QATAR UNIVERSITY

COLLEGE OF HEALTH SCIENCES

PROFILING OF CHRONIC MYELOID NEOPLASMS THROUGH RNA SEQUENCING ANALYSIS

BY

KHOLOUD AHMED ALJASSIM

A Thesis Submitted to

The Faculty of the College of Health

Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

The Master of Sciences in

Biomedical Sciences

June 2018

COMMITTEE PAGE

The members of the Committee approve the Thesis of Kholoud Ahmed Aljassim defended on 10/06/2018

Dr. Pejman Hanifi Moghaddam
Thesis/Dissertation Supervisor

Dr. Nader Al-Dewik
Committee Member

Dr. Joel Malek
Committee Member

Dr. Mashael Al-Shafai
External examiner

Dr. Johan Ericsson
External examiner

Approved

Asma Al-Thani, Dean, College of Health Science

ABSTRACT

ALJASSIM, KHOLOUD AHMED., Master of Sciences: June: 2018, Biomedical Sciences.

Title: Profiling of Chronic Myeloid Neoplasms through RNA Sequencing Analysis.

Supervisor of Thesis: Dr. Pejman Hanifi Moghaddam.

Background: Chronic myeloid leukemia (CML) is a clonal myeloproliferative disorder resulting from the neoplastic transformation of hematopoietic stem cells. A higher number of (54%) CML patients in Qatar failed to respond to the treatment compared to those of (34%) internationally. Classification of the different stages of the disease is done in the clinic but is not sufficient; there is a great need to find markers that can be used to stratify the stages of the CML disease and furthermore monitor those that are undergoing treatment.

Aims and Objectives: For this study, we aimed to identify the disease-specific profile of genes (transcripts) by analyzing the RNA sequencing of patients, recruited from two different stages of CML disease i.e. Chronic Phase (CP) and Complete Remission (CR). Our objective was to see if we can identify a disease-specific group of transcripts.

Methods: A total of 16 subjects were recruited for the study, which included nine patients with chronic phase (CP), three with complete remission and four healthy individuals (Controls, (CNT). RNA extracted from frozen PBMCs or fresh blood was used for RNA sequencing and analysis using Hiseq Illumina 4000.

Results and Discussions: We identified a subset of 23 transcripts and classified them into four different groups. We discovered that eight transcripts: *MAZ_4, RIT1_3,*

BUB1_4, CFLAR_5, PARVG_6, SENP5_6, GATS_1, TAOK3_4 were significantly and differentially expressed in the chronic phase patients only compared to five transcripts that segregated the complete remission from the rest of the cohort study and These are *HNRNPA3_4, SLC4A7_3, TBC1D4_8, CTC1_1, GRAMD1A_6*. The following five transcripts: *U2SURP_5SEPT9_2, DPY19L3_8, CYTH4_1, PLXNB2_8* were differentially expressed in both groups of patients compared to control group. These that are shared between CP and CR group might indicate that study subjects in the complete remission did not fully recover and might need closer monitoring and to validate this current observation will be through a longitudinal follow up for some patients in complete remission using other diagnostic methods to examine the ratio of BCR-ABL1 fusion gene and the expression level of these shared transcripts between CP and CR groups.

Conclusion: This study provides sets of the promising genes (transcripts) that have the potential to be used to stratify CML patients into complete remission and chronic phase groups. These findings have the potential for following-up and to specifically determine the treatment dose. Furthermore, some of these differentially expressed genes (transcripts) might be a potential therapeutic target or could be used as prognostic biomarkers.

DEDICATION

This work is dedicated to my lovely Mother and father

who never stop giving of themselves in a countless manner and who taught me how to face challenges and risks with confidence

To all my family particularly my beloved brothers and sisters

To my friends who encourage and support me

ACKNOWLEDGEMENTS

I would like to thank the supervisor of this project Dr.Pejman Hanifi Moghaddam for his efforts and assistance.

Also, I would like to deliver a special thanks, gratitude, and appreciation to Dr. Nader Al-Dewik for his continuous support, motivation and recommendations since he spent a sufficient time with me for comprehensive reviewing and advising to end up with solid and meaningful outcomes of our project.

Moreover, I would like to extend my gratitude and appreciation to Mr.Mohammed Al-Sharshani for his assistance in covering the practical part of this project through gaining from his experience in processing the RNA extraction and purification in a professional way and for providing me a lot of learning materials and literature. Beside this, I would like to send a special thanks to the team of the molecular diagnostic laboratory at Hamad Medical Corporation-HMC for their collaborations from all aspects.

Besides I would like to express my full thanks and gratitude to the team of the Genome core at the medical college of Weil Cornell represented by Dr.Joel Malek for their excellence in performing the RNA-sequencing of my samples in an organized way.

My special thanks to the Dean of the college of health sciences at Qatar University Dr.Asma Ali Al Thani for her support with Dr.Fatiha and Dr.Shilu at Biomedical Research Center-BRC at QU.

Finally, I would like to thank my wonderful family especially my lovely mom, brothers, and sisters for their countless support to reach this successful and solid-base achievement.

Table of Contents

LIST of TABLES

LIST of FIGURES

LIST OF ABBREVIATIONS

CML	Chronic Myeloid Leukemia
TKIs	Tyrosine Kinase Inhibitors
IM	Imatinib Mesylate
HSCT	hematopoietic stem-cell transplantation
MPDs	Myeloproliferative disorders
MPNs	Myeloproliferative Neoplasms
CP	Chronic Phase
AP	Accelerated Phase
BC	Blast Crisis
CR	Complete remission
CNT	Control group
IRIS	International Randomized Study of Interferon and STI571
Ph-Chromosome	Philadelphia Chromosome
Ph+	Philadelphia Positive
KDa	The unit of protein molecular weight
CDC	Centers for Disease Control and Prevention
FISH analysis	Fluorescent In Situ Hybridization analysis
RT-PCR	Real Time-Polymerase Chain Reaction
ncRNA	Noncoding RNA
PI3-K	Phosphatidylinositol 3-kinase Pathway
LncRNA	Long noncoding RNA
AML	Acute Myelogenous Leukemia

MiRNA	MicroRNA
lincRNA	Long intergenic noncoding RNAs
ERK/MAPK	Extracellular signal-regulated kinase/mitogen-activated protein kinase pathways
XIST	X-inactive specific transcript
PCA3	Prostate cancer antigen 3
NGS	Next generation sequencing
Crk	Adaptor protein that assembles complexes of cellular proteins and facilitates the signal transmission
PBMCs	Peripheral blood mononuclear cells
eRNA	Enhancer RNA
PRC2	Polycomb repressive complex 2
c-Abl	Cytoplasmic Abl gene
miscRNA	miscellanies RNA
rRNA	ribosomal RNA
tRNA	transfer RNA
PSCD family	(pleckstrin homology, Sec7, and coiled-coil domains) family
DNA-PKcs	DNA-dependent protein kinase
TMM	trimmed mean of M values
Edger	Empirical Analysis of Digital Gene Expression
CPM	Count Per Milion
TPM	Transcripts Per Milion
RPKM	Read Per Kilobase Milion
FPKM	Fragments Per Kilobase Milion

NaN or +/- Infinity	Not A Number/ non-finite quantities
indel	Insertion-deletion
EM algorithm	Expectation-Maximization algorithm
A2RE	hnRNP A2 response element
CLK2 and CLK3	The cdc-like kinase CLKs; are an evolutionarily conserved group of dual specificity kinases belonging to cyclin-dependent kinases (CDKs). The CLK family consists of four isoforms namely CLK1, CLK2, CLK3 and CLK4

CHAPTER 1: INTRODUCTION

1.1 Background

According to the World Health Organization (WHO), Myeloproliferative Disorders (MPDs) can be classified into different groups including: chronic myeloid leukemia (CML), chronic neutrophilic leukemia, polycythemia vera (PV), primary myelofibrosis (PMF), essential thrombocythemia (ET), chronic eosinophilic leukemia, mastocytosis and unclassifiable MPDs. Among these hematopoietic neoplasms, CML is the only clonal myeloproliferative disease resulting by a reciprocal translocation between portions of chromosome 22 and chromosome 9 of the BCR and ABL genes respectively (Koptyra et al., 2006) This alteration creates BCR- ABL1 fusion oncogene with shortened chromosome 22 known as Philadelphia chromosome, the unique genetic abnormality of CML (Green, 2010) that is necessary for the manifestation of chronic myeloid leukemia (Herrmann et al., 2014).

The resulted in BCR- ABL1 fusion transcripts are different in sizes and this based on the breakpoint regions in both BCR and ABL1 genes. In BCR genes the major breakpoint cluster regions involving exons (e12- e16) which commonly referred to as (b1-b5), while the breakpoints in ABL gene usually takes place between exons a1 and a2 as a result of alternative splicing. It is found that the fusion transcripts generated with e13a2 (b2a2) and e14a2 (b3a2) junctions are frequently encode for 210 kDa protein known as $p210^{BCR/ABL}$, in which strongly correlated with an active tyrosine kinase observed in initially diagnosed patients with chronic phase of CML as this related to the disease exacerbation (Jain et al., 2016)The size of the resulting BCR-ABL1 transcript encoding 210 kDa protein is larger than the individual ABL1 (145 kDa) and BCR (160 kDa) proteins (Deininger, Goldman, & Melo, 2000) in which the

tyrosine kinase is constitutively active in the resulted BCR-ABL1 fusion protein sized 210 kDa.

Normally, Tyrosine kinases are an important family of enzymes which act as a key mediator in different cellular signaling transduction pathways through phosphorylation of tyrosine residues of the target receptor in response to external and internal stimuli leading to cell growth, differentiation, metabolism, and apoptosis. These roles of Tyrosine kinases are essential for normal cell messages and preservation of cellular homeostasis. However, Tyrosine kinases have been implicated in cancer and acquired mutation which makes these enzymes constitutively active and phosphorylate the downstream cascade and developing neoplastic hallmarks in cancerous cells (Paul & Mukhopadhyay, 2004) It has been shown that the $p210^{BCR/ABL}$ protein is involved in the pathogenesis of human CML by introducing $p210^{BCR/ABL}$ retrovirus protein into the mice bone marrow stem cells, which resulted in subsequent development of a myeloproliferative disease with the CML pathological and clinical features (Di Bacco, Keeshan, McKenna, & Cotter, 2000) Another investigation confirmed the key role of $p210^{BCR/ABL}$ production in CML pathogenesis through identification of particular coding sequence in CML-mRNA when researchers using K562, a human leukemic cell line which representing an earlier stage of granulocyte maturation in their study, thus enable the investigator to measure the $p210^{BCR/ABL}$ activity in vitro in different cellular contexts (Di Bacco et al., 2000) Myeloproliferative disorders (Tefferi & Vardiman), recently known as Myeloproliferative Neoplasms (Klco, Vij, Kreisel, Hassan, & Frater) are heterogeneous group of hematopoietic disease which characterized by excessive clonal production of one or more hematopoietic cell lineages in the bone marrow which might be progressed to acute conditions.

In normal hematopoetic cells, the proliferation, differentiation and survival of these cells are controlled by hematopoietic growth factors including; interleukin 3 (IL-3), stem cell factor, erythropoietin, interleukin 6 (IL-6), thrombopoietin, and colony stimulating factor 2 (CSF-2), which are responsible for activating signaling pathways when adhering to cell surface receptors and resulting in phosphorylation of downstream cascade. However, BCR-ABL1 oncogene displaying aberrant role in cancerous cells by achieving the growth factor autonomy, which in turn impact the capacity of leukemic cells to adhere and invade over continuous stimulating signaling pathways, hence the potential function of this oncogene is the activation of cytokine signal transduction pathway through cytokine recruitment which in turn recruiting kinases proteins in signal transduction pathway to eventually reaching the transcription factors which affect different cellular processes. Moreover, BCR-ABL1 tyrosine kinase causes abnormal expressions of cell cycle control genes through the inhibition of the normal role of c-ABL gene including the regulation of the gene transcription, integrin signaling, and cell cycle. The high activity of BCR-ABL1 tyrosine kinase which demonstrated mainly in the cytoplasm will accelerate the activity of the machinery genes of the cell cycle including Cyclins D2- and D3-associated Cyclin-Dependent Kinase 4 (CdK4) and Cyclin-Dependent Kinase 6 (Cdk6) which allows the progression of the cell cycle phases from G1 to S. This will phosphorylate Rb1 suppressor gene followed by an increase in transcription factors which irreversibly inhibit the activation of Rb1 gene and allows the cell cycle progression to S phase. On the other hand, BCR-ABL1 fusion gene can cause the delay in the cell cycle transition from G2 to mitotic checkpoint (M) which necessary for the DNA repair lead to a failure in apoptosis (Di Bacco et al., 2000).

Currently, many clinical and laboratory findings propose that the abundant existence of Philadelphia chromosome (Ph) in patients sample is strictly correlated with poor prognosis and treatment failure since the BCR-ABL1 oncogene with it is malignant behavior is capable to induce genomic instability leading to multiple mutations which observed in different phases of CML disease as these genetic alterations encoded by BCR-ABL1 genes making cells resistant to therapies that target mutated tyrosine kinase domain. These drugs are known as Tyrosine Kinase Inhibitors (TKIs) in which Imatinib Mesylate (IM), the first drug was discovered for this purpose. It is observed that only 4% of newly diagnosed CML-chronic phase patients per year have a therapeutic resistance. The reason for the treatment failure has been reported to be BCR-ABL1 kinase domain that generates reactive oxygen species (ROS) from mitochondrial respiratory chain (MRC) (Flis, Irvine, Copland, Bhatia, & Skorski, 2012) since BCR-ABL1 oncogenic tyrosine kinase has two major roles in leukemia, the first one is the constitutive activation of the downstream signaling pathways with stimulation of the growth factor independence. The second role of BCR-ABL1 fusion gene is the induction of genomic instability such as chromosomal translocation and mutations which resulted in DNA repair failure leading to DNA damage. In leukemic cells, the increased in oncogene activity with genomic abnormality correlated with the generation of reactive oxygen species (ROS) which considered as an inevitable byproduct of oxidative reduction process in mitochondria. These oxidative molecules combined with the dysregulated pathway of DNA repair causing an accumulation of oxidative DNA damage which in turn cause a resistance to IM (Koptyra et al., 2006).

The function of $p145^{ABL}$ protein, which is the tyrosine kinase (TK) and the product of Abelson gene is controlling of the cell growth, induction of apoptosis and DNA repair. Although $P145^{ABL}$ protein located in the nucleus, some studies revealed its role

in cell signaling for integrin and other cells receptors such as B-cell receptor (eg. CD19). Tyrosine kinase domain in the c-ABL gene is regulated by SH3 and SH2 regions, in which SH3 is a negative regulator of c-ABL activity. Any mutation acquired to SH3 region will constitutively activate the kinase domain, whereas SH2 is the positive regulator (Di Bacco et al., 2000; Shet, Jahagirdar, & Verfaillie, 2002).

CML is not like other cancers. CML is a disease of the bone marrow in which disease phases are assigned based on specific factors such as the patient's age, immature white blood cells count, and the spleen enlargement. CML can be divided into four clinical distinct phases, an initial chronic phase (CP) which characterized by the presence of Philadelphia-chromosome (Ph+) with BCR and c-ABL fusion gene in which cells are responsive to growth factors and reduced apoptosis, show altered adhesion and activation of different hematopoietic signaling pathways and the disease can be easily controlled in this phase. This phase usually associated with a median survival of 4 to 5 years. However, some patients have treatment failure that can undergo to accelerated phase (AP) which last for 4 to 6 months with difficulty in controlling the blood count by myelosuppressive treatment. CML-AP almost constantly progress to Blast Crisis (BC) phase if not treated and can be differentiated from other phases by other molecular changes such as the presence of an extra copy of Ph-chromosome (BCR-ABL1) and high rate of *P53* and Rb1 gene mutation in which these genes are almost implicated in the transformation. Alterations in P53 and Rb1 genes usually occur in 20% to 30% of blast crises cases but their structure remained normal in chronic phase (Di Bacco et al., 2000).

Chronic Myeloid leukemia can be implicated at any age group, however, it accounts for 15 % of adult leukemia with the median age of onset being 67 years" (O'Brien et al., 2012).

Treatment failure with Imatinib Mesylate (IM) reached a high percentage of 54% of CML patients in Qatar where the resistance rate is high compared to that reported by International Randomized Study of Interferon and STI571 (IM) (IRIS) which reaches to 35% A study conducted in Qatar in 2014 concluded that 14% of the treatment failure caused by point mutation with unique tri-nucleotide insertions and the most common cause of IM failure was documented in 50% of cases due to additional chromosomal abnormalities (ACAs) , 28% of cases showed drug resistance for unknown reasons and 14% of patients stopped IM due to intolerance (Al-Dewik, Jewell, Yassin, El-Ayoubi, & Morsi, 2014).

Currently, most of CML patients are effectively treated using both the first and second generation tyrosine kinase inhibitors (TKIs) including IM, Nilotinib, and Dasantinib respectively as therapeutic agents in which IM induces high rate of molecular response and improved overall survival rate and has well-documented safety profile while Nilotinib and Dasantinib cause rapid molecular response and characterized with a reduced risk of the disease transformation to blast crisis Although, all three drugs are used as the first line therapeutic option and they are well-correlated with the time-dependent milestone molecular response of BCR-ABL1 gene after 3, 6 and 12 months, challenges still remain when using these drugs to treat the disease in some patients such as several genetic and epigenetic expression mysteries of CML, no accurate algorithm as no specific biomarkers can be used to identify the risk of the disease progression at diagnosis. In addition to that, the clinical scoring system used at the time of CML diagnoses such as Sokal, Hasford and European Long-Term survival (ELTS) risk scores which can provide a significant prognostic information useful to determine the disease risk, these scoring systems themselves are with limited specificity and sensitivity. Therefore, by investigating novel biomarkers will improve

the prediction of disease progression and enable the selection of the first- line treatment and enhance the optimal responses to TKI therapy (Marum et al., 2017).

Based on the knowledge of central dogma concept by a molecular biologist that the genetic information located only in protein-coding regions of the genome and they believed that any transcript originates from non-coding regions have been considered as a transcriptional noise.

This view persists for a long time period until many evidence and research change their old perspective. It is well-known that cancer is a multistep genetic disease which can be caused by the mutation in protein coding region but several studies indicate that mutations also can be found in non-protein coding regions as the protein-coding gene account for 2% which means that more than 98% of the human genome are non-protein coding. These regions are transcribed into non-coding RNA which can be classified based on the transcript length into small non-coding RNA and long non-coding RNA (lncRNA) (X. Chen, Yan, Zhang, & You, 2017).

Increasing evidence is indicating that lncRNAs are considered as regulators of almost every cellular process, and expression of these non-coding molecules appears to be firmly regulated in physiological conditions while deregulated expression of these molecules drives toward several human diseases phenotypes including cancer (Beermann, Piccoli, Viereck, & Thum, 2016).

A list of lncRNAs discovered by some studies showed that many of which are abundantly expressed in differentiated tissues of specific cancer and they are related to specific cell type more than the mRNA. In the context of the gene sequence, It is known that many functional protein defects resulted by genetic mutations but lncRNAs biology has remained impervious for understanding how sequence affects function in these molecules. Recently, many studies revealed that the molecular

mechanisms of lncRNAs serve to understand the role of these molecules as signals of specific cellular status, prognostic biomarkers and could be exploited as therapeutic targets of cancer. Therefore, emerged technology such as Next Generation Sequencing (NGS) enables the researchers to expand their investigation to identify cancer-associated lncRNAs (Schmitt & Chang, 2016).

During the last decades, multiple studies identified the relevance of non-coding RNAs in gene regulation. However, dysregulation of these molecules was increasingly linked to diverse human diseases and among these dysregulated transcripts of lncRNA is H19, but the profiling of protein-coding genes (mRNAs) and non-protein coding genes particularly, lncRNAs expression in CML patients using high throughput method such as NGS will be the first study of its kind in Qatar which assists the CML patients to overcome the disease complications and treatment failure.

1.2 Aim of the study

The aim of this study is to explore and identify the profile of transcripts expression in different phases of Chronic Myeloid Leukemia (CML) using RNA-sequencing technology which considered as a highly dynamic method with the capability of detecting novel genes (transcripts) from different aspects.

1.3 Research question

Can we use a transcript profile as a tool to stratify the different clinical stages of Chronic Myeloid Leukemia (CML) disease including chronic phase (CP) and complete remission (CR) in Qatar.

CHAPTER 2: LITERATURE REVIEW

2.1. Chronic Myeloid Neoplasm

2.1.1. Definition and Molecular Mechanism of Chronic Myelogenous Leukemia

CML is the most common clonal MPNs, which was first described by Virchow and Benett in 1845 as a case of hypertrophy of the spleen and liver, in which death took place from suppuration of the blood. (Klco et al., 2010) In its initial phase, CML is characterized by unclear symptoms such as abdominal discomfort and weight loss, among others. CML can be identified by an expansion of the cells of the granulocytic lineage. Initially referred to as MPNs, chronic MPDs refer to a class of myeloid hematological malignancies, which occur mainly among the adults (Klco et al., 2010) The disorders have a common characteristic whereby one or more of hematopoietic cell lineages highly extended in the bone marrow. The naming and categorization have been revised by the World Health Organization to emphasize on the stem-cell derived heritage of the disorders.

The chronic phase of CML is characterized by a larger percentage of mature cells, as well as myeloid progenitor cells in the extramedullary tissues and in the blood followed by a maturation arrest in lymphoid or myeloid lineage whereby it transforms within a period of 3-4 years (Shet et al., 2002). The malignancy is also associated with Philadelphia chromosome which is a genetic abnormality resulting from BCR and Abelson (ABL) genes rearrangement.

A study demonstrated that more than 90% of the CML patients have Philadelphia chromosomes. The BCR-ABL1 fusion gene is usually translated into an oncoprotein, $P210^{BCR/ABL}$ which facilitates the malignant transformation of the cells (Shet et al., 2002). However, despite the strides have been undertaken in studying the progression

of the disorder, the process through which CML changes from its chronic form to the blast crisis remained unknown.

In their study, Deininger, Goldman, and Melo (2000) provide a review of the physiologic functions of the translocation partner which results in the development of CML. ABL gene is one of the translocation factors; is a human homolog of the v-abl oncogenes which encodes tyrosine kinase. Human ABL is a form of protein that has two isoforms resulting from splicing of the first exon (Deininger et al., 2000) The protein has several domains including SH1 which is responsible for carrying the tyrosine kinase function. On the other hand, SH2 and SH3 are essential in allowing ABL protein to interact with other proteins. For example, the center of the protein molecule which is rich in Proline interacts with the SH3 domains of proteins like Crk (Deininger et al., 2000). Therefore, in its usual form, ABL is an essential protein in regulating the cell cycle, integrin signaling whereby information is transmitted within the cellular environment as well as providing the cellular response to genotoxic stress (Deininger et al., 2000) On the other hand, BCR is a 160kDa protein which is similar to ABL in terms of expression. The protein has several structural functions. For instance, the N-terminal exon is responsible for encoding serine-threonine kinase. Also, the N-terminus of the protein has a coiled domain for dimer formation (Deininger et al., 2000). The C-terminus has GTPase activity which is essential in regulating actin polymerization.

