"Characterization of mutation in ALK kinase domain and ABL kinase domain of BCR-ABL fused gene in cancerous specimen with NGS"

Dissertation submitted in partial fulfillment for the degree of

Master of Science in Biotechnology

Submitted By

Alisha Pradhan

KIIT School of Biotechnology, Campus- 11

KIIT deemed to be University

Bhubaneswar, Odisha, India

Under the Supervision of

Dr. Sachchida Nand Pandey

MedGenome Labs Ltd.

Bangalore-560099

India

CERTIFICATE

This is to certify the dissertation entitled ***"Characterization of mutation in ALK kinase domain and ABL kinase domain of BCR-ABL fused gene in cancerous specimen with NGS"*** Submitted by ***Alisha Pradhan*** in partial fulfilment of the requirement for the degree of Master of Science in Applied Microbiology, KIIT School of Biotechnology, KIIT University, Bhubaneswar bearing Roll No. 1662001 & Registration No. 1664735147 is a bonafide research work carried out by her under my guidance and supervision from ***8th January*** to ***11th May 2018.***

Name and signature

CERTIFICATE

This is to certify that the dissertation entitled ***"Characterization of mutation in ALK kinase domain and ABL kinase domain of BCR-ABL fused gene in cancerous specimen with NGS"*** submitted by *Alisha Pradhan, Roll No. 1662001, Registration No. 1664735147* to the KIIT School of Biotechnology, KIIT deemed to be University, Bhubaneswar- 751024, for the degree of Master of Science in Applied Microbiology is her original work, based on the results of the experiments and investigations carried out independently by her during the period from ***8th January*** to ***11th May 2018*** of study under my guidance.

This is also to certify that the above said work has not previously submitted for the award of any degree, diploma, fellowship in any Indian or foreign University.

Date:

Place: *Supervisor name & signature*

DECLARATION

I hereby declare that the dissertation entitled ***"Characterization of mutation in ALK kinase domain and ABL kinase domain of BCR-ABL fused gene in cancerous specimen with NGS"*** submitted by me, for the degree of Master of Science to KIIT University is a record of bonafide work carried by me under the supervision of, ***Supervisor name & designation***, ***University, City, State, Country***.

Date:

Place: *Name & Signature*

Abstract

Next-generation DNA sequencing techniques are opening fascinating opportunities in the field of life sciences. It has surpassed its original goal of genomic sequencing and have opened novel fields and its applications in biology and medicine. NGS is making its way through every field of biomedical research. Current sequencing platforms even allow unprecedented view into mixture of RNA and DNA samples. To begin with, NGS requires library preparation with fragments of DNA and RNA molecules being fused with adapters followed by PCR amplification and then sequencing. There are numerous protocols for library preparation, so it is important to look for robust library preparation methods that produce a representative and a non-biased source of nucleic acid material from the genome under investigation.

Here, I examine two kinase domains i.e. ALK kinase domain and ABL1 kinase domain of BCR-ABL fused gene using the NGS platform. Rearrangements in the ALK gene in NSCLC represent a novel molecular target. BCR-ABL fusion gene as well can be targeted for imatinib treatment.

ALK rearrangements are usually assessed by IHC, FISH and RT-PCR, molecular approaches have only recently been taken to account. BCR-ABL fusions that are detected by Sanger sequencing, have also been assessed.

We have evaluated an amplicon based NGS method to detect the tyrosine kinase domain mutations for ALK and ABL of BCR-ABL fused gene.

Standardization and validation of these methods is essential for correct and reproducible results for clinical utilization. The data generated may provide an effective and alternative molecular method of testing to the conventional ones for the detection of known and also novel ALK rearrangements or BCR-ABL fusions. Also with accumulating data showing the advantages and disadvantages of these methods combining these tests for better efficiency maybe beneficial to the patients.

Acknowledgments

On this occasion of submitting my M.Sc thesis, I would like to express my deepest gratitude to the persons who helped me in some way or the other for successful and smooth completion of the project work.

I would like to express my sincere thanks Dr. Sakthivel Murugan, Associate Director, Medgenome Labs Ltd for providing me with my project topic , for his timely support and critical advises.

I would like to express my deepest gratitude to Dr. Sachchida Nand Pandey, Senior Scientist Medgenome Labs Ltd ,for his guidance. His constant advises, providing us with solutions during the whole term helped me to perform at my best. To push our self and yet be calm are the things I learnt from him.

I would fail in my duties if I do not acknowledge my manager Miss. Merina Oliver, Sr. Research Associate, Medgenome Labs Ltd. Her constant efforts to look for solutions, to lead the team yet be working with us inspired us. She never expressed disappointment even when something did not work out. She strives to work harder after every failure. I would like to thank her immensely for teaching us to be positive.

I would like to thank my team member Miss Sarita Singh, Research Associate, Medgenome Labs Ltd for teaching me everything I know about my work today. She has held my hands from the first day and taught me the whole lab flow. My manager has entrusted the correct person as my teacher.

I would like to extend my thanks to my lab mates Miss. Sheuli, Mr Jamir, Mr.Satish, Miss Esha, Miss Ludiya, Miss Manali, Miss Shilpi for giving me advises at time of confusion and for creating a positive working environment.

I would also like to thank the sequencing team- Mr. Joyner, Mr Amol and Mr Harsha for their inputs.

My thanks to all the other teams in the lab and the HR for their help, support and positivity.

Last but not the least I would like to thank Dr.Mrutyunjar Suar, Director & Head , School of Biotechnology, KIIT University for making me capable to join the corporate world.

Table of Contents

Abstract i

Attestation

Acknowledgments ii

Table of Contents iii

Abbreviations iv

List of Figures v

1 Introduction 1

1.1 Background and Context 5

1.2 Scope and Objectives 17

1.3 Achievements 19

1.4 Overview of Dissertation 20

2 Background and Rationale 24

3 What did you learn? 27

3.1 Summary 27

3.2 Evaluation 28

4 How to complement in Corporate? 30

References 31

Abbreviations

1.NGS: Next Generation Sequencing

2.PCR: Polymerase Chain Reaction

3.ALK: Anaplastic Lymphoma Kinase

4.BCR-ABL: Break Point Cluster region-Abelson Murine Leukemia

5.RT-PCR: Real Time-Polymerase Chain Reaction

6.DNA: Deoxyribonucleic Acid

7.RNA: Ribonucleic Acid

8.KRAS: Kirsten Rat Sarcoma virus

9.GTP/GDP:Guanosine triphosphate / Guanosine diphosphate

10.MAPK: Mitogen Activated Protein Kinase

11.TKI: Tyrosine kinase inhibitor

12.ADC: Adeno carcinoma

13.QC:Quality Control

14.cffDNA: Cell free fetal Deoxyribonucleic Acid

15.ALCL: Anaplastic large cell lymphoma

16.IHC: Immuno histochemistry

List of Figures

Figure-1

Workflow chart for ALK kinase domain testing.....3

Figure-2

Work flow chart for ABL kinase domain testing.....4

Figure-3

Treatment algorithm for patients with NSCLC..... 8

Figure-4

Philadelphia chromosome 11

Figure-5

Sequencing by synthesis 15

Figure-6

Work flow for NGS 16

List of Tables

Table-1

List of primers for ALK kinase domain.....21

1 Introduction

The first human genome was sequenced in 2001 after many years of effort by dozens of international laboratories. The growing power and reducing cost of NGS these days have encouraged laboratories to acquire new high-throughput sequencing devices allowing them to analyze tens or hundreds of genes, or even the entire exome. The ability to search for mutations simultaneously in a large number of genes is finding applications in the diagnosis of Mendelian diseases (including at birth), routine screening for hetero-zygotes, and pre-conception diagnosis is allowing it to become a standard technology. NGS is now sufficiently sensitive to analyze circulating fetal DNA in maternal blood (cell-free fetal DNA, cffDNA), enabling applications such as non invasive diagnosis of fetal sex (and X-linked diseases), fetal rhesus among rhesus-negative women, trisomy and also the Mendelian mutations.

