A Comparison of Two Methods of Template Amplification for Next Generation Sequencing: Implications of Polyclonal Formation on DNA Sequence for Several Cancer Tissues

by

Akram Aldilaimi

Thesis

Submitted to the College of Health and Human Services

Eastern Michigan University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Clinical Research Administration

Thesis Committee:

Irwin Martin, Ph.D., Chair

David Kass, Ph.D.

Joseph Paulauskis, Ph.D.

November 15, 2016

Ypsilanti, Michigan

Acknowledgments

I would like to thank Dr. Irwin Martin, Dr. David Kass, and Dr. Joseph Paulauskis for their guidance and support on this research project. Special thanks to Dr. Kass for accepting my application and serving as my research monitor for this project. Thanks to Paradigm Diagnostics, my workplace that allowed me to conduct my research in its laboratories and gave me the authorization to use data to complete this project. I would like also to thank my wife, Afrah, for her support during my educational journeys. Without her support and encouragement, I wouldn't have been able to complete my Master of Science degree. Thanks to my friend Roger Morey at Paradigm Diagnostic Cancer Center for his encouragement and for inspiring me to complete my master's degree. I would thank my former supervisor at Detroit National Testing Laboratory, Jafar Naderi, for his help and also my professor for cell and molecular biology, Dr. Walker for his support. Thanks to graduate school and ORDA for assistance in statistical data analysis. Finally, I would like to acknowledge my many colleagues who played a part in this research study.

Abstract

Next-Generation Sequencing (NGS) technology has advanced the field of personalized medicine by predicting effective treatments for cancer patients using genomic sequence data, including the detection of oncogenes (genes involved in cancer development). Simultaneous amplification of multiple DNA templates (referred to as polyclonal formation) is a primary disadvantage when preparing templates for NGS resulting in redundancies in DNA sequences or nonspecific noise in the sequencing. This study was conducted to compare DNA templates for sequencing prepared using two instruments the Thermo Fisher Ion Chef (IC) and Thermo Fisher OneTouch-2 (OT2). Six sequencing metrics were obtained from 114 sequencing trials for evaluation: polyclonal reads, total sequencing reads, empty microcell well reads, no-template reads, useable number reads, and library number reads. A comparison of mean sequencing metrics between the IC and OT2 methods established that for four of the six metrics, IC was the preferred operation for DNA template preparation due to less polyclonal formation.

Table of Contents

Acknowledgements ii

Abstract iii

Introduction 1

Research Design and Methodology 14

Results 17

Discussion 23

Conclusions 29

References 30

Appendix A: IRB Approval Letter for Conducting the Research Study 35

Appendix B: The Thermo-Fisher Ion Chef and Thermo-Fisher OneTouch-2 36

Appendix C: Example of Summary Report Generated in NGS 37

Appendix D: Ion Torrent Biochips Technology in NGS 38

Appendix E: Metric Variables Used for Analysis of Influence of Polyclonal Formation on NGS 39

Appendix F: Publisher Permission for Use of Ion Torrent Platform Figures 44

List of Figures

Figure 1: Emulsion PCR for DNA library preparation 1

Figure 2: Illustration of the biochip sensor in semiconductor sequencing (318 ion chip)............ 11

Figure 3: Ion Torrent Personal Genome Machine (PGM™) 12

Figure 4: Measurement of H+ during sequencing reactions with a PGM 13

Figure 5: NGS workflow process for sequencing cancer cell DNA 14

Figure 6: The polyclonal formation in NGS 18

Figure 7: The number of templates sequenced by PGM.. 19

Figure 8: The percentage of empty biochip surface... 20

Figure 9: ISP (beads) without DNA templates processed through NGS 21

Figure 10: DNA Library number outcome from NGS.. 21

Figure 11: Usable DNA library outcome from NGS 22

Introduction

Next-Generation Sequencing (NGS) revolutionized the field of personalized medicine by providing better health care outcomes for cancer patients through sequencing their cancer's genetic changes and allowing for individualized treatment approaches. NGS impacted genomic medicine (personalized medicine) by providing a technology that could quickly sequence tumors (days instead of months). NGS allows simultaneous sequencing of millions of DNA fragments called massively parallel sequencing. These fragment sequences are then mapped (placed end-to-end) using the human reference genome (Behjati & Patrick, 2013). The technique is used to rapidly detect single DNA mutations or more extensive mutations including gene fusions (translocation, chromosomal inversion), and somatic copy number variations (SCNV) subsequently used to design a treatment strategy (Koboldt et al., 2013). NGS technology has been increasingly employed in genomic medicine (Koboldt et al., 2013).

Completion of the Human Genome Project (HGP), considered a breakthrough scientific achievement, mapped for the first time the vast majority of the human genome (complete set of DNA), allowing scientists to begin pinpointing the genes (also called oncogenes) involved the onset or progression of cancer (Harrow et al., 2012). Sequencing of these oncogenes by automated, big–dye methodologies was relatively slow compared NGS, which is orders of magnitude faster. HGP reported the first draft of the human genome in 2001 providing initial information about the human genomic landscape (Lander et al., 2001). The HGP determined the nucleic acid sequence order for almost all the base pairs in the human genomes. It also provided linkage maps indicating the location of known and proposed protein-coding genes (n = 20,000-25,000) (Lander et al., 2011). In 1990 the National Human Genome Research Institute (NHGRI) formed a collaboration with other global institutions to sequence the human genome by 2005, but

this project was completed earlier than expected in 2001. Approximately 99% of the genomic sequence was completed, the majority of the human genome is noncoding DNA, also referred to as junk DNA with unknown molecular function (Venter et al., 2001). It took about 11 years to get the first nearly complete human genome sequenced and made available to biomedical researchers.

The HGP sequence is used as a baseline by researchers to align sequences derived by NGS sequencing of tumors. Understanding the particular cancer genome sequence has helped oncologists to develop personalized treatments for many patients. For example, if a patient is predicted, based on DNA sequence, to be resistant to a particular drug (chemotherapy) then a different compound could be administered. Many research studies have theorized that somatic mutations that frequently occur in cancers are involved in cancer progression (Merid et al., 2014). Cancer is known to be a complex molecular disease caused by DNA mutation or epigenetic factors including environmental stimuli (Merid et al., 2014). Cancer is a heterogeneous disease with discoverable genetic variations among tumors. Specific genetic variants have been identified and linked with specific types of cancer (Kandoth et al., 2013).

As we better understand the molecular genetic pathways involved in human diseases like cancer, scientists have identified an increasing number of genomic alterations (DNA mutations) that have potential roles in cancers (Rykunov et al., 2016). Accurate and rapid identification of DNA mutations is the reason NGS has become the platform of choice. Throughput capabilities and the quality of data are two important variables to consider when deciding on technology to use for diagnostic molecular testing. Rating the instruments performance may be based on three key criteria: sequencing throughput (read length of the DNA sequencing), the accuracy of data

(simultaneous sequencing of a single rather than multiple templates), and sequencing depth (DNA templates optimization or coverage sequencing).

NGS technology has shown promise for classifying diseases based on genomic alterations, but ethical issues must be considered to maintain patient confidentiality. It is common for new emerging biotechnologies to raise ethical issues during its implementation. Currently, there is an enormous need for greater focus on how tumors acquire mutations that contribute to disease or drug resistance. Therefore there is a need to analyze genetic changes in tumors isolated from individual patients. Patient consent before using a biospecimen in clinical research is mandatory (Beskow & Dean, 2008). Following consent, a patient's confidentiality could be breached if sequence data was used for purposes other than testing (Beskow & Dean, 2008). Genomic testing should never be shared without a patient's consent. Because NGS technology is so powerful in that patient identifying sequence data are generated so quickly, additional care should be taken to protect these data.

Genomic information derived from NGS may be useful in clinical trials accelerating the lengthy drug development process as well as enhancing its effectiveness by providing genomic data for the drug approval process (Amur et al. 2015). NGS can be a useful tool in drug discovery pipeline from preclinical studies to the clinical research. The Food and Drug Administration (FDA) supports through its Critical Path Initiative (CPI) strategies for driving innovation in genomic biomarkers (Hamburg & Collins, 2010). Additionally, identification of genomic information before treatment has the potential to spare patients from receiving therapies that will be ineffective against their tumors. For example, Oncotype Dx is a relatively new genomic test currently used to predict the chance of breast cancer spreading to another part of the body (metastasis). This genomic test is used to screen early or late stage tumor samples for 21

oncogenes to target therapy treatment (Sparano et al., 2015). Targeted therapy (targeting specific genes or receptors involved in cancer progression) approaches may provide a specific treatment for a particular patient by using the genomic information from his or her tumor contrasted with that of their normal healthy cells (Ha & Shah, 2013). Genomic testing (DNA sequencing) provides a plethora of information strives to understand the genetic involvements in diseases and identifies new approaches for efficient cancer treatments or predictive preventive care (Zhang et al., 2011).

Development and use of reliable biological biomarkers for cancer progression is one of the strategies that FDA is encouraging pharmaceutical companies to employ to accelerate the drug development process. Biomarkers refer to biological markers that can be used as surrogates to measure disease formation and progression. Using genomic or molecular characterization to steer patients to the right drug at the right doses at the right time is the mission of personalized medicine. The success of personalized medicine depends on having accurate molecular diagnostic tests that identify patients who can benefit from targeted therapies (Hamburg & Collins, 2010).

There are two complimentary components of personalized medicine the first is pharmacogenetics, which is the study of inherited gene variants and how the drug response varies in individuals including the function of biological transporters encoded by genes (Jain, 2009). The second is Pharmacogenomics usually refers to the analysis of an array of genes that may determine a drug's behavior as well as be considered as a systematic assessment of gene expression (Jain, 2009). The science of pharmacogenomics studies how the human genome influences drug response within a person or subpopulation. People display differential responses to a drug due to differences in genetic variations that affect gene expression or function of

proteins (often enzymes) targeted by a drug (Goldstein et al., 2003). Personalized medicine allows us to determine the right medical treatments for each patient by understanding the molecular genetic pathways of diseases. Personalized medicine is still an emerging field that is facing many challenges particularly in drug regulatory policy, health insurance reimbursement, and integration into clinical practice (Devon et al., 2015).

In the conventional medical approach to development, a drug is administered to people and then health care professionals monitor its safety and efficacy. Some may have a significant response to a drug, and some may have a minor response. There are many factors affect how people respond to the drug. Notably, some people may absorb, metabolize, and eliminate drugs differently than others due to genetic mutations or variations, liver enzyme function, aging, accumulation of the drug, lifestyle, environmental factors influence, etc. One objective of personalized medicine is to decrease adverse effects while maximizing an individual's therapeutic response to a drug by using NGS. Classifying diseases based on genetic mutations or molecular genetic screening may help in defining more effective treatment approaches. Drugs usually interact with the cellular signaling pathways. Specifically, if an altered gene encodes a modified protein, it may affect drug absorption in the body, or may even cause the formation of a toxic chemical, which could have an impact on human safety (Goldstein et al., 2003).