In order to understand the molecular mechanism of CML, it is important to review the biology of Philadelphia chromosome (Ph+) Ph+ refers to a genetic disorder due to a short chromosome 22. In addition, there is a translocation of genetic material between this chromosome and chromosome 9 (Shet et al., 2002) Thus, c-ABL, proto-oncogenes are transported from chromosome 9 to M-BCR location on chromosome

22. This forms the BCR-ABL fusion gene that encodes $P210^{BCR/ABL}$. The new oncoprotein is associated with abnormal activities in the body including an increase in the rate of binding to the actin skeleton. Also, $P210^{BCR/ABL}$ has an increased tyrosine kinase activity (Shet et al., 2002) Consequently, the abnormal cellular activities contribute to the growth of leukemia cells. As the cells proliferate, coupled with the transgenic expression of $P210^{BCR/ ABL}$, leukemia, lymphomas and CML syndromes are induced. The breakpoints at 9q34 region of ABL gene at chromosome 9 can occur anywhere either upstream toward the first exon of 1b or downstream toward the second exon of 1a or the breakpoints could occur between the two exons in that region. Contrastly, the breakpoints within BCR gene determined by 1 of 3 breakpoint cluster regions (BCR). It is demonstrated that in most of the CML patients, the breakpoint of the BCR gene occurs within a 5.8kb area surrounding BCR gene exons 12 to 16, which originally known as exons b1 to b5 while the breakpoint of BCR gene at this particular region defined as Major breakpoint cluster (M-bcr) (Deininger et al., 2000).

New therapies aimed at arresting CML during the chromosomal translocation phase have been developed. The targeted therapy for Ph chromosome in CML through the synthesis of IM has been used for quite some time. However, the efficiency of this treatment has been put to question because of the persistence and resistance of the disease. Consequently, about 50 mutations are known to have developed from Ph-CML resistance to the drug (Al-Dewik et al., 2014) Thus, it is vital to understand all the dynamics of CML including molecular reaction to various interventions to prevent the development of further mutations in future.

2.1.2. Incidence and Mortality

The incidence rate of the CML is on the rise. According to the American Cancer Society, more than 10% of the current cases of leukemia being diagnosed are chronic myeloid leukemia. The age average at the time of CML diagnosis is 64 years old and is rare among the children. However, it is slightly more prevalent in male than female as demonstrated in figure 1 (Höglund, Sandin, & Simonsson, 2015).

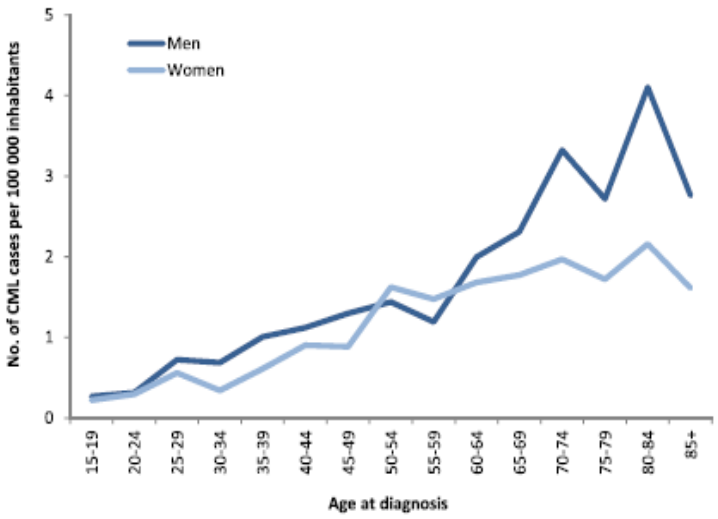


Figure 1. *The yearly incidence of CML per 100, 000 people of different age and gender groups.*

The data also suggests that 1 in every 526 persons in the United States will be diagnosed with CML in their lifetime. The American Cancer Society further indicates that before the end of 2017, more than 8, 000 cases of CML will be diagnosed in the United States with more than 1,000 people dying as a result of CML has a 5-year survival rate depending on the phase of the disease among other factors such as the

biological characteristics and presentation as well as the patient's response to treatment ("Key Statistics for Chronic Myeloid Leukemia," 2018).

The advancement in scientific research has contributed towards the development of sound treatment options which are likely to increase the survival rate. The most available data shows that the rate has doubled from 31% to 66% from the 1990s to 2012. A study conducted in 2014 shows that CML patients constantly taking IM drug and undergoing stem cell transplantation have a high survival rate of 93%. Besides, more than 90% of the patients live for more than five years (Sacha, 2014).

Singh, Henley, and Ryerson (2017) provide an analysis of the cancer incidents and mortality for the year 2013. In addition, the study shows the trends for the years 1999 to 2013 as per the data from the Centers for Disease Control and Prevention (CDC) and the National Cancer Institute (NCI). CML is among the cancers with a high mortality rate in the United States. Since cancers are reported, all states and health organizations, pathology laboratories, and physician offices are required to submit data for the reportable cancer diagnosis to the central cancer registry calculation of incident and mortality rate (Singh, 2017) Part of the reason why it has not been possible to provide accurate and comprehensive incident and mortality rates for most of the cancers is due to incompetent cancer reporting in some states (Jemal, Siegel, Xu, & Ward, 2010) Leukemia is the most common type of cancer which accounts for most of the expected new cancer cases at 43,050 and the estimated number of deaths at 21,840 for both gender in 2010 in the United States. The use of IM as the primary therapy of CML has reduced both the incident and mortality rate significantly ("CHRONIC MYELOGENOUS LEUKEMIA," 2014) Consequently, the treatment has provided a global health effort which is critical in addressing the burden of non-communicable diseases. Another reason the increased use and efficiency of IM is its

low levels of toxicities giving a major therapeutic benefit in both the adult and pediatric population ("CHRONIC MYELOGENOUS LEUKEMIA," 2014). Alternative drugs which are essential in reducing the mortality rate of CML include Nilotinib and Dasatinib. According to the Union for International Cancer Control (2014), CML in the United States has an incidence rate of approximately 1.6/100,000 In addition, the available data has not shown any association between CML and race or ethnicity. Despite the availability of accurate global data on CML, the disorder affects approximately 100,000 people worldwide annually ("CHRONIC MYELOGENOUS LEUKEMIA," 2014). Currently, the prevalence of CML is much higher since treatment with IM prolongs remissions in most of the patients. For instance, in 2010, 4870 cases of CML were diagnosed in the United States with 440 dying of the disease (O'Brien et al., 2012) In 2017, the estimated number of new cases is 8,430 and 1,090 deaths ("Key Statistics for Chronic Myeloid Leukemia," 2018).

Hoglund, Sandin, and Simonsson (2015) explored in their study the epidemiology of CML and the underlying risk factors for developing the disease. The study results were a complement to Union for International Cancer Control (2014) findings that the prevalence of CML is increasing progressively as a result of improvement in therapy regimens such as IM among other tyrosine kinase inhibitors. Another reason suggested to explain the phenomenon is as a result of the general improvement in life expectancy across the population (Höglund et al., 2015) A section of studies and data collected over the years indicated that the prevalence of CML in 1998, 2003 and 2007 was 5.8, 6.8 and 7.3 per 100,000 people respectively (Höglund et al., 2015) As more effective therapeutic regimens are developed, the prevalence of CML is expected to increase as the survival rate increases. For instance, Hoglund, Sandin, and Simonsson

(2015) estimated that the prevalence of CML in the United States will accelerate by more than 35 times the current and the future incidence by 2050.

Although the essential etiology of CML remains unknown, factors such as ionizing radiation have been identified as the key link to CML. For many years, smoking has been regarded as a risk factor in the development of CML. However, recent case-control studies have suggested a weak association between the two variables (Höglund et al., 2015) The introduction of tyrosine kinase inhibitors has paved the way for better understanding of CML especially on progression as well as the survival rates. Most studies confirm that availability of treatment regimen has increased the survival rates of CML dramatically (Höglund et al., 2015) In a study carried out by Kantarjian et al. (2012) to examine the survival rate of CML since the introduction of IM. The medical records for 1,569 of CML patients who had been referred to a health facility within one month of diagnosis between 1965 and 2011 were examined. Out of the 1,569 cases, 1,148 were in the chronic phase, 175 were in the accelerate phase while the remaining 246 were in the blast crisis phase (H. Kantarjian et al., 2012). On average, the study suggested an improved survival rate to about 71% among patients treated with IM since 2001. Among the CML patients in the blast crisis phase, the survival rate also improved with the introduction of IM but has remained seven months since 2001 (H. Kantarjian et al., 2012) Survival has been more significant among the patients in the chronic phase and the accelerated phase.

2.1.3. Therapeutic Options and Treatment Algorithms

CML treatment and management has evolved over the last few years from hematopoietic stem cell transplantation (HSCT) to the current standard regimen. Introduction of tyrosine kinase inhibitors (TKIs), Imatinib Mesylate (IM) has revolutionized the treatment for this form of leukemia. However, with its continued

use, resistances to TKIs occur. IM is on the rise resulting in over 50 different types of mutations. As a result, there is a need to develop alternative treatment modalities for CML patients. Improvement of scientific research has paved the way for better understanding of the molecular biology and progression of CML. Consequently, this has enhanced the efforts to develop therapeutic options to manage the condition. The current treatment modalities are designed to inhibit the activity of tyrosine kinase in BCR-ABL1 (Forrest, Jiang, Eaves, & Smith, 2008).

TKIs (Imatinib, Dasatinib, Nilotinib) have contributed to a milestone in the management of CML. According to Forrest, Jiang, Eaves, and Smith (2008), before the development of these drugs, the use of interferon alpha was the standard treatment. Interferon was given together with cytarabine. Other regimens used include busulfan and hydroxyurea. However, interferon alpha was commonly used because it prolongs the survival rate compared to the alternatives. However, this therapeutic tool had major disadvantages. For example, it is known to produce dose-related toxicities. In addition, the drug's survival benefit was realized in a small number of patients. Unlike TKIs, interferon alpha has major side effects, and patient monitoring is required in order to reduce the dose or discontinue the treatment (Forrest et al., 2008)

Imatinib Mesylate (IM) is the first TKI developed. The drug is useful in treating all the phases of CML. The drug was approved after a clinical trial involving 1106 patients who were newly diagnosed with CML in chronic phase. Follow-up studies on the patients after 60 months showed that the drug was significant both in increasing survival rate and promoting cytogenetic response at the rates of 93% and 87% respectively (Forrest et al., 2008) For newly diagnosed individuals in the chronic phase, a starting dose of 400 mg is recommended on a daily basis. On the other hand, CML patients with additional disorders including cytogenetic abnormality and the

presence of Ph-chromosome require an elevated dose of about 600 mg per day. Patients under IM therapy should be monitored regularly in order to guide on alternative or complementary procedures.

A study carried out by Forrest et al., (2008) stated that the only known curative procedure for CML is HSCT. The risk of relapse following HSCT is 14% (Rizzieri & Moore, 2012) It was a standard treatment for patients who both eligible and had matching donors. In addition, the survival rate is 61%. However, the system has been improved through the introduction of scoring systems which are essential in determining mortality following the transplantation. Since IM is used as the standard treatment for almost all patients, HSCT is only reserved for patients who do not show significant improvement following IM therapy (Jabbour, Fullmer, Cortés, & Kantarjian, 2010).

Rizzieri and Moore (2012) also mentioned the importance of adhering to the developed treatment algorithm especially in decision-making, promoting cost-savings, improving resource management and minimizing practice variation. These guidelines have been used extensively in guiding CML management. Currently, the European Leukemia Net (ELN) and the National Comprehensive Cancer Network (NCCN) are in the process of developing guidelines to target the arrest mutations in the disease. For instance, IM therapy should be stopped if the patient shows a loss of response to the drug, has progressed to the accelerated phase or the blast crisis phase or if the patient exhibits an unsatisfactory initial response (Rizzieri & Moore, 2012) Once the patient commences on the therapy, peripheral blood counts should be monitored. Once they normalize, additional hematological responses should be examined to guide on further clinical decisions (H. M. Kantarjian, Larson, Cortés, Deering, & Mauro, 2013).

IM therapy requires frequent monitoring for resistance. Ideally, a complete hematologic response (CHR) occurs after about three months of therapy. In addition, cytogenetic or FISH analysis is done on bone marrow samples. According to Forrest, Jiang, Eaves, and Smith (2008), drug monitoring in the course of the treatment should be based on Real-Time PCR (RT-PCR) as opposed to the use of cytogenetic analysis which has low sensitivity despite being labor intensive. The cytogenetic analysis also requires a serial bone marrow aspirate which is both painful and tedious to collect. RT-PCR, on the other hand, is more comfortable to conduct because it used peripheral blood and is also more sensitive and by using RT-PCR it is possible to identify a single abnormal Ph+ cell in a group of 10,000 normal cells. Besides, the method allows for detection of the minimal residual disease (Forrest et al., 2008).

2.2. Coding and Non-Coding RNA

Many studies showed that RNA molecules can be classified based on their derivative regions into protein-coding and non-protein coding transcripts. It is known that the majority of well-identified transcripts belong to protein-coding regions of the genome and accounts for 2.3% such as mRNAs. However, few of well-characterized non-protein coding (ncRNAs) transcripts represent the majority of the transcription and annotated as “dark matter” once discovered in the last century. These transcripts including ribosomal (rRNAs), transfer (tRNAs), small nuclear RNA (snRNAs), small nucleolar RNA (snoRNAs) and spliceosomal RNAs (Dinger, Pang, Mercer, & Mattick, 2008) Other categories of ncRNAs classified based on their size into Micro RNA (miRNAs) and long non-coding RNA (lncRNAs). These classes of ncRNAs have a regulatory role and the dysregulated functions of any of these ncRNAs found to be implicated in many diseases especially in cancers. These dysregulated transcripts could be either oncogenes or tumor suppressor genes. mRNA of the CML

has been reviewed extensively in different studies while the identification of ncRNA transcripts in CML have been revealed by limited studies using RNA-sequencing technology particularly in the state of Qatar.

2.2.1. Genes Involved in CML Pathogenesis and Clinical Identification

Progression of the CML is an essential part of the disease pathogenesis. The BCR-ABL1 fusion gene is critical to the pathogenesis of CML. In addition, the gene continues to undergo expression for the cell to proliferate especially in the acute phase. Although the process of evolution to blast crisis is not yet clearly understood, studies have shown that CML cells tend to develop additional molecular defects before the blast crisis phase (Shet et al., 2002) the first cellular event in the progression of CML is decreased apoptosis whereby cells degenerate on their own after sometimes for regeneration of new ones. It is a protective mechanism especially against cancerous cells or cells that acquiring chromosomal rearrangement, deletion, and mutation are critical for continued proliferation of the malignant cells. There are multiple mechanisms which contribute towards reduced apoptosis. One of the ways BCR-ABL1 exerts the antiapoptotic effect is through the phosphorylation and activation of the PI(3)K/Akt pathway (Shet et al., 2002) Another way in which apoptosis is slowed down is through PI(3) kinase activation. BCR-ABL1 fusion gene also increased expression of BCL-2 which contributes to the antiapoptotic effect. Additionally, factors which play antiapoptotic roles such as transcription factor NF-κB are activated in CML. Activation of NF- κB has been demonstrated in transgenic models. Other pathways that inhibit apoptosis include the RAS pathway which elevates the production of cytokines. The level of apoptosis regulation differs depending on the phase of CML. In the chronic phase of CML, there is minimal protection against apoptosis. However, there is evidence of some level of BCR-ABL1

mediated resistance to apoptosis. Lastly, the high rate of BCR-ABL1 mRNA expression is the key to the process of pathogenesis.

Another cellular event in CML progression is blocking the cellular differentiation phase (Shet et al., 2002) during hematopoietic differentiation, the pluripotent stem cells change both morphologically and functionally to distinct blood cells. $P210^{BCR/ABL}$ does not play a direct role in the terminal cell differentiation. However, additional factors with elevated cytokine levels, the formation of new oncogenes and the inactivation of the tumor suppressor genes block the normal cell differentiation (Shet et al., 2002) This phase of disease progression has been useful in the development of treatment regimen against CML. Studies showed that by inducing differentiation with cytokines, normal hematopoiesis is regulated.

During the chronic and the progressive phase of CML, both Major Histocompatibility Complex (MHC) -restricted and (MHC)-unrestricted are critical especially in providing natural control of the (Ph) clone. These mechanisms restrict the regular protective role of the T lymphocytes in eliminating the malignant CML cells. Through the expression of functional Fas-Ligand, CML progenitor cells take the role of cell defense by creating an immune surveillance escape factor. In blast crisis CML, the normal functions of Natural killer (NK) cells and activated killer (AC) cells decline significantly, and this is thought a factor contributing to disease progression (Siegel, Naishadham, & Jemal, 2013).

Another factor which facilitates progression of CML to the blast crisis is the ability of the BCR-ABL1 mRNA to activate several oncogenic pathways. The presence of the gene is also significant in amplifying the activity of BCR-ABL1.

The amplification of BCR-ABL1 gene in vitro increases resistance to IM. The studies confirm the phenomenon that patients in the blast crisis fail to respond to IM. In the

blast crisis, there is additional translocation which is also in high frequency. As a result, Chimeric AML/EVI-1 fusion protein is formed. This protein plays a key role in preventing both cell growth and differentiation. Another significant finding in patient in the final phase of the disease is the homozygous deletion of $p16^{INK4a}$ in more than 50% of the cases, in which $p16^{INK4a}$ is a tumor suppressor gene, located on chromosome 9 which is a key in the progression of CML from the chronic phase to the blast crisis. This gene inhibits cyclin D whose complexes have a role in preventing the arrest of the cell cycle (Shet et al., 2002).

Lastly, CML can progress from the chronic phase to the blast crisis phase due to impairment of the DNA repair mechanism (Forrest et al., 2008) The primary cause of the non-random chromosomal abnormalities is dysfunctional DNA repair. However, this mechanism is not yet well-understood and how it contributes to disease progression. but, Shet et al. (2002) approach the issue based on the core role of BCR-ABL1 and c-ABL. The study suggested that the presence of BCR-ABL1 down-regulates the role of c-ABL in the repair of damaged DNA. The assertion has been proven in a series of in vitro tests on human $p210^{BCR/ABL}$ whereby introduction of BCR-ABL fusion gene indicates a reduced rate of repair on DNA protein DNA-PKcs (Shet et al., 2002) Therefore, in the progressive phase of CML, the level of BCR-ABL protein increases in order to inhibit the rate of DNA repair. Consequently, multiple gene defects accumulate leading to disease progression.

In CML progression, the disease does not progress to Acute Myelogenous Leukemia despite there being a few cases of CML patients developing AML. In addition, the progressive phase of CML is different from acute leukemia. The three main phases in CML include the chronic phase, accelerated phase, and the blast crisis phase ("DISEASE PROGRESSION," 2010) Diagnosis of CML is mainly done during the

chronic phase of the disease. If the condition goes untreated during the chronic phase, the disease progresses to the accelerated and blast crisis phases. The clinical symptoms reported for most patients during the chronic period include fever, bone pain, recurring infections, and anemia. According to the National CML Society (2017), each phase of CML progression can be described based on the number of immature cells in the bone marrow. For example, during the chronic stage of the disease, the bone marrow has less number of blasts (< 10%) On the other hand, the accelerated phase of CML has between 10% - 30% of the blast cells in the bone marrow ("DISEASE PROGRESSION," 2010) Lastly, in the blast crisis phase, more than 30% of the blood cells in the bone marrow are immature cells.

Forrest et al., (2008) stated that the first investigation should be done on patients presenting the clinical symptoms of CML is white blood cell differential count. In addition, a complete blood count (CBC) should be performed immediately after the patients' admission to the hospital. Other tests which are done immediately include bone marrow aspiration and biopsy, serum electrolyte, creatinine, uric acid and liver function tests (Forrest et al., 2008) Examination of the bone marrow is particularly critical in patients presenting the symptoms of leukemia because it provides key information in terms of cellular morphology, marrow cellularity, the presence of fibrosis, and the current number of blast cells (Forrest et al., 2008) Other tests that may be necessary include the detection for the presence or the absence of a deletion in chromosome 9 through fluorescent *in situ* hybridization (FISH) In rare cases, whereby the CML patients are negative for Ph, polymerase chain reaction (PCR) is used to confirm the disorder. It is also used to detect the presence of Ph chromosome. Forrest et al. (2008) pointed out that the importance of PCR not only in the initial diagnosis but also in quantifying BCR-ABL1 transcripts. The strategy is important in the course

of the treatment because it provides a baseline on which patient care is provided. Furthermore, the sensitivity of the test enables the patient to achieve a complete cytogenetic response (CCgR).

Differential diagnosis of CML is required for certainty before the patient is put on therapy. When patients are tested and found positive of a high number of blast cells in the bone marrow, it is not possible to directly determine the type of leukemia present. Differential diagnosis requires additional tests from the bone marrow material or blood samples to identify BCR-ABL1 translocation or the Philadelphia Chromosome through standard Karyotyping. However, only a few cases of acute lymphoblastic leukemic patients have positively diagnosed for Philadelphia chromosome. In most cases, only a slight variation is observed in the breakpoint region which suggests the presence of Philadelphia chromosome ("DISEASE PROGRESSION," 2010) Another diagnostic method in the identification of CML is the examination of the remaining normal cells in either the bone marrow or the blood samples. Presence of a high number of basophils and neutrophils can be suggestive of CML. In addition, the samples show close to normal cell distribution in patients with acute leukemia. Progression of CML to blast crisis marks the final phase of the disease.

2.2.2. Non-Coding RNAs

In the recent years, non-coding RNAs (ncRNA) have gained the attention because of the potential they exhibit as a significant role in a numerous biological regulation (Kung, Colognori, & Lee, 2013) Scientific research has led to the discovery of multiple new RNA types. However, only a small class of the non-protein coding transcripts has shown potentially-derived functions. Transcriptomic studies have also developed knowledge on new classes of noncoding transcripts in various diseases. In a study carried out by Ponting et al. (2009) examined the evolution and the functions

of the noncoding RNAs (ncRNAs). Despite being the least understood of the transcript species, ncRNAs have shown potential in critical areas such as transcription regulation, disease, and epigenetic gene regulation.

Long non-coding RNAs (LncRNAs) are non-protein coding transcripts that are longer than 200 nucleotides. LncRNAs are also referred to as non-protein coding transcripts (Cao, 2014) In the vertebrate genomes, multiple genes encoding lncRNA have been discovered. In addition, it has become apparent that a larger fraction of the genome is represented in as exons in mature RNAs. Currently, more than 14,000 lncRNAs gene units in the human genome are annotated through extensive evidence (Kapusta & Feschotte, 2014) For a vast majority of the gene units, they present the RNA polymerase II transcripts. Large quantities of lncRNAs are known to occur in the mammalian genome. Since these lncRNAs are still under investigation, researchers have different approaches towards understanding them in terms of their origin, filtering, annotation and classification (Kapusta & Feschotte, 2014) The first classification of lncRNAs include the intergenic lncRNAs which do not overlap with known protein-coding loci, epigenic transcription, or post-transcriptional levels (Yang, Junjie, Sanjun, & Ma, 2017) Other categories are dependent on the specific class of organisms (Kapusta & Feschotte, 2014) For instance, in tetrapods, the lncRNAs are categorized based on the specific tissues and cell types they are in. On the other hand, in zebra fish and *Drosophila*, lncRNAs are inventoried in the whole animal but in different stages.

A vast majority of the lncRNAs have been characterized based on experiments and assays both in vivo and in vitro for assessment of the both their functions as well as their evolutionary significance. Despite multiple lncRNAs having defined cellular functions, there is still limited information about the evolutionary path of lncRNAs

either at the individual level or as a group (Kapusta & Feschotte, 2014) The primary reason is that lncRNAs tend to have an extreme heterogeneous sequence. Based on various diverse mechanisms, lncRNAs play an essential role in multiple biological processes (X. Chen et al., 2017) Further research shows that mutations and dysregulations of the lncRNAs involved in both the development and progression of a number of complex human diseases such as colon cancer, lung cancer, cardiovascular diseases, Alzheimer's disease, diabetes and leukemia (X. Chen et al., 2017) The field of cancer research has particularly been promoted by the current advances in understanding the basic mechanisms on how lncRNAs cause or support cancer progression. The current knowledge on changes of lncRNAs is likely to open up new methods including lncRNA-based clinical applications in cancer research.

