MESSENGER RNA AND MICRO-RNA EXPRESSION SIGNATURES IN ENDEMIC BURKITT LYMPHOMA PATIENTS FROM WESTERN KENYA

BY

ODUOR CLIFF ISAYA, MSc

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN MEDICAL GENETICS

SCHOOL OF PUBLIC HEALTH AND COMMUNITY DEVELOPMENT

MASENO UNIVERSITY

DECLARATION

I declare that this thesis is my original work and has not been presented to any other University or Institution for a degree or any other award.

Oduor Cliff Isaya
PHD/PH/00155/2013

Signature..... Date.....

This thesis has been submitted for examination with our approval as supervisors:

Dr. Kiprotich Chelimo (PhD)
Department of Biomedical Sciences and Technology,
Maseno University,
Maseno, Kenya.

Signature…… Date……

Prof. Jeffrey A. Bailey (MD, PhD)
Division of Transfusion Medicine, Department of Medicine and
Program in Bioinformatics and Integrative Biology,
University of Massachusetts Medical School,
Worcester, MA, USA.

Signature…… Date……

ACKNOWLEDGEMENT

This thesis would have not been possible without the help and guidance of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study. My utmost gratitude goes to Prof. Jeffrey Bailey for sponsoring my studies and for his guidance in the course of my project. I am also grateful to Dr. Kiprotich Chelimo, who has been my supervisor and academic mentor, has encouraged me to pursue this topic and spent extra time helping me to achieve a clear structure. Equally, much gratitude goes to Prof. Ann Moormann, who has also been my supervisor, for her steadfast encouragement to complete this study and for the valuable insights she has shared. Together they sacrificed much of their valuable time to discuss the project and thesis at various stages offering constructive criticism and helpful comments. Consequently, their support and encouragement has been priceless to the success of this thesis.

I would wish to recognize and appreciate the efforts of Yasin Kaymaz for his support that made this work a success. Last but not least I thank, my parents, brother, wife and one above all of us, God almighty, for giving me the strength to complete this study.

DEDICATION

I dedicate this thesis to my mother, Joyce Akinyi, in appreciation for her constant encouragement and overwhelming support.

ABSTRACT

Endemic Burkitt lymphoma (eBL) is a B cell tumour common in region with high *Plasmodium falciparum* malaria and is characterized by over expression of the *c-myc* oncogene as a consequence of the *t(8:14)* IGH/MYC translocation. While *MYC* is the seminal event, eBL, like all cancers, is a complex amalgam of genetic and epigenetic changes causing dysregulation of both coding and non-coding transcripts. Emerging evidence suggest that abnormal modulation of mRNA transcription via microRNAs (miRNAs) might be a significant factor in lymphomagenesis. However, the alterations in these miRNAs and their correlations to their putative mRNA targets have not been extensively studied in eBL tumour cells relative to its normal germinal center (GC) B cells. Western Kenya is one of the high malaria transmission regions in Kenya with high cases of eBL in children. Hence comparing microRNA and mRNA expression profiling in eBL tumours from children in western Kenya relative to GC B cells, may partially explain molecular events involved in eBL oncogenesis. While many transcriptional changes in the oncogenic pathway of eBL have been characterized, comparisons of transcripts expression between eBL and GC B cells have not been performed using more sensitive and specific mRNA and micro-RNA transcriptome deep sequencing. This study aimed at determining the mRNA and miRNA expression profile of eBL tumours from children in western Kenya compared to GC B cells and to identify miRNAs that could contribute to eBL clinical presentation, and poor outcomes. Using the generalized linear model implemented in edgeR package in R, this case-control study compared previously published small and long RNA of a set of 5 GC B cells and 28 eBL tumours (12 jaw and 16 abdominal tumours) and identified 211 differentially expressed (DE) mRNAs (79 upregulated and 132 downregulated) and 49 DE miRNAs (22 up-regulated and 27 down-regulated) in eBL tumour cells. Of these DE genes, the study identified the downregulation of *ATM* and *NLK* genes, which represent important regulators in response to DNA damage in eBL tumour cells. These tumour suppressors were targeted by multiple upregulated miRNAs (miR-19b-3p, miR-26a-5p, miR-30b-5p, miR-92a-5p and miR-27b-3p) which could account for their aberrant expression in eBL. One miRNA (miR-10a-5p) was found to be downregulated in eBL jaw tumours and among the non-survivors. Decreased miR-10a expression, relives post-transcriptional regulation to its targets genes (such as CD59 and API5) that would promote tumour cell survival and apoptosis evasion leading to poor eBL patient outcome. Combined loss of p53 induction and function due to miRNA-mediated regulation of *ATM* and *NLK*, together with the observed upregulation of *TFAP4*, may play a central role for human miRNAs in eBL oncogenesis. This would facilitate survival of eBL tumour cells with the IGH/MYC chromosomal translocation and promote MYC-induced cell cycle progression, initiating eBL lymphomagenesis. Given the validated target genes of miR-10a-5p, its altered expression pattern between different tumour sites (jaw versus abdominal tumours) in eBL and in eBL patients with different in-hospital survival outcomes would enhance tumour cell survival and thus, render the tumour less sensitive to chemotherapy. This comprehensive characterization of mRNA and miRNA transcriptomes in eBL relative to GC B cells and among the eBL tumours provides new insights into miRNA-mediated transcript regulation in eBL, which are potentially useful for new improved therapeutic strategies.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

BH	-	Benjamini and Hochberg
cDNA	-	Complementary DNA
CPM	-	Copies per million
DDR	-	DNA damage response
DE	-	Differentially expressed
eBL	-	Endemic Burkitt lymphoma
EBV	-	Epstein Barr virus
edgeR	-	Empirical Analysis of Digital Gene Expression Data in R
FC	-	Fold Change
FDR	-	False Discovery Rate
GC	-	Germinal Center
GTP	-	Guanosine triphosphate
miR	-	Mature microRNA
mir	-	MicroRNA gene
miR*	-	Passenger microRNA
miRNA	-	MicroRNA
MiRNA-seq	-	microRNA sequencing
mRNA	-	Messenger RNA
ncRNA	-	Non-coding RNA
NGS	-	Next generation sequencing
PCR	-	Polymerase chain reaction
pre-miRNA	-	Precursor microRNA
pri-miRNA	-	Primary microRNA

RISC	-	RNA-induced silencing complex
RNA	-	Ribonucleic acid
RNA-seq	-	RNA sequencing
RT-qPCR	-	Quantitative Reverse Transcription Polymerase Chain Reaction
SNP	-	Single Nucleotide Polymorphism
sva	-	Surrogate Variable Analysis
UTR	-	Untranslated region

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1. Background Information

Endemic Burkitt lymphoma (eBL) is a germinal center (GC) B-cell cancer occurring at a high incidence in sub-Saharan Africa and the most prevalent paediatric cancer in western Kenya. This paediatric cancer was first described by Denis Burkitt and was associated with rainfall and increased *P. falciparum* malaria prevalence (Burkitt, 1958; Burkitt and O'Conor, 1961). In Kenya it is highly prevalent at the Coast and Lake Victoria regions where *P. falciparum* malaria is holoendemic. While it is generally sensitive to cytotoxic chemotherapies, some tumours remain or become refractory contributing to poor outcomes in clinical cases (Buckle *et al.*, 2016). It is therefore critical to elucidate molecular mechanisms (such as gene regulation processes) involved in eBL pathogenesis in order to identify molecular targets for both early detection, prognostic indicators, and more effective therapy to improve treatment outcomes.

Burkitt lymphoma is subdivided into an EBV-associated endemic form (eBL) in Africa (also in New Guinea), a sporadic form (sBL) that is most prevalent in developed countries, and an HIV-associated or immunodeficiency-related BL form (id-BL). All forms of BL are characterized by overexpression of the *MYC* gene, a transcription factor and proto-oncogene, which has roles in cell cycle progression, apoptosis and B cell transformation (Ott *et al.*, 2013). This overexpression is most often a consequence of a translocation involving chromosomes 8 and 14 approximating the *IGH* enhancer to an intact *MYC* locus (Dalla-Favera *et al.*, 1982; Zech *et al.*, 1976). Simple overexpression of *MYC* is not in itself transformative in normal cells, as multiple mechanisms and checkpoints exist that counteract aberrant *MYC* expression

by triggering apoptosis (Gabay *et al.*, 2014; Spencer and Groudine, 1991). This suggests that there are likely additional genetic and epigenetic changes that fully potentiate the oncogenic transformation. This multi-factorial concept has been strongly supported by a number of studies demonstrating further driver mutations and epigenetic changes, that play important roles in tumour proliferation, maintenance and abrogating checkpoints in the face of *MYC* overexpression (Kaymaz *et al.*, 2017; Kretzmer *et al.*, 2015; Love *et al.*, 2012; Schmitz *et al.*, 2012). However, the exact pattern and combinations of abnormal epigenetic modulations, via mRNAs and microRNAs (miRNAs), necessary or sufficient for eBL lymphomagenesis has not been fully elucidated.

Endemic BL, like all other forms of BL, is thought to originate from GC B cells based on the expression of V-region genes diversified by somatic mutations in conjunction with its extra-nodal presentation (Klein *et al.*, 1995). A GC program is supported by the detection of somatic mutations in the rearranged V region genes that are characteristic of GC B-cell differentiation (Dave et al., 2006; Klein *et al.*, 1995). While it is unclear if BL cells truly traverse the GC, it is clear that GC B cells are their best normal counterpart and that BL is likely an oncogenically altered GC program (Schmitz *et al.*, 2014) in which GC-restricted transcription factors have powerful oncogenic influence. The expression of protein coding genes and polyadenylated transcripts have provided key insights into tumour dysregulation (Schmitz *et al.*, 2014). However, transcriptome expression differences that would facilitate oncogenesis have not been fully explored in eBL. MicroRNAs are key transcriptome components that have not been examined in primary tumours but, may contribute to altered gene expression leading to lymphomagenesis.

MicroRNAs (miRNAs) are a recently discovered class of small noncoding RNAs with 18 to 24 nucleotides that regulate gene expression post-transcriptionally by binding to mRNAs

with complementarity (Esteller, 2011; Liu *et al.*, 2007). They have been described as managers of gene expression by targeting mRNAs for degradation or translational repression and influences many cellular processes including proliferation, apoptosis, and cell survival that are often key in oncogenesis (Croce, 10/2009). Dysregulation of miRNAs have been found to initiate malignant phenotypes, resulting in development of various cancers (Calin and Croce, 2006; Sassen *et al.,* 2007). MiRNA expression profiling studies can be especially rich in biological information, as variations in expression of hundreds of protein-coding genes may be captured in the expression patterns of one or a few miRNAs that regulate them (Calin and Croce, 2006; Sassen *et al.,* 2007). To date, the global miRNA and mRNA expression patterns of eBL have not been interrogated. An evaluation of aberrant miRNA and mRNA expression changes in eBL, compared to its normal counterpart, could provide an insight into mechanisms involved in eBL genesis and progression. The identification of oncomirs and tumour suppressor miRs, may lead to the development of novel therapeutic agents targeting miRNAs via mimics or antagomirs.

Endemic BL is an aggressive lymphoma that can present in a number of different anatomical locations including the jaw, abdomen, orbital area, central nervous system and breast or a combination of these sites (Mwanda, 2004; Ogwang *et al.,* 2008). However, jaw and abdominal tumours are the most common anatomical sites of presentation (50-80% of cases) in paediatric eBL. Since miRNAs are involved in the initiation, progression and metastasis of human cancers (Liu *et al.*, 2007), miRNA profiling in a variety of cancers has not only identified cancer deregulated miRNAs, but also profiles that correlate with prognosis, staging, and response to treatment (Cho, 2012). However, those miRNAs which are associated with tumour localization, progression and response to treatment in eBL remain unknown. Apart from the obvious tissue difference between the clinical presentations of eBL, aberrant miRNA expression may be involved in molecular differences such as amplified metabolic

processes that could favour survival or be predictive indicators of poor prognosis. These miRNA expression changes could act as novel biomarker targets that would aid in identifying subsets of eBL patients at risk of relapse occurrence.

A combined analysis of detailed expressions patterns of mRNA and miRNA in eBL can be performed using more sensitive and specific next-generation deep sequencing. An integrative analysis of differentially regulated miRNA and mRNA expression in eBL tumours compared to GC B cells and miR expression patterns associated with tumour localization and patient outcome, will help in elucidating the mechanisms involved in eBL oncogenesis and identify key miRNAs and miRNA-mRNA interactions that may underlie eBL lymphomagenesis and patient outcome.

1.2. Statement of the Problem

Endemic Burkitt lymphoma is the most common paediatric cancer in sub-Saharan Africa and occurs at a high incidence in western Kenya, a region that is already burdened by high prevalence of *P. falciparum* malaria (Rainey *et al.,* 2007). This tumour has the highest proliferative rate of any other human cancer with a doubling time of 24 - 26 hours (Chêne, 2009) and has 100% mortality rate if not treated. The *c-myc* oncogene translocation *t(8:14)* results in its overexpression and is the hallmark of BL (Hecht and Aster, 2000). However, it has been shown that even children lacking this translocation still develop this cancer due to deregulation of the microRNAome that initiate *c-myc* overexpression (Onnis *et al.,* 2010). This highlights the importance of the entire transcriptome in eBL pathogenesis, which remains to be fully elucidated.

Aberrant *MYC* overexpression alone, is not in itself transformative in normal cells as multiple mechanisms and checkpoints exist that counteract aberrant *MYC* expressions to initiate apoptotic response (Gabay *et al.,* 2014; Spencer and Groudine, 1991). This suggests that there are likely additional genetic and epigenetic changes, such as aberrant gene expression, that fully potentiate the oncogenic transformation in eBL lymphomagenesis.

The factors underlying the development of relapse and occurrence of death after first line treatment among a small percentage of eBL cases still remain unknown. Consequently, there is need for the identification of prognostic markers to identify the subset of eBL patients, based on tumour presentation site (jaw and abdominal tumours), eligible for more aggressive chemotherapy treatment to prevent relapses and death.

1.3. Objectives

1.3.1. General objectives

To investigate the role of the mRNA and microRNA expression in the molecular pathogenesis, clinical presentation, and prognosis of endemic Burkitt lymphoma in western Kenya.

1.3.2. Specific objectives

i. To identify mRNA expression signatures associated with endemic Burkitt lymphoma in children from western Kenya compared to germinal center B cells.

ii. To identify miRNA expression signatures associated with endemic Burkitt lymphoma in children from western Kenya compared to germinal center B cells.

iii. To compare miRNA expression differences between clinical phenotypes (jaw versus abdomen, and survivors versus non-survivors) in eBL cases from western Kenya.

1.3.3. Research Questions

i. What is the mRNA expression profile associated with endemic Burkitt lymphoma in children from western Kenya compared germinal center B cells?

ii. What is the miRNA expression profile associated with endemic Burkitt lymphoma in children from western Kenya compared to germinal center B cells?

iii. What are the differences in miRNA expression between clinical phenotypes (jaw versus abdomen, and survivors versus non-survivors) in eBL cases from western Kenya?

1.4. Justification of the Study

Almost all tumour cells display differentially expressed miRNA and mRNA expression profiles compared to normal cells. The full extent and function of miRNA expression in most B cell malignancies such as eBL is unknown. The complete description of miRNA and their target gene expression in normal GC B cells and eBL tumour cells is essential in realizing the role of miRNAs in eBL pathogenesis and progression. There is evidence that in eBL, 30% of tumours do not respond to conventional chemotherapy resulting in death of these children. These poor outcomes among eBL patients could be attributed to particular tumour presentation sites. Moreover, 5% of treated children may initially respond to treatment but then relapse within 12 months from primary diagnosis (Hesseling *et al.*, 2008). The molecular events underlying eBL clinical presentation, relapse occurrence and refractory tumours are unknown and was also a major focus of this study.

1.5. Significance of the Study

MicroRNA and mRNAs dysregulated in eBL compared to their normal counterpart provide valuable information which aids in system-level understanding of gene regulatory network initiating eBL's malignant phenotype. Including miRNA and mRNA profiling as an additional tool to understanding the complexities of gene regulatory networks initiating development of eBL, would aid in identifying new targets in its treatment with little or no toxicities and reduced possibilities of relapse occurrences. Furthermore, microRNA expression patterns may be used as biomarkers to identify children at risk of developing eBL and also have prognostic significance in the future to determine and subsets of eBL patients at risk of poor outcome.

CHAPTER TWO

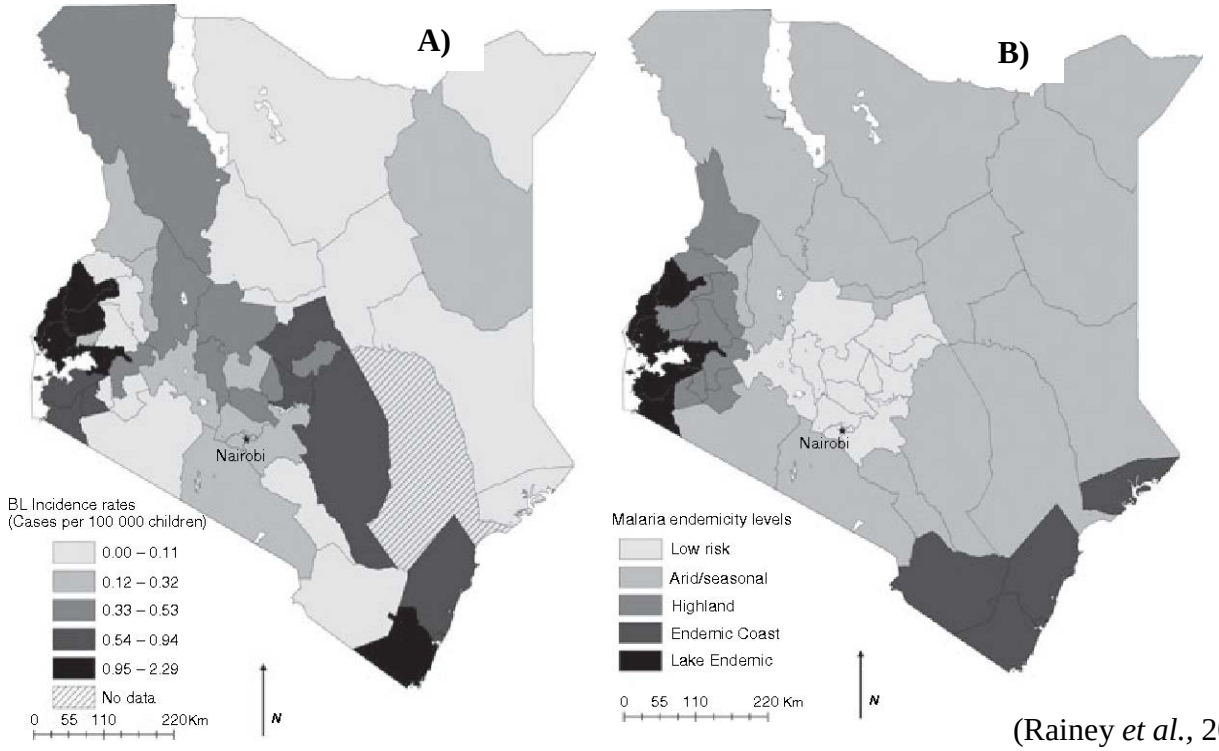
LITERATURE REVIEW

2.1. Burkitt Lymphoma

Burkitt lymphoma is an aggressive germinal center (GC) B cell neoplasm subdivided into three variants according to their epidemiological features; endemic, sporadic and HIV-associated. Endemic Burkitt lymphoma (eBL) was the first variant of BL discovered by Dennis Burkitt (Burkitt and O'Conor, 1961). It is a paediatric malignancy that mainly occurs in equatorial Africa and Papua New Guinea at a high frequency and nearly all its cases are Epstein Barr virus (EBV) infected. In Africa this paediatric neoplasm has the highest incidence in regions where *P. falciparum* malaria is highly prevalent hence malaria infection has long been thought to be a cofactor in lymphomagenesis (Kafuko and Burkitt, 1970; Magrath, 1991). Sporadic BL (sBL) is less frequent than eBL and is predominantly found in patients much older than eBL cases. This BL variant is mainly common in the developed countries and only a minority of sBL cases are EBV infected. The HIV-associated BL (HIV-BL) also referred to as immunodeficiency BL (idBL) most commonly arises in HIV infected patients who are immune compromised and is often a HIV-induced malignancy. Unlike eBL, idBL mainly occurs in adults.

Incidence rates of eBL vary by region in Africa such that children living in malaria holoendemic areas experience higher rates of eBL compared to children from lower transmission areas (Burkitt and O'Conor, 1961; Rainey *et al.*, 2007). Equatorial Africa is a holoendemic malarial region (i.e., regions where essentially the entire population suffers from the disease). In holoendemic regions, >75% of children have splenomegaly and >60% of <5-year-olds have parasitaemia at any given time. Transmission is throughout the year (as opposed to hyperendemic malarial regions, where transmission may be limited to the dry seasons) and

spleen and parasitaemia rates are <70% in children less than 5 years) (Magrath, 2012). In Kenya, a positive trend has been observed between eBL incidence rates and malaria transmission intensity (Figure 2.1) (Rainey *et al.*, 2007)



(Rainey *et al.*, 2007)

Figure 2.1. A) Ten year annual eBL incidence rate in Kenya. B) Malaria endemicity levels in Kenya.

Endemic Burkitt lymphoma accounts for over 74% of all childhood cancers in equatorial Africa (Brady *et al.*, 2007) and has an incidence of 5-15 cases/100,000 children an order of magnitude higher than sBL (Moormann *et al.*, 2007). Boys seem to be more susceptible to eBL than girls with a sex ratio of about 2:1 (Magrath, 1991; Orem *et al.*, 2007) and the majority of the cases are children aged between 2-14 years with a median peak incidence of 6 years (Magrath, 1991). Its tumours can present in a wide range of anatomical sites, although the jaw, neck and abdomen are commonly involved in 50-70% of the cases (Magrath, 1991, 2010). Although therapy of eBL has a high survival rate and can be treated

with chemotherapy, in the absence of surgery or radiotherapy, it is still a great public health concern.

2.1.1. Molecular biology of endemic Burkitt lymphoma

The basic molecular event in BL is translocation of the oncogene (*c-myc*) from its normal position on chromosome 8 to either chromosomes 2, 14 or 22, always in close proximity to the regions that regulate the expression of immunoglobin (Ig) genes (Chêne, 2009). In BL this translocation results in juxtapositioning the DNA coding sequences for *c-myc* with enhancer sequences in the Ig genes. Because Ig enhancer elements are constitutively active in mature B cells, their juxtaposition to *c-myc* in BL cells drives inappropriately high levels of *c-myc* mRNA and protein expression (Hecht and Aster, 2000). As a direct consequence of the chromosomal translocation in BL cells, negative regulatory sequences residing with *c-myc* gene are removed or mutated through other mechanisms, further contributing to increased *c-myc* activity seen in BL cells (Hecht and Aster, 2000). MYC plays a major role in cellular homeostasis and its activity and expression is normally tightly regulated. Posttranscriptional control of this gene are usually overcome in BL through overexpression of *c-myc* protein and other secondary genetic events that enhance transforming activities, antagonize the activity of negative regulatory factors and/or downregulate *c-myc* activities that counteract cellular transformation (Hecht and Aster, 2000). The secondary genetic lesions have neither been fully described nor is there an exhaustive list of candidate genes involved in eBL lymphomagenesis.

Apart from the molecular hallmarks of eBL (*c-myc* translocation) or any other cancer contributing to the survival advantage of the tumour cells, it is known that aberrant gene expression plays a significant role predisposing or even leading to pathogenesis of various cancers such as eBL (Calin and Croce, 2006; Croce, 10/2009). The genetic determinants such as gene expression and regulation leading to eBL development remain to be elucidated. The

t(8:14) alone would not be sufficient for lymphomagenesis of this malignancy, as tumour cells lacking this translocation can be isolated from eBL cases (Onnis *et al.,* 2010). This suggests that additional molecular lesions are required to cooperate with *c-myc* deregulation in the evolution of eBL's malignant phenotype.

Endemic BL, like every other cancer, is a disease involving multi-step changes in the genome. Aberrant transcriptome regulation and expression is one of the changes that would initiate the evolution of a malignant phenotype (Croce, 10/2009). Differential expression, regulation and function of tumour transcriptomes (such as microRNAs and mRNAs) may be involved in lymphomagenesis and progression of eBL. Since a major consequence of uncontrolled *myc* activation, seen in eBL, may be extensive reprogramming of miRNA and mRNA expression patterns (Chang *et al.,* 2008-1; O'Donnell *et al.,* 2005), it implies that the aberrant transcriptome regulation and expression is involved in initiating eBL development but this remains to be elucidated.

2.1.2. Germinal Center reaction in endemic Burkitt Lymphoma development

Many B-cell malignancies, including BL, coopt the gene expression program of their normal GC B-cell counterparts and shape it to execute their oncogenic purposes (Dave *et al.,* 2006). Normal GC B cells and BL tumours express the cytidine deaminase (AID), which mediates both IG somatic hypermutation and IG class switch recombination (CSR). As a consequence, human BLs have somatically mutated IG variable regions, and IG/MYC translocations typically involve IG switch regions, suggesting that they arise by aberrant CSR (Pasqualucci *et al.,* 2004; Ramiro *et al.,* 2004). Initial gene expression profiling of BL have showed that this lymphoma subtype expresses a subset of the genes that characterize the GC stage of differentiation (Dave *et al.,* 2006) such as BCL-6 and PAX5. GC-restricted transcription factors (such as BCL6, which is required for germinal center formation and

recurrently translocated in human lymphomas (Basso and Dalla-Favera, 2012), can have a powerful oncogenic effect when dysregulated because they can simultaneously affect multiple downstream pathways. This results in high dependency addiction of lymphoma cells on lineage-restricted transcription factors (Rui *et al.*, 2011). MYC overexpression in a GC B cell alters cellular metabolism to favour the biosynthesis of various macromolecules that are needed for the malignant phenotype. Other genetic aberrations such an altered gene and miRNA expression patterns coupled with MYC overexpression (as a result of the translocation) in GC B cells could also favour a malignant phenotype. These gene and miRNA expression changes in GC B cells that would initiate eBL lymphomagenesis remain to be unravelled.

2.1.3. Endemic Burkitt lymphoma treatment, relapse and survival

Endemic Burkitt lymphoma is an aggressive B cell lymphoma with tumours doubling in size within 24-26 hours (Chêne, 2009). Due to the aggressive nature of this malignancy it responds well to standard treatment with a 70% in-hospital survival rate (Hesseling *et al*., 2008; Levine, 2002). The standard treatment of eBL is chemotherapy given intravenously, which includes a cyclophosphamide dose and either a high or low dose of methotrexate, and anthracycline (Hesseling *et al.*, 2009). Combination chemotherapy is also used in standard treatment of this malignancy; it consists of prednisolone, doxorubicine, cyclophosphamide, vincristine, methotrexate, cytarabine and hydrocortisone. The combined chemotherapy has a high cure rate but, also has associated toxicities. Therefore, new targeted treatment modalities would be desirable, in order to decrease exposure to cytotoxic chemotherapy and increase the cure rate for eBL.

Endemic Burkitt lymphoma is a curable malignancy with a high survival rate but relapses can occur after standard treatment and the survival chances of the relapse cases is low without the second line treatment (Hesseling *et al.*, 2008), which is more aggressive. However,

second line treatment is not readily available in Kenya and many low income countries (Hesseling *et al.*, 2008). The factors underlying the development of relapse and occurrence of death after first line treatment among a small percentage of eBL cases still remain unknown. Consequently, there is need for the identification of prognostic markers to identify the subset of eBL patients eligible for more aggressive chemotherapy treatment to prevent relapses and death.

The development of relapses after first-line chemotherapy treatment in eBL and low survival rate after relapses highlights the need for identification of prognostic biomarkers and new therapeutic strategies for eBL treatment. Aberrant miRNA and mRNA expression profile have been correlated with metastasis, prognosis and response to treatment in other cancers (Kim *et al.*, 2011; Mallmann *et al.*, 2010-9; Yanaihara *et al.*, 2006). Micro-RNA expression network may have potential prognostic value in eBL relapse/survival and remains to be elucidated.

2.2. Altered Gene expression and cancer

All cancers, like eBL, can be described as a diseases of altered gene expression. They are characterized to have a common phenotype of uncontrollable cell growth and proliferation (Pedraza-Fariña, 2006). Endemic BL, like all cancers, is a multistep process during which cells acquires a series of mutations and other genetic aberrations that eventually lead to aberrant mRNA expression translating to unrestrained cell growth, inhibition of cell differentiation, and evasion of cell death (Ruggero, 2013; Tórtola *et al.*, 1999). A gene that is not normally expressed in a cell can be switched on and expressed at high levels due to genetic aberrations. This can be as a result of gene mutation or changes in gene regulation (epigenetic, transcription, post-transcription regulation mechanisms (such as miRNAs), or post-translation mechanisms).

Knowledge on the gene regulation aberrations that influence the altered gene expression in eBL and other cancers, could be exploited to destroy tumour cells.

In normal cells, some genes function to prevent inappropriate cell growth and proliferation. These are tumour suppressor genes, which are active in normal cells, such as GC B cells, to prevent uncontrolled cell proliferation that would initiate a malignant phenotype. The most studied tumour suppressor gene is P53 gene, which functions as transcription factor and initiator of apoptosis. The gene has been found to be mutated or aberrantly regulated in most cancers (Hong *et al.*, 2014; Olivier *et al.*, 2010). Mutations or aberrant post transcriptional mechanisms that inactivate tumour suppressor genes can functionally inactivate the tumour suppressor gene-encoded protein, or lead to loss or reduction in its expression.

Overexpression of an oncogenes (such as MYC), which are positive cell-cycle and proliferation regulators, can lead to uncontrolled cell growth and proliferation which initiates a malignant phenotype (Poli *et al.*, 2018). Gene expression aberrations is a hallmark of all cancer including eBL, however the gene expression changes in the GC B cells that would initiate the malignant phenotype observed in eBL tumour cells remain to be unravelled.

2.2.1. Post-transcriptional gene regulation

In all cancers, such eBL, an altered microenvironment provides a context similar to that of early embryonic development, marked by the presence of pro-proliferative growth factors and hypoxic growth conditions. Consequently, some tumour cells can acquire embryonic-like/stem cell-like phenotypes characterized by an increased potential for proliferation and survival. Classically, molecular components and changes mediating such processes have been assessed in terms of epigenetic alterations potentiating changes in gene transcription and expression (Postovit *et al.*, 2008). It is now becoming apparent that numerous levels of post-

transcriptional regulation are also integral to the tumour cell response to the microenvironment and initiating a malignant phenotype (Jewer *et al.*, 2012).

Post-transcriptional regulation of gene expression by RNA binding proteins (RBPs) and non-coding RNAs plays an important role in global gene expression. Many post-transcriptional regulators are mis-expressed and mis-regulated in cancers, resulting in altered programs of protein biosynthesis that can drive tumour progression. Post-transcriptional regulation being a primary mechanism by which ncRNAs regulate gene expression, which is extensively involved in biology, evolution, and pathology (Liu *et al.*, 2017). Non coding RNA post-transcriptional regulatory events involved in eBL lymphomagenesis remain to be unravelled. Unravelling part of eBL post-transcriptional regulatory network would aid in identifying biomarkers or future therapeutic targets for this paediatric malignancy.

2.3. Non-coding RNA

The well-studied sequences in the human genome are those of the protein coding genes (mRNA). However, the coding exons of these genes account for about 1.5% of the genome. In recent years, it has become increasingly apparent that the non-protein coding portion of the genome is also of crucial functional importance; for normal development and physiology and also disease development (Esteller, 2011). Recent discoveries have revealed a subtype of noncoding RNAs (ncRNA) such as small ncRNAs which are less than 200 nucleotides long, that are key regulators of transcription and translation of the protein coding regions of the genome. These small ncRNA molecules include small interfering RNA (siRNA), small nucleolar RNA (snRNA), PIWI-interacting RNA (piRNA) and microRNAs (miRNAs) (Esteller, 2011). The relevance of the non-coding genome to human disease has mainly been studied in the context of the widespread disruption of miRNA expression and function that is seen in human cancers and other diseases (Croce, 10/2009; Esteller, 2011). Recent studies on

human diseases, particularly cancer, have shown that genetic defects in miRNAs and their processing machinery are common hallmarks of disease development and progression (Deng *et al.*, 2008). The contribution of ncRNAs (such as miRNAs) to the genesis and progression of eBL still remains unravelled. Hence there is need to determine the full extent of their contribution to eBL genesis and the mechanisms by which ncRNAs exert their pathological effects.

2.4. MicroRNAs

MicroRNAs (miRNAs) are the most extensively studied class of ncRNAs, which are small ~22 nucleotides (nt) in length and mediate post-transcriptional gene silencing by controlling the translation of mRNAs (Croce, 10/2009; Esteller, 2011). MiRNAs were first discovered in 1993 during a study of the lin-4 gene which is crucial for nematode *Caernorhabditis elegans* development (Lee *et al.*, 1993). However, they were not recognized as a distinct class of biological regulators until the early 2000's when the let-7 RNA was associated with triggering transitions to coordinate developmental timing (Reinhart *et al.*, 2000). Since then multiple roles of miRNAs in regulation of gene expression have been revealed (Griffiths-Jones *et al.*, 2006). Currently, over 1000 miRNAs have been experimentally identified in humans and registered in the miRBase database (http://microrna.sanger.ac.uk) (Griffiths-Jones *et al.*, 2006). Bioinformatics prediction of miRNA targets suggest that more than 30% of protein coding genes may be regulated by miRNAs (Lu and Clark, 2012). This makes miRNA molecules one of the most abundant classes of regulatory genes in humans. This class of ncRNAs are now perceived to be key players in post-transcriptional control within the networks of gene regulation (Esteller, 2011).

MicroRNAs play a vital role in regulating many cellular processes, including growth, proliferation, apoptosis, differentiation and development via the negative regulation of over

one-third of all human genes at the translational stage (Croce, 10/2009; Lu and Clark, 2012). MiRNA genes represent 1 to 3% of the known genes in the human genome, a large number of these genes are found in the intron regions, but some can be located in the exons of protein-coding genes (Liu *et al.*, 2007; Saunders *et al.*, 2007).

2.4.1. MicroRNA biogenesis and function

MicroRNAs are processed from longer precursor RNA molecules that are encoded by the miRNA genes (termed *mir*). Biogenesis of miRNAs occurs through a multi-step process that involves the RNase III enzymes; Drosha and Dicer, ultimately resulting in the production of mature miRNAs (miRs) of ~22nt (Liu *et al.*, 2007; Sotiropoulou *et al.*, 2009). The *mir* are transcribed in the nucleus as long stem-loop precursors called primary miRNAs (pri-miRNAs) (>1000nt). The enzyme Drosha cleaves pri-miRNAs into precursors (pre-miRNAs) which are 60-100 nt long. The pre-miRNAs are exported into the cytoplasm by the RNA GTP-dependent transporter Exportin 5 and undergo an additional processing step by Dicer to form a miRNA: miRNA* duplex that is about 15 to 22 nt in length. This RNA duplex is then unwound by Helicase, one RNA strand becomes the mature single-stranded miRNA (miR), while the complementary strand, termed passenger miRNA (miR*), is usually rapidly degraded (Liu *et al.*, 2007; Sotiropoulou *et al.*, 2009). The miR is integrated into an RNA-induced silencing complex (RISC) where it binds to a complementary sequence in the 3'UTR of the target mRNA by base pairing, initiating gene repression.

MicroRNAs recognize their targets based on sequence complementarity. The miR is partially complementary to one or more mRNAs. The complementary sites are usually within the 3'UTR of the target mRNA. Two mechanisms of miRNA repression have been suggested depending on the degree of complementarity between the miRNA and the target mRNA. First, is inhibition of protein translation by binding to imperfect complementary sites within the

3’UTR of the target mRNA. Second, would be endonucleolytic cleavage of the target mRNA by perfect base pairing to the 3’UTR site (Figure 2.1) (Liu *et al.*, 2007).

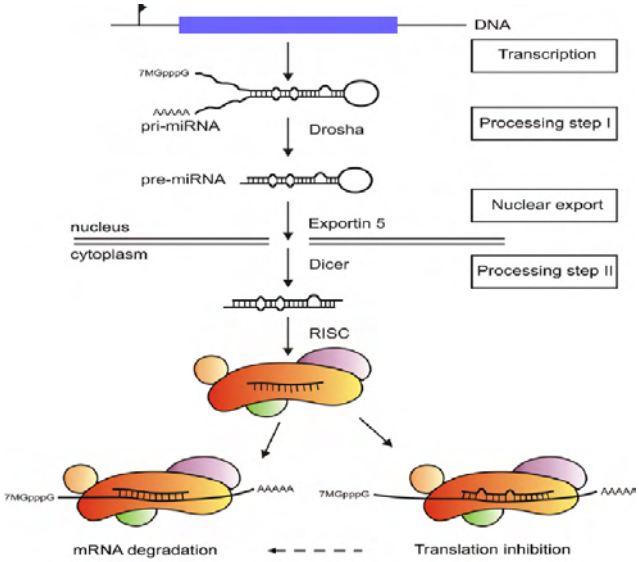


Figure 2.2. Biogenesis and function of miRNAs (Liu *et al.*, 2007)

The biological role and *in vivo* functions of most mammalian miRNAs are still poorly understood. In mammals, miRNAs have been found to play a role in embryogenesis, cell differentiation, brain development, cell proliferation, growth control, and programmed cell death (Croce, 10/2009). Accumulating evidence indicates that miRNAs are involved in the initiation and progression of cancer (Di Lisio *et al.*, 2012; Jin *et al.*, 2013; Wang *et al.*, 2012-5). However, it remains uncertain whether aberrant miRNA expression is a cause or consequence of pathological processes such as eBL development and progression. The underlying mechanisms of why and how miRNAs become deregulated are also largely unknown, and their expression patterns in eBL tumour cells remains to be identified to understand the chain of molecular events that would initiate eBL lymphomagenesis.

