 cell line contained 438 SNPs with common 267 SNPs. Based on SNPs which our cell line possessed, we measured hit rate of our WGA products and saw whether the result can distinguish the cell kind.

There was slight decrease of hit portion between isolating 25 cells and 100 cells and big decrease when isolating 5 cells. (Fig 4.1.6) Maximum sensitivity was 90.04 % and lowest was 31.74 % when the denominator is each cell line's whole SNPS. This value decreased slightly to maximum sensitivity 86.34 % and lowest sensitivity 23.98 % when the denominator is each cell line's unique SNPs. Amazingly there were no hits of different cell types' unique SNPs making specificity and PPV 100 %. Therefore, even sensitivity was not high, distinguishing cell kind using this panel can be assured. Problem of high false positive rate due to low signal had shown same as PCR sequencing. All of uncalled SNPs

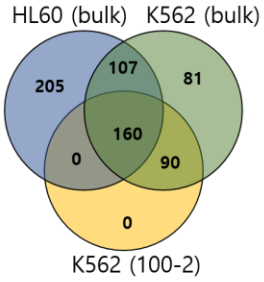
were not due to false hit, but because of they were sequenced in low depth or not read.

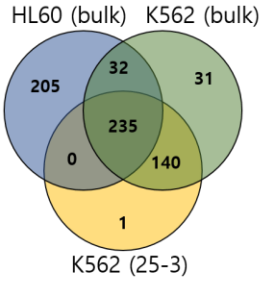


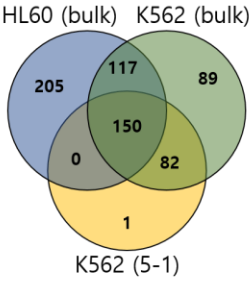




HL60 Samples	HL60 SNPs	n/472	HL60 only	n/205	K562 only	n/171	Exception
HL60_100-1	425	90.04%	177	86.34%	0	0.00%	1
HL60_100-4	393	83.26%	157	76.59%	0	0.00%	1
HL60_25-1	385	81.57%	152	74.15%	0	0.00%	1
HL60_25-2	350	74.15%	137	66.83%	0	0.00%	1
HL60_5-1	222	47.03%	73	35.61%	0	0.00%	1
HL60_5-2	211	44.70%	74	36.10%	0	0.00%	1







K562 Samples	K562 SNPs	n/438	HL60 only	n/205	K562 only	n/171	Exception
K562_100-2	370	84.47%	0	0.00%	137	80.12%	0
K562_100-4	250	57.08%	0	0.00%	90	52.63%	0
K562_25-1	267	60.96%	0	0.00%	95	55.56%	1
K562_25-3	375	85.62%	0	0.00%	140	81.87%	1
K562_5-1	232	52.97%	0	0.00%	82	47.95%	1
K562_5-3	139	31.74%	0	0.00%	41	23.98%	0

Figure 4.1. 6 Target sequencing result of HL60 and K562

Higher depth of sequencing could improve sensitivity. But expensive sequencing cost would limit broad use of our platform. Again, better WGA method or staining reagent which stains nuclei but not decrease WGA efficiency needs to be developed. We did not see any difference between cell kinds showing the potential that this platform is appropriate for wider application regardless of cell kinds. Interestingly, there were one new hit in some groups which were supposed to be each cell line's SNP but not observed at genomic DNA extracted from bulk sample. This SNP seems to have low allele frequency which have might drop out during target capturing, but amplified selectively while WGA. This doesn't imply our platform may have higher sensitivity but have a weakness in detecting heterogeneous SNPs.

WGS was also processed using purified WGA products. As I mentioned before, general depth of WGS is too expensive to conduct in several samples. Therefore, we sequenced each sample in low depth. Using low depth WGS data we plot genome coverage and CNV for accurate estimation of uniformity of amplification.

First of all, to see the uniformity of amplification we plot genome coverage by normalized sequencing depth. (Fig 4.1.7) High

genome coverage in low normalized sequencing depth means low amplification bias appeared while WGA. To obtain genome coverage we used constant bin size. We set bin size as 9000, resulting 343,964 bins along whole genome and aligned NGS reads to every bins which represent the aligned reads. Read representing bin was randomly chosen by random selection of the read. After certain number of reads had been selected, the number of unique bin are counted which represent the genome coverage.

Genomic DNA without WGA appeared to show highest genome coverage. Genome coverage portion was saturated, below 90 % before higher sequencing depth. Different mappability along genome is expected to be the cause of low coverage. As PCR success rate, sequencing result and target sequencing showed, higher number of isolated cells for WGA covers more genome. Slight decrease of genome coverage was shown between 25 cells isolated group and 100 cells isolated group, while big decrement appeared when isolating 5 cells for WGA. For general genome analysis high genome coverage is essential. Platform introduced in this thesis appears to cover 70 % of genome when isolating 25 homogeneous cells which are stained with giemsa staining method.