In this study, the focus was on two next-generation operations that prepare DNA templates (sequencing the desired region of DNA segment) for sequencing primarily through amplification. Two different operations or instruments used for DNA template sample preparation during library creation are the OneTouch-2 (OT2) and the Ion Chef (IC). Both instruments employ the same strategy for sample template development. The workflow of IC is to prepare an amplification of DNA template, perform emulsion polymerase chain reaction

(emPCR), purify samples and load the sample into the biochip (sequencing biochip). In this procedure, magnetic beads attach to the target DNA template only. Therefore, the IC prepares the template for sequencing and loads it on the sequencing biochip. The latter step is not performed by the OT-2 operation.

In this study, the emphasis was on a significant problem in DNA template preparation, polyclonal amplification, that occurs during library creation generated from cancer cell DNA by comparing the work performance of two instruments OT2 and IC. The study was designed to compare sequencing performance of the DNA templates and the consequence of high-percentage polyclonal reads when sequencing 131 oncogenes using the fully-automated IC versus the semi-automated OT2. The hypothesis was that the IC operation would have less polyclonal formation than the OT2 instrument. Preliminary observations support this hypothesis. I also hypothesized a negative association between high polyclonal percentages outcome and overall average total DNA sequencing throughput.

Polyclonal amplification (DNA template duplications) occurs during DNA library preparation and can have a substantial impact on the sequencing outcome (Perrott, 2011). Polyclonal formation interferes with sequencing results by increasing sequencing duplications or simultaneous signaling. Sequencing noise occurs when polymerase chain reaction (PCR) duplications affect DNA sequencing reads and causes unreadable sequencing information. Polyclonal formation before DNA sequencing can negatively impact DNA sequencing shown to results (Schweyen, Rozenberg, Leese, 2014). It is common for a laboratory to repeat analysis of a sample due to many polyclonal reads stochastically reducing sequence coverage. The goal of this study was to make an assessment of the extent of polyclonal formation and determine which

methods of template creation would maximize the coverage of the genome or sequencing output of the PGM.

In this study, PCR used for amplification of a DNA template. The amplification step is necessary to have enough material to sequence. PCR is a lab molecular technique used to amplify a single target DNA template. Another type of PCR that is commonly employed in NGS is emulsion PCR (emPCR), and it refers to clonal amplification in droplets of oil-aqueous solution (reaction bubble) that contains the target DNA template attached to beads. These magnetic beads are designed to attach to the target DNA template that will eventually be sequenced. NGS sequencing requires four major workflow steps: nucleic acid extraction; library construction (DNA template preparation with IC or OT-2); sequencing; and analysis (bioinformatics). DNA extracted from tumor biospecimens is obtained from patients and used for DNA template library preparation before sequencing. DNA library preparation or "library creation" is an essential lab technique that attaches tagged DNA segments to the sequence being interrogated. Successful sequencing requires a sufficient amount of DNA (fewer PCR duplication reads) with sequence adaptors ("barcodes") added to particular ends of the DNA segment. These barcodes act as a red flag to distinguish which DNA region needs to be sequenced. They are also a useful tool in data analysis for alignment purposes to identify the DNA sequence fragments since NGS produced millions DNA segments reads that need to be aligned with the reference human genome.

In this study, six different metrics of performance were obtained from the PGM run summary sequencing reports and used to evaluate sequencing quality: polyclonal reads, total sequencing reads, empty microcell well reads, no-template reads, usable reads and library number reads. Polyclonal reads refer to similar DNA template fragments attaching to the same

bead during library creation causing redundancies in the sequencing results data. Total sequencing reads or total reads relates to the overall number of successful DNA sequencing outcomes. Empty microcells or "empty well reads" refers to an empty well (no bead with DNA) on the biochip-318 arrays. No-template reads refer to microcell wells that had beads but didn't have DNA template attached. Usable reads refer to the number of successful DNA templates sequenced. Library number reads relates to an overall number of library creations during sample DNA templates preparation. A total of 114 pre-run sequencing events from PGM servers were selected to provide study data. Simple random sampling method was used to selected data in this study. The selected samples had assigned a Paradigm Cancer Diagnostics (PCDx) number for sample identification.

The Thermo-Fisher OT2 instrument was first introduced in 2011 for sample template preparation for sequencing exclusively with the Ion Torrent Platform. The OT2 uses an amplification protocol similar to the IC, but it does not have the enrichment or chip loading capabilities built into the instrument. The OT2 system contains two modules: the One Touch ES enrichment and OT2 instrument. OT2 requires manual pipetting after emPCR for template sample preparation. When samples processed through emPCR with OT2 instrument protocol, they need an enrichment lab process for selecting the beads that contained the DNA templates. Following enrichment, the reaction is manually pipetted into the biochip. Each biochip has 5.5 million microcells available to accept beads (with DNA template). Therefore, the OT2 is a semi-automated instrument that does clonal amplification but does not have the capacity to load the biochip. Enrichment refers to a wash step for the beads also called Ion Sphere Particles (ISPs) that removes leftover primers and deoxynucleotide triphosphate (dNTPs) from the PCR process. Figure 1 map out the reaction steps in the process of DNA templates preparation before DNA

sequencing. In contrast, when processing the sample with the IC, the sample enrichment step is automatically performed during the processing. Sample DNA template development in the semi-automated OT2 requires more manual preparation time before loading the sample into the instrument and has no capacity to load biochips.

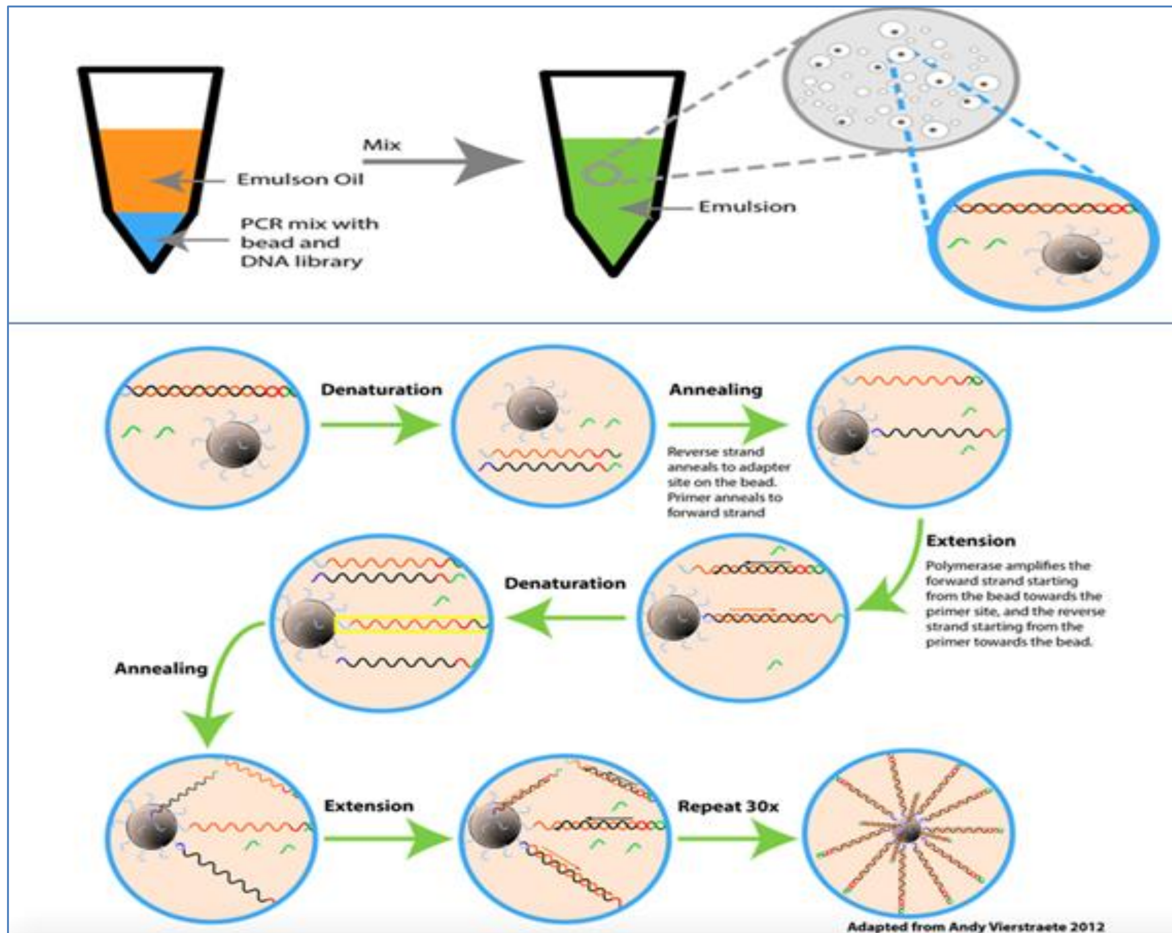


Figure 1. Emulsion PCR for DNA library preparation. There are three major steps in emPCR: denaturation; annealing; and extension. During DNA template preparation emulsion oil is mixed with the DNA library (aqueous solution) and PCR reagent (plus beads) to achieve a clonal amplification of single DNA molecules that contains these elements (DNA library, PCR reagents, beads). The first step in the reaction is denaturation (separation of DNA double strands) from each other at a specific temperature. After denaturation, the temperature is reduced to allowed primers (short synthetic strands of DNA) to attach to the long-strand DNA or beads (annealing), reverse strand anneals to adaptor site on the beads where primer anneals to forward strand. Extension refers to making a copy of the DNA; dNTPs are added to primers by DNA polymerase to make a complementary strand of DNA. Note. From “Emulsion PCR” by A. Vierstraete, 2012 (http://binf.snipcademy.com/lessons/ngs-techniques/emulsion-pcr). Copyright 2016 by binf-snipcademy. Reprinted with permission.

The semiconductor sequencer PGM provides DNA sequencing fast and in a cost-efficient manner. During DNA synthesis hydrogen ions are released when a new nucleotides {adenine (A), guanine (G), cytosine(C), thymine (T)} added to the target DNA strand. Semiconductor sequences detect the release of the hydrogen ion, and then software translates that into a DNA

sequence. During the sequencing process, the PGM reads the chemical changes in the biochip. Figure 2 showing the process of DNA sequencing by synthesized one strand of the DNA template. There is a sensitive layer located at the bottom of the microcell of the biochip, which converts pH (a measure of hydrogen ion concentration) change into voltage. This voltage at specific times determines which base pairs added during primer extension (copying of the DNA).