Yang et al. (2017) draw attention to the importance of understanding the origins of lncRNAs in order to properly outline their functions and classify them accordingly. There are various ways in which the functional noncoding RNAs can develop. The first hypothesis on the origin of lncRNAs is through frame disruptions of a protein-coding gene which leads to the generation of lncRNAs. Secondly, due to chromosomal re-arrangement, two untranscribed and divided sequences can be juxtaposed into a multi-exon lncRNA. The third hypothesis on the origin of lncRNAs is through duplication of the noncoding gene as a result of retro transposition. Also, it is thought that tandem duplication events of neighboring repeats in transcriptions can lead to the formation of lncRNAs (Yang et al., 2017) The last hypothesis on the origin of the lncRNAs suggests that a transposable element is inserted into a gene as the lncRNA. Despite these assumptions being used widely to describe the origin of lncRNAs, there is little knowledge currently to ascertain any of them. However, for some few transcripts such as the X-inactive specific transcript *(Xist)*, the origin is

known. The lncRNAs are found in either sense or antisense alignment to protein-coding genes. Most of the lncRNAs function in cis through their own transcription or through intrinsic RNA-mediated functions in trans (Cheetham, Gruhl, Mattick, & Dinger, 2013) Current knowledge on the clinical importance of lncRNAs is likely to boost understanding of most medical conditions including genetics.

2.2.3. Functional Mechanism of LncRNAs

***Signaling molecules*.** A variety of lncRNA transcripts found to play a significant role in regulating different pathways with respect to cellular and molecular levels. Accordingly, lncRNAs functionally categorized into different classes. The first class of functional mechanism of lncRNA is the signaling molecules such as *HOTAIR* and *lincRNA-p21* (Yang et al., 2017) These molecules are transcribed as a result of diverse stimuli in a time-or space-specific manner. These lncRNA molecules reflect the integrated transcriptional activity. Wang and Chang (2011) identified that lncRNAs usually demonstrate cell type-specific expression and response to stimuli. Thus, it is possible that these responses are under a transcriptional control. Furthermore, due to the specificity of the time and place of transcription of individual RNAs, the molecules are in a position to serve the signaling role efficiently. These signals are indicative of a response to stimuli, interpretation of the cellular context and integration of the cellular signals. Signaling molecules are further divided into by-products of transcription and those possess the regulatory functions. Using the RNA for regulation is advantageous because the processes are performed faster without the need for protein translation (K. C. Wang & Chang, 2011).

LncRNAs have been used as signals for marking space, time, expression of gene regulation and developmental stages. Despite having the two parental alleles expressed equally in some cases, there are specific subsets of genes which signal to

imprint. In such cases, the expression is restricted by an epigenetic mechanism mediated by the lncRNAs. Another signaling role of the lncRNAs is through mediating anatomic specific expression. LncRNAs involved in this process include the *HOTTIP* and *HOTAIR.*

DNA damage also activates lncRNAs such as *PANDA* and *lincRNA-p21* which are mainly detected during times of cellular stress. The signals are important for regulation of *p53* which directly respond to DNA damage through induction of *lincRNA-p21*. Another area where lncRNAs are known to signal for DNA damage is in the mammalian *CDKNIA* promoter whereby several lncRNAs such as *PANDA* are produced and transcribed following DNA damage. During the process, *PANDA* is activated by the *p53* binding to the *CDKNIA* locus. Once produced, *PANDA* interacts with the transcription factor NF-YA in order to limit expression of pro-apoptotic genes thus contributing towards cell cycle arrest. The mechanism demonstrates that several lncRNAs are potentially useful in cell-growth control. The signaling role of lncRNAs is also critical in coordinating cellular activities. These lncRNAs have enhancer-like functions in various human cell lines. Studies have shown that depletion of these lncRNAs resulted in decreased expression of the neighboring protein-coding genes. The activity of master regulatory genes in the process of cell differentiation is also affected. Wang and Chang (2011) described the activity of classical enhancers of lncRNAs as orientation dependent. In addition, this class of lncRNAs requires a minimal promoter in the target gene for transcription enhancement. Lastly, lncRNAs such as *staufen* and RNA-decay are important signaling molecules for repeated sequences. Thus, signal molecules of lncRNA are not only used as markers for downstream transcriptional elements, but they are also crucial in detection and identification of transcript abundances in terms of repetitions.

Molecular decoys. The second class comprises of molecular decoys such as *PTENP1*, *MALAT1, PANDA* and *IPS1* (Yang et al., 2017) These molecular decoys have a role in binding and competitively inhibiting the transcription factor miRNAs. lncRNAs which act as molecular decoys regulate the process of transcription by binding the protein target. Therefore, the molecular decoys act as negatively regulating an effector without exerting additional functions. An example of lncRNA gene in the human which possess the RNA dependent mechanism of transcriptional repression is the *DHFR*. The gene directly inhibits the assembly of the pre-initiation complex through the formation of a stable ncRNA-DNA complex (K. C. Wang & Chang, 2011).

TERRA is another lncRNA which is thought to have a role in regulation and protection of the chromosome ends. Wang and Chang (2011) pointed out that the telomere is a key DNA-protein complex found at the end of the eukaryotic chromosomes which enhances the stability of the chromosome. Another way in which *TERRA* performs the regulatory role is through level changing in a cell cycle-dependent manner (K. C. Wang & Chang, 2011) The levels tend to increase in the early G1 and later decrease in the S phase. During the transition phase, *TERRA* reaches the lowest expression levels especially between late S and G2. In addition to that, Wang and Chang (2011) stated that down-regulation of *TERRA* in the S phase is vital in releasing telomerase and allows the telomeric strand to extend. Another lncRNA which possesses the decoy function is the *PANDA-NF-YA* (K. C. Wang & Chang, 2011) Once the DNA is damaged, and cell arrest or apoptosis is initiated, expression of *PANDA* is induced temporarily to inhibit the apoptotic genes from expression to promote the survival of the cell. Studies are done to assess the decoy function of *PANDA* show that its depletion increases the level of *NE-YA* occupancy at

the target genes. On the other hand, elimination of both *PANDA* and *NE-YA* attenuates the induction of apoptosis and the expression of the apoptotic gene.

Wang and Chang (2011) suggested that the lncRNA decoys are important especially when functioning in the nuclear subdomains or in the cytoplasm. *MALAT1*, an abundant nuclear lncRNA found in the mammalian cells binds to the serine/arginine splicing factors thus altering the activity of pre-mRNAs. *MALAT1* has a particularly important role in the hippocampal neurons whereby it regulates serine splicing factors resulting in the formation of a synapse (K. C. Wang & Chang, 2011).

Molecular guides. The other class of lncRNA based on their functional mechanism is the molecular guides. Examples of these lncRNAs are *Xist, HOTAIR, COLDAIR,* and *lincRNA-p21* (Yang et al., 2017) The molecular guides recruit chromatin-modifying enzymes to various target sequences in either cis or trans. The manner in which the RNAs perform the role cannot be predicted easily based on the lncRNA sequence. There are multiple ways in which the molecular guide lncRNAs function in the process of transcriptional regulation. For example, Enhancer RNA (eRNA) and Air tend to exert their effect in cis (K. C. Wang & Chang, 2011) On the other hand, lncRNAs like *HOTAIR* and *lincRNA-p21* which takes complex forms in the regulation require additional ability to interact with other factors to be effective. For the molecular guides, the gene regulatory components which are carried on by the lncRNA include the activating complexes and the repressive complexes (K. C. Wang & Chang, 2011) The basic mechanisms in which the lncRNAs work is based on the assumption that proper localization of the effector molecule is changed as a result of knockdown of the lncRNA. Another possible event following the knockout includes loss of function of the effector. Also, a knockdown on both the effector and the lncRNA can exacerbate the phenotype (K. C. Wang & Chang, 2011).

Wang and Chang (2011) categorized the molecular guide RNAs based on their course of action in transcriptive regulation. For example, some of the known molecular guides in cis include the *CCND1, HOTTIP, Xist, rDNA, Air* and *COLDAIR* transcripts. On the other hand, the molecular guides in trans include the other *PRC2-bound RNAs, HOTAIR, Jpx,* and *linc-2*. The long noncoding RNA Air silences the transcription of the target gene on the paternal chromosome through interaction between the chromatin, its ncRNA, and the promoter. Also, the ability of the RNA to bind to the complementary strands of DNA sequences may be an indicator of the RNA's capability to guide on the transmission of chromatin (K. C. Wang & Chang, 2011). Another aspect to the cis-regulation by the lncRNAs is the identification of *HOTTIP lncRNA. HOTTIP* serves as an intermediate for transmission of information from a higher order chromosomal looping to chromatin modification.

Molecular scaffolds. Finally, the molecular scaffolds *ANRIL, TERC,* and *HOTAIR* are important in assembling proteins to form ribonucleoprotein complexes. The property of lncRNAs to act as molecular scaffolds is critical to the precise control of the specificity and dynamicity of intermolecular interactions and the signaling events. Wang and Chang (2011) pointed out that there is a possibility of lncRNAs to play a role in various scaffolding complexes despite this being traditionally known as a function of the proteins. Molecular scaffold presents the most complex class of lncRNAs. They have various domains which bind distinct effector molecules (K. C. Wang & Chang, 2011) This property of molecular scaffolds enables them to bind to multiple effector partners simultaneously. Further knowledge of scaffolding complexes is critical in understanding the specific signaling components and how they impact on the cellular behavior.

A major hypothesis on the how the molecular scaffold work suggest that knockdown of the lncRNA changes or interferes with the localization of the effector molecule (K. C. Wang & Chang, 2011) In addition, disassembling the lncRNA-effector scaffold may lead to loss of function of the component effector. Molecular scaffolds such as *ANRIl* and *HOTAIR* represent lncRNAs which are present all the time at the locus and recruit sets of chromatin modifying complexes to the gene responsible for silencing (K. C. Wang & Chang, 2011) Therefore, they indirectly modulate the transcriptional activities.

2.2.4. Integrated Functions

Despite the functional mechanisms which facilitate their classification, lncRNAs have multiple modes of action which are useful in promoting their biological functions. LncRNAs also play an important role in modulating the activity of protein-binding partners (Wilusz, Sunwoo, & Spector, 2009) Many proteins, through the available motifs, bind to RNAs in order to modulate localization, processing and the stability of the bound RNAs. lncRNAs act as co-activators of proteins which are used in the process of transcriptional regulation (Wilusz et al., 2009) Another way in which the lncRNAs modulate the activity of protein-binding partners is by regulating their subcellular localizations. During the process, the transcription factor NFAT is localized in the cytoplasm until it is imported to the nucleus for transcription of the target genes.

Wilusz et al. (2009) in their recent studies suggested that lncRNAs are precursors for small RNAs. Normally, small RNAs usually cluster near the 5' and 3' ends of the gene. Post-transcriptional processing of lncRNAs and protein-coding RNAs produce many small RNAs with a 5' cap structure (Wilusz et al., 2009) In addition, lncRNAs affect the processing of other RNAs. For instance, they are known to modulate the

ability of other RNAs to be cut into small RNAs. Also, lncRNAs affect the processing by changing the pre-mRNA splicing patterns (Wilusz et al., 2009) Furthermore, it is possible for various forms of lncRNAs to base-pair with small RNAs in order to modulate their activities. An example is when lncRNA molecules interact with miRNAs. As a result of competitive inhibition, miRNAs are unable to bind with their target mRNAs (Wilusz et al., 2009).

2.2.5. LncRNAs Pathways in Cancer

Cancer being a disease of aberrant gene expression, it is important to understand mechanisms of gene dysregulation and how they contribute to malignancies. Cheetham et al. (2013) identified the inheritable single-gene disorders contribute to a small percentage of all cancers. On the other hand, a majority of the cancers arise from somatic mutations as a result of interaction between various genetic and environmental factors. With the discovery of multiple non-coding genes including the lncRNAs has revolutionized the current understanding of how the cell's mutation leading to cancer development. In addition, the knowledge has contributed especially in the development of new therapeutic options for specific carcinomas. However, studies on the specific pathway of lncRNAs are still in progress. The current knowledge of the disease is based on a predictive framework used to evaluate the expression patterns of genes (transcripts) and how they affect cellular function (Schmitt & Chang, 2016) Given the extensive role that lncRNAs play in the cell, it is, therefore, possible to predict their involvement in various cancer pathways.

Research showed that various lncRNAs have the potential to signal for specific states of the cell especially in case of pathologies like cancer. Different molecules of lncRNAs signal for specific cellular need. For example, Schmitt and Chang (2016) associated the overexpression of *HOTAIR*, a type of lncRNA with the prediction of

the presence of progressive metastatic disease in patients who had surgically resected breast cancer. In addition, further studies done on lncRNA *HOTAIR* have shown that it can predict differential sensitivity among patients with ovarian cancer. As a result, the findings are helpful especially in guiding the clinical decisions among the health care providers. Other forms of lncRNAs are critical in the development of diagnostic methods for cancers. Schmitt and Chang (2016) assessed the role of *PCA3* which is useful in *PSA* procedure for prostate cancer diagnosis. The test is based on lncRNAs and has become popular due to its non-invasive nature.

Another cancer pathway where the use of lncRNAs is extensive is on the identification of cancer subtypes which are clinically relevant. For example, an increased chance of developing metastatic progression of prostate cancer among men can be predicted by the overexpression of *SChLAP*. Additionally, some lncRNAs are extensively used to predict the effectiveness of various cancer therapies. For instance, studies have shown that ovarian cancer patients who show expression of *HOTAIR* after carboplatin treatment have poor survival rates (Schmitt & Chang, 2016; Zhao et al., 2016).

In a related study, Zhao et al. (2016) break down the various pathways in cancer development where the lncRNAs are heavily involved. Among them is the epigenetic regulation in cancer development. Modification of the cancer-associated RNAs is known to result in loss of imprinting. Alternatively, the lncRNAs can also contribute to epigenetic modification or the functional genes themselves. As a result, these changes can either cause or prevent diseases (Zhao et al., 2016) LncRNAs interact with the nucleosome remodeling complex in the cells and contribute in either the dislocation or rearrangement of the nucleosome. Consequently, repositioned nucleosome can be repackaged in regions which contain tumor suppressor genes and

subsequent repression leading to cancer development (Zhao et al., 2016) A good example is the lncRNA *HNF1A-AS1* whereby recent studies have demonstrated its involvement in esophageal tumorgenesis as a result of the role it plays in modulating the chromatin and assembly of the chromosome. Epigenetic regulation of lncRNAs in cancer development can also take the form of histone modification. In this case, the lncRNAs are involved in the recruitment of histone modifying enzymes in the cis function at the site of transcription. An example is the *CTBP1-AS* in prostate cancer. The lncRNA acts in cis and trans to regulate the epigenetic network thus stimulating cell proliferation (Zhao et al., 2016) Lastly, epigenetic regulation of lncRNAs in cancer can occur through DNA methylation. LncRNAs are involved in the recruitment of DNA methyl-transferases thus inducing demethylation. As a result, the tumor suppressor genes in the cell are silenced contributing to cancer pathogenesis. Although the process is not yet well understood, it is hypothesized that lncRNAs interact with the DNA methyl-transferases. Thus, the new complex guides the protein to various sites resulting in methylation of the promoters and subsequent silencing of the tumor suppressor genes (Zhao et al., 2016).

Zhao et al. (2016) further suggested that lncRNA genes (transcripts) can contribute to cancer development through regulation of the transcription pathway. The process involves the regulation of the transcriptional factors and polymerases. The level of transcription which is mediated by the lncRNAs is dependent on a number of factors such as the sequence features of the lncRNAs and their relative position. In addition, the target gene to be regulated also affects the level of transcription. During early and metastatic phase of the breast cancer, lncRNA *HOTAIR* is overexpressed and interacts with enzymes lysine-specific demethylase 1A and polycomb repressive complex 2

(PRC2) to control methylation process and subsequently regulating gene expression (Zhao et al., 2016).

LncRNAs also act in posttranscriptional regulation to cause cancer. Interaction with complementary base pairs with the antisense lncRNA transcripts forms RNA dimers. This binding resulting in blocks all the binding sites for the transcription factors while promoting those affecting posttranscriptional processes such as transport, ncRNA splicing, degradation, and translation. An example of lncRNA which causes cancer through posttranscriptional regulation pathway is the *MALAT1*. It regulates splicing of mRNAs through modulation of serine/arginine splicing factors.

Signaling pathways regulation by the lncRNAs is another important mechanism by which they contribute to cancer pathogenesis. Although the exact process is not yet understood, lncRNA can either suppress or activate the transduction pathway of the signals either at the transcription or posttranscriptional level (Zhao et al., 2016) There are many signaling pathways which are mediated by the lncRNAs to cause cancer. For instance, the ERK/MAPK pathway affects the important processes of the cell formation leading to unregulated cell growth. The lncRNA contributes to cancer progression by sending signals in the cell to boost its ability to escape apoptosis, enhancing immortalization.

Another signaling pathway through which lncRNAs cause cancer is Wnt/β-catenin signaling pathway which is essential in regulating proto-oncogene *β-catenin* stability (Zhao et al., 2016) Studies have revealed that the gene is activated in various types of cancers. The pathway is evolutionary conserved and it critical in cell proliferation regulation. In addition, the proto-oncogene controls tumor regression activity. Therefore, activation of the Wnt/β-catenin pathway results in genetic alteration and cancers. An example of lncRNA which has recently been associated with cancers

through the Wnt/β-catenin pathway is lncRNA *HOTAIR*. It plays a significant role in the development of squamous cell carcinoma. The pathway is activated following overexpression of the *HOTAIR* and *PRC2*. Activation of the Wnt/β-catenin pathway induces esophageal squamous cell carcinoma (ESCC) metastasis. Other examples of lncRNAs which are associated with various cancer types include *UCA1/Wnt6* in bladder cancer (Zhao et al., 2016).

Another pathway in cancer development is the AKT signaling pathway whereby serine-threonine; Akt kinases affect the signaling network. It is vital in areas such as apoptosis, differentiation, migration, metabolism, and proliferation (Zhao et al., 2016).

Another lncRNA which is common in promoting tumor growth and metastasis is *MALAT1*. The lncRNA is involved in a number of pathways all of which lead to cancer development (Zhao et al., 2016) For example, in gallbladder carcinoma, suppression of *MALAT1* inactivates the ERK/MAPH pathway. The new findings are indicative of the role that *MALAT1* plays a role in the proliferation of metastasis. Example of cancer diseases that develops as a result of direct involvement of lncRNAs is the hepatocellular carcinoma (HCC). Studies done on this cancer type has shown high levels of lncRNA *URHC* which plays a role in signaling and regulation of apoptosis.

Zhao et al. (2016) also described the *p53* signaling pathway which is activated as a result of cellular stress including oncologic signaling and DNA damage. The process then mediates the gene transcription. lncRNAs have been identified as part of the *p53* in these transcriptions. An example of lncRNA in this pathway which has been associated with cancer development is *lincRNA-Pint*. The lncRNA interacts with *PRC2* to form a *p53* transcriptional target. As opposed to most lncRNAs, *lincRNA-*

Pint is a tumor suppressor. In most, colorectal cancer studies, the level of *lincRNA-Pint* is usually downregulated. Protein stability is also important in suppressing cancers. Therefore, some lncRNAs such as *p21* act via *trans* gene regulation to repress the genes which are useful in regulating the stability including *p53* (Zhao et al., 2016).

In the ATM-CHK2 signaling pathway, the core role of the lncRNAs is in the regulation of the cell cycle. As a result, the process corresponds to DNA damage signals and subsequent tumor growth. On the other hand, STAT3 signaling pathway through the cell cycle induced prevention of important cellular activities such as apoptosis, and subsequent development of cancerous cells (Zhao et al., 2016) In PI3K/AKT signaling pathway, lncRNA *HOXD-AS1* is subjected to morphogenic regulation as a result of activation this pathway which develop the oncogenic pathways. Also In this process, lncRNA *UCA1* is expressed in bladder cancer. In such cases, there is a high rate of cell proliferation as well as regulation of the cell cycle progression through *CREB* activation. Finally, the TGF- β signaling pathway has a cytokine transforming growth factor β which is multifunctional. These factors are significant in inducing cell-cycle and tumorigenesis (Zhao et al., 2016).

Studies have been shown that various forms of cancers such as hepatocellular carcinoma are induced through TGF- β signaling pathway which is considered as a complex process. Studies have shown that in some rare cases, as cancer progresses, some forms of lncRNAs are used in regulating the signal transduction pathways (Zhao et al., 2016) In addition, the lncRNA is used as the downstream pathway. As a result, it forms a positive and negative feedback loop.

Table 1. *Listed the LncRNAs expression profiling associated with cancers (Schmitt & Chang, 2016) and (Yang et al., 2017).*

LncRNAs	Expression Patterns	Transcript function
HOTAIR	Upregulated in breast cancer/Ovarian cancer	Tumor oncogene
PTCSC3	Downregulated in thyroid cancer	Tumor suppressor gene
ANRIL	Upregulated in prostate cancer	The scaffold in mediating transcriptional silencing of INK4b-ARF-INK4a
XIST	Dysregulated in multiple types of cancer	The effector of the X inactivation process
PRNCRI	Upregulated in prostate cancer	Tumor suppressor
HULC	Upregulated in hepatocellular carcinoma	Tumor suppressor
H19	Upregulated in hepatocellular carcinoma	Tumor suppressor
KCNQ1OT1	Upregulated in colorectal cancer	Signaling for polycomb repressive complex 2 (PRC2)

2.2.6. The Clinical Implication of LncRNAs in CML

Li and Wang (2016) stated that CML is among the malignant diseases occurring in the clonal hematopoietic stem cells. The specific marker of CML is the breakpoint cluster BCR/ABL1 fusion gene and the Philadelphia chromosome (Kurzrock, Kantarjian, Druker, & Talpaz, 2003) Tumorigenesis induced by the BCR/ABL1 fusion gene involves alteration of various signaling pathways with the aim of facilitating and regulating cell survival and proliferation (Ying Li & Wang, 2016) PI3K/Phosphatase,

tensin homolog *RAS* and Janus Kinases are among the enzymes produced and act both as the signal transducers and activators of the transcription factors. Further studies on CML have shown that in cases of alteration of BCR-ABL1 gene expression, the processes of inducing cellular proliferation are upregulated. Other processes which are downgraded include inhibited cellular differentiation as well as apoptosis. Li and Wang (2016) pointed out that one of the key regulators of cellular transformation medicated by BCR-ABL1 is lncRNA *BGL3*. It has a role in the cell whereby it acts as a competitive endogenous RNA for repressing *PTEN* mRNA and at the same time binding with miRNAs (Ying Li & Wang, 2016) examples of miRNAs in which *BGL3* binds to including *miR-93, miR-17*, *miR-20b, miR-106a,* and *miR-106b*. Once bound, *BGL3* cross-regulates the expression of *PTEN*. Li and Wang (2016) pointed out that inhibition of BCR-ABL1 kinase activity induces the expression of lncRNA *BGL3*. Furthermore, its expression can be induced by following the disruption of BCR-ABL1 expression in either leukemia cells or K562 cells among CML patients (Ying Li & Wang, 2016) Using c-Myc-dependent DNA methylation, BCR-ABL1 is also capable of downgrading the expression of *BGL3*. It is, therefore, possible to develop a potential therapeutic strategy for CML through regulation of lncRNA *BGL3*.

Several studies have demonstrated the significant role which lncRNAs play in various biological processes including the development and progression of the CML. The primary difference between lncRNAs and other protein-coding genes including miRNA is that they are not yet well understood on their specific roles (Ying Li & Wang, 2016) LncRNAs contribute to the carcinogenesis through disruption of the major biological processes. The most common processes disrupted by the lncRNA are chains resulting to cancer development include gene silencing, and DNA methylation in epigenetic modifications (Kukurba & Montgomery, 2015) The main limiting factor

in understanding the specific role of the lncRNAs in the development of CML is lack of sufficient studies on all the lncRNAs involved. Only some of them has been identified so far (Ying Li & Wang, 2016) As a result, this has presented major challenges in fully investigating both their functions and mechanisms. Li and Wang (2016) highlighted that the differential expression of lncRNAs will provide an important marker in diagnosing and prognosis of CML in the future. Also, further research in future will be imperative in the development of lncRNA-mediated therapies.