NGS is becoming increasingly popular because of the quality, robustness and the low noise data it produces. For a successful data to be generated it requires expertise at the wet lab as well the bio-informatics team as to ensure high quality data and data interpretation.

Fluorescence in situ hybridization (FISH) is currently acknowledged as the “gold standard” for detection of ALK rearrangements. The immunohistochemical (IHC) method, which can detect ALK protein expression independently of the underlying mechanism mediating its over expression, is used as a pre-screening test, alongside FISH, to determine ALK status in formalin-fixed paraffin embedded (FFPE) tissue specimens. However, even though IHC is widely implemented in pathology laboratories, easy-to-use, and automatically performed, its interpretation remains difficult to standardize and time-consuming. In addition, FISH is expensive, labor intensive, requires expert pathology assessment, and is not amenable to multiplexing.[1]

The Sanger capillary sequencing technique analyzing BCR-ABL1 is considered the gold standard for mutation detection in a clinical laboratory [2] knowing that this assay has a sensitivity of 15–20 % (i.e., 15–20 mutated transcripts in 100 total BCR-ABL1 transcripts). However, in recent years evidence has been mounting that early detection of mutations may help stratifying patients and predict their therapeutic responsiveness.NGS is said to identify mutations below the lower detection limits of

Sanger sequencing and uncover complex clonal textures in many cases [2].

NGS is currently assisting the more conventional methods and in some cases working as an alternative molecular approach for assessment in routine laboratories. ALK and BCR-ABL testing are amplicon based techniques as an amplicon library improves sensitivity of mutation detection, enhances throughput and allows quantification of relative tumor burden. These molecular approaches strengthen the accuracy of diagnosis by resolving borderline and discordant cases.