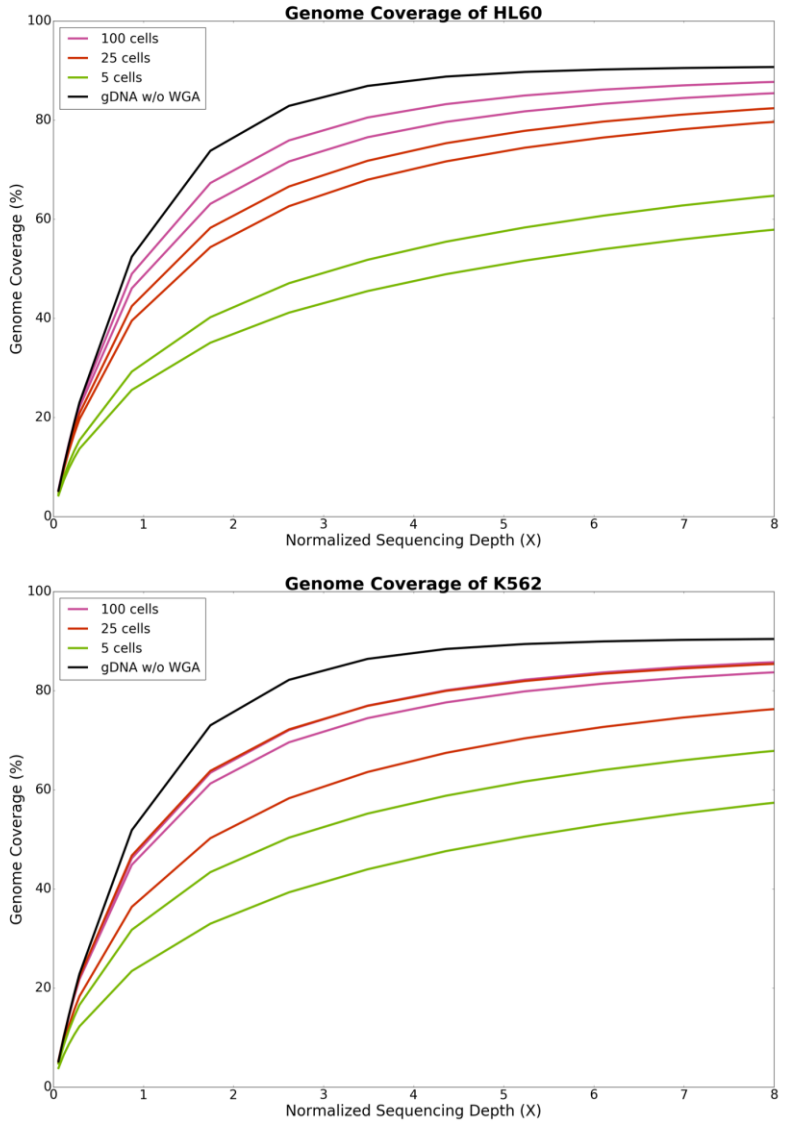


Figure 4.1. 7 Genome coverage of isolated cell lines' WGA product

We also depicted CNV plot to visualize uniformity of amplification. CNV was computed based on variable length adjusting size depends on align-ability of sequences following

Nicholas Navin's work. Entire human chromosomes, except sex chromosome was aligned with WGS data based on hg18 reference human genome. (Fig 4.1.8, 4.1.9) Narrow gray line is read number of each bins and thick blue line is the average of them. This means higher correlation of gray line and blue line expresses higher uniformity of amplification, which is the result of low amplification bias. The reason that blue line does not stays at copy number 2, is that the genomic DNA we used, HL60 and K562, is leukemia cell lines' genomic DNA. Since extracted genomic DNA did not go through any amplification step, there will be no amplification bias and as the result gray line in CNV stays near to the blue line, showing high correlation. There are some variations but these variations are inevitable due to low depth of sequencing and also it's almost negligible.

Isolated cells show higher variations and correlation between blue line and gray line decreases as isolated cell number decreases. Same again, CNV plot of WGA product of 100 isolated cells and 25 isolated cells have little correlation decrement, while 5 isolated cells show immense variations. Even correlation have decreased when amplifying isolated cells, blue line is stably expressing the

copy number of the product. General outline of copy number is expressed properly and unique points of genome, such as high copy number or deletion, were also observed. We did not see any quality difference between WGA product of cell line HL60 and K562.

We demonstrated our novel platform by staining two leukemia cell lines and see uniformity of amplification by focusing on applicability in diagnosis and research field. WGA products were amplified with SNP targeting primers and sequenced and show high specificity at detecting homogenous SNPs. Target sequencing conducted with leukemia panel and show high specificity. Sensitivity was low but can overcome by increasing the number of isolating cells or developing new WGA method for low amplification bias. We also process WGS in low depth. Genome coverage was plotted using constant length of bins and CNV was depicted using various length of bins, which are decided by mappability of the genome locations. We found out isolating more than 25 cells show reliable WGA product for genome analysis. I believe validation of our platform using cell lines expressed the state of ready to use in diagnosis and research fields.