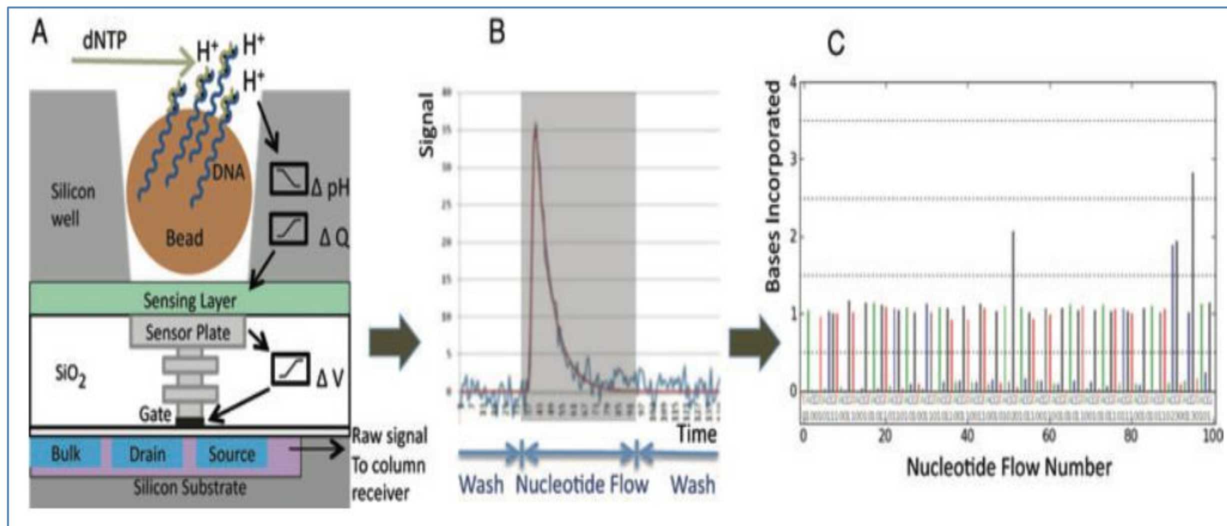


Figure 2. Illustration of the biochip sensor in semiconductor sequencing (318 ion chip). Schematic cross-section of a single microcell well of the biochip where DNA copying reactions occur (A). When dNTP (A, C, G, and T) flows into the chip, a H+ released if there is base pairing (B) and the base is incorporated into the growing DNA chain. The pH of the solution will change (red signal) in this synthesis reaction when incorporation of a nucleotide occurs. An algorithm is used to convert (C) from full series of integration signals from the microcell well into sequencing reads. Green color (A), blue color(C), red color (T), gray color (G). *Figure 2*. Semiconductor pH sensor device. Adapted from “Progress in ion torrent semiconductor chip based sequencing” by B. Merriman and J. Rothberg, 2012, *Electrophoresis, 33,* p. 3401. Copyright 2016 by John Wiley and Sons (Rights Link). Reprinted with permission.

The PGM system works by synthesizing a complementary DNA strand in the presence of one of four nucleotides at a time (wash steps occur between the addition of each new nucleotide). Figure 3 below show the PGM used in NGS. The PGM takes advantage of measuring the release of H+ ions whenever there are two complementary base pairing bonds or matches during the covalent bond formation (A-T and G-C or vice versa T-A and C-G) a H+ ion is released. If a

nucleotide did not incorporate into the growing DNA strand, then there would not be any changes in pH and there wouldn't be a signal generated. Appendix D shows the overview of PGM steps for sequencing data generated in NGS as well as the biochip 318 that commonly used in molecular laboratory.

Figure 3. Ion Torrent Personal Genome Machine (PGM™). The PGM platform used for sequencing DNA. The PGM picks up electrical signals produced by the chemical changes in the biochip (pH) and this signal is converted into the sequence data. Note. From "Ion Personal Genome Machine® (PGM™) system" by Thermofisher Scientific (https://www.thermofisher.com/us/en/home/technical-resources/contact-us.html). Copyright 2016 by Thermo Fisher Scientific. Reprinted with permission.

The biochip advance technology commonly used in NGS. Figure 4 show the detection of H ion released during the sequencing process. Figure 4 illustrated a close look for a single microcell well in the biochip 318 used in the PGM.

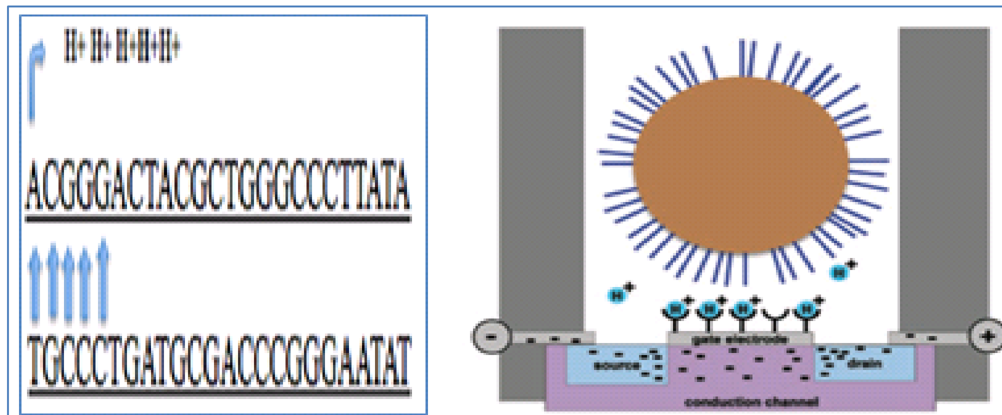


Figure 4. Measurement of H+ during sequencing reactions with a PGM.A microcell well of the biochip (right) is sequences DNA by monitoring the H+ release. Enzymatic addition of dNTPs or nucleotides to the DNA chain causes a H+ ion release. These chemical changes are transformed into an electrical signal indicating the addition of a specific nucleotide. Left image, *Figure 1b*. Semiconductor pH sensor technology. Adapted from “Progress in ion torrent semiconductor chip based sequencing” by B. Merriman and J. Rothberg, 2012, *Electrophoresis, 33,* p. 3400. Copyright 2016 by John Wiley and Sons (RightsLink). Reprinted with permission.

Research Design and Methodology

Template preparation for NGS required five major lab procedures: nucleic acid extraction (DNA and RNA); library creation (barcode added to DNA template), PCR emulsion amplification (emPCR); enrichment; and sequencing. DNA and RNA extracted from cancer samples tissue processed in formalin-fixed paraffin embedded (FFPE) blocks. Nucleic acid was purified using Qiagen kit reagents (QIAamp DNA FFPE tissue kit) per manufacturer's instructions accessed from (https://www.qiagen.com/us/resources/resourcedetail?id=63a84dc7-d904-418e-b71a-1521cf318e82&lang=en). Briefly, FFPE sections were incubated with an excess of xylene to dissolve and remove paraffin wax. The sample was spun in a centrifuge (20,000 x gravity (g) for 2 minutes). The supernatant removed and 1 ml of ethanol added to the sample. The sample was incubated at 37°C for 10 minutes before 180 µl of buffer reagent (ATL) was added to the sample with 20 µl of Proteinase K (mixed by vortexing). Samples were incubated again at 56°C for 3 hours before adding 2 µl of RNase. A combination of 200 µl of buffer (AL) and 200 µl ethanol (95%) was then added together to each sample. The lysate was transferred unto a QIAamp MinElute spin column for elution at (centrifuge at 6000 x g for 1 minute). The purified nucleic acid sample was quantified using a Qubit 2.0 Fluorometer to determine the DNA concentration. A Thermo Fisher Scientific protocol was followed to measure deoxyribonucleic acid concentration.

A QIAgility liquid handling instrument (Qiagen, Inc.) was used for automated mixing of library creation reagents. Briefly, 5 µl of DNA library primer stock was mixed with 5 µl of nucleic acid prepared above. Samples were transferred to a Bio-Rad Rotor-Gene for PCR amplification using the following conditions: 95°C (hold 5 minutes, one cycle), 95°C (15 seconds, 18 cycles), 60°C (30 seconds). Immediately after PCR amplification, samples were

placed in a frozen block (-20°C). Twenty µl of pooled sample for each set (1- 4 libraries) were put in the thermal cycler after adding 8 µl ExoSAP (used to remove the leftover of PCR reagents). The mixture was briefly spun in a micro-centrifuge. The following temperatures conditions were set up for thermal cycles over one hour: 37°C for 45 min, (1 cycle), 80°C for 15 min (1 cycle), and 4°C for 1 min. Amplification products were either transferred to the OT2 or IC instrument for completion of DNA template preparation. (Thermo Fisher Scientific Company User's Guide instruction protocol (https://tools.thermofisher.com/content/sfs/ manuals/ MAN0010902_PGM_HiQ_OT2_Kit_UG.pdf) was used for DNA templates development). The emPCR reactions were performed in the OT2 instrument before enrichment protocol using the Thermo Fisher Scientific protocol. Amplified DNA manually loaded onto 318 biochips. Figure 5 maps out the steps for sample preparation in NGS.

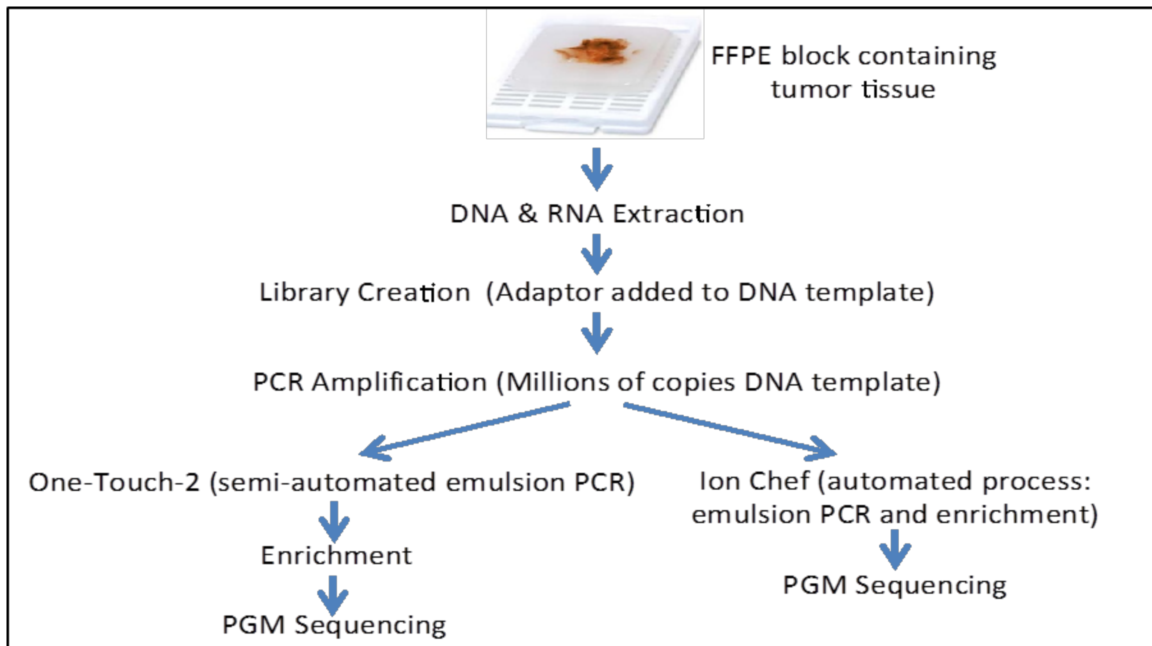


Figure 5. Showing the NGS workflow process for DNA sequencing of the cancer cells from DNA extraction of FFPE sections to sequencing with PGM.