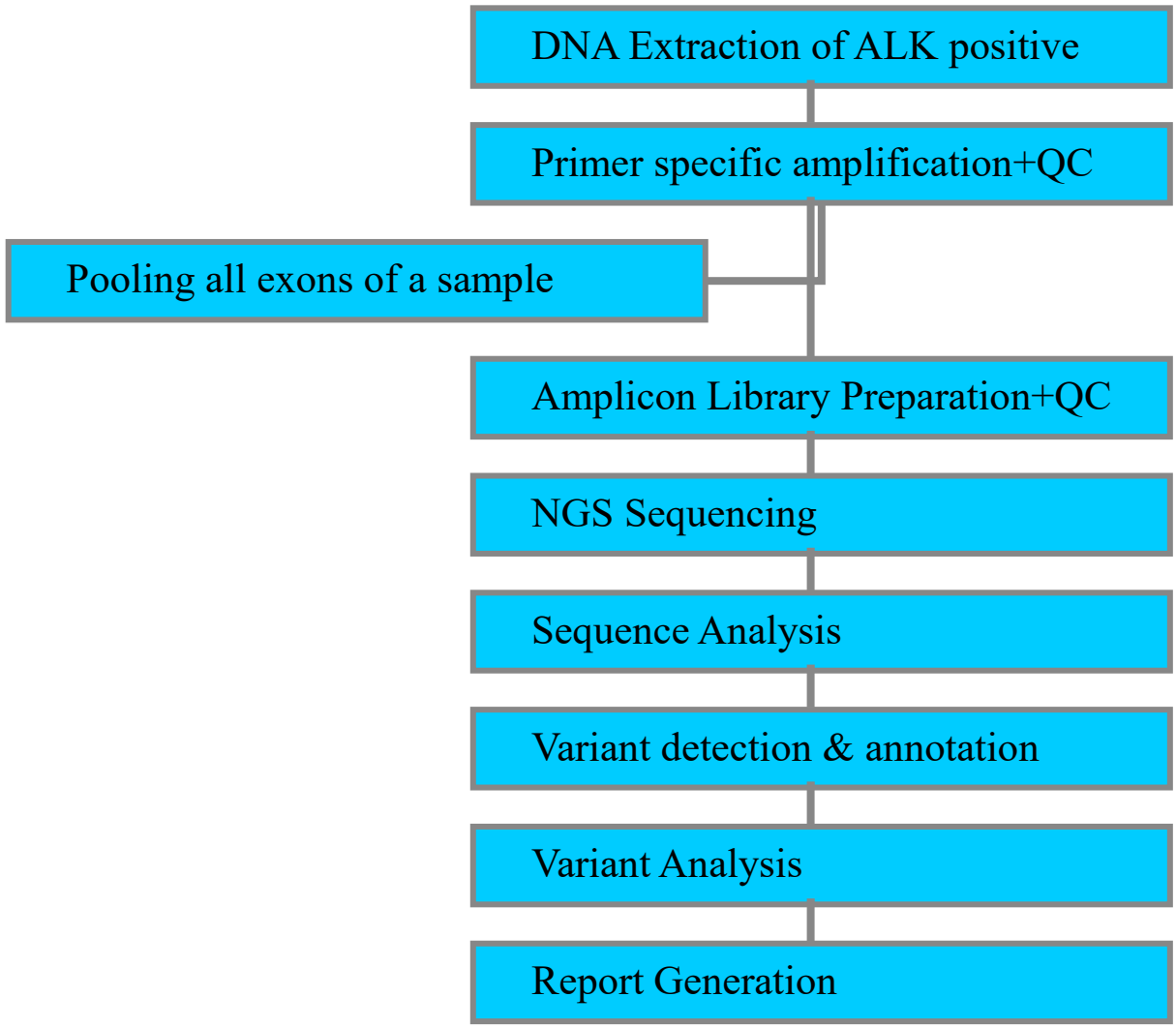

Figure 1. Work flow for ALK Kinase domain testing

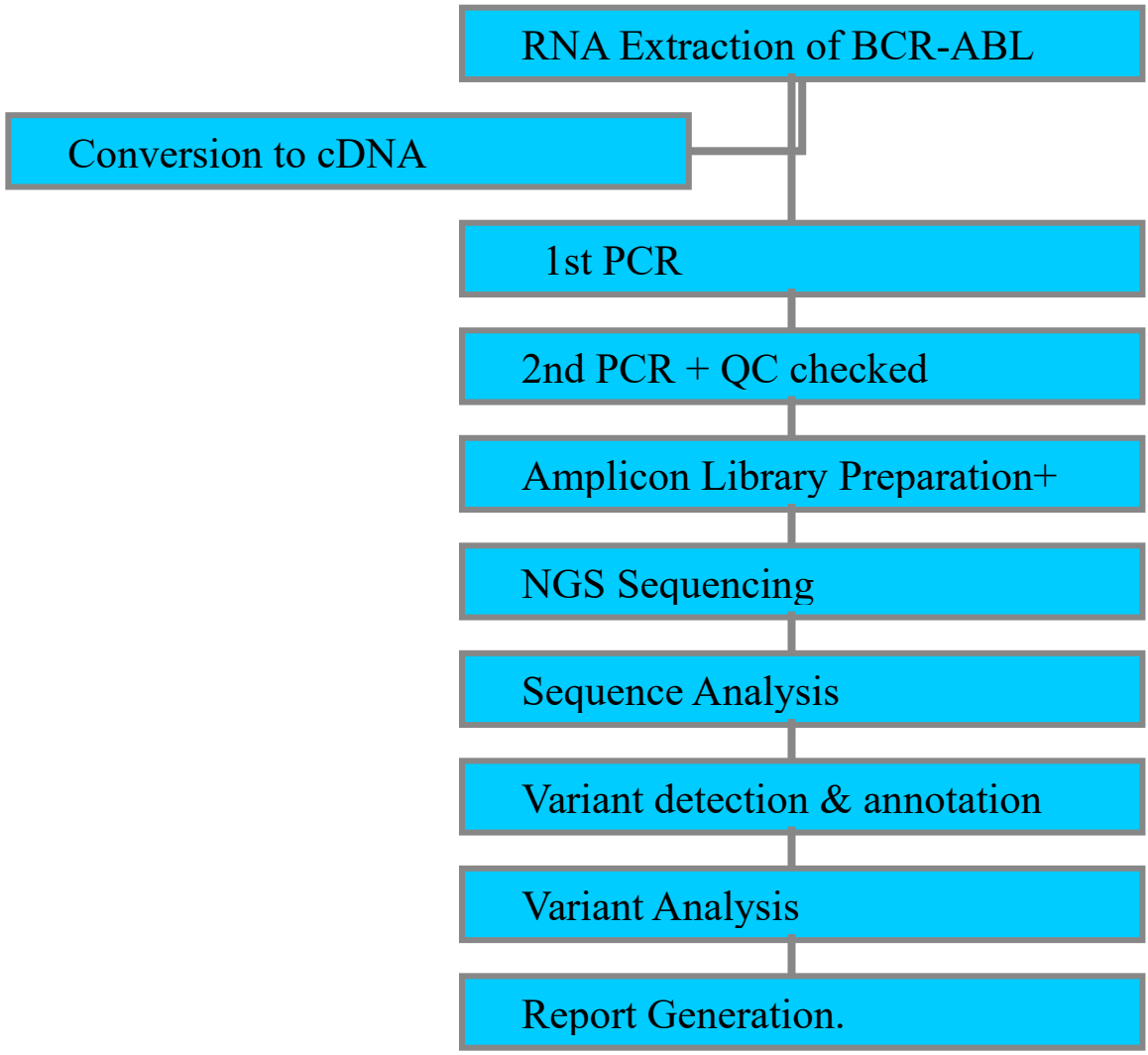

Figure 2. Work Flow chart for ABL kinase domain testing of BCR-ABL fused gene.

1.1 Background and Context

Lung cancer

Lung cancer is the uncontrolled growth of abnormal cells that start off in one or both lungs. The cells lining of the air passages are usually affected. The abnormal cells do not develop into healthy lung tissue instead they divide rapidly and start forming tumors. As tumors become larger and increase in number they weaken the lung's ability to provide oxygen in the bloodstream.

Benign tumors remain in one place and do not spread . It won't invade nearby tissues or spread to other areas of the body i.e.metastasize.

Metastasis refers to cancer spreading beyond its site of origin to other parts of the body.

Malignant tumors spread to other parts of the body either through the bloodstream or the lymphatic system. It can grow into and destroy nearby tissue. It can also metastasize. When cancer starts in lung cells, it is called primary lung cancer.

Primary lung cancer originates in the lungs, while secondary lung cancer starts somewhere else in the body, metastasizes, and reaches the lungs. They are considered different types of cancers and are not treated in the same way.

The two main types of cancer are

i) small-cell lung cancer, and

ii) non-small-cell lung cancer,

They are differentiated according to the shape of cells observed under a microscope. The most common symptoms are coughing, shortness of breath and chest pains.

I) Small cell lung cancer usually starts in cells that line the bronchi in the center of the lungs. The main types of small cell lung cancer are small cell carcinoma and combined small cell carcinoma i.e. mixed tumor with squamous or glandular cells.

ii) Non–small cell lung cancer usually starts in glandular cells on the outer part of the lung. This type of cancer is called adeno carcinoma. Non–small cell lung cancer can also start in flat, thin cells called squamous cells. This type of cancer is called

squamous cell carcinoma of the lung. Large cell carcinoma is another type ofnon–small cell lung cancer, but it is less common.

1.1.1 Molecular targets in lung cancer

The molecular basis of lung cancer is complex and heterogeneous. Development of lung cancers is a multi-step process which includes development of multiple genetic and epigenetic alterations, particularly activation of growth promoting pathways and inhibition of tumor suppressor pathways. Signaling pathways regulated by oncogenes and tumor suppressor genes are often interconnected.

A) KRAS

KRAS is part of the RAS family of proto-oncogenes. Ras proteins are bound to GDP which are inactive in normal quiescent cells. When activated they form GTP bound form and in turn activates a number of upstream growth factor receptors. As well as a number of downstream pathways including MAPK. KRAS plays a critical role in downstream signal transduction induced by a variety of growth factor receptors including EGFR. When mutations occur they alter the GTPase activity of the protein is altered thus causing inactivation of the active RAS-GTP to GDP. This leads to increased signaling through multiple downstream growth promoting pathways. KRAS mutations lead to constitutive activation of pathways downstream of EGFR . A meta-analysis has shown KRAS mutant tumors are resistant to EGFR TKI.[3]

B) EGFR

EGFR encodes a transmembrane tyrosine kinase with an extracellular ligand-binding domain and an intracellular component including a tyrosine kinase domain. Alterations of EGFR are involved in the pathogenesis of many tumors including NSCLC. Binding of the ligand epidermal growth factor leads to receptor homo or hetero dimerisation with other members of the EGFR family and activation of the tyrosine kinase domain . EGFR is involved in regulation of numerous oncogenic

functions such as cell proliferation, survival, differentiation, neovascularisation, invasion and metastasis . Activating mutations in EGFR lead to constitutive tyrosine kinase activation and oncogenic transformation of lung epithelial cells in vitro .[3]

C) BRAF

BRAF encodes a serine/threonine protein kinase and is the downstream effector protein of KRAS and activates the MAPK signal transduction pathway involved in regulation of cell proliferation and survival. While activating BRAF mutations are common in melanoma , they occur in only about 3% of NSCLC. As BRAF and KRAS genes are part of the signaling pathway mediated by EGFR. BRAF mutations in lung cancer occur almost always in ADC . The uncommon, BRAF mutations represent an important therapeutic target to the available targeted therapies in clinical use for melanoma,although there is only limited data about the clinical response this approach in NSCLC . [3]

D) ALK

Rearrangements of the receptor tyrosine kinase ALK result most commonly in fusions of the intracellular kinase domain with the amino terminal end of echinoderm microtubule associated protein-like 4 (EML4) which occurs in a subset of lung cancers . The rearrangement results from a short inversion in chromosome 2p, whereby in the commonest variant, intron 13 of EML4 is fused to intron 19 of ALK. Numerous variants of EML4-ALK fusions have been identified due to differing lengths of EML4, the commonest being exons 1-13 of EML4 joining to exons 20-29 of ALK . [3]

Although ALK rearrangements are mutually exclusive with EGFR and KRAS mutations, cases of coexistence with EGFR mutations have been reported they are known to provide a mechanism for TKI resistance. While ALK inhibition with the tyrosine kinase inhibitor crizotinib produces profound responses, drug resistance develops with evidence of secondary ALK point mutations and activation of EGFR signaling implicated in some cases.

