UNIVERSITY OF HULL

THE UNIVERSITY OF HULL

Identification of the molecular pathways involved in paediatric and adult acute myeloid leukaemia

being a Thesis submitted for the Degree of Masters by Research

at the University of Hull

by

Elliott Brown, BSc (Hons) Biomedical Science

April 2022

Abstract

Acute myeloid leukaemia (AML) affects 3100 people every year in the UK, around 1% of whom are under 19 years of age. Despite improvements in immunotherapeutic treatment opportunities for AML, the rates of survival particularly in adults remain poor. We wanted to determine whether adult and paediatric AML, share common molecular pathways and therefore, the same therapies could be used to treat each.

Previous dataset analysis provided a group of differentially expressed genes in different risk subgroups of AML. Using a variety of analysis tools and a comprehensive literature review, the role of these antigens and microRNAs (miRs) were investigated in various diseases and specifically AML. Models were generated using miRs as biomarkers to predict prognosis grouping based on expression and qPCR was performed to evaluate their use as a biomarker in patient sera.

Nine leukaemia associated` antigens (LAAs) and four miRs were identified for further study due to their effects on patient prognosis. Investigation of the LAAs showed potential roles in cancerous drivers and leukaemogenesis, interacting with known important pathways. Following area under the curve analysis, mRNA-sequencing data from TCGA and TARGET demonstrated that combinations of miRs made powerful predictors of prognosis. qPCR showed no significant difference in expression levels of DEmiRs in sera samples.

BIRC5 may be a key target for specific subtypes of AML and the other LAAs have been found to interact with key pathways in other cancers. However, these often have little to no research in AML patients or cell lines. Furthermore, only miR-1915-5p showed promise as a serum biomarker, however, further investigation is needed following this study.

Acknowledgements

First, I would like to thank both of my supervisors, Dr Barbara Guinn and Dr Stefano Caserta for their continued assistance and support throughout the project. I would also like to thank the other students in the lab group, particularly Leak Cooksey, Kelly Grayson, Eithar Mohamed, and Ellie Beeby for their help in training me. As well as Danny Fletcher and Julliah Javadala for contributing to the lab work presented here. Samantha Girven's knowledge of R was appreciated in the creation of figures.

Next, I would like to thank Dr Pinar Onganer-Uysal and Maria Mortoglou from the University of Westminster who provided advice which was essential to the progression of this study on miRs. I would also like to thank Dr Kim Orchard from the University Hospital Southampton NHS Trust and Professor Ghulam Mufti, King's College Hospital NHS Foundation Trust for patient samples. Ms Lindsay Davis and Professor Ken Mills, Queen's University Belfast were instrumental in the original differential gene expression and correlation analysis that primed this work.

Finally, I am grateful for the opportunity offered by the University of Hull's internship scheme that led me to follow this degree and the continuous support provided by both family and friends.

Declaration

I confirm that this work is original and that if any passage(s) or diagram(s) have been copied from academic papers, books, the internet or any other sources these are clearly identified by the use of quotation marks and the reference(s) fully cited. I certify that, other than where indicated, this is my own work and does not breach the regulations of the University of Hull regarding plagiarism or academic misconduct in examinations. This piece of work is my own and does not contain any unacknowledged work from any other sources. I confirm that any patient information obtained to produce this piece of work was collected following informed consent and has been appropriately anonymised prior to my work commencing on it.

Publications to date

PEER REVIEWED PUBLICATIONS:

Fletcher, D., **Brown, E.,** Javadala, J., Uysal-Onganer, P. & Guinn, B.A. (2022) miRNA expression in acute myeloid leukaemia: new targets for therapy? In Press, *eJHaem*.

Greiner, J., **Brown, E.,** Bullinger, L., Morris, V., Döhner, H., Mills, K.I., & Guinn, B.A. (2021) Survivin' acute myeloid leukaemia – a personalised target for inv(16) patients. Part of a Special Issue titled 'Molecular Mechanisms and Therapies of Myeloid Leukaemia', *International Journal of Molecular Sciences,* 221, 10482.

CONFERENCE ABSTRACTS:

Brown, E., Mortoglou, M., Davis, L., Caserta, S., Orchard, K.H., Mills, K.I., Uysal-Onganer, P. & Guinn, B.A. (2021) miRNAs act as indicators of a poor risk prognosis in adults with acute myeloid leukaemia. Allam Lecture 2021.

Uysal-Onganer, P., Mortoglou, M., **Brown, E.**, Davis, L., Orchard, K.H., Caserta, S., Mills, K.I., & Guinn, B.A. (2021) MiRNA expression in adult acute myeloid leukaemia: biomarkers for risk subgroups? *European Hematology Association (EHA) 2021*.

Brown, E., Davis, L., Mills, K.I., Venney, D., Orchard, K.H. & Guinn, B.A. (2020) Investigation of BIRC5 as a therapeutic target in paediatric and adult acute myeloid leukaemia. *6th Symposium on Advances in Cancer Immunology and Immunotherapy*, Athens, Greece.

Contents

List of Tables

List of Figures

Abbreviations

AGO	Argonaute protein
AKT	Threonine Kinase 1
AML	Acute myeloid leukaemia
ALL	Acute lymphoblastic leukaemia
AUC	Area under the curve
BCAS2	Breast carcinoma-amplified sequence 2
BCL-2	B-cell lymphoma 2
BIRC5	Baculoviral IAP Repeat Containing 5
BM	Bone marrow
CAR-T	Cancer Antigen Receptor Targets
CBF	Core binding factor
CBL	E3 ubiquitin ligase casitas B-lineage lymphoma
CD	Cluster of Differentiation
CDK	Cyclin dependant kinases
cDNA	Complementary DNA
CEACAM3	Carcinoembryonic antigen related cell adhesion molecule 3
CIA	Cytosolic iron–sulphur cluster assembly
CLL	Chronic lymphoblastic leukaemia
C-Myc	MYC Proto-Oncogene
CML	Chronic Myeloid Leukaemia
CPC	Chromosomal passenger complex
Ct	Cycle threshold
DEGs	Differentially expressed genes
DEmiRs	Differentially expressed miRNAs
DGCR8	DiGeorge Syndrome Critical Region 8
DGE	Differential gene expression
D	At diagnosis
DMEM	Dulbecco's Modified Eagle's Medium
DUSP3	Dual specificity phosphatase 3
EDTA	Ethylenediaminetetraacetic acid
EFS	Event free survival
EID	E1A-like Inhibitor of differentiation
EV	Extracellular vesicles
FAB	French-American-British
FLT3	Fms-related tyrosine kinase 3
FOXO1	Forkhead box O1
FU	Follow up
GC	Gastric cancer
GDC	Genomic Data Commons
GO	Gene ontology
GOI	Genes of interest
GSK3	Glycogen synthesis kinase 3
HCC	Hepatocellular carcinoma
HSCT	Hematopoietic stem-cell transplantation
HV	Healthy volunteer
IAP	Inhibitors of apoptosis
ICC	Immunocytochemistry
IF	Immunofluorescence
IL	Interleukin
IRF2BP2	Interferon regulatory factor 2- binding protein 2
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KM	Kaplan-Meier
LAA	Leukaemia associated antigen
mAbs	Monoclonal antibodies
MAGEF1	MAGE Family Member F1
MAPK	Mitogen-activated protein kinase 1
MDS	Myelodysplastic syndrome
MELTF	Melanotransferrin
MILE	Microarray innovations in leukaemia
MORC4	MORC family CW-type zinc finger protein 4
MLL	Mixed-Lineage-Leukaemia
NCI	National cancer institute
NF-κB2	Nuclear factor kappa B2

NPM1	Nucleophosmin 1
NTC	No template control
NV	Normaliser value
OS	Overall survival
PAX2	Paired box 2
PBL	Peripheral blood lymphocytes
PBX2	Pre B cell leukaemia homeobox 2
PD-L1	Programmed death-ligand 1
PI3K	Phosphoinositide 3-kinase
PROM1	Prominin 1
PTEN	Phosphatase and tensin homolog
qPCR	quantitative polymerase chain reaction
RAGE	Receptor for advanced glycation end product
RBCs	Red blood cells
RISC	RNA-induced silencing complex
RNA-seq	RNA-sequencing
ROC	Receiver operator characteristic curve
ROS	Reactive oxygen species
RT	Reverse transcription
RTK	Receptor tyrosine kinases
SAGE1	Sarcoma antigen 1
SLC	Sex-linked chromosome
SLC34A2	Solute carrier family 34 member 2
SMAD2	SMAD family member 2
SOCS2	Suppressor of cytokine signalling 2
SS	Spermatocytic seminoma
STEAP1	STEAP Family Member 1
STRING	Search Tool for the Retrieval of Interacting Genes
SUFU	Suppressor of fused homolog
TARGET	Therapeutically Applicable Research to Generate Effective Treatments
TCF	Transcription Factor
TCGA	The Cancer Genome Atlas
TEAD	TEA DNA binding domain
TGF-β	Transforming growth factor beta
USC	University of Santa Cruz
VGLL4	Vestigial-like 4
WBCs	White blood cells
WHO	World health organisation
YAP	Yes-associated protein 1

Chapter 1. Introduction

1.1 Cancer

Cancer is described as the uncontrolled proliferation of cells and can develop throughout the body. Mutations in genes can cause cancer by inhibiting cell death or promoting accelerated division. There were 367,167 new cases of cancer in the UK between 2015 and 2017 (UK, 2018). Throughout the UK, leukaemia remains one of the most common cancers, with the twelfth highest incidence rate (**Figure 1.1**).

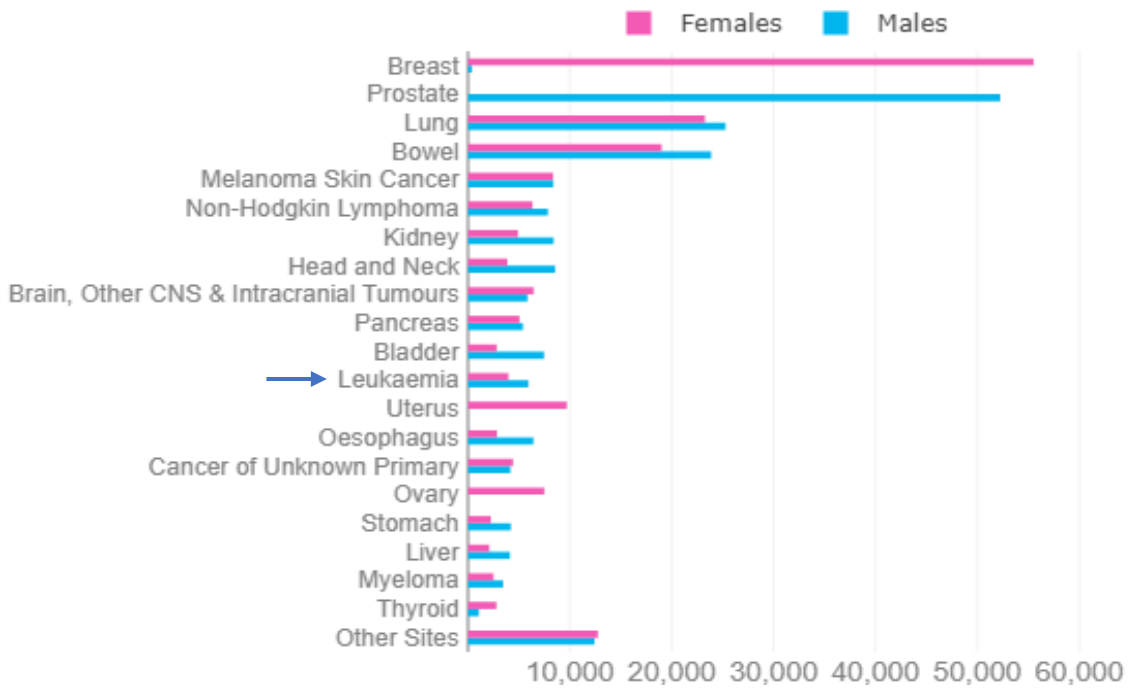


Figure 1.1 Cancer rates in the UK. Most common cancers in the UK between 2016-2018, a total of 9907 cases of leukaemia are diagnosed in this period. Figure from Cancer Research UK [Generated 01/03/22].

1.2 Leukaemia

The most common types of leukaemia are acute myeloid leukaemia (AML), chronic myeloid leukaemia (CML), acute lymphoblastic leukaemia (ALL), and chronic lymphoblastic leukaemia (CLL). Acute and chronic leukaemias differ in their progression, acute leukaemia rapidly progresses and produces an accumulation of immature blast cells in the bone marrow (BM) and blood, these blast cells build up in the BM and prevent the production of healthy blood cells. Chronic leukaemia follows the same path at a slower rate, allowing the cells to mature further along the lineage before the differentiation arrest and are slightly more functional, this often

means chronic leukaemia is diagnosed later in life and is almost exclusively found in older patients.

Myeloid leukaemias originate in myeloid progenitors affecting red blood cells (RBCs), platelets and leukocytes, whereas the lymphocytic/lymphoblastic leukaemias originate exclusively in the lymphoid stem cells that form white blood cells (WBCs) (**Figure 1.2**). AML is defined as a malignant disorder of the BM characterised by the clonal expansion and differentiation arrest of myeloid progenitor cells (Shallis et al., 2019). Mixed lineage leukaemias (MLL) can also occur, where the blasts show phenotypes of myeloid and lymphoid cells. MLL is most commonly found in paediatric patients with translocations on the MLL gene at 11q23 and can indicate worse survival (Muntean & Hess, 2012; Wolach & Stone, 2015). This worse survival is potentially caused by hyperleukocytosis and central nervous system involvement, along with resistance to common chemotherapeutic agents (Winters & Bernt, 2017).

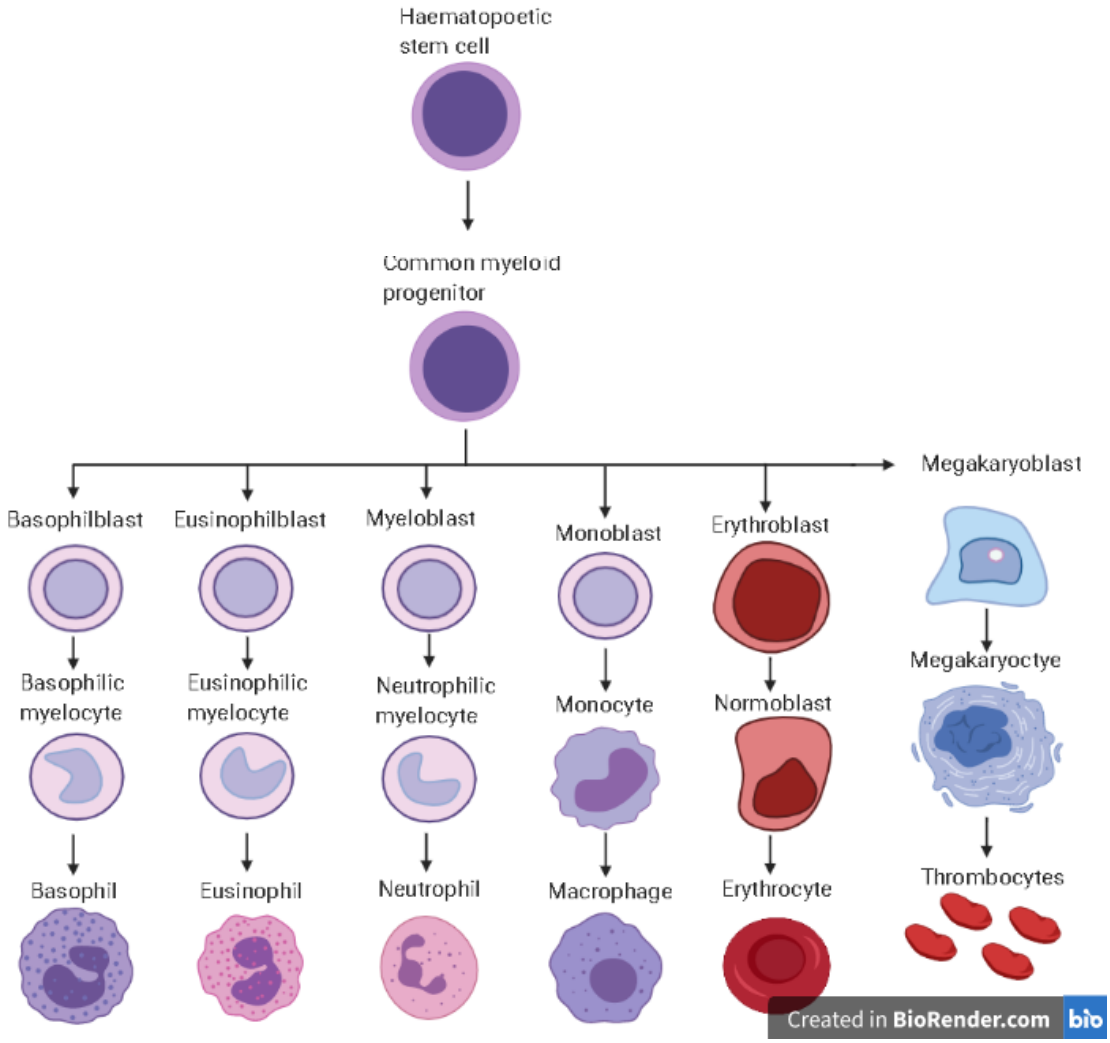


Figure 1.2 Image of erythropoiesis. A diagram showing the stages of haemopoiesis that can be interrupted by AML. Image author's own created using biorender program (https://biorender.com/).

1.2.1 Epidemiology

CLL is the most common form of leukaemia in adults, with roughly 3800 people being diagnosed with it in the UK each year between 2016-18 (UK, 2018). Adult AML has similar occurrence rates as CLL, affecting less than 1 in 10,000 people, and around 3100 adults are diagnosed with it each year. AML makes up around 15-20% of leukaemias in the paediatric population, and survival for paediatrics and young adults is around 70% (Creutzig et al., 2018). Roughly 100 patients each year are diagnosed with paediatric AML in the UK, whilst CML in paediatrics accounts for around five cases a year.

However, AML rates have been increasing over the last 40 years, and this may reflect an ageing population, survival in the face of other risks such as viruses, and exposure to more environmental factors such as greater levels of pollution. Population effects such as the post-war baby boom provided a large increase in the population reaching the age most associated with AML, which could account for some of the dramatic increase in the last 10 years (Hao et al., 2019). In the UK, AML is most often diagnosed between the ages of 75-84 (**Figure 1.3**), with only 1% of those affected being under 19 years of age.

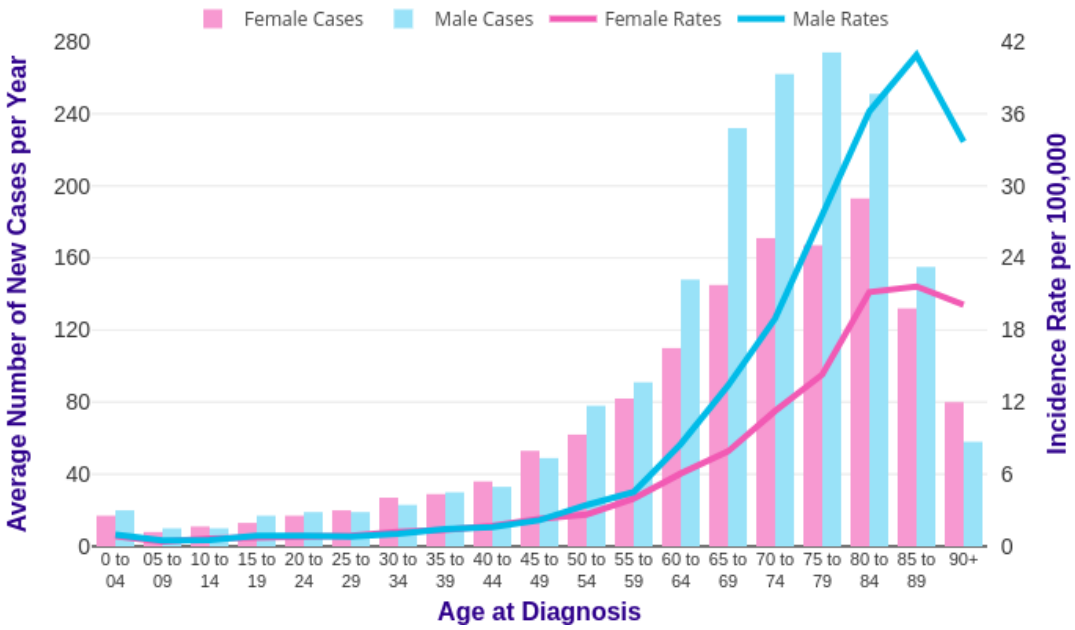


Figure 1.3 AML incidence in UK by age. A small increase in AML cases is shown between 0-4 years of age, followed by a steady increase post 10 years. Significant disparity between males and females appears from 60-64. Data shown taken from Cancer research UK (UK, 2019). [Accessed 12/6/21].

Before the age of 60, AML incidence rates between biological males and females were relatively similar; however, following this the rate in males increased significantly compared to those found in women. This might be a result of the various vocations men have held historically,

which exposed them to more cancer risk factors and encouraged risk-taking behaviours like smoking.

The higher rates could also be due to mutations on the sex-linked chromosome (SLC); having only one X copy of this might make males more susceptible to mutations and the cancers associated with genes on this chromosome. Abnormalities present in the SLC could potentially cause genetic instability and result in the progression of AML (Shahrabi et al., 2018). Another reason more males are diagnosed with AML could be due to blood group distribution; a sex-dependent gene on chromosome 9 may protect blood group O women against leukaemia in certain populations (Jackson et al., 1999).

1.2.2 Aetiology

The origins of AML are not exactly known. In adult patients, it is predicted that the development of AML is due to the accumulation of genetic alterations; this also predicts that the disease will be more complex due to the nature of its cause. In children, it is predicted an in-utero mutation leads to the development of AML following a second genetic event occurs at around 18 months of age, as reviewed in (Greaves et al., 2003).

Environmental factors such as exposure to abnormal levels of radiation, smoking and benzene, along with contact with previous chemotherapeutic agents as well as any other carcinogenic material, can all contribute towards a greater risk of contracting AML (ACS, 2018). Some pre-existing conditions can make a patient more liable to develop AML, other haematological diseases such as myelodysplastic syndrome (MDS), myelofibrosis, and aplastic anaemia. Birth defect diseases such as Down Syndrome and Bloom Syndrome also show an increased prevalence of AML, often presenting in early adulthood. Secondary AML can occur following a previous haematological disorder and has a poor survival rate and a greater risk of relapse, along with conferring greater resistance to treatment.

Certain chromosomal changes such as deletions, inversions, and translocations can be associated with prognosis and can be used to classify leukaemias within the World Health Organisation (WHO) system. Mutations in single genes can also affect patient prognosis, although the most common are only found in around a quarter of patients: nucleophosmin (NPM1), Fms-related tyrosine kinase 3 (FLT3) and DNA methyltransferase 3A (DiNardo & Cortes, 2016). Other gene alterations include: isocitrate dehydrogenase 1 & 2, tet methylcytosine dioxygenase 2, ASXL transcriptional regulator 1, runt-related transcription factor 1, Wilms tumour 1 and P53 (Takahashi, 2011). Currently, at diagnosis, these molecular changes are examined to help inform the treatment of the patient.

1.2.3 Symptoms

AML presents with a wide variety of symptoms, including feeling tired or weak, pallor, breathlessness, recurrent infections, unusual bleeding; and weight loss, amongst others (NHS, 2019). These symptoms are often caused by a lack of healthy blood cells.

1.2.4 Diagnosis and classification

AML is often diagnosed when the patient develops symptoms that lead them to a primary caregiver and a full blood count, which will show irregularities in the blood profile such as a high number of white blood cells or low blood cell count. A peripheral blood smear, in which a sample of blood is examined under a microscope for the presence of blast cells and abnormal/immature cells in the peripheral blood or BM, can also be used to diagnose AML. While evaluating the cell counts and morphology of the cells is useful, other methods can provide a greater depth of information on the condition. Cytochemistry can also be used to stain the cells and determine if it is of myeloid or lymphoid lineage. Following this, the patient will often have a BM biopsy or lumbar puncture to observe the BM.

Cytogenetic testing is performed to determine the prognosis of an individual. This involves incubating blood or BM samples with colcemid™ overnight to prevent mitotic spindle formation and leaving the chromosomes in their condensed state (Czepulkowski, 2001), so Giemsa (G)-banding can facilitate the examination of aneuploidy and chromosomal rearrangements such as translocations, inversions, duplications, and deletions. A karyogram might show large-scale rearrangements such as the creation of a Philadelphia chromosome caused by a t(9;22)(q34;q11) translocation that is often found in CML.

Cluster of Differentiation antigens (CD) detail cell surface receptors that are expressed by leukocytes and can be identified by a single monoclonal antibody (Belov et al., 2001); this allows for the interpretation of antibodies the leukaemia is expressing. Fluorescence in situ hybridization helps diagnose any smaller translocations not found using a karyogram (Wheeler et al., 2018).

Previously, the French-American-British (FAB) method of classification, first proposed in 1976, was used in diagnosing subtypes of AML (Bennett et al., 1976). The FAB method of classification focuses on how leukaemia cells look under a microscope after staining by identifying the type of cell and the maturation (**Table 1.1**). The subtypes M0 to M5 form within immature white blood cells, whilst M6 develops in immature RBCs and M7 progresses in cells that develop into platelets. Some of the FAB subtypes are more common than others (**Table 1.1**). However, the WHO criteria account for more factors that could potentially affect AML prognosis (**Table 1.2**). In 2016, the WHO system was revised and has taken over as the predominant method of

classification, incorporating new diagnostic approaches such as genetic data, biomarkers, morphology, and immunotherapy (Arber et al., 2016).

Table 1.1 Table showing FAB subtype classification and prevalence. Percentage of FAB subtypes upon diagnosis based on a study excluding M3 AML and combining M6 and M7. M2 was the most common form of AML in this cohort (n=1690) (Canaani et al., 2017).

FAB Subtype	AML Type	Occurrence
M0	Undifferentiated acute myeloblastic leukaemia	7.8%
M1	Acute myeloblastic leukaemia with minimal maturation	25.5%
M2	Acute myeloblastic leukaemia with maturation	27.5%
M3	Acute promyelocytic leukaemia (APL)	NA
M4	Acute myelomonocytic leukaemia	14.9%
M4 eos	Acute myelomonocytic leukaemia with eosinophilia	
M5	Acute monocytic leukaemia	19.7%
M6	Acute erythroid leukaemia	4.4%
M7	Acute megakaryoblastic leukaemia	

Table 1.2 Table showing WHO 2016 classifications of AML.

Acute myeloid leukaemia (AML) and related neoplasms	
AML with recurrent genetic abnormalities	AML with myelodysplasia-related changes
AML with t(8;21)(q22;q22.1);*RUNX1-RUNX1T1*	Therapy-related myeloid neoplasms
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22);*CBFB-MYH11*	AML, Not Otherwise Specified (NOS) *
Acute promyelocytic leukaemia (APL) with *PML-RARA*	AML with minimal differentiation
AML with t(9;11)(p21.3;q23.3);*MLLT3-KMT2A*	AML without maturation
AML with t(6;9)(p23;q34.1);*DEK-NUP214*	AML with maturation
AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2*, *MECOM*	Acute myelomonocytic leukaemia
AML (megakaryoblastic) with t(1;22)(p13.3;q13.3);*RBM15-MKL1*	Acute monoblastic/monocytic leukaemia
Provisional entity: AML with BCR-ABL1	Pure erythroid leukaemia
AML with mutated *NPM1*	Acute megakaryoblastic leukaemia
AML with biallelic mutations of *CEBPA*	Acute basophilic leukaemia
Provisional entity: AML with mutated RUNX1	Acute panmyelosis with myelofibrosis
Myeloid sarcoma	Transient abnormal myelopoiesis (TAM)
Myeloid proliferations related to Down syndrome	Myeloid leukaemia associated with Down syndrome

*Not otherwise specified as AML, includes any AML patients whose disease cytogenetic do not fit into another category.

1.2.5 Conventional treatment

Classical treatments for AML include chemotherapy, chemotherapy with stem cell transplant and other drug therapies. Prior to treatment with chemotherapeutic agents, a patient with very

high levels of leukaemic cells in the blood (leukostasis) may need leukapheresis, a process that removes excess white blood cells by cycling the blood through a filter. Whilst this can lower the blood count immediately, it is not a permanent solution; however, it does help to reduce the chance of tumour lysis syndrome; a process in which the dying cancer cells release their contents into the blood and cause spikes in phosphorous, calcium and uric acid (Gupta & Moore, 2018).

Chemotherapy can be separated into distinct phases; induction therapy, which aims to achieve complete remission by destroying the blast cells and consolidation treatment, which is performed to kill the small number of cancerous cells that remain after treatment, thus reducing the risk of relapse. Cycles of therapy are used for consolidation to allow time for the body to recover between treatments. Finally, maintenance therapy can be used to try to prevent relapse and is a common course in ALL, though used less often in AML (Terwilliger & Abdul-Hay, 2017). Chemotherapeutic agents are commonly administered intravenously or via ingestion and include: cytarabine, anthracycline drugs, etoposide, fludarabine, dexamethasone, mitoxantrone, decitabine, azacytidine and others (ACS, 2020).

The current standard care for adult AML is 7+3; this is 3 days of induction with anthracycline and 7 days of infusion with cytarabine (Dombret & Gardin, 2016). Following consolidation, a patient may have hematopoietic stem-cell transplantation (HSCT), depending on how they have responded to treatment. If a patient has already entered remission, then a HSCT isn't required. The stem cells for a HSCT can come from the patient (autologous) or a donor (allogeneic), from peripheral blood, BM or the umbilical cord. In autologous stem cell transplantation, peripheral blood or BM stem cells will have been taken from the patient at an earlier time point and been frozen; the blood is treated to attempt to remove any leukaemic cells before being given back to the patient. In this scenario, autologous HSCT is used to rescue patients when other treatments fail. For allogeneic transplants, a human leukocyte antigen (HLA), a matched donor will be found, often siblings and family. To harvest HSCTs from the PB, the donor will take medication such as granulocyte colony stimulating factors to stimulate stem cell production and release into the blood, which are then collected. The stem cells can also be taken from umbilical cord blood or from BM; however, PB is the most common source. Following this, the stem cells are added into the patients' blood, from where they gather in the BM.

There are also three categories of patients post treatment; complete remission in which there are less than 5% blast cells, with no leukaemia symptoms; normal full blood count and no blast cells found using PCR, known as molecular remission. Minimal residual disease is achieved when the leukaemia cannot be detected with microscopy methods, known as cytogenetic remission, however more sensitive tests (PCR or flow cytometry) will still detect some leukaemic cells.

Active disease involves the presence of clinical signs and symptoms of disease as well as >5% blast cells found in the BM.

Estimating the benefit/risk associated with allogeneic HSCT in first remission for a given patient is one of the most crucial treatment decisions in AML. Despite transplantation being the most effective treatment for preventing AML recurrence, it is associated with a higher rate of treatment-related morbidity and mortality (TRM), particularly in older patients. Patients with poor risk AML, who are unable to receive an HSCT have worse outcomes than patients that can receive a HSCT (Vasu et al., 2018). In M3 AML, vitamin insufficiency is common, and all-trans-retinoic-acid (ATRA) can promote differentiation in malignant erythroid cells (Su et al., 2015).

1.2.5.1 Treatment in Paediatrics

Paediatric patients (from birth to 16 years of age) tend to respond a lot better to treatment; they often do not have any underlying health issues or comorbidities which could complicate treatments (Storey et al., 2017). However, after having AML at a young age, the chance of relapse is increased to around 24-40% (Kaspers, 2014). In paediatric patients, it is predicted that the development of AML is due to one or two genetic changes and that the frequency of somatic mutations is greatly reduced in paediatric AML (Mercher & Schwaller, 2019).

In adults, the use of HSCT is almost always the most effective treatment; however, in the paediatric AML low risk subgroup patients, it was shown to have no effect on survival compared to multi-agent chemotherapy (Gibson et al., 2005). Infants with a high risk of relapse and AML subtypes M0, M6 or M7 should even receive a non-familial HSCT; paediatric patients with MDS or secondary AML should also receive HSCT from a sibling or matched donor (Ljungman et al., 2010). The risk of graft versus host disease is also lower in paediatrics than in adults.

1.2.5.2 Treatment in adults

As older patients tend to have other underlying health conditions which can complicate treatment or mean they cannot survive more intense chemotherapeutic regimes, this can/does lead to mortality rates being higher in adults. If a high risk subgroup patient has an intra-family donor, then an HSCT should always be performed to improve survival rates (Ljungman et al., 2010).

In AML, patients' age is considered when deciding the best treatment approach as well as multi-morbidities. Where patients haven't responded to previous conventional treatment, then more experimental approaches may be considered or the use of palliative care.

1.3 Targeted treatments

Targeted treatments are becoming more common, and the main benefit is that they are more specific for cancer cells as opposed to conventional treatments that can also damage healthy cells within the patient. One type of targeted therapy used clinically since 2005, is monoclonal antibodies (mAbs) where specific antibodies bind to surface receptors on cancer cells (Adams & Weiner, 2005). Targeted therapies can be used to block growth, such as midostaurin (FLT3 inhibitor) (Levis, 2017), encourage maturation with IDH inhibitors (McMurry et al., 2021); destroy the cells like the mAb Gemtuzumab ozogamicin (Yu & Liu, 2019), or prevent pathway function like Glasdegib in the hedgehog pathway (Wolska-Washer & Robak, 2019). The main benefit of mAbs and other targeted treatments is the high specificity to the target cell, leading to less toxicity in non-cancerous cells (Bostrom et al., 2009).

The ideal AML treatment would have to have a high specificity for blasts and would not be present or would be present at considerably lower levels in normal stem cells, circumventing the death of healthy cells.

1.3.1 Immunotherapy

Another promising therapeutic option is immunotherapy, which involves using the body's immune system to target and remove cancer cells. One method uses T-cells, and while this can produce the desired effects of killing leukaemic cells (Daver et al., 2021), T-cells can also cause harmful effects to the host's healthy cells such as cytokine release syndrome (Maude et al., 2014). Immunotherapy covers a wide range of novel treatments (**Table 1.3**).

Table 1.3 Immunotherapy strategies.

Approach	Targets
Adoptive T Cell Therapies	CAR-T Cells (Mardiana & Gill, 2020), CD44v6 CAR-T Cells (Casucci et al., 2013), FLT-3 CAR-T Cells (Jetani et al., 2018), CLL1/CD33 CAR-T Cells (Liu et al., 2018a)
Adoptive NK Cells Therapies	Infusion of Haploidentical NK Cells (Lee et al., 2016) and NK Cell Activation (Xu & Niu, 2020).
T Cell Engagers	Bispecific Antibodies Targeting CD33, Tandem Diabodies IL-3 and CD123 (Guy & Uy, 2018), BiTes and BiKes (Jitschin et al., 2018), Dual Affinity Retargeting Antibodies (Al-Hussaini et al., 2016).
Checkpoint Blockade and Macrophage Checkpoint Blockade	PD-1 inhibitor (Syn et al., 2017), TIM-3 (Acharya et al., 2020), CD47 (Chao et al., 2019).
Monoclonal Antibodies	CD33, CD123, CD16, CD38, CD47, CD25 (Busfield et al., 2014).
Vaccines	Boost immunity and provoke a response (Coppage et al., 2007), Wilms' tumor 1 (WT1) (Van Tendeloo Viggo et al., 2010), P3 peptide vaccination (Qazilbash et al., 2017).

1.3.2 CAR-T Therapy

Recently, treatment has progressed with the use of Cancer Antigen Receptor Targets (CAR-T) (Marofi et al., 2021). CARs are recombinant receptors; these can bind to antigens and activate T-cell responses that would target cancerous cells. The four components of CARs are a hinge region, antigen binding domain, a transmembrane domain, and a signalling domain (Majzner & Mackall, 2018). The second generation of CARs are in development after the use of tri-partite receptors comprising a costimulatory domain and can be used to target surface receptors. Additionally, CARs can bind to carbohydrates, and glycolipid structures as well as proteins. The CAR-T cells can also be combined with chimeric costimulatory receptors, cytokines or, costimulatory ligands (Gill, 2019). For CAR therapy to be successful in AML, there needs to be exclusive AML surface antigens for them to bind to. Unfortunately, cancerous cells can develop resistance to single antigen targeting therapy. Despite initial strong positive reactions by CAR-T cells against target antigen expressing cells, patients that have been given CAR-T cell therapy can lose expression of the target antigen (antigen escape) on their tumour cells (Sterner & Sterner, 2021). Other obstacles related to CAR-T therapy include the challenges in getting the CAR-T cells into a solid tumour, further impeded by the tumour microenvironment and associated toxicity including cytokine release syndrome and neurotoxicity (Majzner & Mackall, 2018).

1.4 Biomarker analysis

During this study, we analysed antigens as potential targets for treatment and microRNAs (miRs) as potential biomarkers of disease risk. An ideal target for treatment would be an antigen that is expressed in most AML patient samples, expressed on the surface of the leukaemic cells and not expressed in healthy BM. Whilst a biomarker for prognosis needs to be expressed in considerably higher levels in all patients and be easily detectable, if it is expressed in healthy cells, it is not as important if they aren't being targeted for destruction by a therapy. Previous research has focused predominantly on adult AML; for this reason, this study aims to evaluate the potential of some leukaemia-associated antigens (LAAs) found to be expressed in adults and/or paediatric patients and act as targets for therapy.

A previous review highlighted the ideal cancer antigen characteristics in leukaemias; the first being specificity for leukaemia (Anguille et al., 2012). Leukaemia-specific antigens are the best targets due to only being expressed in the malignant cells; however, these are harder to find and are often only expressed in 30% of cases (Guinn et al., 2005). LAAs make a promising target for immunotherapy but are not exclusively expressed in the tumour. Following criteria one, the

pattern of expression should be in most AML patients and in most diseased cells. Next, the oncogenicity of the antigen should be observed, and its role within the cancer, as well as its function in normal cells. Penultimately, the immunogenicity of the antigen is vital; a strong response needs to be generated by the immune system towards the antigen. Preferably, it would trigger the activation of cytotoxic T lymphocytes, primed to kill the target cell. Finally, the clinical relevance of the antigen is important; the main objective being to provide benefit to the patient in the form of remission or a decrease in blast count.

1.5 Understanding the molecular mechanisms that underlie paediatric versus adult AML

The current belief is that paediatric and adult AML are two independent diseases. Paediatric AML occurs due to one or two genetic changes, the first often in utero, leading to the peak incidence of AML in the first year of life (Puumala et al., 2013). Adult AML may be due to an accumulation of genetic changes over an extended period (Bolouri et al., 2018). The National Cancer Institute - Therapeutically Applicable Research to Generate Effective Treatments (NCI-TARGET) study highlighted other differences in the genetics of AML. Paediatric patients had diverse and uncommon mutations in genes, while somatic structural variants appeared more often than in adults.

Davis et al examined the expression of 120 cancer testes antigens and LAAs identified as important in cancers including AML (Davis et al., 2020). The aim of this study was to identify which of these antigens were differentially expressed between risk subgroups, thus identifying antigens associated with poorer patient survival. Two databases were used for this analysis to ensure that adults and paediatrics with AML could be directly compared. The Cancer Genome Atlas (TCGA-LAML) (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga) was used to investigate differentially expressed genes (DEGs) in adults with AML whilst the TARGET dataset (https://ocg.cancer.gov/programs/target) was used to examine gene expression in samples from patients with paediatric AML. Only risk subgroups with more than ten patients were examined, which meant that only the TARGET database's standard versus low-risk subgroup categories were compared, whilst the TCGA database's poor versus good, and intermediate versus good, risk subgroups were compared. The TCGA dataset focused on primary untreated tumours and matched patient blood samples; they were then examined by the Biospecimen Core Resource (BCR). Reviewed by a pathologist, the nucleic acids were then isolated, and genotyping performed. This was then sent to a genome sequencing centre for analysis (NCI).

As previously described, the HiSeq 2000 (Illumina) was used to obtain paired-end mRNA sequencing data for the TCGA database (Ley et al., 2013). The mRNA sequences and clinical data came from the National Cancer Institute's Genomic Data Commons (GDC) Portal, including 151 patients. However, Davis et al excluded patients with the FAB classification M3 (APL) and the BCR-ABL1 gene fusion because they were not included in the TARGET AML dataset. This brought the number of patient samples examined to 133. When patients exceeded 18 years of age, they were classified as adults and this dataset contained patients within the age range of 21 to 88. The mRNA sequencing data utilised in the Davis et al study were derived from phs000465 and were created by the TARGET study. The methods used to obtain the mRNA sequencing data are shown here (Bolouri et al., 2018). There were 92 patients' raw mRNA-sequencing data obtained from the TARGET dataset; however, only data from the 65 patients who were sequenced using the HiSeq 2000 were included in the study to match the sequencing machinery used for the TCGA data. In this previous, study the patients were classed as paediatric when their age at diagnosis was < 10 years old, and ages ranged from 0.4 years to 9.7 years. The oldest patient at the final check-up was 17.7 years old. The study used the TCGA and TARGET definitions of risk subgroups (**Table 1.4**).