2.5. MicroRNA and Cancer

Gene deregulation is one of the key mechanisms by which cells can progress into cancer. New mechanisms governing the expression of genes related to cancer development

include posttranscriptional regulation through miRNAs. The novel involvement of this class of ncRNA in human cancers is just beginning to be appreciated but much more remains obscure (Sotiropoulou *et al.,* 2009), for example their role in eBL development.

At their discovery, miRNAs were shown to control fundamental cellular processes, such as differentiation of cells and timing of development of organisms (Liu *et al.,* 2007; Sotiropoulou *et al.,* 2009), and with a large body of evidence documented on miRNA expression being ubiquitously deregulated in cancer cells (Calin and Croce, 2006), this class of ncRNAs may promote eBL tumorigenesis. The first piece of evidence connecting miRNAs and cancer was from a study that showed miR-15 and miR-16 were the target genes of the 13q14 deletion that is common in chronic lymphatic leukaemia (Calin *et al.,* 2002; Sotiropoulou *et al.,* 2009). From this initial finding, more than half of known miRNAs where shown to be located in cancer associated genomic regions and show copy number alterations in cancer (Croce, 10/2009). Aberrant expression of several miRNAs, such as miR-143, miR-145, miR-21, miR-155 and the let-7 family, and many other miRNAs has been detected in various cancers (Croce, 10/2009; Sotiropoulou *et al.,* 2009).

Functional studies of individual miRNAs have shown that they can act as oncogenes (onco-miRs) or as tumour suppressor miRNAs (Croce, 10/2009). Over expression of oncogenic miRNAs and under-expression of tumour-suppressive miRNAs can contribute to tumorigenesis by affecting pathways promoting the acquisition of the hallmarks of cancer. These hallmarks include sustained proliferative signalling, resisting cell death, impaired DNA damage response (DDR), evading growth suppressors, enabling replicative immortality and inducing angiogenesis (Croce, 10/2009; Sotiropoulou *et al.,* 2009). Although targets of most miRNAs have not been identified, some negatively regulate well-known oncogenes, for example miR-15 and miR-16 have been reported to repress the anti-apoptotic factor *BCL2* and

the let-7 miRNA family to target the *RAS* oncogene (Cimmino *et al.*, 2005; Wang *et al.*, 2012-5). MiRNA 17-92 cluster over-expression in B cell lymphomas has been shown to promote *c-myc* mediated oncogenesis (Li *et al.*, 2012; Olive *et al.*, 2010). These studies are evidences of miRNA function in cancer development, but their role specifically in eBL development remains to be elucidated.

MicroRNAs function as regulatory molecules of gene expression with multifaceted activities that exhibit direct or indirect oncogenic properties which promote cancer molecular mechanisms (Liu *et al.*, 2007; Sassen *et al.*, 2007; Sotiropoulou *et al.*, 2009) that can cause the onset of eBL tumour development. Understanding the miRNA signalling network in eBL development, is likely to yield a new class of targeted therapeutic strategies with excellent cure rates but reduced or no toxic effects. MicroRNAs associated poor eBL patient outcome remain to be identified for future targeted therapy development.

2.5.1. MicroRNAs and Burkitt lymphoma

The *t(8:14)* translocation and its variants have been considered the molecular hallmark of BL, for several years now (Hecht and Aster, 2000). However, BL cases lacking this chromosomal aberration have been identified (Leucci *et al.*, 2008; Onnis *et al.*, 2010). Intriguingly, the Burkitt lymphoma (BL) cases express *MYC* at the levels comparable to BL cases with the translocation, despite that the *c-myc* gene is tightly regulated in normal cells (Onnis *et al.*, 2010). Its deregulation is often due to genetic abnormalities or alterations in post-transcriptional regulatory mechanisms mediated by microRNAs (Onnis *et al.*, 2010), who's aberrant expression has been associated with cell transformations and cancer development (Jansson and Lund, 2012). This observation implicates the involvement of these post-transcriptional regulators in the development of BL and also highlights the fundamental importance of the *c-myc* gene in initiating BL tumorigenesis.

Activated expression of the *MYC* gene is one of the most common oncogenic events in BL. Enhanced expression of the *MYC* oncogenic transcription factor reprograms cells to drive proliferation which would initiate this B cell neoplasm (Gavioli *et al.*, 2001; Scheller *et al.*, 2010). It has been shown that a major consequence of *MYC* activation is extensive reprogramming of the miRNA expression pattern in tumour cells (Cairo *et al.*, 2010). Widespread downregulation of some miRNAs is predominantly influenced by *MYC* activation, although direct upregulation of the pro-tumorigenic miR-17-92 cluster has also been observed to promote tumour proliferation and accelerate angiogenesis and tumorigenesis (Chang *et al.*, 2008-1). This observation implies that repression of tumour suppressing miRNAs and upregulation of oncomiRs is a fundamental component of *MYC* tumorigenic program that would initiate eBL development.

The Epstein Barr virus is a highly prevalent virus infecting over 90% of the human population, yet it is associated with the occurrence of several cancers such as BL, but can remain persistent in healthy carriers (Niedobitek *et al.*, 2001). A satisfactory explanation on EBVs role in BL development still remains elusive; however, the interplay between viral products and host microRNA has been proposed (Bornkamm, 2009; Grömminger *et al.*, 2012). In a recent study, miR-127 was shown to be overexpressed due to the presence of EBNA-1 (Onnis *et al.*, 2012). MiR-127 upregulation impaired B cell differentiation (Onnis *et al.*, 2012). This provides a new insight on miRNA involvement in BL pathogenesis and unravels a novel mechanism of human miRNA regulation by a viral product to initiate lymphoma development.

MicroRNAs regulate many genes critical for tumorigenesis (Palmero *et al.*, 2011; Tomasetti *et al.*, 2014); thus studying their expression in endemic BL could enhance the understanding of tumorigenesis pathways and/or helpful for cancer diagnosis and prediction of

clinical outcome, and drug resistance. These alterations of miRNA expression might therefore become a new therapeutic approach in eBL treatment.

2.6. MicroRNA as Predictive and Prognostic Markers in Cancer

MiRNAs are emerging as important modulators in cellular pathways and play a key role in carcinogenesis (Croce, 10/2009). They can change cellular response to a specific class of drugs not only through survival or apoptotic signalling but also by interfering with drug targets. Aberrant miRNA expression may have potential prognostic value in different malignancies (Cho, 2012). Recent studies have identified a number of miRNA as potential biomarkers for cancers; some of them act as oncogenes, tumour suppressors or modulators of metastasis (Flatmark *et al.*, 2016; Lan *et al.,* 2015; Macha *et al.,* 2014; Mishra, 2014; Wang *et al.,* 2016). For example, in pancreatic cancer patients, it has been shown that gemcitabine treated patients with high miR-142-5p expression have significantly longer survival time than those with low miR-145-5p expression (Ohuchida *et al.,* 2011). This miRNA may serve as a predictive marker for gemcitabine response in patients with pancreatic cancer (Ohuchida *et al.,* 2011). Another study on ovarian cancer revealed that patients without complete response to paclitaxel-based treatments have been demonstrated to have lower miR-200c levels than patients with complete response (Leskelä *et al.*, 2011). MiRNA expression has also been correlated with outcomes of lung cancer, its expression has been shown to be an unfavourable prognostic factor in lung cancer (Yanaihara *et al.,* 2006).

These studies collectively show the possibility of using miRNA expression profiles to predict outcomes of various cancers such as eBL. In addition to genetic and epigenetic changes, in the anticancer drug resistant phenotype, miRNAs may be applicable in evaluation of outcome and modification of response in antitumor therapies (Hayes *et al.,* 2014), and would also aid in identifying the subset of eBL patients that would relapse after treatment.

Despite significant advances in cancer therapy, drug resistance is still a major obstacle in cancer treatment (Hong *et al.,* 2013; Nikolaou *et al.,* 2018). Identifying reliable predictors of tumour recurrence following standard treatment to determine whether immediate more aggressive therapy is warranted to minimize relapses, is an important challenge in cancer research. As different cancer therapies are effective in different subgroups of cancer patients (Dancey *et al.,* 2012), like in nasopharyngeal carcinoma (Su et al., 2011), there is a tremendous need for novel biomarkers to predict responses to anticancer treatment in patients, as well as aid in designing future clinical trials that better stratify patients. Given that the molecular events underlying the development for relapse after first line chemotherapy treatment among a small percentage of eBL cases is unknown, there is no way of determining sensitivity or resistance of the tumour to first line chemotherapy. This brings about a need for accurate prognostic indicators to distinguish high risk patients from other patients, to identify the subset of eBL patients eligible for more aggressive chemotherapy to prevent occurrence of relapses in eBL

Endemic Burkitt lymphoma, like every other cancer, is a complex genetic disease that deciphering the role of miRNAs in the signalling pathways responsible for poor patient outcome, despite early treatment, in a subset of eBL patients, would allow better utilization of miRNAs for the prediction and prognosis of eBL treatment.

2.7. MiRNA Target Gene Prediction and Identification

MiRNAs regulate the translational expression of genes. They are thought to bind partially to complementary sites in 3'UTR of the target mRNA in the cytoplasm to repress translation of the target genes (Pasquinelli, 2012). Effective identification of miRNA-mRNA interaction in animal systems remains challenging due to the interaction complexity and limited knowledge of rules governing these processes. To understand the biological function of

miRNAs in disease development, it would be of particular interest to reliably predict/identify potential miRNA targets which may also be involved in diseases (Witkos *et al.*, 2011). However, by virtue of their structure and mode of action, interactions between miRNAs and their mRNA targets are complex and often there are numerous putative miRNA recognition sites in mRNAs. Due to this, computational methods have been devised using target prediction algorithms to find possible miRNA-mRNA interactions (Pasquinelli, 2012; Saito and Sætrom, 2010; Witkos *et al.*, 2011).

The complexity of miRNA-mRNA interactions is not only based on miRNA-mRNA sequence matching, but additional parameters such as orthologous sequence alignment, UTR context or free energy of the complexes to facilitate miRNA-mRNA interaction and also evolutionary conservation of binding and target sites, which need to be taken into account for efficient miRNA target prediction (Witkos *et al.*, 2011). The rules for targeting transcripts by miRNAs has not been fully examined yet but are based mainly on experimentally validated miRNA-mRNA interactions that are a slice of possibility existing *in vivo*. This situation has led to the development of algorithms for miRNA target prediction. Many algorithms have been developed to consider all these aspects in miRNA-mRNA interactions (Table 2.2) (Witkos *et al.*, 2011). Most of them require some degree of sequence complementarity as well as favourable free energy in the miRNA-target duplex, since it is assumed that target site occupancy correlates with the strength of the base pairing (Hamzeiy *et al.*, 2014; Saito and Sætrom, 2010).

Each algorithm developed has its defined criteria of target prediction based on a combination of target features such as seed sequence complementarity, target site accessibility and also the evolutionary conservation of the 3'UTR sequence of candidate genes (Hamzeiy *et al.*, 2014; Saito and Sætrom, 2010). For this reason, these algorithms often create different

target gene lists. However, combining the use of multiple algorithms reduces false positivity making the results more specific although this admittedly comes at a loss in sensitivity (Hamzeiy *et al.*, 2014; Saito and Sætrom, 2010).

Table 2.1. Comparison of miRNA target prediction algorithms.

	miRanda	*DNA-microT*	*PicTar*	*RNAhybrid*	*TargetScan*
Seed-match	×		×	×	×
Free-energy	×	×	×	×	
Conservation	×		×	×	×
Complementarity	×	×		×	
Predict known target	×		×	×	

To understand miRNA-mRNA interactions and role in gene regulation, miRNA target prediction algorithms, that focus on the orthologous sequence alignment and parameters such as free energy of binding (target site accessibility), together with inverse correlation between expression levels of putative miRNA and targeted mRNA would increase the likelihood that the interaction is real and functional (Pasquinelli, 2012). Identifying targets of aberrantly expressed miRNAs in eBL and their functional regulatory network is critical in understanding their role in eBL development.

2.8. MicroRNA and mRNA Expression Profiling

Accurate quantification of genes (mRNAs) and miRNAs in cells and tissues is a critical step in understanding their biological function. MiRNA expression profiling has helped to identify miRNAs that regulate a range of cellular processes. Consequently, miRNAs are being investigated as molecules for the reprogramming of cell fate that would initiate cancer development, as well as being applied as biomarkers for identifying various cancers and also

used for cancer prognosis (Calin and Croce, 2006). The first analysis step is to detect mRNAs and miRNAs which show a significantly different expression pattern in cancer and non-cancer samples. Quantitatively measuring expression levels of miRNAs and their target genes can help in understanding the mechanisms of human disease development and discovering new drug targets.

Considerable efforts have been devoted to developing new methods for high-throughput detection of miRNAs and accurately measure gene and miRNA expression changes. Many properties that are unique to miRNAs pose challenges to their accurate detection and quantification as opposed to mRNAs. For instance, the ~22 nt length of mature miRNAs is insufficient for annealing to traditional primers that are designed for reverse transcription and PCR. Also, due to their short length, very little sequence is available to design complementary microarrays and perform reliable amplification or labelling of each miRNA without introducing signal bias. In addition, miRNAs within a family (for example, the *let-7* family) differ by as little as a single nucleotide, making the ability to discriminate between forms with single-nucleotide differences important in miRNA detection and quantification (Pritchard *et al.*, 2012). Unlike mRNAs, miRNAs lack a common sequence, such as a poly(A) tail, that can be used for selective enrichment or as a universal primer-binding site for reverse transcription. These properties are important because miRNAs represent a small fraction of the total RNA mass and must therefore be selectively detected in a background of other, diverse RNA species, including the pri-and pre-miRNA precursors that also contain the mature miRNA sequence (Pritchard *et al.*, 2012).

Another challenge for profiling hundreds of miRNAs in parallel is that, owing to their short length, variance in GC content leads to a wide variance in melting temperatures for annealing reactions, which could create miRNA-specific biases (Pritchard *et al*., 2012).

Despite these challenges, three major approaches for miRNA detection and quantification are currently well-established: reverse transcription quantitative PCR (RT-qPCR), microarray hybridization, and next generation sequencing (NGS) technology (Pritchard *et al.*, 2012).

The emergence of next generation sequencing (NGS) technology has significantly led to more broad understanding of transcriptomes. RNA sequencing is not only suitable for profiling of known miRNAs as qRT-PCR and microarray can do too, but it is also able to identify novel miRNAs which the other methods are incapable of. Compared to other miRNA expression profiling techniques such as microarrays and RT-qPCR, NGS technology offers precise identification of miRNA sequences by readily distinguishing between miRNAs that differ by a single nucleotide, as well as the identification of isomiRs of varying lengths (Pritchard *et al.*, 2012; Ramsingh *et al.*, 2010). IsomiRs have been identified in previous miRNA sequencing studies that differ in mature miRNA (miR) length or less commonly in nucleotide sequence and are thought to arise from variable miRNA processing or RNA editing (Ramsingh *et al.*, 2010). Although the functional significance of isomiRs is unclear, it would still be important to determine whether there is an aberrant isomiR expression pattern between eBL tumour cells and normal cells (Ramsingh *et al.*, 2010) that may contribute to eBL development.

Next generation sequencing (NGS) was applied in this study to comprehensively assess miRNA expression, identify genetic variants in miRNA genes and screen for alterations in miRNA binding sites in a patient with endemic Burkitt lymphoma. Combining this wealth of information with mRNA profiles and other genome-scale data would be useful for system-level understanding of gene regulatory network initiating eBL's malignant phenotype. Including miRNA profiling as an additional tool to understanding the complexities of gene regulatory

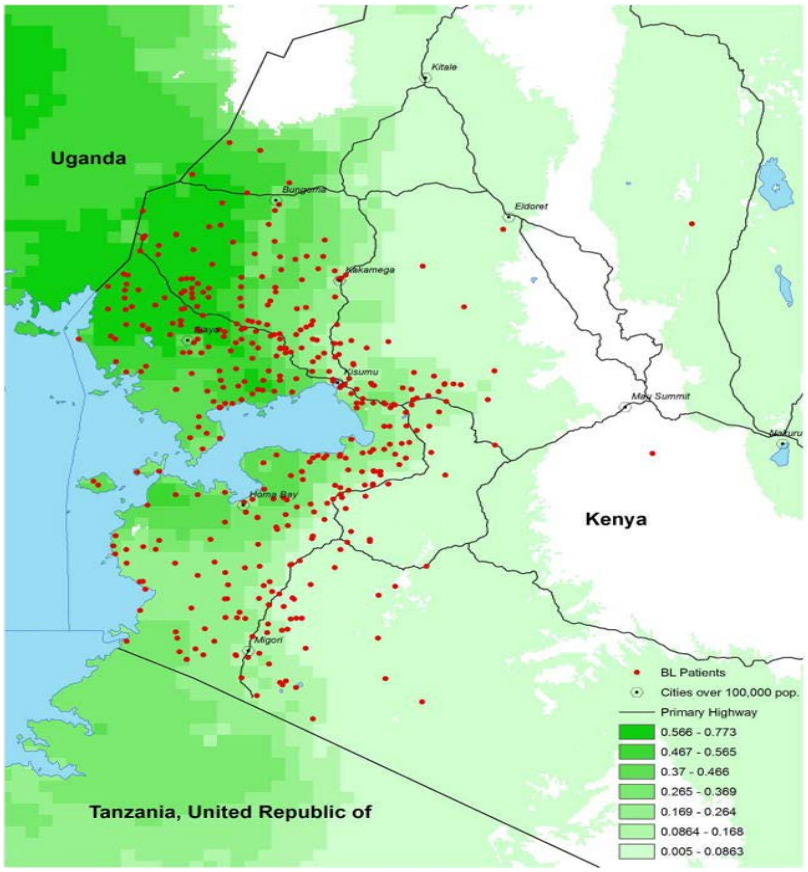
networks initiating development of eBL, would aid in identifying new targets in its treatment with little or no toxicities and reduced possibilities of relapse occurrences.

CHAPTER THREE

MATERIALS AND METHODS

3.1. Study Area

Study participants diagnosed with endemic BL were recruited from Jaramogi Oginga Odinga Teaching and Referral Hospital (JOOTRH) (0˚5′ 17.832″ S and 34˚46′ 13.789″ E) located in Kisumu City. It is the largest hospital and the only regional referral centre for childhood cancer cases in western Kenya. The catchment area for JOOTRH spans western Kenya and displays the expected geographic overlap with malaria transmission.



(Buckle *et al.*, 2016)

Figure 3.1.: Distribution and catchment area in western Kenya of endemic Burkitt lymphoma patients admitted to JOOTRH in Kisumu, Kenya. The map of western Kenya shows P. falciparum malaria transmission intensity (shaded by increasing intensity in green) and home location (red spots) of endemic Burkitt lymphoma patients admitted to JOOTRH in Kisumu, Kenya, from 2003 to 2011.

3.2. Study Population

To determine the role mRNA and miRNA expression in eBL lymphomagenesis among children from western Kenya, a case-control study design was adopted since the mRNA and miRNA expressions profile of eBL tumour cells (Cases) was compared to normal GC B cells (Controls) from healthy children (with no prior history of developing any cancer) between 2-5 years. Endemic BL tumour samples for this study were obtained from eBL patients (n = 28), aged between 5-14 years. eBL patients were diagnosed by use of cytology on the Fine Needle Aspirates (FNA) of the primary tumours. Some of the FNA sample from children diagnosed with eBL were collected in RNAlater prior to chemotherapy treatment at JOOTRH.

In order to compare eBL to their presumed normal GC B cells (Control group), publicly available miRNAseq and mRNAseq GC B-cell datasets were reanalysed in 5 GC B cells (Jima *et al.*, 2010; Koues *et al.*, 2015; Kuchen *et al.*, 2010). The raw miRNAseq and mRNAseq fastq read files of the sorted GC B-cells samples were accessed from the Gene Expression Omnibus (GEO) archive through accession GSE22898 and the Blueprint consortium, dataset ID: EGAD00001002452 (Appendix 5).

3.2.1. Inclusion criteria

The eBL case diagnosis was confirmed by two independent pathologists from touch prep slides. Endemic BL cases were included only if the informed consent was signed and resident in a malaria holoendemic region. Children with confirmed diagnosis of eBL and admitted to JOOTRH were eligible for inclusion in this study. FNA biopsies had to be collected prior to chemotherapy treatment. eBL cases included in the study had to be within ages 2-14 years. eBL being a germinal centre B cell lymphoma, the control group for this study had to be germinal centre B cells for children between ages 2-14 years. The germinal centre B cells had to be from children within the age range of eBL occurrence, which is between 2- 14 years.

3.2.2. Exclusion criteria

Cases were excluded if they didn't sign the informed consent form, were found to have HIV during the hospital visits for their cancer diagnosis, and also if they had been previously diagnosed with another cancer. Cases were also excluded from analysis if they abandon treatment or were lost to follow up making it impossible to monitor their survival (or relapse). Germinal centre B cells from children with a prior history of developing a lymphoma or any cancer were excluded from being part of the control group.

3.3. Sample Size Calculation

Sequencing is a digital counting process, and the total amount of sequence can vary significantly both between runs (technical variability) and between genes within a given run (biological variability), with some genes being invisible (0 counts) in a given run. Therefore, the amount of information in a sequencing run can change between experiments, and this is a critical variation that needs to be accounted for in sample size estimates. To capture the influence of both biological and technical variability, sample size calculations need to be based on a negative binomial (NB) distribution, since it accounts for both aspects. The formula by (Hart *et al*., 2013) includes both sequence based counting error and biological variability, while avoiding the rapidly diminishing returns (and expense) of over-increasing sequencing depth. Therefore, to determine the desired sample size for this study, the power calculator by Hart *et al*., 2013 was used. The power calculator applied was based on the formula:

$$n = 2(Z_{1-\alpha/2} + Z_{\beta})^2 \frac{(1/\mu + \sigma^2)}{(log_e \Delta)^2}$$

The parameters; **α** and **β** are the level of significance of the test and the probability of accepting H_0 when it is false respectively; **z** the corresponding to cut off points, and **Δ** the testing target, were fixed across this study. Their values were:

$z_{1-\frac{0.05}{2}} = 1.96$, corresponding to a two-sided test at α = 0.05; and β = 0.1, corresponding to a 90% power, therefore, $z_{\beta} = 1.28$.

A fold change, Δ= 4.0, corresponds to a log_2FC of 2 change in gene expression between the two groups (cases and controls). Therefore, $log_2FC \geq 2$ would be used as a cutoff to identify the differentially expressed genes (mRNAs) and microRNAs.

The variables $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ depend on the gene and the experiment. $\boldsymbol{\mu}$ represents the depth of coverage of the gene (number of reads assigned to a particular gene), and $\boldsymbol{\sigma}$ represents the coefficient of variation (CV) in the gene between biological replicates (human samples) (Hart et al., 2013).

The coefficient of variation also referred to as the biological variation is a property of the particular gene/population/condition under study. It is expected to be larger for less uniform replicates such as human subjects. Estimation of biological variation by Hart *et al.*, (2013), showed that values of the CV for human samples ranged from 0.32 to 0.74 with a median of 0.43. Therefore, this study used a CV of 0.43.

This sample size calculation formula determines the sample size (cases and controls) and also the fold change to detect a given fixed sample size. Using the power calculator by Hart *et al.*, (2013), assumption of a coefficient of variation of 0.43 between the biological replicates, and the depth coverage, $\boldsymbol{\mu}$**,** of the miRNAs to be 800 (Farazi *et al.*, 2011; Toung *et al.*, 2011-6).

$$n = 2(Z_{1-\alpha/2} + Z_{\beta})^2 \frac{(1/\mu + \sigma^2)}{(log_e \Delta)^2}$$

$$n = 2(1.96 + 1.28)^2 \frac{(1/800 + 0.43^2)}{(log_e 4)^2}$$

$$n = 2(3.24)^2 \frac{(0.18615)}{(1.38629436112)^2}$$

$$n = 2(10.4976) \frac{(0.18615)}{(1.921812)}$$

$$n = (20.9952)(0.0968617118)$$

$$n = 2.03363$$

$$n \approx 3\ samples\ per\ group\ (since\ we\ can't\ get\ an\ 0.03363\ of\ a\ person)$$

Substituting these variables to the formula and into the power calculator, 3 study participants would be the minimum number of samples needed for each group, 3 eBL cases and 3 germinal center B cells (Objective 1 and 2) and 3 eBL Jaw tumours and 3 eBL Abdominal tumours (Objective 3), to draw valid conclusions or inferences about the population. Studying the whole population is usually restricted by time, accuracy and cost constraints however, a complete census would be most appropriate in any given study. Since increasing the number of eBL cases was not constrained by any of the aforementioned factors, a sample size of 28 eBL tumour samples would better control for the biological variations between the samples and also enhance the validity of eBL mRNA and miRNA expression signature for these results.

3.4. RNA and Small RNA Isolation

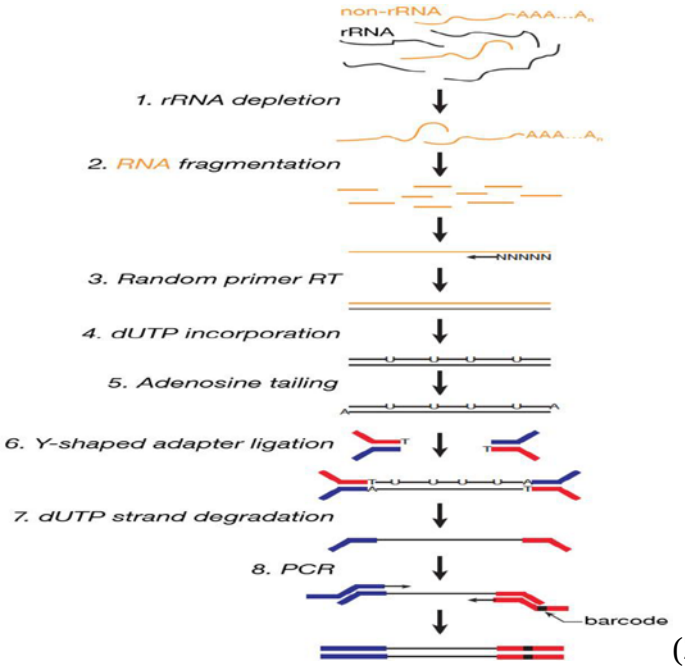
Total RNA and Small RNA molecules were extracted from eBL FNA samples in RNA-later using the AllPrep DNA/RNA/miRNA Universal kit (Qiagen) according to manufacturer's instructions (Appendix 1). The kit comprises a set of specially optimized enzymatic digestion steps which enabled the purification of RNA, miRNA, and DNA. Briefly, the FNA samples were first lysed and homogenized in a highly denaturing guanidine-isothiocyanate-containing

buffer, which immediately inactivates DNases and RNases to ensure isolation of intact DNA and RNA. The lysate was then passed through an AllPrep DNA Mini spin column (Qiagen). This column, in combination with the high-salt buffer, allowed selective and efficient binding of genomic DNA. On-column Proteinase K digestion in optimized buffer conditions allows the purification of high DNA yields from the FNA sample. The columns were then washed and pure, ready-to-use DNA was eluted. The Flow-through from the AllPrep DNA Mini spin column was then digested by Proteinase K in the presence of ethanol. This optimized digestion, together with the subsequent addition of further ethanol, would allow appropriate binding of total RNA, including miRNA, to the RNeasy Mini spin column. A DNase I digestion is then performed to ensure high-yields of DNA-free RNA. Following DNase I digestion, any possible contaminants are efficiently washed away and high-quality RNA (mRNA and miRNA) was eluted. All isolated nucleic acids were stored at -80°C.

Small RNA abundance and integrity were determined after isolation using Nanodrop-ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), respectively. Only samples with a miRNA concentration > 5ng/µl and total RNA RIN (RNA integrity number) > 8.0 were considered for small RNA library preparations and sequencing. Of the 28 samples only 17 were considered for miRNA library preparation.

3.5. RNA Sequencing Library Preparation

Briefly, starting with 1-5µg total RNA, twenty-eight strand-specific RNAseq libraries were prepared following the protocol from Zhang and colleagues (Zhang *et al*., 2012) as illustrated in Figure 3.1, combined with mRNA enrichment with oligo-dT using Dynabeads mRNA purification kit (Life Technologies) (Appendix 2).



(Zhang *et al.*, 2012)

Figure 3.2: Strand specific RNA library preparation protocol workflow.

Final library qualities were confirmed with Bioanalyzer High sensitivity DNA kit (Agilent) and sequenced with paired end read (2×100bp) using multiple lanes of Illumina HiSeq 2000 (Illumina Inc., San Diego, CA, USA). This data can be accessed at dbGAP with accession number phs001282.v1.

3.6. Small RNA Library Preparation and Sequencing.

Seventeen indexed (barcoded) miRNA libraries were prepared using the Illumina Truseq Small RNA Library Preparation Kit (Illumina Inc., San Diego, CA, USA). The small RNA sequencing libraries were prepared according to the manufacturer's protocol (Appendix 3). Briefly, a 3'-adaptor sequence was first added to the 3'end of the small RNA molecules. The 5' adaptor sequences were then added to the 5'- end of the 3' adaptor ligated small RNA molecules. The RNA now had an adaptor sequence at both ends as illustrated in Figure 3.2. The 3' and 5' adaptor ligated RNA was then reverse transcribed into cDNA, and the remaining RNA was removed using RNase.

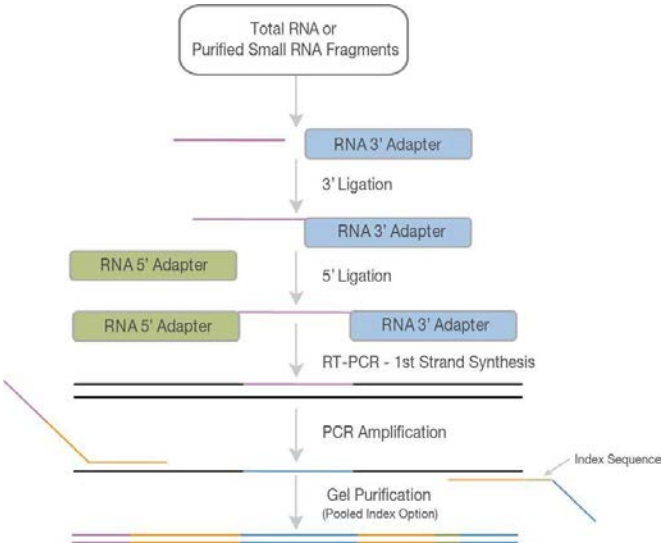


Figure 3.3: Small RNA library preparation protocol workflow

Size fractionation of the miRNA libraries to separate them from adapter dimers was achieved by electrophoresis on 8% TBE polyacrylamide gel (Life Technologies, Carlsbad, CA, USA). The libraries were then purified from the gel and the Agilent High Sensitivity DNA Kit (Agilent Technologies, Colorado Springs, CO, USA) was used to quantify the molarity and size of finished miRNAseq libraries. The purified miRNA libraries were then quantified on the Agilent DNA 1000 chip. The Indexed miRNA libraries were pooled together to equal concentrations and randomised across two lanes of the Illumina flowcell, with 17 indexed samples in each lane. The libraries were sequenced on an Illumina HiSeq 2000 at the University of Massachusetts Deep Sequencing Core with TruSeq v3 reagents. The FASTQ files were produced using the CASAVA pipeline v2.0. The Illumina barcodes and adaptor sequences used in the multiplexed library preparation are provided (Appendix 3.1).

3.7. Bioinformatic Processing of RNA and miRNA Sequencing Data

3.7.1. RNA-seq data processing

The eBL and GC B-cell RNAseq data were in FASTQ (sequence and base quality) format. Since, numerous sequence errors can be introduced during the library preparation, sequencing, and imaging steps, these should be identified and filtered out in the data processing

step. Thus, the initial step in the RNAseq data processing workflow, was the QC of raw sequencing data using FastQC software v0.10.0 (Andrews, 2014), which assess the overall and per-base quality for each read in each sample. Next step, involved the removal of the adapter sequences using Cutadapt v1.1 (Martin, 2011). Low quality bases, as a result of sequencing errors, were also trimmed off using FASTX-Toolkit (Schmieder and Edwards, 2011) since these could prevent proper read alignment to the reference genome. The transcripts were reconstructed and mapped to the reference genome using TopHat (Trapnell *et al.*, 2012, 2009). RSEM software (Li and Dewey, 2011) was then used to quantify the transcript-level abundance from the transcriptome mapping of the RNA-seq reads for all the eBL and GC B-cell data to the reference (Figure 3.4A).

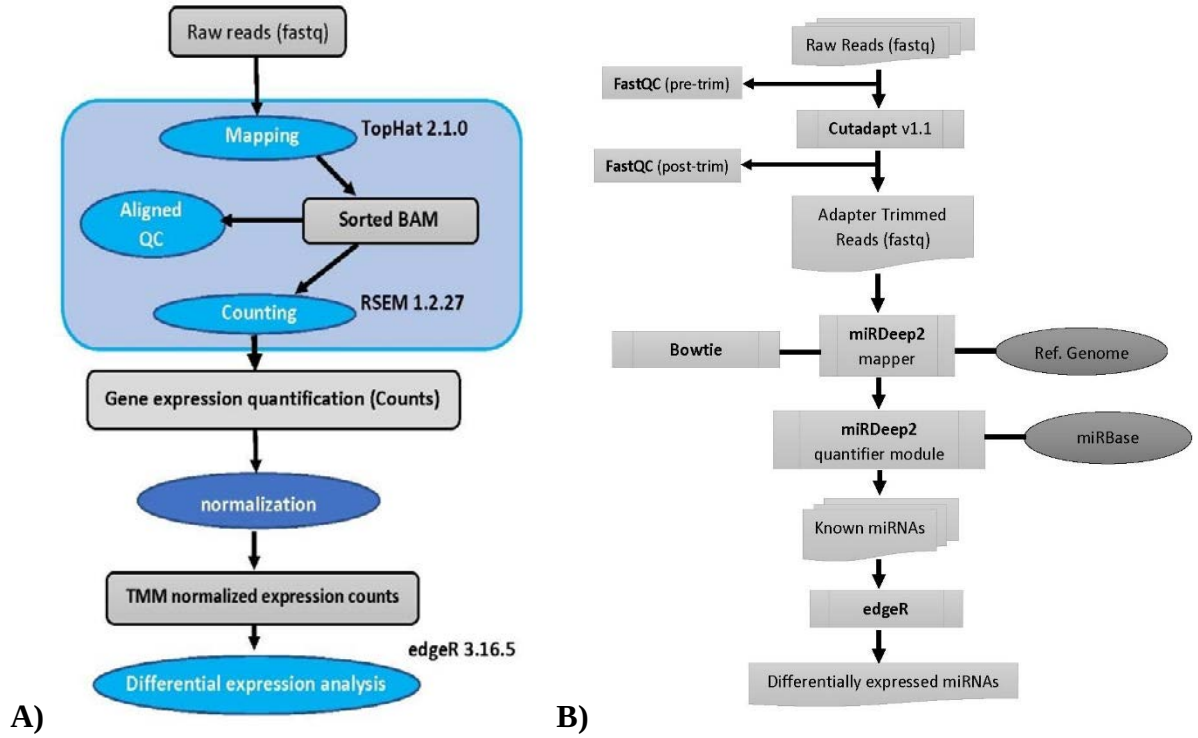


Figure 3.4: Workflow of RNA and miRNA sequencing analysis. **A**) Typical workflow for RNA sequencing (RNA-seq) data processing and analysis. **B**) Workflow of miRNA sequencing (miRNA-seq) data processing and analysis.

3.7.2. Small RNA-seq data processing

Preliminary quality control analysis of the 17 miRNAseq fastq files from the eBL patients and the 4 miRNAseq fastq files from GC B cells obtained from Gene Expression Omnibus (GEO) archive, were carried out with FASTQC software v0.10.0 (Andrews, 2014). Cutadapt v1.1 (Martin, 2011) was then used to trim off the 3'-adaptor sequences from the sequencing reads. Novobarcode ("Novocraft," n.d.) was used to de-multiplex the 17 eBL samples based on the 6-nucleotide barcode that was added to the small RNA sequencing library of each sample. Reads shorter than 18 nucleotides after adaptor trimming and barcode removal were discarded. The trimmed reads were then further checked for the presence of any artificial sequences from the adaptor or barcode using FASTQC. Reads passing all the above filters were aligned to the human genome (hg19) using bowtie (Langmead, 2010). Reads that did not uniquely align to the genome were discarded. The resulting sequences were then analysed using miRDeep2 (Friedländer *et al.,* 2012) to determine the miRNA counts for each of the samples (Figure 3.4B).