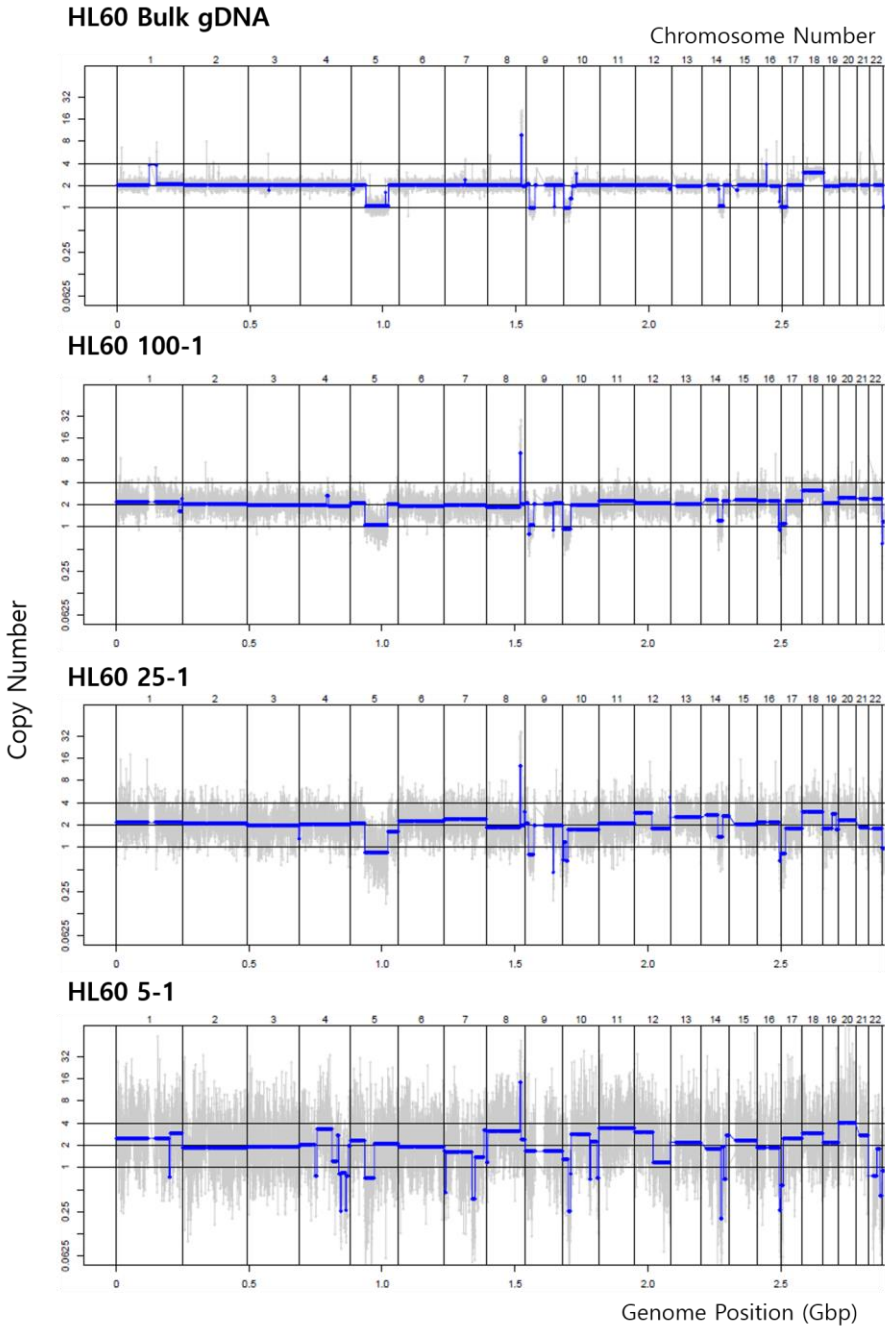


Figure 4.1. 8 CNV of HL60 cell line

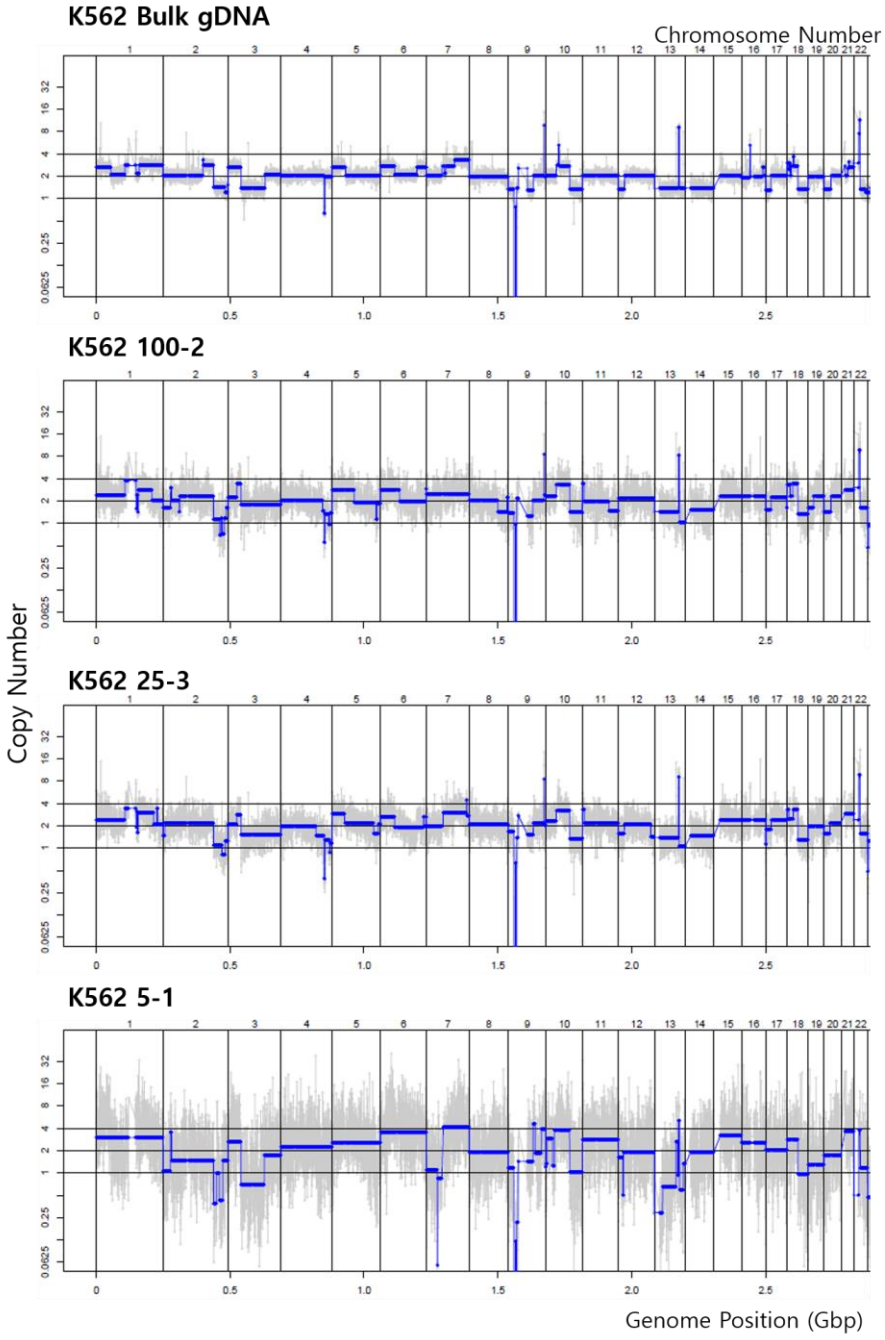


Figure 4.1. 9 CNV of K562 cell line

4.2 Applying on Patients' Blood Sample

We demonstrated our platform using patients' blood sample. Patients' blood sample were provided from prof. Lee in SNUH. Patient with hereditary disorder was chosen. This is because they have specific changes in their genomic DNA which would show our platform properly analysis genome, and every cell have the specific changes which will reduce random noises due to heterogeneous population of cells. Blood samples with two different disorder was collected, one is Down syndrome (DS) which is known as trisomy 21 and other is CATCH22 syndrome (CATCH22), also known as 22q11.2. Both blood samples were smeared on the ITO coated glass in less than 12 hours after being taken out from a body. Fixation and staining with giemsa method was followed right away. Sample was stored in −20 degrees at hospital and sent to laboratory for WGA.

This storage and staining was specially done for our experiment. Generally, blood sample cannot be stained right after

being taken out from body, because there are so many patients in hospital waiting for diagnosis using their blood. We heard, most of the blood samples are stained after 12 hours or 1 day due to busy time schedule. Moreover, since there are a number of blood samples, hospital cannot store them in a refrigerator or deep freezer. Most of the samples are stored in room temperature. We worried about this storage condition. Because to amplify genome by MDA method, longer than 10 kbps of genomic DNA is required for stable amplification. Storing at room temperature degrades genomic DNA to short DNAs[32]. Therefore, we assumed storing patients' sample at room temperature will lower the quality of genomic DNA for WGA.

To verify our assumption, we got 3 slides of each disorder. DS sample was stored at hospital for 1 day while CATCH22 sample was stored for 8 days before sending to our laboratory. One of the slide was used for isolating cells right after it arrived to laboratory and stored at room temperature for 8 days. After storing this glass at room temperature, we isolated the cells again. Room temperature was 20~23 degrees. Another slide was freeze and thaw for 20 times right after it arrived to laboratory. This

slide was used to test the effect of freezing and thawing to quality of genomic DNA. Other slide was stored in −20 degrees for 9 days right after it arrived to laboratory. From each condition, 5 and 25 cells were isolated 3 times and 100 cells were isolated 2 times. All of isolated cells were lysed and amplified with optimized WGA protocol.