Minitab 17 (statistical software, descriptive statistic) was used for performing statistical comparisons of means for the six metrics measured in this study. An independent t-test was used to compare the means. Institutional review board (IRB) approval was requested and appendix A is the IRB approvals letter for permission to do this study.

Results

Polyclonal percentage outcome was assessed for an association with total sequencing reads or the coverage of DNA sequencing. Data indicated that optimal NGS occurs in the absence of identical DNA templates, attached to the same bead while preparing the sample for sequencing. There were 114 DNA specimens sequenced in this study. An independent-samples t-test was used to determine if there were differences between IC and OT2 work performance means. Six-parameter variables were used to assess the DNA templates development performance (polyclonal reads, total sequencing reads, empty microcell well reads, no-template reads, useable number reads, and library number reads) for the two DNA library methods of DNA template preparation, IC, and OT2 (Appendix B). Four of six metrics measured identified IC operation as the preferred instrument for DNA template preparation for NGS with the PGM platform. Appendix A provides raw data for the metric performance variables used for evaluation of the study data.

The polyclonal mean for IC was 31.82 (95% confidence intervals = 28.85, 34.80). The polyclonal mean outcome for OT2's samples group was 43.09, 95% confidence intervals =39.70, 46.48 refer to Figure 6. The polyclonal percentage was significantly higher ($p = 0.032$) than the average IC polyclonal percentage outcome indicating that there are more PCR duplications (polyclonal formation) for OT2 data.

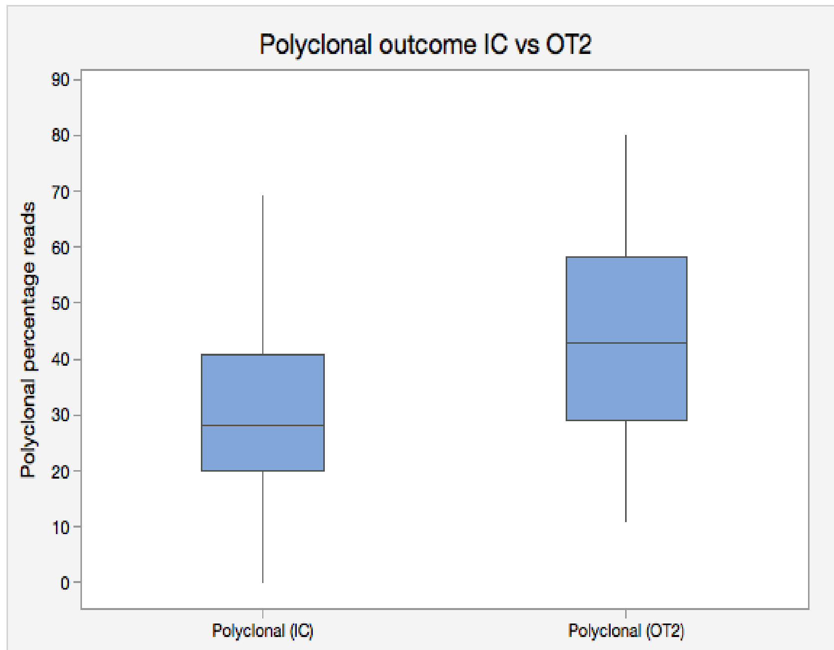


Figure 6. Showing the mean polyclonal formation when sample processed by OT2 was significantly (p = 0.032) higher than ICs mean. The horizontal line in the blue box indicates the median. The vertical line in the top of blue box refer to upper whisker (upper 25% of the points distribution) and the bottom vertical line from the blue box refer to lower whisker which represent the lower 25% of the points distribution.

A significant difference (p <0.05) was observed between mean total numbers of transcripts sequenced for the two treatments. Samples processed through the IC protocol had significantly (p <0.05) higher PGM throughput or more overall DNA sequencing reads than the samples prepared with the OT2 instrument. The total DNA sequenced reads mean for the IC preparation was 3770605 (95% confidence intervals = 3520426, 4020784 (Figure 7). The total DNA sequenced reads mean for the OT2 preparation was 2850276 (95% confidence intervals =2600097, 3100455).

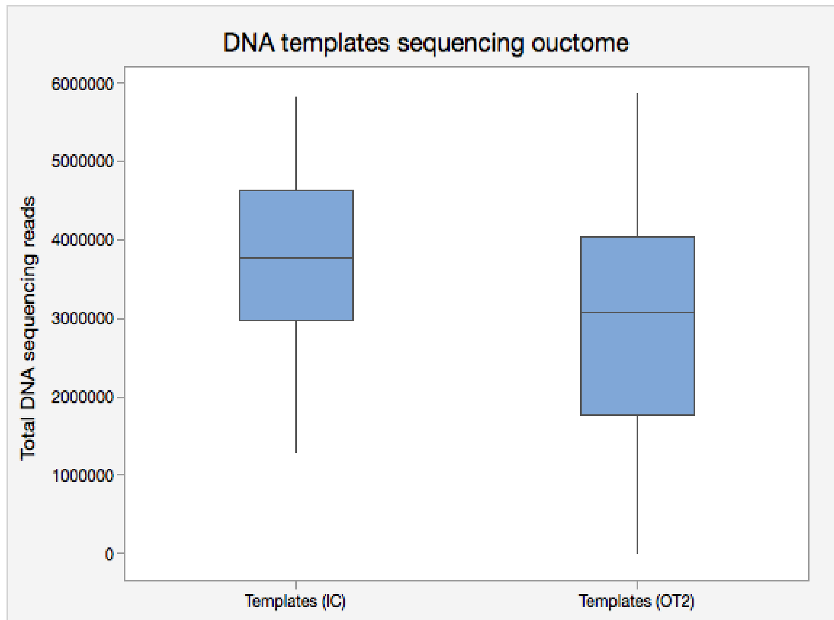


Figure 7. Showing the mean number of template sequenced by PGM. Mean total reads from samples processed for DNA template preparation in IC and OT2 operation. The horizontal line in the blue box indicates the median. The vertical line in the top of blue box refer to upper whisker (upper 25% of the points distribution) and the bottom vertical line from the blue box refer to lower whisker which represent the lower 25% of the points distribution.

Mean empty microcells well (surface of the biochip without a DNA-carrying bead) in IC-processed sample outcome was 33.89 (95% confidence intervals 30.021, 37.769) and the mean for the OT2's samples was 21.6 (95% confidence intervals 29.23, 38.55, Figure 8). The two means of the empty microcells (OT2 and IC) were significantly different (p <0.05).

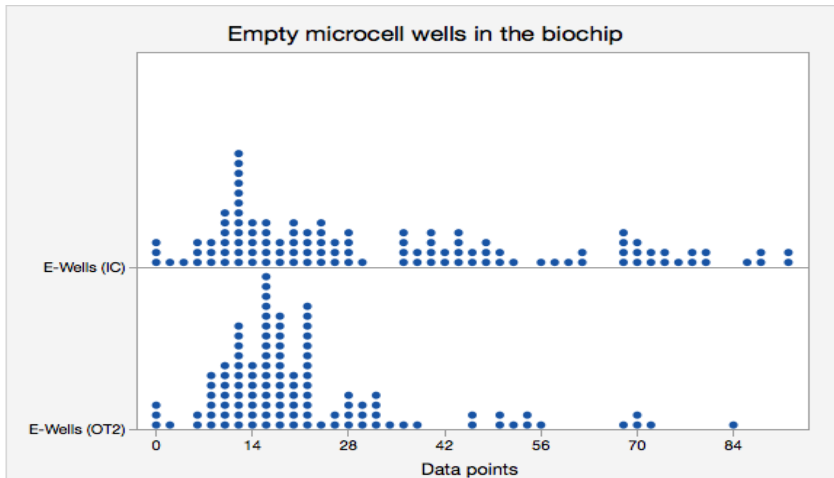


Figure 8. Showing the percentage of empty biochip surface. Empty microcell wells of the biochip that does not contain sample through automation loading (IC) verses manual loading (OT2). Empty microcell wells indicate areas of the biochip that did not got filled with a bead. Each dot in graph represents a data point generated from NGS.

Mean beads without DNA templates (no target DNA attached) was significantly ($p < 0.05$) higher in IC-processed specimens than the OT2 processed ones. The IC sample mean for empty beads or no-DNA template was 9.85 (95% confidence intervals = 6.48, 13.22) and the average for OT2 samples was 1.825, (95% confidence intervals = 0.725, 2.924) refer to figure 9 below.

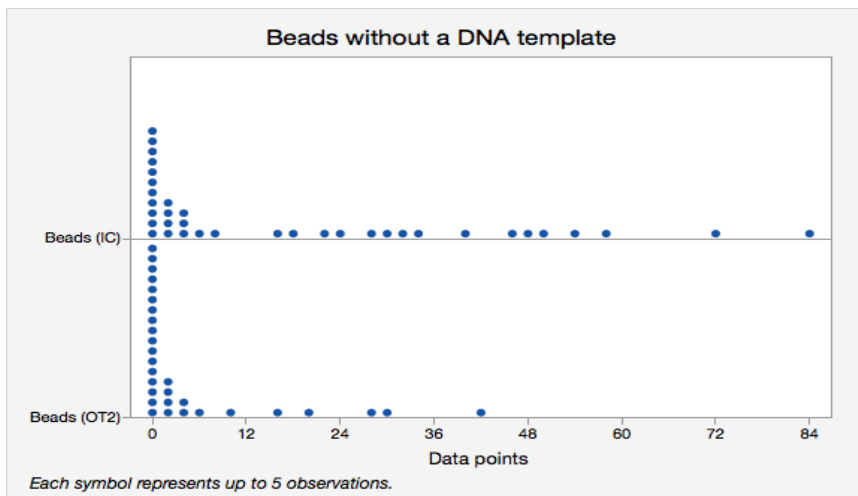


Figure 9. Showing the ISP (beads) without DNA templates processed through NGS. Beads without a DNA template discovered during NGS sequencing and dots represent data points.

The average for IC DNA library number was 72.71 (95% confidence intervals 69.61, 75.80) and the mean for OT2 prepped samples were 59.825 (95% confidence intervals 56.73, 62.91). A significant difference (p <0.05) was observed between IC's operation for DNA library creation and OT2's operation. Figure 10 shows the data distribution for library number detection during the DNA sequencing.