3.8. Statistical Analysis

3.8.1. Differential gene expression analysis

After quality assessment and pre-processing the raw sequencing reads, mRNA read pairs were aligned to a transcriptome index built by RSEM (Li and Dewey, 2011) using Gencode v19 protein coding transcript annotations and hg19 genomic sequence. To perform differential gene expression test between 28 eBL tumours and 5 GC B-cells, edgeR (Robinson *et al.,* 2010) package was used in R computing environment. To be able to account for the batch variables and unknown factors while testing for the differential expression between the eBL tumours and GC B-cell RNA expression data from another dataset, the number of latent factors for every comparison were estimated using svaseq (Leek, 2014) while preserving the variation of interest (Appendix 6). These surrogate variables were then incorporated into the

testing model for edgeR. Since sometimes small p-values (less than 5% or 1%) happen by chance in multiple comparison tests, not adjusting for false positives could lead to incorrectly rejecting the true null hypotheses. P-values were adjusted for multiple testing with the Benjamini and Hochberg (BH) approach (Benjamini and Hochberg, 1995) for adjusting the false discovery rate (FDR) and adjusted *p*-values were filtered at ≤0.01. Significantly differentially expressed (DE) mRNAs had BH multiple test corrected *p*-values < 0.01.

3.8.2. Gene set enrichment analysis

A standard gene set enrichment analysis (GSEA) was performed using the GSEA module implemented by Broad Institute, Cambridge, MA. Gene set enrichment analysis was performed on normalized expression data and on data after surrogate variable analysis to remove the batch effect. For a ranking metric, signal to noise value of each gene was used, and a permutation test for FDR was performed by permuting sample phenotypes (eBL tumour cells and GC B cells). The analysis included standard gene sets of hallmark and oncogenic signatures as well as the curated C2 gene sets from the Molecular Signatures Database (v5.0 MSigDB).

3.8.3. Differential miRNA expression analysis

Differential miRNA expression was performed between the 17 eBL tumour cells and 4 GC B cells. This expression analysis of miRNA-Seq data was also performed using the R/Bioconductor package edgeR (Robinson *et al.*, 2010). First, the number of reads that uniquely mapped to miRNA regions according to the reference database miRbase, were counted. Only miRNAs that had at least 10 counts per million in at least half of the samples were analysed for evidence of differential gene expression. The biological reason for this is that a miRNA must be expressed at some minimal level before it is likely to affect gene regulation. The statistical reason was that very low counts would provide little statistical

information to distinguish between the null and the alternative hypothesis (Kozomara and Griffiths-Jones, 2014). To account for the batch variables and unknown factors while preserving the variation of interest for the differential expression analysis svaseq (Leek, 2014) batch correction package was used. These surrogate variables were then incorporated into the testing model for edgeR. Since sometimes small *p*-values (less than 5% or 1%) happen by chance in multiple comparison tests, not adjusting for false positives could lead to incorrectly rejecting the true null hypotheses. P-values were adjusted for multiple testing with the Benjamini and Hochberg approach (Benjamini and Hochberg, 1995) for adjusting the FDR and adjusted *p*-values were filtered at 0.01. Significantly DE miRNA also had BH multiple test corrected *P*-values <0.01.

3.8.4. Network propagation method to infer the perturbed miRNA regulatory network using differential gene expression data

When a miRNA is dysregulated, it will impact on its direct target genes and subsequently affect the expression of the downstream genes through intracellular molecular regulatory networks, which would finally change the gene expression patterns in a cell. Therefore, a network propagation based model (which is a family of stochastic processes that trace the flow of information through a network over time), should be used to interpret the global transcriptional response to miRNA dysregulation.

The network propagation based method (NP-method) (Wang *et al.*, 2014), was used to infer the key miRNA regulatory networks whose perturbation was most likely to induce the observed gene expression changes in eBL compared to their normal counterpart. By integrating eBL differential gene expression data with prior biological knowledge of miRNA-target interactions ("TargetScanHuman 6.2," n.d.) and the TF (Transcription factor)-gene regulatory network (HTRIdb) (Bovolenta *et al.*, 2012), a network-based random walk with

restart (RWR) plus forward searching algorithm (Ham *et al*., 2013) was carried out to calculate the network perturbation effect score (NPES) of miRNAs. To avoid bias towards miRNAs with a large target set, gene set permutation based analysis repeated 1000 times was performed to normalize the score and estimate the *p*-value for each miRNA.

3.8.5. Differentially expressed (DE) microRNA target identification

MiRNAs regulate expression of specific genes via hybridization to mRNA transcripts to promote RNA degradation, inhibit translation or both (Behm-Ansmant *et al*., 2006). Identification of target genes of the aberrantly expressed miRNAs is important for understanding the regulatory networks associated with eBL development. To investigate the biological relevance of the identified DE miRNAs, all the validated target genes for the DE miRNAs were identified using the validated target module of the miRWalk2.0 (Dweep *et al.*, 2011; Dweep and Gretz, 2015) database.

3.8.5.1. MicroRNA Target Gene Correlation Coefficient Calculation

Using mRNA expression data for the eBL patients in this study, Pearson correlation coefficients between a miRNA and its validated targets were computed using R statistical programming language, to determine whether the expression levels of each miRNA and of its mRNA targets show an inverse correlation.

3.8.6. MicroRNA-mRNA pairs of interest

To identify miRNA-mRNA pairs of interest, the DE validated target genes of the DE miRNAs, that exhibited an inverse expression change to the miRNA were selected. These miRNA-mRNA pairs were tested overall to determine if the number of pairs was more than expected by chance. To achieve this, a permutation test of significance repeated 10,000 times was performed. The permutation tested whether the number of miRNA-mRNA pairs were more than would be expected by chance.

3.8.7. Gene Ontology and Kyoto Encyclopaedia of Genes and Genomes pathway enrichment analysis

For functional analyses of the miRNA targets, gene ontology (GO) term analysis was applied to organize genes into categories based on biological processes, cellular components and molecular functions. Biological pathways defined by Kyoto Encyclopaedia of Genes and Genomes (KEGG) analysis were identified by DAVID (Database for Annotation, Visualization and Integrated Discovery) software (Dennis *et al.*, 2003). DAVID online database was used to provide a set of functional annotations of the majority of genes. *P*-values of each pathway were adjusted using the Benjamini-Hochberg method to control the FDR. In the current study, GO terms and signalling pathways were selected with the threshold of significance being defined as $P < 0.01$ and FDR < 0.05.

3.9. Validation of MiRNA Expression Data using Quantitative Reverse Transcription PCR

MiRNA expression levels were validated by quantitative reverse transcription polymerase chain reaction (RT-qPCR) using TaqMan miRNA assay protocol (Invitrogen, Thermo Scientific, CA, USA.) on the BioRad CFX96 Real-Time System. The experiments were run in triplicates on 17 eBL miRNA samples sequenced. To normalize the expression levels of the target miRNA by correcting for the amount of cDNA loaded to the PCR reaction, the comparative Ct method was used. All Ct-values were normalized to an endogenous control (U54), and ΔCt-values were calculated, where $\Delta Ct = Ct\ (miRNA) - Ct\ (U54)$ (Livak and Schmittgen, 2001; Marabita *et al.*, 2016). Relative expression values $(2^{-\Delta Ct})$ were plotted and compared. To verify mean differences among the groups, the normalized PCR data was analysed using the Wilcoxon rank test in R. A two-sided $p < 0.05$ was considered statistically significant.

3.10. Ethical Considerations

Approval to carry out this study was provided by the School of Graduate Studies (SGS) of Maseno University. Ethical approval was obtained from the Scientific and Ethics Review Unit (SERU) at the Kenya Medical Research Institute (KEMRI), Kenya, and University of Massachusetts Medical School, Institutional Review Committee. The clinical information was obtained under the approval of the Scientific and Ethics Review Unit, of the Kenya Medical Research Institution and written informed consent was obtained from the parents and guardians of eBL patients.

CHAPTER FOUR

RESULTS

4.1. Demographic, Clinical and Sequencing Information of the eBL Samples

To survey the transcriptome expression changes in eBL, mRNAs and miRNAs were sequenced from 28 and 17 primary tumour FNA biopsies, respectively, collected from Kenyan children with a median age of 8.2 years (Table 4.1). For the eBL patients, the tumour-presenting site was 43% (12/28) jaw tumours and 57% (16/28) abdominal tumours. In terms of survival, the eBL samples included 2 patients who died prior to receiving any treatment, 7 patients who died during the initial course of treatment, and 5 patients who were able to complete the recommended chemotherapy treatment with resolution of their tumour and discharged from hospital. For each of the 28 samples, strand specific RNA sequencing was performed, generating on average 14 million paired reads per library (range, 8.9 -53.7 million (M) reads, Appendix 4). Of the 28 eBL tumour samples, only 17 samples had a miRNA concentration >5ng/µl which was the recommended minimum required for sequencing library preparation. Therefore, miRNA sequencing was performed on 17 eBL samples, generating on average 1.4 M reads per library (range, 0.5 – 3M reads, Appendix 4) and 2,042 distinct human miRNAs were detected using miRDeep2. All the 28 samples in the sequencing set showed high expression of BL associated markers, including conventional B cell surface markers CD19, CD20 and CD79A/B (Figure 4.1a) and intracellular markers such as BCL6 and MYC, consistent with the molecular phenotype of BL (Hecht and Aster, 2000).

Table 4.1. Demographic and Clinical characteristics of eBL patients.

Characteristics	Total (N=28)
Age (years), median (range)	8.2 (2-14)
Gender, n (%)	
Male	20 (71%)
Female	8 (29%)
Tumour FNA site, n (%)	
Jaw	12 (42.9%)
Abdomen	16 (57.1%)
In-hospital survival-status, n (%)	
Survived[¥]	5 (29.4%)
Died[¶]	7 (41.2%)
Relapsed	1 (5.9%)
Died prior to chemotherapy[§]	2 (11.8%)
Unknown outcome (patient referred to another hospital)	2 (11.8%)

[¥]eBL patients who completed chemotherapy, were discharged from hospital and still alive after 2 years of follow-up. [¶]Patients who had started chemotherapy but died during the initial course of treatment. [§]eBL patient who died before starting chemotherapy.

4.2. Expression of Germinal Center (GC) B-cell Differentiation Genes in eBL Tumours

To ensure proper gene expression signatures consistent with a GC B-cell phenotype, the expression levels of B cell differentiation genes in eBL tumour cells were checked if they were at comparable levels to the GC derived B cells. RNA expression counts of key GC transcription factors (*BCL6* and *PAX5*) were well expressed while plasma cell genes (*BLIMP1* and *IRF4*) were at low levels. Overall, this supports a GC B-cell like tumour phenotype of the eBL tumour samples and the validity of further comparisons between the eBL tumour cells and normal GC B cells (Figure 4.1b).

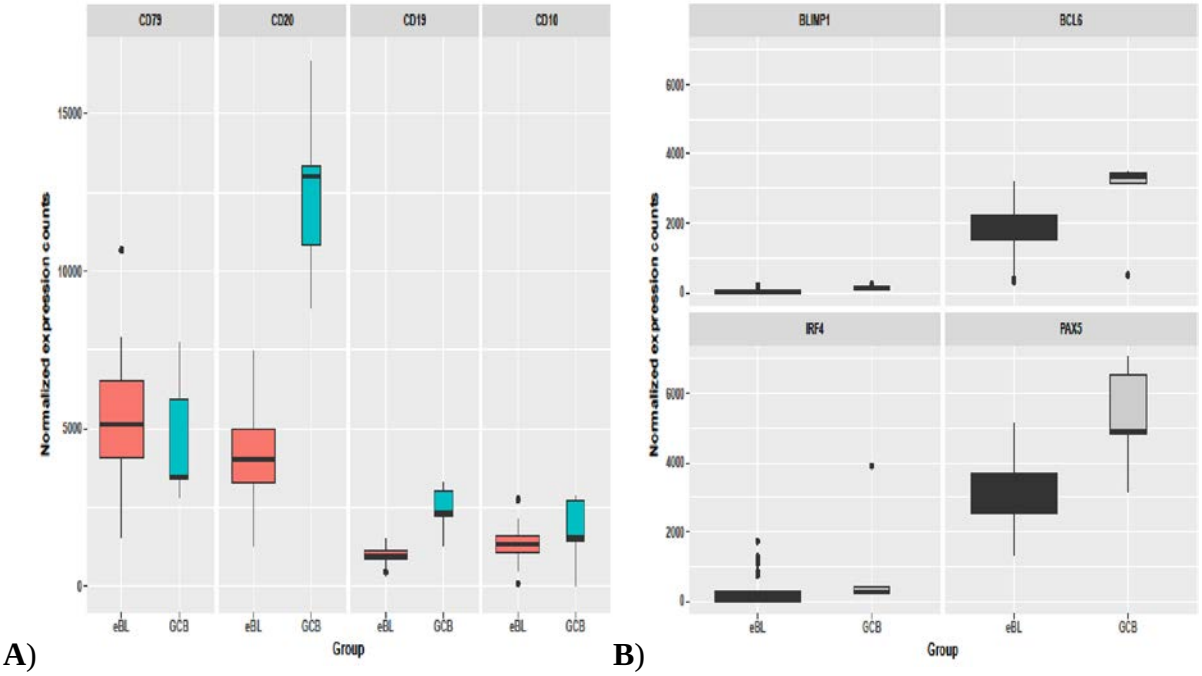


Figure 4.1: Expression of Germinal Center (GC) B-cell Differentiation Genes. **A**) Expression of eBL diagnostic surface markers. **B**) Expression of key transcription factors involved in B-cell differentiation.

4.3. Global Gene Expression Profiling of Endemic BL Tumours.

To identify genes that may contribute to the oncogenic phenotype of eBL, gene expression profiling was conducted on 28 eBL tumour samples and 5 GC B cells. As shown in Appendix 7, hierarchical clustering of most variant genes revealed a clear separation of the two groups where the eBL samples are clearly differentiated from their normal counterparts. Performing differential gene expression analysis between eBL and normal GC B cells, 211 genes were identified to be differentially expressed (DE) using stringent thresholds (logFC>2, *p*-value<0.01 and FDR<0.01) (Figure 4.2, Table 4.2). Of these DE genes, 132 genes were downregulated and 79 were upregulated in eBL compared to their normal GC counterparts (Table 4.2). Among the upregulated genes was *MYC*, whose over-expression is central to BL oncogenesis (logFC=3.07, *p*-value=5.50E-24 and a FDR=1.15E-22) and genes coding for mitochondrial proteins (*MT-ND3, MT-ND4L, MT-ND4, MT-ND2, MT-CYB, MT-ND1, MT-*

CO1 etc), which may be as a result of the elevated metabolism characteristic of cancer cells to sustain their survival. Transcription factor activating enhancer-4 (*TFAP4*), a direct transcriptional target of *MYC* to induce cell cycle progression, was also observed to be upregulated (logFC=2.12, *p*-value=1.18E-35 and FDR= 6.35E-34) in eBL. The protein kinases *ATM* (ataxia-telangiectasia mutated) (logFC=-2.46, *p*-value<0.0001 and FDR<0.00001), an activator of the DNA damage response in the face of DNA double strand breaks (DSBs), and *NLK* (nemo-like kinase) (logFC=-2.55, *p*-value<0.0001 and FDR<0.0001), a p53 activator, were among the downregulated tumour suppressor genes in eBL tumour cells.

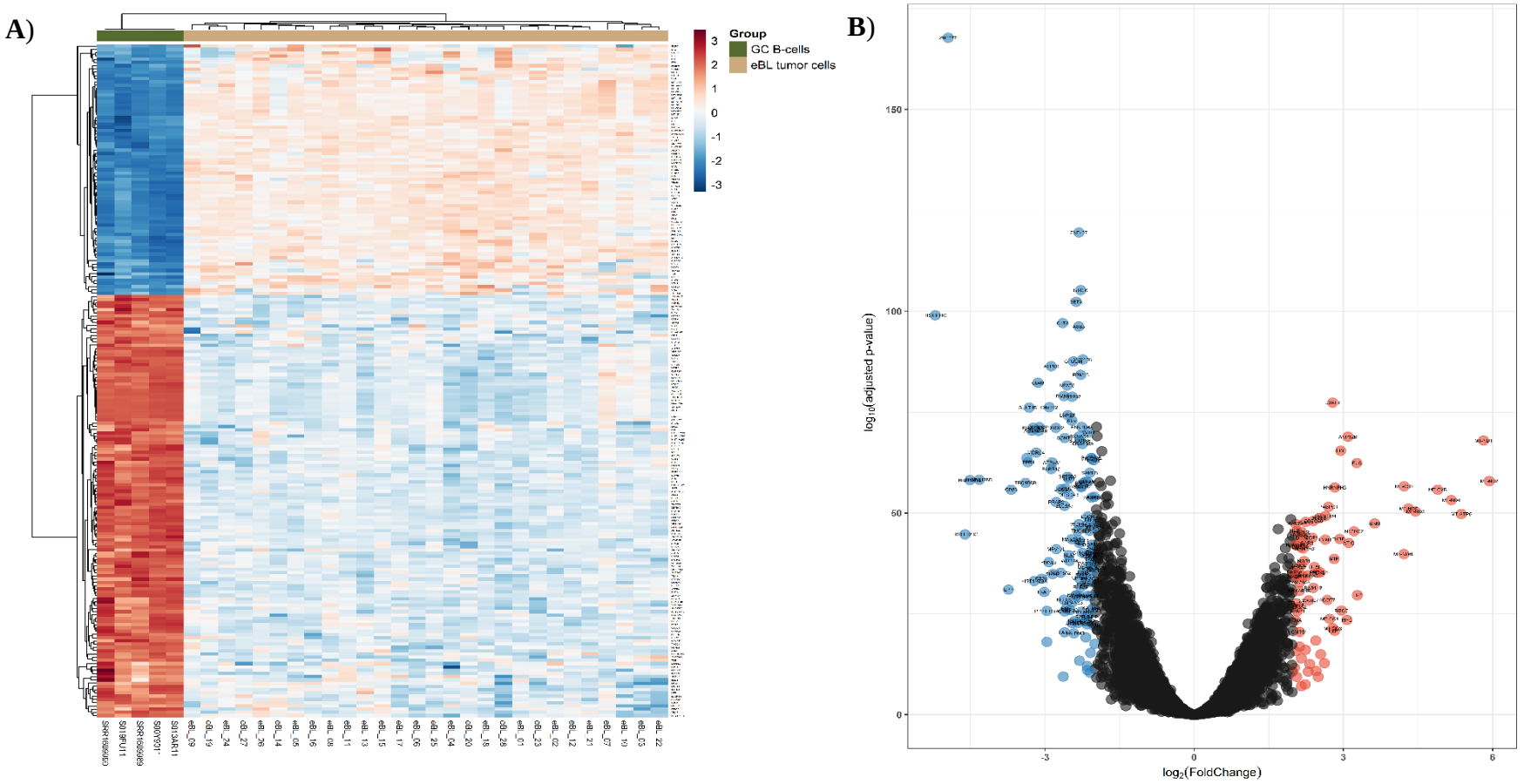


Figure 4.2: Differentially expressed (DE) mRNAs in eBL compared to germinal center (GC) B cells. **A**). Heatmap of DE mRNAs between eBL tumour cells and GC B cells. The heatmap shows the hierarchical clustering based on the expression profiles of the 211 DE genes with at least 2-fold difference in expression compared to their normal counterpart. **B**). Volcano plot representing the significance of the genes (negative-log of the adjusted p-value) vs the fold change difference in eBL compared to GC B-cells. The red and blue coloured circles represent genes which are significantly DE with $p<0.01$ and $FDR<0.01$. The 132 down-regulated genes in eBL are coloured blue (have a negative fold-change value) while the 79 up-regulated genes in eBL are coloured red (positive fold-change value).

Table 4.2: Differentially expressed Genes between eBL tumor cells and GC B cells (logFC>2, *p*-value<0.01 and FDR<0.01).

Gene symbol	Log FC eBL versus GC B cells	BH adjusted *p*-value	FDR
Upregulated genes			
MT-ND2	5.9357	4.60E-44	4.76E-42
MT-ND1	5.8160	1.43E-56	3.78E-54
MT-ATP6	5.4025	1.68E-30	6.24E-29
MT-ND4	5.2228	2.01E-33	9.30E-32
MT-CYB	4.9183	3.80E-34	1.93E-32
MT-ND3	4.4391	1.23E-42	1.15E-40
MT-ND5	4.3117	9.79E-41	7.60E-39
MT-ND4L	4.2542	4.32E-25	1.03E-23
MT-CO2	4.2428	1.71E-32	7.41E-31
SRM	3.5823	1.65E-40	1.23E-38
FUS	3.3145	4.09E-59	1.33E-56
HK2	3.2878	6.02E-27	1.65E-25
NT5DC2	3.1931	1.06E-36	6.01E-35
ATIC	3.1325	1.21E-42	1.15E-40
ANP32B	3.1142	2.94E-65	1.40E-62
MYC	3.0653	5.50E-24	1.15E-22
RGCC	2.9451	2.69E-26	6.93E-25
FKBP4	2.9411	1.86E-36	1.03E-34
FBL	2.9190	1.51E-62	5.67E-60
HNRNPH3	2.8531	2.67E-54	6.36E-52
FTH1	2.8214	1.32E-19	1.87E-18
MT-CO3	2.8068	6.79E-15	6.00E-14
U2AF1	2.7932	2.18E-63	8.65E-61
MIF	2.7535	6.98E-32	2.90E-30
PPDPF	2.7214	5.45E-14	4.32E-13
PABPC1	2.7203	1.72E-43	1.70E-41
MT-CO1	2.7027	1.13E-16	1.19E-15
ACOT7	2.6869	1.99E-23	3.91E-22
LYAR	2.6680	8.60E-36	4.65E-34
CCDC124	2.5947	1.70E-37	1.03E-35
PFDN2	2.5663	6.84E-23	1.30E-21

Table 4.2 (*Continued*)

Official gene symbol	Log FC eBL versus GC B cells	BH adjusted *p*-value	FDR
NPRL3	2.5300	4.83E-33	2.16E-31
HMBS	2.5000	4.87E-39	3.16E-37
GNG7	2.4959	8.30E-15	7.18E-14
TOP1MT	2.4562	1.49E-11	8.87E-11
SOX12	2.4545	3.12E-19	4.29E-18
EIF5B	2.4539	2.83E-38	1.76E-36
TOP1	2.4178	3.53E-43	3.40E-41
CTSD	2.4113	3.08E-08	1.23E-07
CCDC86	2.3837	1.19E-35	6.37E-34
RAB11B	2.3754	6.45E-31	2.45E-29
TXN	2.3506	2.95E-13	2.11E-12
GAR1	2.3093	1.23E-35	6.49E-34
PCBP1	2.2807	5.41E-44	5.52E-42
CD68	2.2736	1.53E-10	8.08E-10
C20orf27	2.2525	2.61E-25	6.26E-24
FTL	2.2334	7.24E-15	6.33E-14
ATP6V0C	2.2316	3.53E-24	7.56E-23
HNRNPA3	2.2102	2.00E-40	1.47E-38
LYZ	2.2069	1.03E-07	3.84E-07
HNRNPD	2.2003	2.40E-38	1.50E-36
UCK2	2.1937	3.11E-10	1.60E-09
NOB1	2.1929	1.81E-36	1.01E-34
NCL	2.1928	3.04E-29	1.01E-27
UTP4	2.1917	1.58E-39	1.07E-37
BANF1	2.1898	1.48E-33	6.99E-32
BMP7	2.1813	9.40E-08	3.52E-07
CMSS1	2.1591	5.69E-17	6.20E-16
PCBP2	2.1560	1.08E-41	9.11E-40
SF3A2	2.1431	1.25E-26	3.29E-25
SLC25A6	2.1372	2.33E-14	1.93E-13
TFAP4	2.1232	1.18E-35	6.35E-34
COPS6	2.0979	5.94E-36	3.24E-34
FAM216A	2.0787	2.56E-29	8.64E-28

Table 4.2 (*Continued*)

Official gene symbol	Log FC eBL versus GC B cells	BH adjusted *p*-value	FDR
CHERP	2.0697	4.87E-42	4.24E-40
LDHA	2.0676	3.68E-19	5.00E-18
TTLL12	2.0642	8.86E-25	2.04E-23
HADH	2.0619	1.52E-29	5.21E-28
NPM3	2.0613	2.31E-15	2.12E-14
FKBP5	2.0583	3.28E-16	3.29E-15
PRKAR1B	2.0560	1.05E-23	2.13E-22
RAB7A	2.0516	1.67E-41	1.34E-39
SRP68	2.0501	5.96E-34	2.96E-32
CYB5A	2.0486	1.93E-09	9.03E-09
TRAP1	2.0446	2.81E-20	4.24E-19
TFDP2	2.0443	9.52E-08	3.56E-07
NME1	2.0346	2.83E-31	1.10E-29
GPATCH4	2.0234	5.09E-18	6.15E-17
NDUFS6	2.0175	4.42E-32	1.88E-30
Downregulated genes			
HIST1H1C	-5.2047	1.08E-102	3.86E-99
ZNF277	-4.9458	5.31E-131	3.79E-127
HIST1H2BC	-4.7110	2.20E-44	2.35E-42
HIST1H1E	-4.5599	7.48E-56	1.84E-53
HIST1H2BD	-4.3218	1.12E-56	3.07E-54
CD83	-3.7611	9.66E-57	2.76E-54
LPP	-3.7011	4.12E-28	1.25E-26
TBC1D8B	-3.4523	1.24E-41	1.01E-39
SULT1B1	-3.3667	7.31E-53	1.49E-50
ASTN2	-3.3030	9.53E-48	1.26E-45
KIAA1328	-3.2938	1.57E-50	2.80E-48
ITPR1	-3.2590	1.64E-48	2.45E-46
HIST1H2BK	-3.2317	1.28E-35	6.72E-34
CD40	-3.2115	1.01E-82	1.45E-79
MORC4	-3.1918	5.71E-44	5.74E-42

Table 4.2 (*Continued*)

Official gene symbol	Log FC eBL versus GC B cells	BH adjusted *p*-value	FDR
SHISA9	-3.1803	4.54E-48	6.49E-46
WDPCP	-3.1656	3.46E-54	7.96E-52
CTSH	-3.0781	5.34E-33	2.37E-31
TRAF1	-3.0247	7.46E-30	2.64E-28
PPP1R15A	-2.9682	2.52E-24	5.51E-23
C4orf32	-2.9633	1.21E-64	5.38E-62
EML6	-2.9589	1.69E-18	2.16E-17
FMNL3	-2.9514	6.22E-34	3.06E-32
PARP15	-2.9012	8.66E-40	6.06E-38
ATP9B	-2.8848	6.99E-81	6.24E-78
ATP8A1	-2.8567	3.86E-49	6.27E-47
SIAH2	-2.8279	1.29E-26	3.39E-25
RPAP2	-2.8176	1.64E-42	1.50E-40
MPV17L	-2.8055	3.65E-33	1.68E-31
DUSP2	-2.7968	9.65E-68	4.92E-65
SLC15A2	-2.7075	1.12E-36	6.29E-35
UGCG	-2.6922	6.17E-50	1.07E-47
FCRL3	-2.6601	3.85E-10	1.95E-09
CNR2	-2.6413	2.43E-24	5.33E-23
SOD2	-2.6365	6.66E-38	4.07E-36
GCNT2	-2.6352	3.23E-64	1.36E-61
LAT2	-2.6183	1.43E-21	2.38E-20
CIITA	-2.6057	2.99E-82	3.56E-79
DMXL1	-2.6054	8.55E-71	5.09E-68
NFAT5	-2.5836	2.10E-47	2.72E-45
KLHL6	-2.5731	2.37E-28	7.43E-27
BMF	-2.5619	1.86E-16	1.91E-15
WDR74	-2.5548	2.07E-37	1.23E-35
SLC9A7	-2.5497	2.25E-44	2.37E-42
NLK	-2.5494	3.82E-22	6.84E-21
MFSD4B	-2.5484	1.95E-44	2.11E-42

Table 4.2 (*Continued*)

Official gene symbol	Log FC eBL versus GC B cells	BH adjusted *p*-value	FDR
LNPEP	-2.5450	3.01E-72	1.95E-69
TSPAN33	-2.5340	7.90E-28	2.34E-26
MEF2C	-2.5067	1.05E-26	2.80E-25
SFT2D2	-2.4950	3.03E-47	3.86E-45
TEP1	-2.4845	1.24E-41	1.01E-39
ATM	-2.4584	7.75E-49	1.18E-46
HLA-DMB	-2.4387	2.48E-21	4.05E-20
LPIN1	-2.4353	2.33E-21	3.81E-20
OTULIN	-2.4171	1.71E-81	1.74E-78
PAG1	-2.4169	5.56E-22	9.77E-21
TMC6	-2.3908	7.61E-34	3.72E-32
ZBTB37	-2.3905	2.24E-53	4.84E-51
MAN2A1	-2.3811	3.03E-30	1.12E-28
BRWD1	-2.3714	8.52E-24	1.75E-22
MACF1	-2.3624	2.23E-23	4.36E-22
ASB3	-2.3477	2.50E-78	1.98E-75
MCTP2	-2.3322	7.89E-28	2.34E-26
OPHN1	-2.3314	1.58E-18	2.03E-17
SETX	-2.3267	1.79E-61	6.07E-59
OXNAD1	-2.3221	1.95E-69	1.07E-66
REL	-2.3205	1.46E-42	1.35E-40
BLOC1S6	-2.3163	3.19E-51	5.83E-49
WDFY4	-2.3045	2.02E-45	2.32E-43
C11orf72	-2.3037	6.66E-26	1.67E-24
VPS13C	-2.2895	7.70E-47	9.48E-45
NEDD9	-2.2881	5.20E-31	1.99E-29
ATAD2B	-2.2862	9.03E-30	3.16E-28
BIRC6	-2.2836	3.20E-97	7.62E-94
RAB8B	-2.2811	2.24E-30	8.28E-29
RBM15	-2.2722	7.50E-78	5.36E-75
SNX11	-2.2719	3.24E-56	8.25E-54
ZNF407	-2.2704	6.55E-88	1.17E-84

Table 4.2 (*Continued*)

Official gene symbol	Log FC eBL versus GC B cells	BH adjusted *p*-value	FDR
KIAA1109	-2.2655	1.23E-51	2.31E-49
CYTH1	-2.2623	9.70E-29	3.15E-27
PTPN6	-2.2551	1.56E-13	1.16E-12
JUNB	-2.2435	9.31E-30	3.24E-28
SLC23A2	-2.2216	9.73E-22	1.65E-20
ANKRD44	-2.2157	5.74E-48	7.73E-46
TMED8	-2.2117	8.92E-28	2.62E-26
STAT3	-2.2113	1.17E-22	2.17E-21
BMP2K	-2.2087	3.70E-48	5.39E-46
ZBTB43	-2.1998	7.75E-42	6.66E-40
VPS13B	-2.1947	6.06E-59	1.88E-56
ZNF141	-2.1742	2.10E-16	2.15E-15
DNAJC10	-2.1648	2.94E-24	6.39E-23
NFKBIE	-2.1622	6.95E-41	5.51E-39
CFLAR	-2.1613	6.25E-32	2.62E-30
SLC35F5	-2.1586	2.77E-29	9.28E-28
IL12RB1	-2.1550	1.31E-23	2.63E-22
SORL1	-2.1449	4.64E-12	2.91E-11
TLE4	-2.1448	4.26E-25	1.02E-23
LCOR	-2.1414	1.18E-45	1.39E-43
CMTM6	-2.1398	2.09E-53	4.67E-51
SMG1	-2.1383	5.90E-62	2.11E-59
HERC1	-2.1363	1.04E-39	7.04E-38
SAMD8	-2.1061	2.44E-45	2.77E-43
ITGAL	-2.0953	4.84E-21	7.67E-20
RAB11FIP1	-2.0934	1.25E-15	1.17E-14
MYO1E	-2.0826	5.37E-09	2.37E-08
PIK3CG	-2.0811	2.55E-27	7.24E-26
ITPKB	-2.0741	7.00E-18	8.38E-17
MALT1	-2.0723	1.32E-29	4.52E-28
TACC1	-2.0676	5.67E-25	1.32E-23
ZNF780A	-2.0604	3.77E-57	1.12E-54

Table 4.2 (*Continued*)

Official gene symbol	Log FC eBL versus GC B cells	BH adjusted *p*-value	FDR
LHFPL2	-2.0599	3.46E-11	1.97E-10
TTLL3	-2.0568	1.96E-08	8.05E-08
PIKFYVE	-2.0545	3.54E-46	4.29E-44
DCP2	-2.0539	9.41E-25	2.15E-23
DOPEY1	-2.0539	4.83E-39	3.16E-37
MTMR14	-2.0514	3.00E-43	2.93E-41
PHF6	-2.0513	5.76E-45	6.43E-43
SNX29	-2.0407	2.64E-40	1.90E-38
SLC12A2	-2.0393	1.84E-33	8.60E-32
KATNAL1	-2.0385	1.79E-26	4.69E-25
RELT	-2.0343	3.31E-29	1.09E-27
IRGQ	-2.0287	7.49E-23	1.41E-21
RNF19A	-2.0278	4.32E-40	3.09E-38
INO80D	-2.0251	1.13E-40	8.59E-39
ORC4	-2.0249	6.13E-30	2.18E-28
PASK	-2.0244	6.28E-23	1.20E-21
CYB561A3	-2.0195	8.56E-27	2.30E-25
LRRC37A2	-2.0179	5.80E-37	3.40E-35
RABGAP1L	-2.0090	1.11E-24	2.51E-23
CDK17	-2.0075	4.62E-33	2.09E-31
PHC3	-2.0040	7.32E-53	1.49E-50
ITSN2	-2.0011	1.09E-21	1.83E-20

Genes differentially expressed (DE) between eBL tumour cells and GC B cells. Genes that meet a cut off of logFC>2, *p*-value<0.01 and FDR<0.01, were considered to be significantly DE. 211 genes were DE in eBL. Some of the DE genes were tumour suppressors such as *ATM*, which downregulated in eBL tumours. Abbreviations: eBL, endemic Burkitt lymphoma; GC, germinal center; BH, Benjamini & Hochberg; FC, Fold Change; FDR, False discovery rate.

4.3.1. Gene sets enriched in eBL tumour cells

To identify more subtle changes in overall pathways and functional sets of genes, a gene set enrichment analysis (GSEA) was performed that looks at the set of genes in concert for deviations in expression. The analysis, using a FDR<0.25 (=25% error rate), detected only one enriched gene set. This was the gene set HALLMARK_MYC TARGETS_V1 (a set of genes regulated by *MYC*) (Figure 4.3), which again highlights MYC's pivotal role in eBL oncogenesis. This comparison confirms that the eBL dataset is consistent with expected differences between normal and cancerous B cells.

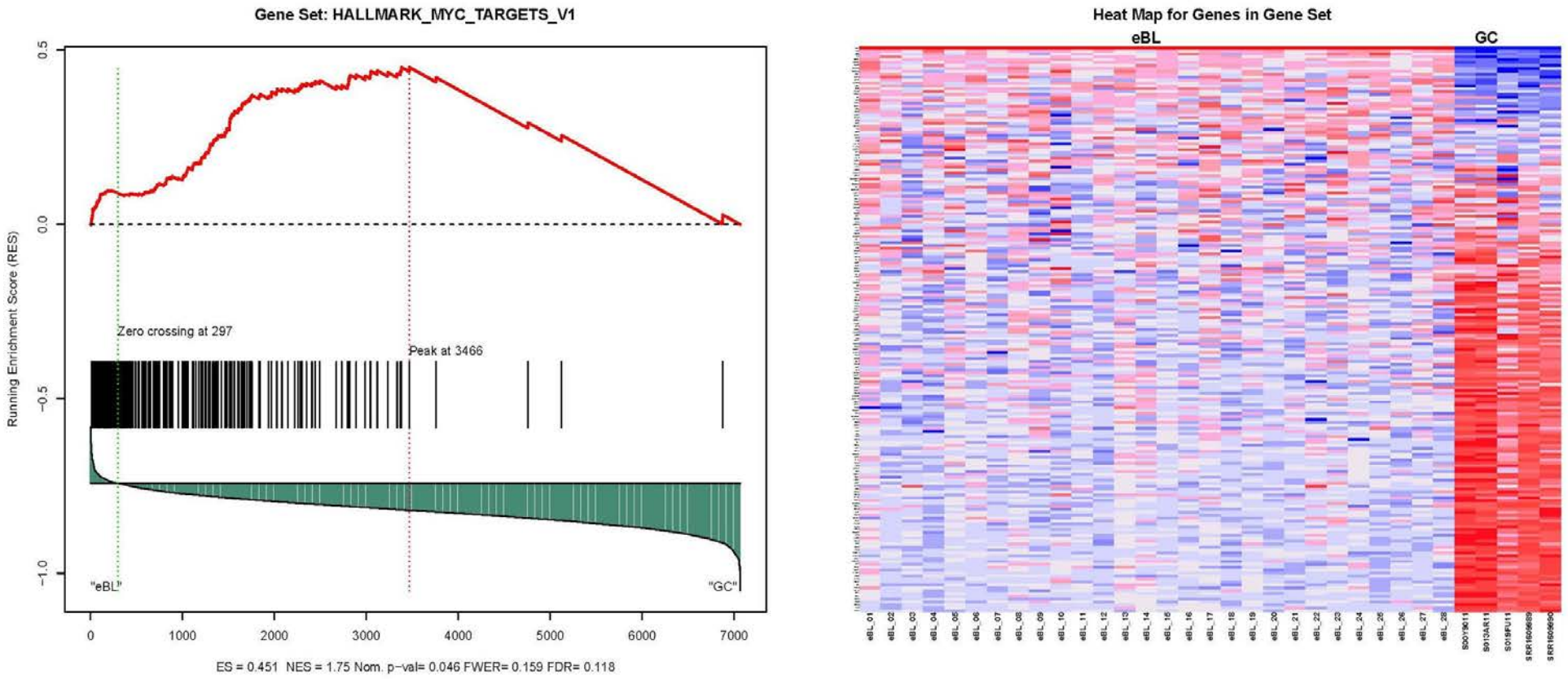


Figure 4.3: Gene set enrichment plot and expression heatmap of corresponding genes in the enriched gene set. Left panels include the running enrichment score throughout the gene set and projection of genes in the gene-set to the complete list of genes rank ordered based on signal to noise ratio. On the expression heatmap (columns are eBL tumours and GC B cells, rows are genes in the gene set), dark red represent higher expression while dark blue lower expression. Genes in this enrichment are a set of genes regulated by MYC in eBLs tumour cells relative to GC B cells (ES = 0.45, Nominal P = 0.046, FDR q = 0.118).