After purifying WGA products, they were amplified with primers listed at Table 4.1. PCR success number dramatically decreased when samples were stored at room temperature. (Fig 4.2.1) Storing samples at room temperature for 8 days, PCR success number was extremely lower compared to samples isolated right after arrival. Unfortunately, this result indicates that our WGA method, MDA, cannot be applied to most of the precious sample stored at hospital, which would have been stored at room temperature. However, we succeeded in amplifying whole genome with high quality for genome analysis using patient's blood samples. Storing condition was critical for maintaining input genomic DNA. There was some decrease of PCR success number while storing at −20 degrees, but still quality of WGA product was good enough for genome analysis. CATCH22 samples show

slightly low PCR success number. 8 more days storing at hospital seems to be the reason of this decrement. Interestingly, freeze and thaw did not affect WGA efficiency. When storing long DNAs, less freeze and thaw is essential for general high quality experiment[33]. DNA are damaged when surrounding water froze to ice. It seems surroundings of smeared cell became dry after fixation, staining and air drying for observation.

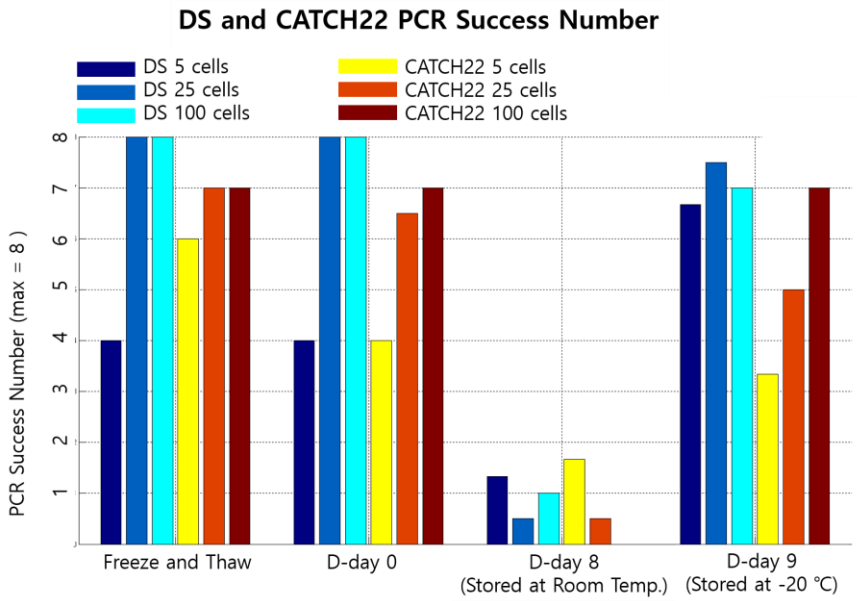


Figure 4.2. 1 PCR success number of patients' blood sample in different storage conditions

For accurate analysis we tried WGS in low depth to see genome coverage and CNV of each whole genome amplified products. (Fig 4.2.2) Analyzing method is same as chapter 4.1.

Three conditions, isolating cells after freeze and thaw 20 times, after 9 days of storage at −20 degrees and immediately when arrived at laboratory, show similar genome coverage. Genome coverage has slightly increased as more cells were isolated from 25 cells to 100 cells. Since coverage curve is not saturated, higher sequencing depth will be required for stable result. Compared to the result at chapter 4.1, demonstrating with patients' blood sample shows lower quality of WGA products. Lower quality is due to inevitable condition of handling patients' blood sample. Longer time period between sample preparation and isolating cells for WGA should be the biggest cause of lower quality, instead of sample difference. If our platform can be instrumented inside the hospital, efficient genome analysis could be done by using high quality input.

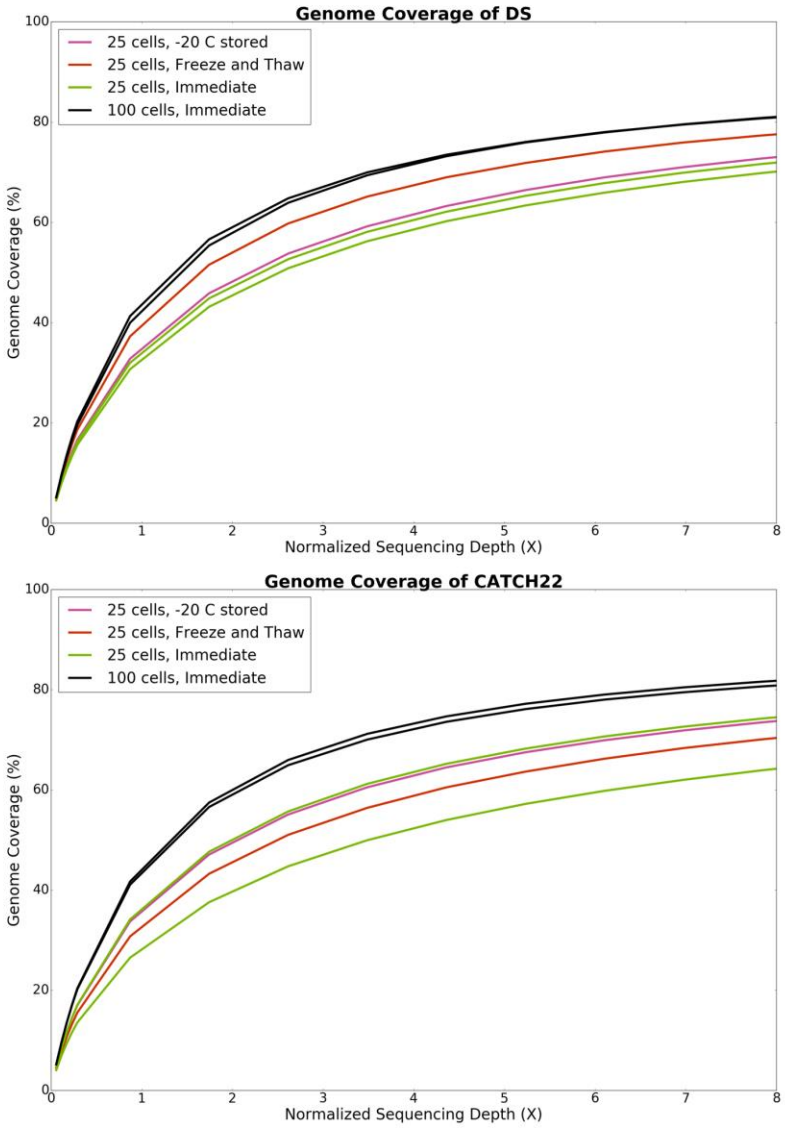


Figure 4.2. 2 Genome coverage of WGA products using blood samples

CNV plot were also displayed for each condition. (Fig 4.2.3, 4.2.4) All CNVs from three different conditions show same

variation when isolating same number of cells. This data implies storing samples at −20 degrees ensures high quality of genomic DNA. Even though storing samples at −20 degrees was better for maintaining high quality of genomic DNA than storing samples at room temperature, we saw higher variations compared to the results which were demonstrated with cell lines. Since blood cell is not much fragile than cell of the cell line, I believe variations can be decreased if duration between drawing blood and staining or storing after stained and isolating shrinks.

