Computational identification of key genes and pathways associated with Acute Myeloid Leukemia

By

MERCY J.

(Roll. No. 20PZO011)

Supervisor

Dr. K.S. SANTHY

Thesis Submitted to the

Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore – 641 043

In Partial Fulfilment of the Requirements for the Degree of Master of Science (M.Sc.)

MAY 2022

Computational identification of key genes and pathways associated with Acute Myeloid Leukemia

By

MERCY J.

(Roll. No. 20PZO011)

Department of Zoology

Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore- 641043

In Partial Fulfilment of the Requirements for the Degree of

Master of Science (M.Sc.)

MAY 2022

30/5/2022

Signature of the Head of the Department

30.5.2022

Signature of the Supervisor

ACKNOWLEDGEMENT

First and foremost, I thank **Almighty** for his grace, which sustained and helped me to complete my thesis work successfully.

I take this opportunity to pay my respectful gratitudeto (Late) **Dr. P.R. Krishnakumar,** former Chancellor, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for all the amenities provided for the conduct of my researchwork.

I wish to express my profound sense of gratitude and sincere thanks to **Dr.S. P. Thyagarajan, D.Sc., Ph.D., M.D., Chancellor**, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for all the amenities provided for the conduct of my research work.

I am extremely grateful to **Dr. (Mrs.) V. Bharathi Harishankar, Ph.D., FRSA**Vice Chancellor, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for supporting me with all facilities needed for the completion of my thesis.

I wish to thank **Dr. (Mrs.) S. Kowsalya, M.Sc., M. Phil., Ph.D.,** Registrar, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for providing all the facilities needed for the completion of my thesis.

It is my pleasure to acknowledge the wholehearted support and encouragement extended to me by (Late)**Dr. P.R. Padma, M.Sc., Ph.D.,** Former Dean, Faculty of Bioscience and **Dr. (Mrs.) A. Vijayalakshmi M.Sc., MPhil., Ph.D.,** Dean of Bioscience, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore.

I am happy to express my hearty thanks to **Dr. (Mrs.). K. Manimegalai, M.Sc., Ph.D.,** Professor and Head of the Department of Zoology, Avinashilingam Institute for Home Science and Higher Education for Women, for her unflinching support and constant encouragement given throughout the tenure of my research.

I wish to express my sincere thanks and deep indebtedness to my guide **Dr. K. S. Santhy, M.Sc., M. Phil., Ph.D., M.C.A., PGDBI,** Professor, Department of Zoology, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for her amiable suggestions, eminent guidance and encouragement throughout the period of my work.

I express my sincere thanks to **all the members of staff**, Department of Zoology, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for their support.

I place on record of my heartfelt thanks to my family and friends for their love and support during my study period.

My thanks are also due to all those who have not been mentioned by name, but nevertheless, have been of tremendous help in their own ways.

Mercy J.

CONTENTS

CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
1	**INTRODUCTION**	1
2	**REVIEW OF LITERATURE**	9
	2.1. Hematopoiesis	9
	2.2. Leukemia	10
	2.3. Epidemiology	11
	2.3.1. India	11
	2.4. Etiology	12
	2.4.1. Pathology	12
	2.4.2. Risk factors	13
	2.4.3. Prognostic factors in AML	14
	2.5. Classification of AML	15
	2.5.1. FAB classification of AML	16
	2.6. Cytochemistry	16
	2.7. Molecular alterations and specific genes involved in AML	17
	2.7.1. Gene function	19
	2.7. 2. NPM1/ALK fusion protein	20
	2.7.3. Biochemical features	21
	2.7.4. Molecular genetics	22
	2.7.4.1. Somatic mutations	22
	2.8. Treatment methods in AML	24
	2.8.1. Induction therapy	24
	2.8.2. Post induction therapy	25
	2.8.3. Remission induction therapy	25