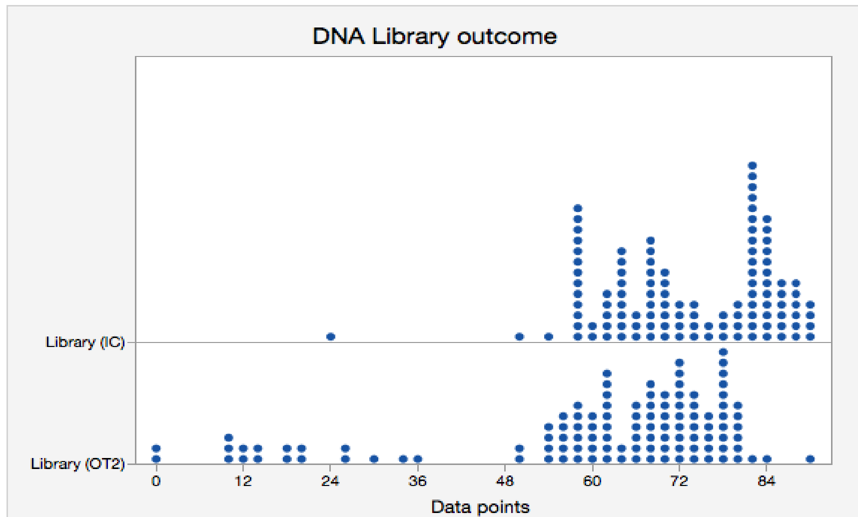


Figure 10. Showing the DNA Library number outcome from NGS. Library number percentage outcome for IC and OT2 for DNA template preparation samples. Dots represent the data point of the empty beads or no-DNA template.

The mean usable DNA library percentage outcome was 53.26 (95% confidence intervals 50.78, 55.73) for IC, 39.47 (95% confident intervals 36.97, 41.92) for OT2 (Figure 11). A significant difference (p <0.05) was observed between IC's operation for usable DNA library and OT2's operation. The usable DNA library percentage in IC operation was significantly higher (p <0.05) than the average OT2 percentage outcome. Appendix C showed an example of the summary reports generated by PGM.

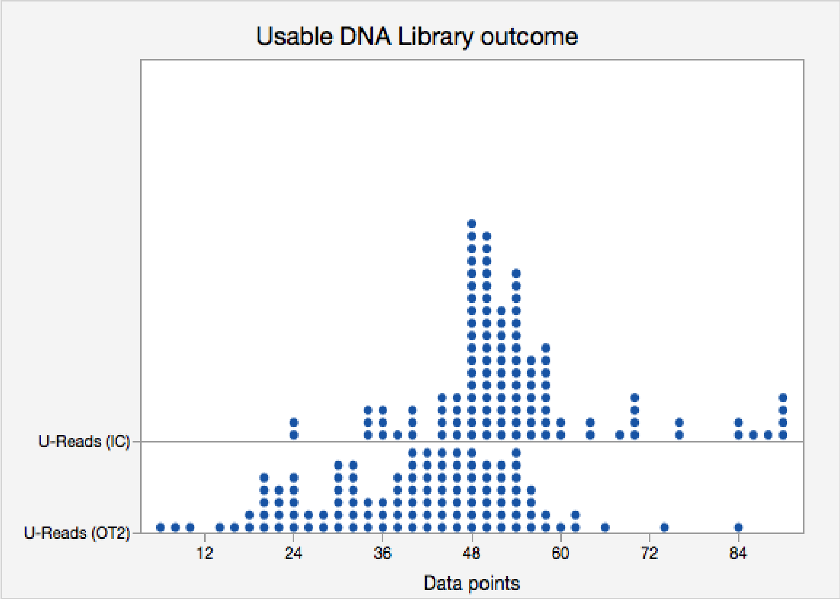


Figure 11. Showing the usable DNA library outcome form NGS. Mean usable library percentage for OT2 and IC prepared DNA libraries. Dots represent the data points for usable library number outcome.

Discussion

The objective of this study was to investigate whether there is efficient DNA library preparation and the consequence of high-percentage polyclonal reads when sequencing targetable genes (screening for 131 oncogenes) using two different instruments. In this study, it was observed that only two minor (not statistically significant) differences in metric performance (the no-templates and empty microcell well) for IC versus OT2 instrument performance. However, these differences probably did not contribute to the final number of total DNA sequence reads. Data presented in this study showed that high polyclonal percentage in NGS negatively affected the overall average sequencing result reads, and that this impacted the depth coverage reads of target DNA sequenced.

The study results supported the hypothesis that sample DNA template preparation performed by IC instrument was better (higher quality and quantity of DNA sequences) than sample DNA template preparations carried out by the OT2 protocol. High polyclonal formation by the OT2 significantly impacted the throughput of the PGM regarding DNA sequencing volume. As stated in the introduction, PCR duplicates can lead to false positive detection in NGS and affects the depth of sequencing outcome (coverage of alleles sequenced). Polyclonal formation occurs when two identical DNA fragments attach to the same beads during emPCR. Notably, the observation of the less polyclonal formation using the IC operation ensured a single DNA template per PCR emulsion mixture as the most probable. Usually, highly stable emPCR can yield a high percentage of beads, or optimal number of beads contained the DNA templates to be sequenced. Additional research will need to be conducted to find if there is a correlation between biochip scaling and algorithm sequencing for the Ion Torrent platform. It could be that a fraction of empty microcell wells contributed to better algorithm sequencing in PGM.

More research is needed to uncover other factors that may affect or influence overall coverage DNA sequencing results. It is possible that more beads aggregate (clump) when processed by the OT2 that generated more polyclonal formation. Also, the stability issue of the emPCR in OT2 operation could contribute to minimize the throughput of the PGM as well as the size of the micro-reaction generated through the emulsion PCR.

The findings of this study are immediately useful for the Paradigm Company, indicating that for better efficiency of downstream sequencing we should use the IC format for emPCR instead of the OT2 platform. Study data showed that higher polyclonal percentage impacted the total sequencing reads (Figure 6). When there was a lower polyclonal formation efficiency of total DNA sequencing reads was higher.

The study results also indicated the performance of emPCR in IC is more stable than emPCR processed in the OT2 instrument. Achievement of monoclonal amplification (beads attaching only DNA templates without duplications) observed during emPCR was at higher levels in the IC process than OT2. The stability of emPCR was more efficient in IC operation than OT2 operation that resulted in less polyclonal formation outcome. Also, the automation process could play a role by providing a highly stable emulsion that contributed to more efficient DNA sequencing. Another factor which could help forma stable emPCR that was beyond the scope of this study was on the size of the micro-reaction generated in the emPCR.

This study data results also showed that DNA sequencing performance for templates prepared by the IC instrument, were maximized to give a better quality of sequencing than the samples prepared in OT2 for tumor DNA template preparation. Unmapped data in NGS (useless information) that may provide poor quality data in a sequencing run usually resulted in higher

polyclonal reads. This useless information may cause consume computing time filtering them out or removing them prior to bioinformatics data analysis.

For no-template, results indicated that the OT2 performed better than the IC. However, this doesn't substantially affect the final number of DNA sequencing reads. Additional research is needed to find out what factors influence bead recovery without DNA template. Also, the results indicated that the loading rate of samples into biochip was greater in OT2 samples processed samples than for IC –prepared samples. It could be that the manual pipetting hand put more pressure in loading the sample into the biochip. Also, it could be a final centrifugation of the sample before loading the sample into the PGM contributed and to better loading than automation operation. However, the manual pipetting or final centrifugation, once again, would probably not directly affect the overall DNA total sequencing reads according to data that showed the sequencing results were better than the sequencing results from the OT2 instrument (Figure 8).

Research is also needed to explore the reasons that manually loading is better than automated loading of samples. There could be some challenges like centrifugation rates that prevent maximization of the loading rate of the sample during IC process into the chip. However, automated loading of the DNA sample template may perform better for completely loading the sample because of the potential of human errors accidently not loading the whole complete sample into biochip were reduced. Also, the metric of empty well in IC process did not show any significant impact on the throughput of the PGM.

Despite the sample loading rate performance, where samples prepared in OT2 was higher than IC's sample loading performance, the total DNA sequencing reads outcome was better with the IC preparation than the OT2's preparation. There could be a barrier to bead recovery when

transferring the samples from the OT2 to the enrichment instrument. This study indicates that the polyclonal formation influences total DNA sequencing reads in NGS.

The optimization of DNA sequencing by NGS is critical for reducing the DNA sequencing duplications (polyclonal) to increase the throughput of DNA sequencing outcome. Oncogenes that cause disease play a role in tumor cell progression and identification of these oncogenes will aid in genomic drug development to cure cancer (Pao et al., 2004). Efficient oncogene detection will also assist in the assessment of anticancer efficacy as well as reducing the time of clinical trial study or drug development. Oncogenes are involved in mutational processes, which lead to the metastatic stage of cancer (Rykunov et al., 2016). Some of these DNA mutations also allow cancer cells to escape being recognized or attacked by our immune cells through producing a particular protein to evade recognition (Rykunov et al., 2016). Identifying these oncogenic genes would aid us in fighting cancer through finding an effective treatment for patients. More extensive research studies need to conduct to link genomic medicine to the health care system.

Despite the increased number of advances in drug development, the lack of drug response or efficacy for a significant proportion of people remains a major issue for drug effectiveness, particularly for oncology. For example, since 1998 only 10 drugs have been approved to treat lung cancer, whereas 167 other lung cancer drugs failed in clinical trial according to a recent report by PhRMA Public Affairs organization (2014). In 2011, FDA approved Crizotinib for patients with non-small lung cancer (NSCLCs) based on genomic testing of patients who contains the ALK (anaplastic lymphoma kinase) gene mutations in NSCLCs. Crizotinib was designed to target EML4-ALK by inhibitor ALK-positive oncogenic fusion gene to block tumor progression (Solomon et al., 2014). Progression-free survival was significantly longer in people

treated with Crizotinib than individuals in a chemotherapy control group; the objective response rate was 74% compared with conventional chemotherapy, 45% (Solomon et al., 2014). Crizotinib was associated with a greater reduction in lung cancer symptoms and more significant improvement in the quality of life (Solomon et al., 2014). An efficient sequencing for these oncogenes could help to improve the quality of patient life or maybe cure them from cancer.

Drug development is a costly and lengthy process, which takes about 12 years of the clinical research study for a single drug marketing approval. An investigational drug must first pass the pre-clinical research study (Phase 0) as FDA requires that all investigational drugs must be tested on animals first before moving it into a human study (Phase 1). In essence, the genomic information or DNA sequencing is a useful tool to optimize drug development and accelerate the research process to get to the clinical endpoint of the investigational drug. A pharmaceutical company could use genomic information to predict cancer progression and provide more efficient treatment options to cancer patients. As stated earlier, NGS provides broad-spectrum analysis of DNA identifying chromosomal inversions, fusion genes, DNA mutations, DNA copy number variations, mRNA expression, and translocations of genes. Furthermore, genomic information could enhance health care quality and increase the efficiency of a patient's treatment. Genomic medicine will not only match a medication that targets specific mutations in a patient but also can identify people who had the mutations for ineffective treatment response and kept them from taking the drug. For example, a KRAS mutation in colon cancer gives health care providers the potential option to determine the right medication for a patient and discontinue prescribing the same drug (Erbitux) found to be ineffective in patients with a mutated KRAS gene (Allegra et al., 2009). In other words, precision medicine will give the physician the option to match the DNA mutations that classify as driver cancer mutation to give a patient a particular

target treatment drug. Also, it can identify people with specific mutations and proteomic alterations (abnormal protein sequence or processing) that cause or affect them to have a different drug response.