Table 1.4 Definition of risk groups (as defined by TCGA or TARGET) via their potential clinical risk derived from molecular characteristics and cytogenetic indicators.

	Risk	Molecular and cytogenetics indicators
TCGA	Poor	Poor risk MLL translocation like t(6;11)(q27;q23), t(11;19)(q23;p13) or the presence of complex cytogenetics and poor risk cytogenetic abnormalities (Ley & Miller, 2013).
	Intermediate	Normal karyotype, the MLL translocation t(9;11)(q23;p13) or other intermediate cytogenetic abnormalities
	Good	t(8;21)(q;22;q22) or inv(16)(p13q22)
TARGET	Low	t(8;21)(q22;q22) inv(16)(p13q22)
	Standard	MLL gene rearrangements such as t(6;11)(q27;q23), t(10;11)(p12;q23), t(9;11)(q23;p13), t(11;19)(q23;p13) or a normal karyotype that didn't pose a change in risk.
	High	FLT3 Internal Tandem Duplication (ITD) (Bolouri et al., 2018).

We wanted to determine whether the molecular pathways subverted in paediatric and adult patients with AML were the same or different. If targeted treatments that have been used predominantly for one age group of patients could be repurposed for the other, this could extend treatment options and hopefully survival times. In order to address the hypothesis, we continued the work initiated by (Davis et al., 2020) using the genes of interest (GOI) that were

differentially expressed between risk sub-groups (**Figure 1.4**). In doing so, we hoped to better understand the disease processes and whether paediatric and adult AML were indeed two separate diseases. We hope this work will identify new targets for therapy and widen our understanding of the biology that underlies AML in both paediatrics and adults.

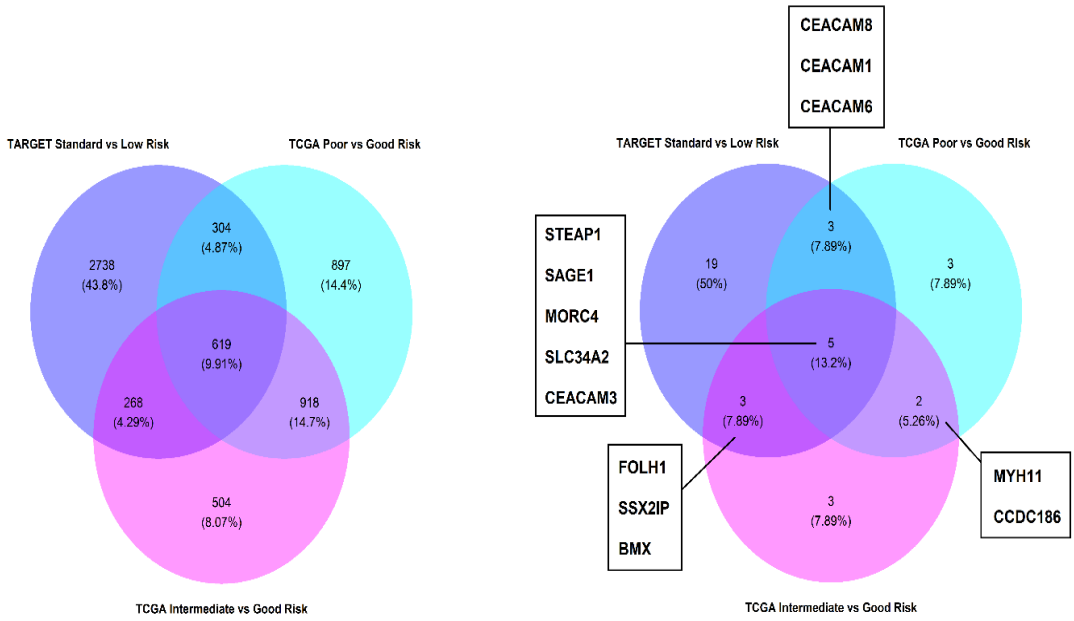


Figure 1.4 Comparisons of risk subgroups and identification of DEGs in AML. Venn diagrams showing the number of genes that were significantly differentially expressed between risk subgroups in the TCGA and TARGET databases, image used from Davis et al.

1.6 microRNAs

Also collated in the DEG list were more than 50 microRNAs (miRs), with very significantly differential expression between poor and good risk subgroups. miRs are small (22 nucleotides on average), non-coding single-stranded oligonucleotides that control gene transcription and translation by silencing or degrading certain messenger RNAs (mRNAs). (Bartel, 2004). Genes that code for miRs are transcribed into primary mRNA transcripts, which then mature in the cytoplasm and nucleus. The mature miRNA binds to the RNA-induced silencing complex (RISC), which then targets the 3' untranslated region of mRNAs for silencing, inhibiting protein creation through lower target stability and translation (Bartel, 2009).

Approximately half of all miRNAs are processed from introns and are therefore intragenic, whilst a few are also processed from exons (Kim & Kim, 2007). miRNAs can also be categorised in two ways by the canonical or non-canonical pathway, the canonical pathway being the most common. Pri-miRNAs are transcribed and developed into pre-miRNAs by a complex of Drosha, a ribonuclease III enzyme, and DiGeorge Syndrome Critical Region 8 (DGCR8), an RNA-binding

protein (Denli et al., 2004). The DGCR8 then recognizes the N6-methyladenylated GGAC motif amongst others in the pri-miRNA before the duplex at the base is cleaved by Drosha. The pre-miRNA is transported into the cytoplasm by the RanGTP/exportin 5 complex, which is then processed by the RNase III endonuclease Dicer that eliminates the terminal loop and creates a mature miRNA duplex (Macfarlane & Murphy, 2010). The orientation of the strand affects whether the miRNA is 3p or 5p. The 3p strand originates from the pre-miRNA hairpin's 3' end, while the 5p strand originates from the 5' end. Following this, the strands are moved into the Argonaute protein (AGO1-4) using ATP (Iwasaki et al., 2010). The amount of 3p or 5p AGO-loaded stands depends on a number of parameters, including the duplex's thermodynamic stability at the 5′ end (O'Brien et al., 2018), with the guide strand having the lowest stability (O'Brien et al., 2018), with the lowest stability strand being the guide strand. The other strand is unwound from this and degraded by AGO2 and other mechanisms (Bail et al., 2010). Non-canonical miRNAs are a subset of miRNAs; they resemble the same formed via the canonical route, however, may skip stages of the this pathway (Abdelfattah et al., 2014). Although Dicer is required to produce miRNAs, Drosha and DGCR8 can be removed because they are only required for the canonical pathway (Okamura et al., 2007; Ruby et al., 2007).

The mechanisms miRNAs can modulate gene expression also differ. One such method is the RISC complex, which includes the guide RNA, the Argonaute protein and glycine-tryptophan repeat-containing protein of 182. This GW182 complex has multiple domains essential for mRNA degradation; the deadenylation starts with the poly-A nuclease 2/3 complex shortening the poly-a tail, which is degraded by the carbon catabolite repression 4 negative on TATA-less (Duchaine & Fabian, 2019). Poly(A)-binding protein also stimulates the cap-dependant mRNA translation by interacting with eukaryotic initiation factor 4 gamma (Kahvejian et al., 2005). Polyadenylated mRNAs are repressed more strongly by miRNAs than non-polyadenylated mRNAs (Moretti et al., 2012).

Whilst research shows that miRNAs can form complexes that restrict protein production, there is also evidence that miRNAs can upregulate protein expression in some situations, possibly as a result of interactions with AGO2 and fragile-x-mental retardation related protein 1 (Vasudevan & Steitz, 2007; Truesdell et al., 2012). It has also been suggested that miRNAs can act within the nucleus to alter expression; however, the exact methods of this remain unclear (Catalanotto et al., 2016). A much smaller RISC complex (160 compared to 3 MDa, without Dicer or transactivation response element RNA-binding protein) may be formed with Importin 8 mediating AGO2 into the nucleus (Ohrt et al., 2008). This complex could potentially regulate RNAs or pre-miRNAs and serve as enhancer triggers to upregulate expression, with multiple models being proposed (Liu et al., 2018b).

miRNAs can accumulate in both endosomes and exosomes found within the reticulum, processing bodies, stress granules, early/late endosomes, the trans-Golgi network, multivesicular bodies, lysosomes, mitochondria and the nucleus (Leung, 2015; O'Brien et al., 2018). miRs are found in a wide range of areas including sera, allowing for potential use as a biomarker (Sohn et al., 2015). It is believed there are two types of miRs identified in human fluids: those found in micro vesicles, exosomes, and apoptotic bodies, and those found with proteins, particularly AGO2 (Chen et al., 2008; Gallo et al., 2012).

miRNAs play a vital function in regulating gene expression via silencing in the processes described above, and their roles in tumours have been demonstrated (Peng & Croce, 2016), as well as the impact of dysregulated miR expression on AML (Fabbri et al., 2008). Current research has investigated the deletion of chromosome 5q suppressing expression of some miRNA-145 and 146a; this reduced expression and in vivo experiments caused myelodysplastic syndrome (Starczynowski et al., 2010).

A review summarised some of the preclinical studies on miRNA-based therapeutics (Wallace & O'Connell, 2017). These are often mimics or antagomiRs that block other molecules from binding with sites on the miRNA (Krützfeldt et al., 2005). There is a range of miRNA mimics that were being tested: miR-22, 29b, 181a and a range of antagomiRs: miR-21/196b and 126. Another recent review by our group analysed the potential of some miRs in AML (Fletcher et., In press, *eJHaem*).

1.7 Thesis Aims

- To determine, through pathway analysis, whether paediatric and adult AML are similar.
- Evaluate the biomarker potential of miRs to categorise AML patients into risk subgroups.

Chapter 2. Materials and Methods

2.1 Characterisation of genes and pathways that play a key role in AML

2.1.1 Identification of GOIs

DEGs were found via a study investigating mRNA-sequencing data in two AML databases (Davis et al., 2020). Davis et al investigated the TCGA adult database (Ley & Miller, 2013) and the TARGET for paediatrics (Bolouri et al., 2018) to look for gene expression difference between AML risk subgroups.

In the TCGA dataset patients were classed as adults when older than 18 years old and the range of ages varied from 21-88 years old. The GDC portal had mRNA-sequencing data with clinical data for 151 patients, but individuals with M3 FAB classification were excluded because they did not appear in the TARGET database. This left 133 patients; however, this number was cut down further to 65 only including patients where cytogenetic and molecular risk classifications were aligned (**Table 1.4**).

The TARGET database patients were classified at age of diagnosis as paediatrics who were less than 10 years old, with ages ranging from 0.4 years to 9.7 years old. The sequencing data was from phs000465, the original group of 92 patients. As some patients were sequenced on both the Illumina HiSeq 2000 and the Illumina HiSeq 2500, it was decided to only include patients who were sequenced on the Illumina HiSeq 2000 to avoid any bias generated by the instruments. This left 65 patients who met the authors criteria and had samples that were sequenced.

With the risk groups created analysis was carried out using the DESeq2 package available as part of the 'Bioconductor' package (Love et al., 2014) within R (Team, 2013). The DESeq2 package accounts for outliers (Cook's distance), low replicates numbers, high ranges, and discreteness, curves are fitted for each gene to move values towards the estimated values (Davis et al., 2020). During the differential gene expression (DGE) analysis the default settings were used with the DESeq function, further Bioconductor packages 'AnnotationDbi' and 'org.Hs.eg.db' were used in the analysis. Following these, results of differentially expressed genes between the subgroups were organised by p value and displayed using Volcano plots from base R and the calibrate package.

2.1.2 Grouping AML with prognostics

To allow for differential expression analysis, patients were grouped via different genetic risk groups (**Table 1.4**) within the clinical data, these were based on cytogenetics that indicated prognosis. These could contribute to event free survival (EFS; risk of relapse) and overall survival

(OS) and would contribute to the grouping on Kaplan-Meier (KM) plots. OS was defined in both datasets as the time from diagnosis to death, or the last contact with the patient while he or she was alive. In TCGA, EFS was defined as the time from diagnosis to death or relapse and in TARGET, EFS was defined as the period from diagnosis to induction failure, death, or relapse.

Survival plots are generated using the raw count data at different quartiles with the survival package (Therneau, 2020), then the 'survdiff' function allowed comparison using log-rank tests comparing EFS between Q1-4 and Q2-3, p-values were considered significant when below 0.05. P-values were collected using the 'Survutils' package (Chan, 2020). This survival information informed the decision on which LAAs to investigate further. The analysis compared prognosis pairs and was accessed in **Table S1A** and **S1B** available online (https://www.mdpi.com/2072-6694/12/10/2769/s1).

2.1.3 STRING analysis

Once we had identified which GOIs to investigate for the functions and pathways they could be involved in we used the Search Tool for the Retrieval of Interacting Genes (STRING) analysis program (Szklarczyk et al., 2019). This publicly available database provided information on protein interactions allowing us to evaluate the GOI and displays a 'web' of other closely related genes. STRING analysis can group genes via various methods such as: co-occurrence, co-expression, text-mining, fusion, experimental, neighbourhood and database evidence. This facilitated a basic understanding of the functions of these genes and the other genes they were related to. STRING was only used to provide support for our understanding of the protein partners and functional roles in the cell.

2.1.4 BIRC5 Expression in Healthy Haematopoietic Cells (Bloodspot)

The affect expression had on OS was deciphered using BloodSpot (www.bloodspot.eu;)(Bagger et al., 2019). BloodSpot encompasses 23 datasets relating to normal and malignant blood formation. It can be used to compare the OS of patients based on differing levels of gene expression. Bloodspot also contains information about the hierarchical differentiation tree of health blood and BM cells, and if certain gene expression levels are elevated in samples from patients with one cytogenetic variation compared with another The Microarray Innovations in Leukaemia (MILE) study database was used, which is a multi-laboratory database with over 3000 whole genome microarray analysis results (Kohlmann et al., 2008), it was organized by the European Leukaemia Network and funded by Roche Molecular Systems, Inc. (Pleasanton, CA, USA).

2.1.5 KEGG pathways

The pathway analysis was further investigated using Kyoto Encyclopaedia of Genes and Genomes (KEGG) (https://www.genome.jp/kegg/(Kanehisa & Goto, 2000)), this gave a good starting point to look at the pathways of the GOI. KEGG is a database of manually collated pathway diagrams that represent known interactions and reactions. Whilst this was very useful for some of the GOI (Baculoviral IAP Repeat Containing 5 (BIRC5) and Six-transmembrane epithelial antigen of prostate family member 1 (STEAP1)), others had no current pathways available on the site. These were: Vestigial Like Family Member 4 (VGLL4), Carcinoembryonic antigen-related cell adhesion molecule 3 (CEACAM3), Solute Carrier Family 34 Member 2 (SLC34A2), Melanotransferrin (MELTF), Melanoma-Associated Antigen F1 (MAGEF1), MORC Family CW-Type Zinc Finger 4 (MORC4) and Sarcoma Antigen 1 (SAGE1). However, KEGG provided a much greater insight into the pathways that were involved with BIRC5, because it is a well-known cancer antigen that has had other studies looking at its function and pathways. To further inform the KEGG pathway analysis map, a literature search was performed out. PubMed was the predominant search engine used and provided most the papers we examined for descriptions of the functions and pathways of the GOIs.

2.1.6 Correlation analysis

Once the pathway diagram was prepared and relationships between the GOI and other genes were established the correlation between these antigens was determined by Professor Ken Mills (Queens University Belfast; DOI: 10.5281/zenodo.4923749) that contained the p values of the relationships between each GOI and 54,000 other genes on the microarray. This was used to reinforce the connection between our antigens and their respective pathways, where available, by applying a p-value. Multiple probe sets were available for some of the genes and so these when multiple probes were available the following probe set numbers were used 202094, 208052 and 212399 for BIRC5, CEACAM3 and VGLL4, respectively.

2.1.7 Association between genes and clinical features

TCGA data was used to investigate the link between each BIRC5 probe set and clinical characteristics of adult AML. M6 and all other FAB subtypes (M0 – M5, M7) or M7 and all other FAB subtypes (M0 – M6), FLT3-ITD and FLT-WT, M6 and all other FAB subtypes (M0 – M5, M7) or M7 and all other FAB subtypes (M0 – M6).

2.1.8 GO and GSEA

There are multiple methods of comparison for large gene sets, two of the main are Gene Set Enrichment Analysis (Reimand et al., 2019) and Gene Ontology (GO) (Zhang et al., 2010). There was a set of DEGs from the Davis et al study and using GO analysis and the PANTHER resource

(http://www.pantherdb.org/) (Mi et al., 2020) the percentage of those genes who were involved in a specific pathway was compared to the total number of genes in that pathway. PANTHER is a database of genes and proteins combined with their functional families that can be used to evaluate gene function, ontology, and the pathways they are involved in. Using the DGE list generated from TCGA and TARGET risk subgroups by Davis et al, we could find the main functions of the gene groups and identify any major differences between the groups.

2.2 miR Analysis

2.2.1 Identifying differentially expressed miRs

As well as certain LAAs being identified in the TCGA and TARGET database as having differential expression between subgroups, we also noticed that a number of miR coding genes were differentially expressed between risk subgroups based on the work performed by (Davis et al., 2020). By identifying miRs with a more than two-fold expression change between the subgroups we focused on a smaller number of miRs that were likely to differ more and be better prognostic markers for risk group. Dr Pinar Onganer (University of Westminster) used the University of Santa Cruz California Santa Cruz (UCSC) Xena browser (https://Xenabrowser.net/) (Goldman et al., 2020) to check the levels of these miRs in TCGA-LAML dataset. Dr Pinar Onganer identified three miRs that had an association with OS as well as fourth with increased copy number in the poor risk subgroup. These were chosen to be further investigated during this study (miR-486-1, miR-1915, miR8086 and miR-378G).

2.2.2 Predicted pathway model

A literature search was performed using PubMed and Scopus to determine the pathways that each miR had already been reported as acting on. The pathway involvement of the miRs was also analysed via KEGG if possible, DAVID (https://david.ncifcrf.gov/) (Huang da et al., 2009) and Bloodspot.

2.2.3 ROC Curves and AUC

We evaluated miRs using Area under the Curve-Receiver Operator Characteristic Curve (AUC-ROC) analysis, which uses the ability of a single or grouped factor to predict the grouping of a patient. First devised in 1941 for operators of military radar receivers (Fukunaga, 2013), it is used to plot the true positive rate against the true negative rate at various threshold settings, in our case 0.5. ROC is a measure of the sensitivity and specificity of a potential biomarker; the proportion of positives correctly detected is referred to as sensitivity, whereas the proportion of negatives correctly identified is referred to as specificity. The perfect biomarker would accurately classify 100% of cases and have an AUC value of 1, a random classifier would get 50%

right and have an AUC of 0.5 (Hajian-Tilaki, 2013). RNA-seq data (TARGET or TCGA) was used to determine which miRs were differentially expressed between risk subgroups (Davis et al., 2020).

ROC is a binary analysis and needs only two possible outcomes, our patients would have either an intermediate or good prognosis or a poor versus good prognosis depending on the dataset. The miRs were first tested individually and then in small groups using SPSS.

2.2.4 Biomarker model

A logistic regression analysis was performed in SPSS to create the combined model of miRs to predict prognosis, first RNA-seq data was imported into columns in accordance with one of the two prognosis options (https://case.truman.edu/files/2015/06/SPSS-Logistic-Regression.pdf). Prognosis was the dependant variable whilst the miRs-486-1, 1915, 8086 and 378G were added in a step wise fashion to the covariates. Probabilities and group membership were saved to create a new output. For model validation the classification table provided the correctly specified group membership as a percentage. To generate a ROC curve and AUC score for the model the predicted probability was used as the test variable.

The R-squared value in a linear regression provides a measure of the proportion of variation in a dependent variable, that can be explained by the model, however this cannot be calculated in binary logistic regression. An alternative R-squared analysis appears in model summary as the Nagelkerke value (Nagelkerke et al., 1984).

To account for any other variables that may have affected the miRNA expression values in the RNA-seq data the models were adjusted for age and gender, this was carried out by adding them in a step wise manor to the model. If either variable were significant then the variables in the equation would return significant B values.

2.2.5 Sample information

Samples from AML006-AML041 were collected from patients attending the clinic of Professor Ghulam Mufti in the Department of Haematology at King's College Hospital (KCH) between 1999-2003 (**Table 2.1**). These samples were collected under the auspices of REC reference 02-04-049 and were frozen at -80^{o}C on receipt rendering them acellular. At that time and in accordance with the Human Tissue Act, these serum samples are classed as being cell-free and as such not "relevant material" or "bodily material" as defined by the act. In addition, the serum samples were given by donors who gave consent for their samples to be retained in perpetuity (KCH LREC 07-03-163 and 08/H0906/94) and there was local research ethics committee approval to use them (KCH LREC 02-04-049 and 08/H0906/94).

Sera from the sample labelled AML042 was collected from a patient attending Dr Kim Orchard's clinic in the Department of Haematology at the University Hospital Southampton NHS Foundation Trust under the auspices of the National Research Ethics Committee approved study titled 'Analysis of Leukaemia Associated Antigen Gene and Protein Expression in Haematopoietic cells and examination of their capacity to induce immune responses' REC reference 07/H0606/88. Samples AML006-41 were originally collected between 1999 and 2001, they had been stored in -80^{o}C since processing however some may have been defrosted during transportation (KCH – Southampton – Luton – Hull) and a single freezer breakdown over this period.

Table 2.1 Table of samples used in this study.

Sample	Age at sampling	Biological Sex	Date on sample tube	Subtype and cytogenetics	Risk Prognosis
AML006	61	Male	24.9.99	M4/tAML	Poor
AML010	79	Male	4.10.99	M2	Good
AML012	34	Female	12.10.99	M4/M5	Poor
AML016	48	Female	29.11.99	M4	Poor
AML022	58	Female	12.1.00	M5	Poor
AML027	85	Female	22.8.00	M3	Good
AML030	53	Male	11.8.00	M4	Poor
AML032	68	Male	14.6.00	tAML	Poor
AML036	69	Male	13.11.00	tAML	Poor
AML038	80	Male	15.12.00	46,XY [6]/45,XY, -7	Poor
AML039	17	Male	15.12.00	Relapsed	Poor
AML040	32	Male	8.1.01	M5	Poor
AML041	43	Male	23.1.01	M4	Poor
AML042*	38	Female	17.08.21	47 XX, +8, t(9;11)(p21;q23); AFLT3-TKD mutation c.2028C>G (Asn676Lys)	Poor
HV005	26	Male	16.05.08	NA	NA

tAML: AML transformed from MDS; *Leukocytes from AML042 were also examined, sera was examined from all AML samples in the table; only leukocytes were examined from healthy volunteer (HV)005.

Whilst there was limited knowledge on the cytogenetics of these patients, their FAB classification allowed us to infer the patients risk subgroups even though they may not be defined in the same way as the Davis et al group (i.e., based on WHO subgrouping) (**Table 2.2**). A previous study analysed the OS and EFS for different FAB subtypes with a patient cohort of 98 people, groups of FAB subtypes were generated with Group I containing M0-M2, group II containing M3 and group III containing M4-M7 (**Table 2.3**) (Padilha et al., 2015).

Table 2.2 Key clinical features of patients in the TARGET and TCGA datasets as analysed by Davis et al, 2020.

	Risk	Number	Cytogenetics	Age at diagnosis	EFS median, days (95% CI)	OS median, days (95% CI)
TARGET	Standard	31	MLL (n=21) t(8;21) (n=10)	637 (137-3542) days	293 (237-637)	1412 (712-NA)
	Low	31	inv(16) (n=15) t(8;21) (n=12) normal (n=4)	2154 (227-3539) days	558 (395 – NA)	NA (NA-NA)
TCGA	Poor	20	complex (n=18) poor risk (n=9) MLL translocation, poor risk (n=5)	60 (33-81) years	152 (119 – 454)	224 (122-801)
	Intermediate	41	normal (n=59) intermediate risk (n=15) MLL translocation t(9;11) (n=1)	60 (21-88) years	292 (244-420)	292 (341-822)

CI: confidence interval; NA: not available

Table 2.3 Table of FAB subtypes and associated risk. Information based on the Padilha et al study.

FAB subtype	Prognosis compared to average for AML
M0, M1, M2	Intermediate
M3	Favourable
M4, M5, M6, M7	Poor

2.2.6 Sample preparation from whole blood

Only one of the AML patient blood samples used in the qPCR was received via Dr Kim Orchard from patients who attend the Department of Haematology at Southampton University Hospitals Trust and processed at the University of Hull; however the other AML samples were all processed in the same way but received from Professor Ghulam Mufti and patients who attended the Department of Haematology at King's College Hospital NHS Foundation Trust (Guinn et al., 2005) at the time of receipt. The process for sample preparation from whole blood in ethylenediaminetetraacetic acid (EDTA) as a source of white blood cells, and clotted blood as a source of sera, was performed following a standard operating procedure/risk assessment/COSHH form for the processing in a class II safety cabinet using a bench top centrifuge.

2.2.6.1 Isolation of leukocytes

1x red cell lysis buffer (155mM NH_4Cl, 10mM $KHCO_3$, 0.1mM EDTA pH8.0) was added to ≤15ml whole blood sample to make the final volume up to 50ml. For 30 minutes, the sample was incubated at room temperature. The supernatant was decanted from the pellet of WBCs after centrifuging the material at 239 x g for 8 minutes or 486-536 x g for 5 minutes. If RBCs were still visible the red cell lysis step was repeated using 10-20ml of 1x RBC lysis buffer for 10 minutes, then centrifuging the sample as before. Again, the supernatant was decanted If there was a high

platelet count then the pellet was washed in 50ml phosphate buffered saline (PBS; pH 8.0: 137 mM NaCl, 3 mM KCl, 1.75 mM KH2PO4, 10 mM Na2HPO4) containing an additional 2 mM EDTA. It is then centrifuged at 200 x g for 10 minutes at room temperature and the supernatant decanted. Resuspend the cell pellet in 50ml of PBS before counting the cells using the trypan blue exclusion method (Section 2.2.8.1).

Cells were resuspended by flicking gently and adding X-VIVO-15^{TM} (BioWhittacker), 1% huAB sera (Sigma, Missouri, US) and 10% DMSO (Sigma) to a concentration of 5 x 10^6 cells/ml. Aliquot 1ml of cells in freezing media per cryovial and place in Mr Frosty which contained isopropanol to the fill line. Place the Mr Frosty in -80^{o}C for 4-16h. For long term storage they should be placed into liquid nitrogen.

2.2.6.2 Isolation of serum

In the absence of anticoagulant, serum samples were obtained and incubated at 370C for 20-30 minutes to allow for coagulation. The liquid portion of the sample was centrifuged at 3000rpm for 5 minutes to separate out the serum. The serum was aliquoted into 10 cryovials and store at -80^{o}C.

2.2.7 Tissue Culture

Cell lines were needed as positive controls while optimising the qPCR, so as not to waste the limited patient samples. For this we identified several cell lines that had been recorded in the literature to express the miRs of interest in this study (**Table 2.4**).

Table 2.4 miRs and their confirmed expression in cell lines.

miRNA expressed	Cell line expressed	Disease type	Paper	Methods
miR-486-1	K562	CML	(Ninawe et al., 2021)	qPCR
	UMUC3	Bladder cancer	(Tang et al., 2020)	Transfections, qPCR
	C2C12	Mouse myoblast	(Holstein et al., 2020)	qPCR
miR-1915	A549	Lung cancer	(Gao et al., 2019)	qPCR
	K562	Acute liver injury	(Qu et al., 2020)	qPCR
	SGC-7901, MKN45, GES-1	Gastric cancer	(Xu et al., 2019)	qPCR
	MDA-MB-231	Breast cancer	(Guo et al., 2018)	qPCR
miR-8086	None found			
miR-378G	MCF-7	Breast cancer	(Schultz et al., 2017)	Luciferase
	A2780	Ovarian cancer	(Liu et al., 2020)	qPCR

The following cell lines were chosen as controls for qPCR analysis due to their availability in our lab and each was sourced originally from ATCC. MDA-MB-231 (ATCCHTB-26) is a triple negative breast cancer cell line that lacks HER2 expression, progesterone receptors and oestrogen

receptors (ECACC, 2017). It was sourced from a pleural effusion of a 51-year-old Caucasian female with a metastatic mammary carcinoma (Cailleau et al., 1978). HL-60 (ATCCCCL-240) is a promyelocytic cell line generated from blood leukocytes collected by leukapheresis from a 36-year-old Caucasian girl with acute promyelocytic leukaemia (Birnie, 1988). Lozzio developed the cell line K562 (ATCCCCL-243) from a pleural effusion of a 53-year-old woman with chronic myelogenous leukaemia in terminal blast crises (Andersson et al., 1979).

Standard procedures were followed when working with cells, Howie style lab coats are worn with gloves. The category 2 laminar flow hood was wiped down twice with virusolve and twice with 70% ethanol before use. Before putting any item in the hood each item was sprayed with 70% ethanol. HL-60 and K562 were both non-adherent, whilst MDA-MB-231 was adherent cell line, and so the latter required an extra trypsinisation step to dissociate it from the flask. The cells were grown in T25, T75 and T175 flasks (Corning) allowing them sufficient space and media to grow. Adherent MDS-MB-231 cells were split when confluency reached around 60-80%, to avoid selective pressure being placed on the cell population by overcrowding and media depletion. HL-60 and K562 were cultured in RPMI 1640 (Gibsco, Thermofisher) containing 10% heat inactivated foetal calf serum (HI-FBS; Thermofisher) and 1% penicillin-streptomycin (PS; Thermofisher), this was referred to as RPMI culture media (CM). MDA-MB-231 cells were cultivated in Dulbecco's Modified Eagle's Medium (DMEM; Thermofisher), 10% HI-FBS and 1% PS, known as DMEM CM. Adherent cells had to be trypsinised before counting and/or passaging. To trypsinised the cells, the media was removed, the cells were washed with 1x PBS and 1 x volume of Trypsin EDTA (Gibsco) added. After 2-3min at 37^{o}C, 10 x the volume of DMEM containing FCS was added to inactivate the trypsin. The cells were then pelleted by centrifuging them for 5 minutes at 1200 x g in a 50ml falcon tube; and after dissociation of the pellet with gentle flicking, the cells are resuspended in the volume of DMEM CM to achieve the required cell concentration.

To count the number of live HL-60 and K562 cells, 10ul of cell culture taken for counting and added to 10ul of trypan blue (Thermofisher). After mixing 10ul of the cell suspension was placed onto a Neubauer Improved Haemocytometer Counting Chamber (Hawksley) (**Figure 2.1**). Clear (live) cells were counted within four squares and the count averaged.

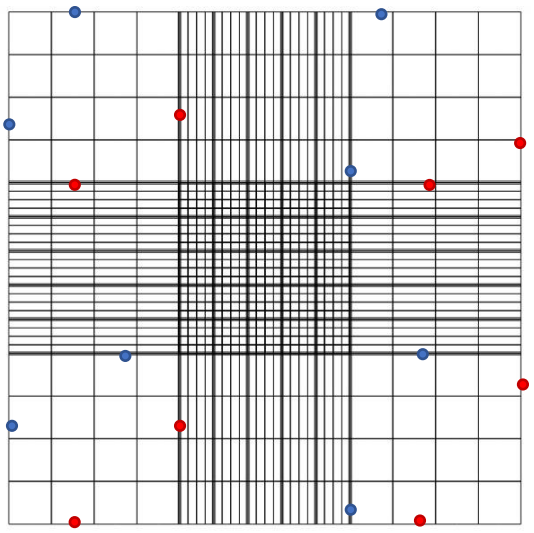


Figure 2.1 Haemocytometer example. Image of haemocytometer used for counting cell concentration. Blue cells are an example of what would have been included and red excluded.

The count was determined as follows:

$$\text{Average cell count per square x dilution factor x } 10^4 \text{ cells/ml}$$

If there were enough cells, then pellets of 3.5 million cells were prepared and immediately frozen in -80^{o}C or used immediately for RNA extraction.

2.2.7 RNA extraction

In anticipation of each experiment, all necessary precautions were taken to minimize contamination of samples; the bench was first wiped with 10% bleach followed by 70% ethanol to remove any DNA or RNA contamination. Gloves were changed regularly; filter tips were used, and RNase-free tubes used when needed.

2.2.8 miRNA extraction from leukocytes

miRNA was extracted from threesamples of leukocytes (**Table 2.1**). 1µl of UniSP4 RNA spike in was added to the 700µl of QIAzol lysis reagent before being added to the cell pellet (3.5 million cells). Following initial disruption via vortexing and centrifugation if the pellet hadn't broken down then the QIAshredder (Qiagen) column was used to further homogenise the cells. The homogenate was incubated for five minutes at room temperature. Then 140µl of chloroform was added to the tube and inverted vigorously for 15 seconds before being incubation for two minutes. After centrifuging for 15 minutes at 40°C, the upper aqueous phase was transferred to a fresh collection tube to ensure no interphase was present. A volume 1.5x of 100% ethanol was added and then mixed thoroughly up and down with a pipette, and up to 700µl was transferred into a RNeasy Mini spin column in a 2ml collection tube before centrifugation at 8000 x g for 15 seconds. The flow through was carefully discarded and the process was repeated with any remaining ethanol and RNA from the upper phase of the mix. 700µl of Buffer RWT was added to

the column and it was again centrifuged at 8000 x g for 15 seconds and the flow through discarded. 500µl of Buffer RPE was added and centrifuged at 8000 x g for 15 seconds. An optional step of inserting the spin column into a new 2ml collection tube and spinning at full speed for 1 minute was conducted to guarantee complete drying of the membrane. After that, the mini column was placed in a clean 1.5ml capped collection tube and 50µl of RNase-free water pipetted directly onto the column membrane before centrifuging at 8000 x g for 1 minute.

miRNA extracts from patient samples and cell lines needed the exact amount of RNA quantitating prior to complementary DNA (cDNA) synthesis, this was performed using the NanoDrop Lite (Thermo Scientific, MA, US). The ng/µl, 260/280 and A260 were all recorded for each sample. To preserve as much sample as possible, 1µl was loaded on to the optical measurement surface after being wiped with lint free tissue and RNAse free water. Each sample was loaded and measured twice to ensure a correct reading, if different then an average was used. The nanodrop functions by measuring the wavelength of light passing through the sample and measures the samples quality as a value of absorbance at OD260nm/absorbance at OD280nm, with ideal values being 1.8-2.0.

2.2.9 miRNA extraction from sera

The manufacturers protocol for the miRNeasy Serum/Plasma Advanced Kit (Qiagen) was followed, as it was a phenol free process and did not need to be performed in a fume hood. 200µl of sera was thawed and transferred to a new 2ml microcentrifuge tube, 60ul of buffer RPL per sample was added to a separate Eppendorf and 1µl of UniSP4 RNA spike-in was added. UniSP4 RNA spike-in was a synthetic RNA target added in known concentration to check for RNA isolation efficiency between samples (RNA Spike-In Kit, For reverse transcription (RT): Qiagen). 60µl of Buffer RPL was added to release and stabilise the RNA from plasma proteins and extracellular vesicles (EVs). 20µl of Buffer RPP was added to the mix before a vortexing for 20 seconds and after a short incubation at room temperature, to precipitate the proteins the material was centrifuged at 12000 x g for three minutes. The supernatant was transferred to a fresh tube, and an equal amount of isopropanol was added along with a short vortex to allow RNA molecules (>18 nucleotides) to bind to the silica membrane. There was then a series of washes to allow RNA to bind to the membrane and wash the contaminants away as follows. The sample was added to a RNeasy UCP MiniElute spin column and centrifuged at 8000 x g for 15 seconds, with the flow through discarded. The column was then filled with 700µl of Buffer RWT and centrifuged for 15 seconds at 8000 x g with the flow through discarded. 500µl of Buffer RPE was added to the column and was spun at 8000 x g for 15 seconds before being discarded, 500µl of 80% ethanol was then added, which was spun at 8000 x g for two minutes to wash the column membrane. On removal, care was taken to remove the column without contacting the flow

through so no carry over of ethanol occurred, to further dry the membrane, it was placed in a new 2ml collecting tube with the lid open and spun for five minutes at full speed (taking care to position the lids in the opposite direction to avoid any snapping). Residual ethanol may interfere with downstream experiments, so it was vital to completely dry the membrane. Finally, to elute the RNA the column was placed in a new 1.5ml collection tube and 20µl of RNase-free water was added directly on the membrane before being spun at full speed.

Following extraction if the total RNA extracted was not being used immediately it was clearly labelled in the RNase free tube it had been collected into in the previous step and stored in the -80^{o}C freezer, for cell samples the quantity of RNA was first measured.

2.2.10 cDNA conversion

The miRCURY LNA RT kit (Qiagen) was used for RT and was done according to the manufacturer's instructions. For the RT reaction, the RNA must be at 5ng/µl, so for samples in which the RNA content was much higher than the RNA had to be diluted with RNAse-free water.

When calculating the amount of template RNA to use from sera RNA extractions we followed the recommended formula which uses elution volume and original serum volume as opposed to the Nanodrop RNA concentration which is used to calculate the volume of RNA extraction added for peripheral blood lymphocytes (PBLs).

Template RNA (µl) = (Elution Volume (µl) / Original Sample Volume (µl)) x 8

0.8ul = (20µl / 200µl) x 8

The reagents (**Table 2.5**) were added to nuclease-free 200µl PCR tubes on ice. To avoid waste, all reagents were mixed and centrifuged briefly before use. All preparation work was performed on ice to preserve the RNA, the 10x miRCURY Enzyme mix was only removed from the freezer prior to immediate use. The RNA was then be added to the PCR machine and the temperature was programmed as listed (**Table 2.6**).

A no template control (NTC) had the same reagents added with more RNase-free water added as a substitute for template RNA, this would then be used as the NTC in RT-qPCR to check for contamination.

Once the RT reaction has completed the cDNA can be placed into a -20^{o}C freezer where it should be used within 5 weeks. The cDNA was aliquoted into 0.5µl PCR tubes for easier use during qPCR.

Table 2.5 Quantities of reagents in cDNA synthesis for sera.

Component	Volume
5x miRCURY RT Reaction buffer	2.0 µl
RNase free water	5.7 µl
10x miRCURY RT Enzyme Mix	1.0 µl
Synthetic RNA spike in (Uni SP6)	0.5 µl
Template RNA	0.8 µl
Total volume	10 µl

Table 2.6 Temperatures of cDNA synthesis.

Step	Time	Temperature (^{o}C)
Reverse-transcription	60 minutes	42^{o}C
Inactivation of reaction	5 minutes	95^{o}C
Cooling	Till removed (Max 1 hour)	4^{o}C
Storage	5 weeks	-20^{o}C

2.2.10 Primers and assays

In the RNA-seq dataset, the Ensemble Gene ID was provided but no information on which variant of the miRNA it could be, so we tested both 3p and 5p variants. On receiving the primers (**Table 2.7**) in their lymphosised format, they are first spun briefly and then 220µl of nuclease-free water was added, followed by vortexing and brief centrifugation to mix the primer and then left for 20 minutes before being aliquoted at 14µl into multiple PCR RNase-free tubes and stored at -20^{o}C.

Table 2.7 Primer sequences

Primer Target	GeneGlobe ID
hsa-miR-486-3p	YP00204107
hsa-miR-1915-3p	YP02101780
hsa-miR-1915-5p	YP02119687
hsa-miR-8086	YP02114444
U6 snRNA	YP00203907
5s rRNA	YP00203906
hsa-miR-Let-7i-5p	YP00204394
hsa-miR-222-3p	YP00204551
UniSP4	YP00203953
UniSP6	YP00203954

2.2.11 Real time-quantitative polymerase chain reaction

The miRCURY LNA SYBR Green PCR kit (Qiagen) was used as the LNA technology can quantify extremely low levels of miRNA without pre-amplification, this was ideal when dealing with serum samples that may have lower RNA content. All reactions were set up on ice to minimise the danger of any deterioration, even though it was not necessary due to the hot start nature of the PCR kit. In our experiment to discount for any interplate variation all 14 samples were tested

for a single miRNA primer on one PCR plate or some had two primers on, however all the samples were on the same plate. A NTC created during the cDNA synthesis was used to allow us to observe any contamination throughout either of the processes.

Thermofisher's StepOne Real-Time PCR machine requires a high 20x concentrated solution of ROX, 0.5µl of ROX dye was used. The protocol for cDNA was to dilute it 1:30 by adding 290µl RNase-free water to 10 µl of RT product, however only a small amount was needed each qPCR reaction and as the cDNA had already been aliquoted into 0.5µl tubes only 14.5µl of RNase free water was added each time. This allows for more long-term storage of each product.

A master mix (**Table 2.8**) for each primer was made containing 10% more reagents than required to account for any pipetting errors, it was mixed and centrifuged briefly. 7µl of the master mix containing everything except the diluted cDNA is added to the plate first, before 3µl of the diluted cDNA was added to the plate and the seal firmly pressed on top and flattened with the edge of a tip. The experiment could then be paused for a small time if needed by being placed in fridge to keep it cool and out of light due to the photosensitive nature of SYBR green. When ready to be placed into the StepOne qPCR machine the plate was first centrifuged in a plate spinner (Labnet, New Jersey, US) to ensure the mixture was at the bottom of the well for better readings.