CHAPTER 3: MATERIALS AND METHODS

3.1. Ethical Approval

Samples were utilized as part of ethical approval by HMC Ethics Committee of Medical Research Centre (MRC) at Hamad Medical Corporation- HMC. Copies of ethical approval with a study 11118/11 and MRC-03-17-0052 are attached in the appendix.

3.2. Sample Size and Collection

A total of 16 blood samples were used for this study that included nine samples of chronic phase (CP), three samples of complete remission (CR) and four samples of healthy individuals (CNT) The samples of the study subjects were collected based on the inclusion and exclusion criteria illustrated in Table 3.Whole blood has been used from each patient for RNA isolation by using QIAamp RNA Blood Mini Kit-based enzymatic reactions. Blood was mixed with erythrocyte lysis buffer (EL) twice: First, by adding 5:1 volumes of (EL: blood) with cold incubation for 15min followed by the centrifugation at 4000 rpm for 5min at 4°C Second, by adding 2:1 volumes of (EL: blood) followed by the centrifugation at 4000 rpm for 5min at 4°C to have a clear supernatant from RBCs and a good precipitated cell pellet (PBMCs) in which the intact RNA was isolated using RLT buffer (lysate) per sample which was stored at -80 for later use. Similarly was done for the control groups. The method for RNA extraction is described below in detail and illustrated in figure 2.

RNA extraction was performed at the molecular diagnostic laboratory at Hamad Medical Corporation-HMC. RNA was extracted from peripheral blood samples (PBMCs) as follow: Manual method was used for RNA extraction by following the manufacturer's instructions of "QIAamp RNA Blood Mini Kit" supplied by

"QIAGEN". For unhealthy blood (High WBCs count leukemic patients) reduce the blood volume to <1. 5ml a maximum of $1x10^7$ leukocytes should be processed while for healthy blood a maximum of 1. 5ml is required which contains $4x10^3$-$7x10^3$/µl. Whole blood was mixed with erythrocyte lysis buffer (EL) at a ratio of 5:1 (V/V) using 15ml falcon tube which incubated in ice for 15 minutes with brief vortex two times (every 5 minutes) during incubation to have a translucent and clear cell suspension, then centrifugation at 4000 RPM for 5 minutes at 4°C followed by decanting the supernatant and reserving the pellet. Add 2 volumes of EL buffer to 1 volume of the whole blood added initially, then cell pellet resuspended by vortexing and then centrifugation for 5 minutes at 4000 RPM at 4°C.

For homogenization step, 1ml of RLT buffer was mixed with 10µl of β - Mercaptoethanol **(β-ME)** then from this mixture (350µl or 600µl) of RLT buffer was added to the leukocyte pellet then lysate was pipetted into QIAshredder spin column in a 2ml collection tube and centrifuged for 2 minutes. The spin column was discarded and the resulted lysate was kept in collection tube either could be stored at -80°C for later using. Based on RLT volume added initially so one volume (350µl or 600µl) of 70% cold ethanol added to the lysate and mixed by gentle pipetting.

New QIAamp spin column in 2ml collection tube was used and centrifuged for 1 minute at 10,000 RPM, here the maximum volume should be added is 700µl. If the sample volume exceeding 700µl, divide this step into two steps after that the QIAamp spin column was transferred into a new collection tube followed by the addition of 500µl of RW1 buffer to the QIAamp spin column and centrifuged for 10 minutes at 10, 000 RPM. QIAamp spin column was placed in a new collection tube and 500µl of RPE buffer was added and the tube was kept in the rack for 1 minute at room temperature then centrifuged for 3 minutes at 10, 000 RPM. A final 1.5 ml

microcentrifuge tube was labeled with sample lab number in which QIAamp spin column was transferred to it and 30- 50µl of RNase-free was added into the QIAamp membrane which kept for 1 minute at room temperature followed by centrifugation for another 1 minute, at 10,000 RPM thus eluted samples were resulted for RNA quantity and quality assessment.

The Concentration and purity of the extracted RNA were measured by the NanoDrop™ (instrument, Thermo Scientific). The concentration of RNA was measured at absorbance ratio of 260/280, which is the acceptable purity ratio of nucleic acid against proteins that range from (1.8-2.2). Simultaneously, the RNA purity was measured at another absorbance ratio of 260/230 which indicates the nucleic acid purity against remained ethanol in the eluted sample, preferably it required >1.7 for this ratio. The optimum volume required per sample for accurate measurement is 2µl.

The RNA concentration was adjusted to 40ng/12ul of that was sent for RNA-Sequencing to the Genomics-core laboratory of Weil Cornell Medical School in Qatar.

In addition to that, RNA integrity number (RIN) was assessed using Agilent 2100 Bioanalyzer system which is important step prior to gene expression analysis, the analysis based on measuring the RNA degradation which usually occur when there is a decrease in the 18S and 28S ribosomal band ratio observed when performing gel electrophoresis and an increase in baseline signal between the two ribosomal peaks because the expression of the ribosomal rRNA is about 30% of total RNA. (Mueller, Lightfoot, & Schroeder, 2004) RIN assessment was done at the Genomics-core laboratory of Weil Cornell Medical School in Qatar. The RNA integrity number

(RIN) values for a set of eight samples were consistently ≥ 8.10 and the RIN values for the other set of eight samples were ≤ 5.90 but all samples were library constructed and sequenced.

Table 2. *The characteristics of the study subjects.*

Patients	Age (year)	Gender	CML phase
CP1	57	M	Chronic Phase
CP2	58	M	Chronic Phase
CP3	30	F	Chronic Phase
CP4	59	M	Chronic Phase
CP5	36	M	Chronic Phase
CP6	66	M	Chronic Phase
CP7	42	M	Chronic Phase
CP8	22	M	Chronic Phase
CP9	30	F	Chronic Phase
CR1	37	M	Complete remission
CR2	40	M	Complete remission
CR3	61	M	Complete remission
CR4	34	M	Complete remission
CNT1	40	M	Healthy Individual
CNT4	30	F	Healthy Individual
CNT5	30	F	Healthy Individual

Table 3. *The inclusion and exclusion criteria of the study subject.*

Inclusion Criteria	Exclusion Criteria
Patient's aged ≥ 20 years old	Normal WBCs count considered at 4-7X10^3 with exception of the Control group
Hematologic, Cytogenetic and Molecular - based diagnosis at any phase of CML	Ph chromosome-negative,BCR-ABL1 negative CML
WBCs count>10X10^3, Ph chromosome-positive CML, *BCR-ABL1* ratio (0% - 100%) Hematology, Cytogenetic and Molecular results	Previous history of IM intolerance
Imatinib Mesylate dose: 400mg/day, Nilotinib dose: 200mg/day (Alternative to (IM, Oral medications)	Pregnant or breastfeeding patients
Treatment response degree: Optimal response or resistance (treatment failure)	
Patients able to take an oral daily capsule	
Patients follow-up every three months, WHO and ELN-based guidelines	

3.2.1. RNA Extraction and Purification

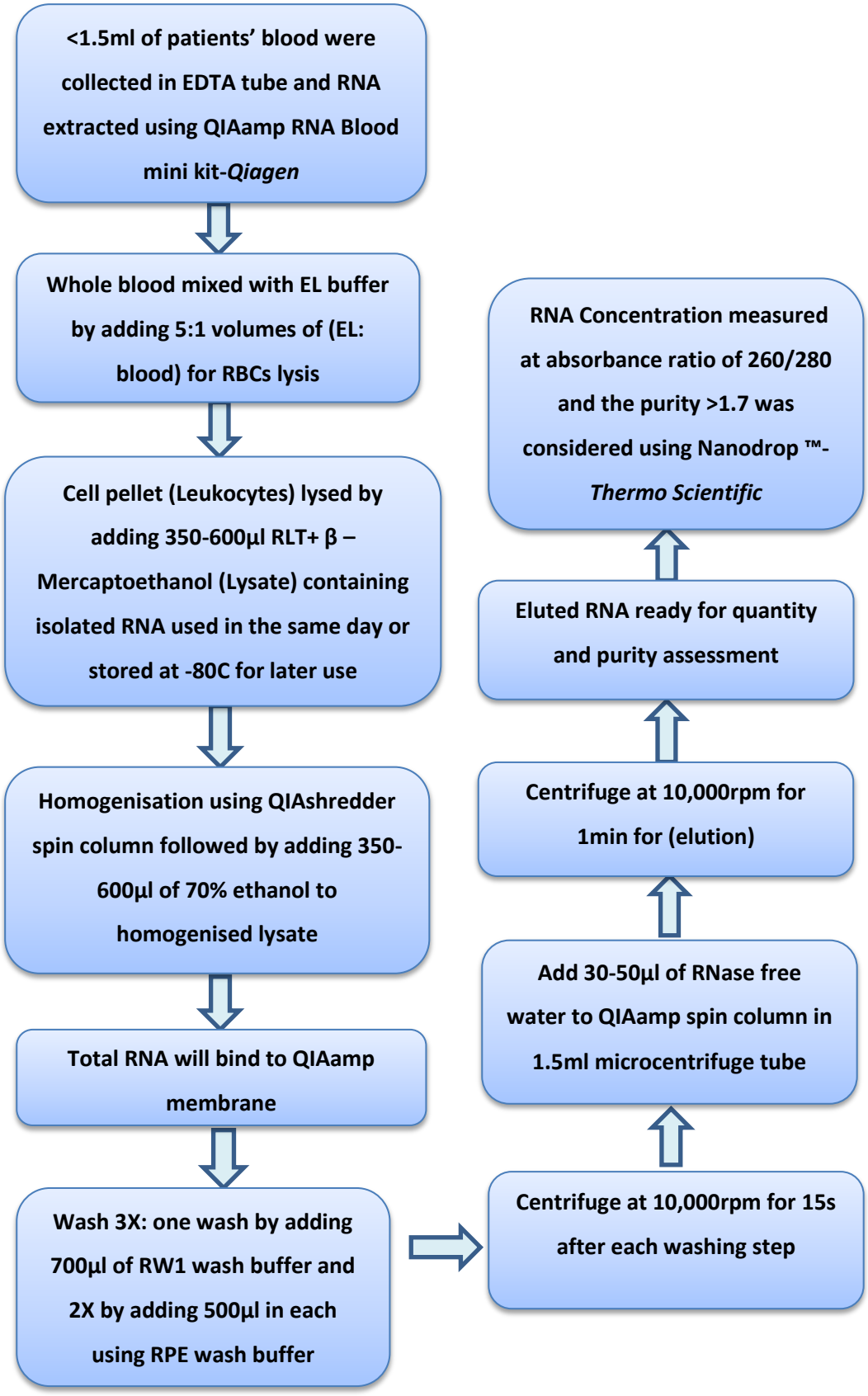


Figure 2. *Flowchart of RNA extraction and purification.*

3.3. RNA-Sequencing (RNA-Seq)

All RNA transcripts (coding and noncoding) longer than 170 bp were sequenced by a total of 1 Gb reads, 10X coverage, paired-end sequencing of 100 bp usingHiseq Illumina 4000. (Paired-end sequencing enable the sequencing of both sides of the DNA fragment in forward and reverse directions because the distance between each paired read is known which can be used for alignment to map the reads over the repetitive regions precisely leading to better alignment of the reads especially across difficult to sequence repetitive regions and improve the quality of the reads.

3.4. Preparing Raw Data

Fastq files of paired reads were imported into CLC Biomedical Genomics Workbench v5. Following the quality control (QC) of the reads which performed using the QC for sequencing reads tool integrated into CLC Biomedical Genomics Workbench software then reads were trimmed and overlapping paired read were merged into one sequence read, whenever relevant.

3.5. RNA-Sequencing Analysis

Paired reads were used to perform RNA-sequencing analysis followed by a calculation to create the fold change based on an annotated reference genome.

3.6. Map Reads to Reference

Reads were mapped to reference database using the following settings. To improve mapping the “quality score” was set at 10 for matching score. Other values were used as default. Reads shorter than 15 bps and less than half with 80% similarity were not mapped. For paired distance 100,000 reads were randomly extracted from the full data and mapped against the reference sequence (Homo_sapiens_refseq_GRCh38.p11) using alignment parameter tool integrated within the CLC Biomedical Genomics

Workbench software in which the QC of mapping criteria were determined. Mapped reads were counted and their distribution across genes and transcripts were calculated (Mortazavi, Williams, McCue, Schaeffer, & Wold, 2008) First, all genes are extracted from the reference genome (Homo_sapiens_refseq_GRCh38. p11_ Genes). Next, all annotated transcripts are extracted from (Homo_sapiens_refseq_GRCh38. p11_ mRNA). If there were several annotated splice variants, they all were extracted and kept for downstream analysis.

3.7. Local Realignment

To improve the alignments of the reads in an existing read mapping a realignment step was used in areas around insertions and deletions relative to the reference (Homer & Nelson, 2010) In this step reads mapped with one end of the read on one side of the insertion-deletion (indel) and the rest mapped on the other side. It used information from the other reads mapping to a region containing an indel, including reads that are located more centered across the indel and thus have been mapped with ends on either side of the indel.

The following steps were taken to normalize the data in order to minimize any differences due to sequencing depth (the number of the unique sequence (reads) shared the same region in a given reference sequence).

3.8. Count Normalization

To remove the possibility of any differences between samples resulted from possible differences in sequencing depth, a TMM (trimmed mean of M values) normalization, similar to that in the empirical analysis of digital gene expression R (EdgeR) was performed prior to comparing of samples (Robinson & Oshlack, 2010).

TMM normalization adjusts library sizes based on the assumption that most genes are not differentially expressed. For the expression visualization tools, (Create Heat Map and PCA) additional filtering and normalization are performed as followings: 1) 'log CPM' (Counts per Million) values are calculated for each gene. 2) The CPM calculation uses the effective library sizes as calculated by the TMM normalization. 3) After this first normalization, a second one is performed for each gene across samples: the read counts for each gene are mean centered and scaled to unit variance. 4) Genes or transcripts with zero expression across all samples or invalid values (NaN or +/- Infinity) were removed.

3.9. Annotation with Genes and Transcripts

The reference sequence, gene, and mRNA (included the splicing) were used for annotation. Expression values, RPKM and TPM were calculated using the lengths of the transcripts calculated from the mRNA track. No spike-in controls were used for this analysis. Reads that matches equally well to more than 10 hits were considered not unique and were not mapped. However, those reads that matched to multiple distinct places, but less than 10 were randomly assigned to one of these places using Expectation-Maximization (EM) algorithm that is similar to RSEM (software package used to measure the transcript quantity and expression level from RNA-sequencing data accurately) (B. Li & Dewey, 2011) and eXpress methods (another application used to quantify the data resulted from RNA-sequencing by using streaming algorithm with linear run time) (Roberts A, 2012) It iteratively estimates the abundance of transcripts, and assigns reads to transcripts according to these abundances.

3.10. Calculating Expression Values

Each intact paired read was counted as one and was used to calculate fragments (FPKM) rather than mapped individual 'reads' (RPKM). The expression value was calculated per gene or transcript. Total counts were calculated for transcripts were the total number of reads mapped per transcript and for genes, the total number of reads mapped per gene.

The biotypes are "as a percentage of all transcripts" or "as a percentage of all genes". For a poly-A enrichment experiment, it is expected that the majority of reads correspond to protein-coding regions. For rRNA depletion protocol, a variety of non-coding RNA regions may also be observed. The percentage of reads mapping to rRNA should usually be <15%

3.11. Calculating Fold Change

For each gene or transcript, the ratio of the expression value in each group was compared to the expression value of the same gene in the control group.

Fold change cut off was set to 1.5 for the raw count. To reduce the false positive, assuming that the large values in fold change can occur by chance if the expression levels are very low in both samples, reads with a minimum expression level of 10 or above were included.

3.12. Principal Component Analysis (PCA)

Principal Component Analysis was used to identify the outlying samples for quality control by reducing the number of transcripts into two or three principle components. Also used to examine if the selected subset of the transcripts can explain the variation between the groups, means that whether these transcripts in 16 samples could be grouped into their clinical stages The first principal component (PC1) specifies the

direction with the largest variability in the data between CP and CNT, the second component (PC2) is the direction with the second largest variation between CP and CR. Principal component analysis (PCA) was calculated based on the normalized log of count per million (CPM) and on a Pearson correlation matrix. To identify differentially expressed transcripts between 16 samples and 3 study groups, statistical analyses were done on the square-root transformed CPM of each transcript. Analysis of variance (ANOVA) was done to examine whether the samples or their groups or their interactions significantly influenced CPM. The Bonferroni correction was used for multiple comparison correction. Shapiro-Wilk and Bartlett's tests were used to test for normal distribution of the data and homogeneity of variances (normalization), respectively. None of the transcript expression levels violated the assumptions of the normal distribution or homogeneity of variances after Bonferroni correction.

3.13. Differential Expression

Multi-factorial statistics, based on a negative binomial generalized linear model (GLM) was used to calculate differential expression in paired groups. Each gene was modeled by a separate GLM to fit curves of expression values without assuming that the error on the values is normally distributed. Similarly to the empirical analysis of digital gene expression R (EdgeR) and DESeq, we assume that the read counts follow a negative binomial distribution similar to `Gamma-Poisson', where the Poisson parameter λ is itself Gamma-distributed. This Gamma distribution is controlled by the dispersion parameter, such that the negative binomial distribution reduces to a Poisson distribution when the dispersion is zero. To correct the problem associated with low GLM dispersion for a gene Cox-Reid adjusted likelihood, as in the multi-factorial EdgeR method was used (Robinson & Oshlack, 2010). Finally, after GLM fit and dispersion the total likelihood of the model given the data, and the uncertainty on each

fitted coefficient was calculated to test if a given coefficient is non-zero, Wald test was used for all group pairs and against control group comparisons.

3.14. Heatmap

Heat Map was created using the TMM normalization (as described before) to make samples comparable, followed by z-score normalization to make transcripts comparable. It also simultaneously clustered samples and features. Each column corresponds to one sample while each row corresponds to a feature (a gene or a transcript) Euclidean distance was used to measure the ordinary distance between two points and data were further filtered by setting the minimum absolute fold change to 1.5 and FDR p-value to less than 0. 01

3.15. Statistical Analysis

Statistical analysis was done using SPSS software and CLC Biomedical Genomic Workbench (Qiagen, Germany). Normalization and statistical methods were described in details in the material and methods of this section. P- value of less than 0.05 was considered for significant.

CHAPTER 4: RESULTS AND DISCUSSION

4.1. Quality of the Reads and Normalization

A total of 1.E+09 reads were obtained (average 7.65E+07, STD 3.03E+07). On average 92.77% (STD 1.79) of reads were mapped to genes vs 7.23 % (STD 1.79) that were mapped to the intergenic region. Over 90% of the reads mapped to mRNA. No rRNA was detected, which is due to the depletion step incorporated in the library construction.

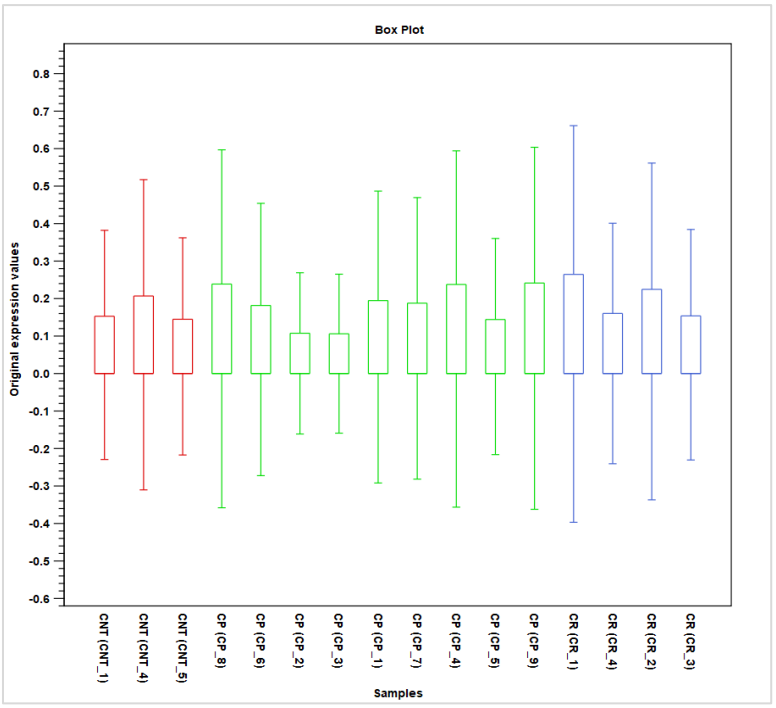
The sequencing depth (total reads) between samples varied as illustrated in figure 3. Even after the log 10 transformation, the variation persisted showed in figure 3. TMM (trimmed mean of M values) normalization, as described in the materials and methods section, similar to that in EdgeR, was performed before comparing of samples to remove differences in sequencing depth which showed in figure 4.

Table 4. *Distribution of biotypes of RNA-seq reads for 16 samples used in this study*

Distribution of Biotypes		
mRNA	90.26%	2.63%
ncRNA	5.44%	2.57%
misc_RNA	4.10%	0.54%
rRNA	0.19%	0.17%
tRNA	0.00%	0.00%

Read Count Normalization across 16 samples

A



B

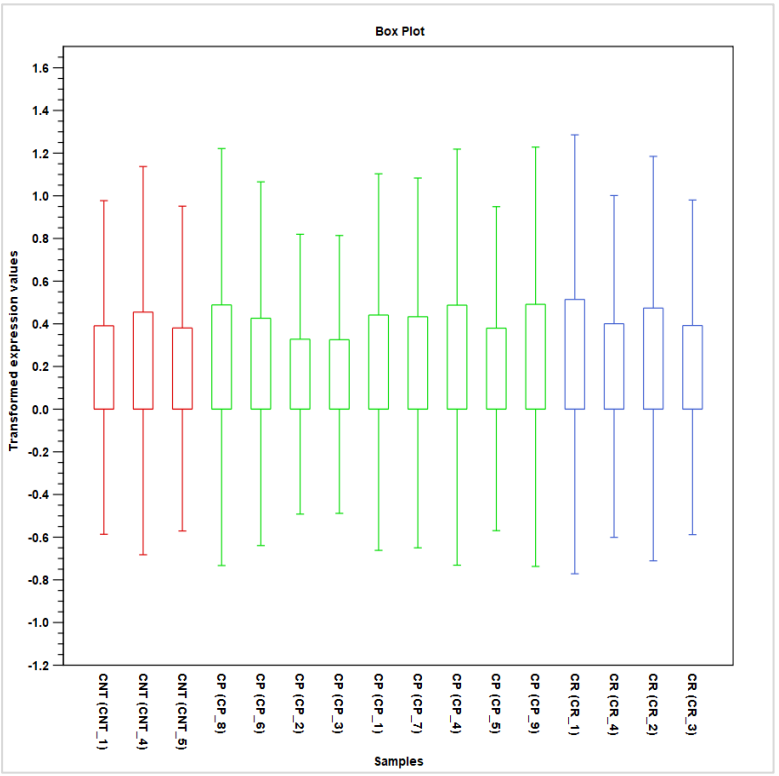


Figure 3. *Box plot of A) raw reads and B) Log 10 transformed paired reads from 16 study samples. CNT: Control group (red); CP: Chronic Phase (green); CR: Complete Remission (blue).*

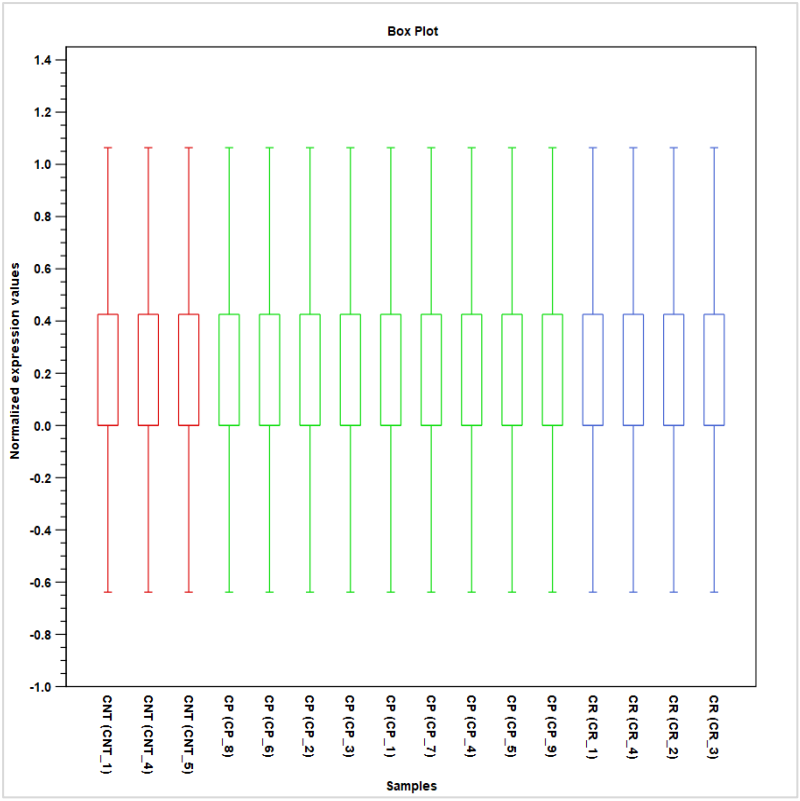


Figure 4. *Box plot of TMM normalized paired reads from 16 study samples. CNT: Control group (red); CP: Chronic Phase (green); CR: Complete Remission (blue).*

4.2. Filtration down and Selection of Transcripts

These databases were used to identify the expression values and the number of transcript variants per gene which generated two types of file including genes expression (GE) file which was developed using gene track database, while the transcripts expression (TE) file which was created using mRNA track database. For this study we used only transcripts expression values. Based on the study objective the TE track was selected, which contains a total of 153,438 transcripts for each one of 16 samples in three clinical phases as CNT: Controls; CP: Chronic Phase; and CR: Complete Remission. The expression values were calculated for each biotype such as ncRNA, mRNA and miscRNA as described in the materials and methods section.