In the drug development industry it is very common to have drugs fail in a clinical research study either due to the safety or lack of efficacy (Hamburg and Collins, 2010). FDA regulations require a drug to be safe and effectiveness before approval. If the drug industry routinely used genomic information in a clinical trials research study, this could impact the quality of medicine as well as reduce the time of investigation and optimize the designed protocol of test screening by enrolling only patients with specific biomarkers rather than symptoms only. A small number of variations in genetic make-up influence how people response to a drug (e.g. DNA copy number variations, proteomic alterations, and gene fusions). An individualized medicine approach may play a crucial role in optimizing drug dosing and reduce the adverse effect of the medicine (Gorre et al., 2001). Advancing genomic medicine may also have a significant effect on the economics of drugs production and distribution.

Personalized medicine not only helps in maximizing drug response but also enables physicians to avoid unnecessary and harmful treatments. Instead of developing a treatment then giving it to many people to see if it works, genomic screening information or molecular stratification could be used in advance to determine which patients will be more likely to benefit from the drug. If genomic sequence data is used to select a targeted therapy approach it may result in a better treatment for patients at the right time with the right dose of a drug. Studies like the one described in this thesis will make sequencing for personalized medicine more efficient and less expensive.

Conclusions

This study validated that polyclonal formation plays an important role by reducing the DNA sequencing reads or the performance of the PGM for cancer cells DNA sequencing. The higher rate of polyclonal formation for the OT2 method contributed to minimizing the performance of the PGM as measured by successful sequencing reads. The IC operation was efficient in DNA template development, which overall indicated a stable emPCR and good bead recovery that contributed to maximizing the DNA sequencing results. In order to have sufficient DNA sequencing using PGM platform there should be less polyclonal formation during DNA library creation. This study will contribute to personalized medicine by helping to detect oncogenes more efficiently.

This study data established that the IC is preferred when preparing DNA templates for NGS due to less production of the polyclonal formations, and an efficient DNA library production contribute to effective final DNA sequencing reads. In addition, the IC consistently produced fewer polyclonal templates during emPCR, even when there were two metrics (empty microcell well and no-DNA template) that showed better performance by the OT2 instrument. Finally, the IC operation method is the preferred platform for DNA template preparation for NGS.

References

Allegra, C., Jessup, M., Somerfield, M.R., Hamiltion, S.R., Hammond, E.H., Hayes, A.,… Schilsky, R.L. (2009). American Society of Clinical Oncology provisional clinical opinion: Testing for KRAS gene mutation in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J Clinical Oncology*, *27*, 2091-2096.

Application Note Amplicon Sequencing. (n.d.) Retrieved October 22, 2015 from https://www3.appliedbiosystems.com/cms/groups/applied_markets_marketing/documents/generaldocuments/cms_094273.pdf

Amur, S., Vange, L.L., Zineh, I., Garner, B.S., Woodcock, J. (2015). Biomarker qualification: Toward a multiple stakeholder framework for biomarker development, regulatory acceptance, and utilization. *Wiley Online Library*, *98*, 34-46.

Behjati, S., Patrick, S.T. (2013). What is next generation sequencing? *Research in Practice*, *98*, 236-238.

Beskow, M.L., Dean, E. (2008). Informed consent for biorepositories: Assessing prospective participants' understanding and opinions. *American Association for Cancer Research*, *17 (6),* 1441-1451.

Bras, J., Guerreiro, R., Hardy, J. (2012). Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease: *Nat Rev Neuroscience*, *13*, 453–464.

Cancer Genome Atlas N. (2012). Comprehensive molecular portraits of human breast tumors. *Nature*, *490*, 61–70.

Clark, M., Chen, R., Lam, H., Karczewski, K.J., Chen, R., Euskirchen, G., … Snyder, M. (2011). Performance comparison of exome DNA sequencing technologies. *Biotechnology*, *29*, 908–914.

Crizotinib Approval Expanded for Lung Cancer. (n.d.). Retrieved May 22, 2016, from http://www.cancer.gov/news-events/cancer-currents-blog/2016/fda-crizotinib-ros1

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., Depristo, MA., … Sherry, ST. (2011). The variant call format and VC F tools. *Bioinformatics, 27*, 2156–2158.

Dausset, J., Cann, H., Cohen, D., Lathrop, M., Lalouel, J.M., White, R. (1990). Centre d'etude du polymorphisme human (CEPH): Collaborative genetic mapping of the human genome. *Genomics, 6*, 575–577.

Devon, K., Lerner, J., Ganai, S., Angelos, E. (2015). Ethic and genomic medicine, how to navigate decisions in surgical oncology. *Journal of Surgical Oncology*, *111*, 18-23.

Emulsion PCR figure from binf snipcademy. (Andy Vierstraete) Retrieved October 12,2015 from http://binf.snipcademy.com/lessons/ngs-techniques/emulsion-pcr

Gorre, M.E., Mohammed, M., Ellowood, K., Hsu, N., Paquette, R., Rao, P.N., Sawyers, C. L. (2001). Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. Science, *293*, 876-880.

Goldstein, D.B., Tate, S.K., Sisodiya, S.M. (2003). Pharmacogenetics goes genomic. *Nature Reviews Genetics*, *4*, 937-947.

Ha, G., Shah, S. (2013). Distinguishing somatic and gremlin copy number event in cancer patient DNA hybridized whole genome SNP genotyping arrays. *Methods in Molecular Biology*, *937*, 355-372.

Hamburg, M.A., Collins, F. (2010). The path to personalized medicine. *The New England Journal of Medicine*, *4*, 301-303.

Harrow, J., Frankish, A., Gonzales, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B. L., … Hubbard, T.J. (2012). Gencode: The reference human genome annotation for the encode project. *Genome Research*, *22*, 1760-1777.

Harismendy, O., N.G, P. C., Strausberg, L. R., Wang, X., Stockwell, B. T., Beeson, C., … Frazer, A.F. (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology.10* (3). doi : 10.1186/gb-2009-10-r32.

Jain, K. (2009). *Textbook of Personalized Medicine*. New York, NY: Springer Science and Business Media, LLC.

Kandoth, C., Mclellan, D.M., Vandin, F., Ye, K., Niu, B., Lu, C., … Ding, L. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*, *502*, 333-339.

Koboldt, D.C., Steinberg, K.M., Larson, E.D., Wilson, K.R., Mardis, R.E. (2013). The Next-generation sequencing revolution and its impact on genomics. *Cell Press*, *155*, 27-38.

Lander, E.S., Lauren, M., Birren, B., Nusbaum, C., Zoby, M.C., Baldwin, J., … Morgan., M. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*, 860-921.

Lung Cancer: A New Era Just Beginning. (n.d.) Retrieved September, 8, 2016 from http://www.phrma.org/sites/default/files/pdf/2014-cancer-setbacks-report.pdf

Meldrum, C., Doyle, M., Tothill, R. (2011). Next-generation sequencing for cancer diagnostics: A practical perspective. *Clinical Biochemistry*, *32*, 177-195.

Merid, K., Goranskaya, D., Alexeyenko, A. (2014). Distinguishing between driver and passenger mutations in individual cancer genomes by network enrichment analysis. *BMC Bioinformatics,15*, 308.

Merriman, B., Rothberg, J., Ion Torrent R&D Team. (2012). Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, *33*, 3397-3417

Myers, J.L., Well, A.D., Lorch, R.F. Jr. (2010). *Research design and statistical analysis* (3rd ed.). New York, NY: Routledge.

Pao, W., Miller, V., Zakowski, M., Doherty, J., Politi, K., Sarkaria, I., … Singh, B. (2004). EGF receptor gene mutations are common in lung cancer from " never smokers" and are associated with sensitivity of tumor to gefitinib and erlotinib. *Proc. Natl.Acad.Sci USA, 36*, 13306-13311.

Perrott, J. (2011). Optimization and improvement of emulsion PCR for the ion torrent next-generation sequencing platform (Masters Dissertation). Retrieved Jun, 20,2015 https://beng.soe.ucsc.edu/sites/default/files/projectreports/Final%20Thesis.pdf

Rykunov, D., Beckmann, N.D., Li, H., Uzilov, A., Schadt, E.E., Reva, B. (2016). A new molecular signature method for prediction of driver cancer pathways form transcriptional data. *Nucleic Acids Research*, *44,*1-11.

Schweyen, H., Rozenberg, A., Leese, F. (2014). Detection and removal of PCR duplicates in population genomic ddRAD studies by addition of a degenerate base region (DBR) in sequencing adaptors. *Biol.Bull*, *227*,146-160.

Sheskin, D.J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures* (5th Ed.). Boca Raton, FL: Chapman & Hall/CRC Press.

Solomon, J.B., Mok, T., Kim, D.W., Yi-long, W., Kazuhiko, N., Mekhail, T., … Blackhall, F. (2014). First-line crizotinib versus chemotherapy in ALK-positive lung cancer *New England Journal of Medicine*, *371*, 2167-2177.

Sparano, J.A., Gray, D.F., Makower, K.I., Pritchard, K.S., Albain, D.F., Hayes, C.E.,. … Sledge, G.W. (2015). Prospective validation of a 21-gene expression assay in breast cancer. *The New England Journal of Medicine*, *373*, 2005-2014.

Thermo Fisher Scientific Ion Chef ™ Instrument. (n.d.) Retrieved Jun 18,2016 from https://www.thermofisher.com/order/catalog/product/4484177?ICID=search-product

Thermo Fisher Scientific Ion 318 ™ Chip kit v2BC (n.d.) Retrieved Jun 25,2016 from https://www.thermofisher.com/order/catalog/product/4488150?ICID=search-product

Thermo Fisher Scientific Ion OneTouch ™ 2 system. (n.d.) Retrieved September 20, 2015 from https://www.thermofisher.com/order/catalog/product/4474779?ICID=search-product

Thermo Fisher Scientific Ion Personal Genome Machine ® PGM™ (n.d.) Retrieved Jun 18,2016 from https://www.thermofisher.com/order/catalog/product/4462921?ICID=search-product

Venter, J.C., Adams, D.M., Myers, W.E., Li, W.P., Mural, J.R., Sutton, G., Smith, O.S., Zhu, X. (2001). The sequence of the human genome. *Science*, *291*,1304-1351.

Walker, L., Morrissey, C. (2013). Bioethics methods in ethical, legal, and socials implications of the human genome project literature. *Bioethics*, *28*, 481-490.

Wilson, J., Nicholls, S. (2015). The Human genome project, and recent advances in personalized genomics. *Dove Press Journal: Risk Management and Healthcare Policy*, *8*, 9-20.

Wolpe, R. (2009). Personalized medicine and its ethical challenges. *World Medical & Health Policy*, *1*, 1944-2858.

Zhang, J., Chiodini, R., Badr, A., Zhang, G. (2011). The impact of next-generation sequencing on Genomic. *Science Direct, 38* (3), 95-109.

Appendix A: IRB approval letter for conducting the research study.