E) ROS1

ROS1 is a proto-oncogene located on chromosome 6q22 which encodes for a transmembrane tyrosine kinase receptor which has high homology with ALK and its protein kinase domain. ROS1 rearrangements like ALK rearrangements are more common in patients who are younger, never smokers or of Asian ethnicity. Furthermore, there is in vitro and early clinical evidence that lung cancers with ROS1 rearrangements are sensitive to kinase inhibitors including the ALK/MET inhibitor crizotinib. [3]

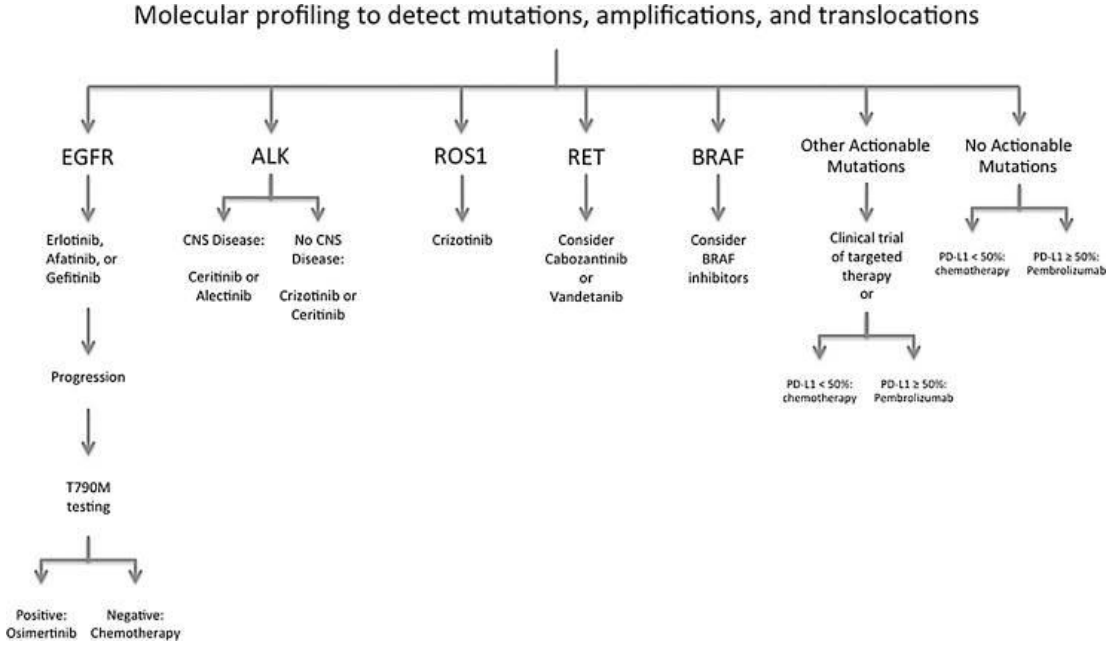

Figure 3. Treatment algorithm for patients with advanced/metastatic non–small cell lung cancer. ALK, anaplastic lymphoma kinase; BRAF, B-Raf proto-oncogene; CNS, central nervous system; EGFR, epidermal growth factor receptor; PD-L1, programmed death-ligand. 1; ROS1, ROS1 proto-oncogene[3]

1.1.2 Non Small Cell Lung Cancer(NSCLC)

The characteristics of NSCLC can be defined by the above molecular targets. Since,

the EGFR mutation and ALK rearrangements are the most clinically relevant at present . There prevalence's can be seen in patients suffering from lung cancer in different regions. However, the presence of these driver mutations is generally seen to be mutually exclusive to the other in the same tumor.

It has been noticed that lung tumors bearing EML4-ALK rearrangement which is the most common type of fusion in NSCLC [4] are non-responsive to conventional chemotherapy or EGFR-tyrosine kinase inhibitors. They maybe sensitive to a specific tyrosine kinase inhibitors. Based on the current understanding of therapeutic molecular targets of EGFR mutation and ALK gene rearrangement in NSCLC and the availability of corresponding targeted agents, a method of testing for molecular targets in NSCLC has been proposed which proposes for a step-wise approach for testing of individual targets which begins with EGFR then, if EGFR is negative, ALK fusion gene or other potential targets whichever is found to be appropriate.

1.1.3 Anaplastic lymphoma kinase(ALK)

Anaplastic lymphoma kinase (ALK) is a receptor tyrosine kinase which belongs to the insulin receptor super-family. It is involved in both normal development and oncogenesis. ALK appears to function in neuronal development and differentiation during embyrogenesis and with time its expression falls to low-levels at age three weeks and there on remains low throughout adult life.

ALK is encoded by a genomic locus found at the chromosomal band 2p23 in humans. ALK is also known by the name cluster designation CD246. ALK is a single-chain trans-membrane protein of 1620 amino acids (aa). ALK possesses an extra-cellular ligand-binding region, a trans-membrane spanning domain and a cytoplasmic kinase catalytic region. The kinase domain (KD) of ALK includes a three-tyrosine-containing motif (tyrosines 1278, 1282 and 1283) within its activation loop. These tyrosine residues represent major auto-phosphorylation sites that regulate the activation loop conformation, occluding access of ATP to the ATP-binding pocket in its non-phosphorylated state and moving outward and away from the binding pocket to allow unobstructed entry of ATP during the kinase-activation process following phosphorylation of the triplet tyrosines.[5,6]

Role of ALK in lung cancer pathogenesis

The anaplastic lymphoma tyrosine kinase (ALK) gene was first described as a driver mutation in anaplastic non-Hodgkin's lymphoma. Dysregulated ALK expression is now an identified driver mutation in nearly twenty different human malignancies, including 4-9% of non-small cell lung cancers (NSCLC). The tyrosine kinase inhtribitor crizotinib is now seen to be more effective than the standard chemotherapeutic agents in treating ALK positive NSCLC, making molecular diagnostic testing for this dysregulated ALK expression a necessary step in identifying optimal treatment procedure.[6]

Leukemia

Chronic lymphocytic leukemia(CLL) and chronic myeloid leukemia(CML) are the most common types of adult leukemia. In fact CML was the first kind of leukemia to be described. CML represents 7–20% of all hematologic malignancies. The frequency of occurrence is higher among adults then children and although both sexes are affected, a slight predominance is seen in males.

CML is a clonal myelo-proliferate neoplasia of cells of the granulocyte lineage, which has the ability to mature and differentiate. Clinically, CML is a triphasic disease. The stages of the disease are differentiated according to their evolution and can be divided into a i)chronic (early) phase (CP) ii)an accelerated phase of the disease and iii)a terminal (blast) phase of CML, similar to acute leukemia. CML invariably progresses from an initial chronic phase (CP) to an accelerated phase (AP) and terminal blast crisis (BC)

1.1.4 CML and the Philadelphia chromosome.

CML is characterized by the acquisition of a cytogenetic abnormality known as the Philadelphia chromosome(Ph). A translocation event that occurs between chromosomes 9 and 22, results in the Philadelphia chromosome, and subsequently the BCR-ABL fusion gene, which is transcribed and translated into a hybrid protein.

The generation of the BCR/ABL fusion gene is necessary for the development of CML. The fusion gene comprises almost the entire coding region of the Abelson (ABL) gene on chromosome 9, and the changeable coding region of the break point cluster region (BCR) gene on chromosome 22. The BCR/ABL gene in the variant Ph chromosome translocation is generated by variant rearrangements with one or several other genomic regions. Compared to CML with the standard Ph chromosome, CML with variant rearrangement do not confer any specific phenotypic or prognostic impact, despite their genetically complex nature. In addition to the involvement of chromosomes 9 and 22, approximately 5% of CML cases are characterized by variant complex translocations involving a third chromosome. [7,8]