These differentially expressed genes were identified and annotated in each biotype category by comparing between each pair of the three clinical phases (CP vs CNT, CR vs CNT and CR vs CP). Accordingly, the heatmap, Venn diagram and Principal Component Analysis (PCA) were used to illustrate the differentially expressed genes in three clinical phases. The initial PCA analysis of 153,438 transcripts failed to show any grouping across the 16 samples. To reduce the number of transcripts from 153,438 the mean differences between each paired group was used to calculate the *p* values. To correct for the number of comparison FDR and Bonferroni values were calculated. The initial section was done by filtering away the transcripts that had a p value less than 0.05. This reduced the number of transcripts to 136 out which 82 transcripts were significantly and differentially expressed in CR vs CP and 54 transcripts in CP vs CNT showed in figure 5.

Venn diagram of 136 differentially expressed transcripts in paired comparisons

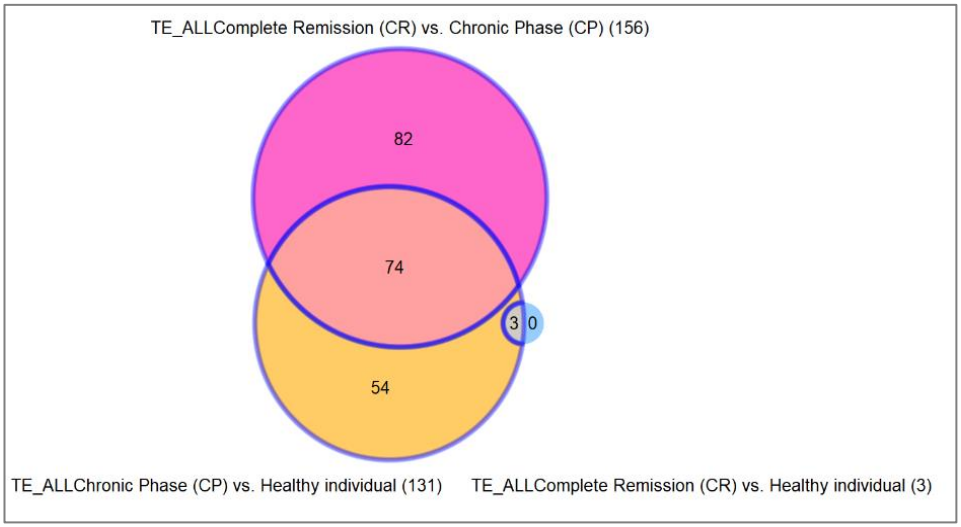


Figure 5. *Venn diagram of 136 significantly differentially expressed transcripts (P< 0.05) in paired comparisons. (CR vs CP, CP vs CNT, and CR vs CNT) used for reducing the 153, 438 transcripts of each of 16 sample to uniquely expressed 136 (82 +54) transcripts.*

To examine if this subset of transcripts can explain the variation between the groups, principal component analysis was performed on these 136 transcripts showed in figure 6. Using these transcripts 16 samples could be grouped into their clinical stages. However, CP2 subject (Chronic phase) was grouped with the control subjects.

3D Principal component analysis (PCA) of 136 transcripts

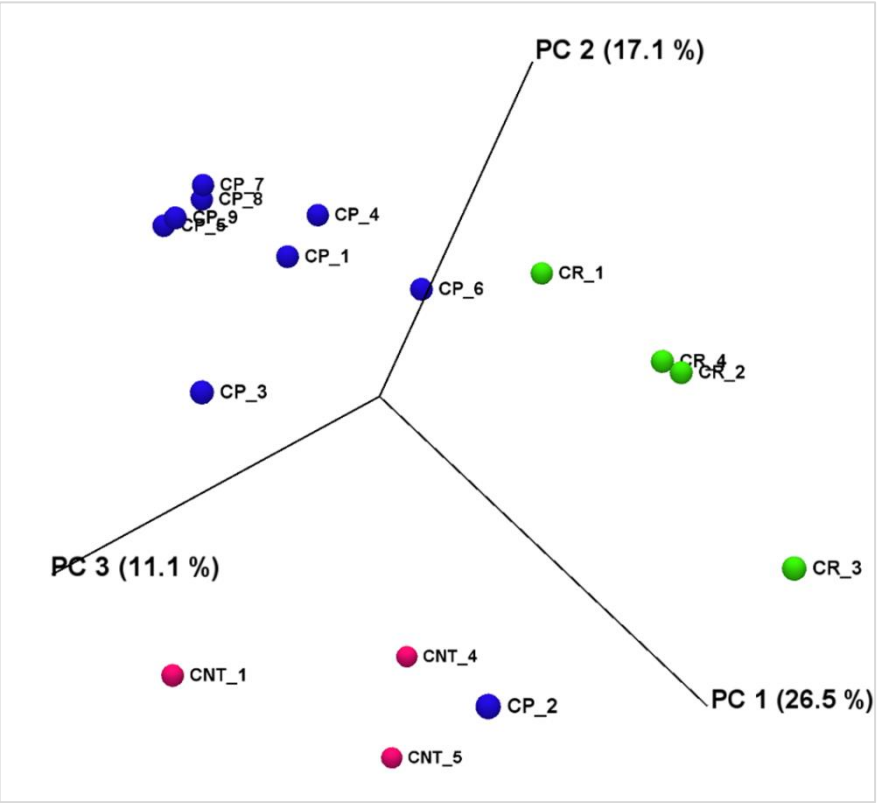


Figure 6. *3D Principal component analysis (PCA) of 136 transcripts. The first PC (PC1), second PC (PC2) and third PC (PC3) explained 26.5, 17.1 and 11.1% of the variances, respectively.*

To further reduce the number of the transcripts we used two methods. First, we clustered them by selecting only those that have a fold change higher than 1.5 Second, Euclidean distance between the transcripts was considered. This produced 39 transcripts that were differentially expressed (P < 0.01 and fold change> 1.5). The characteristics of these 39 transcripts are described in table 5 and 6.

Table 5. *The Description of 39 transcripts selected for this study*

Transcripts Name	Gene name	Transcript ID	Biotype	Status
ADGRG1_27	*ADGRG1*	XM_017023892.1	mRNA	Predicted
ANXA3_2	*ANXA3*	NM_005139.2	mRNA	Known
ATG7_25	*ATG7*	XM_017005546.1	mRNA	Predicted
BUB1_4	*BUB1*	NM_004336.4	mRNA	Known
CD53_1	*CD53*	NM_001040033.1	mRNA	Known
CFLAR_5	*CFLAR*	NR_147253.1	ncRNA	Known
CTC1_1	*CTC1*	NR_046431.1	ncRNA	Known
CYTH4_1	*CYTH4*	XR_001755214.1	ncRNA	Predicted
DPY19L3_8	*DPY19L3*	NM_207325.2	mRNA	Known
FAM208A_7	*FAM208A*	XM_017006030.1	mRNA	Predicted
GATS_1	*GATS*	NR_028040.1	ncRNA	Known
GRAMD1A_6	*GRAMD1A*	XM_011527149.1	mRNA	Predicted
HNRNPA3_4	*HNRNPA3*	NM_194247.2	mRNA	Known
LUZP1_5	*LUZP1*	NM_033631.3	mRNA	Known
MAZ_4	*MAZ*	NM_001042539.2	mRNA	Known
MPRIP_7	*MPRIP*	XM_011523766.2	mRNA	Predicted
MYCBP2_13	*MYCBP2*	XM_017020461.1	mRNA	Predicted
MYCBP2_19	*MYCBP2*	XM_006719786.2	mRNA	Predicted
MYCBP2_3	*MYCBP2*	XM_017020459.1	mRNA	Predicted
NFIX_3	*NFIX*	NM_002501.3	mRNA	Known
NLRC5_20	*NLRC5*	XM_011523376.2	mRNA	Predicted
ORM1_1	*ORM1*	NM_000607.2	mRNA	Known
PARVG_6	*PARVG*	XM_017028908.1	mRNA	Predicted
PLXNB2_8	*PLXNB2*	XM_017028704.1	mRNA	Predicted
PLXNB2_9	*PLXNB2*	XM_005261910.2	mRNA	Predicted
PTPRJ_3	*PTPRJ*	NM_002843.3	mRNA	Known
PXK_20	*PXK*	XM_017006678.1	mRNA	Predicted
RIN3_5	*RIN3*	NM_024832.4	mRNA	Known
RIT1_3	*RIT1*	NM_001256821.1	mRNA	Known
SCAPER_6	*SCAPER*	NM_020843	mRNA	Known
SENP5_6	*SENP5*	XM_011512544.2	mRNA	Predicted
SEPT9_2	*SEPT9*	NM_001293695.1	mRNA	Known
SLC4A7_3	*SLC4A7*	NR_135543.1	ncRNA	Known
STAU1_9	*STAU1*	NM_017453.3	mRNA	Known
TAOK3_4	*TAOK3*	XM_005253897.1	mRNA	Predicted
TBC1D4_8	*TBC1D4*	XM_017020883.1	mRNA	Predicted
U2SURP_5	*U2SURP*	NM_001320219.1	mRNA	Known
UTRN_7	*UTRN*	XM_017011245.1	mRNA	Predicted
VNN1_1	*VNN1*	NM_004666.2	mRNA	Known

Table 6. *The nomenclature of 39 transcripts selected for this study*

Transcripts Name	Transcript ID	Database object name
ADGRG1_27	XM_017023892.1	Adhesion G protein-coupled receptor G1
ANXA3_2	NM_005139.2	Annexin
ATG7_25	XM_017005546.1	ATG7 autophagy related 7 homolog (S. cerevisiae), isoform CRA_b
BUB1_4	NM_004336.4	BUB1 protein
CD53_1	NM_001040033.1	Leukocyte surface antigen CD53
CFLAR_5	NR_147253.1	Homo sapiens CASP8 and FADD like apoptosis regulator (CFLAR), transcript variant 28
CTC1_1	NR_046431.1	Conserved telomere capping protein 1
CYTH4_1	XR_001755214.1	Cytohesin-4
DPY19L3_8	NM_207325.2	Probable C-mannosyltransferase DPY19L3
FAM208A_7	XM_017006030.1	Protein FAM208A
GATS_1	NR_028040.1	CASTOR family member 3 (CASTOR3), transcript variant 4,
GRAMD1A_6	XM_011527149.1	GRAM domain-containing protein 1A
HNRNPA3_4	NM_194247.2	Heterogeneous nuclear ribonucleoprotein A3
LUZP1_5	NM_033631.3	Leucine zipper protein 1
MAZ_4	NM_001042539.2	Myc-associated zinc finger protein
MPRIP_7	XM_011523766.2	Myosin phosphatase Rho-interacting protein
MYCBP2_13	XM_017020461.1	E3 ubiquitin-protein ligase MYCBP2
MYCBP2_19	XM_006719786.2	E3 ubiquitin-protein ligase MYCBP2
MYCBP2_3	XM_017020459.1	E3 ubiquitin-protein ligase MYCBP2
NFIX_3	NM_002501.3	Nuclear factor 1
NLRC5_20	XM_011523376.2	Protein NLRC5
ORM1_1	NM_000607.2	Alpha-1-acid glycoprotein 1
PARVG_6	XM_017028908.1	Parvin, gamma, isoform CRA_b
PLXNB2_8	XM_017028704.1	Plexin-B2
PLXNB2_9	XM_005261910.2	Plexin-B2
PTPRJ_3	NM_002843.3	Protein-tyrosine-phosphatase
PXK_20	XM_017006678.1	PX domain-containing protein kinase-like protein
RIN3_5	NM_024832.4	Ras and Rab interactor 3
RIT1_3	NM_001256821.1	GTP-binding protein Rit1

Table 6 Continue. *The nomenclature of 39 transcripts selected for this study*

Transcripts Name	Transcript ID	Database object name
SCAPER_6	NM_020843	S phase cyclin A-associated protein in the endoplasmic reticulum
SENP5_6	XM_011512544.2	SUMO1/sentrin specific peptidase 5, isoform CRA_a
SEPT9_2	NM_001293695.1	Septin 9, isoform CRA_b
SLC4A7_3	NR_135543.1	Anion exchange protein
STAU1_9	NM_017453.3	Double-stranded RNA-binding protein Staufen homolog 1
TAOK3_4	XM_005253897.1	Serine/threonine-protein kinase TAO3
TBC1D4_8	XM_017020883.1	TBC1 domain family member 4
U2SURP_5	NM_001320219.1	U2 snRNP-associated SURP motif-containing protein
UTRN_7	XM_017011245.1	Utrophin
VNN1_1	NM_004666.2	Pantetheinase

Out of 39 transcripts, 34 are mRNA and 5 are ncRNA. Those with reference number starting with letter X are predicted and unknown. The heatmap of 39 transcripts shows a cluster of the transcripts that are differentially expressed (up-regulated (green) and down-regulated (red)) in one or two of three different groups showed in figure 7.

Heatmap of 39 differentially expressed transcripts clustered in three different groups of CNT: Control group, CP: Chronic Phase, and CR: Complete Remission.

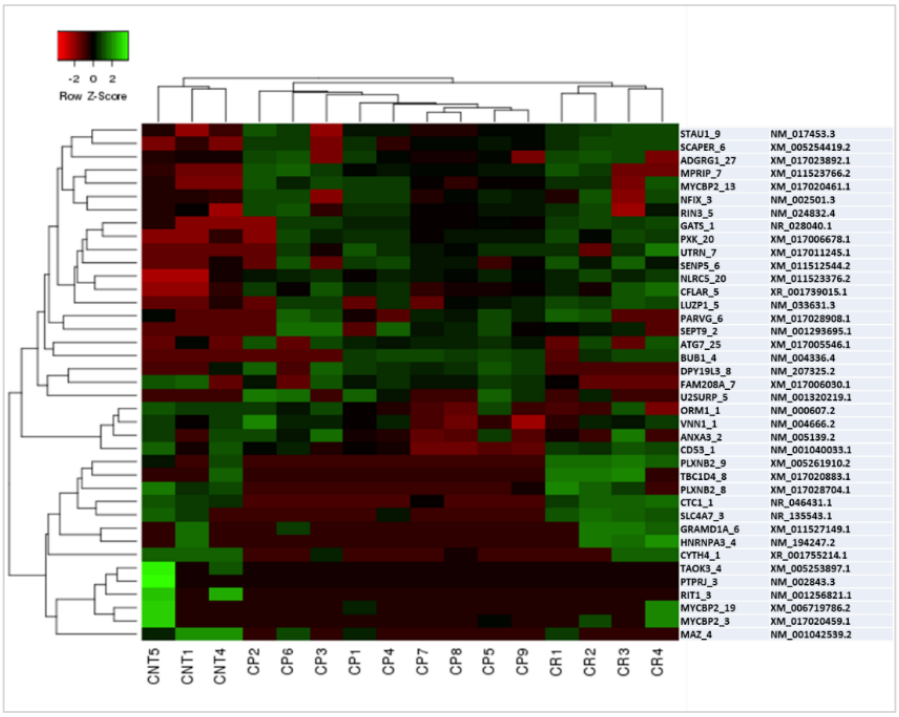


Figure 7. *Heatmap created from Log10 transformed CPM of 16 Samples and 39 transcripts were simultaneously clustered using complete linkage. Each column corresponds to one sample, and each row corresponds to log CPM per transcript. The numbers in front of each transcript are the splice variants. Euclidean distance was used to measure the ordinary distance between two points. Data were further filtered by setting the minimum absolute fold change to 1.5 and FDR p value to less than 0.01*

Further filtration was performed by using the Venn diagram to reduce the number of 39 transcripts to 23 significantly expressed transcripts in paired groups by considering the more stringent FDR p value of less than 0.0001 and the fold change cut off for each transcript at 1.5 illustrated in figure 8. Transcripts with expression value less than 10 were excluded to minimize the false positive reads.

After applying the FDR p value <0.0001, the majority of the resulted genes (transcripts) are protein coding except the four transcripts such as *CFLAR_5, GATS_1, SLC4A7_3, and CTC1_1* are LncRNA.

Venn diagram of 23 significantly expressed transcripts in paired groups comparison CR vs CP, CP vs CNT, and CR vs CNT.

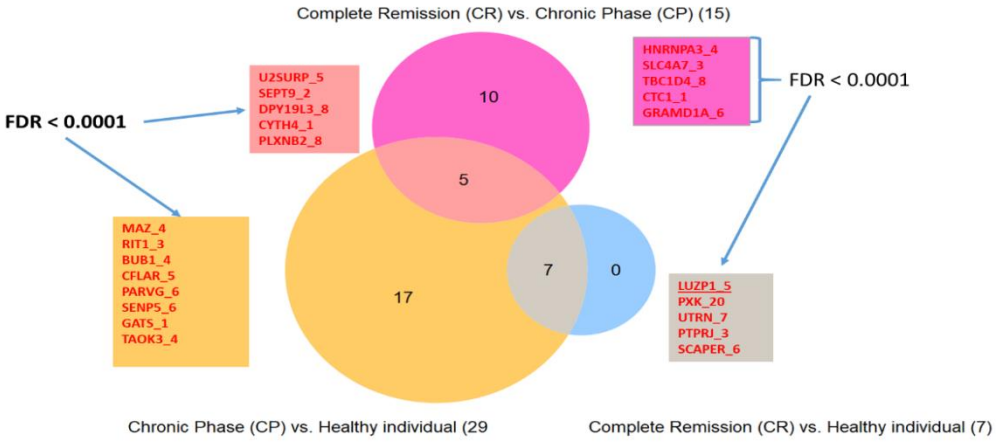


Figure 8. *Venn diagram created for paired groups based on the fold change cut off of 1.5 for raw count per transcript. 136 transcripts were reduced to 39 by filtering FDR p value lower than 0.01 CNT: Controls; CP: Chronic Phase; CR: Complete Remission. 23 transcripts that are marked in red are those that were filtered by using a more stringent FDR of less than 0.0001*

4.2.1. Transcripts that are differentially expressed in patients in chronic phase compared to healthy individuals

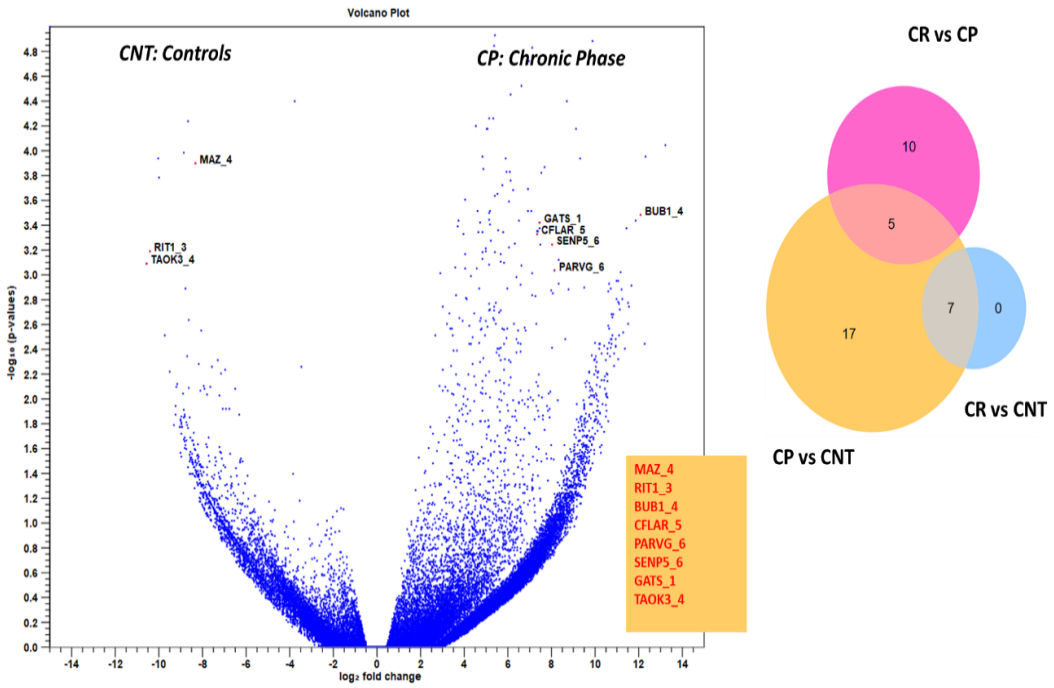


Figure 9. *Volcano plot showing transcripts selected from Venn diagram created for paired groups based on the fold change cut off of 1.5 and FDR of less than 0.0001 CNT: Controls; CP: Chronic Phase; CR: Complete Remission.*

As shown in figure 9. There are 17 transcripts that are differentially expressed in the subjects with chronic phase. We selected eight of these transcripts by setting the cut off of FDR value to less than 0.0001: *MAZ_4, RIT1_3, BUB1_4, CFLAR_5, PARVG_6, SENP5_6, GATS_1, TAOK3_4* All except *MAZ_4, RIT1_3* and *TAOK3_4* are highly expressed in the chronic phase compared to control subjects.

MAZ_ 4 or *Myc*-associated zinc finger protein. It was strongly down regulated in the CP subjects compared to the control group in which it is significantly upregulated with an increased fold of 200 it has been shown to be involved in cellular nitrogen compound metabolic process.

It resides in cytoplasm, nucleoplasm and nucleus and has been reported to have DNA and RNA binding (Bossone, Asselin, Patel, & Marcu, 1992; Robinson & Oshlack, 2010) (Baltz et al., 2012), ion binding (Song et al., 2001) and nucleic acid binding transcription factor activity (Baltz et al., 2012; Castello et al., 2012) Because of these central roles in the cellular function it conceivable to imagine that some of the above activity in the blood cells are suppressed.