In case of DS, we can clearly see that 21^{st} chromosomes are estimated as 3 from every condition even variations were high. 25 cells were enough to detect trisomy for every condition. However, in case of CATCH22, we could not find any unique points from CNV plot. CNV using WGA of 100 isolated cells show suspicious deletion of 22^{nd} chromosome, but not clear. There are two reasons that we failed to detect deletion at 22^{nd} chromosome. First of all, low depth sequencing was not appropriate to detect small abnormal parts which is 3 Mbp. Secondly, bins of various length fail to pin point the certain deletion part. Low depth sequencing, and analyzing strategy for low depth sequencing was not

optimized to detect small deletions of genome. Higher sequencing depth would overcome this problem.

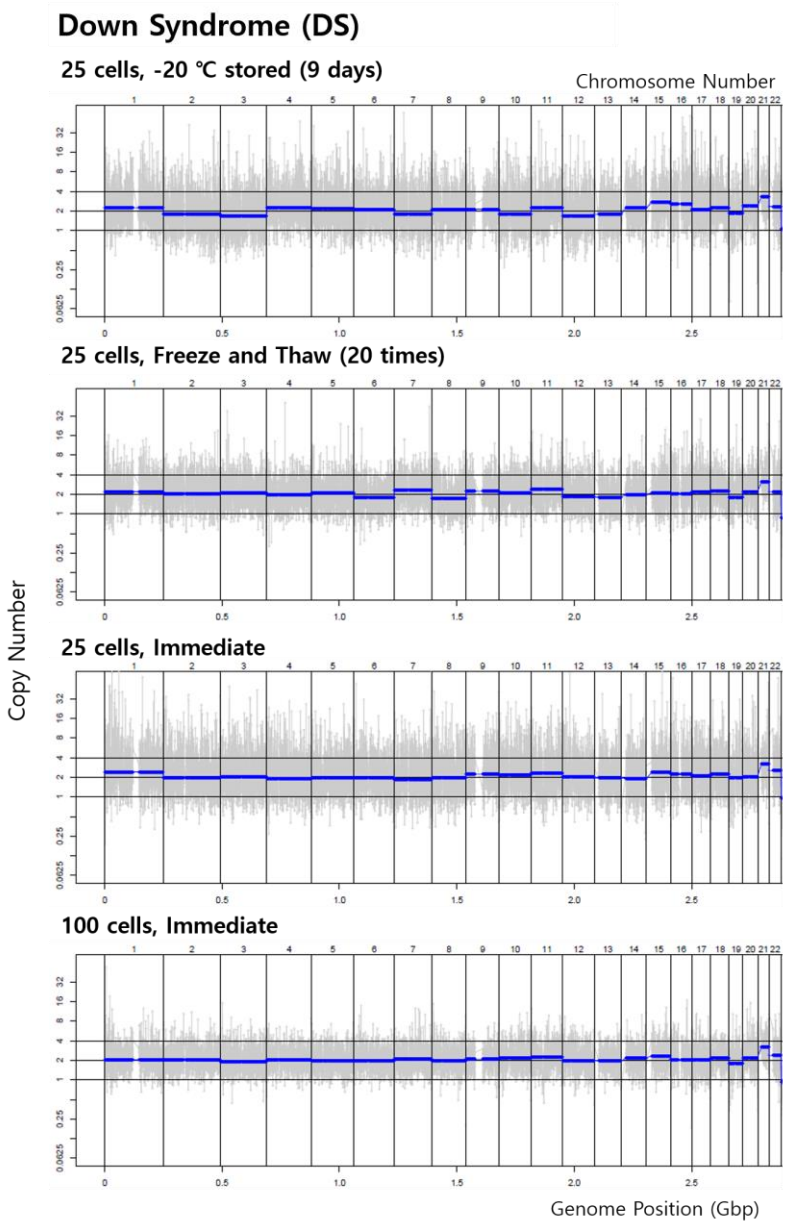


Figure 4.2. 3 CNV plot of WGA DS samples

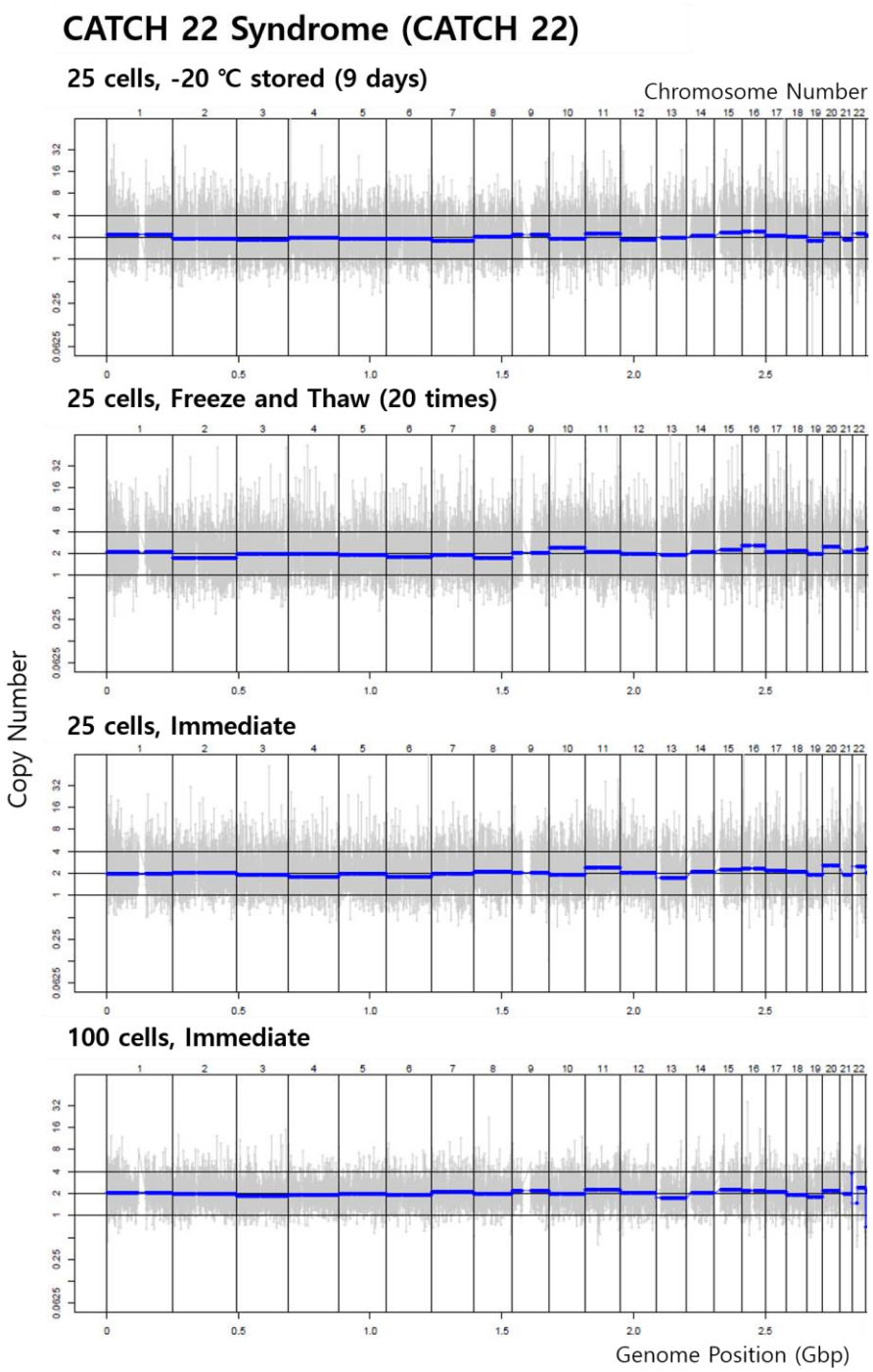


Figure 4.2. 4 CNV plot of WGA CATCH 22 sample

We have demonstrated our platform using patients' blood sample. Storage condition was critical for high quality WGA products. For accelerate development hospital requires cold room to store their patients' sample for future analysis. Storing samples at −20 degrees maintained the quality of genomic DNA. There was slight decrement of PCR success rate, but there was no difference of genome coverage or CNV plot. PCR success rate is more sensitive to stochastic event, so estimating uniformity of amplification is more reliable by estimating via genome coverage or CNV plot.

Applying our platform to patients' blood was not simple. Not only developing WGA method, storage condition, which we cannot modify, required to be optimized for high quality of genomic DNA. Nonetheless, we saw the potential that our platform can be used in real patients' sample if sample provider and analyzer cooperates. Cooperation will be much easier if physical distance gets closer, like instrumenting this platform in the hospital.

Conclusion

In this thesis, I first introduced conventional genome analyzing methods for single cell level. Among conventional methods, LMD, which can isolate targeted cells for further analysis, was proper to combine genomic and morphologic information. However, since LMD requires fragile frame to adhere the cell and to burn around them to isolate the target cells, applications are limited. Moreover, WGA protocol to prepare isolated cells for genome analysis was not optimized. For general and wide genome analysis platform, I and my colleagues developed novel single isolating technology and optimized WGA protocol.

Isolating platform was focused on safe retrieval of isolating cells while maintaining high quality of cell image before isolation to distinguish cell types. We failed to isolate cells on normal slide glass. Instead, the potential of isolating cells on normal slide glass was shown by modeling the absorbance of pulse laser between glass and the cell. We used ITO coated glass and IR pulse laser for safe retrieval. Various sample types were demonstrated to show isolating platform has high applicability.