4.4. MicroRNA Expression Profiling in eBL Tumours

To identify miRNAs that may contribute to the oncogenic phenotype of eBL, miRNA expression profiling was conducted on 17 eBL tumours compared to 4 GC B cells. Hierarchical cluster analysis on the expression profile of the most variant miRNAs clearly separated the GC B cells from the eBL tumour cells (Appendix 7), similar to the gene expression data. The 49 miRNAs were identified to be significantly differentially expressed (logFC>2, *p*-value<0.01 and FDR<0.01 (1% error rate)) between eBL and GC B cells. Of these DE miRNAs, 27 miRNAs were downregulated and 22 were identified to be upregulated in eBL tumour cells compared to their normal counterpart (Figure 4.4, Table 4.3). The downregulated miRNAs included let-7 family members (let-7a-5p, let-7b-5p, let-7c, let-7d-5p, let-7e-5p, let-7f-5p, let-7g-5p) (logFC<-2.5), which all target *MYC* gene for post-transcriptional regulation (Braun *et al*., 2014; Helwak *et al.*, 2013; Kuchen *et al.*, 2010; Lu *et al.*, 2005; Lu and Clark, 2012). Among the upregulated miRNAs in eBL were members of the miR-17~92 cluster (miR-19b-3p, and miR-92a-3p) (logFC>3), which target tumour suppressor genes such as *TP53* (Ricarte-Filho *et al.*, 2009) and *ATM* (ataxia telangiectasia mutated) kinase (Lan *et al.*, 2011; Lu *et al*., 2005), respectively.

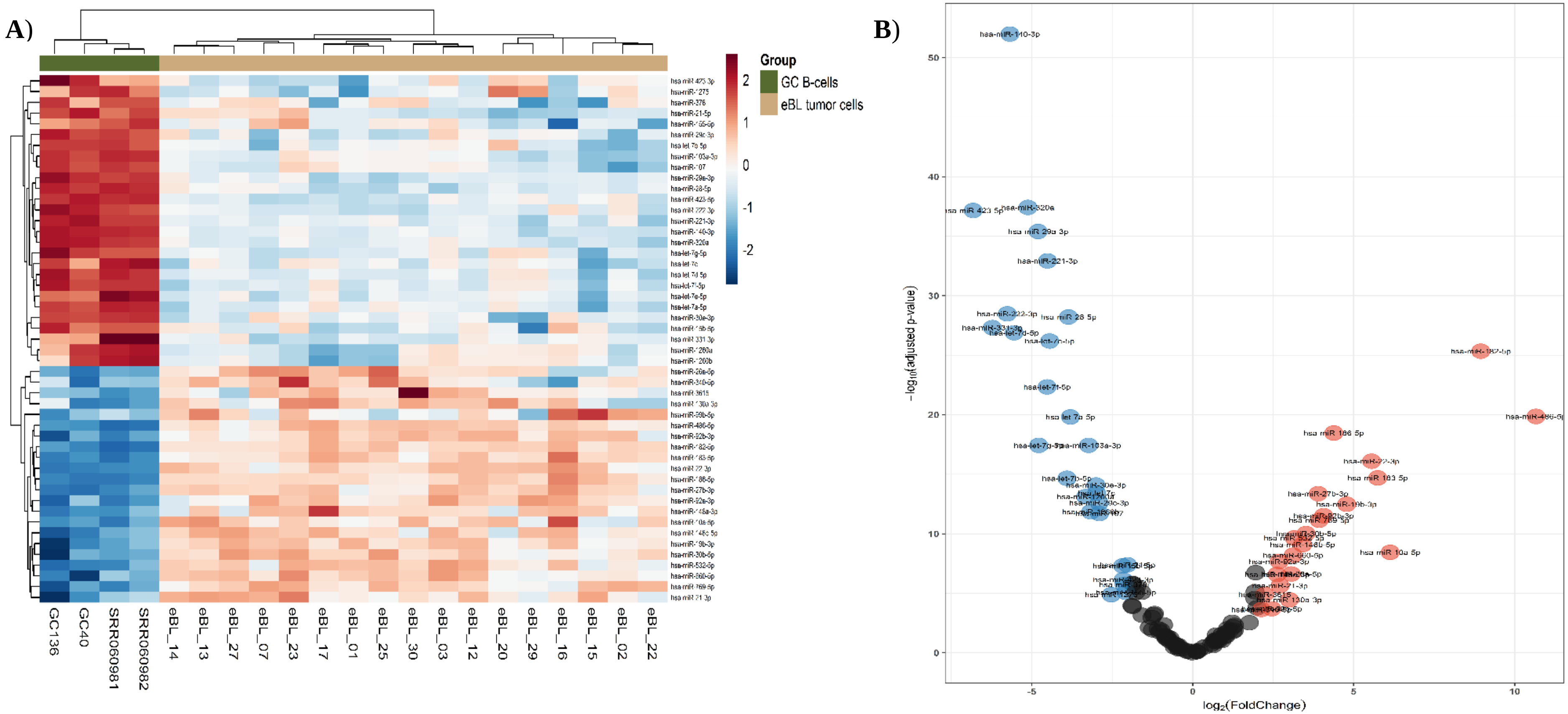


Figure 4.4: Differentially expressed (DE) miRNAs in eBL compared to GC B cells. **A**). Heatmap of DE miRNAs between eBL tumour cells and GC (Germinal center) B cells. The heatmap shows the hierarchical clustering based on the expression profiles of the 49 DE miRNAs with at least 2-fold difference in expression compared to their normal counterpart. **B**). Volcano plot representing the significance of the miRNAs (negative-log of the adjusted p-value) vs the fold change difference in eBL compared to GC B-cells. The red and blue coloured circles represent miRNAs which are DE with p<0.01 and FDR<0.01. The 27 down-regulated miRNAs in eBL are coloured blue (have a negative fold-change value) while the 22 up-regulated miRNAs in eBL are coloured red (positive fold-change value).

Table 4.3: Differentially expressed miRNAs between eBL tumour cells and GC B cells.

miRNA name	Log FC eBL versus GC B cells	BH adjusted *p*-value	FDR	Regulation eBL versus GC B cells
Upregulated miRs				
hsa-miR-486-5p	10.6609	1.38E-20	1.19E-19	Up
hsa-miR-182-5p	8.9478	4.55E-26	4.63E-25	Up
hsa-miR-10a-5p	6.1349	3.60E-09	1.19E-08	Up
hsa-miR-183-5p	5.7463	2.02E-15	1.19E-14	Up
hsa-miR-22-3p	5.5587	7.96E-17	4.95E-16	Up
hsa-miR-19b-3p	4.7807	3.32E-13	1.43E-12	Up
hsa-miR-186-5p	4.3815	3.43E-19	2.56E-18	Up
hsa-miR-92b-3p	4.0673	3.01E-12	1.16E-11	Up
hsa-miR-769-5p	3.9268	7.24E-12	2.70E-11	Up
hsa-miR-27b-3p	3.8964	4.25E-14	2.07E-13	Up
hsa-miR-30b-5p	3.5192	9.55E-11	3.45E-10	Up
hsa-miR-146b-5p	3.4102	8.07E-10	2.74E-09	Up
hsa-miR-532-5p	3.1493	2.26E-10	7.93E-10	Up
hsa-miR-660-5p	3.1183	6.35E-09	2.03E-08	Up
hsa-miR-26a-5p	3.0639	2.38E-07	6.65E-07	Up
hsa-miR-130a-3p	3.0018	3.50E-05	7.13E-05	Up
hsa-miR-21-3p	2.7137	2.29E-06	5.58E-06	Up
hsa-miR-92a-3p	2.6837	2.11E-08	6.58E-08	Up
hsa-miR-148a-3p	2.6221	2.61E-07	7.14E-07	Up
hsa-miR-99b-5p	2.4566	1.97E-04	3.74E-04	Up
hsa-miR-3615	2.2328	1.23E-05	2.65E-05	Up
hsa-miR-340-5p	2.1321	2.30E-04	4.30E-04	Up
Downregulated miRs				
hsa-miR-423-5p	-6.82082	6.92E-38	2.59E-36	Down
hsa-miR-331-3p	-6.22424	4.78E-28	6.70E-27	Down
hsa-miR-222-3p	-5.76142	3.31E-29	6.18E-28	Down
hsa-miR-140-3p	-5.67778	1.03E-52	1.16E-50	Down
hsa-let-7d-5p	-5.54828	1.29E-27	1.61E-26	Down
hsa-miR-320a	-5.12011	3.82E-38	2.14E-36	Down
hsa-miR-29a-3p	-4.80203	4.01E-36	1.12E-34	Down
hsa-let-7g-5p	-4.78494	3.71E-18	2.46E-17	Down
hsa-let-7f-5p	-4.52246	4.60E-23	4.29E-22	Down
hsa-miR-221-3p	-4.50804	1.18E-33	2.64E-32	Down
hsa-let-7e-5p	-4.44258	6.34E-27	7.10E-26	Down

Table 4.3 (*Continued*)

miRNA name	Log FC eBL versus GC B cells	BH adjusted *p*-value	FDR	Regulation eBL versus GC B cells
hsa-let-7b-5p	-3.90908	2.16E-15	1.21E-14	Down
hsa-miR-28-5p	-3.85689	6.16E-29	9.86E-28	Down
hsa-let-7a-5p	-3.79827	1.52E-20	1.22E-19	Down
hsa-miR-1260a	-3.32632	7.61E-14	3.55E-13	Down
hsa-miR-103a-3p	-3.23483	3.74E-18	2.46E-17	Down
hsa-miR-1260b	-3.16842	1.42E-12	5.91E-12	Down
hsa-miR-30e-3p	-3.00521	8.00E-15	4.27E-14	Down
hsa-let-7c	-2.98489	3.79E-14	1.93E-13	Down
hsa-miR-29c-3p	-2.90631	2.48E-13	1.11E-12	Down
hsa-miR-107	-2.89683	2.00E-12	8.00E-12	Down
hsa-miR-1275	-2.52971	1.30E-05	2.74E-05	Down
hsa-miR-15b-5p	-2.17524	4.93E-08	1.45E-07	Down
hsa-miR-423-3p	-2.15549	7.20E-07	1.92E-06	Down
hsa-miR-378i	-2.13256	1.86E-06	4.64E-06	Down
hsa-miR-155-5p	-2.07727	8.29E-06	1.82E-05	Down
hsa-miR-21-5p	-2.02446	4.01E-08	1.22E-07	Down

MiRNAs were considered significantly differentially expressed with a log Fold Change (logFC)>2, *p*-value<0.01 and FDR (False discovery rate) <0.01. Abbreviations: eBL, endemic Burkitt lymphoma; GC, germinal center; BH, Benjamini & Hochberg; FC, Fold Change; FDR, False discovery rate.

4.5. Prediction of MiRNAs Influencing Aberrant Gene Expression in eBL using Network Propagation-Based Method

The correlation between network effect of the miRNA perturbation and gene ranking was evaluated. This identified 12 eBL-related miRNA families significantly enriched (network perturbation effect score (NPES) >2, adjusted *p*-value<0.05 and FDR<0.1) in regulation of the aberrant gene expression profile in eBL (Table 4.4). The top ranked miRNAs (NPES>2.8, *p*=0.001 and FDR<0.02) included, miR-19b-3p (miR-19ab family) and miR-92a/b-3p (miR-

25/32/92abc/363/363-3p/367 family), were significantly upregulated in eBL tumour cells, and are known to target tumour suppressor genes such as *ATM* and *NLK*, which were downregulated in eBL tumours. Overall, the enriched miRNAs (Table 4.4) are more likely to cause the observed differential gene expression in eBL, to supplement the aberrant molecular mechanisms involved in lymphoma development.

Table 4.4: MiRNAs significantly enriched by the network propagation-based method in regulation of the aberrant gene expression profile in eBL

MiRFam	DE miRNAs	NPES.zscore	*p*-value	FDR
miR-19ab	hsa-miR-19b-3p	2.8993	0.001	0.01837
miR-25/32/92abc/363/363-3p/367	hsa-miR-92a-3p, hsa-miR-92b-3p	3.9415	0.001	0.01837
miR-29abcd	hsa-miR-29a-3p, hsa-miR-29c-3p	3.0281	0.002	0.02673
miR-140/140-5p/876-3p/1244	hsa-miR-140-3p	3.1259	0.003	0.0294
miR-183	hsa-miR-183-5p	2.6688	0.005	0.04594
miR-26ab/1297/4465	hsa-miR-26a-5p	2.6475	0.006	0.049
miR-221/222/222ab/1928	hsa-miR-221-3p, hsa-miR-222-3p	2.5557	0.007	0.049
miR-182	hsa-miR-182-5p	2.6168	0.008	0.05113
miR-27abc/27a-3p	hsa-miR-27b-3p	2.4344	0.011	0.06468
let-7/98/4458/4500	let-7a-5p, let-7b-5p, let-7c, let-7d-5p, let-7e-5p, let-7f-5p, let-7g-5p	2.2441	0.013	0.06825
miR-30abcdef/30abe-5p/384-5p	hsa-miR-30b-5p, hsa-miR-30e-5p, hsa-miR-30e-3p	2.3349	0.013	0.06825
miR-17/17-5p/20ab/20b-5p/93/106ab/427/518a-3p/519d	hsa-miR-20a-5p	2.1264	0.019	0.0735

MiRNAs and miRNA families were considered significant with a network perturbation effect score (NPES) >2, adjusted *p*-value<0.05 and False discovery rate (FDR) <0.1. Abbreviations: MiRFam, miRNA Family; DE, differentially expressed; NPES, network perturbation enrichment score; FDR, False discovery rate.

4.6. Validated Target Genes of the Differentially Expressed (DE) MiRNAs

A total of 1,427 genes were identified to be targeted by the DE miRNAs. Of the identified miRNA gene targets, 109 target genes were DE (upregulated or downregulated) in eBL tumours compared to GC B cells. Two hundred and twenty (220) miRNA-mRNA pairs expression identified to be inversely expressed. To test if the observed miRNA-mRNA pair's expression significantly varied and not due to chance, a permutation test repeated 10,000 times was performed. The 181 miRNA-mRNA pairs were then identified, to be of potential biological significance (p-value<0.05) (Table 4.5). Of the DE miRNAs, there was a marked number targeting critical tumour suppressors (*PTEN, AXIN1, ATM, NLK*) and critical proto-oncogenes and tumour promoting genes such as *MYC* (Appendix 9). Figure 4.5 illustrates miRNA-mRNA pairs that could influence *ATM* and *NLK* function in response to DNA damage to facilitate eBL lymphomagenesis.

Table 4.5: Differentially Expressed (DE) validated target genes of the DE miRNAs in eBL

DE miRNA	Validated DE Targets	Permutation test *p*-value
hsa-miR-16-5p	*CDK17, CYB561A3, DNAJC10, HIST1H1C, HIST1H2BC, HIST1H2BK, INO80D, ITPR1, KATNAL1, MACF1, ORC4, OXNAD1, PAG1, PHC3, REL, RELT, SLC12A2, SNX11, TEP1, TLE4*	**<0.00001**
hsa-miR-92a-3p	***ATM**, CMTM6, DCP2, DMXL1, HERC1, HIST1H1E, IRGQ, ITPR1, KIAA1109, LCOR, LHFPL2, LPIN1, LRRC37A2, MACF1, MAN2A1, MEF2C, MTMR14, RAB8B, REL, SMG1, STAT3, TACC1, ZNF277*	**<0.00001**
hsa-miR-20a-5p	*DNAJC10, DUSP2, FMNL3, HIST1H2BD, ITPKB, KATNAL1, NFAT5, PHF6, RAB11FIP1, RABGAP1L, SAMD8, SLC35F5, SOD2, STAT3, UGCG, VPS13C, ZBTB37, ZNF780A*	**0.0001**
hsa-miR-186-5p	*BLOC1S6, HIST1H1C, HIST1H1E, HIST1H2BK, ITPKB, ITSN2, KATNAL1, PHC3, SFT2D2, TLE4, VPS13B, VPS13C*	**0.0002**
hsa-miR-92b-3p	*ITPR1, KIAA1109, LCOR, LHFPL2, MAN2A1, **NLK**, RAB8B, REL, SMG1, TACC1, ZNF277*	**0.0005**
hsa-miR-19b-3p	***ATM**, BRWD1, BRWD1, DCP2, ITPR1, KATNAL1, MACF1, MALT1, RAB8B, SAMD8, SMG1*	**0.0012**
hsa-miR-320a	*ANP32B, C20orf27, MIF, **MYC**, NPM3, PCBP2, SRM, SRP68*	**0.0042**
hsa-let-7b-5p	*BMP7, GPATCH4, **MYC**, NT5DC2, PABPC1, PCBP2, TTLL12, UCK2*	**0.0045**
hsa-miR-30e-5p	*ASB3, **ATM**, LCOR, LHFPL2, NFAT5, OPHN1, TBC1D8B, VPS13B*	**0.0052**
hsa-miR-30b-5p	*ASB3, **ATM**, BLOC1S6, LCOR, LHFPL2, MYO1E, NFAT5, OPHN1*	**0.0057**

Table 4.5 (*Continued*)

DE miRNA	Validated DE Targets	Permutation test *p*-value
hsa-miR-423-5p	*C20orf27, FKBP4, **MYC**, PABPC1, PCBP1, SOX12, SRM*	**0.0077**
hsa-miR-27b-3p	*HIST1H1C, ITSN2, LCOR, LPIN1, **NLK**, UGCG*	**0.0155**
hsa-miR-340-5p	*DUSP2, INO80D, ITPR1, KIAA1109, SOD2, SULT1B1*	**0.017**
hsa-miR-10a-5p	*IRGQ, KLHL6, PIK3CG, SFT2D2, TSPAN33, WDR74*	**0.0187**
hsa-miR-24-3p	*BMP7, LDHA, **MYC**, TFDP2, TOP1, UCK2*	**0.0196**
hsa-miR-221-3p	*FUS, HNRNPD, NCL, NT5DC2, SRP68*	**0.0342**
hsa-miR-26a-5p	***ATM**, BMP2K, CFLAR, MAN2A1, PIKFYVE*	**0.0342**
hsa-miR-183-5p	*BRWD1, ITPKB, RPAP2, SMG1, TMED8*	**0.0369**
hsa-miR-222-3p	*HNRNPD, **MYC**, PCBP2, TRAP1, TXN*	**0.0371**

Validated target genes of the DE miRNAs in eBL compared to GC B cells. These validated target genes were also DE expressed in eBL tumour cells compared to GC B cells. To test if the observed miRNA-mRNA pairs were significant and not due to chance, a permutation test repeated 10000 times was performed. MicroRNA-mRNA pairs with a p-value <0.05 were considered to be of biological importance.

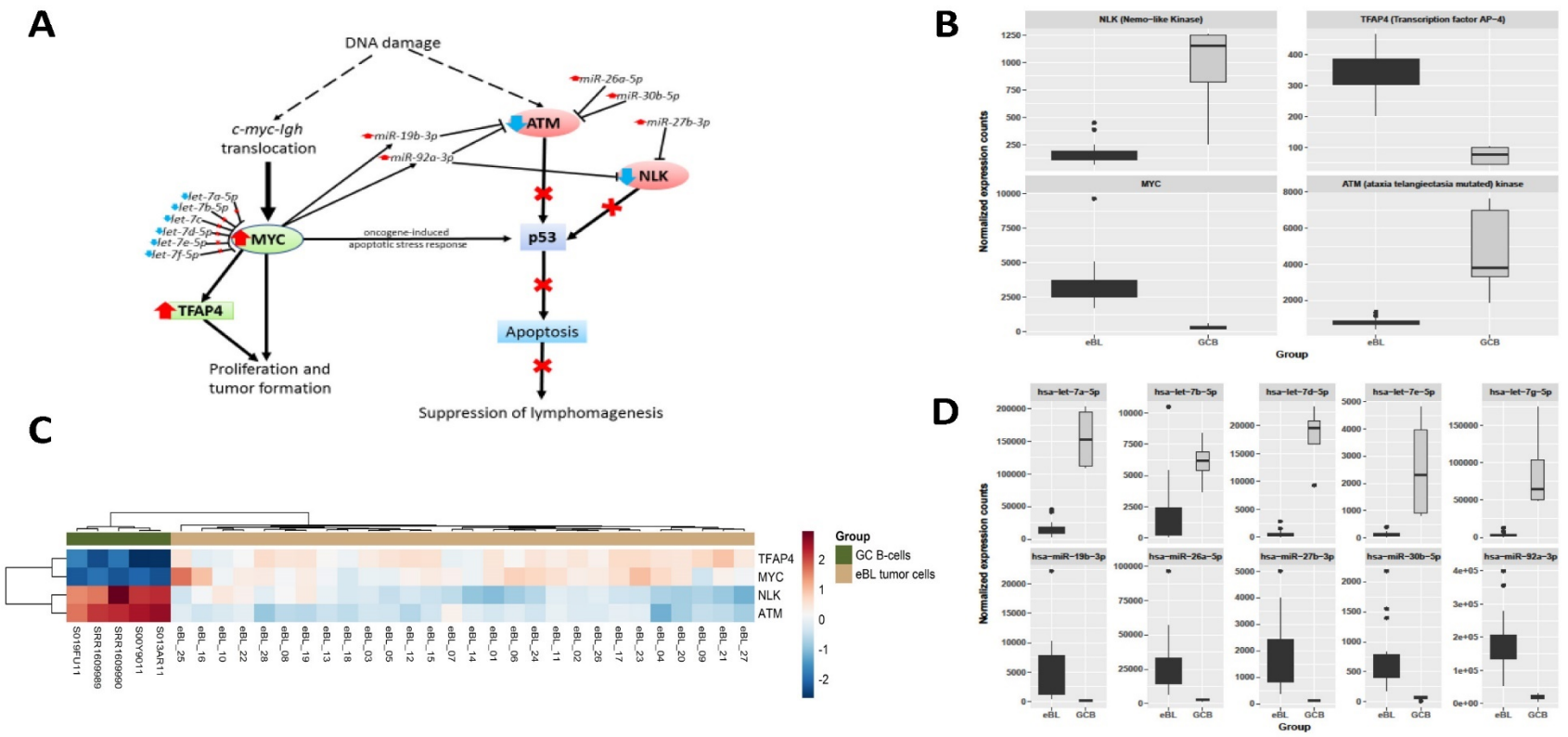


Figure 4.5: Aberrant transcriptome expression pivotal to eBL lymphomagenesis. **A**) Schematic illustration of the aberrant gene expression and miRNA mediated regulatory changes that would initiate lymphomagenesis as a result of DNA damage. Combined loss of p53 function due to small interfering RNA-mediated regulation of ATM and NLK together with upregulation of TFAP4, would facilitate survival of cells with the c-myc-Igh chromosomal translocation and MYC induced cell cycle progression initiating eBL tumor development. ATM checkpoint kinase, transduces genomic stress signals to halt cell cycle progression in response to DNA damage. It is critical in the regulation of apoptosis and lymphomagenesis in c-myc induced lymphomas. ATM is downregulated in eBL and it is targeted by 4 miRs that are upregulated in eBL. NLK is required for the upregulation of P53 expression in response to DNA damage. It interacts with P53 to enhance its stability and activity by abrogating MDM2 mediated degradation. NLK is downregulated in eBL tumor cells and also targeted by 2 miRs that are upregulated in eBL tumor cells. TFAP4/AP4 is a central mediator of cell cycle progression in response to c-MYC activation. **B**) RNA seq. expression counts of MYC, TFAP4, ATM and NLK in eBL tumor cells and GC B cells. **C**) Hierarchical clustering of eBL and GC B cells based on the expression profiles of MYC, TFAP4, ATM and NLK also revealed a clear separation of the two groups. **D**). miRNA seq. expression counts of hsa-miR-26a-5p, hsa-miR-27b-3p, hsa-miR-30b-5p, miR-17~92-cluster members (hsa-miR-19b-3p, and hsa-miR-92a-3p), and let-7-family miRs (hsa-let-7a-5p, hsa-let-7b-5p, hsa-let-7d-5p, hsa-let-7e-5p, and hsa-let-7g-5p) in eBL tumor cells and GC B cells.

4.6.1. Gene ontology and KEGG pathway enrichment

For the 49 DE miRNAs, miRWalk2.0 was used to identify their validated mRNA targets (Appendix 4). Gene ontogeny and pathway enrichment analysis of the validated targets, revealed the enrichment of Pathways in Cancer (p-value<0.01 and FDR<0.01) as the top enriched KEGG pathway, and other cancer associated KEGG pathways (such as hsa05205:Proteoglycans in cancer, hsa05203:Viral carcinogenesis, hsa05220:Chronic myeloid leukaemia) (Figure 4.6).

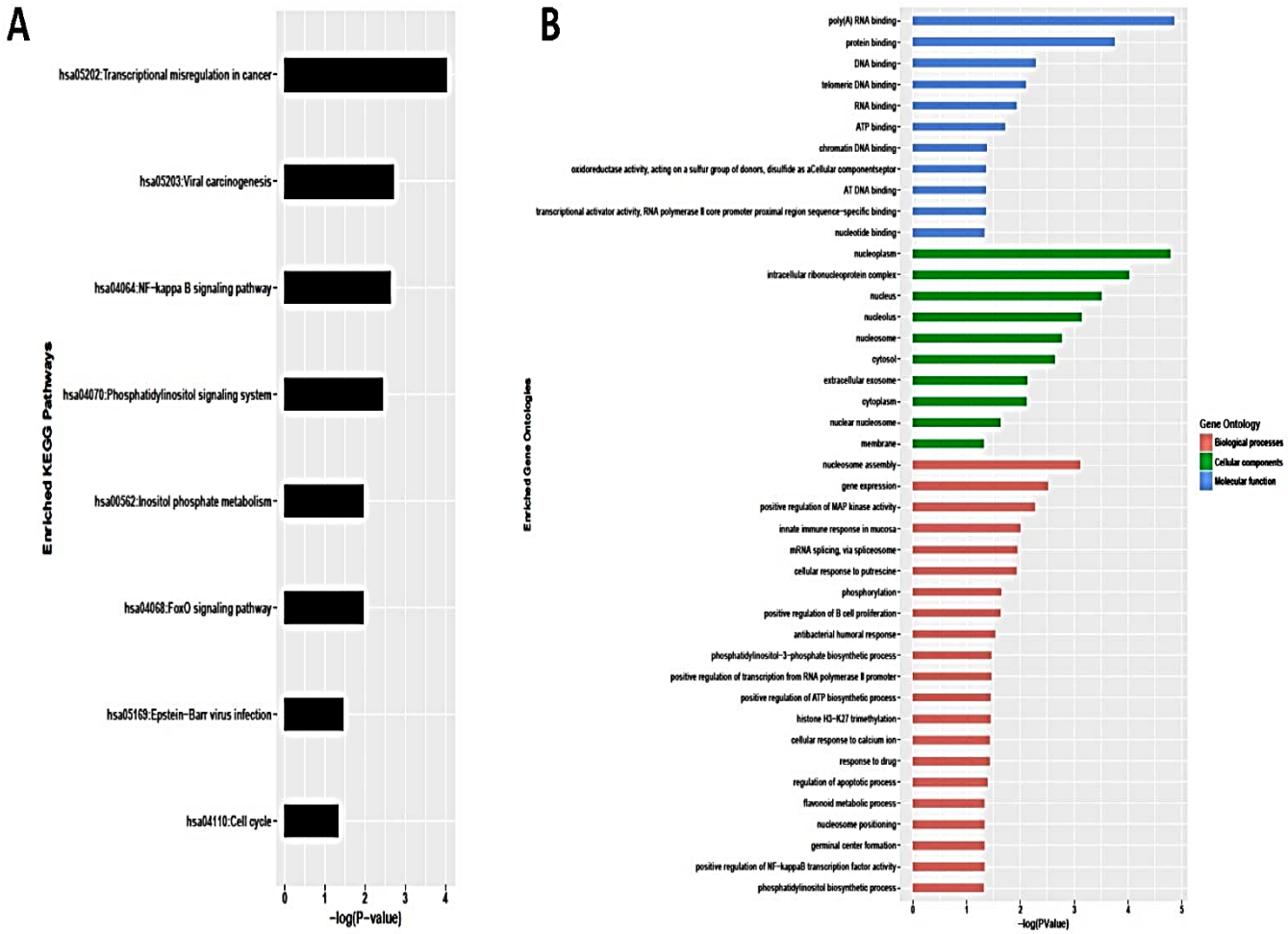


Figure 4.6: Gene Ontology and KEGG Pathway enrichment **A**) The significantly enriched signalling pathways of the validated target genes of the DE miRNAs that showed an inverse expression change. **B**) The significantly enriched gene ontologies (GO's) of the validated target genes of the DE miRNAs that showed an inverse expression change.

4.7. MiRNAs Show Similar Pattern of Expression without Tumour Subtypes by Hierarchical Clustering

To evaluate whether tumour miRNAs expression patterns were suggestive of the existence of eBL subtypes, unsupervised hierarchical clustering of the expression of all expressed miRNAs was performed. The overall correlations among eBL samples were extremely high ($r>0.9$, average) (Figure 4.7). The most disparate eBL samples (eBL_03 and eBL_22 compared to eBL_27 and eBL_16) still showed a high degree of correlation ($r=0.7$-0.8). There was no discernible clustering based on tumour site designation or in-hospital survival. Similarly, principal component analysis and multidimensional scaling showed no discernible separation based on tumour presentation site (Figure 4.7). This suggests that eBL tumours are relatively homogeneous without overt subtypes nor altered miRNA expression signatures based on tumour presentation site consistent with lack of subgroups examining mRNA expression (Kaymaz et al., 2017).

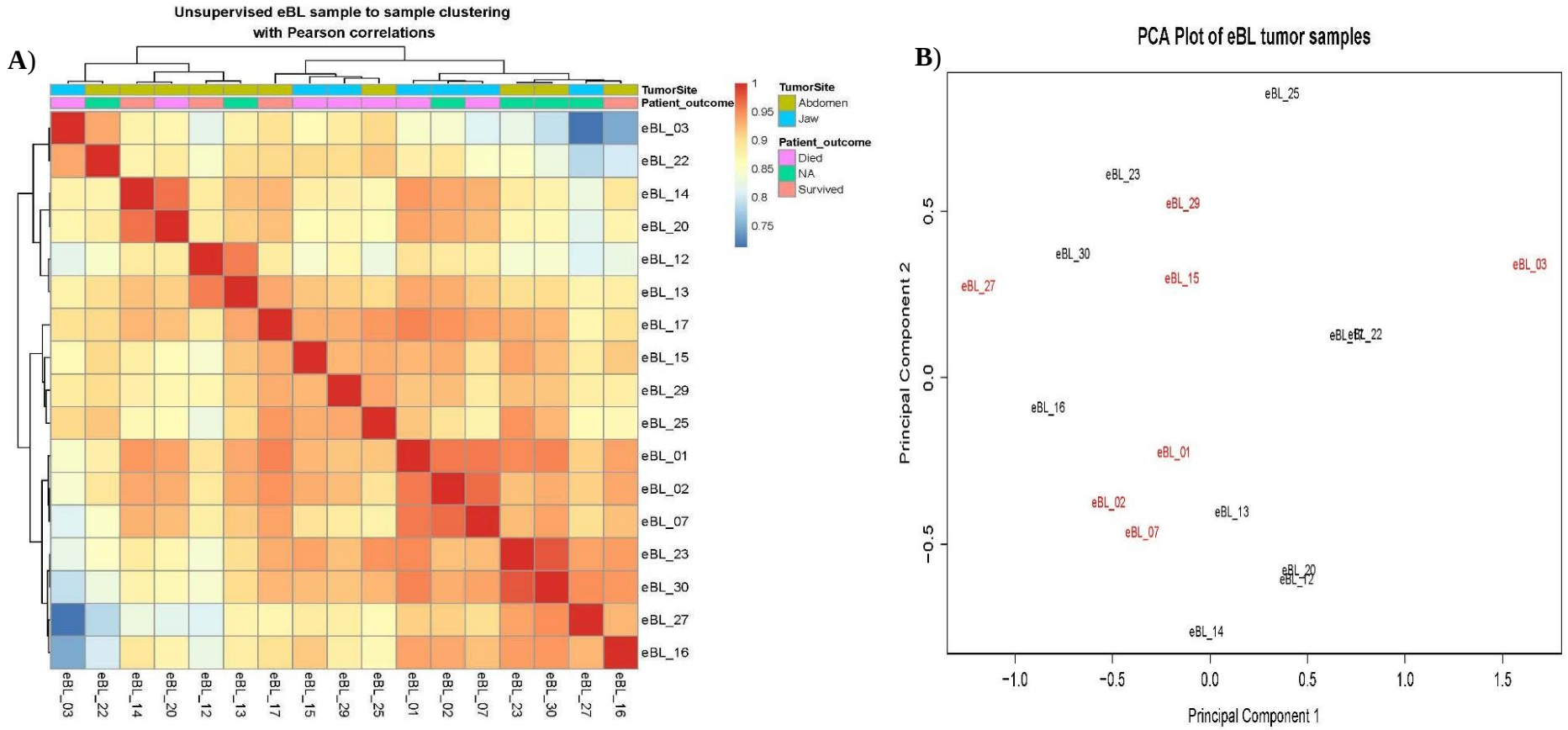


Figure 4.7: Exploratory analysis of miRNA expression in eBL tumours. **A**) Sample to sample hierarchal clustering of BL tumours based on miRNA expression profiles with highest correlation of variation (CV) values (calculated using regularized log transformed all miRNA expression). **B**) PCA plots of eBL tumours sites. The sample names were coloured per their presentation site (red represents the jaw tumours and black represented the abdominal tumours). The tumour samples don't separate/differentiate based on tumour site in the PCA analysis.

4.8. Differentially expressed (DE) miRNAs relative to Tumour Presentation Site and in-hospital Survival

Despite eBL tumours showing no distinctive clusters of overall miRNA expression patterns, a differential expression analysis of individual miRNAs was still performed relative to the tumour presentation site. The expression of human miRNAs was compared between jaw and abdominal eBL tumour samples, and one miRNA, hsa-miR-10a-5p (log2FC= -2.873, and *p*-value = 7.67e-06), was found to be significantly differentially expressed (Table 4.6). This miRNA showed lower expression in eBL jaw tumours compared to the abdominal tumours.

Examining differential expression with regards to in-hospital survival, the miRNA expression profiles of patients who died during the initial course of treatment was compared to patients who successfully completed chemotherapy and were discharged (i.e. in-hospital survivors). The analysis detected only one significant differentially expressed miRNA, hsa-miR-10a-5p (log2FC = -2.935, and *p*-value = 1.12e-05), which was significantly lower in non-survivor's relative to survivors.

Table 4.6. (a) DE miRNAs in eBL jaw tumours compared to abdominal tumours

miRNA	Accession	*p*-value	False Discovery Rate (FDR)	Fold change (FC) log2FC	Jaw tumour vs. abdominal tumour
hsa-miR-10a-5p	MIMAT0000253	7.67e-06	0.001857	-2.873	down-regulated

MiRNAs were considered to be differentially expressed between the jaw and abdominal tumours if the log Fold Change (logFC)>2, *p*-value<0.01 and False discovery rate (FDR)<0.01.

Table 4.6. (b) DE miRNAs in eBL non-survivors compared to survivors

miRNA	Accession	*p*-value	FDR	log2FC	Non-survivors vs. survivors
hsa-miR-10a-5p	MIMAT0000253	1.12e-05	0.0013	-2.935	down-regulated

MiRNAs were considered to be differentially expressed between the eBL non-survivors and survivors if the log Fold Change (logFC)>2, *p*-value<0.01 and False discovery rate (FDR) <0.01.