RIT1_3 is a GTP-binding protein and is also known as *RIBB, RIT* and *ROC1*. *RIT1* encoding a member of *RAS* family proteins. It is a part of cellular component and plasma membrane and is involved in ion binding as participating in neuron stress-mediated survival and signal transduction such as RAS/MAPK pathway (Aoki et al., 2013) ("*RIT1* gene *Ras* like without *CAAX* 1," 2018) A study conducted by Segui, G., et al, 2013 reported that 10% to 50% of somatic mutation in *RAS* subfamily genes particularly *RIT1* gene occurred in myeloid neoplasms including MPNs, mixed myelodysplastic MDS and chronic myelomonocytic leukemia CMML. Also it has been found that 15% to 30% of acute myeloid leukemia cases have somatic mutations in *RAS* family genes (Gomez-Segui et al., 2013) 2% of the cases with lung adenocarcinoma have somatic mutations in *RIT1*. Therefore, RIT1 could be considering as an oncogenic gene (Berger et al., 2014) However, In this particular study, the gene is upregulated in the healthy subjects, who did not have any condition related to chronic myeloid leukemia with an increased expression value of 600 fold compared to chronic phase.

TAOK3_4 or Serine/threonine-protein kinase *TAO3* is involved in DNA metabolic process, cell cycle response to stress, signal transduction and cellular protein modification process .This gene expressed in mitochondria as serine/threonine protein kinase, it activates the p38/MAPK14 stress-activated MAPK cascade but inhibits the basal activity of the MAPK8/JNK cascade (W. Zhang et al., 2000) (Raman, Earnest, Zhang, Zhao, & Cobb, 2007). The encoded protein is a member of the *GCK* subfamily of *STE20*-like kinases. *TAOK3* gene is ubiquitous expressed in the brain and prostate tissues with RPKM values of 12 and 9.7 respectively ("Gene ID: 51347 *TAOK3 TAO* kinase 3 [Homo sapiens (human)]," 2018) In the present research, the downregulation of the gene in cancers and subsequent upregulation in health individuals with an increased expression value of 600 fold suggest that the gene is a tumor suppressor.

The downregulation of *MAZ, RIT1, and TAOK3* genes in chronic phase patients might be an indication of aberrant functions that is required for normal function of blood cells and be a good diagnostic and monitoring tool for those in the chronic phase.

Interestingly, the subjects in the chronic phase were the only one with highly upregulated *BUB1_4*, *CFLAR_5, PARVG_6*, *SENP5_6* and *GATS_1* transcripts.

BUB1_4 is known as budding uninhibited by benzimidazoles encodes a serine/threonine- kinase protein, which acts in phosphorylation of the mitotic checkpoint particularly, the activation of the spindle assembly checkpoint (SAC) ("Gene ID: 699 BUB1 mitotic checkpoint serine/threonine kinase [Homo sapiens (human)]," 2018) On the other hand, this gene has a role in inhibiting the activity of anaphase complex/cyclosome during the mitotic division, DNA repair and chromosomal stability, which is one of the roles of tumor suppressor. Therefore, any

mutations in this gene lead to various cancers and aneuploidy**.** This gene is abundant in testis and lymph node tissues with RPKM values of 25.2 and 11.3 respectively. In this study, the gene showed an irregular pattern of upregulation in the chronic phase with an expression level of 6500 fold compared to healthy individuals. In a study conducted by Hatfield, Reikvam, and Bruserud (2014), stated that seven of 31 patients with acute myeloid leukemia AML had a long-term malignant cell proliferation in vitro showed an altered expression of *BUB1* and other cell cycle regulator genes (*HMMR, NUSAP1, AURKB, CCNF and DLGAP5*), which was suggested to be used as a potential therapeutic target of AML (Hatfield, Reikvam, & Bruserud, 2014) Another study revealed *BUB1* to be one of the targets of *miR-155*, which acts as its negative regulator. Oncogenic *miR-155* was found to be overexpressed in multiple solid and hematological malignancies including breast, lung, colon and leukemia leading to genomic instability by causing defects in DNA mismatch repair, telomere integrity and chromosome stability through binding to 3'-UTR of *BUB1* gene causing the depression of *BUB1* protein levels to (–70%) in normal human dermal fibroblast (HDF) cells injected with lentiviral vector carrying miR-155 and to (–30%) in breast cancer cell line (MCF7) followed by a decreased to (–50%) detected in HCT116 cells using western blot test after ectopic expression of miR-155 to these cells (Pagotto et al., 2018) This is a surprised finding as it is over expression in the chronic phase and to some extent but not significantly up regulated in the complete remission, compared to controls, points to opposite function of *BUB1* as being oncogenic.

PARVG_6. Parvin, gamma, isoform CRA_b is highly expressed in bone marrow, spleen and lymph nodes. It has been shown to be involved in cell adhesion, cytoskeleton organization and fibroblast migration (Korenbaum, Olski, & Noegel, 2001; Yoshimi et al., 2006) In this research, *PARVG* expression was 280 fold

increased in the chronic phase patients compared to control group. It was also elevated in the complete remission phase but was not significant. Given the fact that it is involved in the interaction of leukocytes with the vascular endothelium it might be of great interest to explore the susceptibility of the CML patient to infection.

SENP5_6 Also known as *SUMO*-specific peptidase 5, functions in a reversible posttranslational modification. It has two main functions in the *SUMO* pathways which are important in regulating cellular mechanisms such as the cell division and the transcription process by catalyzing the cleavage of *SUMO1* from a number of mitochondrial substrates, therefore, overexpression of cytosolic *SENP5* gene contribute in regulating the *SUMO1* proteolysis from the mitoconderial substrates which affecting the morphology and metabolism of mitochondria (Zunino, Schauss, Rippstein, Andrade-Navarro, & McBride, 2007) Another pathway of *SENP5* gene is the de-conjugation of *SUMO2* and *SUMO3* from their targeted proteins Unlike ubiquitination, SUMOylation does not guide the protein for degradation. It has been shown to play a role in promoting the tumor progression. However both *SUMO* and ubiquitin proteins are directed by enzymatic cascade. *SENP5* is a *SUMO*-specific protease that generates the precursors of *SUMO*-proteins. It also acts to remove the *SUMO* proteins from *SUMO* conjugated substrates; a process known as isopeptidase. However, *SENP5* would be unable to process the ubiquitin proteins. It has been shown to have a prognostic potential for esophageal carcinoma and ovarian serous cystadecarcinoma. *SENP5* is expressed in testis and brain tissues with RPKM values of 8.4 and 6.3 respectively ("Gene ID: 205564 SUMO specific peptidase 5 [Homo sapiens (human)]," 2018) A study revealed that *SENP5* gene is one of the seven oncogenes that are highly expressed in different tumors. Tumor Genome Cancer Atlas (TCGA) cells which reached to 20% in esophageal carcinoma and lung squamous cell

carcinoma. Also in ovarian serous cystadenocarcinoma samples, the frequency of genetic alteration such as, gene copy number gains (CNG) for *SENP5* was 22% Moreover, the highest expression level of *SENP5* gene correlated with poor prognosis of breast cancer, while the low expression level of this gene in breast cancer associated with good prognosis (Wee, Liu, Lu, Li, & Zhao, 2018) Another study emphasized the role of *SENP5* in promoting the growth of tumor cells. It was reported that SENP5 is highly expressed in hepatocellular carcinoma (HCC), about 3.3 fold change compared to adjacent normal tissues. Also it was detected that *SENP5* regulates the response to DNA damage through deSUMOylation of ATRIP; known as ATR-interacting protein kinase that regulates the response to DNA damage checkpoint pathway, thus *SENP5* could be a promising therapeutic target of HCC (Jin, Pei, Xu, Yu, & Deng, 2016) This study showed a 264 fold expression of *SENP5* gene in the chronic phase group compared to that in complete remission group. *SENP5* could therefore, potentially, be used as a biomarker for chronic phase of CML.

CFLAR_5 (CASP 8 and FADD like apoptosis regulator) and ***GATS_1*** (transcript variant 4 of CASTOR family member 3 (CASTOR3) are both LncRNAs. The former transcript characterized by having some transcript variants containing different isoforms, Therefore *CFLAR* gene is clinically significance with 17 pathogenic variances. *CFLAR* transcript is abundant in fat tissues, bone marrow and lung tissues with RPKM values of 11.3, 9.5 and 9.4 respectively. The latter is a favorable prognostic marker in cervical cancer. This variant of *GATS-1* is generated due to a 3' UTR alternative splice pattern compared to the variant 1. It is non-coding because of its 5'- translational start codon tags it for the nonsense-mediated mRNA decay (NMD) pathway. Not much is known about this LncRNA. However, the results obtained from our study shows a 165 fold significant increase for both LncRNAs in the subjects with

chronic phase compared to healthy individuals. This result could help the physicians in diagnosing the leukemia cells in the early stages. Although they are also elevated in the complete remission subjects but it is not significant. A further study may shed a light on the significance of these two LncRNAs as a biomarker.

4.2.2. Transcripts that are differentially expressed in complete remission only compared to chronic phase

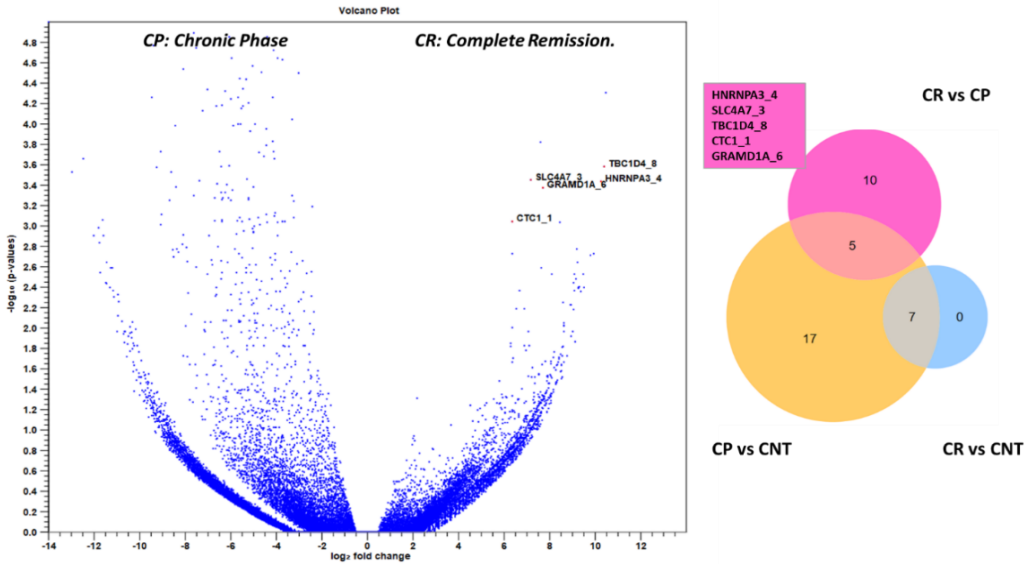


Figure 10. *Volcano plot showing transcripts selected from Venn diagram created for paired groups based on the fold-change cut off of 1.5 and FDR of less than 0.0001 CNT: Controls; CP: Chronic Phase; CR: Complete Remission.*

Figure 10. 10 transcripts found in this study that have a significant expression in the subjects of the complete remission only. We selected five of these transcripts by setting the cut off of FDR value to less than 0.0001: *TBC1D4_8 HNRNPA3_4 SLC4A7_3 GRAMD1A_6 CTC1_1*

HNRNPA3_4 encodes for heterogeneous nucleus ribonuleuprotein A3, this protein is one of *(hnRNPs)* isoforms that includes *hnRNPA0, A1, A2, and B1*. They play a significant role in regulating different cellular processes and are involved in nucleic acid breakdown. This protein is a nucleic acid binding protein mediated in trafficking of RNA molecules in cytoplasm through binding to RNAs that contain *cis*-acting regulatory element known as A2RE. It might have a role in splicing process of the pre-mRNA. This gene is highly abundance in appendix and lymph node with RPMK value of 56.8 and 56 respectively. It has been shown to have a pathogenic variance. Takoya identified 25 proteins that were differentially expressed in Cisplatin- naïve and Cisplatin-resistance bladder cancer cell lines (T24 and T24CDDPR, respectively) *HNRNPA3, PGK1, TK* and *SERPINB2* proteins had an increased expression pattern of more than 1.5 times in T24CDDPR compared to that observed in T24. *HNRNPA3* and *PGK1* were the most highly expressed proteins in the bladder cancer cells regardless of the concentration of the Cisplatin drug used (Taoka et al., 2015) Garcia and Gonzalez showed that transcript level of *HNRNPA3* gene in lung cancer cell lines, under different stress conditions such as hypoxia, serum deprivation was the lowest compared to the highest transcript level of *HNRNPA2, B1, A0* & *A1* (Romero-Garcia, Prado-Garcia, & Lopez-Gonzalez, 2014) In this study, the expression of *HNRNPA3* was significantly up regulated in subjects with complete remission of CML with a significant increase in expression level of 2200 fold compared to control and chronic phase subjects. However, given the fact that one of the control subject (CNT 1) had a comparable expression of *HNRNPA3*, it is more likely that *HNRNPA3* is downregulated in the chronic phase patients than over expression in the complete remission. If that would be the case then *HNRNPA3* can be used as diagnostic tool for

the monitoring the recovery of those that in the complete remission and maybe identify those that are at risk of relapse.

SLC4A7_3. Is the transcript variant number 12 of solute carrier family 4 member 7 and is a LncRNA. This locus encodes a gene for the transmembrane protein of sodium bicarbonate co-transporter. The encoded protein carry both sodium and bicarbonate ions and thus this gene is considered as an electroneutral co-transporter and regulates the intracellular PH necessary in transmission of vision and auditory sensors. In this study this LncRNA was highly expressed with the fold value of 290 in the subjects of the complete remission phase .However, just like the case of *HNRNPA3* it is more likely that the expression is suppressed in the CP since one subject in the control groups had expression as well. This might be a good marker, along *HNRNPA3* for the monitoring of CR group.

TBC1D4_8 known as *Rab-GTPase*-activating protein, acts in the downstream signaling pathway and is characterized by having several binding domains such as *Rab-GTPase,* two types of phosphotyrosine (*PTB1* and *PTB2*), calmodulin domain *CBD* and , multiple *AKT* phosphormotifs , where the phosphorylation of this gene takes place. It is a member 4 of *TBC1* family and is believed to have a role in glucose homeostasis, through controlling the transportation of glucose transporter 4 (GLUT4), activated in the presence of insulin and carry out the glucose into the fat and musculoskeletal tissues where the glucose will be stored as glycogen and used it for producing energy after the glucose oxidization. It has been shown that low expression of *TBC1D4* gene resulted in accumulation of *GLUT4* level at cellular plasma membrane indicating that this gene has a crucial role in saving the *GLUT4* level in cells. The phosphorylation of the protein domain binding to the *TBC1D4 gene* causes this gene to be separated from the glucose transporter vesicles hence these vesicles

will move toward the cell surface. Any structural aberrations in *TBC1D4* gene would lead to manifestation of Type2 diabetes. *TBC1D4* gene is highly expressed in adrenal and ovary with the RPKM values of 28.911 and 15.25 respectively. Clinically, this gene could be either benign or pathogenic since it has 45 pathogenic variances. ("Gene ID: 9882 *TBC1D4 TBC1* domain family member 4 [Homo sapiens (human)]," 2018) It is implicated in prostate cancer progression and cell growth. Increased level of androgen receptor AR would increase the glucose uptake via translocating the glucose transporter *GLUT12* to the plasma membrane. Consequently, this would increase the expression of *TBC1D4* gene, the regulator of the glucose transport. This abnormal signaling pathway of metabolism is one of the hallmarks of cancer (White et al., 2018) Another study shows that silencing of *AKT* substrate for *TBC1D4* gene would reduce the fibroblast and cancer cells proliferation and promote the cell cycle arrest in different cell types. Particularly, mediated by upregulation of cyclin-dependent kinase inhibitor at G1/S phase (P21) However P21 down regulation correlated with *AS160* overexpression which in turn will save the arrested cell and the reduced level of *AS160* in cells It has been known that phosphorylated *AS160* at *T-642* residue *(P-AS160)* correlated with increased cellular proliferation markers and the tumor size since this gene is highly expressed in breast cancer tissues compared to normal one (Gongpan, Lu, Wang, Xu, & Xiong, 2016) Our study shows that *TBC1D4* gene has high expression level of 2200 fold in complete remission. This observation will follow the same argument made for *HNRNPA3_4* and *SLC4A7_3*, that it might be a downregulation in the CP subjects than anything else.

CTC1_1 is the transcript variant number 2 for *CST* telomere replication complex component 1 (CTC1) and is lncRNA. CTC1 forms a heterodimer with *STN1*, which causes the formation of alpha enzyme involved in DNA replication. *CTC1* gene is

known to maintain the telomere structure from degradation. Different pathological conditions occur as a result of the variation of this gene. These conditions including, Cerebroretinal microangiopathy represented by the calcifications and cyst, dyskeratosis congenital also the ductal breast carcinoma ("Gene ID: 80169 *CTC1 CST* telomere replication complex component 1 [Homo sapiens (human)]," 2018) Walsh M. K. et al. (2016) discovered eight genes that were related to the leukocyte telomere length (LTL) polymorphisms. These genes were *ACYP2, TERC, NAF1, TERT, OBFC1, CTC1, ZNF208* and *RTEL1*. They found that the genetic variations associated with longer telomere length were implicated in increased risks of melanoma, glioma and lung cancer in adult, whereas, in children and adolescent, aged 12 to 19, it increase their risk of the neuroblastoma. It was observed that the leukemic cells in children were caused by a mutation in *ACD* gene, which normally acts to telomere protection (Walsh et al., 2016) we observed an expression of lncRNA of *CTC1_1* in CR group with 65 fold compared to CP. Although this gene was not significant it was also expressed in the control group and completely absent in the CP group.

GRAMD1A_6 is a predicted transcript variant (X1) of GRAM domain containing 1A (*GRAMD1A*) It is a paralog of *GRAMD1B* gene. It is highly expressed in lung and spleen tissues with the RPKM value of 13.7 and 13.3 respectively. It has seven pathogenic variances, one of which is pathogenic associated with ductal breast carcinoma. There is a limited amount of information about this gene and its role in leukemia especially in CML. However, our data shows high expression count of 260 fold in CR subjects followed by one control subjects. The absence of expression of this transcript variance in the CP group, in combination with other above mentioned 4 transcripts, might be used as diagnostic marker. Fu et al. (2016) examined the role of

GRAMD1A in the development and progression of Hepatocellular Cellular Carcinoma (HCC) by analyzing the association between the clinical conditions of 78 patients. They included in their study those with advanced stages of HCC (stage III to IV) Furthermore, they considered in their study the link between the survival time and the expression of *GRAMD1A*, using the log-rank test and they detected that patients with elevated levels of *GEAMD1A* gene had a poor prognosis. In addition to these, Fu et al. (2016) studied the relation between the *GRAMD1A* levels and the tumor relapse and growth. They found that HCC stem cells markers such as CD133 and CD90 were positively correlated with *GRAMD1A* overexpression which indicates the role of *GRAMD1A* gene in promoting the self-renewal process of HCC stem cells. Up regulated levels of *GRAMD1A* gene are positively concomitant with HCC tumor expansion and chemo-resistance. This is determined by measuring the apoptotic markers represented by *PARP* and *pro-caspase*, which are cleaved by *GEAMD1A* overexpression and simultaneous increases the anti-apoptotic protein, *BCL-X* (Fu et al., 2016) Although *GRAMD1A* is considered as a prognostic marker for HCC, it can be used as a potential therapeutic target and an indicator for the treatment improvement of psoriasis. The upregulation of *GRAMD1A* gene correlated with effective treatment of skin psoriasis plaques using narrow band-Ultraviolet B radiation (NB-UVB) concurrently with imiquimod (IMQ) ; Toll-Like receptor 7 (TLR7) agonist (Tacastacas et al., 2017).

4.2.3. Transcripts that are differentially expressed in paired comparisons between chronic phase vs complete remission and healthy subjects.

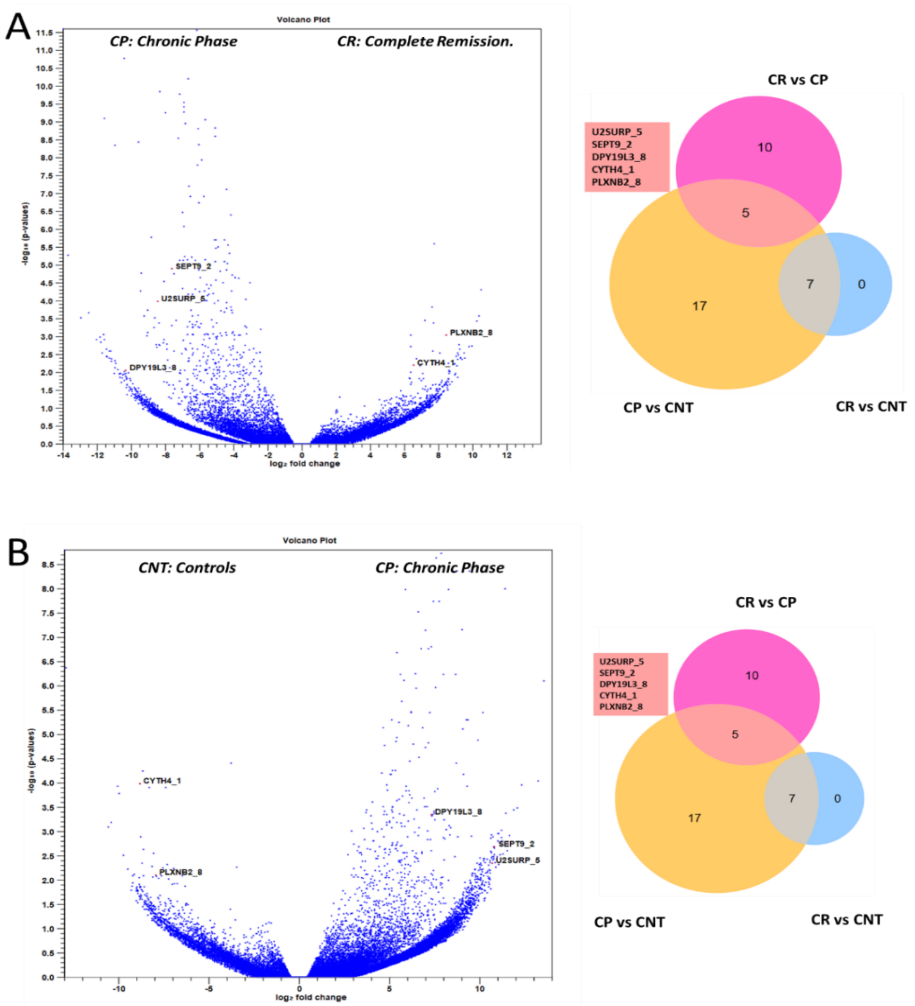


Figure 11. *Volcano plot showing transcripts selected from Venn diagram created for paired groups based on the fold change cut off of 1.5 and FDR of less than 0.0001 CNT: Controls; CP: Chronic Phase; CR: Complete Remission. Log2 fold changes are shown for 5 genes in two group comparisons: A) Chronic Phase vs Controls and B) Complete Remission vs Chronic Phase.*

In Figure 11.There are three transcripts that are differentially upregulated with significant pattern in the subjects with chronic phase only these are *U2SURP_5, SEPT9_2, and DPY19L3_8* while two of which including *CYTH4_1 and PLXNB2_8* were significantly upregulated in complete remission subjects and in healthy individuals compared to chronic phase by setting the cut off of FDR value to less than 0.0001

U2SURP_5 Is a transcript variant number 2 of *U2 snRNP* associated *SURP* domain containing *(U2SURP),* and a predicted to encode intracellular protein.