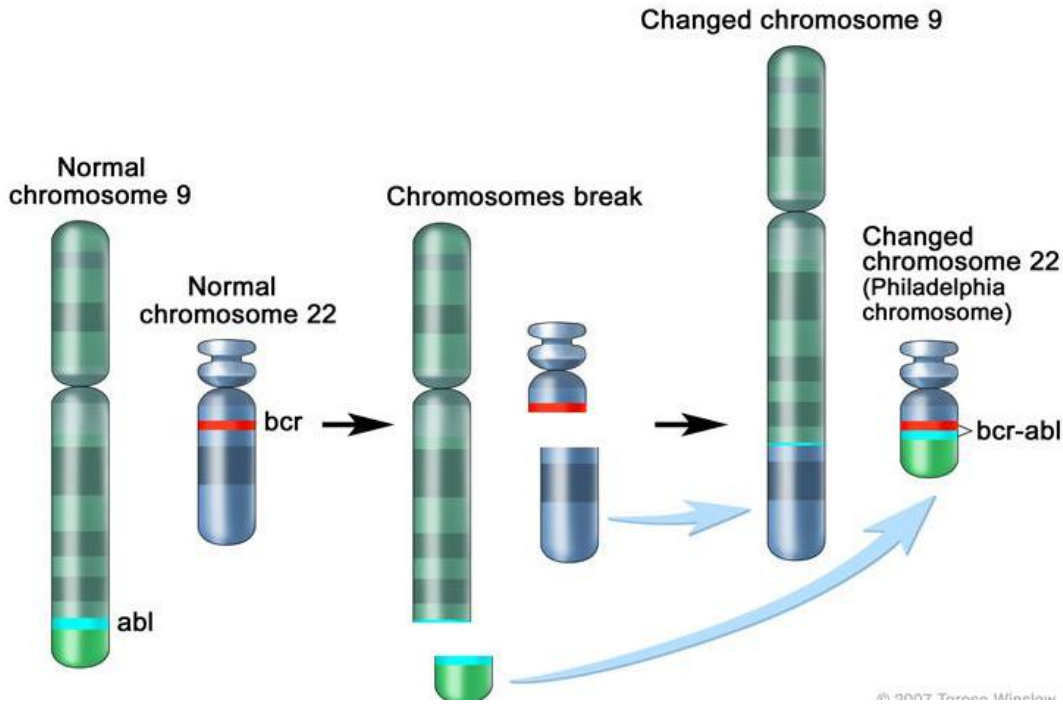

Figure 4. Philadelphia chromosome.

1.1.5 BCR-ABL fusion.

In chronic myelogenous leukemia (CML), the the major break point cluster region (M-bcr) lies between exons 12 and 16. The translocation events involving the M-bcr transcripts encode a 210 kDa protein termed P210. Translocation that occurs within

the minor(m)-bcr region retain only the first exons of the BCR gene transcripts, which yields a smaller P190 gene product. A rare P230 fusion protein, with a micro (μ)-bcr break point between exons 19 and 20, has also been described in a number of variant cases. In addition, some rare breakpoints have also been sporadically reported, which may be grouped into 4 categories:

i)The BCR breakpoints originating within introns that lie outside the M-bcr, m-bcr, or μ-bcr fused to ABL a2.

ii) The BCR breakpoints occurring within exons fused to ABL a2.

iii)The typical BCR breakpoints M-bcr, m-bcr, or μ-bcr fused to ABL breakpoints located downstream of a2.

iv)The transcripts containing intervening sequences between BCR and ABL a2.

The BCR/ABL fusion gene product is a cytoplasmic 210-KDa protein with up regulated tyrosine kinase activity that is considered essential for growth and survival of the leukemic cells. As BCR fuses at the 5 side of SH3 in ABL, the tightly regulated function of SH3 is disturbed and this disables the control over the tyrosine kinase. BCR/ABL with its constitutive tyrosine kinase activity triggers a number of downstream signaling molecules including PI3K, MAPK, NFkB, RAS and STAT5. Supposedly, these signaling molecules and pathways act together to promote malignant transcription, to enhance genetic instability and to suppress apoptosis.[9]

1.1.6 Roots of Imatinib resistance

The fundamental relationship between BCR-ABL and leukemia genesis in CML and ALL prompted the development of Imatinib mesylate which was a rationally designed drug to specifically block BCR-ABL and was found to be remarkably effective.

Imatinib mesylate induces complete cytogenetic response (CCR) in a high proportion of chronic myelogenous leukemia (CML) patients. BCR/ABL kinase domain mutations affecting drug binding can lead to secondary resistance to Imatinib. The BCR-ABL gene product demonstrates constitutively activated tyrosine kinase activity. Imatinib mesylate has proven highly effective in the treatment of CML. Imatinib induces CCR in approximately 80% of newly diagnosed CP patients and has emerged as the first-line therapy for CML. The mechanisms underlying incomplete

elimination of malignant progenitors in imatinib-treated patients are not clear. There is evidence that quiescent primitive CML progenitors may be resistant to apoptosis following imatinib exposure. Primitive progenitors may be resistant to imatinib through mechanisms such as increased drug efflux activity. However, another possibility is that mechanisms known to cause secondary resistance to imatinib could be active in subsets of CML progenitors, allowing their persistence in the setting of overall responsiveness to the drug. [9]

The frequency and durability of Imatinib-induced remissions are significantly worse in progressed phases of CML . Thus, the variable depth of treatment responses seen with IM in distinct Ph+ disease entities strongly suggest that IM resistance is influenced by BCR-ABL-independent factors, which in turn cooperate with BCR/ABL in transformation and therapy resistance. Clinical IM resistance was subsequently found to be quite consistently associated with the detection of BCR-ABL-kinase mutations. However, the variable clinical and genetic features of IM resistance and IM resistance evolution cannot be explained solely by kinase mutations. The other mutations, which synergize with BCR-ABL to enable leukemic stem cell self-renewal could be of causal importance for the evolution of clinical kinase inhibitor resistance. The expansion of mutated clones under selective pressure of IM therapy clearly suggests that mutations confer a BCR-ABL dependent type of IM resistance which were developed to inhibit mutated BCR-ABL- kinases and hence to overcome mutation-dependent IM resistance. However, kinase mutations are neither necessary nor predictive for responsiveness to these novel compounds. Indeed, when kinase mutations are detected prior to, or early after commencing IM treatment they usually, but not always predict upcoming clinical IM resistance and progression.

Unlike Sanger-based sequencing, NGS provides more advanced mutation analysis with larger parallel, deeper target re-sequencing. It is therefore, able to detect lower frequency mutations. Mutation analysis has been much benefited by the development of sequencing technologies, especially next-generation sequencing (NGS).

1.1.7 Diagnosis by Illumina sequencing

Illumina sequencing technology is based on sequencing by synthesis (SBS) method. It is a widely adopted next-generation sequencing (NGS) technology worldwide for generating more than 90% of the world's sequencing data. In principle, the concept behind NGS technology is that the DNA polymerase catalyzes the incorporation of fluorescently labeled dNTPs into a DNA template strand during sequential cycles of DNA synthesis. During each cycle, at the point of incorporation, the nucleotides are identified by fluorophore excitation. However instead of sequencing a single DNA fragment, NGS extends this process across millions of fragments in a massively parallel fashion. [9]

SBS Chemistry

Sequencing by synthesis (SBS) technology uses four fluorescently labeled nucleotides terminal ended to sequence the tens of millions of clusters on the flow cell surface in parallel. During each sequencing cycle, a single labeled deoxynucleotide triphosphate (dNTP) is added to the nucleic acid chain. The nucleotide label serves as a terminator for polymerization, so after each dNTP incorporation, the fluorescent dye is imaged to identify the base and then enzymatically cleaved to allow incorporation of the next nucleotide. Since all four reversible terminator-bound dNTPs (A, C, T, G) are present as single, separate molecules, natural competition minimizes incorporation bias. Base calls are made

directly from signal intensity measurements during each cycle, which greatly reduces raw error rates compared to other technologies. The end result is highly accurate base-by-base sequencing that eliminates sequence-context specific errors, enabling robust base calling across the genome, including repetitive sequence regions and within homo polymers. SBS provides high accuracy, a high yield of error-free reads, and a high percentage of base calls above Q30.[9]