4.9. miR-10a-5p Validation by RT-qPCR

To assess the reproducibility and validity of the miR-10a expression levels by deep sequencing its levels were measured by RT-qPCR. The Wilcoxon rank test confirmed the difference in expression level for miR-10a-5p between the jaw and abdominal tumours, and also between eBL survivors and non-survivors (*p*-values 6.29e-05 and 0.0087 respectively). The boxplot in Figure 4.8, shows the median distribution levels of the log2 normalized relative quantity of miR-10a-5p in the eBL tumours. The RT-qPCR results showed good correlation with sequencing (r=0.77) based on a correlation test between RT-qPCR miR-10a expression levels and miR-10a expression levels from the sequencing experiment. Overall, this further validated the sequencing results (Figure 4.9).

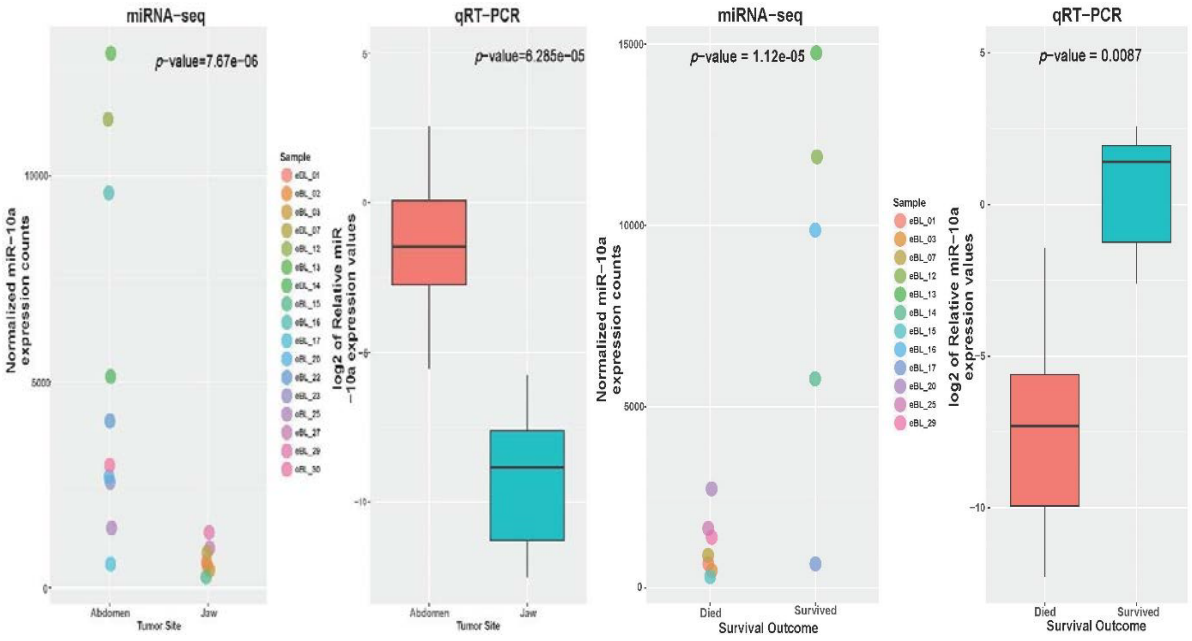


Figure 4.8: RT-qPCR validation of miR-10a-5p expression difference in jaw versus abdominal tumours and in eBL survivors versus non-survivors. The jitter plot shows the normalized expression values from the miRNA sequencing experiment. The box plot shows the log2 median expression levels of miR-10a-5p, confirmed as downregulated in Jaw tumours and in eBL non-survivor patients, estimated in terms of normalized fluorescence intensity.

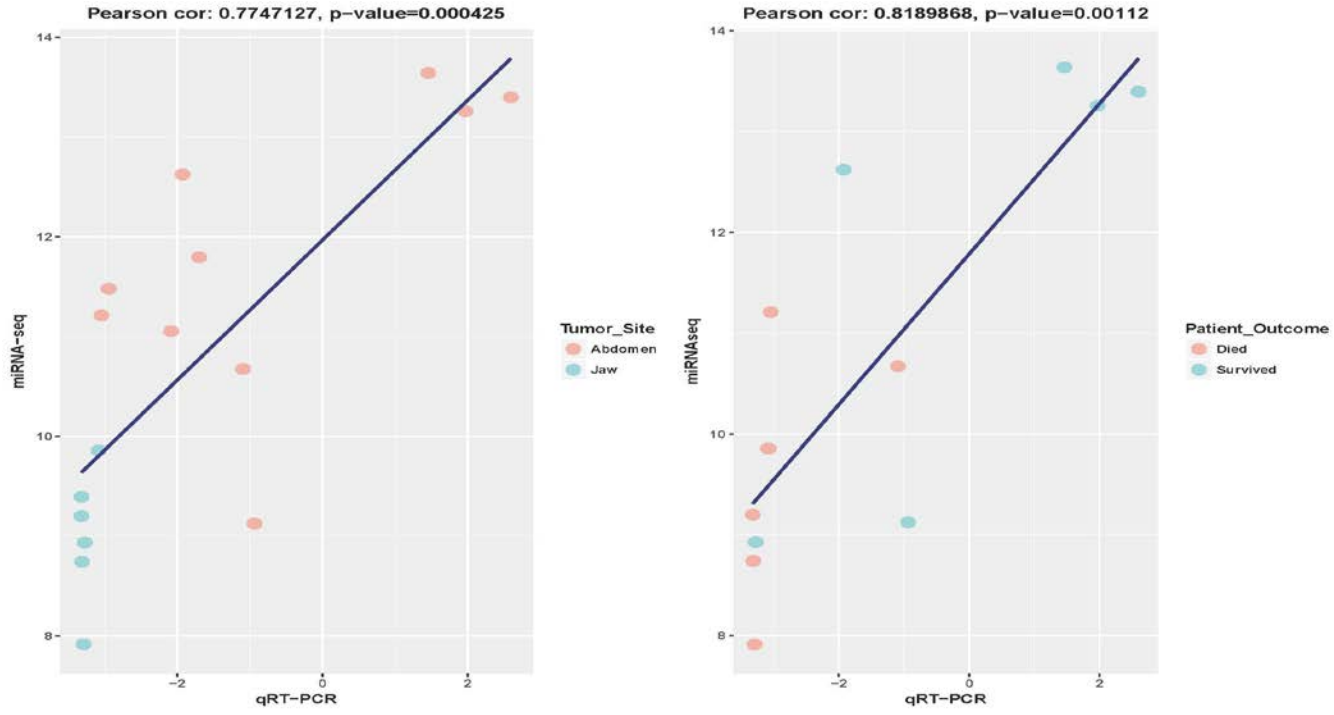


Figure 4.9: Pearson correlation scatter plots showing a positive correlation between RT-qPCR miR-10a-5p expression levels and miRNA-seq miR-10a-5p expression levels.

4.10. Correlation analysis of miR-10a-5p expression and its Validated Target Genes

A pairwise correlation coefficient analysis was performed between miR-10a-5p expression levels and the expression levels of its 251 validated target genes with a median expression count >10 counts per million (cpm) from the RNAseq data set for the eBL patients. Of the validated target gene set, there were multiple genes with some degree of correlation that would explain miR-10a association with patient survival outcome by impairing apoptotic death. *CD59, API5, MDM4,* and *YY1* (validated targets of miR-10a-5p) showed an inverse relationship with miR-10a levels (Appendix 4). Increased levels of each of these genes could impair apoptotic death and influence eBL patient prognosis. While only CD59 showed a significant (p-value < 0.05) inverse correlation with miR-10a levels (Figures 4.10), overall, this suggests that miR-10a could be modulating genes such as CD59, MDM2 and API5 in a way that could promote tumour cell survival among eBL patients with the Jaw tumours.

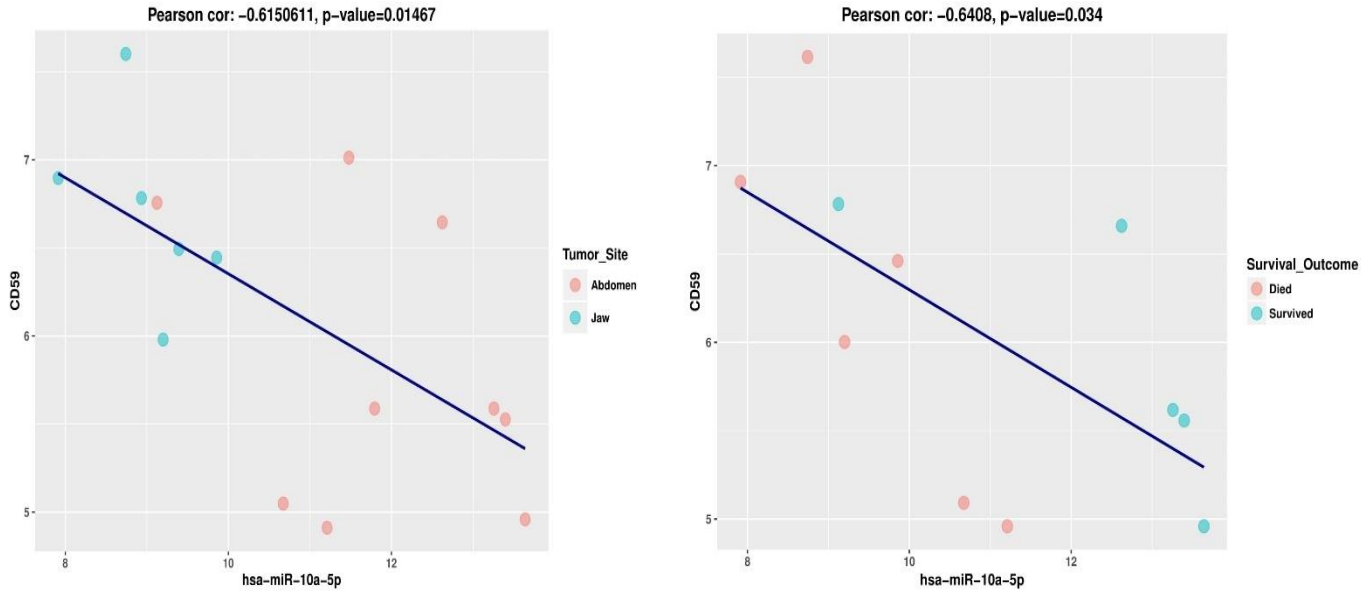


Figure 4.10: Pearson correlation scatter plots showing a negative correlation between hsa-miR-10a-5p and CD59 expression levels in eBL patients based on tumour site and patient survival outcome. The horizontal and vertical axis represents the log2-expression values of the miRNA and mRNA, respectively.

CHAPTER FIVE

DISCUSSION

5.1. Introduction

MicroRNAs regulate the expression of approximately 30% of all genes in the human genome (Nelson and Weiss, 2008). In normal cells, unlike in cancer cells, the interaction of miRNAs and target mRNAs is tightly regulated (Liu *et al.*, 2007; Sotiropoulou *et al.*, 2009). A growing body of evidence suggests that miRNAs are aberrantly expressed in many human cancers and that they play significant roles in the initiation and development of these cancers (Filipowicz *et al.*, 2008). Therefore, to better understand the specific molecular characteristics of eBL, differentially expressed miRNAs and mRNAs in eBL tumour cells compared to germinal center (GC) B cells based on high-throughput, sequencing datasets were identified. This is the first attempt to simultaneously analyse mRNA and miRNA expression profiles in eBL tumour cells compared to their normal counterpart. Two hundred and eleven (211) mRNAs and 49 miRNAs were identified to be differentially expressed (DE) with fold changes >2 and p-values<0.01 in eBL tumour cells compared to GC B cells. Of these, 181 miRNA-mRNA pairs, which appeared to be of genuine biological significance and not by chance, showed an inverse direction of expression change in eBL. With the aim of understanding the transcriptome expression changes pivotal to eBL development, the aberrant expression of genes (such as *ATM* and *NLK*) and miRNAs (such as let-7 family members and miR-17~92 cluster members) that could endorse eBL lymphoma development and sustain survival of tumour cells in the presence of myc translocation were identified. Hsa-miR-10a-5p was also identified to be downregulated in both eBL jaw tumours and among eBL non-survivors, which could favour survival of the tumour cells based on its validated targets.

5.2. MicroRNA Mediated Gene Regulation is involved in eBL Lymphomagenesis

Members of the miR-17~92 cluster gene are the first miRNAs to be implicated in cancer development (Lawrie, 2007; Mendell, 2008). This miRNA gene cluster encodes six distinct miRNAs (miR-17, miR-18a, miR-19a, miR-19b, miR-20a and miR-92) that share the same seed sequence (Mendell, 2008). These miRNAs are frequently over-expressed in other cancers (including multiple B and T cell lymphoid malignancies as well as colorectal cancer, breast cancer, pancreatic cancer, ovarian cancer, lung cancer, and hepatocellular carcinoma) (Lawrie, 2007; Mogilyansky and Rigoutsos, 2013) and in BL compared with other non-Hodgkin lymphomas (NHLs) (Di Lisio *et al.*, 2012; Jima *et al.*, 2010; Mendell, 2008; Musilova and Mraz, 2015). *MYC* overexpression, because of its translocation to the immunoglobulin locus in BL, enhances the expression of miR-17~92 cluster miRNAs by binding directly to its genomic locus (O'Donnell *et al.*, 2005; Schmitz *et al.*, 2014) to accelerate carcinogenesis. MiR-17~92 overexpression has been observed previously in sBL tumours (Schmitz *et al*., 2012). This is consistent with the expression patterns in the current eBL study. The observed elevated expression of MYC, miR-19b-3p, miR-92a-3p and miR-92b-3p in eBL tumour cells compared to GC B-cells, confirms that elevated expression of the miR-17~92 cluster miRNAs is a critical feature facilitating eBL lymphomagenesis.

Human let-7 family members were also observed to be abnormally expressed in eBL compared to GC. These miRNAs act as tumour suppressors, regulators of differentiation and apoptosis, and have been observed to be downregulated in most cancers (Wang *et al.*, 2012-5). Let-7 regulates many transcription factors and oncogenes that play important roles in cell cycle regulation, proliferation and apoptosis. These miRNAs have been shown to repress *MYC*, hence controlling proliferation and tumour development (Sampson *et al.*, 2007). Seven let-7

family members (let-7a, let-7b, let-7c, let-7d, let-7e, let-7f, and let-7g) were observed to be downregulated in eBL tumour cells compared to GC B cells, consistent with their functional role in the genesis and maintenance (Sampson *et al.*, 2007) of eBL tumour cells in the presence of *MYC* deregulation.

5.3. Aberrant Gene Expression in eBL Lymphomagenesis

Constitutive *MYC* activity is necessary for all forms of BL (Schmitz *et al.*, 2014; Spender and Inman, 2014). However, overexpression of this proto-oncogene also induces apoptotic stress responses which are overcome during lymphomagenesis (Spender and Inman, 2014). Following *MYC* translocation and deregulation in eBL, apart from genetic alterations and mutations that would facilitate escape from myc-mediated apoptosis (Kuttler *et al.*, 2001; Spender and Inman, 2014; Vousden *et al.*, 1993), aberrantly expressed miRNAs may also enable a cell to tolerate such oncogene-induced apoptotic stress. *MYC* is known to activate the p53 tumour suppressor pathway to initiate the apoptotic stress response, however tumour cell survival prevails. The observed downregulation of *ATM* gene and *NLK* in eBL, possibly due to small-interfering RNA mediated regulation, would impair P53 induced by *MYC*, initiating lymphoma occurrence.

Aberrancies of the *ATM* gene are among the most commonly occurring somatic mutations in cancer and its loss of function or downregulation has been observed in other cancers (Cremona and Behrens, 2014; Helgason *et al.*, 2015). This checkpoint kinase, transduces genomic stress signals to halt cell cycle progression in response to DNA damage. It is critical in the regulation of the P53 apoptotic pathway and lymphomagenesis in *c-myc* induced lymphomas (Ramiro *et al.*, 2006; Shreeram *et al.*, 2006). *ATM* could be a pivotal tumour suppressor in response to the translocation occurrence characteristic of eBL tumour cells. It is possible that during tumorigenesis a number of GC B cells have low *ATM* levels

due to small interfering RNA-mediated regulation, as a result of irregular expression of miR-27b-3p, miR-26a-5p, miR-30b-5p and myc-dependent activation of miR-17~92 cluster miRNAs. In turn, the levels fall below the threshold to halt cell cycle progression in response to DNA damage and maintain P53 activation. Downregulation of *ATM* gene in eBL tumour cells, implies a defective response to DNA damage and P53 activation to suppress tumour development initiated by the *t(8:14)* chromosomal translocation. Upregulation of miRNAs (miR-27b-3p, miR-26a-5p, miR-30b-5p, miR-19b-3p, and miR-92b-3p) in eBL targeting *ATM* suggests abnormal miRNA mediate regulation of this gene which would lead to *ATM* loss.

The observed *NLK* downregulation in eBL tumour cells could also be critical to aid in tumour cell escape from death initiated by DNA damage (that results in the *c-myc-Igh* chromosomal translocation) and oncogene-induced apoptotic stress. *NLK* has been shown to be an important P53 regulator in response to DNA damage and is critical to P53 stability and function (Kurahashi *et al.*, 2005; Zhang *et al*., 2014). Based on these results, low *NLK* levels in eBL tumours, probably due to miRNA (upregulated miR-92a-3p and miR-27b-3p expression) mediated regulation, could reduce the stability and activation of P53 in suppressing eBL lymphomagenesis. *ATM* and *NLK* genes were also observed to be significantly down-regulated in established BL cell lines (Namalwa, Raji Ramos, Daudi, Thomas, BL41, BL2, BL30, BL70, CA46, and Gumbus) compared to GC B cells (Appendix 8), supporting the notion that loss of these genes are critical to eBL lymphomagenesis and tumour cell survival.

This study also revealed the upregulation of *TFAP4/AP4* (transcription factor AP-4) in eBL tumour cells. Interestingly, *AP4* is a c-MYC inducible transcription factor that has been shown to be elevated in many types of tumours (Hu *et al.*, 2013; Jackstadt *et al.*, 2013; Jackstadt and Hermeking, 2014; Jung *et al.*, 2008; Ramiro *et al.*, 2006) and it also harbours an oncogenic potential (Jackstadt *et al.*, 2013). Therefore, it is likely that the upregulation of AP4 expression

mediates cell cycle progression, probably in response to *MYC* activation, coupled with P53 loss of function due to miRNA regulation of *ATM* and *NLK*, and could facilitate the survival of cells harbouring the *c-myc-Igh* translocation initiating eBL tumour development.

Epstein Barr virus (EBV) is highly associated with eBL diagnosed in Africa and thus the observed KEGG pathway enrichment of infection and viral carcinogenesis pathways was not unexpected. Presence of EBV encoded proteins such as EBNA-1, EBNA-3C and LMP-1 promote genomic instability (Gruhne *et al.*, 2009), which could contribute to eBL pathogenesis. Genomic instability, which would be initiated by EBV latent proteins coupled with loss of *ATM* as observed and impaired P53 activity (as a result of the observed *NLK* loss) due to miRNA repression, would favour the proliferation and survival of eBL tumour cells. EBV miRNA (ebv-miR-bart5), which is expressed in eBL tumour cells (Oduor *et al.*, 2017), and LMP-1 gene can also inhibit *ATM* expression (Gruhne *et al.*, 2009). However, the observed down-regulation of *ATM* in EBV negative BL cell lines (BL2, BL30, BL41, BL70, CA46, Gumbus, and Ramos) (Appendix 8) supports the notion that, irrespective of EBV's association with eBL, other genetic aberrations could lead to *ATM* loss in eBL. Genomic aberrations such as abnormal upregulation of host miRNAs (miR-27b-3p, miR-26a-5p, miR-30b-5p, miR-19b-3p, and miR-92b-3p) targeting *ATM* would favour proliferation, tumour cell survival and occurrences of mutations that would favour oncogenesis.

5.4. MiR-10a-5p as a Predictive Prognostic Biomarker for Endemic BL

Previous studies, have shown that tumours with different anatomical sites have unique patterns of miRNA expression (Petillo *et al.*, 2009). The observed downregulation of miR-10a-5p in jaw tumours and among eBL non-survivors, given that some of its validated targets (such as *CD59*, *MDM2*, *MDM4* and *API5*) potentiate tumour cell survival, resistance to apoptosis and chemotherapy, suggests a tumour suppressive role of this miRNA in eBL tumour

progression. MiR-10a is known to target CD59 (a potent inhibitor of the complement membrane attack complex) (Helwak *et al.*, 2013), therefore its loss could lead to hyperactivation of its target (CD59) which has been shown to be effective at protecting cancer cells from antibody (i.e. rituximab) mediated complement-dependent cytotoxicity ((You *et al.*, 2011). Considering the aggressive nature of the defects leading to eBL development, downregulated miR-10a-5p could also promote the hyperactivation of *API5* (apoptosis inhibitor 5), an apoptosis inhibitory protein targeted by miR-10a-5p (Karginov and Hannon, 2013), which renders tumour cells resistant to T cell initiated apoptosis (Faye and Poyet, 2010; Noh *et al.*, 2014). The observed inverse correlation between miR-10a-5p expression levels and *CD59* and *API5* expression levels, points to a potentially functional miRNA-mRNA interaction which could promote tumour cell survival, chemo-resistance and poor outcomes in eBL patients.

In support of miR-10a-5p's role in cancer progression, downregulation of miR-10a-5p has been implicated in other cancer studies (Jansson and Lund, 2012). For example, downregulation of this miR has been reported in chronic myeloid leukemia (CML) and associated with increased cell growth of the CML cells (Agirre *et al.*, 2008). Experimental re-expression of this miR in CML cells, was shown to decrease cell growth, thus supporting the functional role of hsa-miR-10a in CML disease progression (Agirre *et al.*, 2008).

Despite these results describing mRNA and miRNA expression and miRNA-mRNA interactions influencing eBL lymphomagenesis, having protein expression data for the aberrantly expressed mRNAs would strengthen these findings and demonstrate that the aberrant miRNA expression observed would also influence protein expression initiating lymphomagenesis. However, since next generation RNA sequencing was the only platform used, the study was limited to only RNA expression changes in eBL tumour environment.

Endemic BL tumour cells appear relatively homogeneous without overt subtypes nor altered miRNA expression signatures based on tumour presentation site (Figure 4.7) which is consistent with lack of subgroups examining mRNA expression (Kaymaz *et al*., 2017). However, this observation may be due the sequencing technology used in this study which gets mRNA and miRNA expression changes from the bulk tumour microenvironment which could mask minimal gene and miRNA aberrant events from single cells that could influence tumour cell survival and escape from immune surveillance leading to relapses. Despite this study identifying miR-10a to be associated with poor eBL patient outcome, the molecular factors underlying the development of relapse among a small percentage of eBL cases still remain unknown. This study was limited at studying the miRNA and gene expression changes in eBL relapse cases due to lack of tumour samples from the primary tumour and from the relapse tumour among the relapse patients.

CHAPTER SIX

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

6.1. Summary

In summary, this study represents the first integrative analysis of miRNA and mRNA expression in eBL tumours. A cell with a functional DNA damage response (DDR) mechanism via ATM and P53 induction in case of *MYC* overexpression, would impair eBL lymphomagenesis. The combined loss of p53 function in response to DNA damage and oncogene (MYC) induced stress may be due to miRNA-mediated regulation of *ATM* and *NLK*, impairing DDR signal and oncogene induced stress. Together with the observed upregulation of *TFAP4*, facilitates the survival of eBL tumour cells with the IGH/MYC chromosomal translocation and promotes MYC induced cell cycle progression initiating eBL lymphomagenesis. Given the validated target genes of miR-10a-5p, its altered expression pattern between different tumour sites (jaw versus abdominal tumours) in eBL and in eBL patients with different in-hospital survival outcomes would enhance tumour cell survival and thus, render the tumour less sensitive to chemotherapy.

6.2. Conclusion

1. Loss of *ATM* (a vital initiator of DDR) coupled with *NLK* (a p53 activator and stabilizer) loss, could impair the initiation of an apoptotic response in the presence *MYC* overexpression to initiate eBL lymphomagenesis.
2. Upregulation of miR-17~92 cluster members, miR-27b-3p, miR-26a-5p, and miR-30b-5p targeting *ATM* in eBL tumour cells, explains the some of the aberrant epigenetic changes that would be involved in eBL lymphomagenesis. Downregulation of tumour suppressor miRNAs such as let-7 family members targeting *MYC*, also partially explains some of the molecular events involved in eBL lymphomagenesis.

3. The observed downregulation of miR-10a-5p in eBL jaw tumour biopsies and among the non-survivors implies a survival advantage of tumour cells presenting in the jaw region which would lead to a poor prognosis.

6.3. Recommendations from this Study

1. Develop miRNA antagomirs to counteract *ATM* and *NLK* post-transcriptional regulation thereby initiating apoptosis in response to the *MYC* translocation.

2. *MYC* being the pivotal oncogene in BL oncogenesis and lymphomagenesis, the development of miRNA mimics to target *MYC*, would reduce *MYC* expression levels in eBL tumour cells. Thus, reducing proliferation rate and tumour cell survival.

3. Downregulation or loss of miR-10a-5p can be used as a biomarker for poor prognosis among eBL patients presenting with the jaw tumour.

6.4. Recommendations for Future Studies

1. Exploration of expression levels of *ATM* and *NLK* proteins in eBL tumour cells compared to GC B cells.

2. Studies could explore the expression profile of other short and long RNA regulatory molecules and other epigenetic alterations that may be involved in eBL oncogenesis.

3. Exploration of single cell transcriptome profiling of eBL tumour tissue isolates to allow the characterization of heterogeneous eBL tumour cells that may underlie poor prognosis and relapse occurrences.

REFERENCES

Agirre, X., Jiménez-Velasco, A., San José-Enériz, E., Garate, L., Bandrés, E., Cordeu, L., Aparicio, O., Saez, B., Navarro, G., Vilas-Zornoza, A., Pérez-Roger, I., García-Foncillas, J., Torres, A., Heiniger, A., Calasanz, M.J., Fortes, P., Román-Gómez, J., Prósper, F., 2008. Down-regulation of hsa-miR-10a in chronic myeloid leukemia CD34+ cells increases USF2-mediated cell growth. Mol. Cancer Res. 6, 1830–1840.

Andrews, S., 2014. FastQC. A quality control tool for high throughput sequence data. Babraham Bioinformatics.

Basso, K., Dalla-Favera, R., 2012. Roles of BCL6 in normal and transformed germinal center B cells. Immunol. Rev. 247, 172–183.

Behm-Ansmant, I., Rehwinkel, J., Izaurralde, E., 2006. MicroRNAs silence gene expression by repressing protein expression and/or by promoting mRNA decay. Cold Spring Harb. Symp. Quant. Biol. 71, 523–530.

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Series B Stat. Methodol. 57, 289–300.

Bornkamm, G.W., 2009. Epstein-Barr virus and its role in the pathogenesis of Burkitt's lymphoma: an unresolved issue. Semin. Cancer Biol. 19, 351–365.

Bovolenta, L.A., Acencio, M.L., Lemke, N., 2012. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. BMC Genomics 13, 405.

Brady, G., MacArthur, G.J., Farrell, P.J., 2007. Epstein-Barr virus and Burkitt lymphoma. J. Clin. Pathol. 60, 1397–1402.

Braun, J., Misiak, D., Busch, B., Krohn, K., Hüttelmaier, S., 2014. Rapid identification of regulatory microRNAs by miTRAP (miRNA trapping by RNA in vitro affinity purification). Nucleic Acids Res. 42, e66.

Buckle, G., Maranda, L., Skiles, J., Ong'echa, J.M., Foley, J., Epstein, M., Vik, T.A., Schroeder, A., Lemberger, J., Rosmarin, A., Remick, S.C., Bailey, J.A., Vulule, J., Otieno, J.A., Moormann, A.M., 2016. Factors influencing survival among Kenyan children diagnosed with endemic Burkitt lymphoma between 2003 and 2011: A historical cohort study. Int. J. Cancer 139, 1231–1240.

Burkitt, D., 1958. A sarcoma involving the jaws in African children. Br. J. Surg. 46, 218–223.

Burkitt, D., O'Conor, G.T., 1961. Malignant lymphoma in African children. I. A clinical syndrome. Cancer 14, 258–269.

Cairo, S., Wang, Y., de Reyniès, A., Duroure, K., Dahan, J., Redon, M.-J., Fabre, M., McClelland, M., Wang, X.W., Croce, C.M., Buendia, M.-A., 2010. Stem cell-like micro-RNA signature driven by Myc in aggressive liver cancer. Proc. Natl. Acad. Sci. U. S. A. 107, 20471–20476.

Calin, G.A., Croce, C.M., 2006. MicroRNA signatures in human cancers. Nat. Rev. Cancer 6, 857–866.

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F., Croce, C.M., 2002. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc. Natl. Acad. Sci. U. S. A. 99, 15524–15529.

Chang, T.-C., Yu, D., Lee, Y.-S., Wentzel, E.A., Arking, D.E., West, K.M., Dang, C.V., Thomas-Tikhonenko, A., Mendell, J.T., 2008-1. Widespread microRNA repression by Myc contributes to tumorigenesis. Nat. Genet. 40, 43–50.

Chêne, A., 2009. Impact of Malaria on B-Cell Homeostasis and Epstein Barr Virus Reactivation- Endemic Burkitt's Lymphoma Pathogenesis. Karolinska Institutet.

Cho, W.C.S., 2012. Great potential of miRNAs as predictive and prognostic markers for cancer. Expert Rev. Mol. Diagn. 12, 315–318.

Cimmino, A., Calin, G.A., Fabbri, M., Iorio, M.V., Ferracin, M., Shimizu, M., Wojcik, S.E., Aqeilan, R.I., Zupo, S., Dono, M., Rassenti, L., Alder, H., Volinia, S., Liu, C.-G., Kipps, T.J., Negrini, M., Croce, C.M., 2005. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc. Natl. Acad. Sci. U. S. A. 102, 13944–13949.

Cremona, C.A., Behrens, A., 2014. ATM signalling and cancer. Oncogene 33, 3351–3360.

Croce, C.M., 10/2009. Causes and consequences of microRNA dysregulation in cancer. Nat. Rev. Genet. 10, 704–714.

Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R.C., Croce, C.M., 1982. Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proc. Natl. Acad. Sci. U. S. A. 79, 7824–7827.

Dancey, J.E., Bedard, P.L., Onetto, N., Hudson, T.J., 2012. The genetic basis for cancer treatment decisions. Cell 148, 409–420.

Dave, S.S., Fu, K., Wright, G.W., Lam, L.T., Kluin, P., Boerma, E.-J., Greiner, T.C., Weisenburger, D.D., Rosenwald, A., Ott, G., Müller-Hermelink, H.-K., Gascoyne, R.D., Delabie, J., Rimsza, L.M., Braziel, R.M., Grogan, T.M., Campo, E., Jaffe, E.S., Dave, B.J., Sanger, W., Bast, M., Vose, J.M., Armitage, J.O., Connors, J.M., Smeland, E.B., Kvaloy, S., Holte, H., Fisher, R.I., Miller, T.P., Montserrat, E., Wilson, W.H., Bahl, M., Zhao, H., Yang, L., Powell, J., Simon, R., Chan, W.C., Staudt, L.M., 2006. Molecular diagnosis of Burkitt's lymphoma. N. Engl. J. Med. 354, 2431–2442.

Deng, S., Calin, G.A., Croce, C.M., Coukos, G., Zhang, L., 2008. Mechanisms of microRNA deregulation in human cancer. Cell Cycle 7, 2643–2646.

Dennis, G., Jr, Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4, P3.

Di Lisio, L., Sanchez-Beato, M., Gomez-Lopez, G., Rodriguez, M.E., Montes-Moreno, S., Mollejo, M., Menarguez, J., Martinez, M.A., Alves, F.J., Pisano, D.G., Piris, M.A., Martinez, N., 2012. MicroRNA signatures in B-cell lymphomas. Blood Cancer J. 2, e57.

Dweep, H., Gretz, N., 2015. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. Nat. Methods 12, 697.

Dweep, H., Sticht, C., Pandey, P., Gretz, N., 2011. miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. J. Biomed.

Inform. 44, 839–847.
Esteller, M., 2011. Non-coding RNAs in human disease. Nat. Rev. Genet. 12, 861–874.
Farazi, T.A., Horlings, H.M., Ten Hoeve, J.J., Mihailovic, A., Halfwerk, H., Morozov, P., Brown, M., Hafner, M., Reyal, F., van Kouwenhove, M., Kreike, B., Sie, D., Hovestadt, V., Wessels, L.F.A., van de Vijver, M.J., Tuschl, T., 2011. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. Cancer Res. 71, 4443–4453.
Faye, A., Poyet, J.-L., 2010. Targeting AAC-11 in cancer therapy. Expert Opin. Ther. Targets 14, 57–65.
Filipowicz, W., Bhattacharyya, S.N., Sonenberg, N., 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9, 102–114.
Flatmark, K., Høye, E., Fromm, B., 2016. microRNAs as cancer biomarkers. Scand. J. Clin. Lab. Invest. Suppl. 245, S80–3.
Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., Rajewsky, N., 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 40, 37–52.
Gabay, M., Li, Y., Felsher, D.W., 2014. MYC activation is a hallmark of cancer initiation and maintenance. Cold Spring Harb. Perspect. Med. 4. https://doi.org/10.1101/cshperspect.a014241
Gavioli, R., Frisan, T., Vertuani, S., Bornkamm, G.W., Masucci, M.G., 2001. c-myc overexpression activates alternative pathways for intracellular proteolysis in lymphoma cells. Nat. Cell Biol. 3, 283–288.
Griffiths-Jones, S., Grocock, R.J., Dongen, S. van, Bateman, A., Enright, A.J., 2006. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 34, D140–D144.
Grömminger, S., Mautner, J., Bornkamm, G.W., 2012. Burkitt lymphoma: the role of Epstein-Barr virus revisited. Br. J. Haematol. 156, 719–729.
Gruhne, B., Sompallae, R., Masucci, M.G., 2009. Three Epstein-Barr virus latency proteins independently promote genomic instability by inducing DNA damage, inhibiting DNA repair and inactivating cell cycle checkpoints. Oncogene 28, 3997–4008.
Ham, B., Min, D., Sohn, K., 2013. A generalized random walk with restart and its application in depth up-sampling and interactive segmentation. IEEE Trans. Image Process. 22, 2574–2588.
Hamzeiy, H., Allmer, J., Yousef, M., 2014. Computational methods for microRNA target prediction. Methods Mol. Biol. 1107, 207–221.
Hart, S.N., Therneau, T.M., Zhang, Y., Poland, G.A., Kocher, J.-P., 2013. Calculating sample size estimates for RNA sequencing data. J. Comput. Biol. 20, 970–978.
Hayes, J., Peruzzi, P.P., Lawler, S., 2014. MicroRNAs in cancer: biomarkers, functions and therapy. Trends Mol. Med. 20, 460–469.
Hecht, J.L., Aster, J.C., 2000. Molecular biology of Burkitt's lymphoma. J. Clin. Oncol. 18, 3707–3721.
Helgason, H., Rafnar, T., Olafsdottir, H.S., Jonasson, J.G., Sigurdsson, A., Stacey, S.N., Jonasdottir, A., Tryggvadottir, L., Alexiusdottir, K., Haraldsson, A., le Roux, L.,

Gudmundsson, J., Johannsdottir, H., Oddsson, A., Gylfason, A., Magnusson, O.T., Masson, G., Jonsson, T., Skuladottir, H., Gudbjartsson, D.F., Thorsteinsdottir, U., Sulem, P., Stefansson, K., 2015. Loss-of-function variants in ATM confer risk of gastric cancer. Nat. Genet. 47, 906–910.

Helwak, A., Kudla, G., Dudnakova, T., Tollervey, D., 2013. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153, 654–665.

Hesseling, P.B., Molyneux, E., Kamiza, S., Broadhead, R., 2008. Rescue chemotherapy for patients with resistant or relapsed endemic Burkitt's lymphoma. Trans. R. Soc. Trop. Med. Hyg. 102, 602–607.

Hesseling, P., Molyneux, E., Kamiza, S., Israels, T., Broadhead, R., 2009. Endemic Burkitt lymphoma: a 28-day treatment schedule with cyclophosphamide and intrathecal methotrexate. Ann. Trop. Paediatr. 29, 29–34.

Hong, B., van den Heuvel, A.P.J., Prabhu, V.V., Zhang, S., El-Deiry, W.S., 2014. Targeting tumor suppressor p53 for cancer therapy: strategies, challenges and opportunities. Curr. Drug Targets 15, 80–89.

Hong, L., Yang, Z., Ma, J., Fan, D., 2013. Function of miRNA in controlling drug resistance of human cancers. Curr. Drug Targets 14, 1118–1127.

Hu, B.-S., Zhao, G., Yu, H.-F., Chen, K., Dong, J.-H., Tan, J.-W., 2013. High expression of AP-4 predicts poor prognosis for hepatocellular carcinoma after curative hepatectomy. Tumour Biol. 34, 271–276.