The gene is highly expressed in chronic phase of CML only. *U2SURP* is considered as a prognostic marker in many cancers and its expression is associated with favorable or unfavorable depending on the type of cancer. It is among the genes that have been identified via molecular cloning that includes those encoding spliceosomal proteins novel RNA helicases and protein isomerases and GTPases (Will & Luhrmann, 1997) Part of the 17 S U2SnRNP is Human Splicing Factor (SF3a) that interacts with the pre-mRNA branch site during the formation of the spliceosome. According to Kramer et al, subunits of 60, 66 and 120 kDa are required for the in Vitro formation for SF3a RNA interference with the process formation of the subunits from HeLa cells inhibits the splicing process. Once splicing is inhibited, especially during the chronic phase of the CML, it results in the up regulation of the *U2SURP* gene. Thus, SF3a is a constitutive splicing factor (Kramer et al., 2005).

Around 70% of the human genes have shown to have alternative splice forms. U2SURP gene introduces spliceosomal errors that would result in the formation of aberrant transcripts, not found in normal cells. Instead, they are indicative of the intrinsic property of the cancer cells. Among the cellular alterations include a change

in splice site selection in cancer cells subsequently, affecting the genes implicated in tumor progression and in cancer susceptibility. The splicing defects caused by the *U2SURP* gene can arise from either inherited or somatic mutations in the *cis*-acting regulatory elements can include branch sites, splice donor, intronic and exonic splicing silencers and enhancers. Other changes include concentration, variations in composition, altered activity of the regulatory proteins and localization (Kalnina, Zayakin, Silina, & Line, 2005) In this case, up regulation of the *U2SURP* gene, a splice variant during the chronic phase of CML results in failure in correct recognition of splice sites hence the cancer-specific splice forms. Therefore, the identification of the *U2SURP* gene as a cancer-specific splice form is important as diagnostic and prognostic biomarkers in CML. In this case, the significant up regulation of the *U2SURP_5* splice variant, a cancer-specific splice form during the chronic phase of CML with a significant increase in fold value of 2100 which may interfere with the detection of wild type transcript.

No current studies have exploited the influence of *U2SURP* gene on CML patients. However, the results obtained in the research can provide a background for future investigation on the possibility of the gene expression reducing progression of leukemic cells.

SEPT9_2 Is a transcript variant number 8 for a gene that encodes GTP-binding protein. This gene is a member of the *septin* proteins family characterized with GTPases activity and are associated with cellular cytoskeleton, such as actin and microtubules. This gene plays a role in cellular process such as cytokinesis, motility, polarity as well as controlling the cell cycle. *SEPT9* gene is equally abundant in spleen and lymph node tissues with RPKM value of 19.8 Mutant *SEPT9* gene has been implicated in certain pathogenic conditions such as hereditary neuralgic

amyotrophy (HNA) and rare genetic disease such as muscular atrophy *("Gene ID: 10801 SEPT9 septin 9* [Homo sapiens (human)]," 2018) *SEPT9* gene is implicated in acute myelomonocytic leukemia when translocated with mixed lineage Leukemia *(MLL)* gene that located on chromosome 11. *SEPT9* is one of the *MLL* fusion proteins that partners septins, nuclear and cytoplasmic proteins, and histone acetyletransferases. Furthermore, *SEPT2, SEPT5, SEPT6,* and *SEPT11* are engaged with *MLL* chromosomal rearrangement. Most of *MLL-SEPTIN* genes fusions are identified in hematological malignancies (Cerveira, Bizarro, & Teixeira, 2011) ABL1 gene rearrangements have been implicated in several leukemia especially in chronic myeloid leukemia, other types of leukemia such as T- or B-lymphoblastic leukemia and a rare type of peripheral T-cell prolymorphocytic leukemia (T-PLL) T-PLL is resulted by *SEPT9-ABL1* fusion gene, a novel fusion gene discovered by Suzuki et al., 2014 who identified the resistance of this fusion gene to tyrosine kinase inhibitors (TKIs) (Suzuki et al., 2014) According to the function of *septin* proteins in cytokinesis and controlling the cell cycle, it has been assumed that *SEPT9* gene could be a candidate suppressor genes in ovarian cancer. However, *SEPT9* gene is significantly upregulated in chronic phase of CML and the highly expression patterns of this gene is recognized in almost all chronic phase samples with 2100 fold and those patients that were resistance to tyrosine kinase inhibitors (TKIs) compared to healthy individuals and the subjects in complete remission.

DPY19L3_8 Similar to *U2SURP, DPY19L3* is a prognostic marker. This transcript is variant number 1 for the *dpy-19 like C-mannosyltransferase 3 (DPY19L3)* gene. In this study this transcript is highly upregulated in the chronic phase of CML only and the fold value of this gene was 270. *DPY19L3* has eight pathogenic variances, four of them with uncertain significance. Niwa, Suzuki, Dohmae, and Simizu (2016) suggest

that C-mannosylation is a rare type of glycosylation that is critical in regulating secretion, protein-protein interaction as well as enzymatic activity. Furthermore, this study shows that *DPY19* is a C-mannosyltransferase in Caenorhabditis elegans. In experiments involving gain and loss of functions, *DPY19L3* modifies Rspo1 at W (156) on the basis of the spectrometry. Apart from chronic CML, the gene is expressed in certain human tumor cell lines which is an indication of its malignant roles in tumorigenesis (Morishita, Suzuki, Niwa, Dohmae, & Simizu, 2017)

CYTH4_1 Is a transcript variant number (X2) of gene encoding *Cytohesin-4* protein, a member of cytohesin family known as *PSCD family*. Proteins in the *PSCD family* consist of three domains: N-terminal coiled-coil motif, central *Sec7* domain, which contains the activity of guanine- (nucleotide) exchange protein *GEP*, and C-terminal pleckstrin homology (PH) domain, which interacts with the phospholipids in cellular membrane. This gene acts on *ADP-ribosylation factors (AFRs)* activity through activation of guanine nucleotide-binding proteins participated in proteins sorting and vesicular trafficking pathways ("Gene ID: 27128 *CYTH4 cytohesin 4* [Homo sapiens (human)]," 2018) This gene is expressed ubiquitously in appendix and spleen tissues with RPKM values of 28 and 22.5 respectively. In this study, this transcript variants was highly downregulated in the CP group compared to both complete remission and control groups with a significant high expression value of 270 fold in both subjects.

PLXNB2_8 Is a transcript variant number (X6) of gene encoding *plexin B2*, It is a member of class *B of plexin 2 family*. This gene is a membrane receptor protein involved in cell migration and in axonal growth guidance processes responding to semaphorins. Semaphorins are a large family of proteins that were identified as axonal growth factors and have a role in developing the central nervous system (CNS).

PLXNB2 gene is highly expressed in renal tissue with RPKM value of 34.56 clinically this gene tends to be pathogenic ("Gene ID: 23654 PLXNB2 plexin B2 [Homo sapiens (human)]

", 2018) *PLXNB2* is one of the target genes for *miR-126-3p* and is found to be implicated in cancer pathways. It activates RhoA/ROCK signaling pathway, which plays a role in cell proliferation and transformation in esophageal squamous cell carcinoma (ESCC) and in breast cancer. Therefore, *miR-126* can inhibit the proliferation of cancer cells by targeting *PLXNB2* and decreases the expression of this gene. The downregulation of *PLXNB2* in turn inhibits the RhoA/ROCK signalling pathway. This finding was observed in colon cancer, while downregulated *miR-126-3p* lead to increased expression of *PLXNB2* followed by increased the proliferation and the severity of tumor cells, this pathogenic situation occur in lung adenocarcinoma (Q. Chen et al., 2016) In this study, this transcript variants was highly downregulated in the chronic phase subjects compared to both complete remission and control groups with a highly significant fold value of 1100 in complete remission but this gene has low significant in healthy individuals with a fold value of 130.

4.2.4. Transcripts that are differentially expressed in both chronic phase and complete remission vs healthy subjects.

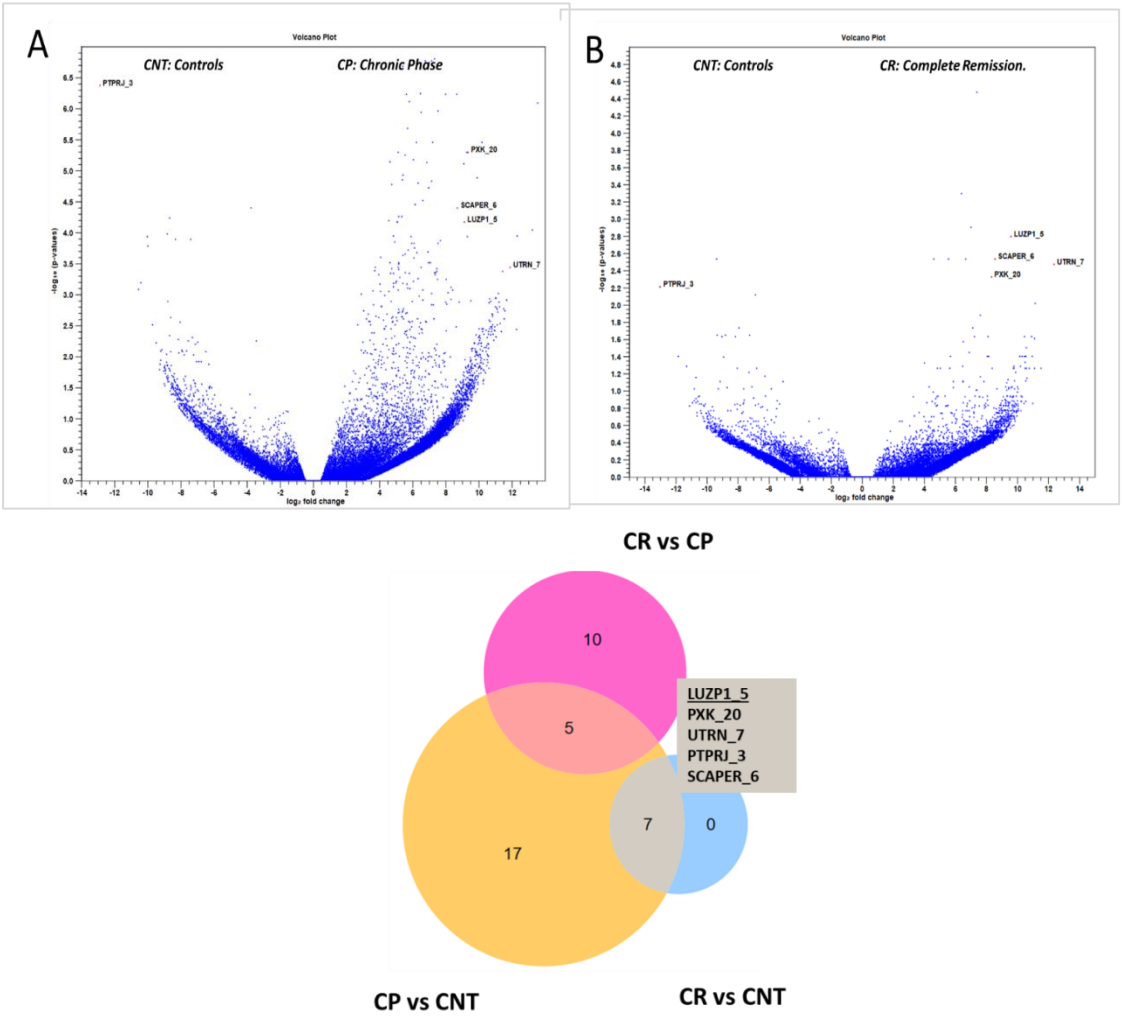


Figure 12. *Volcano plot showing transcripts selected from Venn diagram created for paired groups based on the fold change cut off of 1.5 and FDR of less than 0.0001 CNT: Controls; CP: Chronic Phase; CR: Complete Remission. Log2 fold changes are shown for 5 genes in two group comparisons: A) Chronic Phase vs Controls and B) Complete Remission vs Controls.*

Figure 12.Illustrated seven differentially expressed transcripts in chronic phase and complete remission compared to control group by setting the cut off of FDR value to less than 0.0001, there are only four transcripts which significantly overlapped between chronic phase and complete remission including: *LUZP1_5, PXK_20, UTRN_7, and SCAPER_6* while *PTPRJ_3* is the only transcript showed a stringent fold value with extreme significant in healthy individuals.

LUZP1_5 Is the transcript variant number 1 of gene that encodes *for Leucine zipper protein 1*. There are certain motif-containing protein, in which their lower expression has been associated with disease progression in cancer. *LUZP* is mostly expressed in adult's brain, and in both the small intestine and the duodenum, with RPKM value of 10.5 and 8.9 respectively. Deficiency in *LUZP* protein causes a defect in neural tube closure in mice during the brain development ("Gene ID: 7798 *LUZP1 leucine zipper protein 1* [Homo sapiens (human)]," 2018). Guangbang, Li, and Deiji (2016) pointed out that the protein encoded by *LUZP* gene plays an important role in the developmental stages. Aberrant expression or mutation of this gene could lead to cancer. Its overexpressed pattern has been shown to be correlated with different solid tumors including breast, cervical, kidney and pancreas. Elevated level of *LUZP* mRNA has been reported in Hepatocellular carcinoma (HCC) samples compared to normal liver tissues (1. 87 ± 0. 11 vs 0.58 ± 0.05, $P < 0.01$) The differences of the gene expression between these two groups were measured based on the Tumor Lymph Node metastasis (TNM) stage, disease recurrence and the portal vein invasion. Therefore, the high expression of *LUZP* gene in HCC tissues correlated with the patient's poor prognosis (G. Li, Yuan, Liu, & Liu, 2016).

Viphakone et al., (2015) shows that cancer testis antigens *(CTAs)*, known as *Luzp4*, binds to mRNA export receptor of other genes and stimulates their activity.

This class of *LUZP* protein is upregulated in several tumors particularly in melanoma (Viphakone et al., 2015). Interestingly, in our study population, *LUZP_1* is significantly expressed in both complete remission and chronic phase of chronic myeloid leukemia (CML) subjects with high fold of 1100 in both groups of CML.

PXK_20 Is the transcript variant X2 of mRNA, predicted to code a *phox (PX)* domain containing serine/threonine kinase like *(PXK)* protein which involved in synaptic transmission. Furthermore, the gene is involved in the degradation of epidermal growth factors. The variations in the regulation of *PXK* gene are associated with susceptibility to systemic lupus erythromatosus. In addition, alternative splicing leads to multiple transcript variants. This gene binds to and modulates brain Na, K-ATPase subunits ATP1b1 and ATP1b3. Not much is known about *PXK* gene. However, in our study this variant of *PXK* is significantly expressed in both complete remission and chronic phase of chronic myeloid leukemia (CML) subjects with an increased fold of 1100 in chronic phase compared to complete remission with 800 fold, In terms of significant expression, *PXK* gene has a distinct significant value in chronic phase compared to that in complete remission.

UTRN_7 Is the transcript variant X2 of mRNA, predicted to code *utrophin (UTRN) protein***.** It was annotated using gene prediction methods. *UTRN* is a member of *DMD* gene that encodes a rod-shaped protein known as *Utrophin* or *Dystrophin like protein (DMDL)* It is the large muscular protein that is located at neuromuscular synapse and myotendinous junctions. It is similar in structure and functions to *Dystrophin* gene *(DMD)* and considered as an alternative to *Dystrophin* gene. It has a role in supporting the cytoskeleton of the muscle fibres and links them to extracellular matrix proteins. *Utrophine* can be used as drug to make up the lack of endogenous protein expression of *Utrophin* that result in *Duchenne/Becker muscular dystrophy*. The

UTRN gene is highly expressed in thyroid and fat tissues with an RPKM value of 15.3 and 15.0 respectively, in fetal tissues such as placenta, heart and intestine. *UTRN* is a pathogenic gene with nine pathogenic variances. In Meta-analysis study conducted by Wang X. Y. et al., 2013, they used a total of 1174 gene expression data from two studies that were collected from advanced stages of the prostate cancer. They found eight genes to be differentially expressed and related to the progression of the prostate cancer (*NBL1, C10orf116, SMTN, PARM1, UTRN, SYNPO2, MYLK* and *PTN*) (X. Y. Wang et al., 2013) under hypoxia conditions *CLK1* and *CLK3* are regulators of the alternative splicing process. A study identified that *UTRN* gene to be one of the 12-cancer associated genes that showed a significant increase in *CLK1* and *CLK3* expression in prostate cancer cells (PC3). The elevated levels of *CLK3* expression were detected in different hypoxic cancer cell lines such as colon HT29, breast cancer MCF7 and prostate DU145 (Bowler et al., 2018) It was found that both *UTRN* gene and *SXT11* gene were involved in the pathophysiology of peripheral T-cell Lymphoma (PTCL) and therefore considered as a tumor suppressor gene of (PTCL) but not in other lymphomas such as NK-cell lymphoma and diffuse large B-cell lymphoma (DLBCL) (Yoshida et al., 2015) Another study reported that *UTRN* gene is a candidate tumor suppressor gene, which is downregulated in some primary tumors compared to normal tissues. Any mutation in *UTRN* mainly deletion has been associated with tumor cell growth and tumor progression (Y. Li et al., 2007) our findings showed that this variant of *UTRN* is significantly expressed in both complete remission and chronic phase of chronic myeloid leukemia (CML) subjects with the significant fold value of 8500 in both phases of CML compared to healthy group. No studies have reported the role of *UTRN* in CML disease.

PTPRJ_3 Is the isoform 1 precursor of receptor-type tyrosine-protein phosphatase, receptor type J, *PTPRJ*. This protein is expressed in colon carcinoma. In a study, the gene was found to be pathogenic in two of seven cases. In the present study, this variant of *PTPRJ* was significantly downregulated in both complete remission and chronic phases but it has the optimum level of significant value of 10^{-6} with the highest fold of 4100 in healthy group only. The gene is known for its role in breast cancer, lung cancer and colorectal cancer, where the loss of heterozygosity of *PTPRJ* is a common occurrence. Nevertheless, both the functions and mechanism of the gene in cancers are not well understood. In their study, Zhang, Tu, Li, Ye, and Cui (2017) were able to demonstrate that the ectopic expression of the gene leads to the inhibition of cell growth and mitigates the invasion of the colorectal cancer. When a xenograft tumor model was used, *PTPRJ* inhibited tumorigenecity of HCT116 cell line. Overexpression of *PTPRJ* minimizes the growth-promoting and AKT signaling activation of *miR-155* (X. F. Zhang, Tu, Li, Ye, & Cui, 2017) This property of the gene is an indicator of its role as tumor suppressor and accordingly, might be used as a potential therapeutic target to decrease the proliferation and migration of the cancerous cells (Jayavelu et al., 2016)

SCAPER_6. Is a transcript variant X6 and is a predicted gene encoding S-phase cyclin A associated protein in the ER (SCAPER). This gene encodes for a protein which interact with a complex of *cyclin A* and *cyclin-dependent kinase2 (CDK2)* via cyclin binding motif known as *RX* This interaction occurs at different stages of the cell cycle, particularly the transition process of the cell cycle from S to G2/M phases. Is found in the endoplasmic reticulum (ER) and in nucleus but is also considered as perinuclear protein. The unique function of the *SCAPER* protein is to maintain *cyclin A* homeostasis in the cytoplasm. Ectopic expression of the *SCAPER* protein leads to a

reduction of *cyclin A* level in the nucleus and retardation in cell cycle progress (Tsang, Wang, Chen, Sanchez, & Dynlacht, 2007) Northern blot analysis shows that *SCAPER* gene is ubiquitously expressed in all human tissues with a constant level, suggesting that *SCAPER* gene is housekeeping gene. Aberrations in *SCAPER* gene associated with 12 pathogenic variances and can be implicated in multiple pathogenic conditions such as intellectual disability, Rod-cone dystrophy, Attention deficit hyperactivity disorder and retinitis pigmentosa ("Gene ID: 49855 *SCAPER* S-phase *cyclin A* associated protein in the ER [Homo sapiens (human)]," 2018) Furthermore, the permanent binding of *SCAPER* protein to *cyclin A/CDK2 complex* is concordant with an increase of *cyclin A* level in the cytoplasm, which promotes cells growth and transformation. Therefore, *SCAPER* bounded *cyclin A* might be contributes in carcinogenesis Thus, *SCAPER* protein could be considered as a potential therapeutic target of cancer (Tsang et al., 2007) Flossbach L. et al., 2013 showed that the gain or loss mutation of *SCAPER* gene associated with small cell marginal zone B-cell lymphoma (SC MZBL) However, high level of *SCAPER* gene in this type of B-cell lymphoma can decrease the transition process between the cell cycle phases, thus reduce the cell growth and transformation to large cell lymphoma (Flossbach et al., 2013) In the present study, this variant of *SCAPER* was significantly upregulated in both complete remission and chronic phases and it shows an increased fold of 1100 in both chronic pahse and complete remission . *SCAPER* protein regulates the cell cycle progression through interacting with cyclin A/kinases complex. It is found that this gene has a dual function in chronic myeloid leukemia, such as in promoting the growth of leukemic cells. However its upregulated expression will slow down the proliferation of leukemic cells and tend to transform to complete remission phase of CML disease.

Table 7.*The promising genes (transcripts) in different CML groups of the study subjects.*

Transcript Name	CP	CR	CNT	Significant fold change values
MAZ_ 4			√	200
RIT1_3			√	600
PTPRJ_3			√	4100
BUB1_4	√			6500
U2SURP_3	√			2100
SEPT9_2	√			2100
SENP5_6	√			264
TBC1D4_8		√		2200
HNRNPA3_4		√		2200
UTRN_7	√	√		8500
LUZP1_5	√	√		1100
SCAPER_6	√	√		1100
PXK_20	√	√		CP: 1100, CR: 800
PLXNB2_8		√	√	CR: 1100, CNT:130

CHAPTER 5: CONCLUSION

Due to high rate of the treatment failure among CML patients in Qatar, profiling of genes (transcripts) expression was studied. In our cohort of CML patients, differential transcripts expression was identified within the same group of CP phase which could explain that patients respond to the treatment diversely. These significantly expressed transcripts are *SEPT9, U2SURP, BUB1, and SENP5* which could be used as prognostic biomarkers of the chronic phase of CML.

This study also provides sets of genes as promising findings that have the potential to be used to stratify CML patients into complete remission and chronic phase groups and to monitor those who are at risk of relapse these transcripts are: *LUZP1, SCAPER, UTRN, and PXK*. Transcripts that are significantly upregulted in complete remission subjects only including *TBC1D4, and HNRNPA3* these two transcripts have the potential use for following-up and to determine the treatment dose. Furthermore, some of these differentially expressed genes might be used as potential therapeutic targets. *PLAXNB2* and *CYTH4* transcripts which are significantly upregulated in both complete remission subjects and in healthy individuals could be used for monitoring the CML patients who are under treatment. *PTPRJ_3* is another promising transcript as tumor suppressor which significantly upregulated in healthy individuals only.

Limitations and Strengths

Given the small population of Qatar, close to 2 milion, and low incidence rate of CML, we were successful to recruit 13 patients from two stages of CML disease, Chronic Phase and Complete remission. However, we wished to have an equal number of subjects in each group. One of the limitations of this study was the small sample size with no patients in blast crisis (BC) phase. Another limitation is the scope

of the analysis that only focused on the transcript expression. However, not to be included in this thesis, a number of analysis is under way, where the gene expression and fusion proteins will be mapped and compared to the results presented in this study. The main strength of this project is the number of the patients recruited in the chronic phase.