Whole process starting from cell isolation to WGA were optimized. We found out the pre-process was critical to achieve unbiased genome amplification. Retrieving knowhow, cell lysing methods and random hexamer binding were optimized for high initial amplification rate. High correlation was observed between initial amplification rate and uniformity of amplification which is directly related to quality of input for genome analysis.

Cell lines were used to demonstrate the platform. For ready to go state, cell was stained with giemsa staining method. Giemsa staining method is generally used staining method which stains nuclei to distinguish various cell types. Unfortunately, staining dramatically decrease uniformity of amplification. We found more

than 25 isolated cells are required for reliable genome analysis after WGA. We measured the quality of WGA product by amplifying with 8 different primers targeting SNPs, target sequencing with leukemia panel and whole genome sequencing with low depth and saw genome coverage by normalized sequencing depth and CNV plot.

Moreover, blood samples from hospital were used to see our platform is ready to go state. Following same protocol when demonstrating using cell lines, chromosomal abnormality was observed plotting CNV. Small deletion in chromosome wasn't detected due to low depth of whole sequencing. Storing samples were critical, since genomic DNA degrades when stored at room temperature. Samples require to be stored at −20 degrees for high quality of input genomic DNA for further genome analysis. Efficient WGA method to amplify short fragments of genomic DNA and higher uniformity guaranteed amplification were required for further applications.

Single cell genome analysis accelerates broadening human knowledge of biology. Platform introduced in this dissertation is highly maturated to be applied to diagnosis and research field.

Instrumenting in hospital will improve quality of input genomic DNA which ensures meaningful genome analysis. Isolating cells on bare slide glass and amplifying short fragmented genomic DNA is remaining issue, but I believe many researchers will come out with the solution soon enough. This novel platform will contribute to extend biological knowledge which was unrevealed due to heterogeneous signal or absent of morphologic information while analyzing genomic information.

Bibliography

[1] D Roulston, R III Espinosa, M Stoffel, Graeme I. Bell, MM Le Beau, “Molecular genetic of myeloid leukemia: identification of the commonly deleted segment of chromoseom 20”, Blood, vol. 82, pp. 3424-3429, 1993.

[2] Cinnamon S. Bloss, Dilip V. Jeste and Nicholas J. Schork, “Genomics for Disease Treatment and Prevention”, Psychiatr. Clin. North Am., vol. 34, pp. 147-166, 2011.

[3] Charles G. Mullighan, Letha A. Phillips, Xiaoping Su, Jing Ma, Christopher B. Miller, Sheila A. Shurtleff, James R. Dosning, “Genomic Analysis of the Clonal Origins of Relapsed Acute Lymphoblastic Leukemia”, Science, vol. 322, pp. 1377-1380, 2008.