Jackstadt, R., Hermeking, H., 2014. AP4 is required for mitogen- and c-MYC-induced cell cycle progression. Oncotarget 5, 7316–7327.

Jackstadt, R., Röh, S., Neumann, J., Jung, P., Hoffmann, R., Horst, D., Berens, C., Bornkamm, G.W., Kirchner, T., Menssen, A., Hermeking, H., 2013. AP4 is a mediator of epithelial-mesenchymal transition and metastasis in colorectal cancer. J. Exp. Med. 210, 1331–1350.

Jansson, M.D., Lund, A.H., 2012. MicroRNA and cancer. Mol. Oncol. 6, 590–610.

Jewer, M., Findlay, S.D., Postovit, L.-M., 2012. Post-transcriptional regulation in cancer progression : Microenvironmental control of alternative splicing and translation. J. Cell Commun. Signal. 6, 233–248.

Jima, D.D., Zhang, J., Jacobs, C., Richards, K.L., Dunphy, C.H., Choi, W.W.L., Au, W.Y., Srivastava, G., Czader, M.B., Rizzieri, D.A., Lagoo, A.S., Lugar, P.L., Mann, K.P., Flowers, C.R., Bernal-Mizrachi, L., Naresh, K.N., Evens, A.M., Gordon, L.I., Luftig, M., Friedman, D.R., Weinberg, J.B., Thompson, M.A., Gill, J.I., Liu, Q., How, T., Grubor, V., Gao, Y., Patel, A., Wu, H., Zhu, J., Blobe, G.C., Lipsky, P.E., Chadburn, A., Dave, S.S., Hematologic Malignancies Research Consortium, 2010. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. Blood 116, e118–27.

Jin, H.Y., Oda, H., Lai, M., Skalsky, R.L., Bethel, K., Shepherd, J., Kang, S.G., Liu, W.-H., Sabouri-Ghomi, M., Cullen, B.R., Rajewsky, K., Xiao, C., 2013. MicroRNA-17~92 plays a causative role in lymphomagenesis by coordinating multiple oncogenic pathways. EMBO J. 32, 2377–2391.

Jung, P., Menssen, A., Mayr, D., Hermeking, H., 2008. AP4 encodes a c-MYC-inducible

repressor of p21. Proc. Natl. Acad. Sci. U. S. A. 105, 15046–15051.
Kafuko, G.W., Burkitt, D.P., 1970. Burkitt's lymphoma and malaria. International Journal of Cancer 6, 1–9.
Karginov, F.V., Hannon, G.J., 2013. Remodeling of Ago2-mRNA interactions upon cellular stress reflects miRNA complementarity and correlates with altered translation rates. Genes Dev. 27, 1624–1632.
Kaymaz, Y., Oduor, C.I., Yu, H., Otieno, J.A., Ong'echa, J.M., Moormann, A.M., Bailey, J.A., 2017. Comprehensive Transcriptome and Mutational Profiling of Endemic Burkitt Lymphoma Reveals EBV Type-specific Differences. Mol. Cancer Res. https://doi.org/10.1158/1541-7786.MCR-16-0305
Kim, H.-S., Han, H.D., Armaiz-Pena, G.N., Stone, R.L., Nam, E.J., Lee, J.-W., Shahzad, M.M.K., Nick, A.M., Lee, S.J., Roh, J.-W., Nishimura, M., Mangala, L.S., Bottsford-Miller, J., Gallick, G.E., Lopez-Berestein, G., Sood, A.K., 2011. Functional roles of Src and Fgr in ovarian carcinoma. Clin. Cancer Res. 17, 1713–1721.
Klein, U., Klein, G., Ehlin-Henriksson, B., Rajewsky, K., Küppers, R., 1995. Burkitt's lymphoma is a malignancy of mature B cells expressing somatically mutated V region genes. Mol. Med. 1, 495–505.
Koues, O.I., Kowalewski, R.A., Chang, L.-W., Pyfrom, S.C., Schmidt, J.A., Luo, H., Sandoval, L.E., Hughes, T.B., Bednarski, J.J., Cashen, A.F., Payton, J.E., Oltz, E.M., 2015. Enhancer sequence variants and transcription-factor deregulation synergize to construct pathogenic regulatory circuits in B-cell lymphoma. Immunity 42, 186–198.
Kozomara, A., Griffiths-Jones, S., 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42, D68–73.
Kretzmer, H., Bernhart, S.H., Wang, W., Haake, A., Weniger, M.A., Bergmann, A.K., Betts, M.J., Carrillo-de-Santa-Pau, E., Doose, G., Gutwein, J., Richter, J., Hovestadt, V., Huang, B., Rico, D., Jühling, F., Kolarova, J., Lu, Q., Otto, C., Wagener, R., Arnolds, J., Burkhardt, B., Claviez, A., Drexler, H.G., Eberth, S., Eils, R., Flicek, P., Haas, S., Hummel, M., Karsch, D., Kerstens, H.H.D., Klapper, W., Kreuz, M., Lawerenz, C., Lenze, D., Loeffler, M., López, C., MacLeod, R.A.F., Martens, J.H.A., Kulis, M., Martín-Subero, J.I., Möller, P., Nagel, I., Picelli, S., Vater, I., Rohde, M., Rosenstiel, P., Rosolowski, M., Russell, R.B., Schilhabel, M., Schlesner, M., Stadler, P.F., Szczepanowski, M., Trümper, L., Stunnenberg, H.G., ICGC MMML-Seq project, BLUEPRINT project, Küppers, R., Ammerpohl, O., Lichter, P., Siebert, R., Hoffmann, S., Radlwimmer, B., 2015. DNA methylome analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. Nat. Genet. 47, 1316–1325.
Kuchen, S., Resch, W., Yamane, A., Kuo, N., Li, Z., Chakraborty, T., Wei, L., Laurence, A., Yasuda, T., Peng, S., Hu-Li, J., Lu, K., Dubois, W., Kitamura, Y., Charles, N., Sun, H.-W., Muljo, S., Schwartzberg, P.L., Paul, W.E., O'Shea, J., Rajewsky, K., Casellas, R., 2010. Regulation of microRNA expression and abundance during lymphopoiesis. Immunity 32, 828–839.
Kurahashi, T., Nomura, T., Kanei-Ishii, C., Shinkai, Y., Ishii, S., 2005. The Wnt-NLK signaling pathway inhibits A-Myb activity by inhibiting the association with coactivator

CBP and methylating histone H3. Mol. Biol. Cell 16, 4705–4713.
Kuttler, F., Amé, P., Clark, H., Haughey, C., Mougin, C., Cahn, J.Y., Dang, C.V., Raffeld, M., Fest, T., 2001. c-myc box II mutations in Burkitt's lymphoma-derived alleles reduce cell-transformation activity and lower response to broad apoptotic stimuli. Oncogene 20, 6084–6094.
Lan, F.-F., Wang, H., Chen, Y.-C., Chan, C.-Y., Ng, S.S., Li, K., Xie, D., He, M.-L., Lin, M.C., Kung, H.-F., 2011. Hsa-let-7g inhibits proliferation of hepatocellular carcinoma cells by downregulation of c-Myc and upregulation of p16(INK4A). Int. J. Cancer 128, 319–331.
Langmead, B., 2010. Aligning short sequencing reads with Bowtie. Curr. Protoc. Bioinformatics Chapter 11, Unit 11.7.
Lan, H., Lu, H., Wang, X., Jin, H., 2015. MicroRNAs as potential biomarkers in cancer: opportunities and challenges. Biomed Res. Int. 2015, 125094.
Lawrie, C.H., 2007. MicroRNA expression in lymphoma. Expert Opin. Biol. Ther. 7, 1363–1374.
Leek, J.T., 2014. svaseq: removing batch effects and other unwanted noise from sequencing data. Nucleic Acids Res. 42. https://doi.org/10.1093/nar/gku864
Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843–854.
Leskelä, S., Leandro-García, L.J., Mendiola, M., Barriuso, J., Inglada-Pérez, L., Muñoz, I., Martínez-Delgado, B., Redondo, A., de Santiago, J., Robledo, M., Hardisson, D., Rodríguez-Antona, C., 2011. The miR-200 family controls beta-tubulin III expression and is associated with paclitaxel-based treatment response and progression-free survival in ovarian cancer patients. Endocr. Relat. Cancer 18, 85–95.
Leucci, E., Cocco, M., Onnis, A., De Falco, G., van Cleef, P., Bellan, C., van Rijk, A., Nyagol, J., Byakika, B., Lazzi, S., Tosi, P., van Krieken, H., Leoncini, L., 2008. MYC translocation-negative classical Burkitt lymphoma cases: an alternative pathogenetic mechanism involving miRNA deregulation. J. Pathol. 216, 440–450.
Levine, A.M., 2002. Challenges in the management of Burkitt's lymphoma. Clin. Lymphoma 3 Suppl 1, S19–25.
Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323.
Liu, W., Mao, S.-Y., Zhu, W.-Y., 2007. Impact of tiny miRNAs on cancers. World J. Gastroenterol. 13, 497–502.
Liu, W., Ma, R., Yuan, Y., 2017. Post-transcriptional Regulation of Genes Related to Biological Behaviors of Gastric Cancer by Long Noncoding RNAs and MicroRNAs. J. Cancer 8, 4141–4154.
Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408.
Li, Y., Vecchiarelli-Federico, L.M., Li, Y.-J., Egan, S.E., Spaner, D., Hough, M.R., Ben-David, Y., 2012. The miR-17-92 cluster expands multipotent hematopoietic progenitors whereas imbalanced expression of its individual oncogenic miRNAs promotes leukemia in mice. Blood 119, 4486–4498.

Love, C., Sun, Z., Jima, D., Li, G., Zhang, J., Miles, R., Richards, K.L., Dunphy, C.H., Choi, W.W.L., Srivastava, G., Lugar, P.L., Rizzieri, D.A., Lagoo, A.S., Bernal-Mizrachi, L., Mann, K.P., Flowers, C.R., Naresh, K.N., Evens, A.M., Chadburn, A., Gordon, L.I., Czader, M.B., Gill, J.I., Hsi, E.D., Greenough, A., Moffitt, A.B., McKinney, M., Banerjee, A., Grubor, V., Levy, S., Dunson, D.B., Dave, S.S., 2012. The genetic landscape of mutations in Burkitt lymphoma. Nat. Genet. 44, 1321–1325.

Lu, J., Clark, A.G., 2012. Impact of microRNA regulation on variation in human gene expression. Genome Res. 22, 1243–1254.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R., Golub, T.R., 2005. MicroRNA expression profiles classify human cancers. Nature 435, 834–838.

Macha, M.A., Seshacharyulu, P., Krishn, S.R., Pai, P., Rachagani, S., Jain, M., Batra, S.K., 2014. MicroRNAs (miRNAs) as biomarker(s) for prognosis and diagnosis of gastrointestinal (GI) cancers. Curr. Pharm. Des. 20, 5287–5297.

Magrath, I., 2012. Epidemiology: clues to the pathogenesis of Burkitt lymphoma. Br. J. Haematol. 156, 744–756.

Magrath, I.T., 1991. African Burkitt's Lymphoma: History, Biology, Clinical Features, and Treatment. J. Pediatr. Hematol. Oncol. 13, 222.

Magrath, L., 2010. Denis Burkitt and the African lymphoma. Ecancermedicalscience. https://doi.org/10.3332/ecancer.2009.159

Mallmann, M.R., Staratschek-Jox, A., Rudlowski, C., Braun, M., Gaarz, A., Wolfgarten, M., Kuhn, W., Schultze, J.L., 2010-9. Prediction and prognosis: impact of gene expression profiling in personalized treatment of breast cancer patients. EPMA J. 1, 421–437.

Marabita, F., de Candia, P., Torri, A., Tegnér, J., Abrignani, S., Rossi, R.L., 2016. Normalization of circulating microRNA expression data obtained by quantitative real-time RT-PCR. Brief. Bioinform. 17, 204–212.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12.

Mendell, J.T., 2008. miRiad roles for the miR-17-92 cluster in development and disease. Cell 133, 217–222.

Mishra, P.J., 2014. MicroRNAs as promising biomarkers in cancer diagnostics. Biomark Res 2, 19.

Mogilyansky, E., Rigoutsos, I., 2013. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. Cell Death Differ. 20, 1603–1614.

Moormann, A.M., Chelimo, K., Sumba, P.O., Tisch, D.J., Rochford, R.A., Kazura, J.W., 2007. Exposure to holoendemic malaria results in suppression of Epstein-Barr virus-specific T cell immunosurveillance in Kenyan children. J. Infect. Dis. 195, 799–808.

Musilova, K., Mraz, M., 2015. MicroRNAs in B-cell lymphomas: how a complex biology gets more complex. Leukemia 29, 1004–1017.

Mwanda, O.W., 2004. Burkitt's lymphoma: the initial defining characteristics. East Afr. Med. J. S63–67.

Nelson, K.M., Weiss, G.J., 2008. MicroRNAs and cancer: past, present, and potential future. Mol. Cancer Ther. 7, 3655–3660.

Niedobitek, G., Meru, N., Delecluse, H.J., 2001. Epstein-Barr virus infection and human malignancies. Int. J. Exp. Pathol. 82, 149–170.

Nikolaou, M., Pavlopoulou, A., Georgakilas, A.G., Kyrodimos, E., 2018. The challenge of drug resistance in cancer treatment: a current overview. Clin. Exp. Metastasis. https://doi.org/10.1007/s10585-018-9903-0

Noh, K.H., Kim, S.-H., Kim, J.H., Song, K.-H., Lee, Y.-H., Kang, T.H., Han, H.D., Sood, A.K., Ng, J., Kim, K., Sonn, C.H., Kumar, V., Yee, C., Lee, K.-M., Kim, T.W., 2014. API5 confers tumoral immune escape through FGF2-dependent cell survival pathway. Cancer Res. 74, 3556–3566.

Novocraft [WWW Document], n.d. URL http://www.novocraft.com/ (accessed 11.3.16).

O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V., Mendell, J.T., 2005. c-Myc-regulated microRNAs modulate E2F1 expression. Nature 435, 839–843.

Oduor, C.I., Movassagh, M., Kaymaz, Y., Chelimo, K., Otieno, J., Ong'echa, J.M., Moormann, A.M., Bailey, J.A., 2017. Human and Epstein-Barr Virus miRNA Profiling as Predictive Biomarkers for Endemic Burkitt Lymphoma. Front. Microbiol. 8, 501.

Ogwang, M.D., Bhatia, K., Biggar, R.J., Mbulaiteye, S.M., 2008. Incidence and geographic distribution of endemic Burkitt lymphoma in northern Uganda revisited. Int. J. Cancer 123, 2658–2663.

Ohuchida, K., Mizumoto, K., Kayashima, T., Fujita, H., Moriyama, T., Ohtsuka, T., Ueda, J., Nagai, E., Hashizume, M., Tanaka, M., 2011. MicroRNA expression as a predictive marker for gemcitabine response after surgical resection of pancreatic cancer. Ann. Surg. Oncol. 18, 2381–2387.

Olive, V., Jiang, I., He, L., 2010. mir-17-92, a cluster of miRNAs in the midst of the cancer network. Int. J. Biochem. Cell Biol. 42, 1348–1354.

Olivier, M., Hollstein, M., Hainaut, P., 2010. TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harb. Perspect. Biol. 2, a001008.

Onnis, A., De Falco, G., Antonicelli, G., Onorati, M., Bellan, C., Sherman, O., Sayed, S., Leoncini, L., 2010. Alteration of microRNAs regulated by c-Myc in Burkitt lymphoma. PLoS One 5. https://doi.org/10.1371/journal.pone.0012960

Onnis, A., Navari, M., Antonicelli, G., Morettini, F., Mannucci, S., De Falco, G., Vigorito, E., Leoncini, L., 2012. Epstein-Barr nuclear antigen 1 induces expression of the cellular microRNA hsa-miR-127 and impairing B-cell differentiation in EBV-infected memory B cells. New insights into the pathogenesis of Burkitt lymphoma. Blood Cancer J. 2, e84.

Orem, J., Mbidde, E.K., Lambert, B., de Sanjose, S., Weiderpass, E., 2007. Burkitt's lymphoma in Africa, a review of the epidemiology and etiology. Afr. Health Sci. 7, 166–175.

Ott, G., Rosenwald, A., Campo, E., 2013. Understanding MYC-driven aggressive B-cell lymphomas: pathogenesis and classification. Blood 122, 3884–3891.

Palmero, E.I., de Campos, S.G.P., Campos, M., de Souza, N.C.N., Guerreiro, I.D.C., Carvalho, A.L., Marques, M.M.C., 2011. Mechanisms and role of microRNA

deregulation in cancer onset and progression. Genet. Mol. Biol. 34, 363–370.
Pasqualucci, L., Guglielmino, R., Houldsworth, J., Mohr, J., Aoufouchi, S., Polakiewicz, R., Chaganti, R.S.K., Dalla-Favera, R., 2004. Expression of the AID protein in normal and neoplastic B cells. Blood 104, 3318–3325.
Pasquinelli, A.E., 2012. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nat. Rev. Genet. https://doi.org/10.1038/nrg3162
Pedraza-Fariña, L.G., 2006. Mechanisms of oncogenic cooperation in cancer initiation and metastasis. Yale J. Biol. Med. 79, 95–103.
Petillo, D., Kort, E.J., Anema, J., Furge, K.A., Yang, X.J., Teh, B.T., 2009. MicroRNA profiling of human kidney cancer subtypes. Int. J. Oncol. 35, 109–114.
Poli, V., Fagnocchi, L., Fasciani, A., Cherubini, A., Mazzoleni, S., Ferrillo, S., Miluzio, A., Gaudioso, G., Vaira, V., Turdo, A., Giaggianesi, M., Chinnici, A., Lipari, E., Bicciato, S., Bosari, S., Todaro, M., Zippo, A., 2018. MYC-driven epigenetic reprogramming favors the onset of tumorigenesis by inducing a stem cell-like state. Nat. Commun. 9, 1024.
Postovit, L.-M., Margaryan, N.V., Seftor, E.A., Hendrix, M.J.C., 2008. Role of nodal signaling and the microenvironment underlying melanoma plasticity. Pigment Cell Melanoma Res. 21, 348–357.
Pritchard, C.C., Cheng, H.H., Tewari, M., 2012. MicroRNA profiling: approaches and considerations. Nat. Rev. Genet. 13, 358–369.
Rainey, J.J., Omenah, D., Sumba, P.O., Moormann, A.M., Rochford, R., Wilson, M.L., 2007. Spatial clustering of endemic Burkitt's lymphoma in high-risk regions of Kenya. Int. J. Cancer 120, 121–127.
Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.-T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W., Nussenzweig, A., Nussenzweig, M.C., 2006. Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature 440, 105–109.
Ramiro, A.R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Honjo, T., Nussenzweig, A., Nussenzweig, M.C., 2004. AID is required for c-myc/IgH chromosome translocations in vivo. Cell 118, 431–438.
Ramsingh, G., Koboldt, D.C., Trissal, M., Chiappinelli, K.B., Wylie, T., Koul, S., Chang, L.-W., Nagarajan, R., Fehniger, T.A., Goodfellow, P., Magrini, V., Wilson, R.K., Ding, L., Ley, T.J., Mardis, E.R., Link, D.C., 2010. Complete characterization of the microRNAome in a patient with acute myeloid leukemia. Blood 116, 5316–5326.
Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., Ruvkun, G., 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 403, 901–906.
Ricarte-Filho, J.C.M., Fuziwara, C.S., Yamashita, A.S., Rezende, E., da-Silva, M.J., Kimura, E.T., 2009. Effects of let-7 microRNA on Cell Growth and Differentiation of Papillary Thyroid Cancer. Transl. Oncol. 2, 236–241.
Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Ruggero, D., 2013. Translational control in cancer etiology. Cold Spring Harb. Perspect. Biol. 5. https://doi.org/10.1101/cshperspect.a012336

Rui, L., Schmitz, R., Ceribelli, M., Staudt, L.M., 2011. Malignant pirates of the immune system. Nat. Immunol. 12, 933–940.

Saito, T., Sætrom, P., 2010. MicroRNAs--targeting and target prediction. N. Biotechnol. 27, 243–249.

Sampson, V.B., Rong, N.H., Han, J., Yang, Q., Aris, V., Soteropoulos, P., Petrelli, N.J., Dunn, S.P., Krueger, L.J., 2007. MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. Cancer Res. 67, 9762–9770.

Sassen, S., Miska, E.A., Caldas, C., 2007. MicroRNA—implications for cancer. Virchows Arch. 452, 1–10.

Saunders, M.A., Liang, H., Li, W.-H., 2007. Human polymorphism at microRNAs and microRNA target sites. Proc. Natl. Acad. Sci. U. S. A. 104, 3300–3305.

Scheller, H., Tobollik, S., Kutzera, A., Eder, M., Unterlehberg, J., Pfeil, I., Jungnickel, B., 2010. c-Myc overexpression promotes a germinal center-like program in Burkitt's lymphoma. Oncogene 29, 888–897.

Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864.

Schmitz, R., Ceribelli, M., Pittaluga, S., Wright, G., Staudt, L.M., 2014. Oncogenic mechanisms in Burkitt lymphoma. Cold Spring Harb. Perspect. Med. 4, a014282.

Schmitz, R., Young, R.M., Ceribelli, M., Jhavar, S., Xiao, W., Zhang, M., Wright, G., Shaffer, A.L., Hodson, D.J., Buras, E., Liu, X., Powell, J., Yang, Y., Xu, W., Zhao, H., Kohlhammer, H., Rosenwald, A., Kluin, P., Müller-Hermelink, H.K., Ott, G., Gascoyne, R.D., Connors, J.M., Rimsza, L.M., Campo, E., Jaffe, E.S., Delabie, J., Smeland, E.B., Ogwang, M.D., Reynolds, S.J., Fisher, R.I., Braziel, R.M., Tubbs, R.R., Cook, J.R., Weisenburger, D.D., Chan, W.C., Pittaluga, S., Wilson, W., Waldmann, T.A., Rowe, M., Mbulaiteye, S.M., Rickinson, A.B., Staudt, L.M., 2012. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. Nature 490, 116–120.

Shreeram, S., Hee, W.K., Demidov, O.N., Kek, C., Yamaguchi, H., Fornace, A.J., Jr, Anderson, C.W., Appella, E., Bulavin, D.V., 2006. Regulation of ATM/p53-dependent suppression of myc-induced lymphomas by Wip1 phosphatase. J. Exp. Med. 203, 2793–2799.

Sotiropoulou, G., Pampalakis, G., Lianidou, E., Mourelatos, Z., 2009. Emerging roles of microRNAs as molecular switches in the integrated circuit of the cancer cell. RNA. https://doi.org/10.1261/rna.1534709

Spencer, C.A., Groudine, M., 1991. Control of c-myc regulation in normal and neoplastic cells. Adv. Cancer Res. 56, 1–48.

Spender, L.C., Inman, G.J., 2014. Developments in Burkitt's lymphoma: novel cooperations in oncogenic MYC signaling. Cancer Manag. Res. 6, 27–38.

Su, S.-F., Han, F., Zhao, C., Huang, Y., Chen, C.-Y., Xiao, W.-W., Li, J.-X., Lu, T.-X., 2011. Treatment outcomes for different subgroups of nasopharyngeal carcinoma patients treated with intensity-modulated radiation therapy. Chin. J. Cancer 30, 565–573.

TargetScanHuman 6.2 [WWW Document], n.d. URL http://www.targetscan.org/vert_61/ (accessed 11.3.16).

Tomasetti, M., Santarelli, L., Neuzil, J., Dong, L., 2014. MicroRNA regulation of cancer metabolism: role in tumour suppression. Mitochondrion 19 Pt A, 29–38.

Tórtola, S., Marcuello, E., Risques, R.A., González, S., Aiza, G., Capellà, G., Peinado, M.A., 1999. Overall deregulation in gene expression as a novel indicator of tumor aggressiveness in colorectal cancer. Oncogene 18, 4383–4387.

Toung, J.M., Morley, M., Li, M., Cheung, V.G., 2011-6. RNA-sequence analysis of human B-cells. Genome Res. 21, 991–998.

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578.

Vousden, K.H., Crook, T., Farrell, P.J., 1993. Biological activities of p53 mutants in Burkitt's lymphoma cells. J. Gen. Virol. 74 (Pt 5), 803–810.

Wang, J., Chen, J., Sen, S., 2016. MicroRNA as Biomarkers and Diagnostics. J. Cell. Physiol. 231, 25–30.

Wang, T., Gu, J., Li, Y., 2014. Inferring the perturbed microRNA regulatory networks from gene expression data using a network propagation based method. BMC Bioinformatics 15, 255.

Wang, X., Cao, L., Wang, Y., Wang, X., Liu, N., You, Y., 2012-5. Regulation of let-7 and its target oncogenes (Review). Oncol. Lett. 3, 955–960.

Witkos, T.M., Koscianska, E., Krzyzosiak, W.J., 2011. Practical Aspects of microRNA Target Prediction. Curr. Mol. Med. 11, 93–109.

Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., Stephens, R.M., Okamoto, A., Yokota, J., Tanaka, T., Calin, G.A., Liu, C.-G., Croce, C.M., Harris, C.C., 2006. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell 9, 189–198.

You, T., Hu, W., Ge, X., Shen, J., Qin, X., 2011. Application of a novel inhibitor of human CD59 for the enhancement of complement-dependent cytolysis on cancer cells. Cell. Mol. Immunol. 8, 157–163.

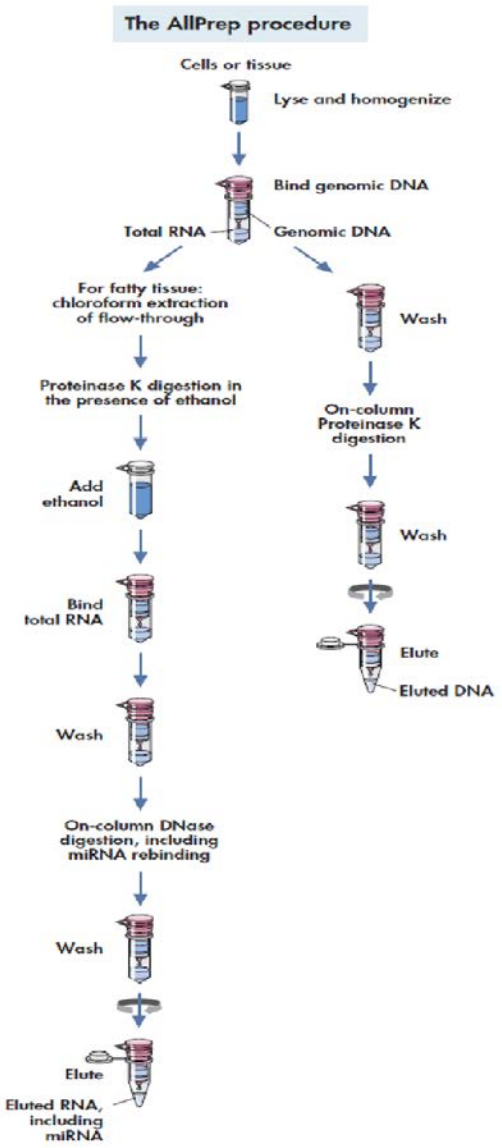
Zech, L., Haglund, U., Nilsson, K., Klein, G., 1976. Characteristic chromosomal abnormalities in biopsies and lymphoid-cell lines from patients with Burkitt and non-Burkitt lymphomas. Int. J. Cancer 17, 47–56.

Zhang, H.-H., Li, S.-Z., Zhang, Z.-Y., Hu, X.-M., Hou, P.-N., Gao, L., Du, R.-L., Zhang, X.-D., 2014. Nemo-like kinase is critical for p53 stabilization and function in response to DNA damage. Cell Death Differ. 21, 1656–1663.

Zhang, Z., Theurkauf, W.E., Weng, Z., Zamore, P.D., 2012. Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. Silence 3, 9.

APPENDIX

Appendix 1: mRNA and miRNA isolation workflow using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen).



Procedure

1. The FNA sample in RNAlater stabilizing solution was spun down at 250xg for 5min. The RNAlater solution was then carefully pipetted out not disturbing the FNA sample pellet at the bottom of the tube.
2. The next step involves disrupting the FNA and homogenize the lysate in Buffer RLT Plus using the QIAshredder.

Note: Ensure that β-ME(mecaptoethanol) is added to Buffer RLT Plus before use.

350ul of Buffer RLT Plus was added to the FNA sample and mixed thoroughly. The lysate was the pipetted directly in the QIAshredder spin column placed in a 2ml collection tube and centrifuged for 2 min at maximum speed.

3. The homogenized lysate was then transferred to an AllPrep DNA Mini spin column placed in a 2ml collection tube. This was then centrifuged for 30s at full speed (20000xg).
4. The AllPrep DNA Mini spin column was then placed in a new 2ml collection tube and stored at 4°C for DNA purification later. The flow-through was transferred into a new 2ml microcentrifuge tube for RNA purification.

Total RNA (including miRNA) purification

5. 50ul of Protienase K was added to the flow-through from step 4 and mixed by pipetting.
6. 200ul of 96% ethanol was then added and mixed well.
7. The mixture was then incubated for 10min. at room temperature.
8. 400ul of 96% ethanol was the added and mixed well.
9. 700ul of the sample, including any precipitate that may have formed, was transferred to an RNeasy Mini spin column placed in a 2ml collection tube. The spin column in the collection tube was then centrifuged for 15 sec. at full speed (20,000xg). The flow-though was then discarded.
10. Step 9 was then repeated until the entire sample had passed through the RNeasy Mini spin column.
11. 500ul of Buffer RPE was added to the RNeasy Mini spin column and centrifuged for 15 sec. at full speed (20,000xg). The flow-through was then discarded. Note: Ethanol needs to be added to Buffer RPE before use.
12. 10ul of DNase I stock solution was added to 70ul of Buffer RDD. This was mixed by gently inverting the tube and centrifuging briefly to collect the residual liquid from the sides of the tube. Note: DNase I is sensitive to physical denaturation, therefore mixing should be carried out by gently inverting the tube.
13. 80ul of DNase I incubation mix was added directly onto the RNeasy Mini spin column membrane and placed on the benchtop for 15 min.
14. 500ul of Buffer FRN was then added to the RNeasy Mini spin column and centrifuged for 15sec. at full speed. The flow-through for this step was saved for step 15. The flow through contains the small RNAs.

    Note: Buffer FRN was supplied as a concentrate, and isopropanol needs to be added to it before use.
15. The RNeasy Mini spin column was placed in a new 2ml collection tube. The flow-through from step 14 was applied to the spin column and centrifuged for 15sec. at full speed (20,000xg). The flow-through for this step was then discarded.
16. 500ul of Buffer RPE was added to the RNeasy Mini spin column and centrifuged at full speed for 15 sec.
17. 500ul of 96% ethanol was added to the RNeasy Mini spin column and centrifuged for 2 min. at full speed to wash the spin column membrane. The long centrifugation dries

the spin column membrane, ensuring that no ethanol is carried over during RNA elution.

18. The RNeasy Mini spin column was placed in a new 1.5ml collection tube. 30 ul of RNase free water was added directly to the spin column membrane. The column in the collection tube was then centrifuged at 8000xg to eluate the RNA.
19. Step 18 was repeated on the sam collection tube since >30ug of RNA was expected to be extracted from the FNA samples.

Appendix 2: Strand specific RNAseq library preparation Protocol (*Zhang et al., 2012*)

Procedure

rRNA depletion

High quality total RNA is essential for efficient rRNA removal.

1. The Ribo-Zero magnetic beads were mixed by gently pipetting. For each total RNA sample, 225 μl Ribo-Zero magnetic beads were dispensed into an RNase-free 1.5 ml centrifuge tube.
2. The tube was placed in the magnetic stand until the supernatant became clear, for approximately 1 minute.
3. With the tube still in the stand, the supernatant was discarded, which contains 0.1% sodium azide.
4. 225 μl RNase-free water was added to the tube, remove the tube from the magnetic stand, and mix the beads by gently pipetting.
5. The tube was returned to the magnetic stand, and let to stand until the solution became clear, and the water was discard.
6. The beads were resuspended in 65 μl Ribo-Zero magnetic bead suspension solution and 1 μl RiboGuard RNase Inhibitor (Illumina). It was mixed well by gently pipetting. The tube was stored at room temperature until step 9.
7. In a 1.5 ml centrifuge tube, the following mixture was prepared for each sample:
   4 μg fresh total RNA
   4 μl Ribo-Zero ‘reaction’ buffer
   10 μl Ribo-Zero rRNA removal solution
   Add water to make a 40 μl total volume.

The unused Ribo-Zero rRNA removal solution and 'reaction' buffer were stored at −80°C.

8. The solution was mixed gently by pipetting and incubated at 68°C for 10 minutes, then the tube were incubated at room temperature for 5 minutes.
9. The magnetic beads from step 6 were gently mixed by pipetting and the RNA solution from step 8 were added to the mixed beads. Using the same pipette tip, the beads were immediately mixed with the RNA by pipetting 10 times. Next, the tube was vortexed for 10 seconds at medium speed. Finally, the mixture was incubated at room temperature for 5 minutes.
10. The tube was vortexed at medium speed for 5 seconds and then incubated at 50°C for 5 minutes.
11. After the 5 minutes incubation, the tube was immediately placed in the magnetic stand for 2 minutes.
12. 84 µl of the supernatant was carefully removed and added to a new 1.5 ml centrifuge tube, and placed in the magnetic stand for 1 minute to get rid of the trace amount of leftover beads from the last step.
13. The supernatant was pipetted into a new 1.5 ml centrifuge tube and 16 µl of water was added.

Size selection and DNase treatment

RNA Clean & Concentrator-5 was used to enrich for RNAs >200 nt, which also removes 5S rRNA and tRNA.

14. 100 µl RNA binding buffer was mixed with 100 µl 100% ethanol. 200 µl of this mixture was then added to the 100 µl RNA from step 13.
15. The buffer/ethanol/RNA mixture was transferred into a Zymo-Spin IC column (Zymo Research) in a collection tube and centrifuged at 17,000 × *g* for 1 minute. The flow-through was then discarded.
16. 400 µl of RNA Wash Buffer was added to the column, spun at 17,000 × *g* for 1 minute. The flow-through was the discarded.
17. To degrade contaminating DNA, the following reagents were mixed:
    3 µl TURBO DNase (2 U/µl)
    3 µl 10× TURBO DNase Buffer
    24 µl RNA Wash Buffer and add the 30 µl mixture to the column.

The column was incubated at 37°C for 30 minutes. The column was centrifuged at 17,000 × g for 1 minute, and the flow-through was discarded.

18. 400 µl RNA Prep Buffer was added to the column, centrifuged at 17,000 × *g* for 1 minute, and the flow-through was discarded.
19. 800 µl RNA Wash Buffer was added to the column, centrifuged at 17,000 × *g* for 1 minute, and the flow-through was discarded.
20. 400 µl RNA Wash Buffer again added to the column, centrifuged at 17,000 × *g* for 1 minute, and the flow-through was discarded.
21. The column was then centrifuged at 17,000 × *g* for 2 minutes.
22. To elute the RNA, the collection tube was replaced with a new 1.5 ml centrifuge tube, and then 10 µl of water was added to the column. This was then incubated at room temperature for 1 minute and centrifuged at 17,000 × *g* for 1 minute to collect the RNA/flow-through.
23. 1 µl RNA was taken to measure the concentration using a NanoDrop spectrophotometer.

Library preparation

1. Fragmenting the RNA. At high temperature, the metal ions within the 5× first strand buffer will hydrolyze the RNA into short fragments. In a 0.2 ml tube, 4 µl rRNA-depleted total RNA was mixed with 4 µl of 5× first strand buffer. The tube were placed into a PCR thermal cycler pre-heated to 94°C and incubated for precisely 4 minutes and 50 seconds. Then quickly chill the tube on ice for at least 1 minute.
2. To reverse transcribe the RNA into first strand cDNA, to the PCR tube these were added:

    1.5 µl 100 mM DTT

    1 µl Random Primer (hexamers, 3 µg/µl)

    7 µl water.

    1 µl dNTP mixture (dATP, dCTP, dGTP, dTTP, 10 mM each)

    0.5 µl 100 mM DTT

    1 µl SuperScript III Reverse Transcriptase (200 U/µl)

    4 µg Actinomycin D (optional, may enhance strand specificity but decrease uniformity of strand coverage).

    The mixture was incubated at 65°C for 3 minutes, and then quickly chilled on ice for 1 minute. Next, it was incubated at 25°C for 5 minutes, then at 50°C for 1 hour. Heat at

70°C for 15 minutes to inactive the reverse transcriptase. Finally, 36 µl AMPure XP beads was used to purify the cDNA, eluting with 22 µl elution buffer.

3. To convert the first strand cDNA to double-stranded cDNA incorporating dUTP instead of dTTP, the following reagents were added to the cDNA from step 2:

   3 µl second strand buffer/10× NEB Buffer 2

   2 µl dUTP mixture (20 mM dUTP, 10 mM dATP, dCTP, dGTP)

   1 µl RNase H (2 U/µl)

   2 µl DNA polymerase I (10 U/µl)

   0.5 µl 100 mM DTT.

   Incubate at 16°C for 2.5 hours.