Recommendations

For the future study, a longitudinal follow up of few patients may help to validate the current observation. It would also help to include those patients that are undergoing treatments. Although, very hard, including their first of kin, this might also help to distinguish inter individual variation from that of disease related. According to apparent results we gained from our study of the differential genes expression sets recognized in three groups of CML using RNA-sequencing, this will emphasize the using of this advanced method in diagnostic fields due to its high sensitivity since this technology considered as the highly dynamic method which has the capability to detect the novel genes and study them from all aspects. Also we are looking forward to recruit patients in blast crisis (BC) phase to study their transcript profile and to increase the number of the study subjects in order to translate the current findings to the clinical field. However, further accurate classification of CML is also recommended.

REFERENCES

Al-Dewik, N., Jewell, A. P., Yassin, M. A., El-Ayoubi, H. R., & Morsi, H. M. (2014). Studying the impact of presence of point mutation, insertion mutation and additional chromosomal abnormalities in chronic myeloid leukemia patients treated with imatinib mesylate in the State of Qatar. *QScience Connect*, 13.

Aoki, Y., Niihori, T., Banjo, T., Okamoto, N., Mizuno, S., Kurosawa, K., Matsubara, Y. (2013). Gain-of-function mutations in RIT1 cause Noonan syndrome, a RAS/MAPK pathway syndrome. *Am J Hum Genet, 93*(1), 173-180. doi:10.1016/j.ajhg.2013.05.021

Baltz, A. G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Landthaler, M. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell, 46*(5), 674-690. doi:10.1016/j.molcel.2012.05.021

Beermann, J., Piccoli, M.-T., Viereck, J., & Thum, T. (2016). Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches. *Physiological reviews, 96*(4), 1297-1325.

Berger, A. H., Imielinski, M., Duke, F., Wala, J., Kaplan, N., Shi, G. X., Meyerson, M. (2014). Oncogenic RIT1 mutations in lung adenocarcinoma. *Oncogene, 33*(35), 4418-4423. doi:10.1038/onc.2013.581

Bossone, S. A., Asselin, C., Patel, A. J., & Marcu, K. B. (1992). MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination. *Proc Natl Acad Sci U S A, 89*(16), 7452-7456.

Bowler, E., Porazinski, S., Uzor, S., Thibault, P., Durand, M., Lapointe, E., Ladomery, M. (2018). Hypoxia leads to significant changes in alternative splicing and elevated expression of CLK splice factor kinases in PC3 prostate cancer cells. *BMC Cancer, 18*(1), 355. doi:10.1186/s12885-018-4227-7

Cao, J. (2014). The functional role of long non-coding RNAs and epigenetics. *Biological procedures online, 16*(1), 42.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Hentze, M. W. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell, 149*(6), 1393-1406. doi:10.1016/j.cell.2012.04.031

Cerveira, N., Bizarro, S., & Teixeira, M. R. (2011). MLL-SEPTIN gene fusions in hematological malignancies. *Biol Chem, 392*(8-9), 713-724. doi:10.1515/BC.2011.072

Cheetham, S., Gruhl, F., Mattick, J., & Dinger, M. (2013). Long noncoding RNAs and the genetics of cancer. *British journal of cancer, 108*(12), 2419-2425.

Chen, Q., Hu, H., Jiao, D., Yan, J., Xu, W., Tang, X., Wang, J. (2016). miR-126-3p and miR-451a correlate with clinicopathological features of lung adenocarcinoma: The underlying molecular mechanisms. *Oncol Rep, 36*(2), 909-917. doi:10.3892/or.2016.4854

Chen, X., Yan, C. C., Zhang, X., & You, Z. H. (2017). Long non-coding RNAs and complex diseases: from experimental results to computational models. *Brief Bioinform, 18*(4), 558-576. doi:10.1093/bib/bbw060

CHRONIC MYELOGENOUS LEUKEMIA. (2014). Retrieved from http://www.who.int/selection_medicines/committees/expert/20/applications/cancer/en/

Deininger, M. W., Goldman, J. M., & Melo, J. V. (2000). The molecular biology of chronic myeloid leukemia. *Blood, 96*(10), 3343-3356.

Di Bacco, A., Keeshan, K., McKenna, S. L., & Cotter, T. G. (2000). Molecular abnormalities in chronic myeloid leukemia: deregulation of cell growth and apoptosis. *The Oncologist, 5*(5), 405-415.

Dinger, M. E., Pang, K. C., Mercer, T. R., & Mattick, J. S. (2008). Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput Biol, 4*(11), e1000176. doi:10.1371/journal.pcbi.1000176

DISEASE PROGRESSION. (2010). Retrieved from http://www.nationalcmlsociety.org/faq/disease-progression

Flis, K., Irvine, D., Copland, M., Bhatia, R., & Skorski, T. (2012). Chronic myeloid leukemia stem cells display alterations in expression of genes involved in oxidative phosphorylation. *Leuk Lymphoma, 53*(12), 2474-2478. doi:10.3109/10428194.2012.696313

Flossbach, L., Holzmann, K., Mattfeldt, T., Buck, M., Lanz, K., Held, M., Barth, T. F. (2013). High-resolution genomic profiling reveals clonal evolution and competition in gastrointestinal marginal zone B-cell lymphoma and its large cell variant. *Int J Cancer, 132*(3), E116-127. doi:10.1002/ijc.27774

Forrest, D. L., Jiang, X., Eaves, C., & Smith, C. (2008). An approach to the management of chronic myeloid leukemia in British Columbia. *Current Oncology, 15*(2), 90.

Fu, B., Meng, W., Zhao, H., Zhang, B., Tang, H., Zou, Y., Zhang, T. (2016). GRAM domain-containing protein 1A (GRAMD1A) promotes the expansion of hepatocellular carcinoma stem cell and hepatocellular carcinoma growth through STAT5. *Sci Rep, 6*, 31963. doi:10.1038/srep31963

Gene ID: 699 BUB1 mitotic checkpoint serine/threonine kinase [Homo sapiens (human)]. (2018). Pubmed Retrieved 31 May 2018 https://www.ncbi.nlm.nih.gov/gene/?term=BUB1

Gene ID: 7798 LUZP1 leucine zipper protein 1 [Homo sapiens (human)]. (2018). Pubmed Retrieved 02 June 2018 https://www.ncbi.nlm.nih.gov/gene/7798

Gene ID: 9882 TBC1D4 TBC1 domain family member 4 [Homo sapiens (human)]. (2018). Pubmed Retrieved 02 June 2018 https://www.ncbi.nlm.nih.gov/gene/9882

Gene ID: 10801 SEPT9 septin 9 [Homo sapiens (human)]. (2018). Pubmed Retrieved 02 June https://www.ncbi.nlm.nih.gov/gene/10801

Gene ID: 23654 PLXNB2 plexin B2 [Homo sapiens (human)]

(2018). Pubmed Retrieved 02 June 2018 https://www.ncbi.nlm.nih.gov/gene/23654

Gene ID: 27128 CYTH4 cytohesin 4 [Homo sapiens (human)]. (2018). Pubmed Retrieved 02 June 2018 https://www.ncbi.nlm.nih.gov/gene/27128

Gene ID: 49855 SCAPER S-phase cyclin A associated protein in the ER [Homo sapiens (human)]. (2018). Pubmed Retrieved 02 June 2018 https://www.ncbi.nlm.nih.gov/gene/?term=SCAPER

Gene ID: 51347 TAOK3 TAO kinase 3 [Homo sapiens (human)]. (2018). Pubmed Retrieved 31 May 2018 https://www.ncbi.nlm.nih.gov/gene/51347

Gene ID: 80169 CTC1 CST telomere replication complex component 1 [Homo sapiens (human)]. (2018). Pubmed Retrieved 02 June 2018 https://www.ncbi.nlm.nih.gov/gene/80169

Gene ID: 205564 SUMO specific peptidase 5 [Homo sapiens (human)]. (2018). Pubmed Retrieved 31 May 2018 https://www.ncbi.nlm.nih.gov/gene/?term=SENP5

Gomez-Segui, I., Makishima, H., Jerez, A., Yoshida, K., Przychodzen, B., Miyano, S., Maciejewski, J. P. (2013). Novel recurrent mutations in the RAS-like GTP-binding gene RIT1 in myeloid malignancies. *Leukemia, 27*(9), 1943-1946. doi:10.1038/leu.2013.179

Gongpan, P., Lu, Y., Wang, F., Xu, Y., & Xiong, W. (2016). AS160 controls eukaryotic cell cycle and proliferation by regulating the CDK inhibitor p21. *Cell Cycle, 15*(13), 1733-1741. doi:10.1080/15384101.2016.1183853

Green, P. A. B. a. A. R. (2010). Myeloproliferative neoplasms. Wendy N.Erber. Retrieved from https://www.cambridge.org/core/books/diagnostic-techniques-in-hematological-malignancies/myeloproliferative neoplasms/700753262D8D064904E44DF6429D5D0F

Hatfield, K. J., Reikvam, H., & Bruserud, O. (2014). Identification of a subset of patients with acute myeloid leukemia characterized by long-term in vitro proliferation and altered cell cycle regulation of the leukemic cells. *Expert Opin Ther Targets, 18*(11), 1237-1251. doi:10.1517/14728222.2014.957671

Herrmann, H., Sadovnik, I., Cerny-Reiterer, S., Rülicke, T., Stefanzl, G., Willmann, M., Herndlhofer, S. (2014). Dipeptidylpeptidase IV (CD26) defines leukemic stem cells (LSC) in chronic myeloid leukemia. *Blood, 123*(25), 3951-3962.

Höglund, M., Sandin, F., & Simonsson, B. (2015). Epidemiology of chronic myeloid leukaemia: an update. *Annals of hematology, 94*(2), 241-247.

Homer, N., & Nelson, S. F. (2010). Improved variant discovery through local re-alignment of short-read next-generation sequencing data using SRMA. *Genome Biol, 11*(10), R99. doi:10.1186/gb-2010-11-10-r99

Jabbour, E., Fullmer, A., Cortés, J. E., & Kantarjian, H. (2010). Clinical algorithms for the treatment of patients with chronic myeloid leukemia: the 2010 perspective. *Clinical Lymphoma Myeloma and Leukemia, 10*, S6-S13.

Jain, P., Kantarjian, H., Patel, K. P., Gonzalez, G. N., Luthra, R., Shamanna, R. K., Kadia, T. M. (2016). Impact of BCR-ABL transcript type on outcome in patients with chronic-phase CML treated with tyrosine kinase inhibitors. *Blood, 127*(10), 1269-1275.

Jayavelu, A. K., Muller, J. P., Bauer, R., Bohmer, S. A., Lassig, J., Cerny-Reiterer, S., . . . Bohmer, F. D. (2016). NOX4-driven ROS formation mediates PTP inactivation and cell transformation in FLT3ITD-positive AML cells. *Leukemia, 30*(2), 473-483. doi:10.1038/leu.2015.234

Jemal, A., Siegel, R., Xu, J., & Ward, E. (2010). Cancer statistics, 2010. *CA: a cancer journal for clinicians, 60*(5), 277-300.

Jin, Z. L., Pei, H., Xu, Y. H., Yu, J., & Deng, T. (2016). The SUMO-specific protease SENP5 controls DNA damage response and promotes tumorigenesis in hepatocellular carcinoma. *Eur Rev Med Pharmacol Sci, 20*(17), 3566-3573.

Kalnina, Z., Zayakin, P., Silina, K., & Line, A. (2005). Alterations of pre-mRNA splicing in cancer. *Genes Chromosomes Cancer, 42*(4), 342-357. doi:10.1002/gcc.20156

Kantarjian, H., O'Brien, S., Jabbour, E., Garcia-Manero, G., Quintas-Cardama, A., Shan, J., . . . Kadia, T. (2012). Improved survival in chronic myeloid leukemia since the introduction of imatinib therapy: a single-institution historical experience. *Blood, 119*(9), 1981-1987.

Kantarjian, H. M., Larson, R. A., Cortés, J. E., Deering, K. L., & Mauro, M. J. (2013). Current practices in the management of chronic myeloid leukemia. *Clinical Lymphoma Myeloma and Leukemia, 13*(1), 48-54.

Kapusta, A., & Feschotte, C. (2014). Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends in Genetics, 30*(10), 439-452.

Key Statistics for Chronic Myeloid Leukemia. (2018). Retrieved from https://www.cancer.org/cancer/chronic-myeloid-leukemia/about/statistics.html

Klco, J. M., Vij, R., Kreisel, F. H., Hassan, A., & Frater, J. L. (2010). Molecular pathology of myeloproliferative neoplasms. *Am J Clin Pathol, 133*(4), 602-615. doi:10.1309/AJCPPPZ1WFVGNE4A

Koptyra, M., Falinski, R., Nowicki, M. O., Stoklosa, T., Majsterek, I., Nieborowska-Skorska, M., Skorski, T. (2006). BCR/ABL kinase induces self-mutagenesis via reactive oxygen species to encode imatinib resistance. *Blood, 108*(1), 319-327. doi:10.1182/blood-2005-07-2815

Korenbaum, E., Olski, T. M., & Noegel, A. A. (2001). Genomic organization and expression profile of the parvin family of focal adhesion proteins in mice and humans. *Gene, 279*(1), 69-79.

Kramer, A., Ferfoglia, F., Huang, C. J., Mulhaupt, F., Nesic, D., & Tanackovic, G. (2005). Structure-function analysis of the U2 snRNP-associated splicing factor SF3a. *Biochem Soc Trans, 33*(Pt 3), 439-442. doi:10.1042/BST0330439

Kukurba, K. R., & Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harbor Protocols, 2015*(11), pdb. top084970.

Kung, J. T., Colognori, D., & Lee, J. T. (2013). Long noncoding RNAs: past, present, and future. *Genetics, 193*(3), 651-669.

Kurzrock, R., Kantarjian, H. M., Druker, B. J., & Talpaz, M. (2003). Philadelphia chromosomepositive leukemias: from basic mechanisms to molecular therapeutics. *Annals of internal medicine, 138*(10), 819-830.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*, 323. doi:10.1186/1471-2105-12-323

Li, G., Yuan, L., Liu, D., & Liu, J. (2016). Upregulation of Leucine Zipper Protein mRNA in Hepatocellular Carcinoma Associated With Poor Prognosis. *Technol Cancer Res Treat, 15*(3), 517-522. doi:10.1177/1533034615587432

Li, Y., Huang, J., Zhao, Y. L., He, J., Wang, W., Davies, K. E., Xiao, S. (2007). UTRN on chromosome 6q24 is mutated in multiple tumors. *Oncogene, 26*(42), 6220-6228. doi:10.1038/sj.onc.1210432

Li, Y., & Wang, X. (2016). Role of long noncoding RNAs in malignant disease. *Molecular medicine reports, 13*(2), 1463-1469.

Marum, J. E., Yeung, D. T., Purins, L., Reynolds, J., Parker, W. T., Stangl, D., Schreiber, A. W. (2017). ASXL1 and BIM germ line variants predict response and identify CML patients with the greatest risk of imatinib failure. *Blood Advances, 1*(18), 1369-1381.

Morishita, S., Suzuki, T., Niwa, Y., Dohmae, N., & Simizu, S. (2017). Dpy-19 like 3-mediated C-mannosylation and expression levels of RPE-spondin in human tumor cell lines. *Oncol Lett, 14*(2), 2537-2544. doi:10.3892/ol.2017.6465

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods, 5*(7), 621-628. doi:10.1038/nmeth.1226

Mueller, O., Lightfoot, S., & Schroeder, A. (2004). RNA integrity number (RIN) standardization of RNA quality control. *Agilent application note, publication*, 1-8.

O'Brien, S., Abboud, C. N., Akhtari, M., Altman, J., Berman, E., DeAngelo, D. J., National Comprehensive Cancer, N. (2012). Chronic myelogenous leukemia. *J Natl Compr Canc Netw, 10*(1), 64-110.

Pagotto, S., Veronese, A., Soranno, A., Lanuti, P., Marco, M. D., Russo, M. V., Visone, R. (2018). Hsa-miR-155-5p drives aneuploidy at early stages of cellular transformation. *Oncotarget, 9*(16), 13036-13047. doi:10.18632/oncotarget.24437

Paul, M. K., & Mukhopadhyay, A. K. (2004). Tyrosine kinase–role and significance in cancer. *International journal of medical sciences, 1*(2), 101.

Raman, M., Earnest, S., Zhang, K., Zhao, Y., & Cobb, M. H. (2007). TAO kinases mediate activation of p38 in response to DNA damage. *EMBO J, 26*(8), 2005-2014. doi:10.1038/sj.emboj.7601668

RIT1 gene Ras like without CAAX 1. (2018). Retrieved from https://ghr.nlm.nih.gov/gene/RIT1

Rizzieri, D., & Moore, J. O. (2012). Implementation of management guidelines for chronic myeloid leukemia: perspectives in the United States. *Pharmacy and Therapeutics, 37*(11), 640.

Roberts A, P. L. (2012). eXpress – a tool for quantification of RNA-Seq data. Retrieved from http://www.rna-seqblog.com/express-a-tool-for-quantification-of-rna-seq-data/

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol, 11*(3), R25. doi:10.1186/gb-2010-11-3-r25

Romero-Garcia, S., Prado-Garcia, H., & Lopez-Gonzalez, J. S. (2014). Transcriptional analysis of hnRNPA0, A1, A2, B1, and A3 in lung cancer cell lines in response to acidosis, hypoxia, and serum deprivation conditions. *Experimental lung research, 40*(1), 12-21.

Sacha, T. (2014). Imatinib in chronic myeloid leukemia: an overview. *Mediterranean journal of hematology and infectious diseases, 6*(1).

Schmitt, A. M., & Chang, H. Y. (2016). Long noncoding RNAs in cancer pathways. *Cancer cell, 29*(4), 452-463.

Shet, A. S., Jahagirdar, B. N., & Verfaillie, C. M. (2002). Chronic myelogenous leukemia: mechanisms underlying disease progression. *Leukemia, 16*(8), 1402-1411. doi:10.1038/sj.leu.2402577

Siegel, R., Naishadham, D., & Jemal, A. (2013). Cancer statistics, 2013. *CA: a cancer journal for clinicians, 63*(1), 11-30.

Singh, S. D. (2017). Surveillance for cancer incidence and mortality—United States, 2013. *MMWR. Surveillance Summaries, 66*.

Song, J., Ugai, H., Ogawa, K., Wang, Y., Sarai, A., Obata, Y., Yokoyama, K. K. (2001). Two consecutive zinc fingers in Sp1 and in MAZ are essential for interactions with cis-elements. *J Biol Chem, 276*(32), 30429-30434. doi:10.1074/jbc.M103968200

Suzuki, R., Matsushita, H., Kawai, H., Matsuzawa, H., Tsuboi, K., Watanabe, S., Ando, K. (2014). Identification of a novel SEPT9-ABL1 fusion gene in a patient with T-cell prolymphocytic leukemia. *Leuk Res Rep, 3*(2), 54-57. doi:10.1016/j.lrr.2014.06.004

Tacastacas, J. D., Oyetakin-White, P., Soler, D. C., Young, A., Groft, S., Honda, K., . . . McCormick, T. S. (2017). Does imiquimod pretreatment optimize 308-nm excimer laser (UVB) therapy in psoriasis patients? *Photodermatol Photoimmunol Photomed, 33*(4), 193-202. doi:10.1111/phpp.12299

Taoka, Y., Matsumoto, K., Ohashi, K., Minamida, S., Hagiwara, M., Nagi, S., Iwamura, M. (2015). Protein expression profile related to cisplatin resistance in bladder cancer cell lines detected by two-dimensional gel electrophoresis. *Biomedical Research, 36*(4), 253-261.

Tefferi, A., & Vardiman, J. W. (2008). Classification and diagnosis of myeloproliferative neoplasms: the 2008 World Health Organization criteria and point-of-care diagnostic algorithms. *Leukemia, 22*(1), 14-22. doi:10.1038/sj.leu.2404955

Tsang, W. Y., Wang, L., Chen, Z., Sanchez, I., & Dynlacht, B. D. (2007). SCAPER, a novel cyclin A-interacting protein that regulates cell cycle progression. *J Cell Biol, 178*(4), 621-633. doi:10.1083/jcb.200701166

Viphakone, N., Cumberbatch, M. G., Livingstone, M. J., Heath, P. R., Dickman, M. J., Catto, J. W., & Wilson, S. A. (2015). Luzp4 defines a new mRNA export pathway in cancer cells. *Nucleic Acids Res, 43*(4), 2353-2366. doi:10.1093/nar/gkv070

Walsh, K. M., Whitehead, T. P., de Smith, A. J., Smirnov, I. V., Park, M., Endicott, A. A., Wiemels, J. L. (2016). Common genetic variants associated with telomere length confer risk for neuroblastoma and other childhood cancers. *Carcinogenesis, 37*(6), 576-582. doi:10.1093/carcin/bgw037

Wang, K. C., & Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Molecular cell, 43*(6), 904-914.

Wang, X. Y., Hao, J. W., Zhou, R. J., Zhang, X. S., Yan, T. Z., Ding, D. G., & Shan, L. (2013). Meta-analysis of gene expression data identifies causal genes for prostate cancer. *Asian Pac J Cancer Prev, 14*(1), 457-461.

Wee, Y., Liu, Y., Lu, J., Li, X., & Zhao, M. (2018). Identification of novel prognosis-related genes associated with cancer using integrative network analysis. *Sci Rep, 8*(1), 3233. doi:10.1038/s41598-018-21691-5

White, M. A., Tsouko, E., Lin, C., Rajapakshe, K., Spencer, J. M., Wilkenfeld, S. R., . . . Frigo, D. E. (2018). GLUT12 promotes prostate cancer cell growth and is regulated by androgens and CaMKK2 signaling. *Endocr Relat Cancer, 25*(4), 453-469. doi:10.1530/ERC-17-0051

Will, C. L., & Luhrmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr Opin Cell Biol, 9*(3), 320-328.

Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes & development, 23*(13), 1494-1504.

Yang, Y., Junjie, P., Sanjun, C., & Ma, Y. (2017). Long non-coding RNAs in Colorectal Cancer: Progression and Future Directions. *Journal of Cancer, 8*(16), 3212.

Yoshida, N., Tsuzuki, S., Karube, K., Takahara, T., Suguro, M., Miyoshi, H., Seto, M. (2015). STX11 functions as a novel tumor suppressor gene in peripheral T-cell lymphomas. *Cancer Sci, 106*(10), 1455-1462. doi:10.1111/cas.12742

Yoshimi, R., Yamaji, S., Suzuki, A., Mishima, W., Okamura, M., Obana, T., Ishigatsubo, Y. (2006). The gamma-parvin-integrin-linked kinase complex is critically involved in leukocyte-substrate interaction. *J Immunol, 176*(6), 3611-3624.

Zhang, W., Chen, T., Wan, T., He, L., Li, N., Yuan, Z., & Cao, X. (2000). Cloning of DPK, a novel dendritic cell-derived protein kinase activating the ERK1/ERK2 and JNK/SAPK pathways. *Biochem Biophys Res Commun, 274*(3), 872-879. doi:10.1006/bbrc.2000.3244

Zhang, X. F., Tu, R., Li, K., Ye, P., & Cui, X. (2017). Tumor Suppressor PTPRJ Is a Target of miR-155 in Colorectal Cancer. *J Cell Biochem, 118*(10), 3391-3400. doi:10.1002/jcb.25995

Zhao, Z.-J., Yuan, J., Su, J., Wang, X.-Y., Liu, M.-M., Yuan, X.-R., Wu, M.-H. (2016). Role of long non-coding RNAs in cancer and associated signal transduction pathways. *INTERNATIONAL JOURNAL OF CLINICAL AND EXPERIMENTAL PATHOLOGY, 9*(8), 7779-7792.

Zunino, R., Schauss, A., Rippstein, P., Andrade-Navarro, M., & McBride, H. M. (2007). The SUMO protease SENP5 is required to maintain mitochondrial morphology and function. *J Cell Sci, 120*(Pt 7), 1178-1188. doi:10.1242/jcs.03418