[4] “When Cancer Comes Back: Cancer Recurrence”, American Cancer Society, 2015.

[5] David J Drasin, Tyler P Robin and Heide L Ford, “Breast cancer epithelial-to mesenchymal transition: examining the functional consequences of plasticity”, Breast Cancer Research, 13:226,

2011.

[6] Devon A. Lawson, Nirav R. Bhakta, Kai Kessenbrock, Karin D. Prummel,Ying Yu, Ken Takai, Alicia Zhou, Henok Eyob, Sanjeev Balakrishnan, Chih-Yang Wang, Paul Yaswen, Andrei Goga and Zena Werb, "Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells, Nature, vol. 526, pp. 131-135, 2015.

[7] Ornatsky O, Bandura D, Baranov V, Nitz M, Winnik MA and Tanner S, "Highly multiparametric analysis by mass cytometry", J. Immunol. Methods, vol. 361, pp. 1-20, 2010.

[8] Nicholas E. Navin, "The first five years of single-cell cancer genomics and beyond", Genome Res., vol. 25, pp. 1499-1507, 2015.

[9] Adam K. White, Michael Vanlnsberghe, Oleh I. Petriv, Mani Hamidi, Darek Sikorski, Marco A. Marra, James Piret, Samuel Aparicio and Carl L. Hansen, "High-trhoughput microfluidic single-cll RT-qPCR", PNAS, vol. 108, pp. 13999-14004, 2011.

[10] Richard J. Zaino, James Kauderer, Cornelia Liu Trimble, Steven G. Silverberg, John P. Curtin, Peter C. Lim and Donald G. Gallum, "Reproducibility of the Diagnosis of Atypical Endometrial

Hyperplasia", American Cancer Society, vol. 106, pp. 804-811, 2006.

[11] K. Schutze, H. Posl and G. Lahr, "Laser micromanipulation system asuniversal tools in cellular and molecular biology and in medicine", Cellular and Molecular Biology, vol. 44, pp. 735-746, 1997.

[12] Kristina E Aaltonen, Anna Ebbesson, Caroline Wigerup and Ingrid Hedenfalk, "Laser capture microdissection (LCM) and whole genome amplification (WGA) of DNA from normal breast tissue- optimization for genome wide array analyses", BMC Research Notes, vol. 4, 2011.

[13] Martha S. Rook, Sctt M. Delach, Galina Deyneko, Andrew Worlock and Jia Liu Wolfe, "Whole Genome Amplification of DNA from Laser Capture-Microdissected Tissue for High-Throughput Single Nucleotide Polymorphism and Short Tandem Repeat Genotyping", American Journal of Pathology, vol. 164, pp. 23-33, 2004.

[14] Pavel Katsel, William Byne, Panos Roussos, Weilun Tan, Larry Sieverand Vahram Haroutunian, "Astrocyte and Glutamate Markers in the Superficial, Deep, and White Matter Layers of the

Anterior Cingulate Gyrus in Schizophrenia",
Neuropsychopharmacology, vol. 36, pp. 1171-1177, 2011.

[15] Jose M. Lage, John H. Leamon, Tanja Pejovic, Stefan Hamann, Michelle Lacey, Deborah Dillon, Richard Segraves, Bettina Vossbrinck, Antonio Gonzalez, Daniel Pinkel, Donna G. Albertson, Jose Costa and Paul M. Lizardi, "Whole Genome Analysis of Genetic Alterations in Small DNA Sample using Hyperbranched Strand Displacement Amplification and Array-CGH", Gemome Researhce, vol. 13, pp. 294-307, 2003.

[16] Qianqian Zhu, Qiang Hu, Lori Shepherd, Jiamin Wang, Lei Wei, Carl D.Morrison, Jeffrey M. Conroy, Sean T. Glenn, Warren Davis, Marilyn L. Kwan, Isaac J. Ergas, Janise M. Roh, Lawrence H. Kushi, Christine B. Ambrosone, Song Liu and Song Yao, "The Impact of DNA Input Amount and DNA Source on the Performance of Whole-Exome Sequencing in Cancer Epiemiology", Cancer Epidemiol. Biomarkers. Prev., vol.24, pp. 1207-1213, 2015.

[17] Jongsuk Chung, Dae-Soon Son, Hyo-Jeong Jeon, Kyoung-Mee Kim, Gahee Park, Gyu Ha Ryu, Woong-Yang Park and Donghyun Park, "The minimal amount of starting DNA for Agilent's hybrid

capture-based targeted massively parallel sequencing", Scientific Reports, vol. 6, 2016.

[18] Hakan Telenius, Adele H. Pelmear, Alan Tunnacliffe, Nigel P. Carter, Annemarie Behmel, Malcolm A. Ferguson-Smith, Magnus Nordenskjold, Roswitha Pfragner and Bruce A.J. Ponder, "Cytogenetic Analysis by Chromosome Painting Using DOP-PCR Amplified Flow-Sorted Chromosomes", Genes, Chromosome & Cancer, vol. 4, pp. 257-263, 1992.

[19] Heike Fiegler, Philippa Carr, Eleanor J. Douglas, Deborah C. Burford, Sarah Hunt, James Smith, David Vetrie, Patricia Gorman, Ian P.M. Tomlinson and Nigel P. Carter, "DNA Microarrays for Comparative Genomic Hybridization Based on DOP-PCR Amplification of BAC and PAC Clones", Genes, Chromosome & Cancer, vol. 36, pp. 361-374, 2003.

[20] Robert Pinard, Alex de Winter, Gary J Sarkis, Mark B Gerstein, Karrie R Tartaro, Ramona N Plant, Michael Egholm, Jonathan M Rothberg and John H Leamon, "Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing", BMC Genomics, vol. 7, 2006

[21] Frank B. Dean, Seiyu Hosono, Linhua Fang, Xiaohong Wu, A.

Fawad Faruqi, Patricia Bray-Ward, Zhenyu Sun, Qiuling Zong, Yuefen Du, Jing Du, Mark Driscoll, Wanmin Song, Stephen F. Kingsmore, Michael Egholm and Roger S. Lasken, "Comprehensive human genome amplification using multiple displacement amplification", PNAS, vol. 99, pp. 5261-5266, 2002.

[22] Chenghang Zong, Sijia Lu, Alec R. Chapman and X. Sunney Xie, "Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell", Sicence, vol. 338, pp. 1622-1626, 2012..