Table 2.8 RT-qPCR reagents and volumes.

Component	Volume
RNase free water	0.5 µl
ROX Reference dye	0.5 µl
2x miRCURY SYBR Green Master mix	5 µl
Cdna template	3 µl (diluted 1:30)
PCR primer mix	1 µl
Total volume	10µl

The qPCR machine then followed the cycles on a maximal/fast ramp rate with a run lasting around 1 hour 45 minutes (**Table 2.9**). Melt curves were performed for each run to evaluate performance of primers, ideally there was only a single product being detected and the melt curve slowly increases the temperature each cycle until the fluorescence decreases as the bonds break.

Table 2.9 RT-qPCR thermocycling conditions.

Step	Time	Temperature (°C)
PCR initial heat reaction	2 minutes	95°C
2-step cycling: Denaturation	10 seconds	95°C
Combined annealing/extension	60 seconds	60°C
Number of cycles	40	
Melt curve	60-95°C	

2.2.12 Normalization selection

Normalisation was required to minimise inter-sample variation between biological and technical replicates. There is no agreed upon standard for miR normalisation, however the use of commonly expressed endogenous controls was standard practice when trying to observe miR expression in PBLs. Whilst having different compositions to miRNAs, 5s rRNA and U6 snRNA are often used in cell samples due to their consistent expression. However, for use as normalisers analysing serum samples the previously mentioned normalizers U6 snRNA (Benz et al., 2013) and 5s rRNA (Lim et al., 2011b) would not work and so we used Qiagen's recommended stably expressed normalizers miR-222-3p and mir-Let-7i-5p (*Guidelines_for_profiling_biofluid_miRNAs.pdf*). To further normalise the data, UniSP4 spike in was added during RNA extraction to observe the efficiency of RNA extraction and a spike in of known volume was added (UniSP6), this could be used as another variable to observe the cDNA efficiency. In the miRCURY LNA miRNA PCR starter kit miR-103-3p was included, that could also have been used to normalise the results, however the two used were already consistent.

2.2.13 Analysis of results

In this experiment we analysed the Cycle threshold (Ct) value or the point in which the fluorescence crosses that of the background fluorescence crated by the ROX dye, the lower a Ct value the greater amount of miRNA in the sample. Technical replicates (triplicate wells) were first checked for any significant differences between the three wells to allow adjustment of the mean if one value was significantly different to the others. Any results with more than one Ct value as undetermined or above 37 are excluded as was carried out in previous tests with miRNA following the protocol in (Caserta et al., 2018). Ideally with a large enough array of miRs being analysed, those with consistent expression would be used to create a Normaliser value (NV). Within the constraints of this experiment two endogenous control miRs and a spike in has been used.

The Ct value for miRs of interest was normalized to the average Ct of both endogenous controls (miR-222-3p and miR-Let-7i-5p) and the UniSP6 spike in Ct to calculate the NV, this should

account for any variation in the RT-efficiency. The Ct of miRs would then be subtracted from the NV on a sample basis to provide the ΔCt, which equals the difference between Ct of the miR assay and the NV.

Despite the limited number of samples, it was possible to group the patients via their risk subgroup based on FAB subtype and compare the ΔCt values between the groups. Using GraphPad Prism and SPSS, the results were analysed using correlation and linear regression analysis to determine the impacts of three variables (patient age, time since sample collection, and biological sex) on the expression levels of all miRs and the NV. The correlation analysis is performed to test if there is a relationship between two variables such as the age of patient or date sample taken versus the miR expression, the linear regression analysis between these two variables shows the same however visualises the relationship with a predictive line. The non-parametric Mann-Whitney test was employed to check significant differences between two groups such as poor versus good or male versus female due to the nature of our groupings (biological sex and risk group) not having enough samples to run parametric testing.

Chapter 3. Examination of the pathways utilised by differentially expressed genes in AML

Data shown in this chapter has been published in part in Greiner J, Brown E, Bullinger L, Hills RK, Morris V, Döhner H, Mills KI, Guinn BA. Survivin' Acute Myeloid Leukaemia-A Personalised Target for inv(16) Patients. Int J Mol Sci. 2021 Sep 28;22(19):10482. doi: 10.3390/ijms221910482 (appendix II).

3.1 Introduction

Davis et al discovered that the expression of nine LAAs BIRC5, CEACAM3, MAGEF1, STEAP1, VGLL4, MORC4, SLC34A2, SAGE1 and MELTF differed significantly between standard and low risk paediatric AML patient subgroups, as well as between poor, intermediate and good risk adult AML patient subgroups. Furthermore, when comparing standard versus low-risk subgroups and quartile expression levels, BIRC5, MAGEF1, MELTF, STEAP1 and VGLL4 were shown to be differentially expressed with reference to EFS in paediatric AML patients. There were significant differences in EFS in adult AML patients when comparing intermediate versus good quartile expression levels of MORC4, and poor versus good risk quartile expression levels of SAGE1. Following the examination of KEGG pathway data, there were several genes altered in AML cells involved in key processes in AML such as the control of cell proliferation (SSX Family Member 2 Interacting Protein, AML1-ETO) and the evasion of apoptosis (BIRC5, WNT1). Based on these findings we examined these proteins and the pathways further.

3.2 Aims

To determine the pathways that the GOI are involved in, where possible, and specifically those relating to AML.

- Discover common and unique pathways utilised in adult and paediatric AML.
- Identify potential targets for therapeutic treatment in both cohorts

3.3 Results – analysis of nine LAAs

3.3.1 Review of the Literature regarding the GOI

To determine which were worthy of further investigation and to see whether they shared molecular pathways indifferent of subtype and whether they were important in paediatric or adult AML, a comprehensive search of the literature was used to understand their role in cancer. If there was no literature on the effects in leukaemias then other diseases were evaluated.

3.3.1.1 BIRC5

BIRC5 was discovered to play an important role in cell cycle progression and mitosis in healthy cells, interacting with the chromosomal passenger complex (CPC) and Aurora-B kinase in the nucleus to ensure proper mitotic spindle formation. Supported by the findings of STRING analysis revealing substantial gene correlations to other proteins involved in the production of mitotic spindles (**Figure 3.1**), as well as data provided by Professor Ken Mills on correlation of genes and background reading (Vader et al., 2006).

Aurora-B kinase activity is essential for cytokinesis to occur, and its activity is tightly regulated by its binding partners. It shows a high expression level in the G_2M phase of mitosis (Altieri, 2003), survivin depleted cells exit mitosis without correct chromosomal alignment (Castedo et al., 2004). The incorrect chromosomal alignment would then lead to the eventual cell death. BIRC5 has higher expression in BM cells compared to peripheral blood. BIRC5 belongs to the Inhibitor of Apoptosis (IAP) family of proteins, which binds to caspases 3 and 7, inhibits their activity, and prevents cell death induced by CD95 (Tamm et al., 1998). Its expression has found to be regulated by p53 with its downregulation leading to greater levels of apoptosis (Mirza et al., 2002; Wang et al., 2004). The activation of p53 leads to enhanced activation of apoptotic signalling pathways. Survivin also competes with the cyclin dependant kinase (Cdk)4/p16INK4a complex, suggesting that survivin starts cell cycle entry following nuclear translocation and an interaction with CDK4 and survivin suppresses the ability of p16INK4a to arrest the cell cycle (Suzuki et al., 2000).

When Wnt binds with the frizzled receptor on the surface of the cell it stops the degradation of β-catenin and causes an accumulation inside the cell (Liu et al., 2022). The Wnt/-catenin pathway stimulates transcription of several downstream targets, one of which is thought to be BIRC5. If this pathway gets disrupted, survivin expression would be reduced as it won't be transcribed (Chen et al., 2016). As a result of the cytokinetic disturbances, this could lead to higher levels of apoptosis and cell death.

A previous study showed evidence that MYC Proto-Oncogene (C-Myc) and Sp-1 helped modulate drug resistance via survivin expression regulation, achieved through the RK/MSK mitogen-activated protein kinase (MAPK) controlling C-Myc and Sp-1 levels (Zhang et al., 2015). Independent studies also found BIRC5 to be a poor prognostic indicator in paediatric AML (Tamm et al., 2004; Esh et al., 2011). Survivin and its splice variants were found in 83.8% of adult AML (Wagner et al., 2006).

(A)

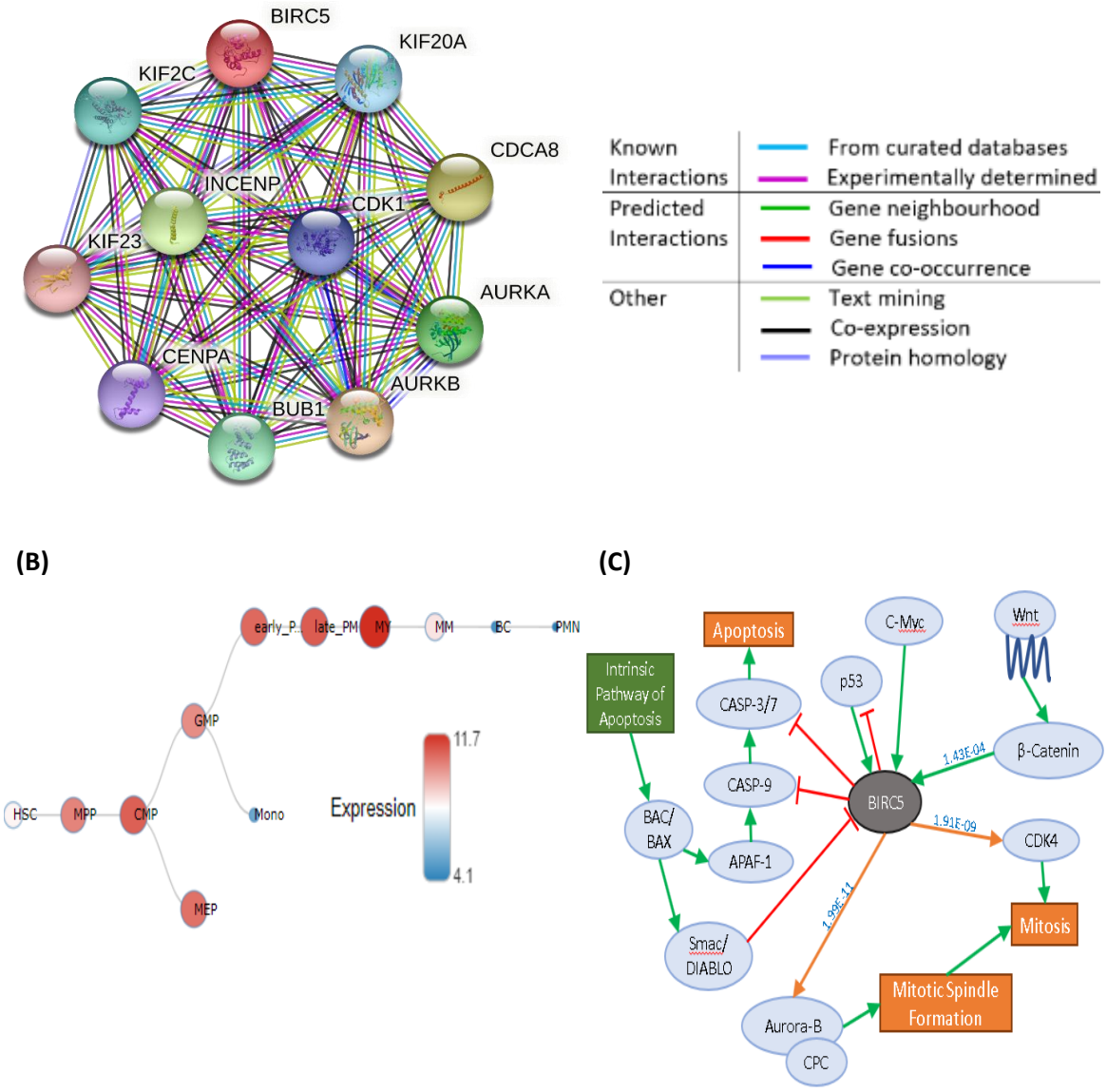


Figure 3.1 BIRC5 expression in health and disease. (A) STRING research revealed that BIRC5 interacts with other genes involved in mitotic regulation (CDK1), particularly the CPC (INCENP, KIF20A, CDCA8, AURKB, and AURKA), and microtubule formation (KIF2C and KIF23). As well as mitotic checkpoint partner (BUB1). [Accessed 14/02/22]**; (B)** The BIRC5 expression in healthy FACs sorted blood cells was examined using the BloodSpot programme. The highest levels of expression were seen in myelocytes, late promyelocytes, early promyelocytes, common myeloid progenitors, megakaryocyte/erythroid precursors, and multipotent progenitors. Band cells and polymorphonuclear cells showed reduced expression, while haematopoietic stem cells and metamyelocytes showed no expression (PMN); **(C)** Interactions between BIRC5 and other proteins based on peer-reviewed published research: caspases (Li et al., 1998; Tamm et al., 1998; Shin et al., 2001); p53 (Mirza et al., 2002; Wang et al., 2004; Sah et al., 2006); C-Myc (Fang et al., 2009); Wnt/β-catenin (Stewart, 2014); CDK4; (Suzuki et al., 2000); mitotic spindle formation (Castedo et al., 2004; Vader et al., 2006); BAC/BAX/DIABLO (Tamm et al., 1998; Song et al., 2003). The p-values of correlation analysis between some of the genes after analysis in the MILE dataset are shown above the arrows. (Haferlach et al., 2008)(DOI: 10.5281/zenodo.4923748).

3.3.1.2 CEACAM3

CEACAM3 is involved in bacterial phagocytosis, which is vital for healthy cell function. CEACAM3 does not facilitate cell-cell adhesion like the rest of the CEACAM/CD66 family, but it does

mediate the opsonin-independent identification and destruction of pathogenic bacteria (Pils et al., 2008). Once bound, the Syk cytoplasmic tyrosine kinase is recruited, and both proteins are phosphorylated (Sarantis & Gray-Owen, 2007). Tyrosine protein kinase HCK is part of the Src family of tyrosine kinases (**Figure 3.2A**), CEACAM3 triggers its recruitment and activation (Bonsignore et al., 2020). It is a type of immunoglobulin-related receptor expressed on the surface of human granulocytes (Buntru et al., 2011). It recognises gram negative bacteria and once bound promotes rapid internalisation and destruction of the CEACAM binding bacteria (Buntru et al., 2012). Phosphatidylinositol 3 kinase (PI3K) binds to CEACAM3 in phosphorylated axis upon binding with bacteria, stimulating the formation of reactive oxygen species (ROS) necessary for destruction of the bacteria (Buntru et al., 2011). Another group found CEACAM3 to be differentially expressed in leukocyte membranes, following PCR they found elevated CEACAM3 expression during CML blastic crisis (Luo et al., 2012). Using the results of microarray analysis, a group was able to determine that growth factor receptor-bound protein 14 (Grb14) bound to CEACAM3 with the SH2 domain and negatively regulated it, reducing phagocytosis of bacteria (Kopp et al., 2012).

3.3.1.3 MAGEF1

MAGEF1 is part of a family of MAGE proteins that have similar functions and are all cancer-testis antigens, MAGE family members were found to bind to the tripartite motif containing (TRIM) family (Doyle et al., 2010). The activity of the RING-type zinc finger-containing E3 ubiquitin protein ligase has been discovered to be enhanced by MAGEF1, it also promotes the ubiquitination and degradation of MMS19, a crucial component of the cytosolic iron-sulfur protein assembly (CIA). MMS19 down regulation inhibits DNA repair enzymes such excision repair 2 (ERCC2), which rely on iron-sulfur clusters (Weon et al., 2018). It has been correlated with genome wide demethylation and has higher expression in the testes (De Smet et al., 1996). Unlike the MAGEG1 family member, MAGEF1 does not play a part in the structural maintenance of the chromosome (SMC5-6) complex, although it can form complexes with interacting inhibitor of differentiation (EID) proteins (**Figure 3.2B**) (Hudson et al., 2011). MAGEF1 could control the CIA pathway by the degradation and ubiquitination of CIA-targeting protein MMS19 (Weon et al., 2018).

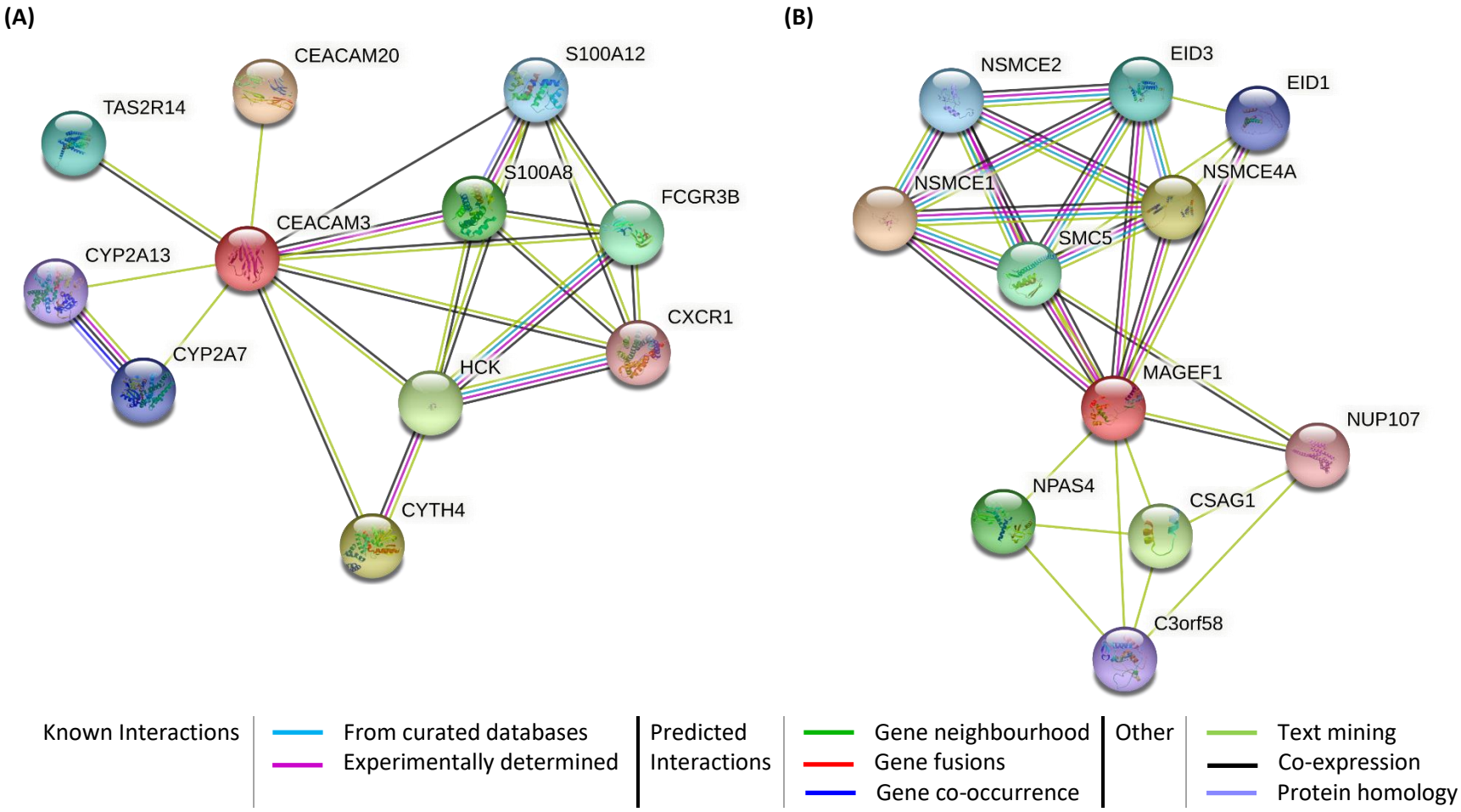


Figure 3.2 Protein-protein interactions between CEACAM3 and MAGEF1. (A) CEACAM3 is predicted to interact with of carcinogen related adhesion molecules (CEACAM20). It is also co-expressed with a tyrosine protein kinase (HCK) found in haematopoietic cells. It is co-expressed with two calcium and zinc binding proteins (S100A8 and S100A12) that regulate inflammatory processes and immunological responses, as well as an immunoglobulin gamma Fc receptor (FCGR3B) and CYTH4, that promotes activation of ARF factors. [Accessed 14/02/22]. **(B)** MAGEF1 has predicted interactions with genes involved in the maintenance of the chromosome elements (MSCMCE4A, NSCMCE1 and SMC5) and inhibitors of differentiation (EID3 and EID1). [Accessed 14/02/22].

3.3.1.4 STEAP1

STEAP family members are metalloreductases **(Figure 3.3A**) with primary roles in the endocytic and exocytic pathways (Gomes et al., 2012). Most of the STEAP family of proteins are able to reduce iron and copper using NAD+ as an acceptor however STEAP1 is missing the FNO-like domain that allows them to do so (Ohgami et al., 2006), (Kleven et al., 2015). STEAP1 does still have a single beta heme and so is still a novel metal reductase but not to the same extent as other family members (Kim et al., 2016). Although STEAP1 is significantly expressed in a number of malignancies including bladder, colon, ovarian, prostate, and Ewing sarcoma, no other evidence for its overexpression in AML has been reported (Moreaux et al., 2012). STEAP1 was found to be co-expressed with trophoblast glycoprotein in lung cancer (Zhuang et al., 2015). There is evidence that phosphorylated eukaryotic translation initiation factor 4E regulates STEAP1 at a cap-dependent level (Jiang et al., 2020). Silencing of STEAP1 in colorectal cancer cells increased production of ROS and slowed cell growth, it was identified as acting along the nuclear erythroid 2-related factor NRF2 pathway with inhibition of STEAP1 causing less antioxidants regulated NRF2 (Nakamura et al., 2019).

3.3.1.3 MORC4

The MORC4 protein has a coiled-coil region and a HATPase-c domain with a CW zinc finger motif, nuclear localization signals, and a nuclear-binding matrix (Hong et al., 2017). The zinc finger motif may facilitate protein-protein or protein-nucleic acid interactions, and the N-terminus may signal a necessity for ATP hydrolysis (**Figure 3.3B**). The CW zinc finger could indicate that it can bind to methylated DNA and MORC4 may act as a transcription factor with a role in DNA recombination or repair, it was also overexpressed in 66% of breast cancers (Liggins et al., 2007). MORC4 potentially reduces miR-193B-3p mediated inhibition of apoptosis in breast cancer cells, this would link to higher expression correlating with patient who had lower OS rates (Yang et al., 2019). The serine proteases (PRSS) family and genetic variations in CLDN2-MORC4 were discovered to be related with acute pancreatitis (Weiss et al., 2018). MORC4 inhibition resulted in decreased enrichment of STAT3 binding midline 2 promoter regions in luminal A/B breast cancer tumour cell lines, while boosting MORC4 expression boosted B-cell lymphoma 2 (BCL-2) expression and increased treatment resistance (Luo et al., 2020). Whilst testing the properties of a natural drug, Baicalin, MORC4 was discovered to be a target of miR-338-3p, overexpression of this miR suppressed cell metastasis and lead to apoptosis (Duan et al., 2019).

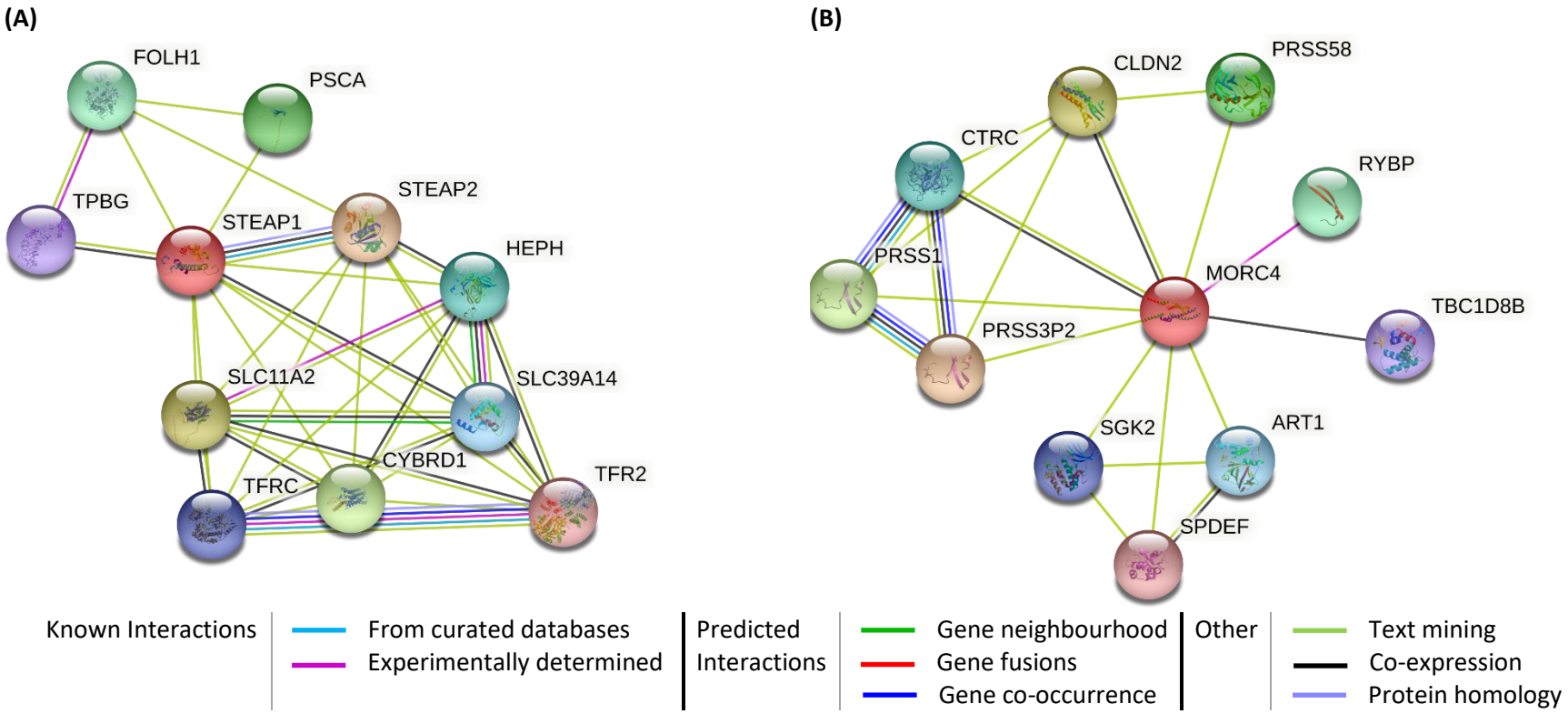


Figure 3.3 Protein interactions between STEAP1 and MORC4. (A) STEAP1 has potential interactions with genes involved in the reduction of iron (STEAP2, CYBRD1 and HEPH) and other metal ion transporters (SLC11A2 and SLC39A14). Along with some cell proliferation regulators (PSCA and TPBG). [Accessed 27/03/21]. (B) MORC4 is known to interact with YY1 associated functional partner (RYBP), some members of the S1 peptide family (PRSS3P2, PRSS1 and PRSS58) as well as associated with a gene playing a major role in the junction-specific obliteration of intracellular space (CLDN2). Text mining also found links between MORC4 and a protein that regulates trypsinogens (CTRC). [Accessed 27/03/21].

3.3.1.3 VGLL4

The biological functions of VGLL4 occur via interaction with TEA domain (TEAD) transcription factors (Deng & Fang, 2018). It is a transcriptional cofactor that competes with Yes-associated protein (YAP) to interact with TEADs, influencing the rate of transcription from DNA to mRNA and regulating downstream genes (**Figure 3.4A**) (Zeng et al., 2017b). VGLL4 is also phosphorylated by CDK1 during G2M arrest and normal mitosis (Zeng et al., 2017b). At the transcription factor level, Wnt/β-catenin and Hippo-YAP signalling are closely connected, VGLL4 can target the TEAD4-Transcription Factor 4 (TCF4) complex to co-regulate both pathways (Jiao et al., 2017). In gastric cancer, VGLL4 inhibits B-catenin and TCF, negatively regulating the Wnt/B-catenin signalling pathway and VGLL4 overexpression increases E-cadherin (CDH1) levels while decreasing B-catenin levels, suppressing epithelial-mesenchymal transition (Li et al., 2015). Inhibition of miR-130a prevented YAP-induced tumorigenesis and prevented YAP-induced liver enlargement typically caused by the Hippo pathway inactivation, this was thought to be related to the repression of VGLL4 (a YAP inhibitor), which would then boost YAP signalling further (Shen et al., 2015). USP11 interacts with VGLL4 and promotes its deubiquitylation, this enhances its stability (Zhang et al., 2016). It was recently shown that VGLL4 stabilises interferon regulatory factor 2-binding protein 2 (IRF2BP2) and a decrease in VGLL4 leads to a decrease of programmed death-ligand 1 (PD-L1), this suggests it regulates PD-L1 in some manner (Wu et al., 2019). A lack of IRF2BP2 prevents transcription of Cd274 caused by the IFN-γ cytokine, suggesting VGLL4 is an inhibition target that indirectly promotes Cd274 transcription by stabilising IRF2BP2 (Pastor et al., 2021).

3.3.1.3 SLC34A2

Solute Carrier Family 34 Member 2 (SLC34A2) protein is a sodium dependant phosphate transport protein, it actively transports phosphate into cells via Na(+) cotransport, it is predominantly active when there's low levels of Pi in the diet (Marks, 2019). In hepatocellular carcinoma (HCC) cells, knocking down SLC34A2 dramatically reduced the levels of phosphorylated PI3K and Threonine Kinase 1 (AKT). Taken together, these findings imply that silencing SLC34A2 decreases HCC cell proliferation and migration through inhibiting the PI3K/AKT signalling pathway (Li et al., 2016). The suppression of this pathway may be due to interactions with phosphatase and tensin homolog (PTEN) (Liu et al., 2017). In glioma cells, knocking down SLC34A2 inhibited the activation of the epidermal growth factor receptor/PI3K/AKT signalling pathway (Bao et al., 2019). There is evidence SLC34A2 is a fusion partner for ROS1 (a receptor tyrosine kinase (RTK)) in a small number of non-small cell lung cancers (**Figure 3.4B**) (Lin & Shaw, 2017) and in a lung tumour model cell line (A549), SLC34A2 inhibited tumour growth and metastatic ability (Wang et al., 2015b).

In neuroblastoma, SLC34A2 was shown to be implicated in a pathway that inhibits Wnt signalling; it binds to a promoter of miR-25, which then binds to glycogen synthesis kinase 3 (Gsk3) (Chen et al., 2019). Overexpression of SLC34A2 is a prognostic marker in bladder cancer, following its inbibtion tumour growth was reduced potentially due to a decrease in expression of C-Myc (Ye et al., 2017), this correlates with the correlation values we generated.

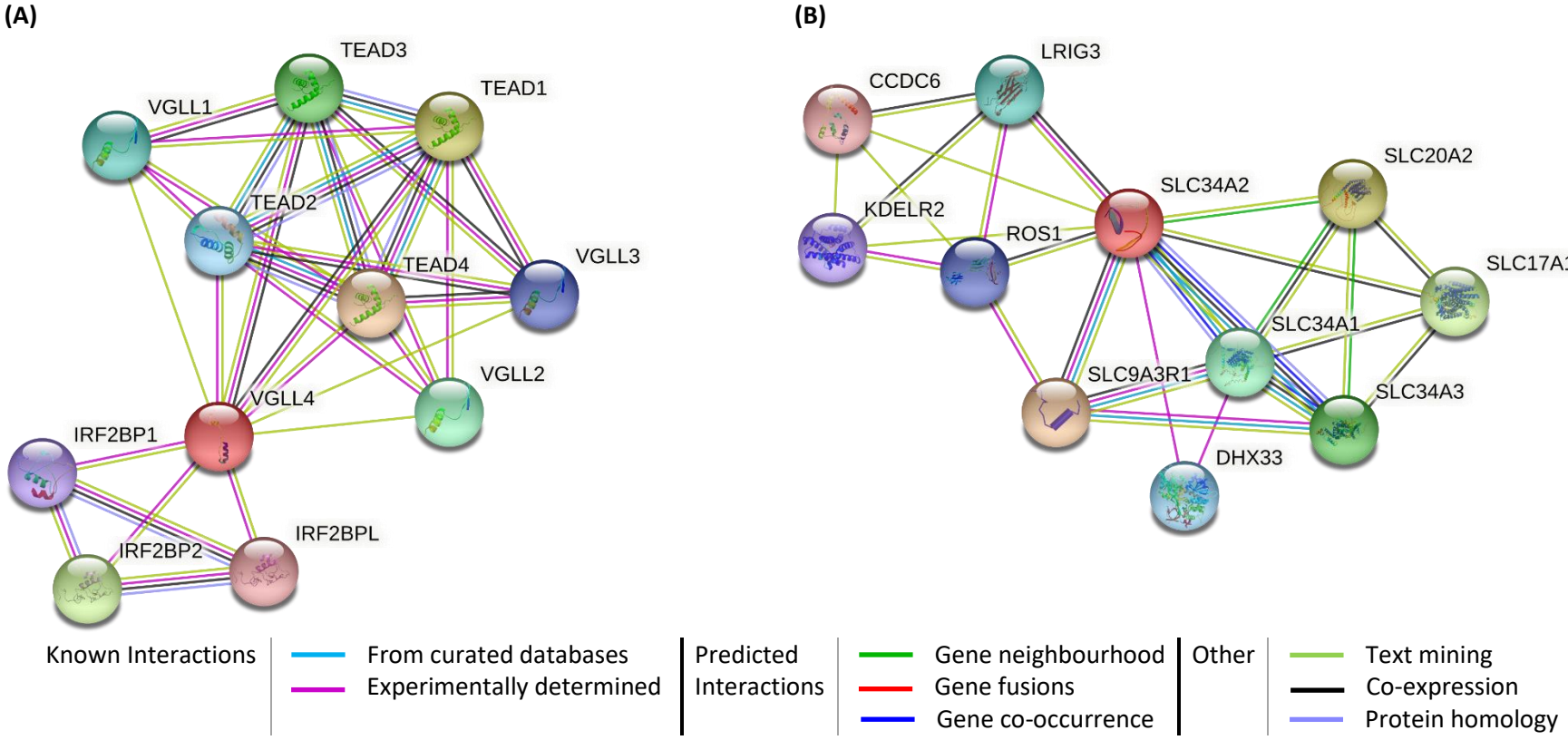


Figure 3.4 Protein interactions of VGLL4 and SLC34A2. **(A)** VGLL4 interacts with genes involved in the Hippo signalling pathway (TEAD1, 2, 3 and 4), other coactivators of TEFs (VGLL1, 2 and 3) and interferon binding proteins acting as transcriptional inhibitors (IRF2BPL1, 2 and L). [Accessed 27/03/21]; **(B)** SLC34A2 interacts with functional partners involved in other sodium dependant transport (SLC20A2, SLC17A1, SLC34A1, SLC9A3R1 and SLC34A3). It also is predicted to be co-expressed with RTK proteins (ROS1). [Accessed 24/11/21].

3.3.1.3 SAGE1

Sarcoma antigen 1 expression (SAGE1) is restricted to male germ-line cells and tumour cells this means it potentially codes for tumour-specific antigens recognized by T lymphocytes (Maheswaran et al., 2015). Activation of MAGE genes in tumour cells could be caused by genome-wide demethylation that often occurs in many cancers (De Smet et al., 1996). STRING analysis found SAGE1 to co-express with sperm surface binding proteins, supported by the origin of spermatocytic seminoma (SS) (a testicular germ cell tumour) and provides evidence for heterogeneity of this tumour (**Figure 3.5A**) (Lim et al., 2011a). SAGE1 expression was found in 50% of non-small cell lung cancer samples (Zhang et al., 2021) and in 33% of oesophageal carcinomas (Chen et al., 2014), NXF2 was also found to be expressed in here, hence why STRING has highlighted them being co-expressed.

3.3.1.3 MELTF

Melanotransferrin (MELTF) is a transferrin homolog discovered attached to the cell membrane, its function is to bind iron via a single high affinity iron (III) (Dunn et al., 2006). The binding site is attached via a glycosylphosphatidylinositol anchor, however the binding of iron may not be its main role (Suryo Rahmanto et al., 2007). It can only bind one Fe atom; a study published in 2000 found that removing MELTF had no effect on Fe uptake. Its alternate function could be binding with urokinase-type plasminogen activator to generate plasmin; the plasmin produced could cause extracellular degradation and endothelial cell detachment, leading to tumorigenesis and causing the increased cell metastasis observed (**Figure 3.5B**) (Michaud-Levesque et al., 2005). MELTF may also play a role in the vascularization of solid tumours and the progression of melanoma (Tiklová et al., 2010). It could be a potential biomarker in gastric cancer (GC), inhibition of MELTF reduced the invasiveness of GC cells and increased mRNA expression in tissue leading to shorter survival and a poorer prognosis (Sawaki et al., 2019).

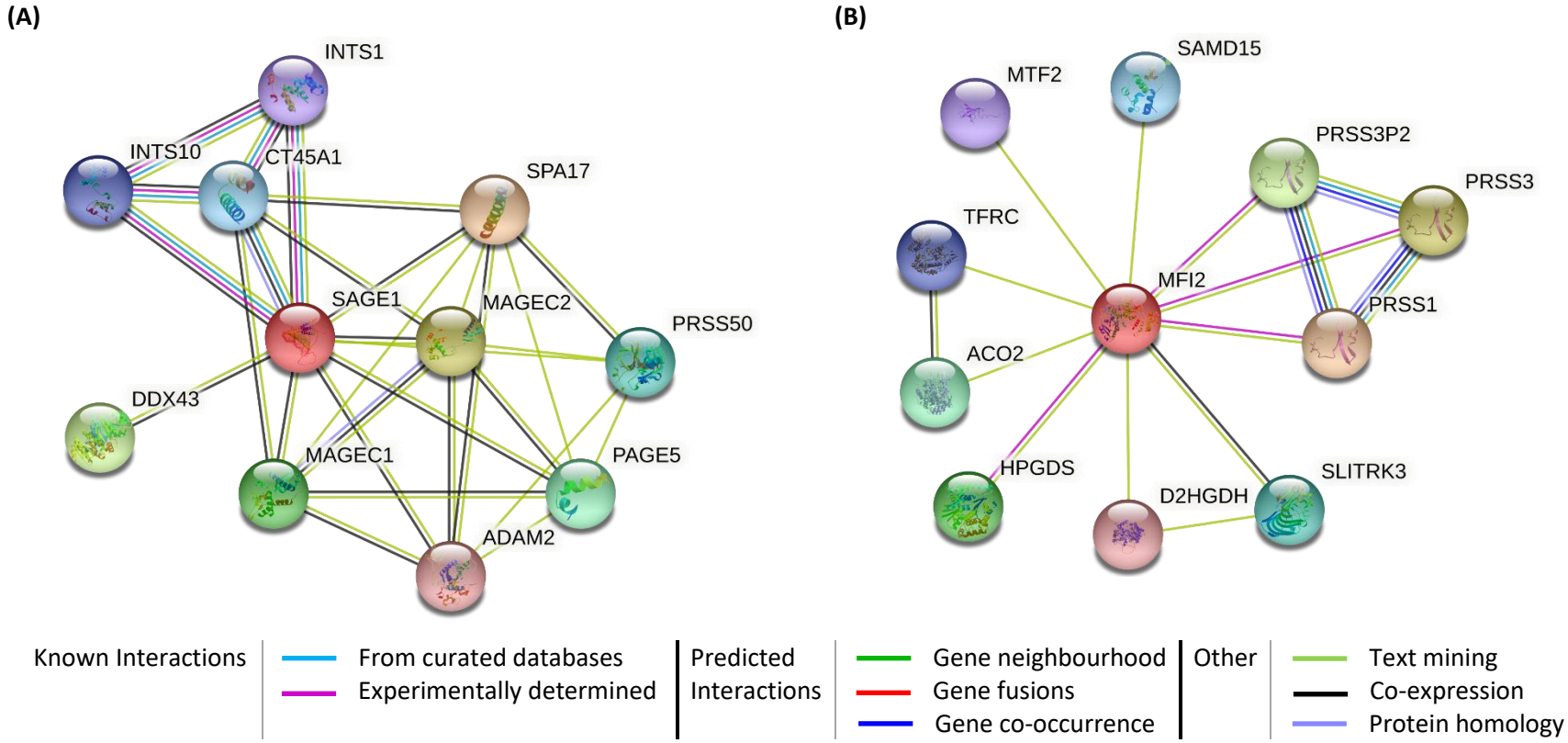


Figure 3.5 STRING images of SAGE1 and MELTF. (A) SAGE1 is predicted to interact with some sperm surface binding proteins (SPA17 and ADAM2), components of integrator complex involved in snRNA transcription (INTS1 and INTS10) and some melanoma associated antigens proposed to enhance ubiquitin ligase activity (MAGEC2 and MAGEC1). [Accessed 11/02/21]. **(B)** MELTF was found to interact with members of the S1 peptide family, which are involved in trypsin activities (PRSS1, PRSS3 and PRSS3P2). MELTF has also been identified to interact with HPGDS, an enzyme that catalyses the conversion of PGH2 to PGD2, as well as being co-expressed with SLITRK3, a synaptic cell adhesion molecule. [Accessed 11/02/22].