	2.8.4. Post remission therapy	26
	2.8.5. Hematopoietic stem cell transplantation	27
	2.8.6. Autologous stem cell transplantation	27
	2.9. Targeted therapy in AML	28
	2.9.1. FLT3 Inhibitors	28
	2.9.2. Midostaurin	29
	2.9.3. Quizartinib	30
	2.9.4. RAS	30
	2.9.5. JAKs	31
	2.9.6. Monoclonal antibodies and immunoconjugates	32
	2.10. System biology approach	32
3	**MATERIALS AND METHODS**	34
	3.1. RNA - Sequencing data	34
	3.2. Identification of DEGs	34
	3.3. Gene set enrichment analysis	36
	3.4. Functional annotation and pathway enrichment analysis of DEGs	37
	3.5. Identification of associated miRNAs with target DEGs	40
	3.6. Identification of transcription factors for targeting DEGs	41
	3.7. Integration of protein-protein interaction	42
	3.8. Hub gene validation	45
4	**RESULTS**	47
	4.1. Gene expression dataset	47
	4.2. Gene set enrichment analysis	48
	4.3. GO and KEGG enrichment analysis of DEGs	50
	4.4. Identification of miRNA - target gene network	52
	4.5. Transcription factors and survival analysis	53

	4.6. PPI network and module analysis	54
	4.7. Expression level and prognostic value of ten hub genes	56
5	**DISCUSSION**	60
6	**SUMMARY AND CONCLUSION**	68
	BIBLIOGRAPHY	

LIST OF TABLES

TABLE No.	TITLE	PAGE No.
I	Top ten up and down regulated genes between NPM1 mutation and wild-type	48
II	Gene ontologies of DEGs involved in NPM1 mutation associated AML	50
III	Enriched pathways of DEGs involved in NPM1 mutation associated AML	51

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
1	Hematopoiesis	10
2	Gene level differential expression analysis of workflow	35
3	The homepage for GSEA	37
4	The homepage for DAVID	38
5	The homepage for miRNet	41
6	The homepage for network analyst	42
7	The homepage for STRING	43
8	Cytoscape	44
9	Metascape	45
10	UALCAN	46
11	Volcano plot of comparing all of the significant DEGs from NPM1 mutation and NPM1 wild-type samples	47
12	GSEA results of NPM1 mutations in AML patients	49
13	Reconstructed KEGG pathway of cancer signaling by using KEGG mapper tool	51
14	Most associated miRNAs with DEGs	52
15	Most associated transcription factors with DEGs	53
16	Confidence view of protein-protein interaction network of DEGs predicted using Cytoscape software	54
17	PPI network of ten hub genes with high degree of connectivity	55
18	GO term analysis of top ten hub genes by Metascape	55
19	Expression level and prognostic values of ten hub genes	56

LIST OF ABBREVIATIONS

ABMT - Allogeneic or autologous blood or marrow stem cell transplantation

ALK - Anaplastic lymphoma kinase

ALL - Acute Lymphoblastic Leukemia

ALM - Acral lentiginous melanoma

AML - Acute Myeloid Leukemia

AML-MRC - Acute Myeloid Leukemia With Myelodysplasia-Related Changes

ara- C - cytosine arabinoside

ASXL1 - ASXL Transcriptional Regulator 1

BCAR1- Breast CAncer gene 1

BCLL - B-cell lymphocytic leukaemia

CBF - Core binding factor

CDK2- Cyclin-dependent kinase 2

CEBPA - CCAAT enhancer binding protein alpha

CLL - Chronic Lymphocytic

CML - Chronic Myeloid Leukemia

CMML - Chronic myelomonocytic leukemia

CR - Complete Remission

CREBBP - Cyclic adenosine monophosphate Response Element Binding protein

CSC - Cancer stem cells

DAVID - Database for Annotation, Visualization, and Integrated Discovery

DBGET - Integrated Database Retrieval System

DDX31- DEAD-Box Helicase 31

DEGs - Differentially expressed genes

DFS - Disease Free Survival

DGE - Digital gene expression

DNA - Deoxyribonucleic acid

DNMT3A - DNA Methyltransferase 3 Alpha

ECOG - Eastern Cooperative Oncology Group

EFS - Event free survival

ELN - Elastin

ERK - Extracellular signal-regulated kinase

ERK1 - extracellular signal-regulated kinase 1

ERK2 - extracellular signal-regulated kinase

EZH2 - Enhancer of zeste homolog 2

FAB - French-American-British

FGF2 - Fibroblast growth factor 2

FLT3 - fms-like tyrosine kinase 3

GATA2 - GATA Binding Protein 1

GLUL - Glutamate-Ammonia Ligase

GMP - Granulocyte-macrophage progenitors

GNAS - Guanine Nucleotide binding protein, Alpha Stimulating activity polypeptide