[23] Chen Minfeng, Pengfei Song, Dan Zou, Xuesong Hu, Shancen Zhao, Shengjie Gao and Fei Ling, "Comparison of Multiple Displacement Amplification (MDA) and Multiple Annealing and Looping-Based Amplification Cycles (MALBAC) in Single-Cell Sequencing", PLoS ONE, vol. 9, 2014.

[24] Sinuhe Hahn, Alex M. Garvin, Edoardo di Naro and Wolfgang Holzgreve, "Allel Drop-Out Can Occur in Alleles Differing by a Single Nucleotide and Is Not Alleviated by Preamplification on Minor Template Increments", Genetic Testing, vol. 2, pp. 351-355, 2009.

[25] Howon Lee, Hyoki Kim, Sungsik Kim, Taehoon Ryu,

Hwangbeom Kim,Duhee Bang and Sunghoon Kwon, “A high-throughput optomechanical retrieval method for sequence-verified clonal DNA from the NGS platform”, Nat. Commun. 6:6073, 2015.

[26] Samuel S. Mao, “Laser Ablation, Fundamentals & Applications”, 2005.

[27] Zhuangchun Wu, Zhihong Chen, Xu Du, Jonathan M. Logan, Jennifer Sippel, Maria Nikolou, Katalin Kamaras, John R. Reynolds, David B. Tanner, Arthur F. Hebard and Andrew G. Rinzler, “Trnasparent, Conductive Carbon Nanotube Films”, Science, vol. 305, pp. 1273-1276, 2004.

[28] Oleksandr Malik and F. Javier De la Hidalga-W, “Physical and Technological Aspects of Solar Cells Based on Metal Oxide-Silicon Contacts with Induced Surface Inversion Layer”, Application of Solar Energy, ch. 5, 2013.

[29] Jeffrey C Hanson, Michael A Tangrea, Skye Kim, Michael D Armani,Thomas J Pohida, Robert F Bonner, Jaime Rodriugez-Canales and Michael R Emmert-Buck, “Expression microdissection adapted to commercial laser dissection instruments”, Nat. Protocol, vol. 6, pp. 457-467, 2011.

[30] Juan M. Gonzalez, M. Carmen Portillo and Cesareo Salz-Jimenez, “Multiple displacement amplification as a pre-polymerase chain reaction (pre-PCR) to process difficult to amplify samples and low copy number sequences from natural environments”, Environmental Microbiology, vol. 7, pp. 1024-1028, 2005.

[31] Nicholas Navin, Jude Kendall, Jennifer Troge, Peter Andrews, Linda Rodgers, Jeanne McIndoo, Kerry Cook, Asya Stepansky, Dan Levy, Diane Esposito, Lakshmi Muthuswamy, Alex Krasnitz, W. Richard McCombie, James Hicks and Michael Wigler, “Tumour evolution inferred by single-cell sequencing”, Nature, vol. 472, pp. 90-95, 2011.

[32] Alexandra Berger, Michaela Bruschek, Claude Grethen, Wolfgang Sperl and Barbara Kofler, “Poor Storage and Handling of Tissue Mimics Mitochondrial DNA Depletion”, Diagnostic Molecular Pathology, vol. 10, pp. 55-59, 2001.

[33] Wen Shao, Sonny Khin and William C. Kopp., “Characterization of Effect of Repeated Freeze and Thaw Cycles on Stability of Genomic DNA Using Pulsed Field Gel Electrophoresis”, Biopreservation and Biobanking, vol. 10, pp. 4-11, 2012.

국문 초록

본 논문에서는 채취 혈액에서 단일 세포 유전체 분석을 위한 플랫폼을 소개한다. 본 플랫폼을 이용하여 실제 환자의 혈액에서 5개의 세포를 이용하여 다운 신드롬의 유전체 변이를 관찰하였다. 유전체 분석에 필요한 세포수는 매우 중요한데, 이는 적은 수의 세포가 필요할수록 높은 확률로 특이한 마커를 찾을 확률이 높아지기 때문이다. 본 플랫폼을 개발하기 위해 단일 세포 분리 방법을 개량, 유전체 증폭 방법을 최적화하였다.

첫번째로 목표한 세포를 분리하기 위해서는 현미경으로 세포를 관찰한다음 이를 바로 분리해야 한다. 기존 방법은 플라스틱 판 위의 놓인 목표한 세포 주위를 태워 분리하였다. 본 논문에서는, Indium tin oxide 를 희생층으로, 적외선 펄스 레이저를 사용하여 안전하게 분리하였다. 사용했다. 그와 동시에 희생층(sacrificial layer)은 유리 슬라이드(slide glass)와 같이 높은 물리적, 화학적 견고성을 갖고 있어 기존 세포 관측 방법에 변화를 줄 필요가 없었다.

유전체 증폭 방법의 최적화를 통해 적은 유전체를 시료로 사용할 때 균일한 증폭물을 얻을 수 있었다. 편향이 적게 전체 유전체 증폭을 하는 것은 높은 수준의 유전체 분석을 위해 중요하다. 기존 방법의 전체 유전체 증폭은 높은 수준의 유전체 분석에 충분하지 않았다. 세포 수거,

세포 용해, 사전 DNA 변성과정은 증폭의 효율을 증대 시켰다. 이는, 균일한 증폭물을 얻게 해주었다.

본 플랫폼이 다양한 분야에 적용될 수 있음을 보여주기 위해, 두 종류의 세포주와 유전질환이 있는 두 환자로부터 혈액을 채취하여 입증하였다. 염색과 보관하는 과정에서 유전체의 질이 떨어지고 이는 비균일한 증폭물을 만들기 때문에, 신뢰성 높은 증폭물을 얻기 위해 필요한 최소한의 세포를 찾았다. 초기 증폭 속도, PCR 증폭률, 그리고 표적 시퀀싱(target sequencing) 양성율(positive rate)을 보았다. 그리고 낮은 깊이의 전유전체 서열 분석(low depth whole genome sequencing)을 통한 직접적인 유전체 분석비율(genome coverage)도 보았다. 이를 통해 김자 염색(giemsa staining)이 된 세포의 유전체 분석을 위해서는 최소 25개의 세포가 필요함을 확인하였다. 본 기술이 여러 분야에서 다양한 세포 분포로 인해 밝혀지지 않았던 생물학적으로 중요한 요소를 찾을 준비되었다고 생각한다.

주요어: 단일 세포 분리, 희생층, 전체 유전체 증폭, 유전체 분석 비율, 증폭 편향도

학번: 2012-20863