   After incubation, the double-stranded cDNA were purified using 45 µl AMPure XP beads and elute the cDNA into a 1.5 ml centrifuge tube using 33 µl elution buffer. After this, continue or store the sample at -20°C.

4. The ends of the double-stranded cDNA were then repaired. DNA polymerase I, which is used for second strand cDNA synthesis, uses as primers the RNA leftover from RNase H digestion. Consequently, the double-stranded cDNAs generated in step 3 have 3′ overhanging ends. Step 4 converts the sticky ends into blunt ends. The following mixture was added to the DNA from step 3: 5 µl 10× T4 DNA Ligase buffer

   2 µl dNTP mixture (10 mM each)

   5 µl T4 DNA Polymerase (3 U/µl)

   1 µl Klenow DNA Polymerase (5 U/µl)

   5 µl T4 PNK (10 U/µl).

   The mixture was then incubated at 20°C for 30 minutes.

5. To establish a library with the narrow size range (200 bp to 350 bp) required for successful high throughput sequencing, the cDNA is purified using AMPure XP beads, which exploit the finding that carboxyl coated magnetic beads bind distinct DNA size ranges depending on polyethylene glycol (PEG) and salt concentration. After end repair, the reaction was mixed with 35 µl AMPure XP beads and incubated at room temperature for 5 minutes. Then the tube was placed in the magnetic stand for 3 minutes. The supernatant was transferred into a new tube and the beads were discarded. To the new tube with the supernatant, an additional 40 µl of AMPure XP beads was added and then the standard AMPure XP bead purification protocol was followed, The DNA was eluted with 33 µl elution buffer.

6. The PCR products were tailed with adenosine to facilitate adapter ligation. Klenow Fragment was used with D355A and E357A mutations (Klenow 3′ to 5′ exo⁻, 5 U/µl), a DNA polymerase lacking both 3′ to 5′ and 5′ to 3′ exonuclease activities, to add a single adenosine to the 3′ ends of the DNA. To the DNA from step 5, add:

   5 µl second strand buffer/10× NEB Buffer 2

   1 µl dATP (10 mM)

   3 µl Klenow 3′ to 5′ exo⁻ (5 U/µl)

   9 µl water.

   Incubate at 37°C for 30 minutes. After incubation, purify the DNA using 60 µl AMPure XP beads; elute with 24 µl elution buffer.

7. The Y-shaped adapters (oligonucleotide sequences © 2007–2011 Illumina, Inc. All rights reserved) were added. To prepare the Y-shaped adapter, 25 µl adapter oligo 1 and oligo 2 were mixed (each at 50 µM stock concentration). These were heated at 95°C for 2 minutes, then ramp down slowly to room temperature. The oligo mixture in an aluminium heat block were heated for 2 minutes. Then the block was removed from the heater and let to cool down to room temperature, for approximately 30 minutes. The adapters were ligated to the purified double-stranded cDNA by adding:

   25 µl 2× rapid ligation buffer

   1 µl adapter (10 µM)

   1.5 µl T4 DNA Ligase (600 U/µl).

   This mixture was incubated at room temperature for 15 minutes. After incubation, 50 µl AMPure XP beads was used to purify the DNA. The DNA was eluted with 30 µl elution buffer.

8. Treat with 5 U/µl UDG. Add 2 µl UDG to the DNA from step 7, and incubate at 37°C for 30 minutes.

9. PCR amplify the cDNA. Add the following mixture to the DNA from step 8:

   10 µl 5× HF Buffer (Phusion Polymerase, NEB)

   1 µl 10 µM PCR Primer 2 (one of the twelve to provide the barcode)

   1.5 µl dNTP (10 mM each)

   0.5 µl Phusion High-Fidelity DNA Polymerase (2 U/µl)

   5 µl water.

   5 µl 10× AccuPrime *Pfx* Reaction Mix (Invitrogen)

   1 µl 10 µM PCR Primer 2 (one of the twelve to provide the barcode)

1 µl AccuPrime *Pfx* DNA Polymerase (2.5 U/µl)

11 µl water.

10. Incubate the tube at 98°C for 40 seconds, 65°C for 30 seconds and 72°C for 30 seconds. After the incubation, pause the PCR machine, and then add 1 µl 10 µM PCR Primer 1. Continue the PCR with 10 cycles of 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds, followed by incubation at 72°C for 3 minutes. Then purify the library with 50 µl AMPure XP beads. Finally, elute the DNA with 20 µl elution buffer.
11. Alternatively, the PCR can be performed using AccuPrime *Pfx* DNA Polymerase: Incubate the tube at 95°C for 40 seconds, 65°C for 30 seconds and 68°C for 30 seconds. After the incubation, pause the PCR machine, and then add 1 µl 10 µM PCR Primer 1. Continue the PCR with 10 cycles of 95°C for 15 seconds, 65°C for 30 seconds, 68°C for 30 seconds, followed by incubation at 68 °C for 3 minutes. Then purify the library with 50 µl AMPure XP beads. Finally, elute the DNA with 20 µl elution buffer.

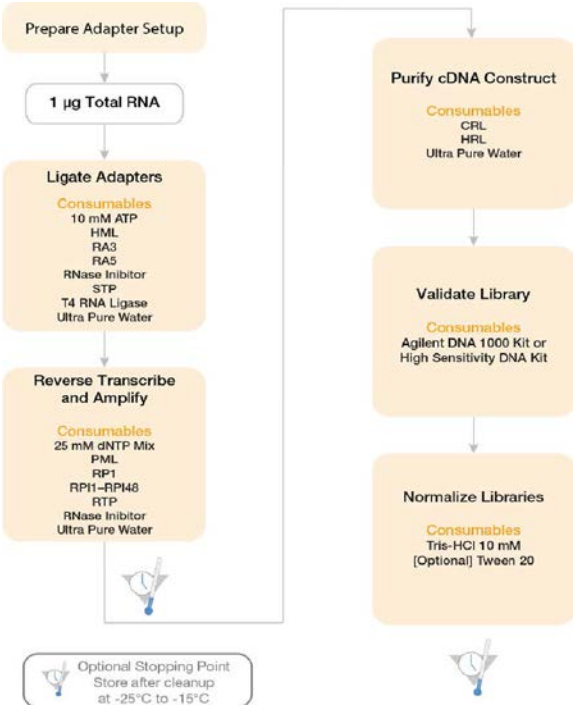
Additional protocol *using AMPure XP beads to purify DNA*

1. Warm the AMPure XP beads to room temperature, mix well, and pipette the required volume from the stock bottle to the sample tube.
2. Mix the beads and DNA sample by gently pipetting, and then incubate at room temperature for 5 minutes.
3. Place the tube in the magnet stand for 5 minutes until the supernatant appears clear.
4. Discard the supernatant.
5. Keep the tube in the stand and add 180 µl of 70% (v/v) ethanol into the tube without disturbing the beads.
6. Wait for 30 seconds and discard the ethanol supernatant.
7. Repeat steps 5 and 6.
8. To remove any ethanol remaining on the sides of the tube, centrifuge the tube at 1000 × *g* for 1 minute.
9. Place the tube in the magnetic stand for 30 seconds, and then remove any residual ethanol using a 10 µl pipette.
10. Add the specified volume of elution buffer to the beads and pipette to mix.
11. Wait 3 minutes, and then place the tube in the magnetic stand for 3 minutes.
12. Use a 10 µl pipette to carefully transfer the eluted DNA to a new tube (sacrifice 1 to 2 ul to avoid carrying over any beads).

Appendix 3: MicroRNA Library Preparation protocol.

Procedure

The diagram illustrates a single TruSeq Small RNA Library Prep Kit workflow.



a) Ligation of 3′ Adapter and 5′ Adapter to MicroRNAs

- A 3′-ligation reaction was set up in a nuclease-free 200µL PCR tube using 5.2µL of 1µg total RNA and 1 µL diluted 3′ RNA adaptors. The tube was then incubated in a thermal cycler at 70˚C for 2 min., and then transferred immediately to ice.
- 1 µL of 10× ligation buffer; 0.8 µl 100mM $MgCl_2$; 1.5 µL T4 RNA ligase, truncated and 0.5 µL RNase Out (40U/µl) was added to the tube and mixed well. Then the tube was incubated in a thermal cycler at 22˚C for 1 hour.
- The 5′ adaptors was prepared for ligation by heating it at 70˚C for 2 min. then transferring to ice.
- 1 µL 10mM ATP, 1µl SRA 5′ adaptor, and 1µL T4 RNA ligase was added to the 3′ ligation mixture. The tube was then incubated in a thermal cycler at 20˚C for 1 hour and was later kept at 4˚C until the next reaction of reverse transcription and amplifying the ligated microRNA products.

b) Reverse Transcription and PCR Amplification of Adaptor-Ligated MicroRNAs

- 25mM dNTP was diluted to 12.5mM by mixing 1µL of 25mM dNTP with 1 µL of nuclease-free water. A reverse transcription reaction was then set up in a nuclease-free 200µl PCR tube containing 4 µL 3′ and 5′ adaptor-ligated RNA; 1µL diluted SRA RT primer.
- The tubes were then incubated at 70˚C for 2 min. in a thermocycler then transferred immediately to ice.

- 20 µL 5× first strand buffer, 0.5 µL 12.5mM dNTP, 1µL 100mM DTT, 0.5µL RNase Out (40U/µL) and 1.0 µL Superscript II RT (200U/µL) was added to the reaction tube and mixed well.
- The tube was then incubated in a thermocycler at 48˚C for 30 min. then at 44˚C for 1 hour.
- After incubation more reagents: 10µL 5× clonal Phu buffer, 0.5µL primer GX1; 0.5µL primer GX2; 0.5 µL 25mM dNTP mix; 0.5 µL phusion polymerase and 28µL H_2O were added to the reaction tube.
- The material was then amplified using PCR cycling conditions of 1 cycle of 30s at 98˚C followed by 12 cycles of 10s at 98˚C, 30s at 60˚C, 15s at 72˚C then followed 10 min at 72˚C.

c) Gel Purification of cDNA construct (miRNA Library)

The amplified cDNA constructs were purified using gel electrophoresis in preparation for subsequent cluster generation. After gel purification, the cDNA's were eluted and were concentrated by ethanol precipitation. Ethanol precipitation resulted in a more concentrated final small RNA library. Individual libraries with unique indexes were pooled and gel purified together. Equal volumes of the library or molar amounts of the library were combined and then the samples were loaded on the gel.

- An appropriate volume of 1× TBE buffer was needed and prepared from a 5× TBE buffer stock solution for use in electrophoresis.
- 1 µL of a 25bp ladder was mixed with 1 µL of 6× DNA loading dye and loaded onto a well in the 6% PAGE gel mounted on a gel electrophoresis apparatus.
- 50 µL of the amplified cDNA construct was mixed with 10 µL of 6× DNA loading dye, then 25 µL of the mixed amplified cDNA construct and loading dye was loaded on two wells each on the 6% PAGE gel. The gel was then run for 30-35 min. at 200V until the front of the dye exits the gel.
- 10× gel elution buffer was diluted into a fresh tube by mixing 1 volume of the 10× gel elution buffer stock with 9 volumes of nuclease-free water.

- A sterile, nuclease-free, 0.5 ml micro-centrifuge tube was then punctured 4-5 times with a 21-gauge needle. This 0.5 ml micro-centrifuge tube was then placed into a sterile, round-bottomed, nuclease-free, 2 ml micro-centrifuge tube.
- The gel was then viewed on a Dark Reader Illuminator. Using a clean scalpel the bands corresponding to approximately the adapter-ligated constructs derived from the 22 and 30 nt small RNA fragments were cut out. The bands containing the 22nt RNA fragments with both adaptors attached is a total of 93 nt in length, while those containing 30nt RNA fragments with adaptors is 100 nt in length.
- The bands of interest were then placed into a 0.5ml punctured micro-centrifuge tube.
- The tubes were then centrifuged at 14,000 rpm in a micro-centrifuge for 2 min at room temperature to move the gel through the holes into the 2 ml tubes.
- 100 µl of the 1× gel elution buffer was added to the gel debris into the 2ml tube. The DNA were then eluted by rotating the tube gently at room temperature for 2 hours.
- The eluate and the gel debris will then be transferred to the top of a spin-X filter, then the filter centrifuged for 2 min at 14,000 rpm.
- 1 µl of glycogen, 10 µl of 3M NaOAc, and 325 µl of -20˚C 100% ethanol was then added. Immediately the micro-centrifuge contents were again centrifuged at 14,000 rpm for 20min,
- The supernatant was later on removed and discarded, leaving the pellet intact. The pellet was then washed with 500 µl of room temprature70% ethanol. The supernatant was again removed and discarded, leaving the pellet intact. The pellet was then dried and later resuspended in 10 µl of resuspension buffer.
- 1 µl of the DNA was then loaded onto an Agilent Technologies Bioanalyzer using a DNA-1000 chip to determine the size and concentration of the prepared library. The prepared DNA sample libraries was diluted to about 2nM and then denatured for clustering.

d) Sequencing the MicroRNA library with the Illumina Genome Analyzer System

The DNA fragments from the microRNA library were denatured into single strands, and annealed to complementary oligonucleotides that are pregrafted on the Illumina flow cell

surface. The single DNA molecule were then amplified *in situ* by solid face isothermal bridge amplification to form a cluster, which is ~1 µm in diameter. Millions of clusters were formed simultaneously on the flow cell surface. The DNA in each cluster were linearized by cleavage within one adaptor sequence and denatured, generating single-stranded templates which was used for sequencing.

In sequencing, the sequencing primer was hybridized to the template, and fluorescence-labeled reversible terminator deoxyribonucleotide is incorporated by a modified DNA polymerase. Images generated for the fluorescence of each cluster were later on analyzed to generate the sequences of the microRNAs.

Appendix 3.1.: Index (Barcode) Sequences

The TruSeq Small RNA Library Prep Kit contains the following indexed adapter sequences:

Indices A Box – Index Sequences 1-12

Index	Sequence		Index	Sequence
RPI1	ATCACG		RPI7	CAGATC
RPI2	CGATGT		RPI8	ACTTGA
RPI3	TTAGGC		RPI9	GATCAG
RPI4	TGACCA		RPI10	TAGCTT
RPI5	ACAGTG		RPI11	GGCTAC
RPI6	GCCAAT		RPI12	CTTGTA

Indices B Box – Index Sequences 13-24

Index	Sequence		Index	Sequence
RPI13	AGTCAA		RPI19	GTGAAA
RPI14	AGTTCC		RPI20	GTGGCC
RPI15	ATGTCA		RPI21	GTTTCG
RPI16	CCGTCC		RPI22	CGTACG
RPI17	GTAGAG		RPI23	GAGTGG
RPI18	GTCCGC		RPI24	GGTAGC

Appendix 4: Endemic BL Sample Information and Number of sequencing reads

Sample ID	Sex	Diagnosis Report	Tumour site	Survival Status	mRNAseq status	No. of paired end mRNA reads	miRNA conc. (pg/µl)	% of total Small RNA	miRNAseq status	No. of miRNA reads
eBL_23	M	CONFIRMED BL	NA	unknown	Seq'ed	11,308,986	32602.3	31%	Seq'ed	1,993,092
eBL_11	M	CONFIRMED BL	Abdomen	Survivor	Seq'ed	52,646,949			Not Seq'ed	
eBL_09	M	CONFIRMED BL	Abdomen	Survivor	Seq'ed	9,720,944			Not Seq'ed	
eBL_10	M	CONFIRMED BL	Abdomen	Died during remission	Seq'ed	11,065,910			Not Seq'ed	
eBL_14	F	CONFIRMED BL	Abdomen	Survivor	Seq'ed	11,017,129	9856.2	43%	Seq'ed	3,272,539
eBL_04	M	CONFIRMED BL	Jaw	Died during remission	Seq'ed	11,996,151			Not Seq'ed	
eBL_12	M	CONFIRMED BL	Abdomen	Survivor	Seq'ed	53,727,557	49651	44%	Seq'ed	961,769
eBL_13	F	CONFIRMED BL	Abdomen	Survivor	Seq'ed	11,868,132	25659.4	47%	Seq'ed	939,731
eBL_01	M	CONFIRMED BL	Jaw	Died in hospital	Seq'ed	10,295,415	122691.8	49%	Seq'ed	1,255,858
eBL_02	M	CONFIRMED BL	Jaw	Died in hospital	Seq'ed	8,955,776	21903	25%	Seq'ed	1,572,454
eBL_16	F	CONFIRMED BL	Abdomen	Survivor	Seq'ed	12,439,032	7485.2	38%	Seq'ed	1,279,133
eBL_25	M	CONFIRMED BL	Abdomen	Died in hospital	Seq'ed	12,670,654	18396.7	42%	Seq'ed	1,828,963
eBL_17	F	CONFIRMED BL	Abdomen	Survivor	Seq'ed	10,474,949	10970.5	42%	Seq'ed	1,798,499
eBL_26	F	CONFIRMED BL	Jaw	Survivor	Seq'ed	12,836,849			Not Seq'ed	
eBL_27	M	CONFIRMED BL	Jaw	Survivor	Seq'ed	9,241,469	177670	86%	Seq'ed	81,791
eBL_03	M	CONFIRMED BL	Jaw	Died in hospital	Seq'ed	11,295,180	11566.2	36%	Seq'ed	1,774,226

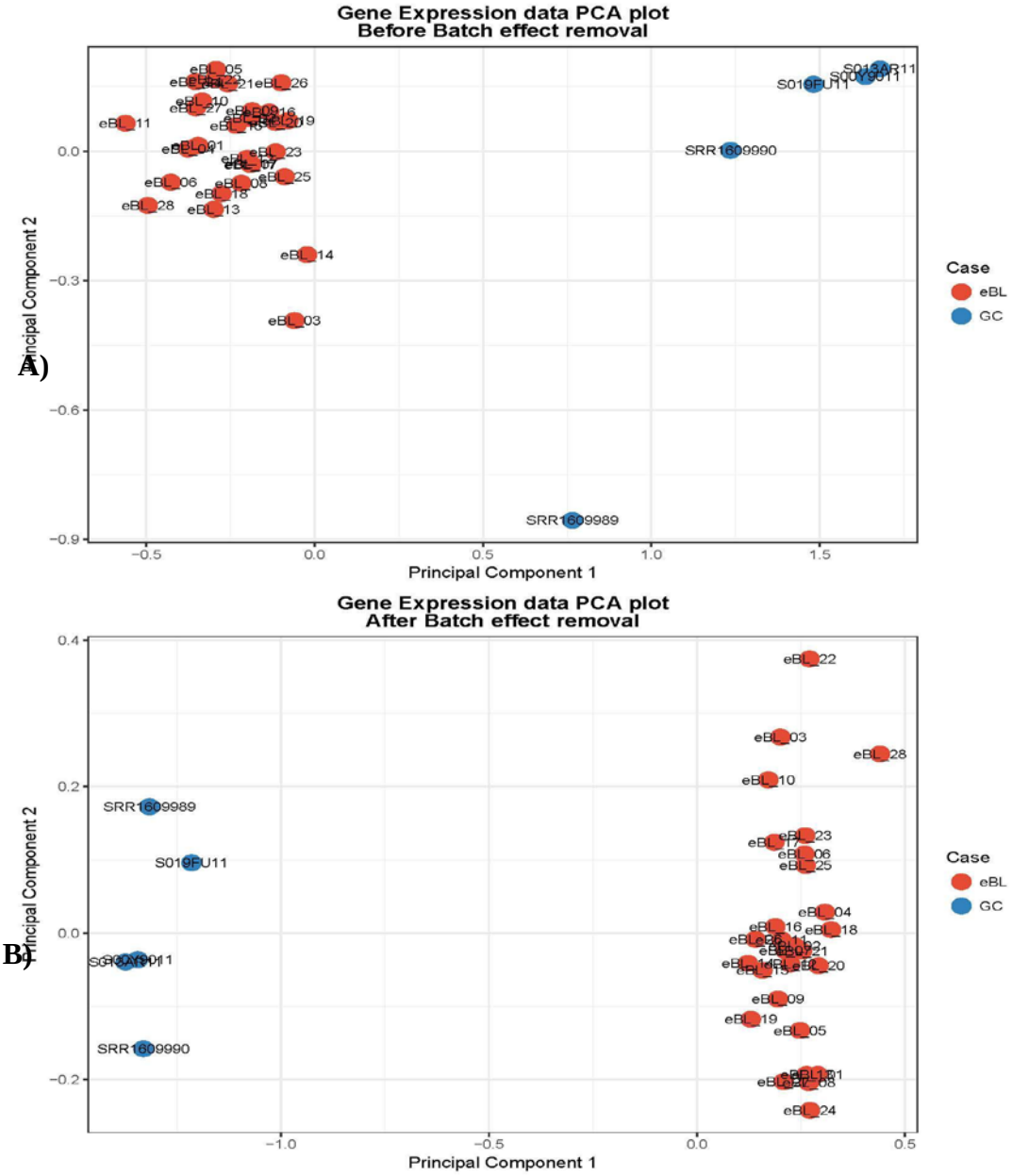
Appendix 4 (*Continued*)

Sample ID	Sex	Diagnosis Report	Tumour site	Survival Status	mRNAseq status	No. of paired end mRNA reads	miRNA conc. (pg/µl)	% of total Small RNA	miRNAseq status	No. of miRNA reads
eBL_18	M	CONFIRMED BL	Abdomen	Died in hospital	Seq'ed	11,118,925			Not Seq'ed	
eBL_05	M	CONFIRMED BL	Jaw	Survivor	Seq'ed	13,554,532	55273.4	3%	Not Seq'ed	
eBL_19	M	CONFIRMED BL	Abdomen	Died in hospital	Seq'ed	11,477,167	7935	1%	Not Seq'ed	
eBL_06	F	CONFIRMED BL	Jaw	Survivor	Seq'ed	11,729,827	18348	1%	Not Seq'ed	
eBL_20	M	CONFIRMED BL	Abdomen	Died in hospital	Seq'ed	9,919,996	43663.8	14%	Seq'ed	2,387,482
eBL_28	M	CONFIRMED BL	Orbital	Survivor	Seq'ed	7,636,841			Not Seq'ed	
eBL_15	M	CONFIRMED BL	Jaw	Survivor	Seq'ed	10,663,577	8848.1	31%	Seq'ed	378,850
eBL_07	M	CONFIRMED BL	Jaw	Died in hospital	Seq'ed	13,740,469	29108.6	44%	Seq'ed	1,190,289
eBL_21	F	CONFIRMED BL	Abdomen	Survivor	Seq'ed	12,355,220			Not Seq'ed	
eBL_22	F	CONFIRMED BL	Abdomen	Survivor	Seq'ed	11,135,655	18763.5	27%	Seq'ed	1,239,796
eBL_08	M	CONFIRMED BL	Jaw	Survivor	Seq'ed	13,320,261			Not Seq'ed	
eBL_29	M	CONFIRMED BL		NA	Not Seq'ed		7450.4	31%	Seq'ed	882,138
eBL_30	M	CONFIRMED BL	Abdomen	NA	Not Seq'ed		44012.8	32%	Seq'ed	566,480
eBL_24	M	CONFIRMED BL	NA	NA	Seq'ed	11,653,236			Not Seq'ed	

Appendix 5: Germinal center (GC) B cell sample Information and Number of sequencing reads

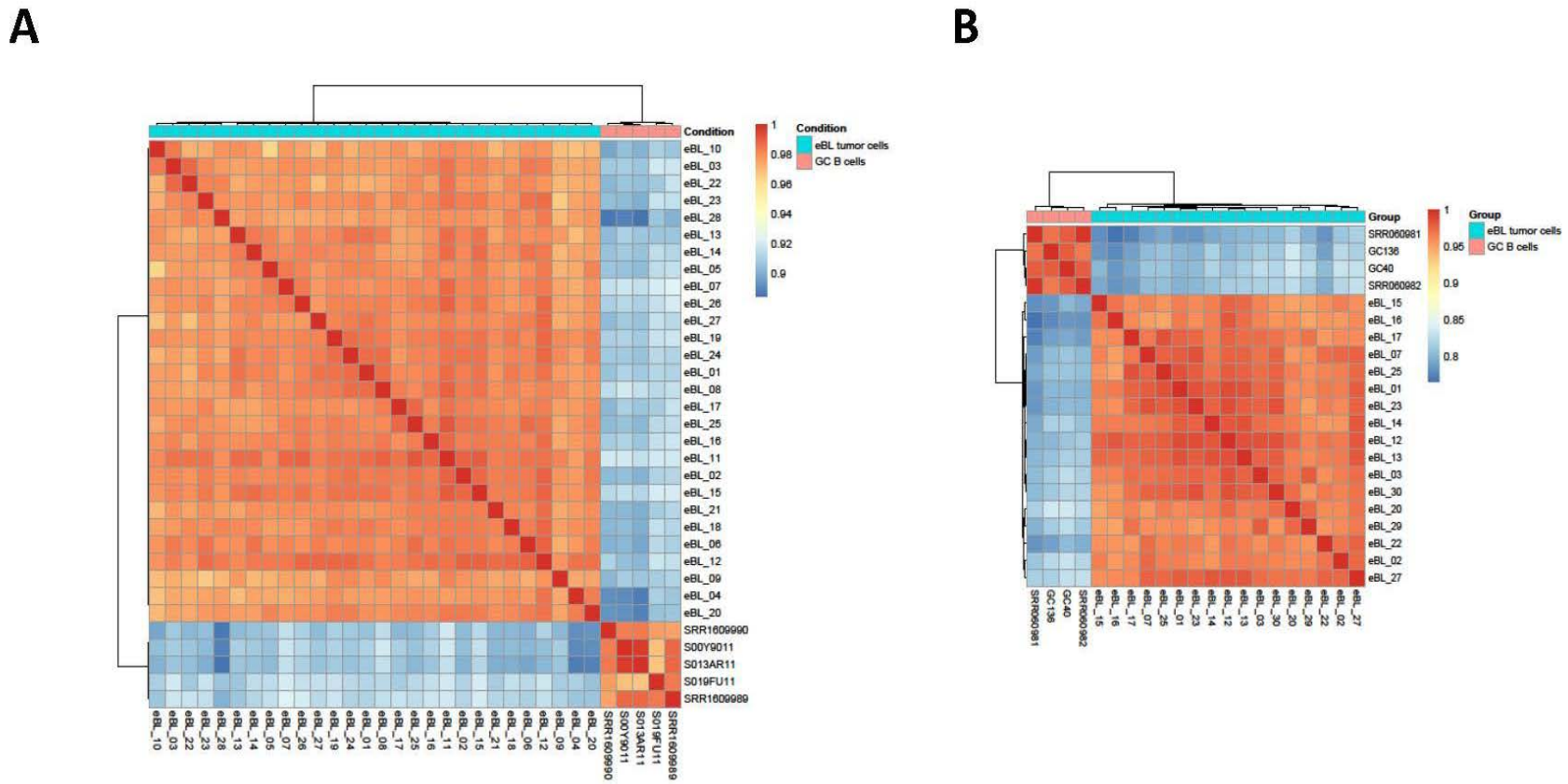
Sample ID	Gender	Age	Health status	Tissue type	Cell type	Markers	Sample data	No. of paired end mRNA reads	No. of miRNA reads
S00Y9O11	F	5 - 10 yrs	Healthy	tonsil	GC B cells	CD19+/CD20hi/CD38++	mRNAseq	113,967,029	NA
S013AR11	M	0 - 5 yrs	Healthy	tonsil	GC B cells	CD19+/CD20hi/CD38++	mRNAseq	102,574,943	NA
S019FU11	F	5 - 10 yrs	Healthy	tonsil	GC B cells	CD19+/CD20hi/CD38++	mRNAseq	104,729,799	NA
SRR1609989	NA	NA	Healthy	tonsil	GC B cells	CD19+/CD10+/CD44low/CXCR4-	mRNAseq	18,102,029	NA
SRR1609990	NA	NA	Healthy	tonsil	GC B cells	CD19+/CD10+/CD44low/CXCR4-	mRNAseq	16,225,477	NA
GC40	NA	NA	Healthy	Lymphoid organ	GC B cells	CD19+/IgD−/CD38++	miRNAseq	NA	1,232,232
GC136	NA	NA	Healthy	Lymphoid organ	GC B cells	CD19+/IgD−/CD38++	miRNAseq	NA	844,514
SRR060981	F	3 yrs	Healthy	tonsil	GC B cells	CD19+/IgD-/CD38++/intCD77+	miRNAseq	NA	3,681,838
SRR060982	F	3 yrs	Healthy	tonsil	GC B cells	CD19+/IgD-/CD38++/intCD77-	miRNAseq	NA	3,133,401

Appendix 6: Principal Component Analysis (PCA)



Appendix 6: **A**) PCA plot showing clustering of GC B cells and eBL tumour cells before batch effect removal. **B**) PCA plot showing clustering of GC B cells and eBL tumour cells after batch effect/noise removal using svaseq. Better clustering of GC B cells based on cell type was observed and not clustering based the different study sample datasets

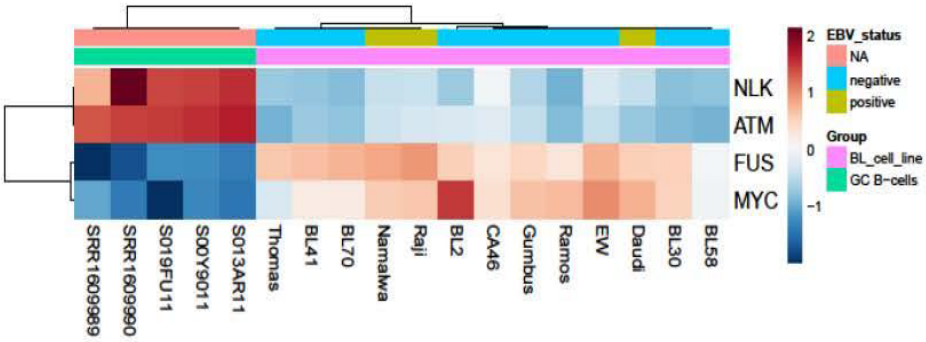
Appendix 7: Sample to sample hierarchal clustering



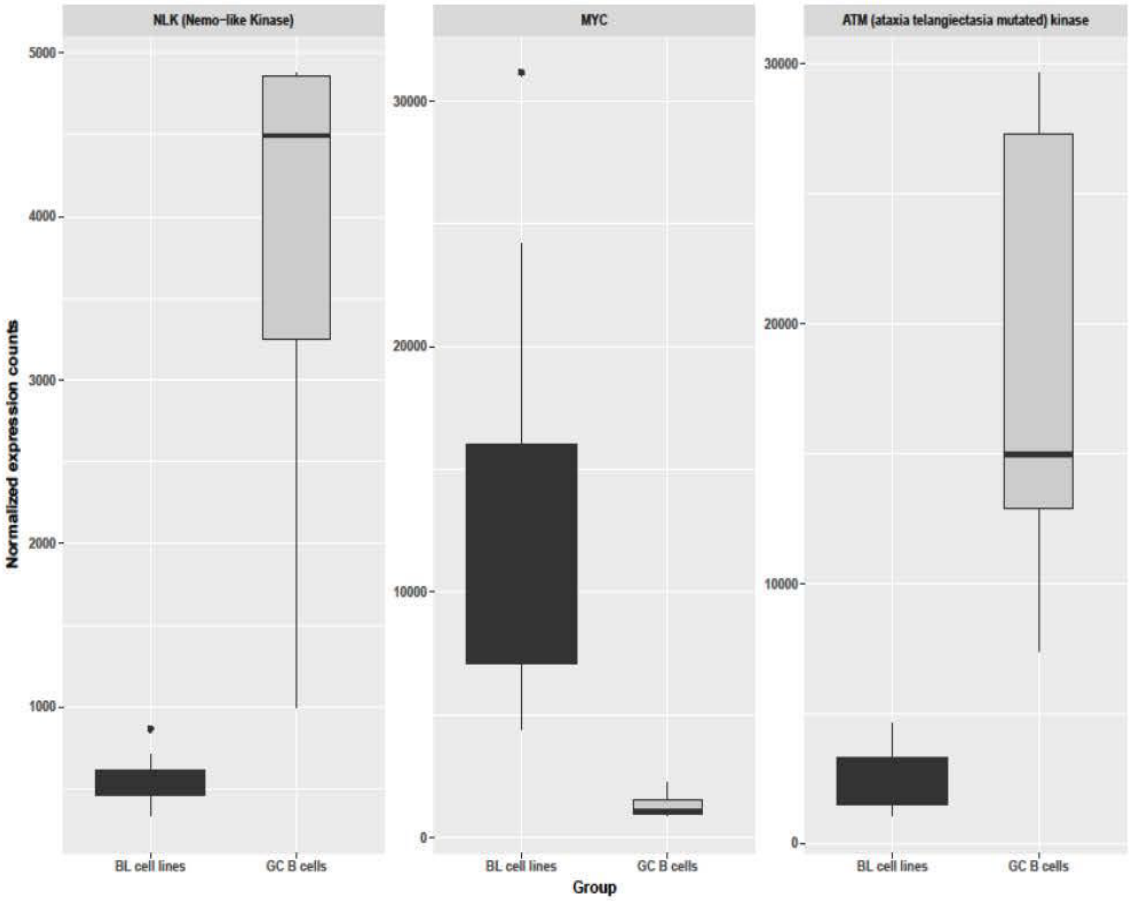
Appendix 7: Sample to sample hierarchal clustering of eBL tumour cells and GC B cells based on **A**) mRNA expression profiles, **B**) miRNA expression profiles with highest correlation of variation (CV) values (calculated using regularized log transformed mRNA and miRNA expression values).

Appendix 8: *ATM* and *NLK* downregulation in BL cell lines**.**

A.) Hierarchical clustering of BL cell lines and germinal center (GC) B cells based on the expression of *MYC*, *NLK* and *ATM* genes.



B.) Expression changes of *ATM* and *NLK* in BL cell line compared to GC B cells.