GO - Gene Ontology

GTP - Guanosine-5'-triphosphate

HCT - Hematopoietic Cell Transplantation

HLA - Human Leukocyte Antigens

HSC - Hematopoietic Stem Cells

HSCT - Haematopoietic stem cell transplantation

HTML - Hypertext Markup Language

IDH1 - Isocitrate dehydrogenase 1

IDH2 - Isocitrate dehydrogenase 2

IRES - Impaired internal ribosome entry site

ITD - Internal tandem duplication

JAK2 - Janus Kinase 2

KEGG - Kyoto Encyclopedia of Genes and Genomes

LSC - Leukemic Stem Cells

MAPK - Mitogen-activated protein kinase

MDS - Myelodysplastic Syndrome

MEP - megakaryocyte–erythroid progenitors

miRNA – microRNA

MLL-PTD - A partial tandem duplication (PTD) of the mixed lineage leukemia (MLL)

MPN - Myeloproliferative neoplasms

MPO - Myeloperoxidase

MRD - Minimal Residual Disease

MRO - Maestro

MRPL - Mitochondrial ribosomal protein L1

MRPS5 - Mitochondrial Ribosomal Protein S5

NGS - Next-generation sequencing

NIFK - Nucleolar Protein Interacting With The FHA Domain Of MKI67

NOD-SCID - Nonobese diabetic/severe combined immunodeficiency

NPM1- Nucleophosmin

NSCLC - non-small cell lung cancer

OS - Overall Survival

PB - Peripheral Blood

PPI - Protein-Protein Interaction

PPM1L - Protein phosphatase, Mg2+/Mn2+ dependent 1L

RAF - rapidly accelerated fibrosarcoma

RASSF9 - Ras association domain-containing protein 9

RNA - Ribonucleic acid

RNA-Seq - RNA sequencing

RPL7L1- Ribosomal protein L7 like 1

RTKs - Receptor tyrosine kinases

RUNX1- Runt-related transcription factor 1

sAML- Secondary Acute Myeloid Leukemia

SLC25A33 - Solute Carrier Family 25 Member 33

STAT3 - Signal transducer and activator of transcription 3

STRING - Search Tool for the Retrieval of Interacting Genes/Proteins

TCGA - The Cancer Genome Atlas

TCL – T-cell Lymphomas

TET2 - Tet Methylcytosine Dioxygenase 2

TF - Transcription factor

TP53- Tumor protein 53

WHO - World Health Organization

WT1 - Wilms' tumor gene

INTRODUCTION

1

INTRODUCTION

Cancer, the uncontrolled growth of cells, is a major cause of death throughout the world (Shewach and Kuchta, 2009). Cancer is the second leading cause of mortality worldwide. Overall, the prevalence of cancer has actually increased. Therefore, cancer is a serious problem affecting the health of all human societies. Unfortunately, it is a variety disease at the tissue level and this variety is a major challenge for its specific diagnosis, followed by efficacy of treatment. In men, the highest percentages of cancer types occur in the prostate, lung and bronchus, colon and rectum, and urinary bladder, respectively. In women, cancer prevalence is highest in the breast, lung and bronchus, colon and rectum, uterine corpus and thyroid, respectively (Hassanpour and Dehghani, 2017).

Cancer is a complex disease involving numerous tempo-spatial changes in cell physiology, which ultimately lead to malignant tumors. Abnormal cell growth (neoplasia) is the biological endpoint of the disease. Tumor cell invasion of surrounding tissues and distant organs is the primary cause of morbidity and mortality for most cancer patients. The biological process by which normal cells are transformed into malignant cancer cells has been the subject of a large research effort in the biomedical sciences for many decades (Seyfried and Shelton, 2010).

The question of "what causes cancer" has intrigued people for generations. In 1950, the World Health Organization sponsored an international symposium, and the attendees were intrigued by the dramatic variations in the types of cancer found in different areas of the world. It was learned that people who migrated to other countries, developed types of cancer common to their adopted countries, rather than their homelands. This implied that most cancers were caused by exposures in the environment, rather than inherited genetic factors (Blackadar, 2016). Among various diseases, cancer has become a big threat to human beings globally. As per Indian population census data, the rate of mortality due to cancer in India was high and alarming with about 806000 existing cases by the end of the last century. All types of cancers have been

reported in Indian population including the cancers of skin, lungs, breast, rectum, stomach, prostate, liver, cervix, esophagus, bladder, blood, mouth etc (Ali *et al*., 2011).