UHSRC Determination: **EXEMPT**

DATE: **December 20, 2015**

TO: **Akram Aldilaimi, M.S**
Eastern Michigan University

Re: **UHSRC: # 845183-1**
Category: Exempt category *4*
Approval Date: December 20, 2015

Title: **Performance Comparison of Targeted Next-Generation Sequencing Platforms and Implications of Polyclonal Formation on Genomic Cancers Sequencing Data Outcome**

Your research project, entitled **Performance Comparison of Targeted Next-Generation Sequencing Platforms and Implications of Polyclonal Formation on Genomic Cancers Sequencing Data Outcome,** has been determined **Exempt** in accordance with federal regulation 45 CFR 46.102. UHSRC policy states that you, as the Principal Investigator, are responsible for protecting the rights and welfare of your research subjects and conducting your research as described in your protocol.

Renewals: Exempt protocols do not need to be renewed. When the project is completed, please submit the **Human Subjects Study Completion Form** (access through IRBNet on the UHSRC website).

Modifications: You may make minor changes (e.g., study staff changes, sample size changes, contact information changes, etc.) without submitting for review. However, if you plan to make changes that alter study design or any study instruments, you must submit a **Human Subjects Approval Request Form** and obtain approval prior to implementation. The form is available through IRBNet on the UHSRC website.

Problems: All major deviations from the reviewed protocol, unanticipated problems, adverse events, subject complaints, or other problems that may increase the risk to human subjects **or** change the category of review must be reported to the UHSRC via an **Event Report** form, available through IRBNet on the UHSRC website

Follow-up: If your Exempt project is not completed and closed after **three years**, the UHSRC office will contact you regarding the status of the project.

Please use the UHSRC number listed above on any forms submitted that relate to this project, or on any correspondence with the UHSRC office.

Good luck in your research. If we can be of further assistance, please contact us at 734-487-3090 or via e-mail at human.subjects@emich.edu. Thank you for your cooperation.

Sincerely,

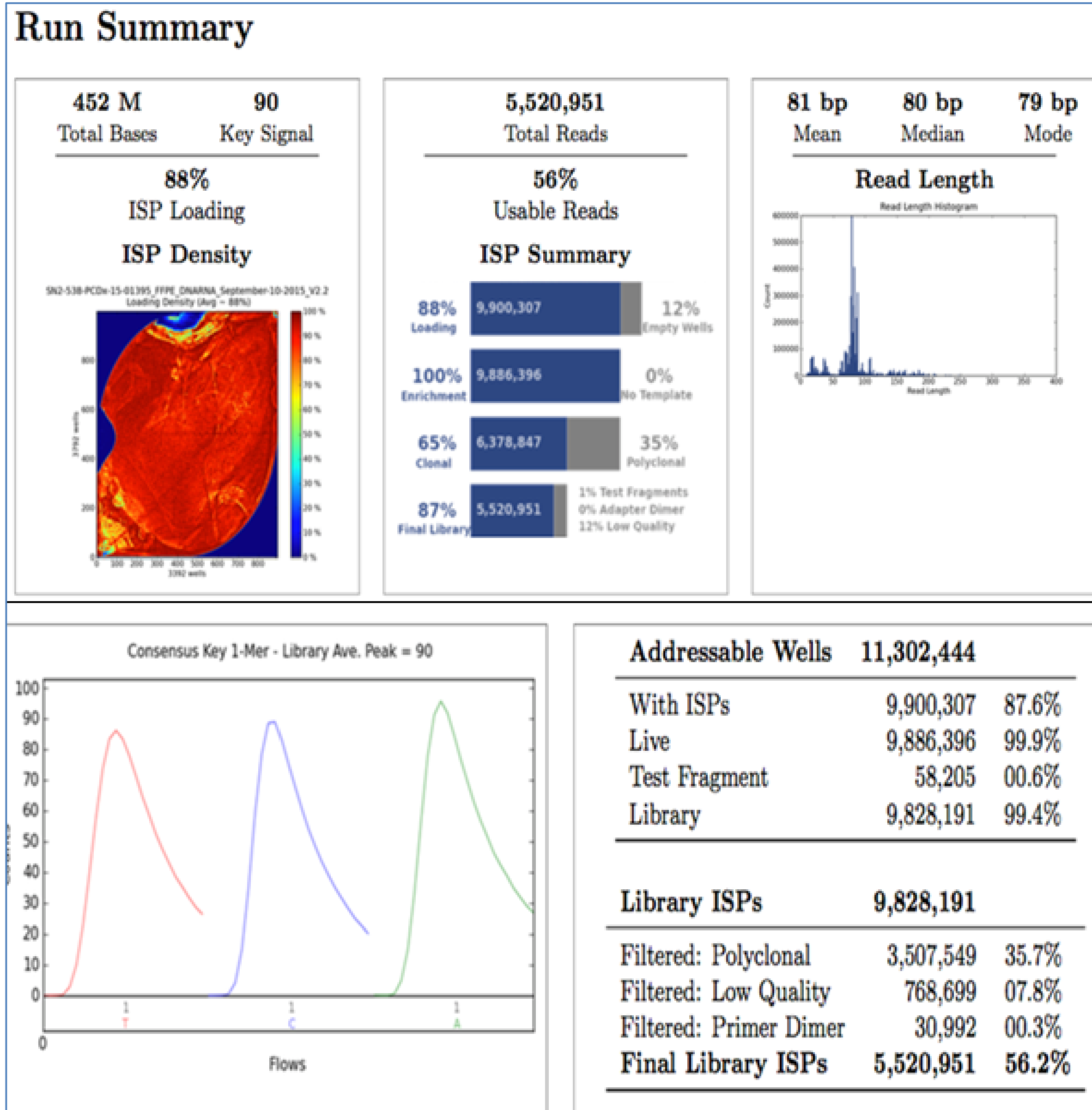
Jennifer Kellman-Fritz, PhD
Chair
University Human Subjects Review Commitee

Appendix B: The Thermo-Fisher Ion Chef and Thermo-Fisher OneTouch-2.

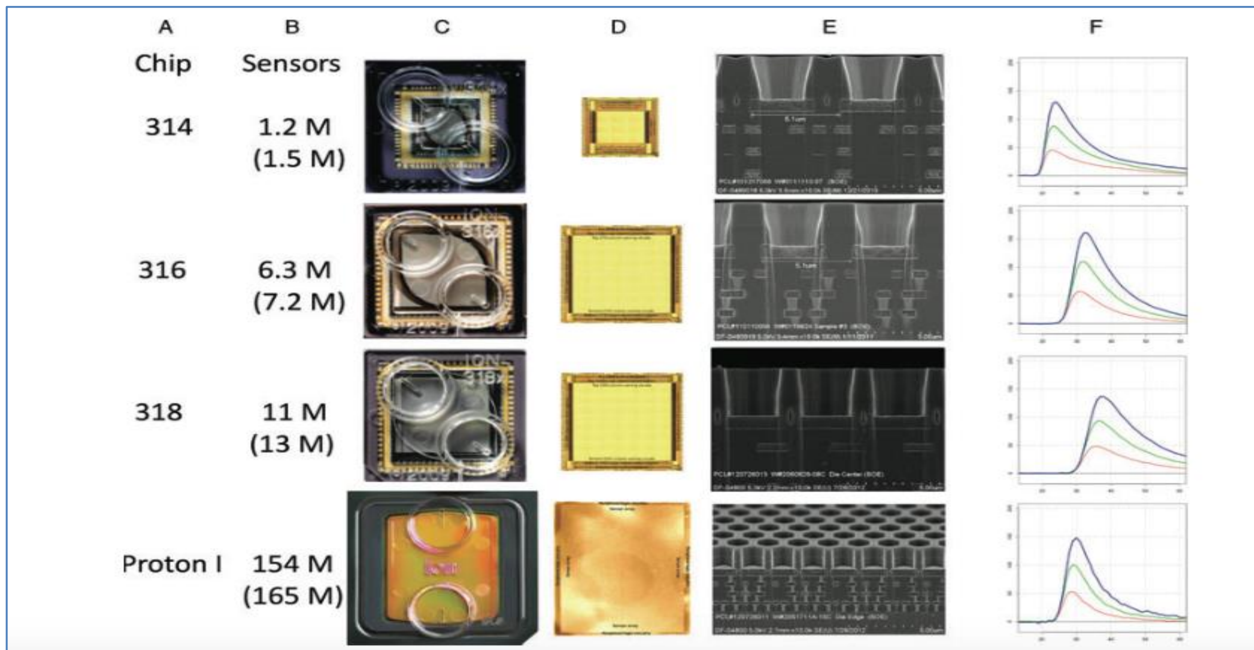
The above picture is for Thermo-Fisher Ion Chef ™ used for DNA template preparation. Note. From " Ion Personal Genome Machine® (PGM™) system" by Thermofisher Scientific (https://www.thermofisher.com/order/catalog/product/4484177?ICID=search-product). Copyright 2016 by Thermofisher Scientific. Reprinted with permission.

The OneTouch-2 (OT2) for sample template preparation (right) and enrichment unit (left) that used to amplify and enrich DNA templates before loading into PGM for DNA sequencing. Note. From "Ion One Touch ™ 2 system" by Thermofisher Scientific (https://www.thermofisher.com/order/catalog/product/4474779?ICID=search-product) Copyright 2016 by Thermofisher Scientific. Reprinted with permission.

Appendix C: Example of Summary Report generated by Ion Torrent PGM sequencing software for an NGS run.



Appendix D: Ion Torrent biochips technology in NGS.



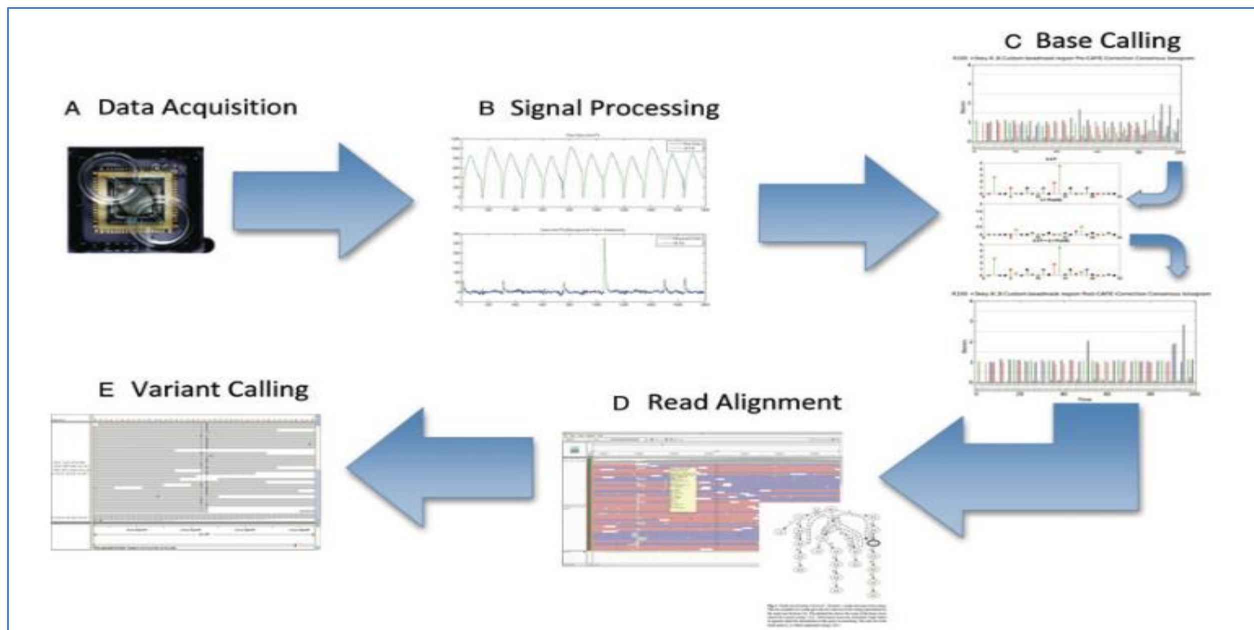


Figure 4 and 5. Chip scaling and Informatics pipeline. Adapted from " Progress in ion torrent semiconductor chip based sequencing" by B. Merriman and J. Rothberg, 2012, *Electrophoresis, 33*,P. 3404 and 3413.Copyright 2016 by John Wiley and Sons (RightsLink). Reprinted with permission.