Appendix 9: Validated targets of DE miRNAs

Appendix 9: Table 1- Validated target genes of the differentially expressed miRNAs

miRNA	MIMATid	Gene	EntrezID	PMID
hsa-miR-320a	MIMAT0000510	ANP32B	10541	23622248
hsa-miR-30b-5p	MIMAT0000420	ASB3	51130	22473208
hsa-miR-30e-5p	MIMAT0000692	ASB3	51130	22473208
hsa-miR-19b-3p	MIMAT0000074	ATM	472	20371350
hsa-miR-26a-5p	MIMAT0000082	ATM	472	24211747
hsa-miR-30b-5p	MIMAT0000420	ATM	472	22473208
hsa-miR-30e-5p	MIMAT0000692	ATM	472	22473208
hsa-miR-92a-3p	MIMAT0000092	ATM	472	23622248
hsa-miR-186-5p	MIMAT0000456	BLOC1S6	26258	19536157
hsa-miR-30b-5p	MIMAT0000420	BLOC1S6	26258	20371350
hsa-miR-26a-5p	MIMAT0000082	BMP2K	55589	20371350
hsa-let-7b-5p	MIMAT0000063	BMP7	655	23622248
hsa-miR-24-3p	MIMAT0000080	BMP7	655	23824327
hsa-miR-130a-3p	MIMAT0000425	BRWD1	54014	23313552
hsa-miR-130a-3p	MIMAT0000425	BRWD1	54014	21572407
hsa-miR-182-5p	MIMAT0000259	BRWD1	54014	23313552
hsa-miR-183-5p	MIMAT0000261	BRWD1	54014	23313552
hsa-miR-19b-3p	MIMAT0000074	BRWD1	54014	23313552
hsa-miR-19b-3p	MIMAT0000074	BRWD1	54014	21572407
hsa-miR-103a-3p	MIMAT0000101	C20orf27	54976	20371350
hsa-miR-320a	MIMAT0000510	C20orf27	54976	23622248
hsa-miR-423-5p	MIMAT0004748	C20orf27	54976	23592263
hsa-miR-21-3p	MIMAT0004494	C4orf32	132720	23446348
hsa-miR-30e-3p	MIMAT0000693	CCDC86	79080	23592263
hsa-miR-486-5p	MIMAT0002177	CD40	958	19475450
hsa-miR-16-5p	MIMAT0000069	CDK17	5128	20371350
hsa-miR-26a-5p	MIMAT0000082	CFLAR	8837	23622248
hsa-miR-92a-3p	MIMAT0000092	CMTM6	54918	23622248
hsa-let-7f-5p	MIMAT0000067	COPS6	10980	21530537

Appendix 9: Table 1 *(Continued)*

miRNA	MIMATid	Gene	EntrezID	PMID
hsa-miR-16-5p	MIMAT0000069	CYB561A3	220002	20371350
hsa-miR-19b-3p	MIMAT0000074	DCP2	167227	22473208
hsa-miR-92a-3p	MIMAT0000092	DCP2	167227	23622248
hsa-miR-92a-3p	MIMAT0000092	DMXL1	1657	23622248
hsa-miR-16-5p	MIMAT0000069	DNAJC10	54431	21572407
hsa-miR-20a-5p	MIMAT0000075	DNAJC10	54431	23446348
hsa-miR-21-3p	MIMAT0004494	DNAJC10	54431	20371350
hsa-miR-20a-5p	MIMAT0000075	DUSP2	1844	22012620
hsa-miR-340-5p	MIMAT0004692	DUSP2	1844	22012620
hsa-miR-423-5p	MIMAT0004748	FKBP4	2288	23622248
hsa-miR-182-5p	MIMAT0000259	FMNL3	91010	23622248
hsa-miR-20a-5p	MIMAT0000075	FMNL3	91010	22473208
hsa-miR-1260b	MIMAT0015041	FUS	2521	23622248
hsa-miR-221-3p	MIMAT0000278	FUS	2521	23622248
hsa-miR-331-3p	MIMAT0000760	FUS	2521	23622248
hsa-let-7b-5p	MIMAT0000063	GPATCH4	54865	18668040
hsa-miR-92a-3p	MIMAT0000092	HERC1	8925	23622248
hsa-miR-16-5p	MIMAT0000069	HIST1H1C	3006	23622248
hsa-miR-186-5p	MIMAT0000456	HIST1H1C	3006	23622248
hsa-miR-27b-3p	MIMAT0000419	HIST1H1C	3006	23622248
hsa-miR-186-5p	MIMAT0000456	HIST1H1E	3008	23622248
hsa-miR-92a-3p	MIMAT0000092	HIST1H1E	3008	23622248
hsa-miR-16-5p	MIMAT0000069	HIST1H2BC	8347	23622248
hsa-miR-20a-5p	MIMAT0000075	HIST1H2BD	3017	21572407
hsa-miR-16-5p	MIMAT0000069	HIST1H2BK	85236	23622248
hsa-miR-186-5p	MIMAT0000456	HIST1H2BK	85236	23622248
hsa-miR-1260b	MIMAT0015041	HNRNPA3	220988	23622248
hsa-miR-221-3p	MIMAT0000278	HNRNPD	3184	20371350
hsa-miR-222-3p	MIMAT0000279	HNRNPD	3184	23622248

Appendix 9: Table 1 *(Continued)*

miRNA	MIMATid	Gene	EntrezID	PMID
hsa-miR-16-5p	MIMAT0000069	INO80D	54891	23622248
hsa-miR-340-5p	MIMAT0004692	INO80D	54891	22291592
hsa-miR-10a-5p	MIMAT0000253	IRGQ	126298	21572407
hsa-miR-92a-3p	MIMAT0000092	IRGQ	126298	23622248
hsa-miR-183-5p	MIMAT0000261	ITPKB	3707	23622248
hsa-miR-186-5p	MIMAT0000456	ITPKB	3707	23622248
hsa-miR-20a-5p	MIMAT0000075	ITPKB	3707	22473208
hsa-miR-16-5p	MIMAT0000069	ITPR1	3708	22473208
hsa-miR-19b-3p	MIMAT0000074	ITPR1	3708	22473208
hsa-miR-340-5p	MIMAT0004692	ITPR1	3708	20371350
hsa-miR-340-5p	MIMAT0004692	ITPR1	3708	23592263
hsa-miR-92a-3p	MIMAT0000092	ITPR1	3708	20371350
hsa-miR-92b-3p	MIMAT0003218	ITPR1	3708	20371350
hsa-miR-186-5p	MIMAT0000456	ITSN2	50618	23622248
hsa-miR-27b-3p	MIMAT0000419	ITSN2	50618	23592263
hsa-miR-16-5p	MIMAT0000069	KATNAL1	84056	20371350
hsa-miR-186-5p	MIMAT0000456	KATNAL1	84056	20371350
hsa-miR-19b-3p	MIMAT0000074	KATNAL1	84056	20371350
hsa-miR-20a-5p	MIMAT0000075	KATNAL1	84056	22473208
hsa-miR-340-5p	MIMAT0004692	KIAA1109	84162	21572407
hsa-miR-92a-3p	MIMAT0000092	KIAA1109	84162	22473208
hsa-miR-92b-3p	MIMAT0003218	KIAA1109	84162	22473208
hsa-miR-10a-5p	MIMAT0000253	KLHL6	89857	23622248
hsa-miR-27b-3p	MIMAT0000419	LCOR	84458	23622248
hsa-miR-30b-5p	MIMAT0000420	LCOR	84458	21572407
hsa-miR-30e-5p	MIMAT0000692	LCOR	84458	21572407
hsa-miR-92a-3p	MIMAT0000092	LCOR	84458	22012620
hsa-miR-92b-3p	MIMAT0003218	LCOR	84458	22012620
hsa-miR-24-3p	MIMAT0000080	LDHA	3939	23824327

Appendix 9: Table 1 *(Continued)*

miRNA	MIMATid	Gene	EntrezID	PMID
hsa-miR-331-3p	MIMAT0000760	LDHA	3939	23622248
hsa-miR-30b-5p	MIMAT0000420	LHFPL2	10184	22473208
hsa-miR-30e-5p	MIMAT0000692	LHFPL2	10184	22473208
hsa-miR-92a-3p	MIMAT0000092	LHFPL2	10184	20371350
hsa-miR-92b-3p	MIMAT0003218	LHFPL2	10184	20371350
hsa-miR-148a-3p	MIMAT0000243	LNPEP	4012	20371350
hsa-miR-27b-3p	MIMAT0000419	LPIN1	23175	22815788
hsa-miR-92a-3p	MIMAT0000092	LPIN1	23175	23622248
hsa-miR-92a-3p	MIMAT0000092	LRRC37A2	474170	23622248
hsa-miR-16-5p	MIMAT0000069	MACF1	23499	18668040
hsa-miR-19b-3p	MIMAT0000074	MACF1	23499	22473208
hsa-miR-92a-3p	MIMAT0000092	MACF1	23499	23622248
hsa-miR-182-5p	MIMAT0000259	MALT1	10892	22012620
hsa-miR-19b-3p	MIMAT0000074	MALT1	10892	22012620
hsa-miR-26a-5p	MIMAT0000082	MAN2A1	4124	22473208
hsa-miR-532-5p	MIMAT0002888	MAN2A1	4124	21572407
hsa-miR-92a-3p	MIMAT0000092	MAN2A1	4124	22473208
hsa-miR-92b-3p	MIMAT0003218	MAN2A1	4124	22473208
hsa-miR-92a-3p	MIMAT0000092	MEF2C	4208	23622248
hsa-miR-320a	MIMAT0000510	MIF	4282	23622248
hsa-miR-92a-3p	MIMAT0000092	MTMR14	64419	23622248
hsa-let-7a-5p	MIMAT0000062	MYC	4609	20033209
hsa-let-7b-5p	MIMAT0000063	MYC	4609	23622248
hsa-let-7d-5p	MIMAT0000065	MYC	4609	24510096
hsa-let-7e-5p	MIMAT0000066	MYC	4609	24510096
hsa-let-7f-5p	MIMAT0000067	MYC	4609	19956384
hsa-let-7g-5p	MIMAT0000414	MYC	4609	20309945
hsa-miR-222-3p	MIMAT0000279	MYC	4609	24510096

Appendix 9: Table 1 *(Continued)*

miRNA	MIMATid	Gene	EntrezID	PMID
hsa-miR-24-3p	MIMAT0000080	MYC	4609	19748357
hsa-miR-30e-3p	MIMAT0000693	MYC	4609	20371350
hsa-miR-30e-3p	MIMAT0000693	MYC	4609	21572407
hsa-miR-320a	MIMAT0000510	MYC	4609	23622248
hsa-miR-320b	MIMAT0005792	MYC	4609	23622248
hsa-miR-423-5p	MIMAT0004748	MYC	4609	23622248
hsa-miR-30b-5p	MIMAT0000420	MYO1E	4643	20371350
hsa-miR-221-3p	MIMAT0000278	NCL	4691	23622248
hsa-miR-20a-5p	MIMAT0000075	NFAT5	10725	20371350
hsa-miR-30b-5p	MIMAT0000420	NFAT5	10725	22473208
hsa-miR-30e-5p	MIMAT0000692	NFAT5	10725	22473208
hsa-miR-27b-3p	MIMAT0000419	NLK	51701	23592263
hsa-miR-92b-3p	MIMAT0003218	NLK	51701	23416699
hsa-miR-320a	MIMAT0000510	NPM3	10360	23446348
hsa-miR-320b	MIMAT0005792	NPM3	10360	23446348
hsa-let-7b-5p	MIMAT0000063	NT5DC2	64943	23622248
hsa-let-7e-5p	MIMAT0000066	NT5DC2	64943	23622248
hsa-miR-221-3p	MIMAT0000278	NT5DC2	64943	23622248
hsa-miR-30b-5p	MIMAT0000420	OPHN1	4983	20371350
hsa-miR-30e-5p	MIMAT0000692	OPHN1	4983	20371350
hsa-miR-16-5p	MIMAT0000069	ORC4	5000	20371350
hsa-miR-16-5p	MIMAT0000069	OXNAD1	92106	18668040
hsa-let-7b-5p	MIMAT0000063	PABPC1	26986	23622248
hsa-miR-423-5p	MIMAT0004748	PABPC1	26986	23622248
hsa-miR-16-5p	MIMAT0000069	PAG1	55824	20371350
hsa-miR-423-5p	MIMAT0004748	PCBP1	5093	23622248
hsa-let-7b-5p	MIMAT0000063	PCBP2	5094	23622248
hsa-let-7e-5p	MIMAT0000066	PCBP2	5094	23622248
hsa-miR-222-3p	MIMAT0000279	PCBP2	5094	23622248

Appendix 9: Table 1 *(Continued)*

miRNA	MIMATid	Gene	EntrezID	PMID
hsa-miR-320a	MIMAT0000510	PCBP2	5094	23622248
hsa-miR-16-5p	MIMAT0000069	PHC3	80012	20371350
hsa-miR-186-5p	MIMAT0000456	PHC3	80012	22012620
hsa-miR-20a-5p	MIMAT0000075	PHF6	84295	22473208
hsa-miR-10a-5p	MIMAT0000253	PIK3CG	5294	24522205
hsa-miR-26a-5p	MIMAT0000082	PIKFYVE	200576	23622248
hsa-miR-130a-3p	MIMAT0000425	RAB11FIP1	80223	23824327
hsa-miR-20a-5p	MIMAT0000075	RAB11FIP1	80223	22473208
hsa-miR-19b-3p	MIMAT0000074	RAB8B	51762	22473208
hsa-miR-92a-3p	MIMAT0000092	RAB8B	51762	22473208
hsa-miR-92b-3p	MIMAT0003218	RAB8B	51762	22473208
hsa-miR-20a-5p	MIMAT0000075	RABGAP1L	9910	23313552
hsa-miR-16-5p	MIMAT0000069	REL	5966	20371350
hsa-miR-92a-3p	MIMAT0000092	REL	5966	21572407
hsa-miR-92b-3p	MIMAT0003218	REL	5966	21572407
hsa-miR-16-5p	MIMAT0000069	RELT	84957	22473208
hsa-miR-183-5p	MIMAT0000261	RPAP2	79871	23622248
hsa-miR-130a-3p	MIMAT0000425	SAMD8	142891	21572407
hsa-miR-19b-3p	MIMAT0000074	SAMD8	142891	20371350
hsa-miR-20a-5p	MIMAT0000075	SAMD8	142891	21572407
hsa-miR-10a-5p	MIMAT0000253	SFT2D2	375035	23824327
hsa-miR-186-5p	MIMAT0000456	SFT2D2	375035	21572407
hsa-miR-16-5p	MIMAT0000069	SLC12A2	6558	18668040
hsa-miR-20a-5p	MIMAT0000075	SLC35F5	80255	22473208
hsa-miR-183-5p	MIMAT0000261	SMG1	23049	21572407
hsa-miR-183-5p	MIMAT0000261	SMG1	23049	23592263
hsa-miR-19b-3p	MIMAT0000074	SMG1	23049	20371350
hsa-miR-92a-3p	MIMAT0000092	SMG1	23049	22473208
hsa-miR-92b-3p	MIMAT0003218	SMG1	23049	23622248

Appendix 9: Table 1 *(Continued)*

miRNA	MIMATid	Gene	EntrezID	PMID
hsa-miR-16-5p	MIMAT0000069	SNX11	29916	22473208
hsa-miR-20a-5p	MIMAT0000075	SOD2	6648	22012620
hsa-miR-340-5p	MIMAT0004692	SOD2	6648	22012620
hsa-miR-423-5p	MIMAT0004748	SOX12	6666	23592263
hsa-miR-320a	MIMAT0000510	SRM	6723	23622248
hsa-miR-423-5p	MIMAT0004748	SRM	6723	20371350
hsa-miR-221-3p	MIMAT0000278	SRP68	6730	23622248
hsa-miR-320a	MIMAT0000510	SRP68	6730	23622248
hsa-miR-320b	MIMAT0005792	SRP68	6730	23622248
hsa-miR-20a-5p	MIMAT0000075	STAT3	6774	23059786
hsa-miR-92a-3p	MIMAT0000092	STAT3	6774	23622248
hsa-miR-340-5p	MIMAT0004692	SULT1B1	27284	22012620
hsa-miR-92a-3p	MIMAT0000092	TACC1	6867	22473208
hsa-miR-92b-3p	MIMAT0003218	TACC1	6867	22473208
hsa-miR-30e-5p	MIMAT0000692	TBC1D8B	54885	23622248
hsa-miR-16-5p	MIMAT0000069	TEP1	7011	23622248
hsa-miR-1260a	MIMAT0005911	TFDP2	7029	19536157
hsa-miR-1260b	MIMAT0015041	TFDP2	7029	19536157
hsa-miR-24-3p	MIMAT0000080	TFDP2	7029	23824327
hsa-miR-16-5p	MIMAT0000069	TLE4	7091	20371350
hsa-miR-186-5p	MIMAT0000456	TLE4	7091	20371350
hsa-miR-183-5p	MIMAT0000261	TMED8	283578	23824327
hsa-miR-24-3p	MIMAT0000080	TOP1	7150	19748357
hsa-miR-769-5p	MIMAT0003886	TRAF1	7185	23824327
hsa-miR-222-3p	MIMAT0000279	TRAP1	10131	23622248
hsa-miR-331-3p	MIMAT0000760	TRAP1	10131	23622248
hsa-miR-10a-5p	MIMAT0000253	TSPAN33	340348	23622248
hsa-let-7b-5p	MIMAT0000063	TTLL12	23170	18668040
hsa-let-7e-5p	MIMAT0000066	TTLL12	23170	23622248

Appendix 9: Table 1 *(Continued)*

miRNA	MIMATid	Gene	EntrezID	PMID
hsa-miR-140-3p	MIMAT0004597	TTLL12	23170	23592263
hsa-miR-222-3p	MIMAT0000279	TXN	7295	23622248
hsa-let-7b-5p	MIMAT0000063	UCK2	7371	18668040
hsa-miR-140-3p	MIMAT0004597	UCK2	7371	23592263
hsa-miR-24-3p	MIMAT0000080	UCK2	7371	23592263
hsa-miR-20a-5p	MIMAT0000075	UGCG	7357	23592263
hsa-miR-27b-3p	MIMAT0000419	UGCG	7357	23592263
hsa-miR-27b-3p	MIMAT0000419	UGCG	7357	23824327
hsa-miR-660-5p	MIMAT0003338	UGCG	7357	23592263
hsa-miR-186-5p	MIMAT0000456	VPS13B	157680	23622248
hsa-miR-30e-5p	MIMAT0000692	VPS13B	157680	23622248
hsa-miR-186-5p	MIMAT0000456	VPS13C	54832	23622248
hsa-miR-20a-5p	MIMAT0000075	VPS13C	54832	22473208
hsa-miR-10a-5p	MIMAT0000253	WDR74	54663	23622248
hsa-miR-20a-5p	MIMAT0000075	ZBTB37	84614	20371350
hsa-miR-92a-3p	MIMAT0000092	ZNF277	11179	20371350
hsa-miR-92b-3p	MIMAT0003218	ZNF277	11179	20371350
hsa-miR-20a-5p	MIMAT0000075	ZNF780A	284323	23446348

Abbreviation & Definitions: MIMATid, Accession number for the mature miRNA; PMID, PubMed ID miRNA-mRNA interaction; DE, differentially expressed.

Upregulated

Downregulated

PUBLICATIONS

Manuscript #1- Human and Epstein-Barr virus miRNA profiling as predictive biomarkers for Endemic Burkitt lymphoma. *Published 28th March 2017*

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 March 2017
doi: 10.3389/fmicb.2017.00501

Human and Epstein-Barr Virus miRNA Profiling as Predictive Biomarkers for Endemic Burkitt Lymphoma

*Cliff I. Oduor[1,2†], Mercedeh Movassagh[3†], Yasin Kaymaz[3], Kiprotich Chelimo[2], Juliana Otieno[4], John M. Ong'echa[1], Ann M. Moormann[5†] and Jeffrey A. Bailey[3,6†]**

[1] Center for Global Health Research, Kenya Medical Research Institute, Kisumu, Kenya, [2] Department of Biomedical Sciences and Technology, Maseno University, Maseno, Kenya, [3] Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA, USA, [4] Jaramogi Oginga Odinga Teaching and Referral Hospital, Ministry of Medical Services, Kisumu, Kenya, [5] Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA, USA, [6] Division of Transfusion Medicine, Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA

OPEN ACCESS

Edited by:
Sam Mbulaiteye,
National Cancer Institute, USA

Reviewed by:
Rebecca Skalsky,
Oregon Health & Science University, USA
Erle S. Robertson,
University of Pennsylvania, USA

***Correspondence:**
Jeffrey A. Bailey
jeffrey.bailey@umassmed.edu

†These authors have contributed equally to this work.

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: 12 November 2016
Accepted: 10 March 2017
Published: 28 March 2017

Citation:
Oduor CI, Movassagh M, Kaymaz Y, Chelimo K, Otieno J, Ong'echa JM, Moormann AM and Bailey JA (2017) Human and Epstein-Barr Virus miRNA Profiling as Predictive Biomarkers for Endemic Burkitt Lymphoma. Front. Microbiol. 8:501. doi: 10.3389/fmicb.2017.00501

Endemic Burkitt lymphoma (eBL) is an aggressive B cell lymphoma and is associated with Epstein-Barr virus (EBV) and *Plasmodium falciparum* malaria co-infections. Central to BL oncogenesis is the over-expression of the MYC proto-oncogene which is caused by a translocation of an Ig enhancer in approximation to the myc gene. While whole genome/transcriptome sequencing methods have been used to define driver mutations and transcriptional dysregulation, microRNA (miRNA) dysregulation and differential expression has yet to be fully characterized. We hypothesized that both human and EBV miRNAs contribute to eBL clinical presentation, disease progression, and poor outcomes. Using sensitive and precise deep sequencing, we identified miRNAs from 17 Kenyan eBL patient tumor samples and delineated the complement of both host and EBV miRNAs. One human miRNA, hsa-miR-10a-5p was found to be differentially expressed (DE), being down-regulated in jaw tumors relative to abdominal and in non-survivors compared to survivors. We also examined EBV miRNAs, which made up 2.7% of the miRNA composition in the eBL samples. However, we did not find any significant associations regarding initial patient outcome or anatomical presentation. Gene ontology analysis and pathway enrichment of previously validated targets of miR-10a-5p suggest that it can promote tumor cell survival as well as aid in evasion of apoptosis. To examine miR-10a-5p regulatory effect on gene expression in eBL, we performed a pairwise correlation coefficient analysis on the expression levels of all its validated targets. We found a significant enrichment of correlated target genes consistent with miR-10a-5p impacting expression. The functions of genes and their correlation fit with multiple target genes impacting tumor resilience. The observed downregulation of miR-10a and associated genes suggests a role for miRNA in eBL patient outcomes and has potential as a predictive biomarker that warrants further investigation.

Keywords: endemic Burkitt lymphoma, microRNA expression, EBV, RNAseq, miR-10a-5p

Manuscript #2- Integrative microRNA and mRNA deep-sequencing expression profiling in endemic Burkitt Lymphoma- *Published 13 November 2017*

Oduor *et al. BMC Cancer* (2017) 17:761
DOI 10.1186/s12885-017-3711-9

BMC Cancer

RESEARCH ARTICLE Open Access

Integrative microRNA and mRNA deep-sequencing expression profiling in endemic Burkitt lymphoma

Cliff I. Oduor[1,2], Yasin Kaymaz[3], Kiprotich Chelimo[2], Juliana A. Otieno[4], John Michael Ong'echa[1], Ann M. Moormann[5] and Jeffrey A. Bailey[3,6*]

Abstract

Background: Burkitt lymphoma (BL) is characterized by overexpression of the *c-myc* oncogene, which in the vast majority of cases is a consequence of an IGH/MYC translocation. While *myc* is the seminal event, BL is a complex amalgam of genetic and epigenetic changes causing dysregulation of both coding and non-coding transcripts. Emerging evidence suggest that abnormal modulation of mRNA transcription via miRNAs might be a significant factor in lymphomagenesis. However, the alterations in these miRNAs and their correlations to their putative mRNA targets have not been extensively studied relative to normal germinal center (GC) B cells.

Methods: Using more sensitive and specific transcriptome deep sequencing, we compared previously published small miRNA and long mRNA of a set of GC B cells and eBL tumors. MiRWalk2.0 was used to identify the validated target genes for the deregulated miRNAs, which would be important for understanding the regulatory networks associated with eBL development.

Results: We found 211 differentially expressed (DE) genes (79 upregulated and 132 downregulated) and 49 DE miRNAs (22 up-regulated and 27 down-regulated). Gene Set enrichment analysis identified the enrichment of a set of MYC regulated genes. Network propagation-based method and correlated miRNA-mRNA expression analysis identified dysregulated miRNAs, including miR-17~95 cluster members and their target genes, which have diverse oncogenic properties to be critical to eBL lymphomagenesis. Central to all these findings, we observed the downregulation of *ATM* and *NLK* genes, which represent important regulators in response to DNA damage in eBL tumor cells. These tumor suppressors were targeted by multiple upregulated miRNAs (miR-19b-3p, miR-26a-5p, miR-30b-5p, miR-92a-5p and miR-27b-3p) which could account for their aberrant expression in eBL.

Conclusion: Combined loss of p53 induction and function due to miRNA-mediated regulation of *ATM* and *NLK*, together with the upregulation of *TFAP4*, may be a central role for human miRNAs in eBL oncogenesis. This facilitates survival of eBL tumor cells with the IGH/MYC chromosomal translocation and promotes MYC-induced cell cycle progression, initiating eBL lymphomagenesis. This characterization of miRNA-mRNA interactions in eBL relative to GC B cells provides new insights on miRNA-mediated transcript regulation in eBL, which are potentially useful for new improved therapeutic strategies.

Keywords: Endemic Burkitt lymphoma, miRNA, mRNA, RNA sequencing, Lymphomagenesis

* Correspondence: jeffrey.bailey@umassmed.edu
[3]Department of Bioinformatics & Integrative Biology, University of Massachusetts Medical School, Worcester, MA, USA
[6]Division of Transfusion Medicine, Department of Medicine, University of Massachusetts Medical School, 368 Plantation St. Albert Sherman Building 41077, Worcester, MA 01605, USA
Full list of author information is available at the end of the article

ETHICAL APPROVALS

KENYA MEDICAL RESEARCH INSTITUTE

P.O. Box 54840 - 00200 NAIROBI, Kenya
Tel: (254) (020) 2722541, 2713349, 0722-205901, 0733-400003; Fax: (254) (020) 2720030
E-mail: director@kemri.org; info@kemri.org Website: www.kemri.org

KEMRI/RES/7/3/1 **MAY 23, 2008**

FROM: **SECRETARY, KEMRI/National Ethical Review Committee**

THRO': **Dr. J Vulule,**
CENTRE DIRECTOR, CGHR,
KISUMU

FORWARDED
DIRECTOR
27/5/08

TO: **Dr. Ann M Moormann (Case Western Reserve University) (Principal Investigator)**

RE: **SSC No.1381** – The effect of Plasmodium falciparum malaria on T cell immunity and Endemic Burkitts lymphona

Dear Madam,

This is to inform you that the abovementioned protocol has undergone expedited review.

We acknowledge receipt of the following documents:

1. The study proposal version 10 March 2008
2. The Informed Consent Document (ICD) form 1 for healthy Kenyan children in English
3. Form 1.1: Healthy child enrollment questionnaire and venous blood sample collection
4. The ICD form 2: for children diagnosed with Burkitts lymphoma
5. Form 2.1 BL patient enrollment questionnaire and venous blood sample collection
6. The ICD form 3: for children who have been diagnosed with BL but are in remission for a repeat venous blood sample
7. Form2.2 BL patient discharge summary and venous blood sample collection
8. Form 2.3: BL patient follow-up and venous blood sample collection
9. The ICD form 4: for healthy Kenyan adults to optimize laboratory assays
10. Form 3.1 Healthy adult enrollment questionnaire and venous blood sample collection
11. The ICD form 5: for healthy US adults never exposed to malaria
12. Form 4.1 Healthy US adult enrollment questionnaire and venous blood collection

Thank you for your informative study proposal that aims to investigate the mechanisms of malaria-induced dysregulation of EBV-specific T cell immunity and its relationship to eBL. This will by a prospective study of healthy Kenyan children with divergent malaria exposure histories and by examining children with eBL compared to healthy Kenyan and US adults who have robust immune responses to EBV.

KENYA MEDICAL RESEARCH INSTITUTE

P.O. Box 54840-00200, NAIROBI, Kenya
Tel (254) (020) 2722541, 2713349, 0722-205901, 0733-400003; Fax: (254) (020) 2720030
E-mail: director@kemri.org info@kemri.org Website:www.kemri.org

21 OCT 2014

KEMRI/RES/7/3/1

October 13, 2014

FORWARDED
DIRECTOR
CENTRE FOR GLOBAL HEALTH RESEARCH

TO: **DR. JOHN ONG'ECHA,**
PRINCIPAL INVESTIGATOR

THROUGH: **DR. STEPHEN MUNGA**
THE DIRECTOR, CGHR
KISUMU

Dear Sir,

RE: **SSC PROTOCOL NO. 2844(*INITIAL SUBMISSION*): IMPACT OF MALARIA ON SHAPING IMMUNITY TO EBV AND ENDEMIC BURKITT LYMPHOMA**

This is to inform you that during the 231st meeting of the KEMRI/ERC held on September 16, 2014, the above study was reviewed.

The Committee noted that the above referenced study aims to determine how malaria infection influences the differentiation and survival of EBV-specific T cell responses and loss of t-cell control over EBV.

Due consideration has been given to ethical issues, this study is therefore granted approval for implementation effective this day **October 13, 2014**. Please note that authorization to conduct this study will automatically expire on **October 12, 2015**.

If you plan to continue with data collection or analysis beyond this date please submit an application for continuing approval to the ERC secretariat by **August 31, 2015**. You are also required to submit any proposed changes to this protocol to the SSC and ERC prior to initiation and advise the ERC when the study is completed or discontinued.

You may embark on the study.

Yours faithfully,

PROF. ELIZABETH BUKUSI,
ACTING SECRETARY,
KEMRI/ETHICS REVIEW COMMITTEE

CONSENT FORM

KEMRI/UMMS
CONSENT FOR INVESTIGATIONAL STUDIES
Project Title: Impact of Malaria on Shaping Immunity to EBV and Endemic Burkitt Lymphoma
Principal Investigators: Ann Moormann, Ph.D., MPH and John M. Ong'echa, Ph.D.
Consent 4: BL biopsy for research request

INVESTIGATOR'S STATEMENT
Dr. Ann Moormann from the University of Massachusetts Medical School (UMMS) in the USA, Dr. John M. Ong'echa at the Kenya Medical Research Institute (KEMRI), and Dr. Juliana Otieno at Jaramogi Oginga Odinga Teaching and Referral Hospital (JOOTRH), Ministry of Health, Kenya are doing a research study on Burkitt lymphoma (BL). We are doing this study because the causes of BL not fully understood. You are being asked to participate in this study because your child may have BL. As part of medical care at JOOTRH, when a patient is thought to have this cancer, a biopsy is used to confirm the diagnosis. A biopsy is a procedure used to remove a small piece (about the size of the print made by your littlest finger) of the tumor from the body so that it can be looked at under a microscope. We are asking for another small piece of your child's tumor to be used in a research study. The purpose of this consent form is to give you the information you will need to help you decide whether a tumor biopsy from your child can be included in this study or not. Please read the form carefully. You may ask questions about the purpose of the research, what we would ask you and your child to do, the possible risks and benefits, your rights and your child's rights as a volunteer, and anything else about the study or this form that is not clear. When we have answered all your questions, you can decide if you want to sign this consent form or not. This process is called "informed consent." We will give you a copy of this form for your records.

PURPOSE OF THE STUDY
The purpose of this study is to learn more about Burkitt lymphoma (BL). You are being asked to sign this consent form because your child has been admitted to JOOTRH in Kisumu, Kenya and might have BL. At the end of this study, we hope to find ways to prevent BL and ways to treat BL tumors that may not be cured using the standard medicines.

NUMBER OF PARTICIPANTS
Tumor biopsies will be requested from patients suspected of having Burkitt lymphoma until we have 80 specimens collected and confirmed as BL. Biopsies from both jaw and abdominal tumors will be collected.

STUDY DURATION
It is expected to take about 30 minutes for you to read this consent form and have time to ask any questions before signing the consent form.

A biopsy will not be done for research purposes only. The biopsy will be requested by the doctor taking care of your child to confirm the diagnosis. When the biopsy procedure is being done, the doctor will take another small piece and put it into a tube for research. This will add less than 5 additional minutes to the procedure.

STUDY PROCEDURES
Since taking a tumor biopsy is part of medical care at JOOTRH, the procedure will be explained to you by the nurse and doctors taking care of your child. The procedure used depends on where the tumor is on the body. This study does not change the hospital procedure in any way. After the biopsy, if you have any concerns about the biopsy site and how it is healing, please contact one of the doctors for an evaluation or nurse on the ward caring for your child.

POSSIBLE RISKS OF STUDY PARTICIPATION
There are no added risks to your child if you consent for a biopsy to be used for research.

However, normal side effects of the biopsy procedure may include pain (stinging and burning), bleeding and infection. The biopsy areas will be covered with a bandage and may be tender for several days. It should heal within one or two weeks.

ALTERNATIVES TO TAKING PART IN THIS RESEARCH STUDY
You do not have to sign this consent form. You do not have to participate in this study.
Your child will still get regular care at JOOTRH, even if you decide not to participate in this study.

POSSIBLE BENEFITS OF STUDY PARTICIPATION
Your child will not benefit directly from results of this study. The diagnosis and treatment will be decided by the doctors at JOOTRH caring for your child. This study will in no way interfere with the care of your child. It is possible that what we learn from this study will help doctors and scientists learn how to prevent BL in other children and to improve our ability to cure this cancer.

SUMMARY OF YOUR RIGHTS AS A PARTICIPANT IN A RESEARCH STUDY
Your participation in this research study is voluntary. Refusing to participate will not alter your usual health care or involve any penalty or loss of benefits to which you are otherwise entitled. If you decide to join the study, you may withdraw at any time and for any reason without penalty or loss of benefits. If information generated from this study is published or presented, your identity or your child's identity will not be revealed. In the event new information becomes available that may affect the risks or benefits associated with this study or your willingness to participate in it, you will be notified so that you can decide whether or not to continue participating.

CONFIDENTIALITY OF RECORDS
Your identity, your child's identity and his or her records will remain confidential. The biopsy specimen will be identified by a code number. Any information about your child to be used in this study will be marked by this code number. If you agree to participate in this study we will ask for some basic information to be recorded about your child. This information will include your child's age, sex, site of tumor(s). During the course of your child's illness, we would also like permission to know if your child is responding well to treatment and which treatment regimen was used. This will be determined by Dr. Otieno or another doctor at JOOTRH caring for your child. This information can be found in your child's medical records. Signing this consent form gives the study staff permission to look at your child's chart and record this information for the purpose of the study. This information will be helpful to learn more about BL.

The study code number assigned to your child will be linked to your child's hospital record number. The log book that will match the code number with your child's hospital number will be kept in a locked file cabinet at Center for Global Health Research, KEMRI and on a password protected file. Links between the study code and information that could identify your child will be

kept for five years unless you provide consent that they may be kept indefinitely. Research data is kept indefinitely but will only be labeled by the study code number.

We will do everything we can to protect your privacy. We cannot guarantee absolute privacy, however. Your personal information may be disclosed if required by law. If your confidential information is released, you will be told. Some people may need to review the study records. They may include the UMMS or KEMRI Ethical Review Boards and the study staff from UMMS-KEMRI. The results of this study will be published. Any publication will not use your name, your child's name or identify you or your child personally.

SPECIMEN STORAGE
Specimens collected from your child will be tested by laboratories associated with the University of Massachusetts, USA. It is anticipated that the entire specimen will be used for the studies proposed by this protocol. If you consent, any remaining specimens will be kept indefinitely. Any future testing of your child's specimens will not be done unless written approval has been given from the ethical review committees in Kenya and UMMS. If you circle "yes' on this consent form below, you are allowing your child's specimen to be used for future research, if it is related to cancer, without seeking re-approval from you individually.

If you change your mind in the future, you may contact Dr. John Ong'echa (0733-447920) or the KEMRI/National Ethical Review Committee (ERC), PO Box 54840, Nairobi 00200 at (020) 272-2541 or the Director of KEMRI, PO Box 54840, Nairobi at (020)272-2541.

Consent for use of child's samples for future studies yes no
(Please circle parent's response)

REASONS FOR REMOVING YOUR CHILD FROM THE STUDY WITHOUT YOUR CONSENT
If the doctor cancels the request for a biopsy then your child will no longer be eligible to be in this study. This may occur if your child's doctor receives other information that means your child does not have BL.

COSTS TO YOU
There is no cost to you for being in this study. There are no added tests or procedures being done only for research purposes. You will be asked to pay for medical procedures requested by the doctor in order to find out why your child is ill, the same as for children not in the study. All other medical costs outside of this study will be paid by you or your health insurance carrier (if you have insurance).

IF YOU THINK YOUR CHILD HAS A RESEARCH-RELATED INJURY
There are no procedures being done for study purposes only. All procedures are for the care and diagnosis of your child. However, you may call Dr. Otieno with any question you may have regarding the study or your child's participation in this study.

No funds are available to provide compensation for non-physical injury such as lost work or pain and suffering. You and/or your health insurance carrier will continue to be responsible for costs for your child's medical care or for medical expenses determined not directly related to study procedures. You will not be giving up any of your legal rights by signing this consent form.

IF YOU SIGN THIS CONSENT FORM, YOU DO NOT NEED TO DO ANYTHING ELSE.
Once you sign this consent form, the study coordinator will show the nurse-in-charge and the doctor taking care of your child that your child is enrolled in this study. This will allow the nurse to tell the study coordinator when the biopsy procedure is scheduled. Then the surgeon will be notified that during the procedure, a small piece of the tumor can be put in a tube for research purposes. Project staff will be there during the procedure to collect the specimen and will bring it to the UMMS-KEMRI lab.

PAYMENT
You will not be paid to participate in this research study. However, in appreciation for your participation the project can pay for transport home when your child is discharged from hospital. We will also assist with transport reimbursement for out-patient clinic visits to make sure your child is recovering well. If your child is experiencing symptoms that indicate that the cancer may be coming back, we will transport you and your child to JOOTRH for evaluation. Please contact Mrs. Pamella Omolo for assistance with at 0722890318.

Contact information

One of our team members named _____ has described to you what is going to be done, the risks, hazards, and benefits involved. Further information with respect to illness or injury resulting from a research procedure as well as a research subjects' rights is available from KEMRI/National Ethical Review Committee (ERC), PO Box 54840, Nairobi 00200 at (020) 272-2541 or the director of KEMRI, PO Box 54840, Nairobi at (020)272-2541, or the Coordinator for the Committee for the Protection of Human Subjects in Research in the United States at (508) 856-4261 or write to Committee for the Protection of Human Subjects in Research, University of Massachusetts Medical Center, 55 Lake Avenue North, Worcester, MA 01655 U.S.A.

For study coordinator:

Study assigned unique identification number: BL - __ __ __ __
(number given at time of enrollment on Form 2.1)

SIGNATURE PAGE

_____
Study staff conducting consent discussion (print name)

_____ _____
Study staff signature **Date (dd/mm/yyyy)**

PARTICPANT'S STATEMENT
The study described above has been explained to me. I agree to volunteer to participate in the study. I have had a chance to ask questions. I have been told that if I have future questions about the research I can ask one of the contacts listed above or the study staff named above. I give permission to the researchers to use my child's medical records as described in this consent form and to collect a biopsy specimen for research purposes. I will receive a copy of this consent form. If you have read this consent form (or had it explained to you), understand it and agree for your child to take part in this study, please sign your name below.

_____
Child's name (print name)

Study Code number: UMMS-KEMRI BLB-00__ __
(Study code assigned in order of enrollment, first number is UMMS-KEMRI BLB-0001, etc).

_____
Parent or legal guardian's name (print name)

_____ _____
Parent or legal guardian's signature **Date (dd/mm/yyyy)**
Or mark (right thumb unless otherwise indicated)

_____
Cell phone contact of parent or nearest neighbor/relative

_____
Witness's name, if necessary (print – this name should be a different UCI staff member than the person who conducted the consent discussion)

_____ _____
Witness's signature **Date (dd/mm/yyyy)**

Original to: Participant's study file.
Copy to: Participant. If participant declines a copy of this CF, then check this box and initial: [ ]