In recent years remarkable progress has been made towards the understanding of proposed hallmarks of cancer development and treatment. However with its increasing incidence, the clinical management of cancer continues to be a challenge for the 21st century. Treatment modalities comprise of radiation therapy, surgery, chemotherapy, immunotherapy and hormonal therapy (Baskar *et al*., 2012). The future of cancer management is expected to be profoundly dependent upon the use of biomarkers that will guide physicians at every step of disease management. Cancer biomarkers can be used for the accurate evaluation and management of the disease in different stages. They can be useful for predicting several outcomes during the course of disease including early detection, outcome prediction and detection of disease recurrence. Most importantly, with the clinical appearance of many new therapeutic agents, appropriate markers can be used to determine which tumors will respond to which treatments in order to predict the likelihood of drug resistance (Chatterjee and Zetter, 2005).

Radiation therapy is an essential element of curative treatment of cancers of the breast, prostate, cervix, head and neck, lung, brain, as well as sarcomas. In more advanced disease stages, radiation therapy is used before, during, or after surgery and is frequently combined with chemotherapy, either as concurrent or adjuvant treatment (Jaffray and Gaspodarowicz, 2015). Radiation is a physical agent, which is used to destroy cancer cells. The radiation used is called ionizing radiation because it forms ions and deposits energy in the cells of the tissues it passes through. This deposited energy can kill cancer cells or cause genetic changes resulting in cancer cell death (Baskar *et al*., 2012).

Chemotherapy is a treatment that uses drugs that kill rapidly dividing cancer cells to prevent them from growing and making more cells. The era of chemotherapy began in the 1940s with the first uses of nitrogen mustards and antifolate drugs. Cancer drug development since then has transformed from a low-budget, government-supported research effort to a high-stakes, multi-billion dollar industry. The targeted-therapy revolution has arrived, but the principles and limitations of chemotherapy discovered by the early researchers still apply (Chabner and Roberts, 2005).

Major advances have been made in the field of immunology in the past two decades. A better understanding of the molecular and cellular mechanisms controlling the immune system has opened the door to many innovative and promising new cancer therapies that manipulate the immune response (Borghaei *et al*., 2009). Harnessing the immune system to treat chronic infectious diseases or cancer is a major goal of immunotherapy. Among others, impediments to this aim include host failure to identify tumor antigens, tolerance to self and negative immunoregulatory mechanisms. But with recent progress, active and passive immunotherapy are proving themselves as effective therapeutic strategies (Waldmann, 2003).

Hormone therapy is an effective and non-toxic therapy for oestrogen and progesterone receptor-positive breast cancer and prostate cancer. Serum levels of oestradiol and testosterone are controlled by the hypothalamic–pituitary–gonadal pathway. Hormone therapy can be used to reduce the size of the primary cancer prior to radical surgery or radiotherapy or to reduce the risk of recurrence. Hormone therapy is highly effective in patients with locally advanced or metastatic disease, with a high response rate. Most patients eventually relapse with 'castrate-refractory' disease, for which increasing numbers of active agents are entering clinical practice (Abraham and Staffurth, 2016).

Radiotherapy and surgery are local, or locoregional, approaches to cancer treatment. As long as the cancer is localized to its site of origin or has spread to the regional lymph nodes only, there is a chance of curing the disease using a locoregional approach. However, the natural history of many forms of cancer has taught us that in many situations the cancer cells are not confined locally or regionally, although they appear to be so using currently available diagnostic tools. In these situations, diagnostic and staging tools show only local or locoregional disease, but in reality malignant cells have already escaped the locoregional boundaries and will eventually induce disease recurrence and possibly distant metastasis. In this scenario, local or locoregional therapies are clearly not enough, and the treatment of cancer requires a systemic approach involving chemotherapy, hormonal therapy or targeted therapy (Rosenblatt and Zubizarreta, 2017).

Recent advances in understanding the biologic mechanisms underlying cancer development have driven the design of new therapeutic approaches, termed 'targeted therapies', that selectively interfere with molecules or pathways involved in tumor growth and progression.

Inactivation of growth factors and their receptors on tumor cells as well as the inhibition of oncogenic tyrosine kinase pathways and the inhibition of molecules that control specific functions in cancer cells constitute the main rational bases of new cancer treatments tailored for individual patients (Ciavarella *et al*., 2010). Although targeted therapies can be achieved directly by altering specific cell signaling by means of monoclonal antibodies or small molecules inhibitors (Pérez-Herrero and Fernández-Medarde, 2015).