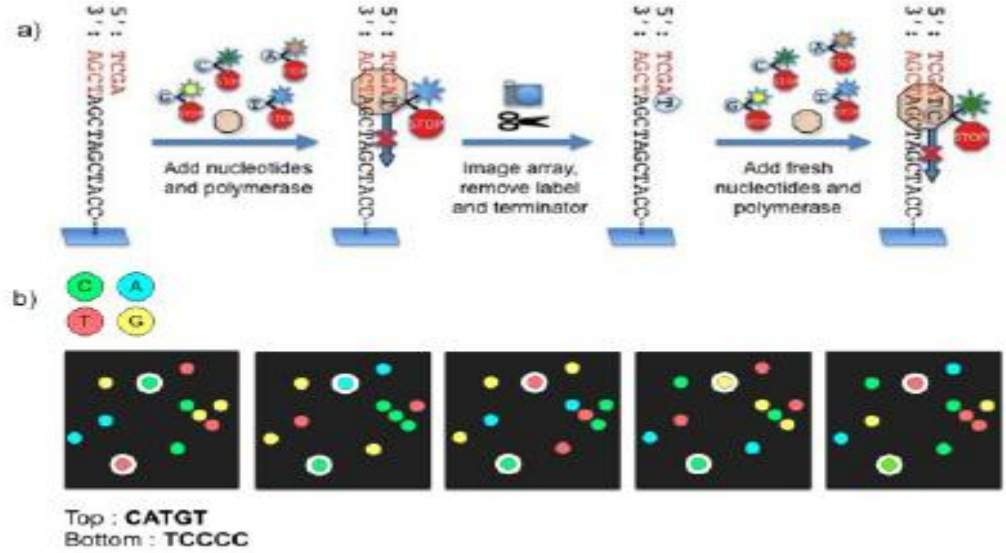

Figure 5. Sequencing by synthesis. (a) Illumina sequencing chemistry. (b) The sequencing cycles are repeated to determine the sequence of bases.

Work Flow for NGS

The NGS work flow includes the following steps:

1. **Library Preparation**— The library that has to be sequenced is prepared by fragmenting the DNA or cDNA sample. This is followed 5′and 3′ adapter ligation. Some kits facilitate “tagmentation” that means it combine sthe fragmentation and ligation reactions in a single step. This greatly increases the efficiency of the library preparation process. Adapter-ligated fragments are then PCR amplified and gel purified.

2. **Cluster Generation**—For cluster generation, the prepared library is loaded into a flow cell where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment is then amplified into distinct,

clonal clusters through bridge amplification. When cluster generation is complete, the templates are ready for sequencing.

3. **Sequencing**—Illumina SBS technology uses a proprietary reversible terminator–based method that detects single bases as they are incorporated into DNA template strands. As all four reversible terminator–bound dNTPs are present during each sequencing cycle, natural competition is minimized eliminating incorporation bias and greatly reducing raw error rates.

4. **Data Analysis**—During data analysis and alignment, the newly identified sequence reads are aligned to a reference genome. Following alignment, many variations of analysis are possible.[9]

Figure 6. Work Flow for NGS

1.2 Scope and Objectives

For our regions of interest which were the two kinase domains ALK domain and BCR-ABL domain we used amplicon sequencing technique.

This NGS technique was used as to identify all the possible regions of the kinase domain in ALK which are involved during rearrangements and to identify any mutations in the BCR-ABL fused domain which might be responsible for the drug resistance. Though both are kinase domains the library preparation work for both the gene panels were different.

1.2.1 Amplicon Sequencing

To facilitate rapid production of whole exome sequencing data, new amplicon based methods have been designed that simplify DNA preparation and utilize smaller inputs of DNA. To analyze the genetic variation in specific genomic regions amplicon sequencing can be performed. It is a highly targeted approach where the PCR products which are amplicon after ultra-deep sequencing allows efficient variant identification and characterization. This method uses oligo nucleotide probes designed to target and capture regions of interest, followed by next-generation sequencing (NGS). Amplicon sequencing is useful for the discovery of rare somatic mutations in complex samples such as tumors mixed with germ line DNA. Amplicon Sequencing enables to efficiently discover, validate, and screen genetic variants. It also supports multiplexing of hundreds to thousands of amplicons per reaction to achieve high coverage. It delivers highly targeted re-sequencing even in difficult-to-sequence areas, such as GC-rich regions. Most importantly it can reduce sequencing costs and turn around time when compared to broader approaches such as whole-genome sequencing.

Amplicon sequencing has been the method of choice in many high-throughput DNA sequencing applications. There has been great focus on the means by which to analyze the burgeoning amount of data afforded by high-throughput DNA sequencing.

In contrast, there has been a distinct lack of attention paid to considerations surrounding the importance of sample preparation and the fidelity of library generation. No amount of high-end bio informatics can compensate for poorly prepared samples and it is therefore imperative that careful attention is given to sample preparation and library generation within work flows, especially those involving multiple PCR steps.[11]

Applications of the technology, platform evaluations and bio informatic approaches to data analysis . While all three are extremely important in the generation of high fidelity data, a heavy focus on these aspects fails to address the need to pay close attention to the implementation of protocols and procedures at the bench. The data one has to work with is, and will only ever be, as good as the quality of experimental procedures implemented and no amount of high-end bio informatics can compensate for poorly prepared samples, artifacts or contamination. It is therefore imperative that careful consideration is given to the ways in which samples are screened for sequencing, in addition to the method used to generate the amplicon sequencing library.

1.3 Achievements

After months of amplification and library preparation with DNA samples for ALK kinase domain and cDNA samples for ABL1 kinase domain of BCR-ABL. We were able to amplify our regions of interest with which we were able to identify mutations and also look at fusions.

ALK is undergoing further validation to accumulate more amount of data so that ALK testing can be validated and performed on a molecular and genetic level. These tests in the future can accompany the conventional methods for better and accurate data.

As for the ABL kinase domain, it is also in a screening stage where we are trying to prepare more NGS libraries so as to detect the mutations in different samples. As Imatinib resistance has become associated with BCR-ABL positive samples it would be of great knowledge if the mutations causing resistance could be identified. This information could also help in identification of fusions that occurs in the BCR-ABL region causing those mutations. Therefore as more sample libraries are prepared, more data are produced. This would facilitate for analysis and for future purposes where NGS approach would provide a standard for testing.

The amplification at the beginning using the specific primers for the regions of interests in both ALK and BCR-ABL kinase domains was the first successful stage. Thereon, the libraries produced also showed appropriate sizes. The most important part that is the analysis, is still going on and the results once obtained will help us decide the next step.

A validation plan is being designed for ALK as we await results for BCR-ABL samples.

1.4 Overview of Dissertation

The ALK and BCR-ABL kinase domains NGS library had to be prepared for it to go for sequencing. Prior to library preparation, the received samples had to be amplified. For this, when the samples were first received the QC reports were obtained. After getting the QC reports the DNA quantity was verified. It is to note that DNA was taken for ALK positive samples, as for BCR-ABL positive samples cDNA was taken. This was done as in case of ALK we were looking for mutations so genomic DNA could be used but in case of BCR-ABL which is a fused gene, mutations were being looked at the fused region for which cDNA was appropriate.

As fusions are the result of translocation. These translocation can occur anywhere in the genome, including introns and other non-coding sequences. They can also occur within the coding regions of genes with limited expression patterns. This means that many of the translocation that occur in a cell may not be expressed and thus have little or no biological relevance. For this reason, DNA is not the ideal substrate to search for oncogenic fusions. RNA, on the other hand, is the intermediate product of gene expression and is ideal for detecting fusions, because then we are only looking at those that are expressed and potentially oncogenic.

The next step was to amplify the regions of interest in both the domains. However two different approaches were taken as to amplify the samples.

For the ALK positive samples we had designed primers that would target the exons which were our regions for interest.

As for BCR-ABL positive samples amplicons were prepared using nested PCR approach.