3.4 Association between genes and clinical features

Analysing the relationship between BIRC5 probesets and adult AML clinical features revealed significant differences in BIRC5 expression between genders, NPM mutation and wild type, FLT3-ITD and FLT-WT, M6 and all other FAB subtypes (M0 – M5, M7) and M7 and all other FAB subtypes (M0 – M6) (**Table 3.1**). However, no association was found between above and below median BIRC5 expression in adults with AML and OS (data not shown) except when looking only at the patients with an inv(16) (Greiner et al., 2021).

Table 3.1 Association between patient clinical features and BIRC5 in adults with AML (MILE).

BIRC5	P-values															
		vs. M6						Vs. M7								
Probeset ID	FAB	M0	M1	M2	M3	M4	M5	M0	M1	M2	M3	M5	M6	Gender	NPM	FLT3
1555826_at	0.0017	3.7E-06	6.4E-06	1.4E-05	4.3E-05	1.6E-05	1.2E-04	NS	NS	NS	NS	NS	1.5E-04	NS	NS	NS
202094_at	1.9E-04	4.2E-07	1.3E-06	1.0E-06	1.4E-05	6.1E-06	6.6E-06	0.028	NS	NS	NS	NS	0.019	0.0013	NS	NS
202095_s_at	NS	0.0022	0.013	0.0078	0.011	0.011	0.0060	0.034	NS	NS	NS	NS	NS	0.0093	0.021	NS
210334_x_at	6.9E-04	3.2E-05	6.5E-05	1.1E-05	1.6E-05	8.5E-05	8.9E-05	0.023	0.047	0.016	0.016	0.049	NS	0.0023	0.043	0.028

NS: not significant

3.5 Pathway analysis of nine LAAs

Multiple programs such as KEGG (Section 2.1.5), STRING (Section 2.1.3 and accessed on/up until 11/02/22), WEB-based Gene Set Analysis Toolkit as well as a comprehensive literature review were used to establish proteins that interacted with the LAAs. As different probe sets were available for BIRC5, when possible, the following were used 202094, 208052 and 212399. A pathway map was made with predicted or known functional partners of the LAAs and connecting pathways in leukaemogenesis (**Figure 3.6**), correlation analysis was carried out to confirm the strength of the relationship between some LAAs and other genes using the MILE dataset (GSE13159).

3.6 PANTHER analysis of differentially expressed genes – from each subgroup comparison

Using the PANTHER gene set analysis tool, we compared the DEG lists between risk subgroups in AML patients as identified by Davis et al, 2020. We did this for molecular functions (**Figure 3.7**), biological processes (**Figure 3.8**) and pathways (**Figure 3.9**). All groups are involved in very similar functions molecular (predominantly binding and catalytic activity), biological (predominantly cellular, metabolic and biological) and pathways (primarily Wnt signalling).

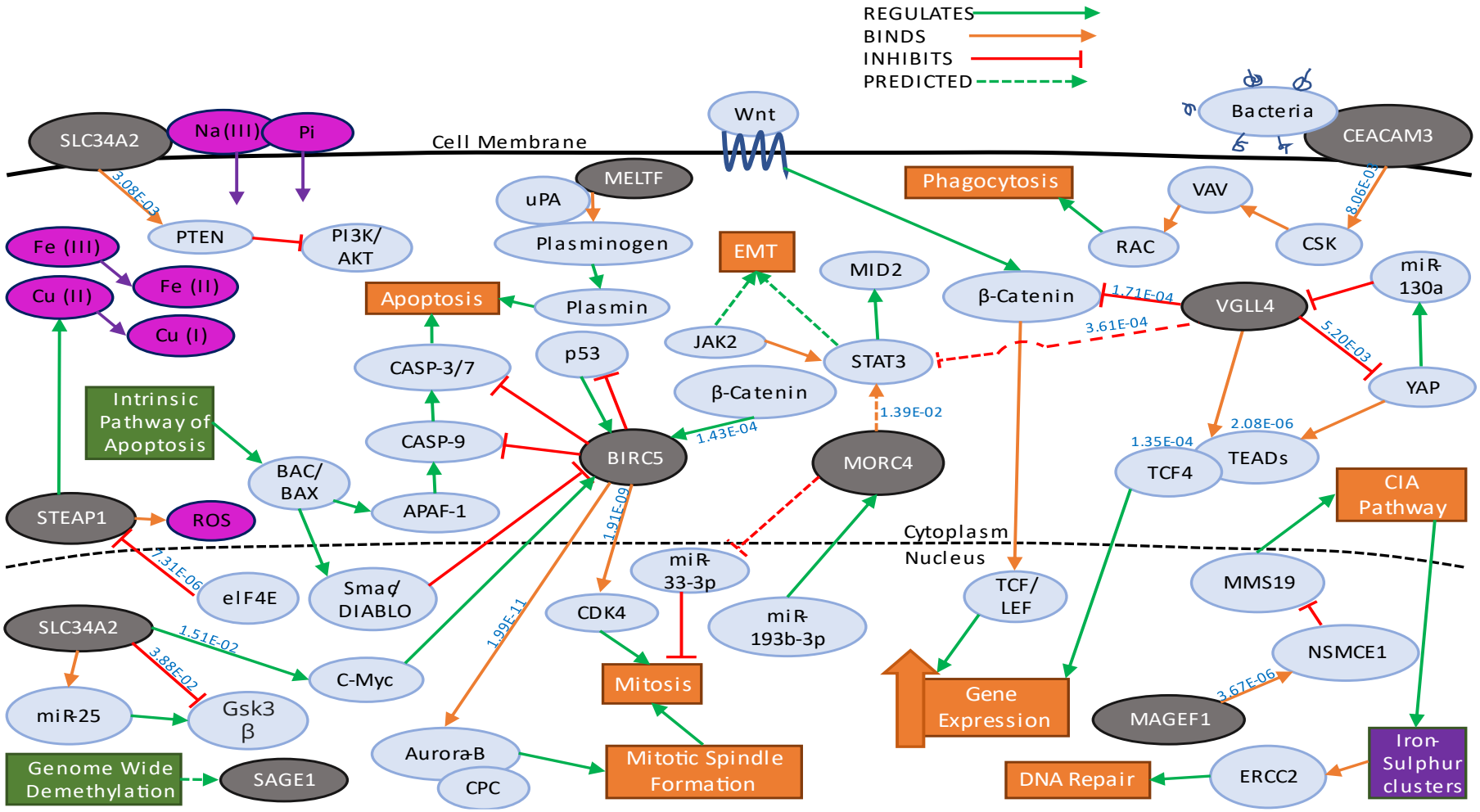


Figure 3.6 Predicted pathway map of nine LAAs. Generated with a literature search and the use of KEGG and STRING this highlights some of the pathways the GOI potentially act on. Correlation values generated by Professor Ken Mills, highlighting the relationship between the GOI and the downstream pathways are shown in pale blue font. Diagram authors own.

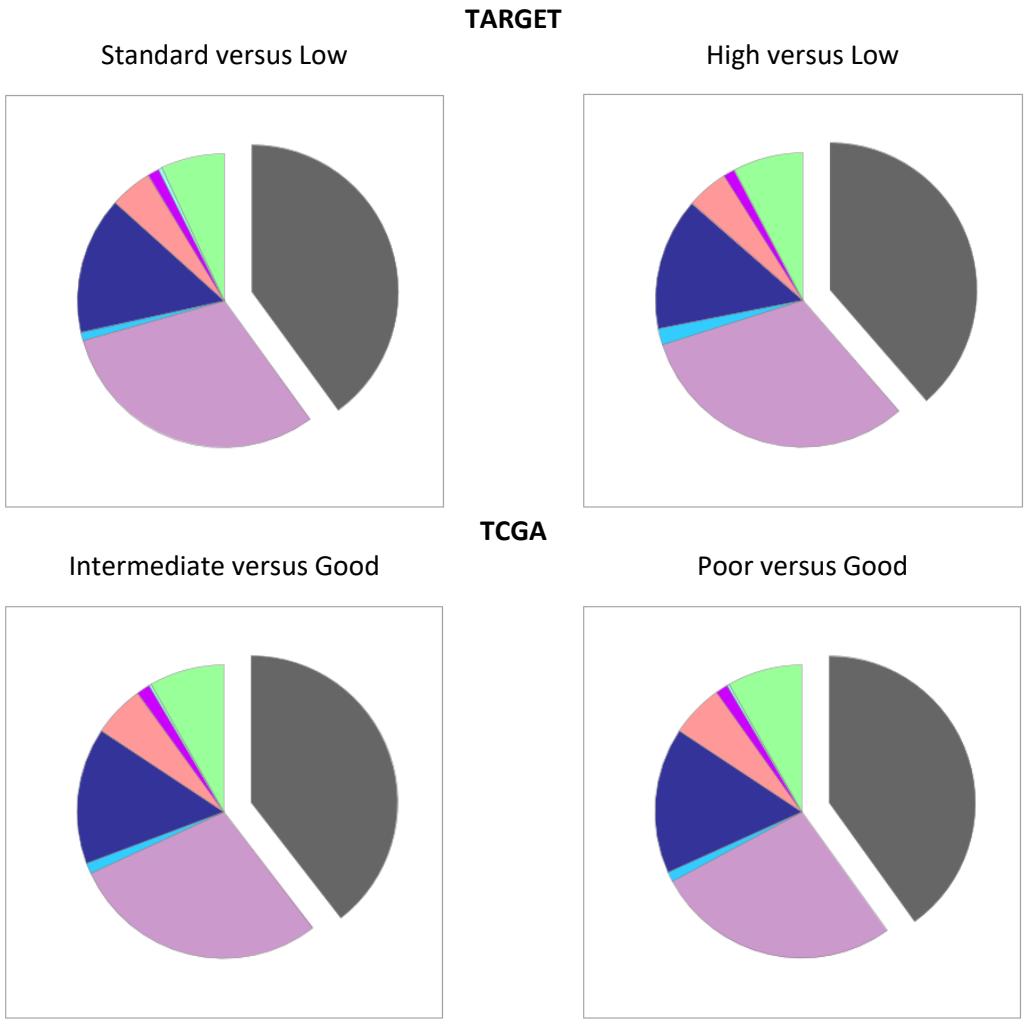


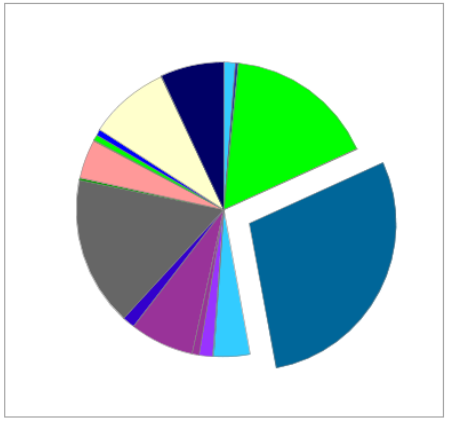
Figure 3.7 The molecular functions of the differentially expressed gene lists. Despite representing the molecular functions of the DEGs identified by Davis et al, whereby AML risk subgroup comparisons were made, similar molecular functions are utilised by DEGs in each. In each comparison the majority of genes were involved in binding.

Key:

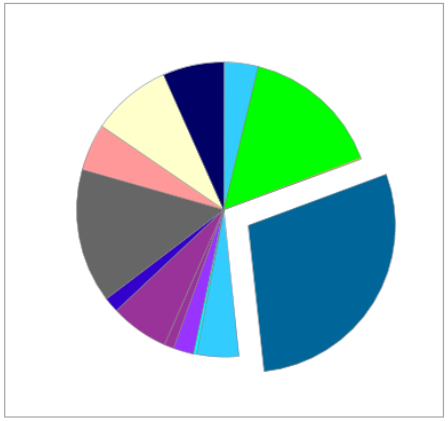
- Binding
- Catalytic activity
- Molecular function regulator
- Transporter activity
- Molecular transducer activity
- Structural molecule activity
- Molecular adaptor activity
- Translation regulator activity

TARGET

Standard versus Low

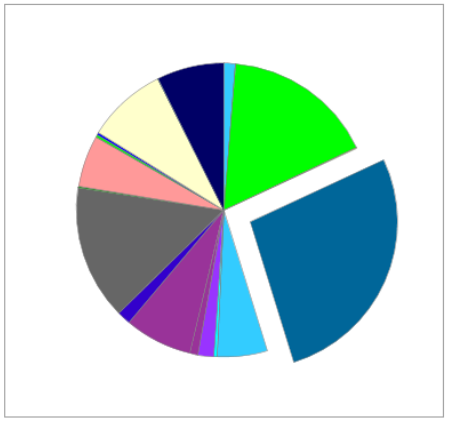


High versus Low

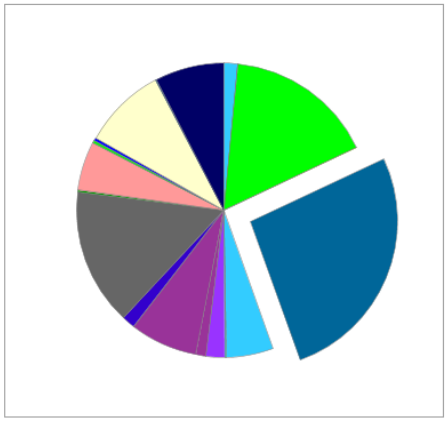


TCGA

Intermediate versus Good



Poor versus Good



TARGET

Figure 3.8 The biological processes of the differentially expressed gene lists. Evaluating the biological processes of the genes differentially between risk subgroups, the largest portion of genes are involved in cellular processes.

Key:

- Cellular process
- Metabolic process
- Biological regulation
- Response to stimuli
- Signalling
- Localisation
- Multicellular organismal process
- Biological adhesion
- Developmental process
- Immune system process
- Locomotion
- Interspecies interaction between organisms
- Growth

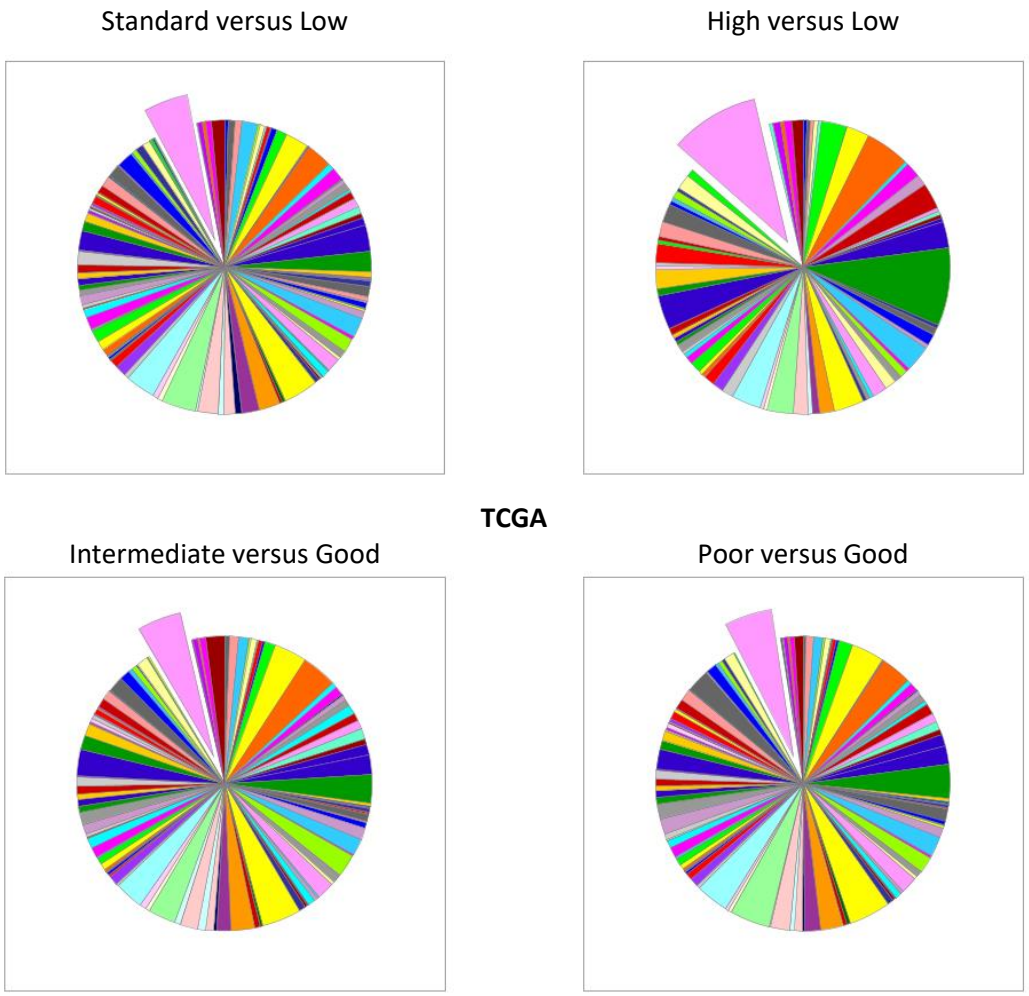


Figure 3.9 The pathways of the differentially expressed gene lists. In all subgroups the Wnt signalling pathway occupies the largest proportion of genes and a considerably larger portion in the high versus low paediatrics along with the Cadherin signalling pathway.

Key:

- Wnt signalling pathway
- Gonadotrophin-releasing hormone receptor pathway
- Inflammation mediated by chemokine and cytokine signalling pathway
- Integrin signalling pathway
- CCKR signalling map
- Angiogenesis
- Alzheimer disease-presenilin pathway
- Heterotrimeric G-protein signalling pathway-Gi alpha and Gs alpha mediated pathway
- Cadherin signalling pathway
- Huntingdon disease
- EGF receptor signalling pathway
- PGDF signalling pathway

Chapter 4. Investigation of the miRs utilised by AML cells to control gene expression

4.1 Introduction

We focussed on the miRs that were found to be differentially expressed when comparing risk subgroups between paediatric and adult AML patients (**Table 4.1**). These had all been identified by Davis et al, 2020 and the miRs with a >2-fold difference in expression were identified. Two of these miRs (miR486-1 and miR1915) were found to have a significant effect on survival depending on expression levels, whilst a third (miR-8086) was close to significant (**Figure 4.2**). miR-486-1 had been found in the original analysis in both the intermediate versus good, and poor versus good risk subgroups, miR-1915 was found in intermediate versus good risk subgroup comparisons as well as TARGET standard versus low (Davis et al., 2020). miR-8086 was differentially expressed in the intermediate versus good subgroup comparison. In separate analysis a fourth miR (miR-378G) showed considerable expression in the poor risk cytogenetic group and was differentially expressed in the poor versus good risk subgroup. These four miRs were chosen to be investigated further in models and qPCR.

Table 4.1 miRs differentially expressed in TARGET and TCGA groups.

TARGET Standard vs low risk				TCGA Intermediate vs Good risk		TCGA Poor vs good risk	
Symbol	p-value	Symbol	p-value	Symbol	p-value	Symbol	p-value
MIRLET7DHG	5.09E-33	**MIR624**	0.000305	**MIR3186**	8.14E-22	**MIR3186**	1.09E-10
MIR1282	5.63E-18	**MIR4308**	0.00032	MIRLET7BHG	5.67E-10	**MIR4500HG**	1.47E-07
MIR497HG	3.02E-17	**MIR133A1HG**	0.000326	**MIR4740**	1.21E-08	MIR4729	5.38E-06
MIR3661	1.88E-13	**MIR4512**	0.000388	**MIR210HG**	5.14E-08	MIR6731	7.38E-06
MIR4482	1.35E-09	**MIR326**	0.000776	**MIR1-1HG-AS1**	6.68E-08	**MIR4539**	1.43E-05
MIR3198-1	1.82E-09	MIR4748	0.000964	MIR1-1HG	2.42E-07	**MIR4537**	1.74E-05
MIR3918	4.54E-09	MIR3945	0.000969	**MIR8086**	8.70E-07	MIR4312	2.17E-05
MIR616	9.46E-07	**MIR604**	0.001142	MIR3619	2.70E-06	**MIR486-1**	2.25E-05
MIR17HG	8.88E-06	MIR103A2	0.001441	MIR5191	8.43E-06	**MIR378G**	2.99E-05
MIR1972-1	9.07E-06			MIR7152	2.03E-05	MIR7853	4.63E-05
MIR548C	9.88E-06			**MIR486-1**	4.79E-05	MIR210HG	4.65E-05
MIR1915	1.40E-05			**MIR1915**	5.38E-05	**MIR4740**	0.000104
MIR1296	3.77E-05	**TARGET High vs low risk**		MIR3681HG	7.16E-05	MIR3936	0.000105
MIR1254-1	4.39E-05	Symbol	p-value	MIR155HG	9.61E-05	MIR4538	0.000121
MIR374B	6.99E-05	**MIR3198-1**	2.65E-07	MIR3181	0.000173	MIR3936HG	0.00016
MIR4284	7.23E-05	**MIR644A**	0.00015	MIR4304	0.000183	**MIR6772**	0.00023
MIR3128	8.38E-05			MIR548I2	0.000216	MIR4304	0.000308
MIR4755	0.000112			MIR6731	0.000228	MIRLET7BHG	0.000498
MIR644A	0.000135			MIR4439	0.000298	MIR7152	0.000498
MIR221	0.000141			MIR4537	0.00039	MIR3681HG	0.000728
MIR421	0.000199			MIR4539	0.000595	MIR938	0.000804

MIRs indicated in bold text showed >2-fold difference in expression

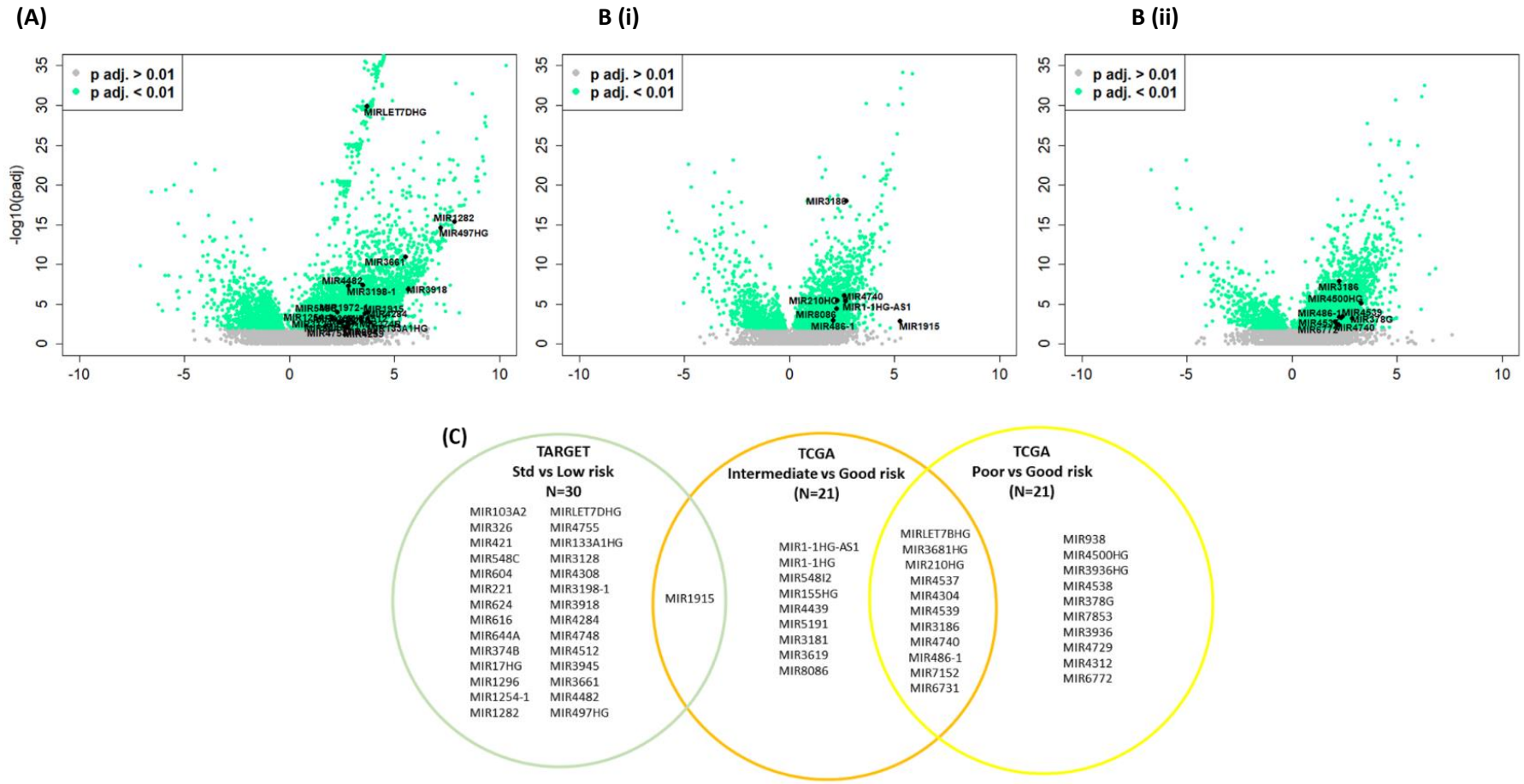


Figure 4.1 miRs identified as differentially expressed between risk subgroups in AML. Volcano plots show the up regulation of miR expression when comparing AML risk subgroups in datasets from **(A)** TARGET standard versus low and **(B)** TCGA (**i**) intermediate versus good (**ii**) poor versus good. In the standard versus low-risk volcano plot, the genes which have a positive log2FC value have increased expression in the standard risk group compared to the low risk. The good/low risk subgroups were used as references to which the intermediate/poor risk subgroups were compared for DGE analysis. **(C)** The degree of overlap between DEmiRs in intermediate versus good and poor versus good risk subgroup comparisons in the TCGA dataset. There was an absence of overlap between differentially expressed miRs identified in the TCGA (intermediate or poor versus good) and TARGET (standard versus low) risk subgroup comparisons.

Following analysis in UCSC Xena browser by Dr Pinar Uysal-Onganer two miRs (miR-486-1 and miR-1915) were found to affect OS depending on above and below median expression levels, whilst a third is nearly significant (miR-8086) (**Figure 4.2**).

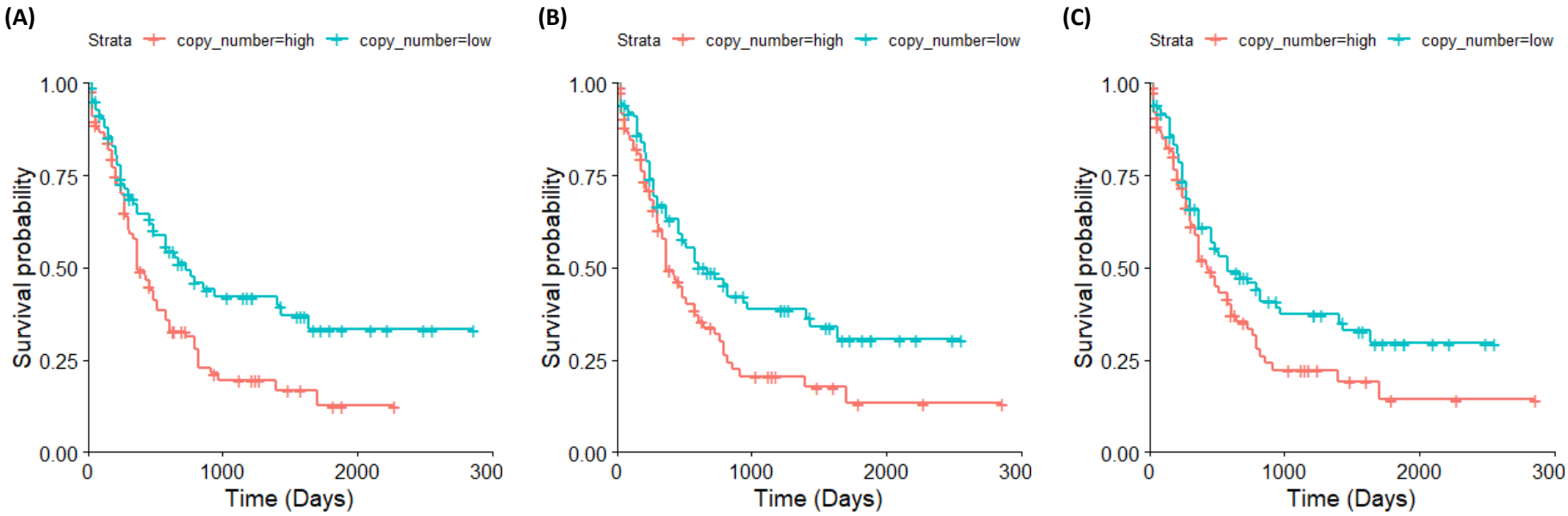


Figure 4.2 Overall Survival compared to miR copy number. Below averages cellular expression levels of (**A**) miR-486-1, (**B**) miR-1915 contributed to poorer OS (days survived after diagnosis) (p=0.006 and p=0.02 respectively), whilst (**C**) miR-8086 was close but not significant (p=0.06), code example created in R by Samantha Girvan with the 'dplyr', 'survminer' and 'survival' packages using copy numbers from the GDC TCGA-LAML dataset on UCSC Xena browser.

Having identified miRs that were differentially expressed between risk subgroups in adult AML and having identified three that were associated with OS, we decided to examine the cellular levels of the miRNAs in adult AML patient samples to match that of the mRNA-seq data. On finding that we had a limited number of cell samples we examine the larger cohort of serum samples available to us.

4.2 Aims

- To evaluate the capacity of miRs to identify AML patient risk subgroup at disease diagnosis based on their expression in cells and serum.

4.3 Significantly differentially expressed miRs

Using the Bloodspot analysis tool, we examined whether any of the DEmiRs with a >2 fold difference in expression were known to have elevated expression in association with any specific cytogenetic abnormalities in adults with AML (**Table 4.2**). Only three miRs we had identified (miR-Let-7DHG, miR-210HG and miR-4500HG) were associated with cytogenetic abnormalities and were represented by probe sets 559129_at, 230710_at and 232488_at, respectively. It was interesting to note that examining the TCGA dataset in bloodspot miR-Let-7DHG had the highest expression in del(7q)/7q- while in the MILE dataset its expression was highest in healthy BM.

Table 4.2 Association between risk subgroups and cytogenetic abnormalities. Association between specific cytogenetic abnormalities and miRs as indicated by BloodSpot data are shown in bold.

	Risk subgroup	Molecular and cytogenetic indicators	miR association
TCGA	Poor	Poor risk MLL translocation like t(6;11)(q27;q23), t(11;19)(q23;p13) or the presence of complex cytogenetics and poor risk cytogenetic abnormalities (Ley and Miller, 2013).	MIR378G, MIR4500HG, MIR 4537, MIR4539, MIR6772, MIR486-1, MIR3186, MIR4740
	Intermediate	Normal karyotype, the MLL translocation t(9;11)(q23;p13) or other intermediate cytogenetic abnormalities	MIR1915, MIR210HG, MIR8086, MIR486-1, MIR3186, MIR4740, MIR1-1HG-AS1
	Good	t(8;21)(q;22;q22) or inv(16)(p13q22)	
TARGET	Low	t(8;21)(q22;q22) or inv(16)(p13q22)	
	Standard	MLL gene rearrangements such as t(6;11)(q27;q23), t(10;11)(p12;q23), t(9;11)(q23;p13), t(11;19)(q23;p13) or a normal karyotype that didn't pose a change in risk.	MIRLET7DHG, MIR133A1HG, MIR326, MIR374B, MIR421, MIR497HG, MIR548C, MIR604, MIR624, MIR1254-1, MIR1282, MIR1972-1, MIR3128, MIR3661, MIR3918, MIR4284, MIR4308, MIR4482, MIR4512, MIR3198-1, MIR1915
	High	FLT3 ITD	

To further visualise the different expression levels of the DEmiRs in each risk subgroup in which they were differentially expressed, the raw count data from TCGA was shown (**Figure 4.3**). Further analyses using the UCSC Xenabrowser highlighted the copy number of these four miRs were almost identical in terms of copy numbers in favourable and intermediate AML cytogenetic risk subgroups but elevated in poor risk subgroups.

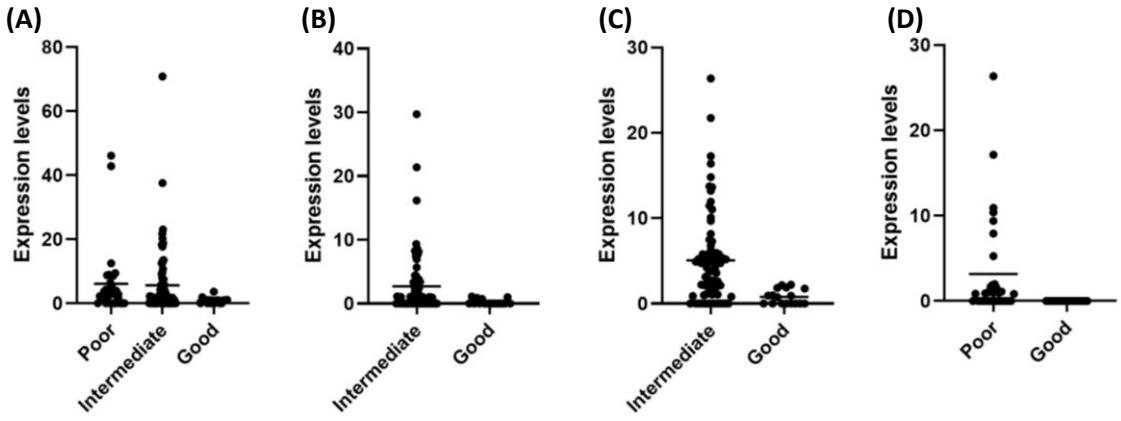


Figure 4.3 Expression levels of miRs in TCGA risk groups. (**A**) miR-486-1 in poor (n=32) versus intermediate (n=75) versus good (n=17), (**B**) miR-1915 in intermediate (n=75) versus good (n=17), (**C**) miR-8086 in intermediate (n=75) versus good (n=17), (**D**) miR-378G in poor (n=32) versus good (n=17). Significantly raised raw count levels were found in the worse risk groups.

Using DAVID to analyse the pathways that the DEmiRs were involved in, only one of the identified DEmiRs was found to map to the KEGG pathway for MicroRNAs in cancer [05206: Accessed 15/9/21]. miR-326, working with miR-34a, acts to inhibit Notch activity in glioma cells as part of the tumourigenesis process (Kefas et al., 2009).

4.4 Literature review of miRNA gene regulation in health and disease

To determine what was already known about the genes that each miR regulates, the literature was comprehensively searched. PubMed and Scopus were mainly used with no limits (in years, disease or patient age) and the search term was solely that of the miR.

4.4.1 miR-1915

In HCC cell lines under oxidative stress, miR-1915-3p levels were drastically reduced following the knockdown of TP53, leading to the assumption that p53 regulates miR-1915-3p in some manner (Wan et al., 2017). BCL-2 is a proto-oncogene that typically binds to pro-apoptotic proteins (PUMA, BAK, and BAX), inhibiting their activity and preserving the integrity of the mitochondria and through this enhancing cell survival (Kontos et al., 2014). Following DNA damage, p53 was found to upregulate miR-1915 expression, however, the levels of pri-miR-1915 did not change unlike the levels of both pre and mature miR-1915, this suggests p53 may be inhibiting the processing of primary to precursor miRNA (Nakazawa et al., 2014). The upregulation of miR-1915 may then inhibit BCL-2 and prevent apoptosis. Within colorectal cancer, elevated levels of miR-1915 caused by mimics reduced the BCL-2 level and sensitized the

cells to anticancer drugs (Xu et al., 2013), providing more evidence for the effect of miR-1915 on BCL-2.

MiR-1915-3p has been demonstrated to inhibit lung cancer cell apoptosis by downregulating developmentally regulated GTP binding protein 2 and pre B cell leukaemia homeobox 2 (PBX2), whereas overexpression prevented etoposide-induced apoptosis (Xu et al., 2016).

Overexpression of miR-1915-3p was found in the serum of breast cancer patients and its overexpression repressed dual specificity phosphatase 3 (DUSP3) which activated extracellular signal-regulated protein kinase 1/2 (Guo et al., 2018).

In adult renal progenitor cells, miR-1915 was found to regulate the expression of markers of renal progenitors such as CD133 and Paired box 2 (PAX2). For the expression of these stem cell markers lows levels of miR-1915 were needed, suggesting that it may inhibit their expression (Sallustio et al., 2013).

Following the transfection of miR-1915 mimics and inhibitors into GC cells, miR-1915 had tumour suppressive effects on the cells, reducing the levels of proliferation and migration of HP-infected GC cells. This may have been caused by the targeting of receptor for advanced glycation end product (RAGE) (Xu et al., 2019).

4.4.2 miR-486

In AML, miR-486 was shown to regulate JAK-STAT signalling by inhibiting suppressor of cytokine signalling 2 (SOCS2), via the silencing of SOCS2 (a negative regulator of the JAK/STAT pathway) it enhanced cell proliferation (Sha et al., 2021). The knockdown of miR-486 was shown to increase JAK-STAT3 activity and increase proliferation.

In muscle cells, miR-486 was shown to regulate PI3K/AKT signalling by targeting PTEN and Forkhead Box O1 (FOXO1), the levels of PTEN and FOXO1 were reduced by overexpression of miR-486 (Small et al., 2010; Cristofoletti et al., 2013; Shen et al., 2019).

miR-486-5p has been reported to negatively impact tumour suppressor pathways in prostate cancer (Yang et al., 2017). It also functions as a tumour suppressor in oesophageal cancer by targeting CDK4/ breast carcinoma-amplified sequence 2 (BCAS2) (Lang & Zhao, 2018). miR-486-5p was also found to target the transforming growth factor beta (TGF-β)/ SMAD family member 2 (SMAD2) signalling pathway, which controls transcription of downstream targets (Yang et al., 2017). miR-486-3p was found to regulate γ-globin expression via modulating BAF chromatin remodeling complex subunit in erythroid cells (Lulli et al., 2013).

miR-486-5p was discovered to be downregulated in HCC tissues and cell lines and the overexpression of miR-486-5p reduced the proliferation and migration of the cells. In cell lines with overexpression of miR-486-5p the levels of E3 ubiquitin ligase casitas B-lineage lymphoma (CBL) were reduced with similar results observed in tissue samples, an increased expression of CBL reversed the effects of the miR on the rate of proliferation and so miR-486-5p could be a tumour suppressor acting via the inhibition of CBL (He et al., 2019).

Polymorphonuclear myeloid-derived suppressor cells (PMN-MDSCs) have been shown to decrease the levels of miR-486-3p and accordingly activate nuclear factor kappa B2 (NF-κB2) in thyroid cancer, resulting in greater invasion rates (Chen et al., 2020). By recovering the levels of the miR this group were able to reduce these effects suggesting it may be a tumour suppressor. A further study in AML found that most patients categorised with normal cytogenetics displayed increased miR-486, specifically in the M2 subtype (Seyyedi et al., 2016a).

With the TF-1 erythroleukaemia cell line, an increase in miR-486 expression was shown to be caused by hypoxic conditions. Sirt1 was also identified as a target of miR-486 which can modulate differentiation of erythroid TF-1 cells, inhibition of the miR suppressed growth and differentiation (Shi et al., 2017).

4.4.3 miR-378G

miR378G has been shown to be involved in the suppression of metastasis in colon cancer cells via the inhibition of SDA1 Domain Containing (SDAD1), overexpression of miR-378G was found to inhibit β-catenin expression (Zeng et al., 2017a). It is also found to act in a similar manner, interacting with BRAF one to achieve suppression of proliferation in colorectal cancer (Wang et al., 2015c). The expression of VEGF in GC cell lines was inhibited by miR-378 (Deng et al., 2013), VEGF expression was also found to be controlled by miR-378G in laryngeal cancers (Hua et al., 2006).

miR-378 has been found to inhibit the expression of suppressor of fused (SUFU), a negative regulator of the hedgehog pathway that is involved in signalling systems specifying cell growth and differentiation during vertebrate development (Cheng & Yue, 2008). This might suggest a role in tumour cell survival along with the inhibition of Fus-1, which could increase the survival of the cell (Lee et al., 2007). This group also found transfection of miR-378 decreased caspase 3 expression and would therefore reduce apoptosis and further improve cell survivability.

In oral squamous cell carcinoma it has been suggested that LncRNA HOXC13 antisense RNA functioned as a ceRNA to make the cells more malignant and this could have been done by the inhibition of miR-378G, demonstrated by the expression of the miR being negatively correlated with that of LncRNA HOXC13 antisense RNA (Li et al., 2020).

The overexpression of miR-378 in cardiomyocytes might inhibit their proliferation via the suppression of the Ras GTPase-MAPK pathway by targeting three components of this pathway: MAPK1, insulin-like growth factor receptor 1 and kinase suppressor of RAS 1 (KSR1) (Ganesan et al., 2013).

Following analysis of the potential roles of these miRs, a pathway map was created highlighting potentially important processes and interactions (**Figure 4.4**). At the time of this study little to no information is available about the role of miR-8086 so it was not covered in the previous section. All the miRs evaluated have links to control of the apoptosis and other vital signalling pathways.