Hematopoiesis means the formation of blood cellular components that occurs during embryonic development and throughout adulthood to produce and replenish the blood system. Studying hematopoiesis can help scientists and clinicians to understand better the processes behind blood disorders and cancers. Furthermore, hematopoietic stem cells (HSCs) can be used as a model system for understanding tissue stem cells and their role in ageing and oncogenesis (Jagannathan Bogdan and Zon, 2013). Continuous blood cell production throughout the lifetime of an individual is ensured by hematopoietic stem cells (HSCs), which are rare bone marrow cells that possess extensive self-renewal capacity and ability to differentiate to all blood cell lineages. Due to the unique properties of HSCs, they can regenerate the entire hematopoietic system of a recipient upon transplantation and thereby provide a cure for inherited and acquired blood diseases. As such, HSCs are of substantial therapeutic interest (Gekas *et al*., 2010).

The hematopoietic system remains robust with regards to extrinsic perturbations, in sharp contrast with the stochastic behavior of hematopoeitic stem cells (HSCs) at the single cell level, suggesting that stability may be achieved within a stem cell system that undergoes constant self-renewal, commitment to differentiation and generates cell type diversification (Hoang, 2004). When the processes of self-renewal and differentiation become deregulated or uncoupled, leukemias can result, characterized by an accumulation of immature blast cells that fail to differentiate into functional cells. Like other neoplasms, leukemia arises from the clonal expansion of a single cell and is sustained by a leukemic stem cell (LSC) (Warner *et al*., 2004).

Leukemia is an amalgam of cancers and arises due to the malignancy of the any elements of blood and bone marrow. In other terms, they are abnormal white blood cells, which are not fully developed and are called blasts or leukemia cells. The growth of the Leukemia cells are rapid than Normal cells. As with time, they replace the population of the normal WBCs and RBCs and may spread to the lymph nodes and other organs. As in 2012, 3, 52,000 people were

affected by Leukemia and 2, 65,000 deaths occurred. It is the most common type of cancer in children and three quarters of Cases being) about 90% of all leukemias are diagnosed in adults, with acute lymphoblastic leukemia (ALL) and chronic lymphocytic leukemia (CLL) being most common in adults (Chapla Ugandhar *et al*., 2015).

Acute myeloid leukemia (AML) is the most common acute leukemia in adults, with an incidence of over 20, 000 cases per year in the United States alone. Large chromosomal translocations as well as mutations in the genes involved in hematopoietic proliferation and differentiation result in the accumulation of poorly differentiated myeloid cells. AML is a highly heterogeneous disease; although cases can be stratified into favorable, intermediate and adverse-risk groups based on their cytogenetic profile, prognosis within these categories varies widely (De Kouchkovsky and Abdul Hay, 2016).

Leukemia cells usually invade the blood fairly quickly. They can then spread to other parts of the body, including the lymph nodes, liver, spleen, central nervous system (brain and spinal cord), and testicles. Other types of cancer also can start in these organs and then spread to the bone marrow, but these cancers are not leukemia. The term “acute” means that the leukemia can progress quickly, and if not treated, would probably be fatal within a few months (Chapla Ugandhar *et al*., 2015). Although many patients with AML have a response to induction chemotherapy, refractory disease is common, and relapse represents the major cause of treatment failure (Papaemmanuil *et al*., 2016).

Accordingly, the rules for AML classification require collection of the patient’s history, including previous cytotoxic therapies, which define “therapy-related myeloid neoplasms,” or a prior history of Myelodysplastic syndrome (MDS) defining “AML with myelodysplasia-related changes” (“AML-MRC”). The second field of investigation for classifying a case of AML is the presence of specific gene mutations or rearrangements defining the category of “AML with recurrent genetic abnormalities”. The detection of balanced or unbalanced cytogenetics aberrations considered associated with MDS and/or detection of multilineage dysplasia by morphology, defines the disease as “AML-MRC.” When the disease cannot be classified in another category, the morphologic exam of bone marrow and peripheral blood is the only parameter useful in the subcategorization of AML, not otherwise specified (Voso *et al*., 2016).