Sequence Name	Sequence
Exon 1F	5'-CC TCA TTA TTG TGG CCT GT-3'
Exon 1R	5'-AA GGG CAG GCT CAA GAG T-3'
Exon 2F	5'-AG TTC TCA GCT CAC AGC CT-3'
Exon 2R	5'-AC CCT CTC CAG GTT CTT TGG-3'
Exon 3F	5'-GA CTC AGC TCA GTT AAT TTT GG-3'
Exon 3R	5'-CA GCA AAG ACT GGT TCT CAC-3'
Exon 4F	5'-TT CCC TCC TCT CAC TGA CAA-3'
Exon 4F	5'-GA CAG GAA GAG CAC AGT CAC-3'
Exon 5F	5'-GC ATT TCC TTT CTT CCC AG-3'
Exon5R	5'-CA CAC CCC ATT CTT GAG G-3'
Exon 6F	5'-GC AGG GCA GAT GCT TAA T-3'
Exon 6R	5'-GG AGG ATG ATG GCT GAC TT-3'
Exon 7F	5'-GG TGT GTC TAT ATC CAT CTC CA-3'
Exon 7R	5'-AG CAT ATG TGG CTC TGG ATA-3'

Table 1. ALK kinase domain primers for 7 exons

After successful amplification of all the regions which were confirmed by running them on gel and checking their gel images for correct sizes.

The ALK samples were now pooled together and moved forward for library preparation.

For BCR-ABL kinase domain was amplified using a 2-step RT-PCR procedure. Complementary DNA was generated by reverse transcription followed by a first-step PCR reaction to isolate a 1.5-kilobase (kb) cDNA fragment, which included the BCR-ABL junction and the ABL kinase domain. The primers used for reverse transcription and first-step PCR were CM10 (5-GAAGCTTCTCCCTGACATCCGT-3) and 3 ABL (5-GCCAGGCTCTCGGGTGCAGTCC-3). A second-step PCR reaction was performed to isolate the ABL kinase domain using the primers 5 ABL (5-GCGCAACAAGCCCACTGTCTATGG-3) and 3 ABL positive samples the nested PCR products were used for library preparation.

The first step prior to library preparation is fragmentation but in this case as we were already working with amplicons, fragmentation was not required. The sizes were already small enough to be taken up for NGS. The library had to be prepared for its whole genome.

For the whole genome library preparation appropriate library preparation kit which are commercially available was used. The Library Prep Kit for Illumina contained all the enzymes and buffers required to convert a broad range of input amounts of DNA into high quality libraries for next-generation sequencing on the Illumina platform.

The first step was end repair which was done using end repair buffer and enzyme in a thermocycler for an hour. Next step was adaptor ligation which was performed with the adaptors provided in the kit along with some enzyme and buffer for about 30 minutes. Adaptors are the oligos bound to the 5′and 3′end of each DNA fragment in a sequencing library. The adaptors are complementary to the lawn of oligos present on the surface of Illumina sequencing flow cells.

Next size selection was done followed by a cleanup using Ampure XP beads. The final PCR was then kept. To which universal primers and indexes were added. This was followed by a final clean up using the same beads.

Finally the size distribution was checked on the Tape station or the Bio-analyzer. After checking for the results when the correct sizes were seen it was then sent for sequencing.

After sequencing it was sent to the bio-informatics team following which it would be sent to a team of genome analysts who would identify the variants and then prepare a report.

The report would then help us identify the variants and all other important factors as coverage and read on targets. This if the results are satisfactory the data is stored for future reference. These results give us a confidence on the fact that NGS gives satisfactory or even more better results when compared to the conventional methods.

The ALK samples in our case gave positive results, all our regions of interest could be amplified leading us to the conclusion that it worked. Currently, a validation plan is being set for ALK positive sample testing.

As for the first set of BCR-ABL samples it worked well. A second set of libraries were prepared for which the results are awaited. Once the results arrive we plan on working with one more set of samples so that we can gain enough confidence on the test.

2 Background and Rationale

The idea of personalized targeted therapy is a promising arena in any cancer treatment. The current attention is now focused on detection of oncogenic drivers responsible for tumor initiation and maintenance and development of drug resistance.[12]

The standardization and validation of ALK rearrangement detection methods is essential for accurate and reproducible results. The choice of diagnostic methodology for ALK rearrangement detection however remains a matter of debate. Each technique gives different quantities/volumes of diagnostic material and is associated with its associated risks and positive and negative predictive values.

I)Fluorescent in situ hybridization (FISH) assay has been considered the "gold standard" for ALK rearrangement testing. The ALK FISH assay utilizes break-apart probes to label the ALK rearrangement break point at the 3′ and the 5′ ends with orange/red and green fluorochromes, respectively. It has its own advantages and disadvantages:

a)The type of ALK rearrangement can produce false-negative and false-positive detection .

b)The inability of this method to identify specific translocation types.

c)Attention has to be paid with the nature of examined samples.

d)FISH is an expensive and time consuming technique requiring labor intensiveness and technical and interpretative expertise.

The usage of the FISH assay also remains advantageous over other assays like RT-PCR

as it has the ability to identify ALK fusions with variant partners and to determine gene copy number changes.

II)IHC is a clinically applicable test routinely used in diagnostic pathology labs. Currently a number of antibody are that can detect ALK protein expression in NSCLC using FFPE tissue. However, like other IHC detection assays, there are some technical challenges.

a)Antigen retrieval, affinity of the primary antibody, optimization of the IHC

procedures, protein degradation in paraffin blocks over time, and the relative low expression of ALK protein in ALK-rearranged NSCLC compared to the ALCL.

b)ALK IHC testing is yet to be standarized as a screening technique. The ALK1 IHC antibody was originally used for ALK detection in ALCL . Subsequently, it was determined that ALK-rearranged NSCLC exhibit lower ALK expression compared to ALCL .

c)The sensitivity of ALK1 antibody in IHC to detect ALK protein fusion was found to be lower compared to the FISH testing in some studies.

d)The need for positive and negative controls is essential for accurate assessment.

There remains much debate and controversy over test selection.

As a result, different PCR-based techniques have emerged to detect ALK rearrangement including next generation sequencing (NGS) technique and multiplex RT-PCR. In NGS, the genome can be assessed at different levels. Technically, NGS involves the DNA amplification utilizing DNA library template prepared from a patient's tumor sample followed by single run sequencing in a massively parallel fashion . Unlike multiplex RT-PCR, where at least two known sequences are targeted in one single reaction, the knowledge of recurrent genomic alterations is not required in NGS. Another important application of NGS technique is to assess the complex genomic alterations not only from tissue specimens but also the other types of samples. The application of NGS testing to liquid biopsy strategies provides the new opportunity for diagnosis and monitoring disease of patients with ALK fusion.[12]

Currently, the Sanger capillary sequencing technique analyzing BCR-ABL1 is considered the gold standard for mutation detection in a clinical laboratory . This assay has a sensitivity of 15–20 % (i.e., 15–20 mutated transcripts in 100 total BCR-ABL1 transcripts).

Next-generation sequencing (NGS) of an amplicon library improves sensitivity of mutation detection, enhances throughput and allows quantification of relative mutation burden. NGS is thus predicted to replace, in future, conventional Sanger sequencing for routine diagnostics. In CML and Philadelphia chromosome-positive (Ph+) acute lymphoblastic leukemia (ALL) patients who failed multiple lines of TKI therapy, NGS could identify mutations below the lower detection limits of Sanger

sequencing and uncover complex clonal textures in many cases. It is important to determine the sequencing depth and the lower detection limit below which true low-level BCR-ABL1 mutations/variants cannot be distinguished from the background error noise occurring. It is reported that sometimes mutations cannot be detected by Sanger sequencing during imatinib therapy and this later leads to development of resistance under second line or subsequent lines of treatment. The mechanisms underlying persistence of small numbers of malignant progenitors in imatinib-sensitive patients are unclear. The BCR-ABL mutations were associated with varying levels of imatinib resistance. BCR-ABL kinase domain mutations affecting drug binding can lead to secondary resistance to imatinib.[13]

The NGS coupled with a robust bio informatics approach for mutation calling would ensure reliable detection of emerging BCR-ABL mutations allowing therapy switching and selection of the most appropriate therapy.