Appendix E: Metric variables used for analysis of influence of polyclonal on NGS.

Appendix A: Metric variables used for evaluation of sequencing performance of IC and OT2 templates. Blue variables represent IC data and green variables represent OT2 data.

Poly.	Poly.	Em.	Em.	T. Reads	T. Reads	N.Temp.	N.Temp.	L ib.	L ib.	Usab. L	Usab. L
16	59	43	11	3727144	188170	3	0	81	77	48	31
47	15	6	52	1285939	344198	0	10	84	46	47	41
24	39	27	15	4618076	188170	1	0	65	86	44	20
30	44	15	13	2792170	536018	0	0	82	85	48	47
16	65	36	17	4201664	542199	2	0	58	69	48	83
16	27	43	17	4619073	3946	3	0	84	80	48	59
22	41	69	22	2310154	542199	53	3	84	82	56	48
52	14	10	68	4482017	559109	0	27	84	61	37	54
23	37	26	10	5520951	3956	0	0	85	82	47	52
23	53	26	1	2881483	559109	1	1	85	80	49	38
20	39	27	0	3771806	677654	1	0	85	89	52	54
29	55	91	16	3543503	687151	58	0	86	65	48	30
20	15	46	31	3419140	865185	3	0	86	48	40	41
54	47	7	11	2347865	912614	0	0	87	76	35	41
18	65	30	5	4123390	912614	3	0	89	82	57	28
40	52	13	15	2521604	931594	0	0	81	53	49	26
41	52	10	7	1354992	3220934	0	0	67	82	53	40
49	25	10	84	1364657	3308888	0	42	57	71	70	57
33	29	19	25	1479156	3543503	0	4	81	69	54	49
34	17	19	45	1480372	509542	0	6	75	46	54	39

Poly.	Poly.	Em.	Em.	T. Reads	T. Reads	N.Temp.	N.Temp.	L ib.	L ib.	Usab. L	Usab. L
52	30	11	18	1496414	3141117	0	0	68	71	33	50
44	70	10	7	1546009	1796591	0	0	58	78	47	23
42	43	12	22	1644517	4779951	0	0	82	61	47	35
22	16	69	50	1932945	3204810	53	6	90	52	84	45
34	36	12	15	1945237	4161907	0	0	79	86	53	55
24	45	76	19	2003583	1487148	48	0	74	58	48	42
69	50	27	13	2357980	2931643	1	0	61	58	53	39
18	15	74	32	2358110	2988692	46	3	68	59	56	46
29	51	91	10	2369115	3842791	58	0	80	60	48	41
67	19	14	31	2406386	3185633	0	1	76	60	24	50
56	14	11	54	2438033	3088224	0	4	78	61	34	33
33	23	12	16	2566967	4625177	0	0	81	61	56	57
38	18	16	49	2746871	5788493	0	6	73	61	51	51
25	31	57	33	2844533	3171902	4	1	81	61	51	39
22	32	67	29	2908174	2088083	33	1	69	61	54	42
28	22	18	21	2913596	3245131	0	0	88	61	58	55
43	42	11	16	2913596	3578035	0	0	62	61	70	45
55	49	8	19	2979591	3159315	0	0	68	61	36	35
19	65	37	21	2979591	5468838	2	0	61	62	54	7
38	47	15	18	3041459	5849615	0	1	58	63	52	44
21	35	38	28	3058352	4444907	2	0	64	64	45	35
68	54	23	15	3069090	2834369	1	0	65	65	53	37
67	55	5	14	3082522	366225	0	0	84	65	70	33
55	71	8	27	3099400	4461573	0	1	82	65	36	46

Poly.	Poly.	Em.	Em.	T. Reads	T. Reads	N.Temp.	N.Temp.	L ib.	L ib.	Usab. L	Usab. L
22	31	1	19	3102120	2658686	1	1	84	66	50	54
16	31	36	19	3154839	5258964	1	1	84	67	50	54
33	70	80	5	3191111	3466020	48	0	81	68	47	22
18	66	70	17	3360650	2211089	31	0	87	55	67	19
20	18	50	38	3368599	4621121	6	3	68	68	55	49
46	62	18	23	3385869	2659006	0	0	72	68	43	21
20	77	73	17	2003583	1487148	23	0	74	58	48	42
41	46	12	20	2357980	2931643	0	0	61	58	53	39
15	22	48	16	2358110	2988692	5	0	68	59	56	46
18	12	56	70	2369115	3842791	16	30	80	60	48	41
19	39	42	19	2406386	3185633	2	0	76	60	24	50
31	58	13	22	2438033	3088224	0	0	78	61	34	33
14	28	45	14	3412419	2030678	1	0	87	68	63	5
0	14	86	36	3429187	2710144	40	1	61	68	52	37
2	11	72	72	3437062	1571432	27	19	87	68	49	62
21	43	48	15	3455296	2800492	5	0	57	69	48	66
25	37	23	17	3532977	3746956	0	0	81	60	50	38
20	46	40	15	3542274	4882380	3	0	63	69	59	21
42	21	13	27	3550358	2206944	0	0	57	69	50	61
34	35	87	12	3624667	5036363	84	0	71	60	75	52
28	41	22	13	3717449	4080673	72	0	63	69	75	53
34	22	87	22	3726083	3651436	33	1	80	70	46	48
28	53	27	12	3742113	635254	27	0	64	70	50	47
33	55	79	17	3772882	3185820	45	0	82	71	53	39

Poly.	Poly.	Em.	Em.	T. Reads	T. Reads	N.Temp.	N.Temp.	L ib.	L ib.	Usab. L	Usab. L
25	43	23	15	3772882	3772587	0	0	69	71	51	48
14	19	35	32	3788236	1360051	0	2	81	71	87	56
20	30	50	25	3867012	3707029	6	1	70	71	52	43
2	29	72	21	3868777	2069477	27	0	58	71	56	52
21	34	40	22	3883411	5295435	2	0	67	72	50	32
16	54	68	15	4026027	1782042	17	0	78	72	43	31
52	15	11	45	4030951	3528697	0	2	70	72	50	48
73	15	24	56	4035689	4882380	1	16	64	72	53	50
28	36	22	30	4044963	3184642	0	2	81	73	55	43
18	27	44	16	4236974	550144	3	0	70	73	84	52
33	46	23	18	4262882	5145426	0	0	53	73	46	42
33	57	0	0	4312548	2596995	0	0	74	73	45	34
31	39	4	27	4360090	3054396	4	1	81	73	86	44
31	71	0	8	4360726	2030678	0	0	87	68	63	5
36	51	0	22	3412419	2710144	0	1	61	68	52	37
22	60	35	11	3429187	1571432	2	0	87	68	49	62
23	54	51	18	3437062	2800492	4	0	57	69	48	66
18	55	60	9	3455296	3746956	18	0	81	60	50	38
24	72	18	10	3532977	4882380	1	0	63	69	59	21
23	69	25	12	4380399	1739437	1	0	83	74	40	50
18	61	40	18	4391232	3822208	6	0	50	74	52	44
17	64	39	15	4391232	4456910	3	0	63	75	52	56
27	67	19	11	4493318	4625177	0	0	81	75	53	44
40	67	14	12	4501500	2191923	0	0	77	76	59	29

Poly.	Poly.	Em.	Em.	T. Reads	T. Reads	N.Temp.	N.Temp.	L ib.	L ib.	Usab. L	Usab. L
20	44	67	11	4504740	3499463	22	0	63	76	51	45
57	64	11	9	4561055	471043	0	0	57	0	58	19
34	70	21	11	4563829	4045339	1	0	64	77	49	30
41	15	15	54	4563829	3377443	0	6	59	50	48	27
65	57	12	12	4573709	4223708	0	0	67	77	44	32
33	34	21	69	4574456	4641896	0	3	61	77	53	31
19	41	43	14	4674061	3839094	2	0	23	53	63	15
86	33	5	21	4765069	3067548	0	0	89	77	58	17
28	43	77	21	4799348	2451206	49	0	58	77	55	30
18	23	42	29	4804004	4461573	3	1	84	78	53	23
27	76	19	10	4811830	2715554	0	0	69	78	58	19
27	67	67	7	4837779	3103272	47	0	88	78	50	22
44	66	12	15	4847687	2197951	0	0	58	78	57	47
48	41	9	14	4853440	2195294	0	0	68	79	34	26
18	33	61	21	4893104	2453734	7	0	57	0	49	17
22	43	62	21	4903503	1709281	30	0	63	80	47	44
48	80	9	0	4941549	2242032	0	0	79	20	24	31
53	69	11	9	4974652	3983784	0	0	85	80	50	14
34	60	20	7	4994411	4038780	0	0	83	80	49	46
14	60	77	7	4380399	3934044	17	0	71	25	90	53
61	61	48	17	4391232	1739437	4	0	83	74	40	50
30	36	15	16	4391232	3822208	0	0	50	74	52	44

Appendix F: Publisher Permission for use of Ion Torrent platform figures.

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Oct 02, 2016

This Agreement between Akram A Aldilaimi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	3957260949716
License date	Sep 27, 2016
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Electrophoresis
Licensed Content Title	Progress in Ion Torrent semiconductor chip based sequencing
Licensed Content Author	Barry Merriman,Ion Torrent R&D Team,Jonathan M. Rothberg
Licensed Content Date	Dec 4, 2012
Licensed Content Pages	21
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Electronic
Portion	Figure/table
Number of figures/tables	4
Original Wiley figure/table number(s)	Figure 1B, Figure 2,Figure 4,and Figure 12
Will you be translating?	No
Title of your thesis / dissertation	A Comparison of Two Methods of Template Amplification for Next Generation Sequencing: Implications of Polyclonal Formation on DNA Sequence for Several Cancer Tissues
Expected completion date	Dec 2016
Expected size (number of pages)	50
Requestor Location	Akram A Aldilaimi 5713 Appoline DEARBORN, MI 48126 United States Attn: Akram A Aldilaimi
Publisher Tax ID	EU826007151
Billing Type	Invoice
Billing Address	Akram A Aldilaimi 5713 Appoline