4.5 miR expression and survival

We found that none of the miRs identified as promising (bolded in **Table 4.1**) were indicative of survival (>6 years) when expressed at above or below median levels in the MILE database (miR-LET7DHG p=0.531; miR-210HG p=0.23; miR-4500HG p=0.711) (n=172); however when we analysed 175 adults with AML in GDC AML TCGA dataset available on UCSC Xena browser, above median expression of miR-486-1 and miR-1915 that had been identified as DEmiRs in the intermediate versus good subgroup comparison in the TCGA dataset showed a significant association with reduced survival suggesting again that upregulation of these two miRs were associated with the poorer risk subgroups in adult AML (**Figure 4.2**). miR-8086 whilst not significant showed a trend towards increased expression affecting the OS of patients.

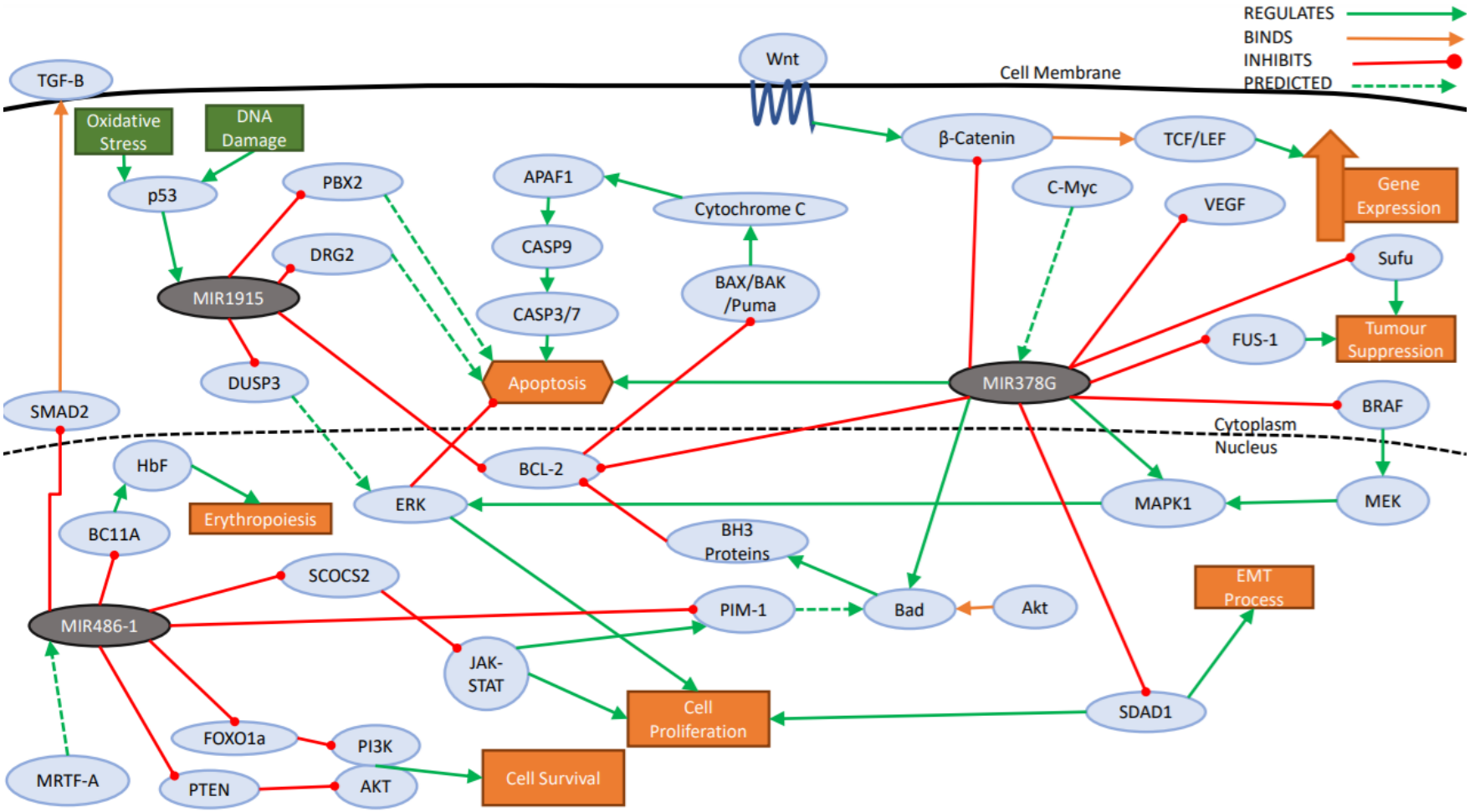


Figure 4.4 Predicted pathway map of DEmiRs of interest. Generated via a literature search this pathway map focuses on some miRs showing diagnostic promise. The pathways are not necessarily in AML and could have been investigated in a variety of diseases and conditions. miR-8086 had no previous research on any pathways it could potentially be involved in.

4.6 ROC Curves, AUC and models

The miRs of interest were further investigated for their ability to distinguish between subgroups using the mRNA sequencing data available from the TCGA. When evaluating the intermediate versus good risk subgroups there is n=17 in the good group and n=75 in the intermediate group. Following ROC analysis of the intermediate versus good groups, curves were generated highlighting the specificity and sensitivity of combinations of miRs (**Figures 4.5, 4.6 and 4.7**). An AUC score above 0.7 is acceptable, whilst a score above 0.8 is excellent and above 0.9 is even better (Mandrekar, 2010).

When analysing the mRNA sequencing data using ROC analysis age and gender were added in a stepwise manor and were found to have no significant effect on any of the models, proving that they did not affect the miR expression in that cohort.

4.6.1 Single miRs in the intermediate versus good subgroup

Individual miRs could still make potential classifiers to distinguish between risk groups (**Figure 4.5**).

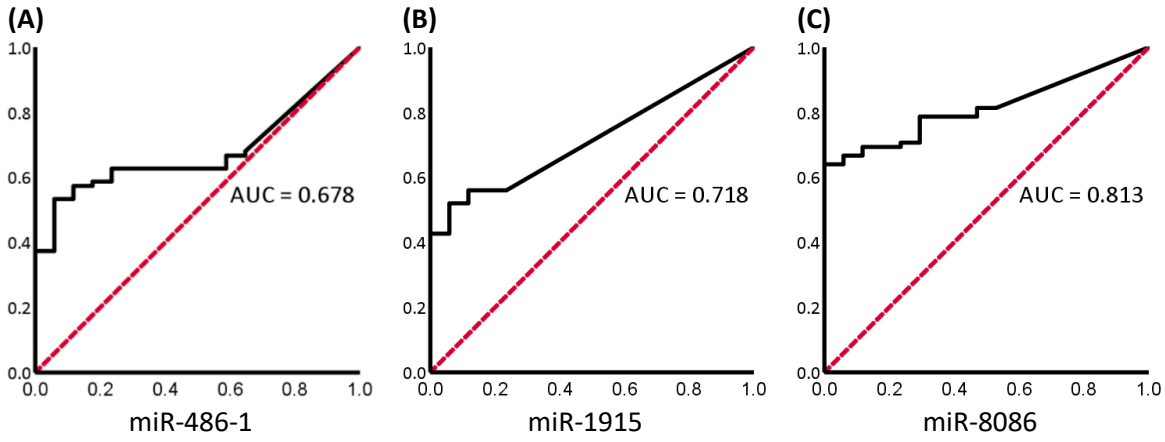


Figure 4.5 Intermediate versus good risk groups for single miRs. (A) miR-486-1 (AUC Value = 0.678, Standard error = 0.0561, 95% Confidence interval = 0.5684 to 0.7884, p-value = 0.0221), **(B)** miR-1915 (AUC Value = 0.718, Standard error = 0.0550, 95% Confidence interval = 0.6102 to 0.8259, p-value = 0.0052) and **(C)** miR-8086 (AUC Value = 0.8133, Standard error = 0.0437, 95% Confidence interval = 0.7277 to 0.8990, p-value = <0.0001).

Even on its own, miR-8086 is potentially a powerful classifier for subgroups. When analysing the raw count data for miR-8086 patients often had either no expression or very high expression, this could contribute towards the high AUC score.

4.6.2 Combination of two miRs

Using any combination of two miRs yielded higher AUC scores, allowing for greater predictive ability of a model to classify patients based on mRNA-sequencing expression (**Figure 4.6**).

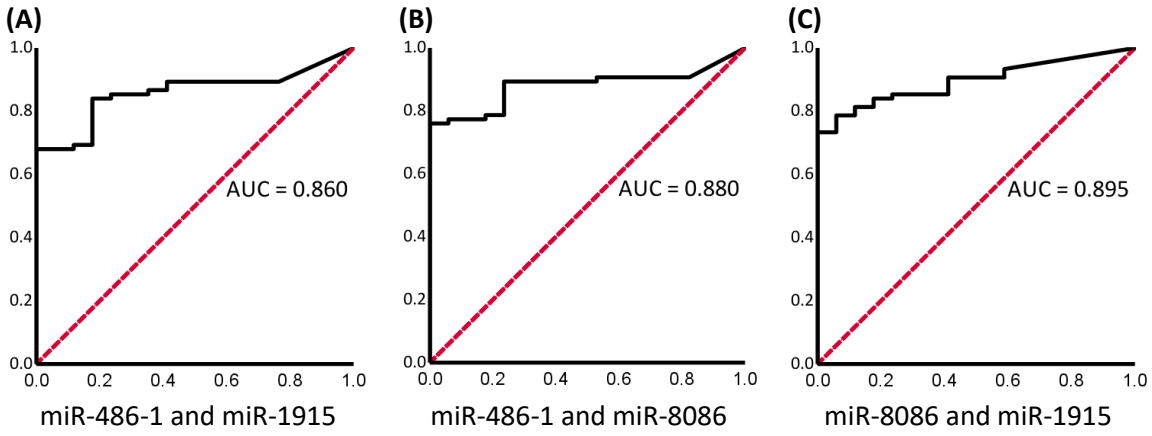


Figure 4.6 Intermediate versus good risk groups for combinations of miRs. (A) miRs-486-1 and 1915 (AUC Value = 0.860, Standard error = 0.039, 95% Confidence interval = 0.784 to 0.935, p-value = <0.0001), **(B)** miRs-486-1 and 8086 (AUC Value = 0.880, Standard error = 0.035, 95% Confidence interval = 0.811 to 0.949, p-value = <0.0001) and **(C)** miRs-8086 and 1915 (AUC Value = 0.895, Standard error = 0.032, 95% Confidence interval = 0.832 to 0.959, p-value = <0.0001).

4.6.3 All miRs combined

With three miRs involved in the ROC analysis the AUC score increases further, showing a more specific and sensitive model (**Figure 4.7**).

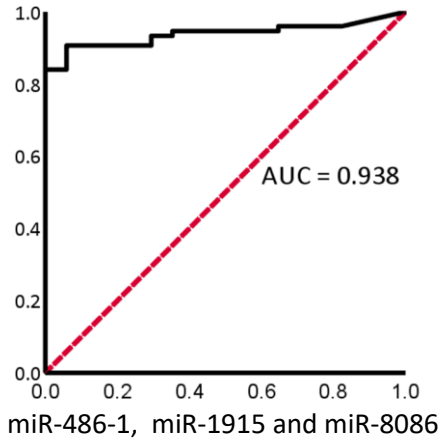


Figure 4.7 Intermediate versus good risk groups for all miRs combined. miRs-486-1, 1915 and 8086 (AUC Value = 0.938, Standard error = 0.025, 95% Confidence interval = 0.890 to 0.987, p-value = <0.0001).

When analysing the power of miRs to distinguish between intermediate versus good risk group a combined model (**Figure 4.7**) had an AUC of 0.938 compared to the best individual miR with an AUC of 0.813.

4.6.4 miRs in poor versus good subgroup TCGA

When evaluating the poor versus good risk subgroups there were n=32 in the good group and n=32 in the poor group. The potentially more important subgroups to distinguish between is poor versus good, due to the greater differences in OS. A model able to predict the poor risk patients would allow greater attention to be given to them or more severe treatment course applied. After ROC analysis of the poor versus good subgroups, curves were created showing the models (**Figures 4.8 and 4.9**).

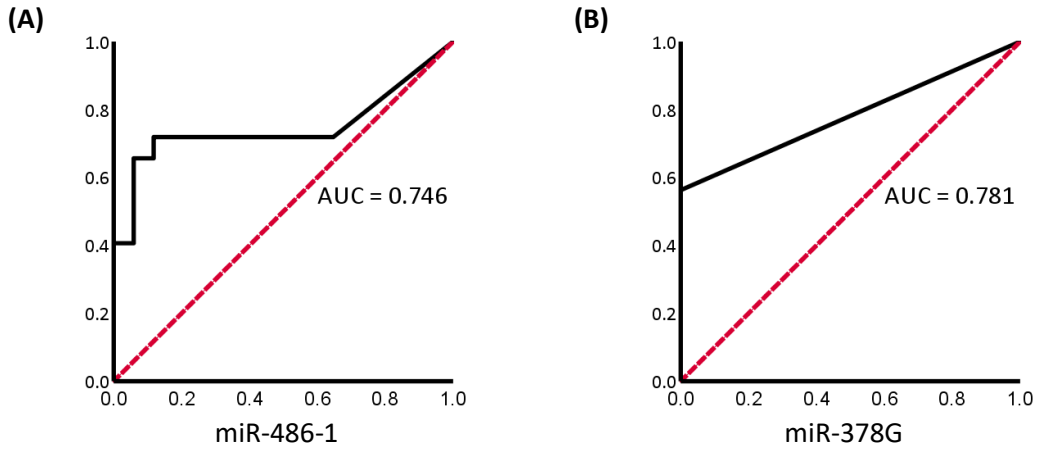


Figure 4.8 Individual miRs in the poor versus good risk subgroups. **(A)** miR-486-1 (AUC Value = 0.746, Standard error = 0.039, 95% Confidence interval = 0.784 to 0.935, p-value = <0.0001) and **(B)** miR378G (AUC Value = 0.781, Standard error = 0.035, 95% Confidence interval = 0.811 to 0.949, p-value = <0.0001).

With two miRs included in the model it becomes a better predictive marker, however not quite the level of classification the model for intermediate versus good achieved (**Figure 4.9**).

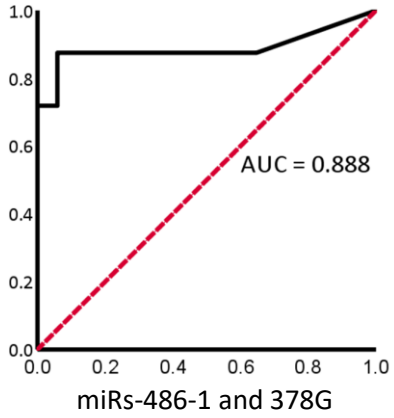


Figure 4.9 Poor versus good risk groups miRs-486-1 and 378G. miRs-486-1 and 378G (AUC Value = 0.888, Standard error = 0.050, 95% Confidence interval = 0.789 to 0.987, p-value = <0.0001).

miR-486-1 and miR-378G are good predictive biomarkers on their own however when combined are a more powerful tool for predicting poor prognosis, compared to the combination of three miRs in the intermediate versus good group the AUC score is lower (0.888 to 0.938), by adding a third a miR into this model the power could be improved further.

4.7 qPCR Results for positive controls and normalisers

To test the consistency of the RNA extraction kit, RT process and to test the miR assays, cell lines with predicted expression of the DEmiRs were analysed (**Table 2.4**). Cell sample normalisers (U6 snRNA and 5s rRNA) did not show any expression in the positive cell lines used. Positive controls used were K562, HL-60 and MDA-MB-231. None of the miRs investigated showed any significant expression, with different Ct values and too much variance in the results. This could have been due to inexperience in the methodology or the miRs and normalisers were not present in the cell lines.

4.8 qPCR Results for PBLs

With limited information on the performance of the assays and the normalisers due to the insufficient data gathered from the cell lines, some patient PBL samples were used to test efficiency. With only three PBL samples (two from a single patient at different time points and one healthy volunteer). The results are limited when analysing qPCR, however it validated the assays for our use and gave more experience in the processes required (**Figure 4.10**).

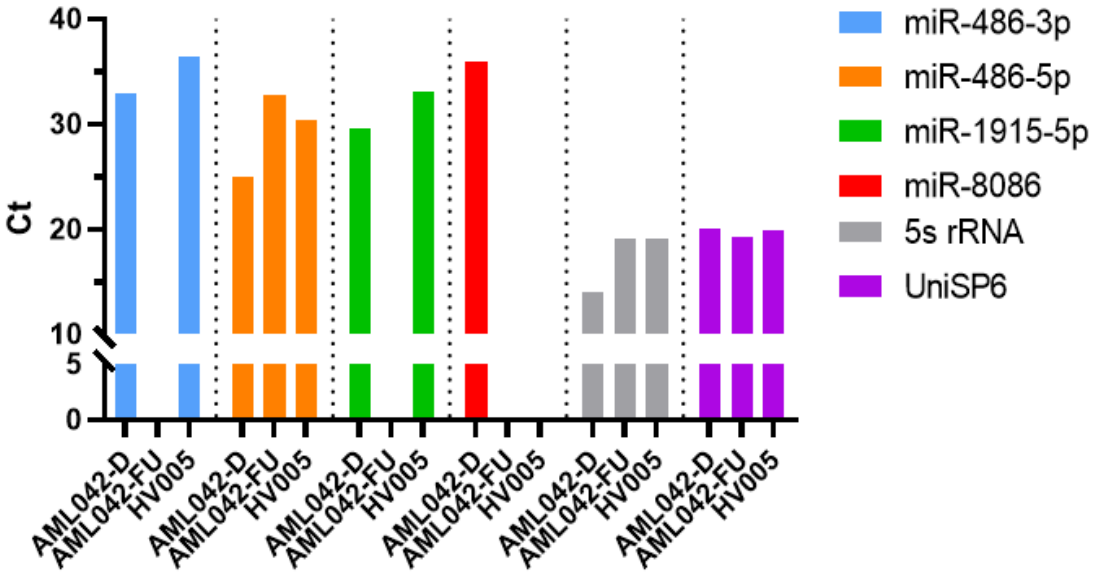


Figure 4.10 Raw Ct values for three PBL samples. qPCR analysis of AML042 at diagnosis and follow up (FU) compared with a HV, the normalisers are stable between samples but expression of some miRs is limited.

Creating a NV from the UniSP6 and 5s rRNA allowed for better comparisons of expression between samples by allowing the calculation of ΔCt (**Figure 4.11**). There were no discernible differences between any of the samples other than the complete lack of expression of miR-8086 in the healthy volunteer sample and the follow-up sample.

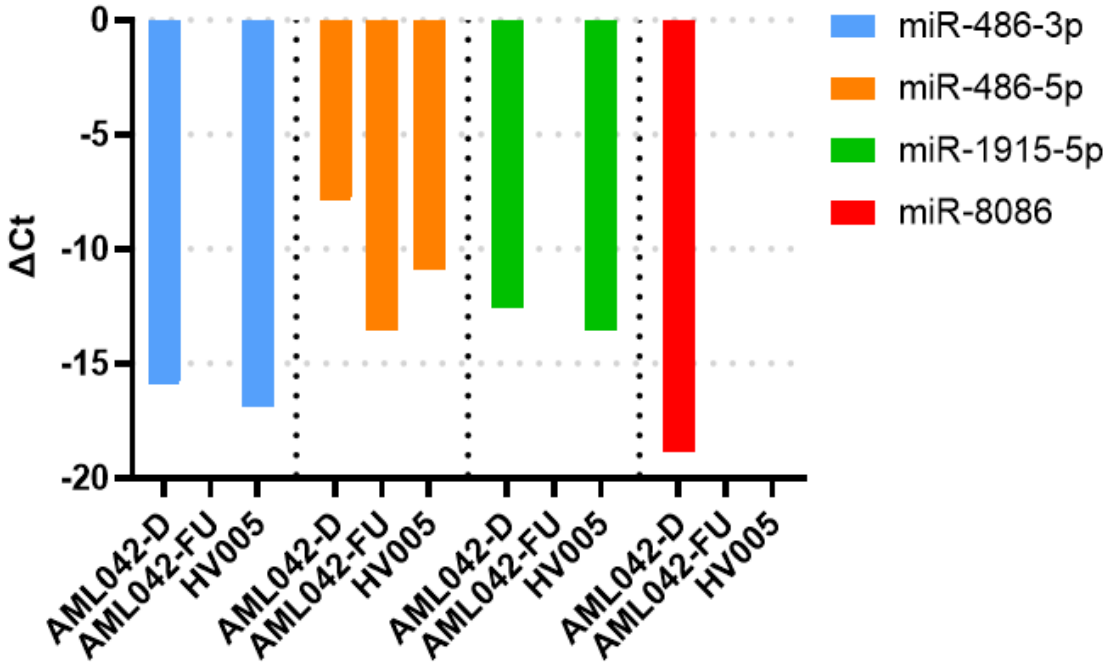


Figure 4.11 ΔCt values for three PBL samples. Samples were adjusted using the NV created with UniSP6 and 5s rRNA, miR-8086 showed no expression in the FU or HV.

With limited data to analyse, a point of interest is the lack of expression of miR-8086 in both the FU and HV sample. A patient in FU would hopefully have characteristics more akin to a healthy individual then one at the point of diagnosis.

4.9 qPCR serum

The availability of a greater number of serum samples allowed for more thorough analysis of miRs based on clinical features. The normalisers used to analyse miR expression in sera are different from those used to examine cellular presence of miRs in cell lines and patient cell samples. The combined use of two endogenous controls and a spike-in proved reliable for this small study of sera samples (**Figure 4.12**).

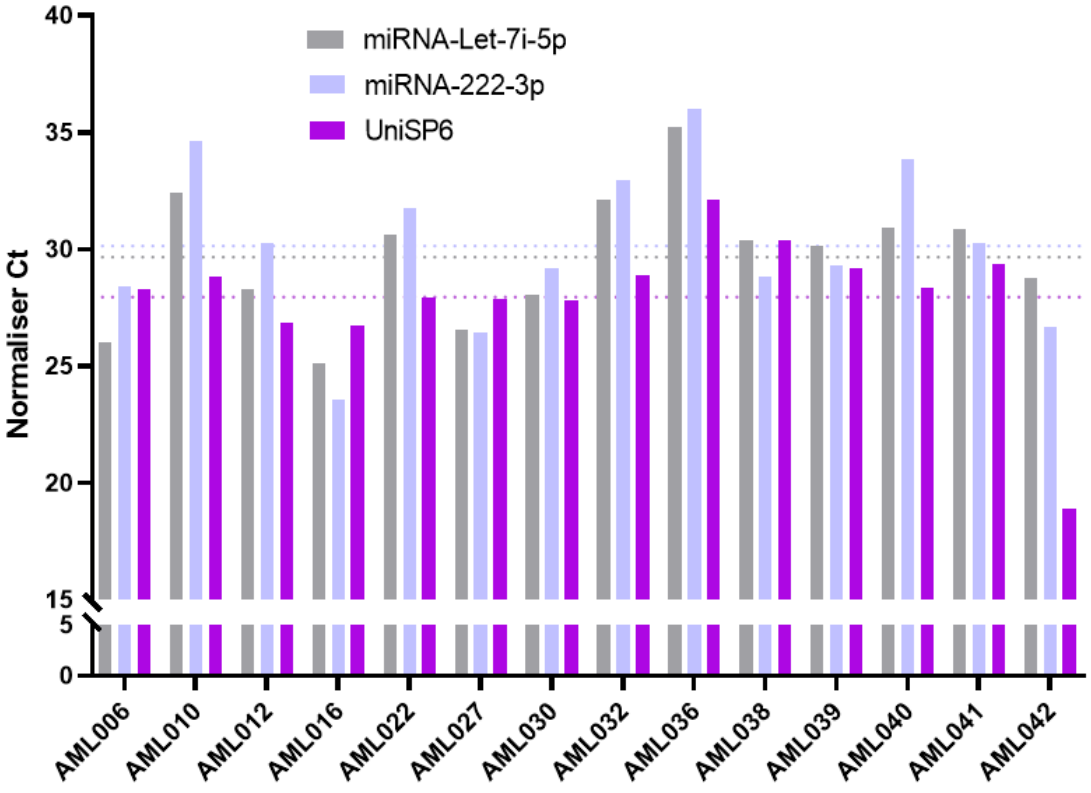


Figure 4.12 Raw Ct of normalisers. Endogenous controls and UniSP6 raw Ct values in each sera patient sample analysed. Without adjustment both miR-222-3p and miR-Let-7i-5p are both within a close range in a single patient however there is a large amount of variance between patients. UniSP6 is more consistent throughout the whole cohort of patients and should be used to help form the NV. The dotted line signifies the mean of each normaliser for the cohort.

There was still some variance across the sample range, however the use of a NV reduces this significantly (**Figure 4.13**).

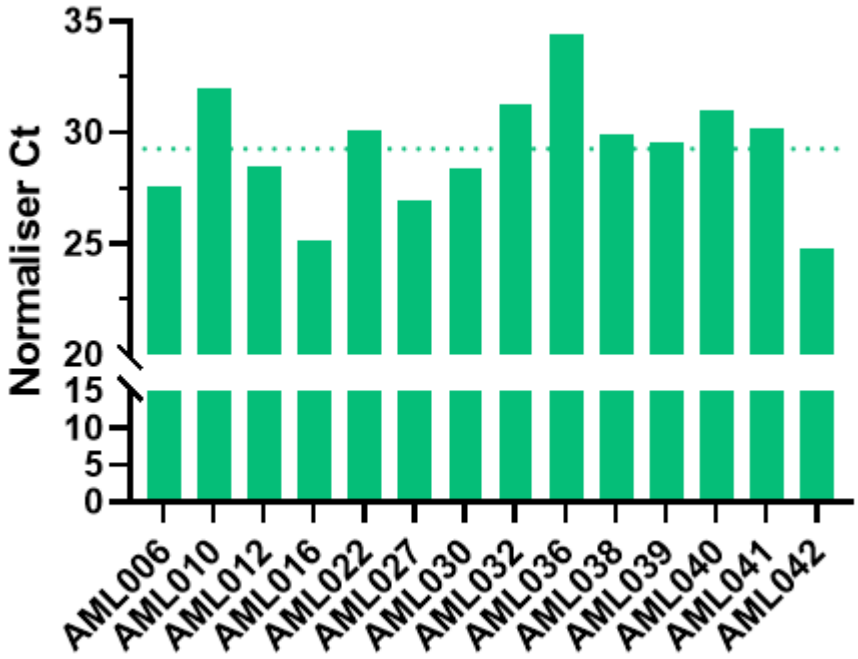


Figure 4.13 Ct of endogenous controls combined with UniSP6 to form a NV. Endogenous controls miR-Let-7i-5p and miR-222-3p were combined with UniSP6 spike-in to form a NV that was used to calculate ΔCt for each miR investigated. There is still a high degree of variance between all the samples. Dotted line signifies the mean NV for the cohort.

Across the cohort miR-486-5p had the most consistent expression in all samples (**Figure 4.14**). The expression of miR-8086 and miR-1915-3p did not show sufficient amplification for further data analysis.

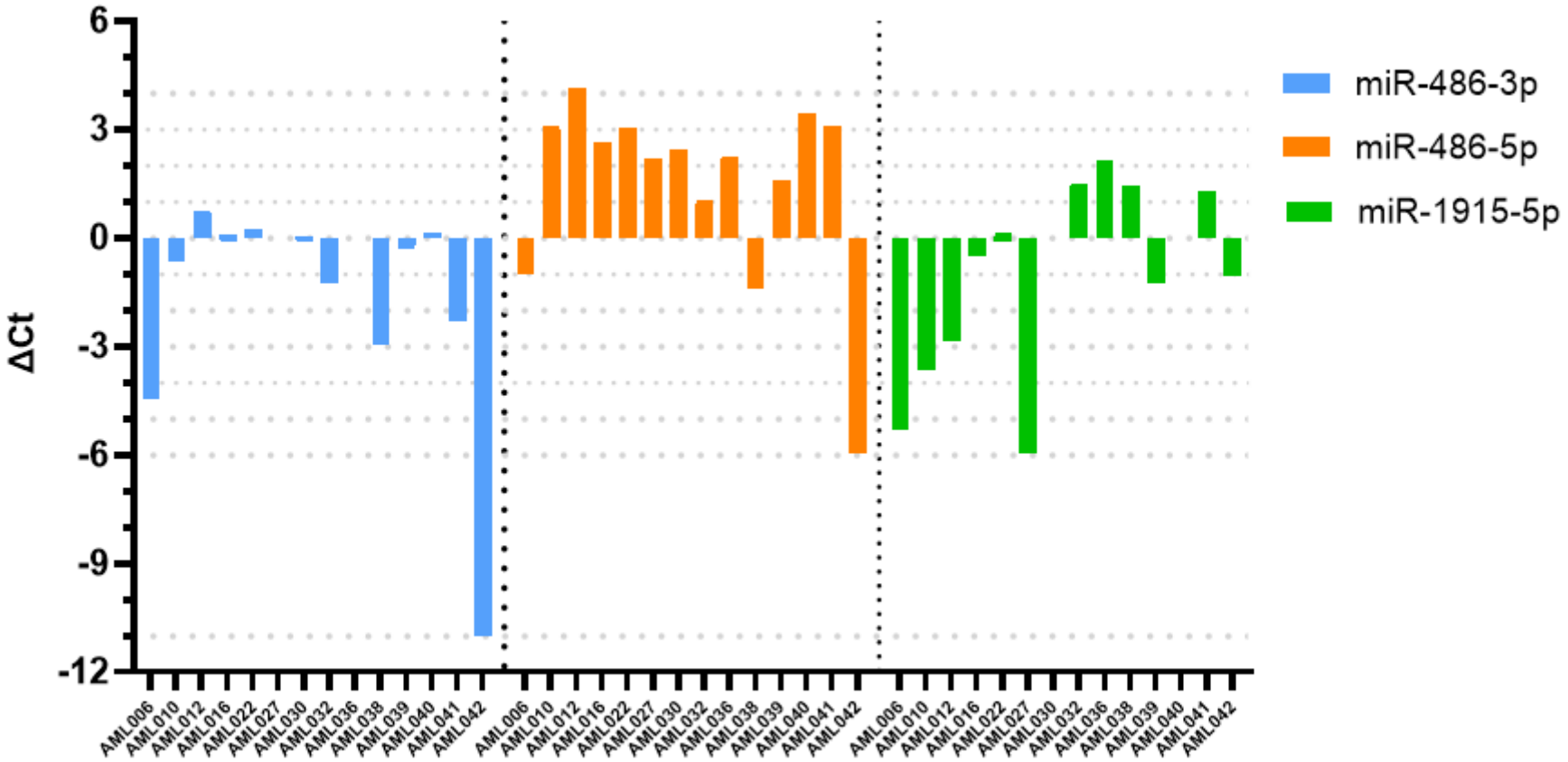


Figure 4.14 ΔCt of miRs of interest in all patient samples analysed. The Ct values have been normalised using the NV to bring them into a closer range of each other. Some patients showed no identifiable expression of one miR. In miR-8086 and miR-1915-3p no expression was shown was recorded during the qPCR; this aligns with the positive controls showing no expression of miR-8086.

Grouping the patients using their FAB subtype (**Table 2.5**), allowed the observation that there was no significant difference in expression of miRs between the risk subgroups (**Figure 4.15**). However, there was a trend towards a lower level of miR-1915-5p in the two favourable risk sera samples.

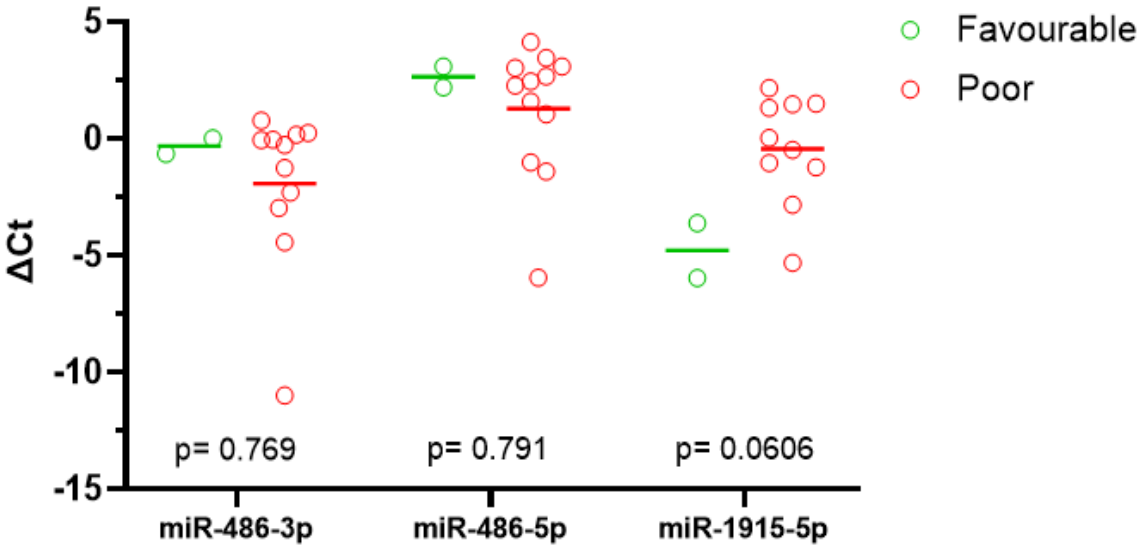


Figure 4.15 ΔCt values of all miRs tested via qPCR in sera. miR-486-3p and miR-486-5p showed no significant differences between the subgroups with p values of 0.77 and 0.79 respectively. ΔCt values display a trend (p=0.061) towards increased expression of miR-1915-5p in the poor risk subgroups, but we cannot call these trends significant due to the small sample size (n=2) in the good risk group. Mann-Whitney test used to generate p-values.

Using non-parametric tests due to the small group sizes there are no significant results, however miR-1915-5p does show a trend towards increased expression in the poor risk subgroup (**Figure 4.15**).

4.10 Significance of age, biological sex, and date of sample on qPCR

To evaluate the relationship age and gender may have on serum miR expression linear regression, correlation analysis and Mann-Whitney tests was performed where suitable. The first analysis observes the effect of age on miR expression (**Figure 4.16**).

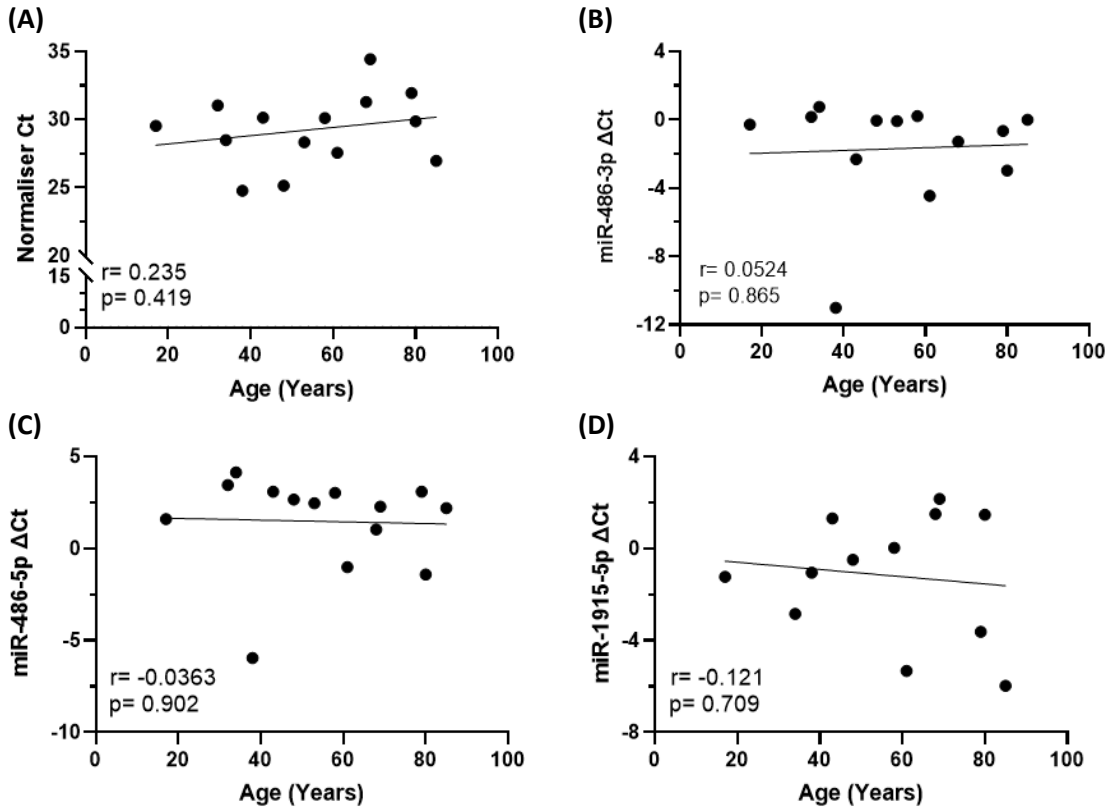


Figure 4.16 Linear regression of age on miR expression. **(A)** Normaliser Ct value relationship with age (n=14). **(B)** miR-483-3p ΔCt (n=14). **(C)** miR-483-5p ΔCt (n=13). **(D)** miR-1915-5p ΔCt (n=12). Correlation and linear regression analysis, values reported are correlation.

The age of the patient at diagnosis did not influence DEmiR expression nor change the NV in our patient cohort, as evidenced by the fact that none of the p values were significant ($p<0.05$, correlation analysis).

Analysis of the impact of patient biological sex on miR expression was examined (**Figure 4.17**). There was a small trend in the NV between males and females, with females showing a slightly higher expression of the NV (**Figure 4.17A**). The trend of higher expression of miRs also continues in miR-486-3p (**Figure 4.17B**) and miR-486-5p (**Figure 4.17C**). This trend was reversed in miR-1915-5p, with males showing slightly higher expression (**Figure 4.17D**).

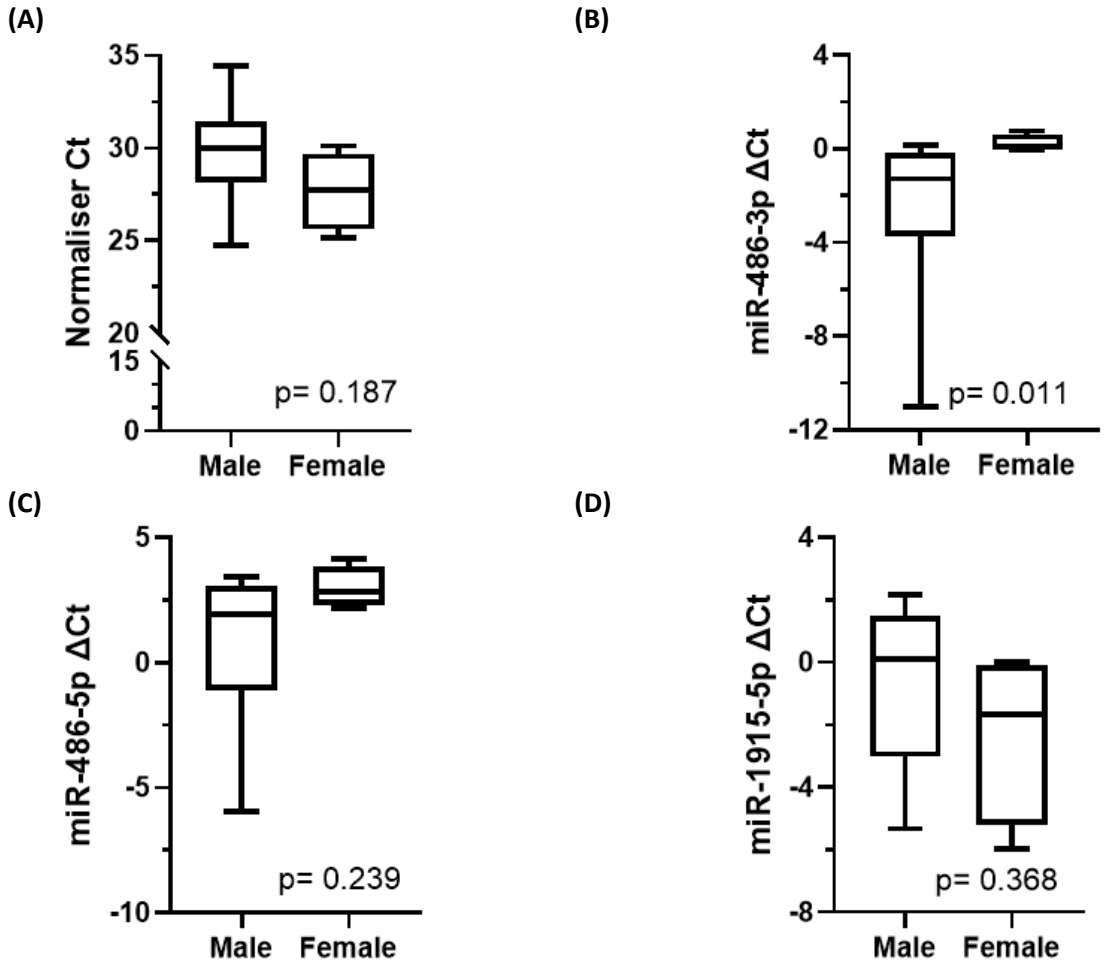


Figure 4.17 Analysis of gender on miR expression. **(A)** Normaliser Ct value relationship with age (10 men and women). **(B)** miR-486-3p ΔCt (9 men and 4 women). **(C)** miR-483-5p ΔCt (10 men and 4 women). **(D)** miR-1915-5p ΔCt (8 men and 4 women). Mann-Whitney tests used due to small sample size in some groups not allowing parametric tests.