3 What did you learn?

As mentioned in previous chapters, our main aim was to be able to use NGS technique for ALK rearrangement testing and BCR-ABL mutation detection. Several studies have been done in the same subject and with some reference we have also begun our work.

The successful completion to this work would help us establish a method of testing that involves NGS providing us with better and accurate results. There are some limitations to this work. We have to check the reproducibility, repeatibility and sensitivity of this method. Along with this we have to make sure about the expenses involved. Though NGS can prove to be more accurate it is also expensive. So, it is important to improve the cost effectiveness of the method. As we are still working on all these aspects it is important to optimize these tests for all kinds of samples. Also if there could be some inputs where we can reduce any amount of steps without hampering the results wherein, we would also be benefiting cost wise would be helpful.

3.1 Summary

The development of an optimization and validation plan for ALK rearrangements and BCR-ABL mutation detection tests are among several other panels we are working on.

We plan on optimizing the protocols following which the development of a validation plan would occur.

The project started with the amplification of our regions of interest using specific primers for each exon in case of ALK and using a nested PCR approach using two sets of primers in case of BCR-ABL.

In case of ALK as we were working with 7 different exons in a single sample. So, to amplify each exon at one go was a challenge. At first exon 4 did not show any amplification in the gel indicating that the primers were not working. After designing a new primer for that region we were able to amplify that particular exon to the PCR conditons we had optimized for the rest of the exons. We then tried amplifying a set

of samples for all 7 exons which included the new primer. This time all exons except exon 25 did not show any amplification. As exon 5 had earlier amplified and was showing negative results only for some samples we concluded that the primer was not a problem. We then tried to eliminate the manual errors which could have occured. We then increased the number of cycles in the PCR with a new set. It worked initially but then again we obtained negative results. Finally we used polymerase from a different kit and finally were able to amplify all the exons at a go. Thus we had optimized the PCR amplification for all the exons. After different sets of samples were amplified, library preparation was done and then sent for sequencing. Next, a validation plan was set using a designed cell-line named Kelly P4.

For BCR-ABL nested PCR was kept for the cDNA samples. Two sets of primers were used for reverse transcription and first-step PCR were

CM10 (5-GAAGCTTCTCCCTGACATCCGT-3) and,

3 ABL (5-GCCAGGCTCTCGGGTGCAGTCC-3).

A second-step PCR reaction was performed to isolate the ABL kinase domain using the primers

5 ABL (5-GCGCAACAAGCCCACTGTCTATGG-3) and,

3 ABL (5-GCCAGGCTCTCGGGTGCAGTCC-3) positive.

For the first set of IRMA positive samples we had to repeat the sets using two different amplification kits. The second kit gave good results and the samples amplified using this second kit were taken for library preparation. The libraries were then sent for sequencing. The results were positive for those samples. However we are yet to receive details from the bio informatics team.

3.2 Evaluation

We had placed a few objectives ahead for the optimization plan that had to occur. There were few objectives added as we began working on it.

- ✓ To begin with we had to achieve successful amplification of all the regions of interest in both the kinase domains.(ALK & BCR-ABL).
- ✓ To check that the newly designed primers were working and that the old ones had not degraded.

- ✓ To optimise the PCR reactions along with the kit.
- ✓ To optimise the library preparation protocol to obtain the required sizes.
- ✓ To prepare libraries of correct sizes.
- ✓ To send for sequencing on time.
- ✓ To follow-up with the results.
- ✓ To set up a validation plan.
- ✓ Prepare more libraries with different samples for data storage.

We were able to complete all our objectives in a sequential manner and within the time frame given to us for completion of the project. The success of the optimization plan will be evaluated once the results come and if positive will be sent for commercialization.

4 How to complement in corporate?

In science even a negative result is a result. This being said, in a corporate lab we need to learn fast and give results even faster. We cannot afford to make mistakes or take things lightly as even if we are dealing with project samples those are real patient samples. Any mistake or carelessness in a single person's work at any step will be affecting the final result altogether. Corporate world is a circle, mistakes are re-traceable. A single mistake can become a crime in the long run as we would be playing with a life. Even an unexplained delay can be thought as delaying a patient at risk of a possible treatment.

Everyday we go to our labs get new samples some for routine tests and some as a part of a new project for the lab to grow. But we treat each and every sample with the same amount of attention. Sometimes we face failure even in routine tests but we cannot afford to waste time on mourning about the time loss. Instead we trace our steps back to where possibly things might have gone wrong and be even more careful the next time as the sample amount is also limited.

A single fail of a batch of samples means loss. Loss in terms of money, energy and time.

We need to be careful about the instruments and following the SOP's. Every safety training is important, even disposals are a serious matter.

It is a fierce world and corporate world is no different. It is fierce, we have to create our own niches as we are replaceable. Even in this hectic time it is the people who we work with make the system go smoothly. We get the treatment we give others. So it is important to be professional yet respectful. Professionalism does not mean being rude or selfish. From what I have learnt it is being able to work in a team.

Personal growth is important but it is also important that the team grows so that the company's growth also happens in the long run.

References

- 1. Julie A.Vendrell . Et al , Detection of known and novel *ALK* fusion transcripts in lung cancer patients using next-generation sequencing approaches, *J. Scientific Reports,* 2;7(1):12510, Oct 2017.

2. Soverini S . Et all ,Unraveling the complexity of tyrosine kinase inhibitor-resistant populations by ultra-deep sequencing of the BCR-ABL kinase domain, *J. Blood ,*122(9):1634-48, Aug 2013.

3. Ryan Griffin and Robert A. Ramirez , Molecular Targets in Non–Small Cell Lung Cancer. J.Ochsner, 17(4): 388–392, 2017 Winter.

4. Lorena Lobo de Figueiredo-Pontes. Et al, Identification and characterization of ALK kinase splicing isoforms in non-small-cell lung cancer, *J. Thorac Oncol,* 9(2): 248–253,Feb 2014

5. Carminia Maria Della Corte et al, Role and targeting of anaplastic lymphoma kinase in cancer, *J .Molecular Cancer*,17:30, 2018

6. Rodney E. Shackelford et al, ALK-rearrangements and testing methods in non-small cell lung cancer, *J. Genes & Cancer*, Vol. 5 (1-2), January 2014

7. Andreas Burchert, Roots of imatinib resistance: A question of self-renewal? , *J. Science Direct,*152–161, June 2017

8. Wikipedia, https://en.wikipedia.org/wiki/Philadelphia_chromosome,

9. Stephanie Salesse & Catherine M Verfaillie, BCR/ABL: from molecular mechanisms of leukemia induction to treatment of chronic myelogenous leukemia, *J.Oncogene* , volume21 pages8547–8559, December 2002

10. Illumina, https://sapac.illumina.com/techniques/sequencing/dna-sequencing

11. Dáithí C. Murray et al, From Benchtop to Desktop: Important Considerations when Designing Amplicon Sequencing Workflows, *J. Plos* , 22;10(4):e0124671, Apr 2015

12. Xiaomin Niu et al, Anaplastic Lymphoma Kinase Testing: IHC vs. FISH vs. NGS, *J. Curr. Treat. Options in Oncol*, 18: 71 ,2017

13. Katerina Machova Polakova et al, Next-generation deep sequencing improves detection of BCR-ABL1 kinase domain mutations emerging under tyrosine kinase inhibitor treatment of chronic myeloid leukemia patients in chronic phase, *J.Cancer Res Clin Oncol,* 141:887–899, 2015