Examination of the impact of time since the sample was taken on miR levels were performed. Most samples were taken around 20 years ago and may have been subject to considerable degradation as demonstrated by previous studies (Ibberson et al., 2009). A single sample, AML042 was received much more recently and had significantly different expression that may have caused the elevated levels of UniSP6 detected in the sample (**Figure 4.18**).

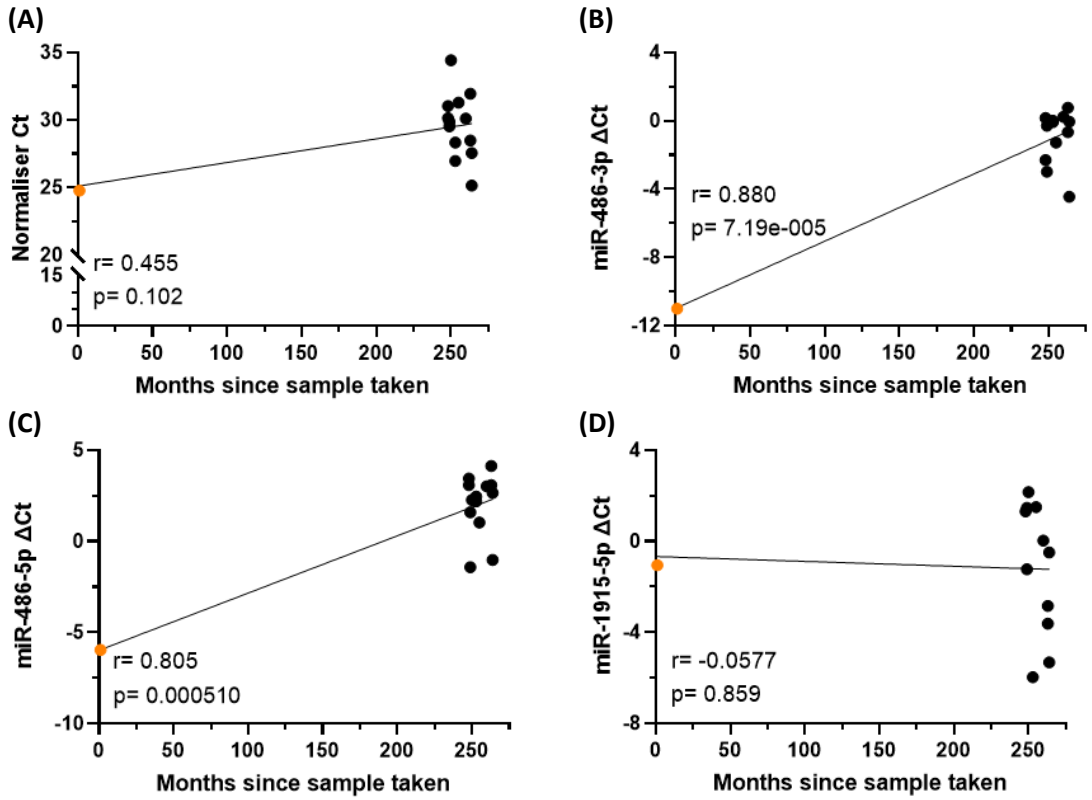


Figure 4.18 Linear regression of time since sample extracted on miR expression including AML042. **(A)** Normaliser Ct value relationship with age (n=14). **(B)** miR-483-3p ΔCt (n=13). **(C)** miR-483-5p ΔCt (n=14). **(D)** miR-1915-5p ΔCt (n=12).

Excluding the most recently received sample (AML042) leads to more homogeneous expression values for all miR and the NV (**Figure 4.19**).

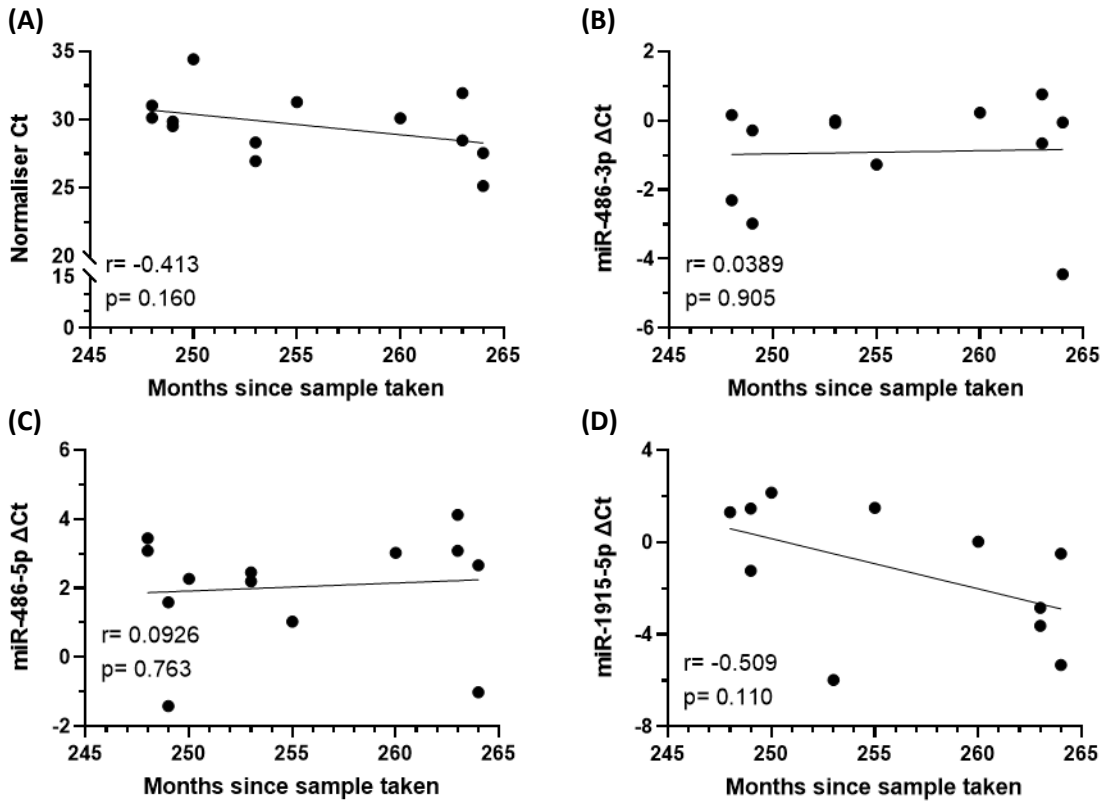


Figure 4.19 Linear regression of time since sample extracted on miRNA expression excluding AML042. **(A)** Normaliser Ct value relationship with age (n=13). **(B)** miR-483-3p ΔCt (n=12). **(C)** miR-483-5p ΔCt (n=13). **(D)** miR-1915-5p ΔCt (n=11). No significant differences shown between miR and endogenous control expression when analysing the time of RNA extraction from time of sample extraction.

There is still a trend in miR-1915-5p when observing the time since sample taken, this miR may degrade faster than the others (**Figure 4.19D**).

4.11 Melt Curve Analysis

Melt curves were performed with every plate to evaluate the primer performance. All normalisers (miR-222-3p and miR-Let-7i-5p) had very uniformed melt curves (**Figure 4.20**) along with the UniSP4 and 6 (spike in products). When product was detected, there was a uniform peak at 70^{o}C other than in miR-1915-5p where the peak was at 80^{o}C. Primers were all predesigned and wet lab validated by Qiagen.

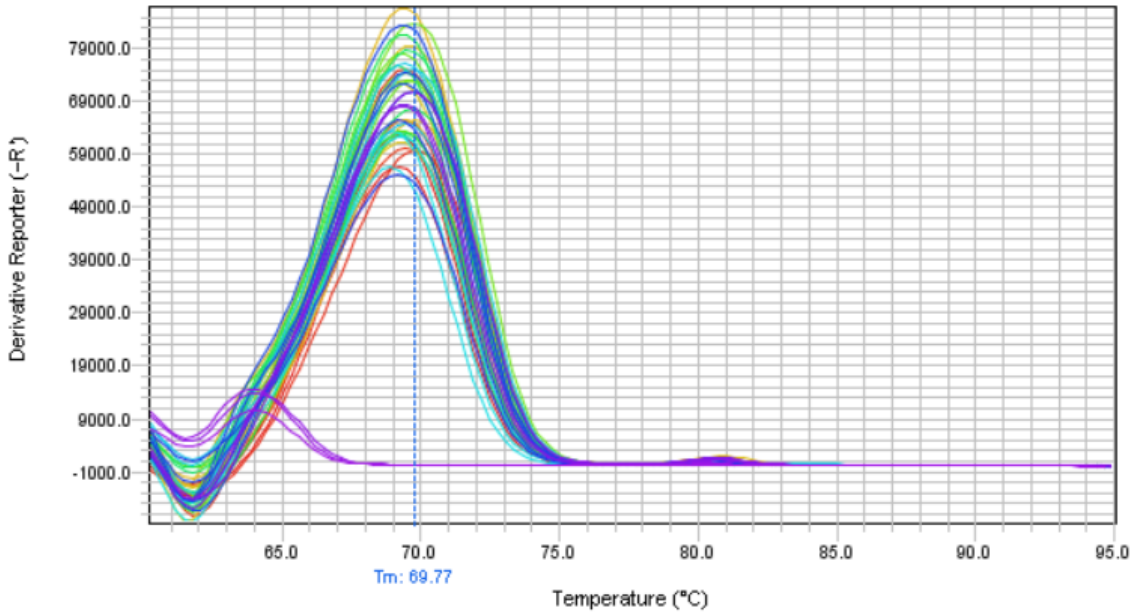


Figure 4.20 Melt curve of miRNA-222-3p. Singular peak at the 70^{o}C, caused by the dissociation of the double-stranded DNA during the ramping up of heating. This image was generated from melting the miR-222-3p primers. As there is only a single peak, we can presume only one product has been formed. Of note is the purple peak at 64^{o}C, this is the control recording from a tube of all qPCR reagents except cDNA template.

Chapter 5. Discussion

Whilst being a rare cancer type, AML is one of the most common leukaemias, and around 3100 patients are diagnosed each year in the UK. Paediatric AML is treated effectively compared to adults suffering from the disease (nearly 70% chance of survival in some paediatric cohorts high relapse rates of up to 35% (Zwaan et al., 2015)), potentially due to differences in the mechanisms of the disease as well as different responses to and alternative treatment options. Within adults and children there are indicators of the patients' prognosis, that can predict response to treatments and post diagnosis survival. This study hoped to identify novel targets for therapeutic treatment and establish whether miRs could be used as biomarkers of prognosis circumventing the need for cytogenetic tests and speeding up the time to treatment. Previously a list of genes had been derived from the analyses of RNA-seq datasets and needed further investigation to evaluate their potential to act as targets for treatment. Following analysis of the pathways these miRs were involved in, ROC analysis was performed to establish the power of these DEmiRs when combined in a model and used as a predictor of risk group. Additionally, qPCR was carried out on serum samples from a variety of AML patients to see whether the increased expression of miRs in mRNA-seq data translated into increased expression in the sera.

5.1 Potential roles of LAAs in AML

Differential gene expression analysis was performed; analysis of the results provided large gene lists of genes differentially expressed when comparing risk subgroups in AML. Nine of these genes were identified as having differential expression that correlates with OS and on this basis were investigated further using STRING, gene ontology and a literature-based searches to determine their function and the signalling pathways they were involved in.

PANTHER analysis showed that the greatest proportion of the DGE list was involved in binding, followed by catalytic activity, acted as molecular function regulators, had transporter and/or transducer activity. Numerically fewer of the DEGs were involved in structural, adaptor and translational regulatory activity. All gene lists (TCGA: poor vs good, intermediate vs good, TARGET: high vs low and standard vs low) were involved in very similar molecular functions (**Figure 3.7**), with a high amount involved in binding and catalytic activity, there were no noticeably significant differences between the groups. There were also similarities between all groups biological processes; most genes played roles in cellular processes, followed by metabolic processes and biological regulation. Unlike the molecular functions, there were noticeable differences in the biological processes that differed between the high versus low TARGET risk subgroups compared with the standard versus low-risk subgroups. The high-risk subgroup

comparison has a larger proportion of DEGs involved in developmental processes, while TARGET high versus low-risk subgroups also have a larger percentage of genes that are involved in biological adhesion and multicellular organism processes. When evaluating the key pathways from the gene lists, the proportion of the genes involved in the Wnt and CHD1 signalling pathway in the poor versus good TARGET risk subgroups is much larger proportion than both TCGA risk subgroup comparisons and intermediate versus good from the TARGET study. This suggest that in the highest risk paediatric patients their diseased cells were more likely to experience some alterations in the Wnt or cadherin pathway, and these pathways could make these better pathways to try and target for that group of patients.

BIRC5 is the LAA which has been researched extensively with regards to its functions and molecular interactions. Examining some of these studies demonstrates BIRC5s vital role in cancers and potential roles in AML. Targeting some of these pathways to neutralise the impact of BIRC5 may provide a new avenue for therapy for AML patients. BIRC5 is essential for successful mitosis, binding to both CPC and Aurora-B kinases, whilst in the nucleus it leads to correct mitotic spindle formation. If BIRC5 was inhibited, the rate of division would be reduced, and this could lead to cell death. STRING analysis highlighted the relationship between the other partners involved in the regulation of mitosis along with very strong correlation analysis with Aurora-B.

BIRC5 is a member of the IAP family, that reduces apoptosis by binding to and inhibiting caspases 3 and 7, both of which are involved in the programmed cell death pathway. p53, a critical protein in cellular control known as the "Guardian of the Genome," was also recently found to be regulating BIRC5 expression (Feroz & Sheikh, 2020). Downregulation by p53 leads to an increase in cell death within the pathways it acts upon, it is therefore reasonable to see why downregulation of BIRC5 could lead to higher levels of apoptosis. One of the hallmarks of cancer is resisting cell death which high levels of BIRC5 would cause through the prevention of apoptosis, this could be why it is found in high levels in many cancers (Li et al., 2019). Furthermore, BIRC5 was found to be a downstream target of the Wnt pathway, an important pathway in the regulation of transcription as well as one of the main pathways highlighted via PANTHER analysis of the datasets. Another axis of BIRC5 function is the competitive inhibition of the CDK4/p16INK4a complex, which prevents p16INK4a causing the arrest of the cell cycle and is another potential way BIRC5 prevents cell death (Suzuki et al., 2000).

As well as preventing apoptosis BIRC5 may confer resistance to therapeutic drugs, particularly when regulated by C-Myc and Sp-1 which are also regulated by MAPK (Zhang et al., 2015). This pathway increases the level of BIRC5 and therefore decreases cell death and in CML BIRC5 has been shown to increase imatinib resistance (Carter et al., 2006). Overexpression of BIRC5 has

been shown to confer resistance to FLT3 inhibitors in cell lines (Zhou et al., 2009), such that if BIRC5 levels could be reduced then FLT3 inhibitors may become more effective in the treatment of patients. However, there could be difficulty in targeting therapies specifically to cancerous cells.

Due to BIRC5's roles in key cell processes such as angiogenesis, higher levels of it are often found in cancer. BIRC5 is found in most adult AML patient samples, with increased expression leading to chemotherapeutic resistance and having an association with a poorer prognosis. BIRC5 could potentially be expressed in cancers due to global genome wide DNA hypomethylation (Lyu et al., 2018). It would make an ideal target due to its consistent expression in cancers and lack of expression in normal healthy tissue, however due to the great number of pathways it acts upon it may be difficult to control any upstream or downstream co-targets of it.

Following the evaluation of BIRC5 expression and patient clinical features, we found patients with core binding factor (CBF) AML and in particular those with inv(16) and higher levels of BIRC5 had lower survival rates. Chromosomal 16 alterations are found in 5-8% of AMLs (Lv et al., 2020). As such BIRC5 may be a target for treatment in a small subset of AML patients with AML inv(16) (Greiner et al., 2021). Expression of BIRC5 was also found to be increased in cases with monosomy 7/loss of 7q or a t(15;17), which could provide another small subset of patients that BIRC5 targeted treatment would benefit.

Whilst CEACAM3 has predominantly been shown to assist the phagocytosis of bacteria by triggering tyrosine kinases, this process isn't a known driver of cancer or leukaemia. Its co-expression with calcium and zinc binding proteins, could just be an indicator of the relationship with immune response due to bacteria. Whilst a reduction in phagocytosis may not help in AML, lower levels of ROS could influence cancer cell death caused by oxidative stress. Being expressed on the surface of granulocytes may make CEACAM3 a target for immunotherapy however there could be difficulty in specifically targeting the cancerous cells, having been found in high levels during blastic crisis in CML, testing with HVs would need to be carried out to determine whether the expression was restricted to AML cells. Grb14 functions as a negative regulator of CEACAM3 and offers a potential method of targeting the pathways CEACAM3 acts upon. Observing the expression of CEACAM3 in GRB14 transfected cell lines could be valid way to confirm the interaction as well as monitor any downstream effects.

As the MAGE family are CT antigens, expression is often found in a variety of cancers due to global promoter hypomethylation. In addition, CT antigens are often restricted in their expression in healthy tissues to immune privileged sites such as the testes and placenta that lack MHC class I. This cancer-specific expression makes them ideal targets for immunotherapeutic

treatment. MAGEF1 was found to cause ubiquitination and degradation of MMS19, and through MMS19 interacted with the CIA pathway. The control of iron sulphur clusters is important for ERCC2 (a DNA repair enzyme) and may be why STRING analysis showed a cluster of proteins involved in chromosomal maintenance. If knockdown or increased expression of MAGEF1 could be achieved with treatment, then the regulation of genomic integrity through the CIA pathway may be possible. MAGEF1 was also found to bind with EID proteins although it is unclear which is regulating the other as EID1 can bind and inhibit EP3OO which is essential for the differentiation of cells. In AML the blast cells grow uncontrollably and do not differentiate and so targeting this pathway could be viable to force erythropoiesis and ensure only functional red blood cells released from the BM. MAGE binding with TRIM family members could possibly inactivate p53 and further destabilise the cells, although the relationship may be with other MAGE family members and not MAGEF1.

STEAP1 was shown to be expressed in a multitude of cancer types and might stimulate cancerous proliferation by the modulation of ions, although it has only a single beta heme group compared to other STEAP family members and so this may not be its primary function. In colorectal cancer a reduction of STEAP1 caused an increase in ROS, indicating different oxidative stress could assist cancerous cell development. However, STEAP1 slowed the growth of cells due to the regulation of NRF2 that protects against oxidative damage (Ma, 2013). Regulation of STEAP1 by eIF4E may provide a way to control the downstream applications, its accumulation in the cell membrane makes it a good target for immunotherapy.

If the same findings apply in AML as they do in A/B breast cancer cell lines then reducing the level of MORC4 would lead to lower expression of STAT3 and increased expression of MORC4 showing higher levels of BCL-2; a known anti-apoptotic protein, that also confers resistance to treatment. However, there are already treatment options available that can target BCL-2 in untreated patients over 75 with a poor prognosis (Venetoclax) (Wei et al., 2020). The identification that miR-338-3p may inhibit MORC4 opens possibilities to encourage cell death and slow migration, by increasing the expression of miR-338-3p, however the other pathways and interactions pertaining to this miR need to be examined.

As VGLL4 is an inhibitor that competes with YAP to bind with TEADs in the Hippo pathway, it could be used to cause a reduction in Hippo signalling that would cause increased proliferation and anti-apoptotic effects. However, these anti-apoptotic effects may not be desirable in AML treatment, where the cells need help differentiating. Whilst this does highlight a potential route to alter the Hippo pathway by inhibiting the expression of VGLL4 and therefore enabling more apoptosis providing space for healthy BM cells to mature. Correlation analysis revealed a close relationship with β-catenin and further research in GC supports the idea that VGLL4 inhibits β-

catenin and TCF4, ultimately lowering gene expression controlled by the Wnt pathway which are responsible for some aspects of the cell cycle. Overexpression of VGLL4 increased levels of CDH1, whilst a study of AML identified lower levels of CDH1 correlating with shorter OS (Zhang et al., 2017). This potentially could be reversed to increase the expression of VGLL4 to increase CDH1 levels to try and achieve better OS rates. VGLL4 was also discovered to reduce the expression of PD-L1, a protein that is typically found in tumours and whose attachment to its receptors on T-cells suppressed anti-tumour immunity by blocking T-cell activation signals (Sun et al., 2018). This was supported by the correlation values between VGLL4 and PD-L1, providing another way VGLL4 could be targeted for immunotherapeutic benefit by preventing PD-L1 from inhibiting the immune response through the disease course of AML (Giannopoulos, 2019). The stabilization of IRF2BP2 by VGLL4 might repress IL-1ß/TNFα signalling via NF-κB, leading to AML cell death (Ellegast et al., 2021), as such an increase in VGLL4 may be used to clear the BM of immature blast cells and provide space for healthy cells to mature. There is evidence VGLL4 may be inhibited by miR-130a and provides options for treatment by addition of a miR-130a mimic although other effects of this would need to be evaluated further and whether the effects would be helpful in treatment of AML due to the various pathways VGLL4 is involved with, it may not be specific enough.

SLC34A2 functions as a transport protein that moves phosphates inside the cells in a sodium dependant manor, whilst ion transport is important in cancers other functions of SLC34A2 may be more relevant to AML. It has been shown that SLC34A2 interferes with the PI3K and AKT signalling pathway which triggers downstream activation of RTKs, this pathway is a highly controlled pathway that finishes by phosphorylating ribosomal protein S6 that then promotes cell proliferation and synthesis of proteins (Hemmings & Restuccia, 2012). SLC34A2 may influence this pathway due to the conversion of PIP3 to PIP2 by PTEN that antagonises the AKT pathway (Stambolic et al., 1998), there was a strong correlation with PTEN in our analysis. This would be a key pathway to target in AML, by reducing the amount of phosphorylating ribosomal protein S6 it might be possible to slow the production of immature cells that characterise AML. Interactions in neuroblastomas were found between SLC34A2 and miR-25 which binds with Gsk3β and antagonises the Wnt pathway, this is supported by our correlation analysis with Gsk3β. Research in bladder cancer found overexpression of SLC34A2 to be a poor prognostic biomarker and a relationship between it and C-Myc was found and supported in our correlation analysis. C-Myc is a family of proteins that regulate apoptotic cell death and an important target in cancers and provides another important pathway for targeting.

Little research has been done on the actions and pathways of SAGE1 and as such it is an area of potential interest for further investigation. As SAGE1 is restricted to male germ line cells and

tumour cells it could potentially be activated by global promoter hypomethylation which could explain why it is often found expressed in AML and other cancers. STRING analysis showed co-expression with NXF2 in digestive tract carcinomas, however no shared pathways were found.

MELTF partly functions as a method of iron binding in the cell membrane, although this may not be its main role. Iron is involved in a variety of processes and could influence AML by affecting cell proliferation or forming ROS. Iron chelating agents are already used as a form of induction therapy in AML and proteins associated with iron are commonly found as co-factors for proteins involved in AML which can have a range of negative effects when dysregulated (Weber et al., 2021), it is possible that MELTF may be affecting AML on this axis. The other axis MELTF may be using to influence cancers is through the generation of plasmin, in gastric cancer the inhibition of MELTF suppressed the invasiveness of the GC cells, this supports the idea that the plasmin generated by MELTF could cause extracellular degradation along with detachment of endothelial cells and lead to cell death by this avenue. The plasmin may also be aiding leukaemic cells in their metastasis (Scherrer et al., 1999).

5.2 Pathways of DEmiRs

The DEG analysis also identified several miRs, the pathways of some of these DEmiRs were investigated based on having >2-fold change, a literature search, and their effect on OS as determined using UCSC Xena browser.

p53 has been demonstrated to regulate miR-1915 in several investigations (Nakazawa et al., 2014; Wan et al., 2017), in HCC its levels were significantly decreased when p53 was knocked down and cells were under oxidative stress. This could have been in response to DNA damage, the p53 regulates the processing from pri to pre miR and so reduces its levels. Through this axis p53 negatively regulates BCL-2 expression by modulating miR-1915 expression, thus preventing BCL-2 binding with the apoptotic proteins and avoiding cell death. In the context of AML this could contribute to the overcrowding of the BM. The therapeutic area to target would be altering miR-1915 expression to allow BCL-2 to carry out its normal functions and lead to apoptosis of AML cells.

The apoptosis of lung cancer cells could be prevented by the downregulation of developmentally regulated GTP binding protein 2 and PBX2 by inhibition from miR-1915-3p. DRG and PBX2 are typically involved in VP16-induced apoptosis and higher expression of miR-1915-3p prevented etoposide induced apoptosis, within AML the downregulation of this miR could prove useful to encourage apoptosis of the cells. In renal progenitor cells, miR-1915 inhibited the expression of Prominin 1 (PROM1) and PAX2. PROM1 is a known cancer biomarker it also may have a role in

the Wnt/β-catenin pathway, angiogenesis is encouraged by the activation of Wnt and increased expression of VEGF and IL-8 (Barzegar Behrooz et al., 2019), the Wnt pathway was also highlighted during panther analysis as a key pathway in the gene set. By altering the expression of miR-1915 it may present an opportunity to target the Wnt via the interactions of PROM1. PAX2 may maintain the tissue specific stem cells by preventing terminal differentiation and apoptosis (Lang et al., 2007), and could play a role in the development of cancers including AML where blast cells do not differentiate fully. A decrease in the expression of PAX2 may be useful to encourage differentiation of erythroid cells.

In breast cancer patients miR-1915-3p was upregulated and its overexpression inhibited DUSP3. The normal role of DUSP3 is to inhibit extracellular signal-regulated protein kinase which prevents cell proliferation and tumour growth, when inhibited by miR-1915 it would increase proliferation in the cell and may explain the miRs high expression in AML.

miR-1915 was found to target RAGE, which resulted in tumour suppression and could be a viable target for therapy. RAGE is typically found in stress environments and its upregulation increases the survival of the cell, enhancing autophagy while diminishing apoptosis (Sparvero et al., 2009). A study investigating RAGE expression in a cohort of AML patients found that overexpression is common but has no effect on the prognosis (Chai et al., 2013), suggesting an avenue of limited value for further investigations.

Using miR-1915 as a therapeutic target may be difficult due to all the pathways it is found in; cell line experiments would need to be carried out to determine the effects of mimics or inhibitors. Whilst none of the experiments or research previously has been done in AML or AML cell lines it is possible it acts in the same manner in BM cells and that requires further investigation. In conclusion miR-1915 is a key component of multiple pathways and a potential driver of leukemogenesis.

Following a study of AML patients, 40 of 45 cytogenetically normal patients showed increased miR-486 expression, especially in M2 but still increased in M0 and M1 (Seyyedi et al., 2016b). Involved in the JAK-STAT signalling pathway by the silencing of SOCS2, the JAK-STAT pathway leads to an increased level of proliferation in cells. By silencing SOCS2 (inhibitor of JAK-STAT), miR-486 can lower cell proliferation, which is useful in AML where the continuous production of immature cells progresses the disease. Furthermore, in muscle cells it has been shown to regulate the PI3K/AKT pathway via an interaction with the upstream targets PTEN and FOXO1 that inhibits their expression. Their inhibition would lead to less cell growth and lower cell survival, and this was found to occur in CML (Wang et al., 2015a). Whilst AML and CML are not

the same disease, they do share some clinical similarities especially when considering CML blast crisis and AML (ie >20% blasts in the BM).

In the TF-1 erythroleukaemia cell line, miR-486 showed increased expression under hypoxic conditions. Hypoxic conditions are thought to provide a survival advantage to AML cells in the BM (Jiang et al., 2021), whilst increased expression of miR-486 by HIFs may provide a further advantage and should be explored more. SIRT1 was found to be a target of miR-486 and interacts with signalling pathways like Wnt (featured in PANTHER analysis) and Notch and as such can have an effect on the proliferation, differentiation and apoptosis of cells (Ren et al., 2019), this could provide opportunities to interrupt these pathways. miR-486-3p was also shown to inhibit BAF chromatin remodeling complex subunit expression in erythroid cells, which also reduces γ-globin expression. γ-globin expression can lead to increased foetal haemoglobin production, and this could be a predictor for a better prognosis in AML and MDS patients treated with Decitabine (Lübbert et al., 2017).

miR-486 can act as a tumour suppressor in oesophageal cancer by targeting CDK4 and BCAS2, and overexpression of miR-486 decreases the EC9706 cells' ability to migrate and invade. When miR-486 levels were increased and CDK4 and BCAS2 were being suppressed, there was a reduction in their downstream targets' caspase-3 and the signalling molecule p21. This reduction of caspase 2 and p21 would reduce the levels of apoptosis in the cells. Inhibition of the SMAD2 protein was also found to be caused by miR-486, this further affected TGF-β and TGF-β-dependent gene expression which in normal haematopoiesis is a negative regulator of proliferation that can also stimulate differentiation and apoptosis (Dong & Blobe, 2006). In leukaemia the homeostatic effects of TGF-β can be seen. One of the mechanisms behind this may be the dysregulation of miR-486 which could influence this pathway via SMAD2 and prevent TGF-β performing its anti-apoptotic role.

When miR-486-5p was overexpressed in HCC cell lines there was an inhibition of proliferation and migration within those cells. This could be caused by the downregulation of CBL expression, suggesting miR-486 acts as a tumour suppressor by inhibiting this multifunctional protein that can control growth, apoptosis, and invasion in various diseases (Li et al., 2018). If these findgins were the same in AML then miR-486-5p may be able to manipulate the expression of CBL for therapeutic benefit. Finally, in thyroid cancer reduction in the expression of miR-486-3p can lead to the activation of NF-κB2. This protein is a transcription factor, and its transcription leads to higher rates of invasion and induction of anti-apoptotic gene expression (Tian et al., 2005). In AML NF-Kb2 activity was found to allow leukaemic cells to stimulate proliferation and avoid

apoptosis (Zhou et al., 2015), as such artificially increasing the levels of this miR may be a way of reducing the effects of NF-κB2.

miR-378G was found to be involved in a variety of pathways and diseases, although it hasn't been investigated in a leukaemic setting. In colon cancer, evidence was found that it decreased cancerous activities like proliferation and migration via the inhibition of SDAD1 (believed to act upstream from a leukaemia inhibitory factor). Overexpression of miR-378G in colon cancer was also shown to inhibit the expression of β-catenin, suggesting its role in the Wnt pathway, a key pathway identified by our PANTHER research into the DEGs. Suppression of proliferation in colorectal cancer was also shown through miR-378G interacting with BRAF a key component of the RAS/MAPK pathway that functions to regulate growth an division, however BRAF can already be targeting through the use of other drugs such as vemurafenib, dabrafenib or the MEK inhibitor trametinib (Geyer et al., 2019).

In gastric cancer cell lines, it was found to be under expressed, and its expression was restored via the addition of demethylation reagent. When restored to higher levels it suppressed the growth of cancerous cells and restored growth of healthy cells, this provides a key area to be able to target as it also restores correct production of cells. This could be due to the regulation of CDK6 and VEGF signalling. In AML VEGF encourages blast cell survival, proliferation and therapeutic resistance (Kampen et al., 2013), making it a key driver in the development of AML. CDKs can regulate the cell cycle, DNA repair and transcription as well as often being overexpressed in AML, making them a key target for any treatment (Lee & Zeidner, 2019), miR-378G could provide another option for lowering CDK expression.

miR-378G inhibits SUFU, a negative regulator of the hedgehog pathway, showing that higher levels of miR-378G can lead to elevated levels of proliferation and cell growth. The hedgehog pathway can lead to resistance of AML to treatment via chemotherapy or radiotherapy (Terao & Minami, 2019), targeting it via the regulation of SUFU could assist other treatments be more effective. It was also shown to decrease the expression of caspase-3, a key protein in programmed cell death. In AML this could mean that higher levels of miR-378G would lead to less apoptosis and more uncontrolled cell growth leading to immature blast cells being forced out the BM.

miR-378 in cardiomyocytes has been found to target components of the MAPK pathway, involved in the signal transduction pathways that can regulate a variety of key processes such as cell death, proliferation, and differentiation. Expression of the miR inhibit proliferation of the cells and if it has the same function in AML then high levels of it would potentially stop over proliferation.

Examining miR-8086 in literature and the pathways its involved proved difficult, with very few studies investigating it.

5.3 miRs as biomarkers using mRNA-sequencing data

Following ROC analysis on mRNA-seq data, we demonstrated that a combination of miR models could be used to determine the risk group of a patient depending on the expression levels of various miRs. When attempting to account for any variables other than risk prognosis that could have affected the DEmiRs expression; two variables were accounted for, patient age and gender. Through adding these into the model we can conclude that they had no effect on the expression of the cellular miRNAs.

miR-486 and 1915 whilst not as powerful as miR-8086 on their own do have more consistent expression throughout all the cohort analysed. Despite the very promising AUC score (.080) of miR-8086, the results should be taken with caution based on the nature of the data. miR-8086 values may have been artificially high due to a few very high readings in the intermediate risk groups. Potentially miR-8086 does have much higher expression in a small number of cases but due to the inconsistent readings it may not have much value as a biomarker or a therapeutic target. miR-378G showed differential expression in the poor versus good risk group but not in the intermediate versus good. This may be of more use in a clinical setting to help determine the patients with a worse prognosis earlier. Following the step wise addition of multiple miRs into the model, the predictive power increased significantly with each miR added, thus exemplifying the ability to distinguish between risk groups based on expression.

If a stable biomarker signature like those used as hypoxia biomarkers were found in AML, then it could be used as a more rapid alternate method then those currently used to determine risk groups. Whole blood could be taken, spun and the fractions of PMBCs used for RNA extraction, cDNA conversion and qPCR. If the levels of miR were found to be elevated, then the patient risk group could be classified and inform treatment decisions. To be useful in a clinical setting more miRs should need to be added to the model. This would reduce the rate of false positives and negatives caused by any of the miRs not being expressed as expected.

After the evaluation of the GDC TCGA AML dataset utilising the UCSC Xena browser the copy number reinforces the difference in expression across the cytogenetic risk groups that the Davis et al paper also found.

5.4 miRs as biomarkers using qPCR data

Experiments were performed to identify a cell line that could be used as a positive control for analysis, however no cell lines were found that could provide consistent expression of any of the DEmiRs or the normalisers. Two PBL samples were available for analysis but due to the limited sample size they were not used further in analysis and did not demonstrate any significant (below Ct of 37) or consistent miR expression. With no paediatric samples available for analysis, 14 adult serum samples were used for qPCR analysis. Without information on the cytogenetics of all the patients, FAB subtype was used to infer risk categories. Spike-ins (UniSp4 and UniSP6) combined with endogenous controls (miR-Let-7i-5p and miR-222-3p) were used to normalise the samples, their expression was expected to be stable in all samples, provide equal cDNA conversion and allowed for comparisons between samples. Normalisers acted as positive controls by the observation of their presence. Assays needed a Ct value below 37 to be used in the analysis and any above were not included. Technical triplicates all needed to show consistent levels of amplicons for use in analysis. With limited resources available, miR-378G was not examined in the serum samples and so the four miRs investigated in serum samples were: miR-486-3p and -5p, miR-1915-3p and -5p, and miR-8086.

A key component of successful qPCR analysis was the calculation of the NV. In this study two endogenous controls and a spike-in were averaged to provide a value that was applied to all samples to reduce inter-sample variation. Despite the two endogenous control CT values being close in each sample there was still a considerable range across the sample (Ct= 11.3), with the addition of the UniSP6 spike in, the NV the range across the cohort was reduced to (Ct=9.7). This is a significant difference; however, no other normalisers were available to try and reduce the variation, but because of the variation the results should be taken cautiously and symbolise trends rather than a statistical significance. The greatest variation was likely caused by a much higher expression of UniSP6 in AML042 which could have been caused by the diminished age of the sample. The UniSP6 Ct in this sample was 18.9, much lower than the next highest sample (AML012) with a Ct of 26.7. The low UniSP6 Ct value caused the NV for this sample to be much lower and increased the range of the NV between all samples, hence the exclusion from some analysis to observe trends more effectively.

Despite the very promising results following the ROC analysis, miR-8086 did not display the same strong results in the serum qPCR. It showed no consistent expression in any samples, whether that was cells or sera. These results slightly correlate with the mRNA-seq data from TCGA, in that when the miR was present it was found at much higher levels and with more base reads however it was not found very often. This may be why it performed so well in ROC analysis and influenced the models heavily due to a very few samples having a large expression of it. This was differently

accounted for when analysing the mRNA-seq data, with outliers being excluded from the analysis. Examining the miR-8086 reads on miRbase it appears the stem loop sequence is found often but not the mature form of the miR, the primer used for identification of this miR may need altering to accommodate this and the melt curves were difficult to analyse with no product being formed. Following evaluation of the limited data gathered from the cell samples, miR-8086 was only expressed in AML032 at disease diagnosis and not in the follow-up samples taken from patients post-treatment. It was also not detectable in the HV, potentially because it is only found in AML patients but this finding should be confirmed with more cell samples due to its limited expression in sera samples.

miR-486-3p was expressed in all 14 samples, however the non-parametric Mann-Whitney test returned a p value that showed the groups were not significantly different (p=0.77). Once adjusted with the NV, AML042 showed considerably lower ΔCt of all miRs then other samples though this may be affected by AML042 having a lower NV. Interestingly miR-486-3p showed significantly more expression in the female group (p=0.01). No obvious reasons are presented for this, however studies have found some miRs vary in expression between biological sexes (Guo et al., 2017). In conclusion miR-486-3p did not follow the expression pattern expected with regards to the risk subgroups, however the small sample size means more investigation is needed to see if a trend emerges. The primer performed well with a single peak being created in the melt curves, showing a single product was being formed.

When comparing the miR-486 variants, the 5p mature sequence had considerably higher reads on miRbase suggesting its the more common variant. This concurs with the qPCR data showing on average higher levels than the 3p variant. It was found to be expressed in 13 of the samples and the melt curve showed a single uniform product was being created.

Only AML016 showed limited expression of this DEMIR with a raw Ct of 34.9 and a ΔCt of 9.9 and so this is most likely an anomalous result as no other samples showed any other variation in expression. This could have been due to not enough cDNA in this sample, yet this sample showed strong expression with other DEmiRs and so most likely this isn't the case. It could be due to poor primer performance, but the melt curve showed only one product being formed when it was present even if the curve was slightly less consistent than the other primers. This could also have been anticipated by evaluating the miRbase site with the 1915-3p variant having significantly fewer base reads in studies compared to the 5p variant.

Both favourable prognosis patients (AML010 and 027), showed reduced expression levels of miR-1915-5p in line with the mRNA-sequencing data in which intermediate prognosis patients had higher expression then those in the good risk group. However, there were only two

favourable prognosis samples and so whilst there appears to be a trend, it cannot be classed as significant. Using the non-parametric Mann-Whitney test the p value for significant difference (p=0.061) between the groups was close to appearing significant and was the most promising of all DEmiRs analysed. Expression was only recorded in 12 of the 13 sera samples.

Further variables needing to be considered when analysing the results of serum qPCR were; the age of patient at diagnosis, the time since sample taken and biological sex. Potentially the lack of expression in sera, cells and cell lines could have been genuine and the Ct values may accurately depict the levels within those samples. Alternatively, the integrity of the samples could have diminished. Whilst the samples were unlikely to have undergone RNA degradation after being extracted, as they were used very quickly, the samples may have had their integrity diminished prior to the RNA extraction, this could be due to the nature of the samples in the cohort being over 20 years old. The time since sample taken was analysed as a variable to evaluate the effects on miR expression depending on how long ago a sample was collected. The results of linear regression when including the AML042 sample (1 month old sample at time of RNA extraction) showed a significant difference in levels of miR-486-3p and 5p, with a strong but insignificant difference in the NV. AML042 also had a considerably higher expression of UniSP6, suggesting the cDNA conversion was more efficient. This suggests that there was some degradation in the samples analysed, despite some sources claiming proper storage and preparation of samples (Huang et al., 2017) will cause minimal differences for up to 17 years (Matias-Garcia et al., 2020). It should be noted that the exact conditions that the samples had been subject to was an unknown variable. It is possible in the moves between universities, by the freezer, and between freezers at university sites, the serum samples were subjected to some degree of freeze thaw which could have disrupted the RNA content. Whilst gender was not found to significantly alter most of the miRs examined in our cohort of patients, there was a slight trend towards women having slightly higher expression of miR-486-3p. There was also evidence to suggest that older women may express higher levels of some miRs (Sredni et al., 2011), and lower levels of others (Noren Hooten et al., 2013) although no explanations have been provided for why this might be.

There were no significant differences in miR expression with different ages of patient in any of the miRs or in the NV, though there was a very slight trend for miR expression to decrease with age (r=-0.05, -0.04 and -0.12), the values were not significant. There was evidence that some miR expression could be age related (Noren Hooten et al., 2013), however the miRs we chose appeared not to be.

Another critical note is that the NV and the DEmiRs were not compared to expression in a healthy donor cohort and so whether lower levels in low and intermediate risk patients are the same in healthy donors remain to be discerned.

The cell lines that we attempted to use as positive controls were all cultured for this study and had been reported in the literature to express these miRs (**Table 2.4**). They had been grown in accordance with manufacturer's guidance and pellets snap frozen for RNA extraction and stored at -80^{o}C until use. Very few of the assays showed any expression in the cell lines but this could have been due to user inexperience in the process as assays were performed on these samples first.

Limited prognostic information on the patients made grouping them into favourable and poor prognostic groups difficult, with FAB subtypes being all that was available for many of the samples. Being a slightly dated system, it does not consider all the various cytogenetic markers a hospital would consider now. With WHO subgrouping, rather than FAB subtyping, the miRs may have shown differential expression between risk subgroups, however this needs further analysis.

It is important to note that the original investigation of the DEmiRs was due to DGE from mRNA-sequencing data done on whole cells and not in serum like the second part of this study, for this reason our results could be different from the PBL analysis. Another potential reason for the variation in the samples may be due to the difference in the number of WBCs, as a larger number of WBCs could lead to more cell lysis and more spillage which might account for larger amounts of miR in the sera.

Chapter 6. Future directions

The aim of this study was to identify the key pathways in adult and paediatric AML. Previous work had discovered DEG and DEmiRs. Further investigation of these DEG and the DEmiRs is needed to realise their potential as biomarkers and therapeutic targets.

6.1 Antigens role in AML

Whilst analysis of the literature for these GOI found many pathways they were involved with, there was often no research specifically looking at these proteins in AML. To get a more definitive picture of their role in AML, tests should be carried out in cell lines using mimics and knockdowns to observe their effect on cell behaviour.

Further actions on the GOIs identified through this study could involve performing qPCR on patient samples to identify their expression in subgroups of AML, with subsequent comparisons to datasets and determining whether mRNA sequencing data would support the findings. It should be possible to assess these findings when compared to other adult mRNA-sequencing cohorts, this could potentially narrow down the number of antigens that need further investigation and allow a group to investigate only the most significantly DEGs.

BIRC5 had some of the best results when evaluating the OS and DGE analysis, however a number of groups have already investigated this gene in AML, including ourselves. By investigating some of the less well-known genes a novel insight into AML biology could be made.

Any possibilities for use of these GOI as a targeted therapy could be investigated using knockdown and mimics, to analyse the impact of the increase and decrease of these GOI on cancer cell function. Knockdowns would provide knowledge of the protein interactions and their downstream effects; this would help increase our understanding of the cancerous drivers of AML and provide new opportunities for therapeutic targets.

ICC could be performed on PBLs or BM cells to provide more information about the expression of these antigens in the cells or immunofluorescence may be a better option to allow clarification of where the protein accumulates in the subcellular compartments of the cell and proteins it interacts with. This would allow comparisons with what has been found in literature and the corresponding pathways and functional roles. A western blot could be carried out to determine the levels of expression of the proteins inside the cells, if knockdown or mimics were added then this would also be a method of evaluating the levels inside the cells without relying only on qPCR.

In this thesis only a selection of the most promising DEGs with the greatest influence on EFS and OS were further investigated but other genes even more differentially expressed could be analysed for their role in various cellular pathways. If they are more significantly differentially expressed between subgroups, then they could make better predictors of risk groups.

6.2 miRNAs as biomarkers

A way to add value to the analysis performed would be to investigate another mRNA-sequencing cohort, potentially the Vizome BEAT AML dataset could be used with more than 500 patients. For the miRs we have identified to be used as biomarkers they would have to be shown to be specific to AML and sensitive for diseased cells. Using a larger validation cohort of patients including some healthy volunteers would provide more evidence of their utility.

Machine learning is a valuable tool that could be used to objectively assess the DEmiRs. A program could be developed to determine the miR signature of each AML patient. After the machine has been trained using patient samples classified by current methods, then we could potentially find different DEmiRs to include as part of a growing biomarker model. The more miRs that are included, the better the accuracy of the predictions which would reduce false positives. However, it could take longer to collect the results if testing involved a great number of miRs and qPCR.

The wide variety of pathways these DEmiRs are involved in provide plenty of opportunities for therapeutic targets. However, this is also a problem as DEmiRs affect multiple pathways which may make it difficult to balance beneficial and adverse effects in the body. Further work on cell lines could be performed to determine exactly which pathways the miRs affect and if the use of mimics and inhibitors is a viable of strategy or if the DEmiRs are just restricted to being a prognostic biomarker.

6.3 qPCR Results

A simple way to increase the quality of results for this investigation would be to investigate a larger cohort with more detailed information on the karyotype of these patients to draw trends and conclusions more accurately. In a cohort with greater than three patients in each group parametric tests could be used to perform to achieve more significant analyses. More detailed cytogenetic information such as: translocations, inversions and deletions would equate to better real world risk prognosis subgrouping, rather than the FAB type correlations used in this thesis, and would allow for a more accurate correlation between the associated risk of an AML patient

and the expression of miRs. Furthermore, more recent samples would allow for the reduction in variables that could affect the miR expression values in qPCR despite our samples showing no significant difference when comparing the date the sample was taken. Another benefit of newer samples would be in the analysis of miRs as biomarkers in a hospital environment. The sample would be collected and analysed in a much shorter time frame.

Other differentially expressed miRs could also be investigated to see whether they show greater expression differences between prognosis subgroups. If enough miRs were analysed (>30), then it is predicted that at least one would be consistently expressed in every patient sample and could be used as a more accurate NV then the combination of endogenous controls and a spike in used here. The inclusion of healthy volunteers could provide an additional control group for assessment by qPCR. The Guinn lab continues to receive AML samples from Dr Kim Orchard at SUTH and is applying for ethical approval to receive additional adult AML samples from Dr David Allsup's group at HUTH. Further adult AML samples would allow for a more recent cohort of samples to be analysed, as well as more detailed patient information to facilitate their assignment to risk subgroups.

Paediatric AML samples would allow us to investigate differentially expressed miRs in younger patients. miR-378G showed promising results via ROC analysis in adults with poor versus good risk features, along with being in the paediatric gene list, however we lacked funding and access to a collaborator with these rare paediatric AML samples to pursue this arm of investigation. In the future peripheral blood could be collected and would allow qPCR to be performed on PBLs. This would allow studies on purified blasts or analysis of the cells the mRNA-sequencing data was originally performed on.

After qPCR, knockdown models could be used to evaluate the effects of these miRs on cell function and identify pathway partners by monitoring the expression of predicted downstream targets following knockdown. We could perform mRNA-sequencing on the miRNA knocked-down cells to determine the impact of changes in expression on other genes. If mimics were available these could be used to simulate a higher expression level inside cells, if the predicted downstream targets were altered accordingly, it would reinforce being a functional pathway partner that could be targeted via therapy.

Following Daniel Fletcher's review 'MiRNA expression in acute myeloid leukaemia: new targets for therapy?' (Fletcher et al., In press, eJHaem), there were promising targets identified, that were involved in key pathways subverted in AML. This group of miRs could be investigated further with the use of qPCR to validate the findings. Upregulation of the genes targeted by miRs could be checked using ICC to analyse the targets expression. ICC would not show an increase in

the miR expression but could show the levels and cellular localisation of the proteins they act upon.

Finally, if the miRs do show promise and reliably categorise AML patients into the correct prognosis subgroup then it could be used in hospitals to provide rapid results regarding risk subgroup, survival and best practise targeted treatment. Whilst all the samples analysed in this study were pre-treatment, there is potential to evaluate patients' post-treatment to see whether there is a difference in miR expression during rounds of treatment and response to therapy. This could facilitate the monitoring of minimal residual disease and the prediction of relapse. With the potential of using miRs in a therapeutic setting to treat those most likely to experience relapse, miRs may also provide a more rapid and cheaper alternative to cytogenetic analysis especially appealing when there are so many possible translocations in AML patients.

Chapter 7. References

Abdelfattah, A. M., Park, C. & Choi, M. Y. (2014) Update on non-canonical microRNAs. *Biomolecular concepts*, 5(4), 275-287.

Acharya, N., Sabatos-Peyton, C. & Anderson, A. C. (2020) Tim-3 finds its place in the cancer immunotherapy landscape. *J Immunother Cancer*, 8(1).

ACS (2018) *Risk factors for Acute Myeloid Leukemia.* Available online: https://www.cancer.org/cancer/acute-myeloid-leukemia/causes-risks-prevention/risk-factors.html [Accessed.

ACS (2020) *Chemotherapy for Acute Myeloid Leukaemia* Available online: https://www.cancer.org/cancer/acute-myeloid-leukemia/treating/chemotherapy.html [Accessed.

Adams, G. P. & Weiner, L. M. (2005) Monoclonal antibody therapy of cancer. *Nature Biotechnology*, 23(9), 1147-1157.

Al-Hussaini, M., Rettig, M. P., Ritchey, J. K., Karpova, D., Uy, G. L., Eissenberg, L. G., Gao, F., Eades, W. C., Bonvini, E., Chichili, G. R., Moore, P. A., Johnson, S., Collins, L. & DiPersio, J. F. (2016) Targeting CD123 in acute myeloid leukemia using a T-cell–directed dual-affinity retargeting platform. *Blood*, 127(1), 122-131.

Altieri, D. C. (2003) Survivin, versatile modulation of cell division and apoptosis in cancer. *Oncogene*, 22(53), 8581-8589.

Andersson, L. C., Nilsson, K. & Gahmberg, C. G. (1979) K562--a human erythroleukemic cell line. *Int J Cancer*, 23(2), 143-7.

Anguille, S., Van Tendeloo, V. F. & Berneman, Z. N. (2012) Leukemia-associated antigens and their relevance to the immunotherapy of acute myeloid leukemia. *Leukemia*, 26(10), 2186-2196.

Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., Le Beau, M. M., Bloomfield, C. D., Cazzola, M. & Vardiman, J. W. (2016) The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*, 127(20), 2391-2405.

Bagger, F. O., Kinalis, S. & Rapin, N. (2019) BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. *Nucleic Acids Research*, 47(D1), D881-D885.

Bail, S., Swerdel, M., Liu, H., Jiao, X., Goff, L. A., Hart, R. P. & Kiledjian, M. (2010) Differential regulation of microRNA stability. *RNA (New York, N.Y.)*, 16(5), 1032-1039.

Bao, Z., Chen, L. & Guo, S. (2019) Knockdown of SLC34A2 inhibits cell proliferation, metastasis, and elevates chemosensitivity in glioma. *J Cell Biochem*, 120(6), 10205-10214.

Bartel, D. P. (2004) MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*, 116(2), 281-297.

Bartel, D. P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2), 215-233.

Barzegar Behrooz, A., Syahir, A. & Ahmad, S. (2019) CD133: beyond a cancer stem cell biomarker. *J Drug Target*, 27(3), 257-269.

Belov, L., de la Vega, O., dos Remedios, C. G., Mulligan, S. P. & Christopherson, R. I. (2001) Immunophenotyping of leukemias using a cluster of differentiation antibody microarray. *Cancer Res*, 61(11), 4483-9.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R. & Sultan, C. (1976) Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol*, 33(4), 451-8.

Benz, F., Roderburg, C., Vargas Cardenas, D., Vucur, M., Gautheron, J., Koch, A., Zimmermann, H., Janssen, J., Nieuwenhuijsen, L., Luedde, M., Frey, N., Tacke, F., Trautwein, C. & Luedde, T. (2013) U6 is unsuitable for normalization of serum miRNA levels in patients with sepsis or liver fibrosis. *Experimental & molecular medicine*, 45(9), e42-e42.

Birnie, G. D. (1988) The HL60 cell line: a model system for studying human myeloid cell differentiation. *The British journal of cancer. Supplement*, 9, 41-45.

Bolouri, H., Farrar, J., Triche, T., Ries, R., Lim, E., Alonzo, T., Ma, Y., Moore, R., Mungall, A., Marra, M., Zhang, J., Ma, X., Liu, Y., Liu, Y., Auvil, J., Davidsen, T., Gesuwan, P., Hermida, L., Salhia, B. & Meshinchi, S. (2018) The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nature Medicine*, 24.

Bonsignore, P., Kuiper, J. W. P., Adrian, J., Goob, G. & Hauck, C. R. (2020) CEACAM3—A Prim(at)e Invention for Opsonin-Independent Phagocytosis of Bacteria. *Frontiers in Immunology*, 10.

Bostrom, J., Lee, C. V., Haber, L. & Fuh, G. (2009) Improving antibody binding affinity and specificity for therapeutic development. *Methods Mol Biol*, 525, 353-76, xiii.

Buntru, A., Kopp, K., Voges, M., Frank, R., Bachmann, V. & Hauck, C. R. (2011) Phosphatidylinositol 3'-kinase activity is critical for initiating the oxidative burst and bacterial destruction during CEACAM3-mediated phagocytosis. *J Biol Chem*, 286(11), 9555-66.
Buntru, A., Roth, A., Nyffenegger-Jann, N. J. & Hauck, C. R. (2012) HemITAM signaling by CEACAM3, a human granulocyte receptor recognizing bacterial pathogens. *Arch Biochem Biophys*, 524(1), 77-83.
Busfield, S. J., Biondo, M., Wong, M., Ramshaw, H. S., Lee, E. M., Ghosh, S., Braley, H., Panousis, C., Roberts, A. W., He, S. Z., Thomas, D., Fabri, L., Vairo, G., Lock, R. B., Lopez, A. F. & Nash, A. D. (2014) Targeting of acute myeloid leukemia in vitro and in vivo with an anti-CD123 mAb engineered for optimal ADCC. *Leukemia*, 28(11), 2213-21.
Cailleau, R., Olivé, M. & Cruciger, Q. V. (1978) Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. *In Vitro*, 14(11), 911-5.
Canaani, J., Beohou, E., Labopin, M., Socié, G., Huynh, A., Volin, L., Cornelissen, J., Milpied, N., Gedde-Dahl, T., Deconinck, E., Fegueux, N., Blaise, D., Mohty, M. & Nagler, A. (2017) Impact of FAB classification on predicting outcome in acute myeloid leukemia, not otherwise specified, patients undergoing allogeneic stem cell transplantation in CR1: An analysis of 1690 patients from the acute leukemia working party of EBMT. *Am J Hematol*, 92(4), 344-350.
Carter, B. Z., Mak, D. H., Schober, W. D., Cabreira-Hansen, M., Beran, M., McQueen, T., Chen, W. & Andreeff, M. (2006) Regulation of survivin expression through Bcr-Abl/MAPK cascade: targeting survivin overcomes imatinib resistance and increases imatinib sensitivity in imatinib-responsive CML cells. *Blood*, 107(4), 1555-1563.
Caserta, S., Mengozzi, M., Kern, F., Newbury, S. F., Ghezzi, P. & Llewelyn, M. J. (2018) Severity of Systemic Inflammatory Response Syndrome Affects the Blood Levels of Circulating Inflammatory-Relevant MicroRNAs. *Frontiers in Immunology*, 8.
Castedo, M., Perfettini, J. L., Roumier, T., Andreau, K., Medema, R. & Kroemer, G. (2004) Cell death by mitotic catastrophe: a molecular definition. *Oncogene*, 23(16), 2825-37.
Casucci, M., Nicolis di Robilant, B., Falcone, L., Camisa, B., Norelli, M., Genovese, P., Gentner, B., Gullotta, F., Ponzoni, M., Bernardi, M., Marcatti, M., Saudemont, A., Bordignon, C., Savoldo, B., Ciceri, F., Naldini, L., Dotti, G., Bonini, C. & Bondanza, A. (2013) CD44v6-targeted T cells mediate potent antitumor effects against acute myeloid leukemia and multiple myeloma. *Blood*, 122(20), 3461-72.
Catalanotto, C., Cogoni, C. & Zardo, G. (2016) MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions. *International journal of molecular sciences*, 17(10), 1712.
Chai, H. Y., Qian, J., Lin, J., Yang, J., Li, Y., Wang, C. Z., Chen, X. X., Chen, Q., Deng, Z. Q., Yao, D. M. & Ma, J. C. (2013) [Expression of RAGE-1 gene in acute myeloid leukemia]. *Zhongguo Shi Yan Xue Ye Xue Za Zhi*, 21(1), 20-4.
Chan, F. C. (2020) *Survutils package.* Available online: https://github.com/tinyheero/survutils
Chao, M. P., Takimoto, C. H., Feng, D. D., McKenna, K., Gip, P., Liu, J., Volkmer, J. P., Weissman, I. L. & Majeti, R. (2019) Therapeutic Targeting of the Macrophage Immune Checkpoint CD47 in Myeloid Malignancies. *Front Oncol*, 9, 1380.
Chen, J., Wang, P., Cai, R., Peng, H., Zhang, C. & Zhang, M. (2019) SLC34A2 promotes neuroblastoma cell stemness via enhancement of miR-25/Gsk3β-mediated activation of Wnt/β-catenin signaling. *FEBS Open Bio*, 9(3), 527-537.
Chen, L., Xiong, L., Hong, S., Li, J., Huo, Z., Li, Y., Chen, S., Zhang, Q., Zhao, R., Gingold, J. A., Zhu, X., Lv, W., Li, Y., Yu, S. & Xiao, H. (2020) Circulating Myeloid-derived Suppressor Cells Facilitate Invasion of Thyroid Cancer Cells by Repressing miR-486-3p. *J Clin Endocrinol Metab*, 105(8).
Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., Guo, J., Zhang, Y., Chen, J. & Guo, X. (2008) Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell research*, 18(10), 997-1006.
Chen, X., Duan, N., Zhang, C. & Zhang, W. (2016) Survivin and Tumorigenesis: Molecular Mechanisms and Therapeutic Strategies. *J Cancer*, 7(3), 314-23.
Chen, Y. T., Panarelli, N. C., Piotti, K. C. & Yantiss, R. K. (2014) Cancer-testis antigen expression in digestive tract carcinomas: frequent expression in esophageal squamous cell carcinoma and its precursor lesions. *Cancer Immunol Res*, 2(5), 480-6.
Cheng, S. Y. & Yue, S. (2008) Role and regulation of human tumor suppressor SUFU in Hedgehog signaling. *Adv Cancer Res*, 101, 29-43.
Coppage, M., Belanger, T., Zauderer, M. & Sahasrabudhe, D. (2007) In vitro generation of tumor specific T cells that recognize a shared antigen of AML: molecular characterization of TCR genes. *Leuk Res*, 31(2), 195-202.
Creutzig, U., Kutny, M. A., Barr, R., Schlenk, R. F. & Ribeiro, R. C. (2018) Acute myelogenous leukemia in adolescents and young adults. *Pediatr Blood Cancer*, 65(9), e27089.

Cristofoletti, C., Picchio, M. C., Lazzeri, C., Tocco, V., Pagani, E., Bresin, A., Mancini, B., Passarelli, F., Facchiano, A., Scala, E., Lombardo, G. A., Cantonetti, M., Caprini, E., Russo, G. & Narducci, M. G. (2013) Comprehensive analysis of PTEN status in Sezary syndrome. *Blood*, 122(20), 3511-20.
Czepulkowski, B. H. (2001) *Analyzing chromosomes / Barbara Czepulkowski, [editor]*. Oxford: BIOS Scientific.
Daver, N., Alotaibi, A. S., Bücklein, V. & Subklewe, M. (2021) T-cell-based immunotherapy of acute myeloid leukemia: current concepts and future developments. *Leukemia*, 35(7), 1843-1863.
Davis, L., Mills, K. I., Orchard, K. H. & Guinn, B.-A. (2020) Identification of Genes Whose Expression Overlaps Age Boundaries and Correlates with Risk Groups in Paediatric and Adult Acute Myeloid Leukaemia. *Cancers*, 12(10), 2769.
De Smet, C., De Backer, O., Faraoni, I., Lurquin, C., Brasseur, F. & Boon, T. (1996) The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. *Proceedings of the National Academy of Sciences*, 93(14), 7149-7153.
Deng, H., Guo, Y., Song, H., Xiao, B., Sun, W., Liu, Z., Yu, X., Xia, T., Cui, L. & Guo, J. (2013) MicroRNA-195 and microRNA-378 mediate tumor growth suppression by epigenetical regulation in gastric cancer. *Gene*, 518(2), 351-9.
Deng, X. & Fang, L. (2018) VGLL4 is a transcriptional cofactor acting as a novel tumor suppressor via interacting with TEADs. *Am J Cancer Res*, 8(6), 932-943.
Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F. & Hannon, G. J. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014), 231-235.
DiNardo, C. D. & Cortes, J. E. (2016) Mutations in AML: prognostic and therapeutic implications. *Hematology. American Society of Hematology. Education Program*, 2016(1), 348-355.
Dombret, H. & Gardin, C. (2016) An update of current treatments for adult acute myeloid leukemia. *Blood*, 127(1), 53-61.
Dong, M. & Blobe, G. C. (2006) Role of transforming growth factor-beta in hematologic malignancies. *Blood*, 107(12), 4589-4596.
Doyle, J. M., Gao, J., Wang, J., Yang, M. & Potts, P. R. (2010) MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases. *Mol Cell*, 39(6), 963-74.
Duan, X., Guo, G., Pei, X., Wang, X., Li, L., Xiong, Y. & Qiu, X. (2019) Baicalin Inhibits Cell Viability, Migration and Invasion in Breast Cancer by Regulating miR-338-3p and MORC4. *Onco Targets Ther*, 12, 11183-11193.
Duchaine, T. F. & Fabian, M. R. (2019) Mechanistic Insights into MicroRNA-Mediated Gene Silencing. *Cold Spring Harb Perspect Biol*, 11(3).
Dunn, L. L., Sekyere, E. O., Suryo Rahmanto, Y. & Richardson, D. R. (2006) The function of melanotransferrin: a role in melanoma cell proliferation and tumorigenesis. *Carcinogenesis*, 27(11), 2157-69.
Ellegast, J. M., Alexe, G., Hamze, A., Lin, S., Pimkin, M., Ross, L., Dharia, N. V., Robinchaud, A. L., Conway, A., Khalid, D., Wunderlich, M., Benajiba, L., Nabet, B., Gray, N. S., Orkin, S. H. & Stegmaier, K. (2021) Unleashing Cell-Intrinsic Inflammation As a Strategy to Kill AML Blasts. *Blood*, 138, 3305.
Esh, A. M., Atfy, M., Azizi, N. A., El Naggar, M. M., Khalil, E. E. & Sherief, L. (2011) Prognostic significance of survivin in pediatric acute lymphoblastic leukemia. *Indian J Hematol Blood Transfus*, 27(1), 18-25.
Fabbri, M., Garzon, R., Andreeff, M., Kantarjian, H. M., Garcia-Manero, G. & Calin, G. A. (2008) MicroRNAs and noncoding RNAs in hematological malignancies: molecular, clinical and therapeutic implications. *Leukemia*, 22(6), 1095-1105.
Fang, Z. H., Dong, C. L., Chen, Z., Zhou, B., Liu, N., Lan, H. F., Liang, L., Liao, W. B., Zhang, L. & Han, Z. C. (2009) Transcriptional regulation of survivin by c-Myc in BCR/ABL-transformed cells: implications in anti-leukaemic strategy. *J Cell Mol Med*, 13(8b), 2039-52.
Feroz, W. & Sheikh, A. M. A. (2020) Exploring the multiple roles of guardian of the genome: P53. *Egyptian Journal of Medical Human Genetics*, 21(1), 49.
Fukunaga, K. (2013) *Introduction to statistical pattern recognition*.Elsevier.
Gallo, A., Tandon, M., Alevizos, I. & Illei, G. G. (2012) The majority of microRNAs detectable in serum and saliva is concentrated in exosomes. *PloS one*, 7(3), e30679.
Ganesan, J., Ramanujam, D., Sassi, Y., Ahles, A., Jentzsch, C., Werfel, S., Leierseder, S., Loyer, X., Giacca, M., Zentilin, L., Thum, T., Laggerbauer, B. & Engelhardt, S. (2013) MiR-378 Controls Cardiac Hypertrophy by Combined Repression of Mitogen-Activated Protein Kinase Pathway Factors. *Circulation*, 127(21), 2097-2106.
Gao, J., Gao, L., Li, R., Lai, Z., Zhang, Z. & Fan, X. (2019) Integrated analysis of microRNA-mRNA expression in A549 cells infected with influenza A viruses (IAVs) from different host species. *Virus Res*, 263, 34-46.
Geyer, M. B., Abdel-Wahab, O. & Tallman, M. S. (2019) BRAF in the cross-hairs. *Expert review of hematology*, 12(3), 183-193.

Giannopoulos, K. (2019) Targeting Immune Signaling Checkpoints in Acute Myeloid Leukemia. *Journal of clinical medicine*, 8(2), 236.
Gibson, B. E. S., Wheatley, K., Hann, I. M., Stevens, R. F., Webb, D., Hills, R. K., De Graaf, S. S. N., Harrison, C. J. & for the United Kingdom Childhood Leukaemia Working Party, a. t. D. C. O. G. (2005) Treatment strategy and long-term results in paediatric patients treated in consecutive UK AML trials. *Leukemia*, 19(12), 2130-2138.
Gill, S. I. (2019) How close are we to CAR T-cell therapy for AML? *Best Pract Res Clin Haematol*, 32(4), 101104.
Goldman, M. J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A. N., Zhu, J. & Haussler, D. (2020) Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology*, 38(6), 675-678.
Gomes, I. M., Maia, C. J. & Santos, C. R. (2012) STEAP proteins: from structure to applications in cancer therapy. *Mol Cancer Res*, 10(5), 573-87.
Greaves, M. F., Maia, A. T., Wiemels, J. L. & Ford, A. M. (2003) Leukemia in twins: lessons in natural history. *Blood*, 102(7), 2321-2333.
Greiner, J., Brown, E., Bullinger, L., Hills, R. K., Morris, V., Döhner, H., Mills, K. I. & Guinn, B. A. (2021) Survivin' Acute Myeloid Leukaemia-A Personalised Target for inv(16) Patients. *Int J Mol Sci*, 22(19).
Guinn, B. A., Bland, E. A., Lodi, U., Liggins, A. P., Tobal, K., Petters, S., Wells, J. W., Banham, A. H. & Mufti, G. J. (2005) Humoral detection of leukaemia-associated antigens in presentation acute myeloid leukaemia. *Biochem Biophys Res Commun*, 335(4), 1293-304.
Guo, J., Liu, C., Wang, W., Liu, Y., He, H., Chen, C., Xiang, R. & Luo, Y. (2018) Identification of serum miR-1915-3p and miR-455-3p as biomarkers for breast cancer. *PLoS One*, 13(7), e0200716.
Guo, L., Zhang, Q., Ma, X., Wang, J. & Liang, T. (2017) miRNA and mRNA expression analysis reveals potential sex-biased miRNA expression. *Scientific Reports*, 7(1), 39812.
Gupta, A. & Moore, J. A. (2018) Tumor Lysis Syndrome. *JAMA Oncology*, 4(6), 895-895.
Guy, D. G. & Uy, G. L. (2018) Bispecific Antibodies for the Treatment of Acute Myeloid Leukemia. *Current hematologic malignancy reports*, 13(6), 417-425.
Haferlach, T., Kohlmann, A., Basso, G., Béné, M.-C., Chiaretti, S., Downing, J. R., Hernández, J. s.-M. a., Hofmann, W.-K., Kipps, T. J., Koay, E. S., Kronnie, G. T., Liu, W.-m., MacIntyre, E., Mills, K. I., Papenhausen, P., Preudhomme, C., Rassenti, L. Z., Shurtleff, S. A., De Vos, J., Yeoh, A. E., Williams, P. M., Wieczorek, L. & Foà, R. (2008) The Clinical Utility of Microarray-Based Gene Expression Profiling in the Diagnosis and Sub-Classification of Leukemia: Final Report on 3252 Cases from the International MILE Study Group. *Blood*, 112(11), 753-753.
Hajian-Tilaki, K. (2013) Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian journal of internal medicine*, 4(2), 627-635.
Hao, T., Li-Talley, M., Buck, A. & Chen, W. (2019) An emerging trend of rapid increase of leukemia but not all cancers in the aging population in the United States. *Scientific Reports*, 9(1), 12070.
He, J., Xiao, B., Li, X., He, Y., Li, L. & Sun, Z. (2019) MiR-486-5p Suppresses Proliferation and Migration of Hepatocellular Carcinoma Cells through Downregulation of the E3 Ubiquitin Ligase CBL. *Biomed Res Int*, 2019, 2732057.
Hemmings, B. A. & Restuccia, D. F. (2012) PI3K-PKB/Akt pathway. *Cold Spring Harbor perspectives in biology*, 4(9), a011189-a011189.
Holstein, I., Singh, A. K., Pohl, F., Misiak, D., Braun, J., Leitner, L., Hüttelmaier, S. & Posern, G. (2020) Post-transcriptional regulation of MRTF-A by miRNAs during myogenic differentiation of myoblasts. *Nucleic Acids Res*, 48(16), 8927-8942.
Hong, G., Qiu, H., Wang, C., Jadhav, G., Wang, H., Tickner, J., He, W. & Xu, J. (2017) The Emerging Role of MORC Family Proteins in Cancer Development and Bone Homeostasis. *J Cell Physiol*, 232(5), 928-934.
Hua, Z., Lv, Q., Ye, W., Wong, C.-K. A., Cai, G., Gu, D., Ji, Y., Zhao, C., Wang, J., Yang, B. B. & Zhang, Y. (2006) MiRNA-directed regulation of VEGF and other angiogenic factors under hypoxia. *PloS one*, 1(1), e116-e116.
Huang da, W., Sherman, B. T. & Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1), 44-57.
Huang, L.-H., Lin, P.-H., Tsai, K.-W., Wang, L.-J., Huang, Y.-H., Kuo, H.-C. & Li, S.-C. (2017) The effects of storage temperature and duration of blood samples on DNA and RNA qualities. *PLOS ONE*, 12(9), e0184692.
Hudson, J., Zabrady, K., Kozakova, L., Liao, C., Guerineau, M., Colnaghi, R., Vidot, S., Marek, J., Sreenivas Reddy, B., Lehmann, A. & Palecek, J. (2011) Interactions between the Nse3 and Nse4 Components of the SMC5-6 Complex Identify Evolutionarily Conserved Interactions between MAGE and EID Families. *PloS one*, 6, e17270.

Ibberson, D., Benes, V., Muckenthaler, M. U. & Castoldi, M. (2009) RNA degradation compromises the reliability of microRNA expression profiling. *BMC biotechnology*, 9, 102-102.
Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T. & Tomari, Y. (2010) Hsc70/Hsp90 Chaperone Machinery Mediates ATP-Dependent RISC Loading of Small RNA Duplexes. *Molecular Cell*, 39(2), 292-299.
Jackson, N., Menon, B. S., Zarina, W., Zawawi, N. & Naing, N. N. (1999) Why is acute leukemia more common in males? A possible sex-determined risk linked to the ABO blood group genes. *Ann Hematol*, 78(5), 233-6.
Jetani, H., Garcia-Cadenas, I., Nerreter, T., Thomas, S., Rydzek, J., Meijide, J. B., Bonig, H., Herr, W., Sierra, J., Einsele, H. & Hudecek, M. (2018) CAR T-cells targeting FLT3 have potent activity against FLT3(-)ITD(+) AML and act synergistically with the FLT3-inhibitor crenolanib. *Leukemia*, 32(5), 1168-1179.
Jiang, F., Mao, Y., Lu, B., Zhou, G. & Wang, J. (2021) A hypoxia risk signature for the tumor immune microenvironment evaluation and prognosis prediction in acute myeloid leukemia. *Scientific Reports*, 11(1), 14657.
Jiang, J.-N., Wu, Y.-Y., Fang, X.-D. & Ji, F.-J. (2020) EIF4E regulates STEAP1 expression in peritoneal metastasis. *Journal of Cancer*, 11(4), 990-996.
Jiao, S., Li, C., Hao, Q., Miao, H., Zhang, L., Li, L. & Zhou, Z. (2017) VGLL4 targets a TCF4-TEAD4 complex to coregulate Wnt and Hippo signalling in colorectal cancer. *Nat Commun*, 8, 14058.
Jitschin, R., Saul, D., Braun, M., Tohumeken, S., Völkl, S., Kischel, R., Lutteropp, M., Dos Santos, C., Mackensen, A. & Mougiakakos, D. (2018) CD33/CD3-bispecific T-cell engaging (BiTE®) antibody construct targets monocytic AML myeloid-derived suppressor cells. *J Immunother Cancer*, 6(1), 116.
Kahvejian, A., Svitkin, Y. V., Sukarieh, R., M'Boutchou, M. N. & Sonenberg, N. (2005) Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev*, 19(1), 104-13.
Kampen, K. R., Ter Elst, A. & de Bont, E. S. (2013) Vascular endothelial growth factor signaling in acute myeloid leukemia. *Cell Mol Life Sci*, 70(8), 1307-17.
Kanehisa, M. & Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1), 27-30.
Kaspers, G. (2014) How I treat paediatric relapsed acute myeloid leukaemia. *British Journal of Haematology*, 166(5), 636-645.
Kefas, B., Comeau, L., Floyd, D. H., Seleverstov, O., Godlewski, J., Schmittgen, T., Jiang, J., diPierro, C. G., Li, Y., Chiocca, E. A., Lee, J., Fine, H., Abounader, R., Lawler, S. & Purow, B. (2009) The neuronal microRNA miR-326 acts in a feedback loop with notch and has therapeutic potential against brain tumors. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(48), 15161-15168.
Kim, K., Mitra, S., Wu, G., Berka, V., Song, J., Yu, Y., Poget, S., Wang, D. N., Tsai, A. L. & Zhou, M. (2016) Six-Transmembrane Epithelial Antigen of Prostate 1 (STEAP1) Has a Single b Heme and Is Capable of Reducing Metal Ion Complexes and Oxygen. *Biochemistry*, 55(48), 6673-6684.
Kim, Y.-K. & Kim, V. N. (2007) Processing of intronic microRNAs. *The EMBO Journal*, 26(3), 775-783.
Kleven, M. D., Dlakić, M. & Lawrence, C. M. (2015) Characterization of a single b-type heme, FAD, and metal binding sites in the transmembrane domain of six-transmembrane epithelial antigen of the prostate (STEAP) family proteins. *J Biol Chem*, 290(37), 22558-69.
Kohlmann, A., Kipps, T. J., Rassenti, L. Z., Downing, J. R., Shurtleff, S. A., Mills, K. I., Gilkes, A. F., Hofmann, W. K., Basso, G., Dell'orto, M. C., Foà, R., Chiaretti, S., De Vos, J., Rauhut, S., Papenhausen, P. R., Hernández, J. M., Lumbreras, E., Yeoh, A. E., Koay, E. S., Li, R., Liu, W. M., Williams, P. M., Wieczorek, L. & Haferlach, T. (2008) An international standardization programme towards the application of gene expression profiling in routine leukaemia diagnostics: the Microarray Innovations in LEukemia study prephase. *Br J Haematol*, 142(5), 802-7.
Kontos, C. K., Christodoulou, M. I. & Scorilas, A. (2014) Apoptosis-related BCL2-family members: Key players in chemotherapy. *Anticancer Agents Med Chem*, 14(3), 353-74.
Kopp, K., Buntru, A., Pils, S., Zimmermann, T., Frank, R., Zumbusch, A. & Hauck, C. R. (2012) Grb14 is a negative regulator of CEACAM3-mediated phagocytosis of pathogenic bacteria. *J Biol Chem*, 287(46), 39158-70.
Krützfeldt, J., Rajewsky, N., Braich, R., Rajeev, K. G., Tuschl, T., Manoharan, M. & Stoffel, M. (2005) Silencing of microRNAs in vivo with 'antagomirs'. *Nature*, 438(7068), 685-689.
Lang, B. & Zhao, S. (2018) miR-486 functions as a tumor suppressor in esophageal cancer by targeting CDK4/BCAS2. *Oncol Rep*, 39(1), 71-80.
Lang, D., Powell, S. K., Plummer, R. S., Young, K. P. & Ruggeri, B. A. (2007) PAX genes: roles in development, pathophysiology, and cancer. *Biochem Pharmacol*, 73(1), 1-14.

Lee, D. A., Denman, C. J., Rondon, G., Woodworth, G., Chen, J., Fisher, T., Kaur, I., Fernandez-Vina, M., Cao, K., Ciurea, S., Shpall, E. J. & Champlin, R. E. (2016) Haploidentical Natural Killer Cells Infused before Allogeneic Stem Cell Transplantation for Myeloid Malignancies: A Phase I Trial. *Biol Blood Marrow Transplant*, 22(7), 1290-1298.
Lee, D. J. & Zeidner, J. F. (2019) Cyclin-dependent kinase (CDK) 9 and 4/6 inhibitors in acute myeloid leukemia (AML): a promising therapeutic approach. *Expert Opinion on Investigational Drugs*, 28(11), 989-1001.
Lee, D. Y., Deng, Z., Wang, C.-H. & Yang, B. B. (2007) MicroRNA-378 promotes cell survival, tumor growth, and angiogenesis by targeting SuFu and Fus-1 expression. *Proceedings of the National Academy of Sciences*, 104(51), 20350.
Leung, A. K. L. (2015) The Whereabouts of microRNA Actions: Cytoplasm and Beyond. *Trends in cell biology*, 25(10), 601-610.
Levis, M. (2017) Midostaurin approved for FLT3-mutated AML. *Blood*, 129(26), 3403-3406.
Ley, T. & Miller, C. (2013) Ding LCancer Genome Atlas Research Network, et al., Cancer Genome Atlas Research Network Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*, 368, 2059-74.
Ley, T. J., Miller, C., Ding, L., Raphael, B. J., Mungall, A. J., Robertson, A., Hoadley, K., Triche, T. J., Jr., Laird, P. W., Baty, J. D., Fulton, L. L., Fulton, R., Heath, S. E., Kalicki-Veizer, J., Kandoth, C., Klco, J. M., Koboldt, D. C., Kanchi, K. L., Kulkarni, S., Lamprecht, T. L., Larson, D. E., Lin, L., Lu, C., McLellan, M. D., McMichael, J. F., Payton, J., Schmidt, H., Spencer, D. H., Tomasson, M. H., Wallis, J. W., Wartman, L. D., Watson, M. A., Welch, J., Wendl, M. C., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y., Chiu, R., Chu, A., Chuah, E., Chun, H. J., Corbett, R., Dhalla, N., Guin, R., He, A., Hirst, C., Hirst, M., Holt, R. A., Jones, S., Karsan, A., Lee, D., Li, H. I., Marra, M. A., Mayo, M., Moore, R. A., Mungall, K., Parker, J., Pleasance, E., Plettner, P., Schein, J., Stoll, D., Swanson, L., Tam, A., Thiessen, N., Varhol, R., Wye, N., Zhao, Y., Gabriel, S., Getz, G., Sougnez, C., Zou, L., Leiserson, M. D., Vandin, F., Wu, H. T., Applebaum, F., Baylin, S. B., Akbani, R., Broom, B. M., Chen, K., Motter, T. C., Nguyen, K., Weinstein, J. N., Zhang, N., Ferguson, M. L., Adams, C., Black, A., Bowen, J., Gastier-Foster, J., Grossman, T., Lichtenberg, T., Wise, L., Davidsen, T., Demchok, J. A., Shaw, K. R., Sheth, M., Sofia, H. J., Yang, L., Downing, J. R. & Eley, G. (2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*, 368(22), 2059-74.
Li, F., Aljahdali, I. & Ling, X. (2019) Cancer therapeutics using survivin BIRC5 as a target: what can we do after over two decades of study? *Journal of Experimental & Clinical Cancer Research*, 38(1), 368.
Li, F., Ambrosini, G., Chu, E. Y., Plescia, J., Tognin, S., Marchisio, P. C. & Altieri, D. C. (1998) Control of apoptosis and mitotic spindle checkpoint by survivin. *Nature*, 396(6711), 580-4.
Li, H., Wang, Z., Zhang, W., Qian, K., Liao, G., Xu, W. & Zhang, S. (2015) VGLL4 inhibits EMT in part through suppressing Wnt/β-catenin signaling pathway in gastric cancer. *Med Oncol*, 32(3), 83.
Li, W., Xu, L., Che, X., Li, H., Zhang, Y., Song, N., Wen, T., Hou, K., Yang, Y., Zhou, L., Xin, X., Zeng, X., Shi, S., Liu, Y., Qu, X. & Teng, Y. (2018) C-Cbl reverses HER2-mediated tamoxifen resistance in human breast cancer cells. *BMC Cancer*, 18(1), 507.
Li, W., Zhu, Q., Zhang, S., Liu, L., Zhang, H. & Zhu, D. (2020) HOXC13-AS accelerates cell proliferation and migration in oral squamous cell carcinoma via miR-378g/HOXC13 axis. *Oral Oncol*, 111, 104946.
Li, Y., Chen, X. & Lu, H. (2016) Knockdown of SLC34A2 Inhibits Hepatocellular Carcinoma Cell Proliferation and Invasion. *Oncol Res*, 24(6), 511-519.
Liggins, A. P., Cooper, C. D., Lawrie, C. H., Brown, P. J., Collins, G. P., Hatton, C. S., Pulford, K. & Banham, A. H. (2007) MORC4, a novel member of the MORC family, is highly expressed in a subset of diffuse large B-cell lymphomas. *Br J Haematol*, 138(4), 479-86.
Lim, J., Goriely, A., Turner, G. D., Ewen, K. A., Jacobsen, G. K., Graem, N., Wilkie, A. O. & Rajpert-De Meyts, E. (2011a) OCT2, SSX and SAGE1 reveal the phenotypic heterogeneity of spermatocytic seminoma reflecting distinct subpopulations of spermatogonia. *J Pathol*, 224(4), 473-83.
Lim, J., Zhou, L., Ho, Y. K. & Wan, G. (2011b) SnoU6 and 5S RNAs are not reliable miRNA reference genes in neuronal differentiation. *Neuroscience*, 199, 32-43.
Lin, J. J. & Shaw, A. T. (2017) Recent Advances in Targeting ROS1 in Lung Cancer. *Journal of Thoracic Oncology*, 12(11), 1611-1625.
Liu, F., Cao, Y., Pinz, K., Ma, Y., Wada, M., Chen, K., Ma, G., Shen, J., Tse, C. O. & Su, Y. (2018a) First-in-human CLL1-CD33 compound CAR T cell therapy induces complete remission in patients with refractory acute myeloid leukemia: update on phase 1 clinical trial. *Blood*, 132, 901.
Liu, H., Lei, C., He, Q., Pan, Z., Xiao, D. & Tao, Y. (2018b) Nuclear functions of mammalian MicroRNAs in gene regulation, immunity and cancer. *Molecular Cancer*, 17(1), 64.

Liu, J., Xiao, Q., Xiao, J., Niu, C., Li, Y., Zhang, X., Zhou, Z., Shu, G. & Yin, G. (2022) Wnt/β-catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Signal Transduction and Targeted Therapy*, 7(1), 3.
Liu, W., Yang, Y. J. & An, Q. (2020) LINC00963 Promotes Ovarian Cancer Proliferation, Migration and EMT via the miR-378g /CHI3L1 Axis. *Cancer Manag Res*, 12, 463-473.
Liu, X., Zhou, X., Xu, H., He, Z., Shi, X. & Wu, S. (2017) SLC34A2 Regulates the Proliferation, Migration, and Invasion of Human Osteosarcoma Cells Through PTEN/PI3K/AKT Signaling. *DNA Cell Biol*, 36(9), 775-780.
Ljungman, P., Bregni, M., Brune, M., Cornelissen, J., Witte, T. d., Dini, G., Einsele, H., Gaspar, H. B., Gratwohl, A., Passweg, J., Peters, C., Rocha, V., Saccardi, R., Schouten, H., Sureda, A., Tichelli, A., Velardi, A., Niederwieser, D. & for the European Group for Blood and Marrow, T. (2010) Allogeneic and autologous transplantation for haematological diseases, solid tumours and immune disorders: current practice in Europe 2009. *Bone Marrow Transplantation*, 45(2), 219-234.
Love, M. I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.
Lulli, V., Romania, P., Morsilli, O., Cianciulli, P., Gabbianelli, M., Testa, U., Giuliani, A. & Marziali, G. (2013) MicroRNA-486-3p regulates γ-globin expression in human erythroid cells by directly modulating BCL11A. *PLoS One*, 8(4), e60436.
Luo, J., Zeng, S. & Tian, C. (2020) MORC4 Promotes Chemoresistance of Luminal A/B Breast Cancer via STAT3-Mediated MID2 Upregulation. *Onco Targets Ther*, 13, 6795-6803.
Luo, X., Zhang, J., Liu, X., Du, Q., Xu, N., Xu, L., Huang, B. & Xiao, X. (2012) [Identification of differentially expressed genes related to blastic crisis in chronic myeloid leukemia]. *Nan Fang Yi Ke Da Xue Xue Bao*, 32(6), 840-2.
Lv, L., Yu, J. & Qi, Z. (2020) Acute myeloid leukemia with inv(16)(p13.1q22) and deletion of the 5'MYH11/3'CBFB gene fusion: a report of two cases and literature review. *Molecular Cytogenetics*, 13(1), 4.
Lyu, H., Huang, J., He, Z. & Liu, B. (2018) Epigenetic mechanism of survivin dysregulation in human cancer. *Science China. Life sciences*, 61(7), 808-814.
Lübbert, M., Ihorst, G., Sander, P. N., Bogatyreva, L., Becker, H., Wijermans, P. W., Suciu, S., Bissé, E. & Claus, R. (2017) Elevated fetal haemoglobin is a predictor of better outcome in MDS/AML patients receiving 5-aza-2'-deoxycytidine (Decitabine). *Br J Haematol*, 176(4), 609-617.
Ma, Q. (2013) Role of nrf2 in oxidative stress and toxicity. *Annual review of pharmacology and toxicology*, 53, 401-426.
Macfarlane, L.-A. & Murphy, P. R. (2010) MicroRNA: Biogenesis, Function and Role in Cancer. *Current genomics*, 11(7), 537-561.
Maheswaran, E., Pedersen, C. B., Ditzel, H. J. & Gjerstorff, M. F. (2015) Lack of ADAM2, CALR3 and SAGE1 Cancer/Testis Antigen Expression in Lung and Breast Cancer. *PLoS One*, 10(8), e0134967.
Majzner, R. G. & Mackall, C. L. (2018) Tumor Antigen Escape from CAR T-cell Therapy. *Cancer Discov*, 8(10), 1219-1226.
Mandrekar, J. N. (2010) Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology*, 5(9), 1315-1316.
Mardiana, S. & Gill, S. (2020) CAR T Cells for Acute Myeloid Leukemia: State of the Art and Future Directions. *Frontiers in oncology*, 10, 697-697.
Marks, J. (2019) The role of SLC34A2 in intestinal phosphate absorption and phosphate homeostasis. *Pflugers Arch*, 471(1), 165-173.
Marofi, F., Rahman, H. S., Al-Obaidi, Z. M. J., Jalil, A. T., Abdelbasset, W. K., Suksatan, W., Dorofeev, A. E., Shomali, N., Chartrand, M. S., Pathak, Y., Hassanzadeh, A., Baradaran, B., Ahmadi, M., Saeedi, H., Tahmasebi, S. & Jarahian, M. (2021) Novel CAR T therapy is a ray of hope in the treatment of seriously ill AML patients. *Stem Cell Research & Therapy*, 12(1), 465.
Matias-Garcia, P. R., Wilson, R., Mussack, V., Reischl, E., Waldenberger, M., Gieger, C., Anton, G., Peters, A. & Kuehn-Steven, A. (2020) Impact of long-term storage and freeze-thawing on eight circulating microRNAs in plasma samples. *PLOS ONE*, 15(1), e0227648.
Maude, S. L., Barrett, D., Teachey, D. T. & Grupp, S. A. (2014) Managing cytokine release syndrome associated with novel T cell-engaging therapies. *Cancer J*, 20(2), 119-22.
McMurry, H., Fletcher, L. & Traer, E. (2021) IDH Inhibitors in AML—Promise and Pitfalls. *Current Hematologic Malignancy Reports*, 16(2), 207-217.
Mercher, T. & Schwaller, J. (2019) Pediatric Acute Myeloid Leukemia (AML): From Genes to Models Toward Targeted Therapeutic Intervention. *Frontiers in pediatrics*, 7, 401-401.

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.-P., Mushayamaha, T. & Thomas, P. D. (2020) PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research*, 49(D1), D394-D403.
Michaud-Levesque, J., Demeule, M. & Béliveau, R. (2005) Stimulation of cell surface plasminogen activation by membrane-bound melanotransferrin: a key phenomenon for cell invasion. *Exp Cell Res*, 308(2), 479-90.
Mirza, A., McGuirk, M., Hockenberry, T. N., Wu, Q., Ashar, H., Black, S., Wen, S. F., Wang, L., Kirschmeier, P., Bishop, W. R., Nielsen, L. L., Pickett, C. B. & Liu, S. (2002) Human survivin is negatively regulated by wild-type p53 and participates in p53-dependent apoptotic pathway. *Oncogene*, 21(17), 2613-2622.
Moreaux, J., Kassambara, A., Hose, D. & Klein, B. (2012) STEAP1 is overexpressed in cancers: A promising therapeutic target. *Biochemical and Biophysical Research Communications*, 429(3), 148-155.
Moretti, F., Kaiser, C., Zdanowicz-Specht, A. & Hentze, M. W. (2012) PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding. *Nature Structural & Molecular Biology*, 19(6), 603-608.
Muntean, A. G. & Hess, J. L. (2012) The pathogenesis of mixed-lineage leukemia. *Annual review of pathology*, 7, 283-301.
Nagelkerke, N. J. D., Oosting, J. & Hart, A. A. M. (1984) A Simple Test for Goodness of Fit of Cox's Proportional Hazards Model. *Biometrics*, 40(2), 483-486.
Nakamura, H., Takada, K., Arihara, Y., Hayasaka, N., Murase, K., Iyama, S., Kobune, M., Miyanishi, K. & Kato, J. (2019) Six-transmembrane epithelial antigen of the prostate 1 protects against increased oxidative stress via a nuclear erythroid 2-related factor pathway in colorectal cancer. *Cancer Gene Ther*, 26(9-10), 313-322.
Nakazawa, K., Dashzeveg, N. & Yoshida, K. (2014) Tumor suppressor p53 induces miR-1915 processing to inhibit Bcl-2 in the apoptotic response to DNA damage. *The FEBS Journal*, 281(13), 2937-2944.
NHS (2019) Overview of AML.
Ninawe, A., Guru, S. A., Yadav, P., Masroor, M., Samadhiya, A., Bhutani, N., Gupta, N., Gupta, R. & Saxena, A. (2021) miR-486-5p: A Prognostic Biomarker for Chronic Myeloid Leukemia. *ACS Omega*, 6(11), 7711-7718.
Noren Hooten, N., Fitzpatrick, M., Wood, W. H., 3rd, De, S., Ejiogu, N., Zhang, Y., Mattison, J. A., Becker, K. G., Zonderman, A. B. & Evans, M. K. (2013) Age-related changes in microRNA levels in serum. *Aging*, 5(10), 725-740.
O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. (2018) Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology*, 9(402).
Ohgami, R. S., Campagna, D. R., McDonald, A. & Fleming, M. D. (2006) The Steap proteins are metalloreductases. *Blood*, 108(4), 1388-1394.
Ohrt, T., Mütze, J., Staroske, W., Weinmann, L., Höck, J., Crell, K., Meister, G. & Schwille, P. (2008) Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Res*, 36(20), 6439-49.
Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M. & Lai, E. C. (2007) The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell*, 130(1), 89-100.
Padilha, S. L., Souza, E. J. d. S., Matos, M. C. C. & Domino, N. R. (2015) Acute myeloid leukemia: survival analysis of patients at a university hospital of Paraná. *Revista Brasileira de Hematologia e Hemoterapia*, 37(1), 21-27.
Pastor, T. P., Peixoto, B. C. & Viola, J. P. B. (2021) The Transcriptional Co-factor IRF2BP2: A New Player in Tumor Development and Microenvironment. *Frontiers in Cell and Developmental Biology*, 9.
Peng, Y. & Croce, C. M. (2016) The role of MicroRNAs in human cancer. *Signal Transduction and Targeted Therapy*, 1(1), 15004.
Pils, S., Gerrard, D. T., Meyer, A. & Hauck, C. R. (2008) CEACAM3: an innate immune receptor directed against human-restricted bacterial pathogens. *Int J Med Microbiol*, 298(7-8), 553-60.
Puumala, S. E., Ross, J. A., Aplenc, R. & Spector, L. G. (2013) Epidemiology of childhood acute myeloid leukemia. *Pediatric blood & cancer*, 60(5), 728-733.
Qazilbash, M. H., Wieder, E., Thall, P. F., Wang, X., Rios, R., Lu, S., Kanodia, S., Ruisaard, K. E., Giralt, S. A., Estey, E. H., Cortes, J., Komanduri, K. V., Clise-Dwyer, K., Alatrash, G., Ma, Q., Champlin, R. E. & Molldrem, J. J. (2017) PR1 peptide vaccine induces specific immunity with clinical responses in myeloid malignancies. *Leukemia*, 31(3), 697-704.
Qu, M., Zou, X., Fang, F., Wang, S., Xu, L., Zeng, Q., Fan, Z., Chen, L., Yue, W., Xie, X. & Pei, X. (2020) Platelet-derived microparticles enhance megakaryocyte differentiation and platelet generation via miR-1915-3p. *Nat Commun*, 11(1), 4964.

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., Merico, D. & Bader, G. D. (2019) Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature Protocols*, 14(2), 482-517.
Ren, Z., He, H., Zuo, Z., Xu, Z., Wei, Z. & Deng, J. (2019) The role of different SIRT1-mediated signaling pathways in toxic injury. *Cellular & molecular biology letters*, 24, 36-36.
Ruby, J. G., Jan, C. H. & Bartel, D. P. (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature*, 448(7149), 83-6.
Sah, N. K., Khan, Z., Khan, G. J. & Bisen, P. S. (2006) Structural, functional and therapeutic biology of survivin. *Cancer Lett*, 244(2), 164-71.
Sallustio, F., Serino, G., Costantino, V., Curci, C., Cox, S. N., De Palma, G. & Schena, F. P. (2013) miR-1915 and miR-1225-5p regulate the expression of CD133, PAX2 and TLR2 in adult renal progenitor cells. *PLoS One*, 8(7), e68296.
Sarantis, H. & Gray-Owen, S. D. (2007) The specific innate immune receptor CEACAM3 triggers neutrophil bactericidal activities via a Syk kinase-dependent pathway. *Cell Microbiol*, 9(9), 2167-80.
Sawaki, K., Kanda, M., Umeda, S., Miwa, T., Tanaka, C., Kobayashi, D., Hayashi, M., Yamada, S., Nakayama, G., Omae, K., Koike, M. & Kodera, Y. (2019) Level of Melanotransferrin in Tissue and Sera Serves as a Prognostic Marker of Gastric Cancer. *Anticancer Res*, 39(11), 6125-6133.
Scherrer, A., Wohlwend, A., Kruithof, E. K., Vassalli, J. D. & Sappino, A. P. (1999) Plasminogen activation in human acute leukaemias. *Br J Haematol*, 105(4), 920-7.
Schultz, D. J., Muluhngwi, P., Alizadeh-Rad, N., Green, M. A., Rouchka, E. C., Waigel, S. J. & Klinge, C. M. (2017) Genome-wide miRNA response to anacardic acid in breast cancer cells. *PLoS One*, 12(9), e0184471.
Seyyedi, S. S., Soleimani, M., Yaghmaie, M., Ajami, M., Ajami, M., Pourbeyranvand, S., Alimoghaddam, K. & Akrami, S. M. (2016a) Deregulation of miR-1, miR486, and let-7a in cytogenetically normal acute myeloid leukemia: association with NPM1 and FLT3 mutation and clinical characteristics. *Tumor Biology*, 37(4), 4841-4847.
Seyyedi, S. S., Soleimani, M., Yaghmaie, M., Ajami, M., Pourbeyranvand, S., Alimoghaddam, K. & Akrami, S. M. (2016b) Deregulation of miR-1, miR486, and let-7a in cytogenetically normal acute myeloid leukemia: association with NPM1 and FLT3 mutation and clinical characteristics. *Tumour Biol*, 37(4), 4841-7.
Sha, C., Jia, G., Jingjing, Z., Yapeng, H., Zhi, L. & Guanghui, X. (2021) miR-486 is involved in the pathogenesis of acute myeloid leukemia by regulating JAK-STAT signaling. *Naunyn Schmiedebergs Arch Pharmacol*, 394(1), 177-187.
Shahrabi, S., Khodadi, E., Saba, F., Shahjahani, M. & Saki, N. (2018) Sex chromosome changes in leukemia: cytogenetics and molecular aspects. *Hematology*, 23(3), 139-147.
Shallis, R. M., Wang, R., Davidoff, A., Ma, X. & Zeidan, A. M. (2019) Epidemiology of acute myeloid leukemia: Recent progress and enduring challenges. *Blood Rev*, 36, 70-87.
Shen, J., Nie, X., Huang, S. Y., Qin, Y. Q., Pan, L. L. & Wang, X. T. (2019) Neuromuscular electrical stimulation improves muscle atrophy induced by chronic hypoxia-hypercapnia through the MicroRNA-486/PTEN/FoxO1 pathway. *Biochem Biophys Res Commun*, 509(4), 1021-1027.
Shen, S., Guo, X., Yan, H., Lu, Y., Ji, X., Li, L., Liang, T., Zhou, D., Feng, X. H., Zhao, J. C., Yu, J., Gong, X. G., Zhang, L. & Zhao, B. (2015) A miR-130a-YAP positive feedback loop promotes organ size and tumorigenesis. *Cell Res*, 25(9), 997-1012.
Shi, X. F., Wang, H., Kong, F. X., Xu, Q. Q., Xiao, F. J., Yang, Y. F., Ge, R. L. & Wang, L. S. (2017) Exosomal miR-486 regulates hypoxia-induced erythroid differentiation of erythroleukemia cells through targeting Sirt1. *Exp Cell Res*, 351(1), 74-81.
Shin, S., Sung, B.-J., Cho, Y.-S., Kim, H.-J., Ha, N.-C., Hwang, J.-I., Chung, C.-W., Jung, Y.-K. & Oh, B.-H. (2001) An Anti-apoptotic Protein Human Survivin Is a Direct Inhibitor of Caspase-3 and -7. *Biochemistry*, 40(4), 1117-1123.
Small, E. M., O'Rourke, J. R., Moresi, V., Sutherland, L. B., McAnally, J., Gerard, R. D., Richardson, J. A. & Olson, E. N. (2010) Regulation of PI3-kinase/Akt signaling by muscle-enriched microRNA-486. *Proc Natl Acad Sci U S A*, 107(9), 4218-23.
Sohn, W., Kim, J., Kang, S. H., Yang, S. R., Cho, J.-Y., Cho, H. C., Shim, S. G. & Paik, Y.-H. (2015) Serum exosomal microRNAs as novel biomarkers for hepatocellular carcinoma. *Experimental & Molecular Medicine*, 47(9), e184-e184.
Song, Z., Yao, X. & Wu, M. (2003) Direct interaction between survivin and Smac/DIABLO is essential for the anti-apoptotic activity of survivin during taxol-induced apoptosis. *J Biol Chem*, 278(25), 23130-40.
Sparvero, L. J., Asafu-Adjei, D., Kang, R., Tang, D., Amin, N., Im, J., Rutledge, R., Lin, B., Amoscato, A. A., Zeh, H. J. & Lotze, M. T. (2009) RAGE (Receptor for Advanced Glycation Endproducts), RAGE Ligands, and their role in Cancer and Inflammation. *Journal of Translational Medicine*, 7(1), 17.

Sredni, S., Gadd, S., Jafari, N. & Huang, C.-C. (2011) A Parallel Study of mRNA and microRNA Profiling of Peripheral Blood in Young Adult Women. *Frontiers in Genetics*, 2.
Stambolic, V., Suzuki, A., de la Pompa, J. L., Brothers, G. M., Mirtsos, C., Sasaki, T., Ruland, J., Penninger, J. M., Siderovski, D. P. & Mak, T. W. (1998) Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell*, 95(1), 29-39.
Starczynowski, D. T., Kuchenbauer, F., Argiropoulos, B., Sung, S., Morin, R., Muranyi, A., Hirst, M., Hogge, D., Marra, M., Wells, R. A., Buckstein, R., Lam, W., Humphries, R. K. & Karsan, A. (2010) Identification of miR-145 and miR-146a as mediators of the 5q– syndrome phenotype. *Nature Medicine*, 16(1), 49-58.
Sterner, R. C. & Sterner, R. M. (2021) CAR-T cell therapy: current limitations and potential strategies. *Blood cancer journal*, 11(4), 69-69.
Stewart, D. J. (2014) Wnt signaling pathway in non-small cell lung cancer. *J Natl Cancer Inst*, 106(1), djt356.
Storey, S., Gray, T. F. & Bryant, A. L. (2017) Comorbidity, Physical Function, and Quality of Life in Older Adults with Acute Myeloid Leukemia. *Current geriatrics reports*, 6(4), 247-254.
Su, M., Alonso, S., Jones, J. W., Yu, J., Kane, M. A., Jones, R. J. & Ghiaur, G. (2015) All-Trans Retinoic Acid Activity in Acute Myeloid Leukemia: Role of Cytochrome P450 Enzyme Expression by the Microenvironment. *PloS one*, 10(6), e0127790-e0127790.
Sun, C., Mezzadra, R. & Schumacher, T. N. (2018) Regulation and Function of the PD-L1 Checkpoint. *Immunity*, 48(3), 434-452.
Suryo Rahmanto, Y., Dunn, L. L. & Richardson, D. R. (2007) The melanoma tumor antigen, melanotransferrin (p97): a 25-year hallmark – from iron metabolism to tumorigenesis. *Oncogene*, 26(42), 6113-6124.
Suzuki, A., Hayashida, M., Ito, T., Kawano, H., Nakano, T., Miura, M., Akahane, K. & Shiraki, K. (2000) Survivin initiates cell cycle entry by the competitive interaction with Cdk4/p16(INK4a) and Cdk2/cyclin E complex activation. *Oncogene*, 19(29), 3225-34.
Syn, N. L., Teng, M. W. L., Mok, T. S. K. & Soo, R. A. (2017) De-novo and acquired resistance to immune checkpoint targeting. *Lancet Oncol*, 18(12), e731-e741.
Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J. & Mering, C. V. (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, 47(D1), D607-d613.
Takahashi, S. (2011) Current findings for recurring mutations in acute myeloid leukemia. *Journal of hematology & oncology*, 4, 36-36.
Tamm, I., Richter, S., Oltersdorf, D., Creutzig, U., Harbott, J., Scholz, F., Karawajew, L., Ludwig, W.-D. & Wuchter, C. (2004) High Expression Levels of X-Linked Inhibitor of Apoptosis Protein and Survivin Correlate with Poor Overall Survival in Childhood de Novo Acute Myeloid Leukemia. *Clinical Cancer Research*, 10(11), 3737.
Tamm, I., Wang, Y., Sausville, E., Scudiero, D. A., Vigna, N., Oltersdorf, T. & Reed, J. C. (1998) IAP-Family Protein Survivin Inhibits Caspase Activity and Apoptosis Induced by Fas (CD95), Bax, Caspases, and Anticancer Drugs. *Cancer Research*, 58(23), 5315-5320.
Tang, Q., Jiang, X., Ma, S., Wang, L., Li, R. & Ma, J. (2020) MIR22HG regulates miR-486/PTEN axis in bladder cancer to promote cell proliferation. *Biosci Rep*, 40(6).
Team, R. C. (2013) R: A language and environment for statistical computing.
Terao, T. & Minami, Y. (2019) Targeting Hedgehog (Hh) Pathway for the Acute Myeloid Leukemia Treatment. *Cells*, 8(4), 312.
Terwilliger, T. & Abdul-Hay, M. (2017) Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer J*, 7(6), e577.
Therneau, T. M. (2020) *A Package for Survival Analysis in R.* Available online: https://CRAN.R-project.org/package=survival [Accessed.
Tian, B., Nowak, D. E. & Brasier, A. R. (2005) A TNF-induced gene expression program under oscillatory NF-kappaB control. *BMC genomics*, 6, 137-137.
Tiklová, K., Senti, K. A., Wang, S., Gräslund, A. & Samakovlis, C. (2010) Epithelial septate junction assembly relies on melanotransferrin iron binding and endocytosis in Drosophila. *Nat Cell Biol*, 12(11), 1071-7.
Truesdell, S. S., Mortensen, R. D., Seo, M., Schroeder, J. C., Lee, J. H., LeTonqueze, O. & Vasudevan, S. (2012) MicroRNA-mediated mRNA Translation Activation in Quiescent Cells and Oocytes Involves Recruitment of a Nuclear microRNP. *Scientific Reports*, 2(1), 842.
UK, C. R. (2018) Cancer Research UK. Available online: https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence#heading-Zero [Accessed 28/01/2021].
UK, C. R. (2019) AML Incidence by age.

Vader, G., Medema, R. H. & Lens, S. M. (2006) The chromosomal passenger complex: guiding Aurora-B through mitosis. *J Cell Biol*, 173(6), 833-7.
Van Tendeloo Viggo, F., Van de Velde, A., Van Driessche, A., Cools, N., Anguille, S., Ladell, K., Gostick, E., Vermeulen, K., Pieters, K., Nijs, G., Stein, B., Smits Evelien, L., Schroyens Wilfried, A., Gadisseur Alain, P., Vrelust, I., Jorens Philippe, G., Goossens, H., de Vries, I. J., Price David, A., Oji, Y., Oka, Y., Sugiyama, H. & Berneman Zwi, N. (2010) Induction of complete and molecular remissions in acute myeloid leukemia by Wilms' tumor 1 antigen-targeted dendritic cell vaccination. *Proceedings of the National Academy of Sciences*, 107(31), 13824-13829.
Vasu, S., Kohlschmidt, J., Mrózek, K., Eisfeld, A. K., Nicolet, D., Sterling, L. J., Becker, H., Metzeler, K. H., Papaioannou, D., Powell, B. L., Kolitz, J. E., Moore, J. O., Baer, M. R., Roboz, G. J., Stone, R. M., Byrd, J. C., Carroll, A. J. & Bloomfield, C. D. (2018) Ten-year outcome of patients with acute myeloid leukemia not treated with allogeneic transplantation in first complete remission. *Blood Adv*, 2(13), 1645-1650.
Vasudevan, S. & Steitz, J. A. (2007) AU-Rich-Element-Mediated Upregulation of Translation by FXR1 and Argonaute 2. *Cell*, 128(6), 1105-1118.
Wagner, M., Schmelz, K., Wuchter, C., Ludwig, W.-D., Dörken, B. & Tamm, I. (2006) In vivo expression of survivin and its splice variant survivin-2B: Impact on clinical outcome in acute myeloid leukemia. *International Journal of Cancer*, 119(6), 1291-1297.
Wallace, J. A. & O'Connell, R. M. (2017) MicroRNAs and acute myeloid leukemia: therapeutic implications and emerging concepts. *Blood*, 130(11), 1290-1301.
Wan, Y., Cui, R., Gu, J., Zhang, X., Xiang, X., Liu, C., Qu, K. & Lin, T. (2017) Identification of Four Oxidative Stress-Responsive MicroRNAs, miR-34a-5p, miR-1915-3p, miR-638, and miR-150-3p, in Hepatocellular Carcinoma. *Oxid Med Cell Longev*, 2017, 5189138.
Wang, L.-S., Li, L., Li, L., Chu, S., Shiang, K.-D., Li, M., Sun, H.-Y., Xu, J., Xiao, F.-J., Sun, G., Rossi, J. J., Ho, Y. & Bhatia, R. (2015a) MicroRNA-486 regulates normal erythropoiesis and enhances growth and modulates drug response in CML progenitors. *Blood*, 125(8), 1302-1313.
Wang, Y., Yang, W., Pu, Q., Yang, Y., Ye, S., Ma, Q., Ren, J., Cao, Z., Zhong, G., Zhang, X., Liu, L. & Zhu, W. (2015b) The effects and mechanisms of SLC34A2 in tumorigenesis and progression of human non-small cell lung cancer. *J Biomed Sci*, 22(1), 52.
Wang, Z., Fukuda, S. & Pelus, L. M. (2004) Survivin regulates the p53 tumor suppressor gene family. *Oncogene*, 23(49), 8146-8153.
Wang, Z., Ma, B., Ji, X., Deng, Y., Zhang, T., Zhang, X., Gao, H., Sun, H., Wu, H., Chen, X. & Zhao, R. (2015c) MicroRNA-378-5p suppresses cell proliferation and induces apoptosis in colorectal cancer cells by targeting BRAF. *Cancer Cell Int*, 15, 40.
Weber, S., Parmon, A., Kurrle, N., Schnütgen, F. & Serve, H. (2021) The Clinical Significance of Iron Overload and Iron Metabolism in Myelodysplastic Syndrome and Acute Myeloid Leukemia. *Frontiers in Immunology*, 11.
Wei, Y., Cao, Y., Sun, R., Cheng, L., Xiong, X., Jin, X., He, X., Lu, W. & Zhao, M. (2020) Targeting Bcl-2 Proteins in Acute Myeloid Leukemia. *Frontiers in Oncology*, 10.
Weiss, F. U., Hesselbarth, N., Párniczky, A., Mosztbacher, D., Lämmerhirt, F., Ruffert, C., Kovacs, P., Beer, S., Seltsam, K., Griesmann, H., Böhme, R., Kaune, T., Hollenbach, M., Schulz, H. U., Simon, P., Mayerle, J., Lerch, M. M., Cavestro, G. M., Zuppardo, R. A., Di Leo, M., Testoni, P. A., Malecka-Panas, E., Gasirowska, A., Głuszek, S., Bugert, P., Szentesi, A., Mössner, J., Witt, H., Michl, P., Hégyi, P., Scholz, M. & Rosendahl, J. (2018) Common variants in the CLDN2-MORC4 and PRSS1-PRSS2 loci confer susceptibility to acute pancreatitis. *Pancreatology*, 18(5), 477-481.
Weon, J. L., Yang, S. W. & Potts, P. R. (2018) Cytosolic Iron-Sulfur Assembly Is Evolutionarily Tuned by a Cancer-Amplified Ubiquitin Ligase. *Mol Cell*, 69(1), 113-125.e6.
Wheeler, F. C., Kim, A. S., Mosse, C. A., Shaver, A. C., Yenamandra, A. & Seegmiller, A. C. (2018) Limited Utility of Fluorescence In Situ Hybridization for Recurrent Abnormalities in Acute Myeloid Leukemia at Diagnosis and Follow-up. *American journal of clinical pathology*, 149(5), 418-424.
Winters, A. C. & Bernt, K. M. (2017) MLL-Rearranged Leukemias—An Update on Science and Clinical Approaches. *Frontiers in Pediatrics*, 5.
Wolach, O. & Stone, R. M. (2015) How I treat mixed-phenotype acute leukemia. *Blood*, 125(16), 2477-2485.
Wolska-Washer, A. & Robak, T. (2019) Glasdegib in the treatment of acute myeloid leukemia. *Future Oncol*, 15(28), 3219-3232.
Wu, A., Wu, Q., Deng, Y., Liu, Y., Lu, J., Liu, L., Li, X., Liao, C., Zhao, B. & Song, H. (2019) Loss of VGLL4 suppresses tumor PD-L1 expression and immune evasion. *The EMBO Journal*, 38(1), e99506.
Xu, C., Li, H., Zhang, L., Jia, T., Duan, L. & Lu, C. (2016) MicroRNA-1915-3p prevents the apoptosis of lung cancer cells by downregulating DRG2 and PBX2. *Mol Med Rep*, 13(1), 505-512.

Xu, J. & Niu, T. (2020) Natural killer cell-based immunotherapy for acute myeloid leukemia. *Journal of Hematology & Oncology*, 13(1), 167.
Xu, K., Liang, X., Cui, D., Wu, Y., Shi, W. & Liu, J. (2013) miR-1915 inhibits Bcl-2 to modulate multidrug resistance by increasing drug-sensitivity in human colorectal carcinoma cells. *Mol Carcinog*, 52(1), 70-8.
Xu, X. C., Zhang, W. B., Li, C. X., Gao, H., Pei, Q., Cao, B. W. & He, T. H. (2019) Up-Regulation of MiR-1915 Inhibits Proliferation, Invasion, and Migration of Helicobacter pylori-Infected Gastric Cancer Cells via Targeting RAGE. *Yonsei Med J*, 60(1), 38-47.
Yang, Y., Ji, C., Guo, S., Su, X., Zhao, X., Zhang, S., Liu, G., Qiu, X., Zhang, Q., Guo, H. & Chen, H. (2017) The miR-486-5p plays a causative role in prostate cancer through negative regulation of multiple tumor suppressor pathways. *Oncotarget*, 8(42), 72835-72846.
Yang, Z., Zhuang, Q., Hu, G. & Geng, S. (2019) MORC4 is a novel breast cancer oncogene regulated by miR-193b-3p. *J Cell Biochem*, 120(3), 4634-4643.
Ye, W., Chen, C., Gao, Y., Zheng, Z. S., Xu, Y., Yun, M., Weng, H. W., Xie, D., Ye, S. & Zhang, J. X. (2017) Overexpression of SLC34A2 is an independent prognostic indicator in bladder cancer and its depletion suppresses tumor growth via decreasing c-Myc expression and transcriptional activity. *Cell Death Dis*, 8(2), e2581.
Yu, B. & Liu, D. (2019) Gemtuzumab ozogamicin and novel antibody-drug conjugates in clinical trials for acute myeloid leukemia. *Biomarker Research*, 7(1), 24.
Zeng, M., Zhu, L., Li, L. & Kang, C. (2017a) miR-378 suppresses the proliferation, migration and invasion of colon cancer cells by inhibiting SDAD1. *Cell Mol Biol Lett*, 22, 12.
Zeng, Y., Stauffer, S., Zhou, J., Chen, X., Chen, Y. & Dong, J. (2017b) Cyclin-dependent kinase 1 (CDK1)-mediated mitotic phosphorylation of the transcriptional co-repressor Vgll4 inhibits its tumor-suppressing activity. *J Biol Chem*, 292(36), 15028-15038.
Zhang, E., Shen, B., Mu, X., Qin, Y., Zhang, F., Liu, Y., Xiao, J., Zhang, P., Wang, C., Tan, M. & Fan, Y. (2016) Ubiquitin-specific protease 11 (USP11) functions as a tumor suppressor through deubiquitinating and stabilizing VGLL4 protein. *Am J Cancer Res*, 6(12), 2901-2909.
Zhang, S., Cao, J., Kong, Y. M. & Scheuermann, R. H. (2010) GO-Bayes: Gene Ontology-based overrepresentation analysis using a Bayesian approach. *Bioinformatics*, 26(7), 905-911.
Zhang, T. J., Zhou, J. D., Ma, J. C., Deng, Z. Q., Qian, Z., Yao, D. M., Yang, J., Li, X. X., Lin, J. & Qian, J. (2017) CDH1 (E-cadherin) expression independently affects clinical outcome in acute myeloid leukemia with normal cytogenetics. *Clin Chem Lab Med*, 55(1), 123-131.
Zhang, Y., Chen, H. X., Zhou, S. Y., Wang, S. X., Zheng, K., Xu, D. D., Liu, Y. T., Wang, X. Y., Wang, X., Yan, H. Z., Zhang, L., Liu, Q. Y., Chen, W. Q. & Wang, Y. F. (2015) Sp1 and c-Myc modulate drug resistance of leukemia stem cells by regulating survivin expression through the ERK-MSK MAPK signaling pathway. *Mol Cancer*, 14, 56.
Zhang, Y., Yu, X., Liu, Q., Gong, H., Chen, A. A., Zheng, H., Zhong, S. & Li, Y. (2021) SAGE1: a Potential Target Antigen for Lung Cancer T-Cell Immunotherapy. *Mol Cancer Ther*, 20(11), 2302-2313.
Zhou, J., Bi, C., Janakakumara, J. V., Liu, S.-C., Chng, W.-J., Tay, K.-G., Poon, L.-F., Xie, Z., Palaniyandi, S., Yu, H., Glaser, K. B., Albert, D. H., Davidsen, S. K. & Chen, C.-S. (2009) Enhanced activation of STAT pathways and overexpression of survivin confer resistance to FLT3 inhibitors and could be therapeutic targets in AML. *Blood*, 113(17), 4052-4062.
Zhou, J., Ching, Y. Q. & Chng, W.-J. (2015) Aberrant nuclear factor-kappa B activity in acute myeloid leukemia: from molecular pathogenesis to therapeutic target. *Oncotarget*, 6(8), 5490-5500.
Zhuang, X., Herbert, J. M., Lodhia, P., Bradford, J., Turner, A. M., Newby, P. M., Thickett, D., Naidu, U., Blakey, D., Barry, S., Cross, D. A. & Bicknell, R. (2015) Identification of novel vascular targets in lung cancer. *Br J Cancer*, 112(3), 485-94.
Zwaan, C. M., Kolb, E. A., Reinhardt, D., Abrahamsson, J., Adachi, S., Aplenc, R., De Bont, E. S. J. M., De Moerloose, B., Dworzak, M., Gibson, B. E. S., Hasle, H., Leverger, G., Locatelli, F., Ragu, C., Ribeiro, R. C., Rizzari, C., Rubnitz, J. E., Smith, O. P., Sung, L., Tomizawa, D., van den Heuvel-Eibrink, M. M., Creutzig, U. & Kaspers, G. J. L. (2015) Collaborative Efforts Driving Progress in Pediatric Acute Myeloid Leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 33(27), 2949-2962.

Chapter 8. Appendices

Appendix I

International Journal of *Molecular Sciences*

MDPI

Article

Survivin' Acute Myeloid Leukaemia—A Personalised Target for inv(16) Patients

Jochen Greiner [1,2], Elliott Brown [3], Lars Bullinger [4,5], Robert K. Hills [6], Vanessa Morris [3], Hartmut Döhner [2], Ken I. Mills [7] and Barbara-ann Guinn [3,*]

1 Department of Internal Medicine, Diakonie Hospital Stuttgart, 70176 Stuttgart, Germany; greiner@diak-stuttgart.de
2 Department of Internal Medicine III, University of Ulm, Helmholtzstr. 10, 89081 Ulm, Germany; hartmut.doehner@uniklinik-ulm.de
3 Department of Biomedical Sciences, University of Hull, Hull HU6 7RX, UK; Elliott.Brown-2016@hull.ac.uk (E.B.); V.S.Morris-2017@hull.ac.uk (V.M.)
4 Department of Hematology, Oncology and Tumor Immunology, Charité–Universitätsmedizin Berlin, 13353 Berlin, Germany; lars.bullinger@charite.de
5 German Cancer Consortium (DKTK), Partner site Berlin, 13353 Berlin, Germany
6 Nuffield Department of Population Health, Richard Doll Building, University of Oxford, Oxford OX3 7LF, UK; robert.hills@ndph.ox.ac.uk
7 Patrick G. Johnson Centre for Cancer Research, Queen's University Belfast, Lisburn Road, Belfast BT9 7AE, UK; K.Mills@qub.ac.uk
* Correspondence: B.Guinn@hull.ac.uk; Tel.: +44-1482-466543

Abstract: Despite recent advances in therapies including immunotherapy, patients with acute myeloid leukaemia (AML) still experience relatively poor survival rates. The Inhibition of Apoptosis (IAP) family member, survivin, also known by its gene and protein name, Baculoviral IAP Repeat Containing 5 (BIRC5), remains one of the most frequently expressed antigens across AML subtypes. To better understand its potential to act as a target for immunotherapy and a biomarker for AML survival, we examined the protein and pathways that BIRC5 interacts with using the Kyoto Encyclopedia of Genes and Genomes (KEGG), search tool for recurring instances of neighbouring genes (STRING), WEB-based Gene Set Analysis Toolkit, Bloodspot and performed a comprehensive literature review. We then analysed data from gene expression studies. These included 312 AML samples in the Microarray Innovations In Leukemia (MILE) dataset. We found a trend between above median levels of BIRC5 being associated with improved overall survival (OS) but this did not reach statistical significance ($p = 0.077$, Log-Rank). There was some evidence of a beneficial effect in adjusted analyses where above median levels of BIRC5 were shown to be associated with improved OS ($p = 0.001$) including in Core Binding Factor (CBF) patients ($p = 0.03$). Above median levels of BIRC5 transcript were associated with improved relapse free survival ($p < 0.0001$). Utilisation of a second large cDNA microarray dataset including 306 AML cases, again showed no correlation between BIRC5 levels and OS, but high expression levels of BIRC5 correlated with worse survival in inv(16) patients ($p = 0.077$) which was highly significant when datasets A and B were combined ($p = 0.001$). In addition, decreased BIRC5 expression was associated with better clinical outcome ($p = 0.004$) in AML patients exhibiting CBF mainly due to patients with inv(16) ($p = 0.007$). This study has shown that BIRC5 expression plays a role in the survival of AML patients, this association is not apparent when we examine CBF patients as a cohort, but when those with inv(16) independently indicating that those patients with inv(16) would provide interesting candidates for immunotherapies that target BIRC5.

Keywords: BIRC5; overall survival; survivin; acute myeloid leukaemia; Core Binding Factor (CBF); inv(16)

check for updates

Citation: Greiner, J.; Brown, E.; Bullinger, L.; Hills, R.K.; Morris, V.; Döhner, H.; Mills, K.I.; Guinn, B.-a. Survivin' Acute Myeloid Leukaemia—A Personalised Target for inv(16) Patients. *Int. J. Mol. Sci.* **2021**, *22*, 10482. https://doi.org/10.3390/ijms221910482

Academic Editors: Alessia Lignetti and Barbara Guinn

Received: 13 September 2021
Accepted: 22 September 2021
Published: 28 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.