Recurring chromosome abnormalities are strongly associated with certain subtypes of leukemia, lymphoma and sarcomas. Chromosome translocations are frequently observed in both de novo and therapy-related acute AML and MDS (Zhang and Rowley, 2006). Chromosomal translocations regulate the behavior of leukemia. They not only predict outcome but they define therapy. There is a great deal of knowledge on the products of leukemic translocations, yet little is known about the mechanism by which those translocations occur. Given the large number of DNA double-strand breaks that occur during normal progression through the cell cycle, especially from V(D)J recombination, stalled replication forks or failed decatenation, it is surprising that leukemogenic translocations do not occur more frequently. Fortunately, hematopoietic cells have sophisticated repair mechanisms to suppress such translocations. When these defenses fail leukemia becomes far more common, as seen in inherited deficiencies of DNA repair. Analyzing translocation sequences in cellular and animal models, and in human leukemias, has yielded new insights into the mechanisms of leukemogenic translocations (Nickoloff *et al*., 2008).

All cancers arise as a result of changes that have occurred in the DNA sequence of the genomes of cancer cells. Over the past quarter of a century much has been learnt about these mutations and the abnormal genes that operate in human cancers. We are now, however, moving into an era in which it will be possible to obtain the complete DNA sequence of large numbers of cancer genomes (Stratton *et al*., 2009). Genome sequencing is an important step toward correlating genotypes with phenotypic characters. Sequencing technologies are important in many fields in the life sciences, including functional genomics, transcriptomics, oncology, evolutionary biology, forensic sciences, and many more (Verma *et al*., 2017).

The human genome contains approximately 3.2 billion nucleotides and about 23,500 genes. Each gene has protein-coding regions that are referred to as exons. The human genome contains about 180,000 exons, which are collectively called an exome. An exome comprises about 1% of the human genome and hence is about 30 million nucleotides in size. Today's technologies afford the opportunity to sequence all nucleotides in the human exome and even in the human genome. Given that more than three-quarters of the known disease-causing variants are located in the exome (Marian, 2014).

New technologies for DNA sequencing, coupled with advanced analytical approaches, are now providing unprecedented speed and precision in decoding human genomes. This combination of technology and analysis, when applied to the study of cancer genomes, is revealing specific and novel information about the fundamental genetic mechanisms that underlie cancer's development and progression (Mardis, 2012). Discoveries from cancer genome sequencing have the potential to translate into advances in cancer prevention, diagnostics, prognostics, treatment and basic biology (Mwenifumbo and Marra, 2013).

Although great advances have been made in the therapeutic methods for AML, still prognosis is far from ideal. However, there are some comprehensive studies, many AML patients with low cure rate are drug resistant and will eventually experience early recurrence. Such recurrence rate is increasing every year, more likely to cause resistance to chemotherapy and failed remission of leukemia stem cells. Furthermore, the incidence of AML had been more and a subset of patients died within one year due to distinct metastasis. Therefore, the molecular association in the occurrence and prognosis of AML needs to be explored, which will contribute to the finding of diagnosis and prognostic markers, and therapeutic targets of AML. Due to the revolution in techniques and the cost reduction of next-generation sequencing (NGS) there are several studies which have categorized the AML based on the cytogenetics and molecular mutations (Renneville *et al*., 2008).

Over the years, bioinformatic study based on gene expression of RNA-seq has seemed as an effective new approach to diagnose the important molecular pathogenesis of tumors. A few publicly accessible databases from large patient associates have been built, which provide the chance to distinguish biomarkers in correlation with disease expansion and treatment response. Signaling pathways such as RTKs (Matthews *et al*., 2003), MAPK (Platanias, 2003), Ras/RAF/MEK/ERK (Parcells *et al*., 2006; Shaw and Cantley, 2006) and P13K/AKT pathway (Martelli *et al*., 2006) were associated in pathogenesis of AML. In any case, a positive and definite outcome regarding Nucleophosmin (NPM1) mutation in AML is still inadequate. Therefore, the aim of the present study was to find the key genes and pathways involved in NPM1 mutations of AML.

In the present study, by means of a series of bioinformatic tools, we acknowledged differentially expressed genes in RNA-seq data with NPM1 mutation. Then the differentially

expressed genes (DEGs) were exposed to the pathway enrichment and Gene Ontology (GO) analysis. A Protein-Protein Interaction (PPI) network and clusters were created, visualized and analyzed, and hub genes were identified. Identification of a characteristic signature of mutant NPM1 AML may help to determine the chemotherapeutic response to other chemotherapeutic agents for the individualized selection of therapies and regimens. Considering all the above facts the present study was conducted with the following objectives:

OBJECTIVES:

- To identify Differentially Expressed Genes (DEGs) between Nucleophosmin gene (NPM1) mutation and normal patients
- To assess the cellular, molecular, and pathogenic characteristics of Acute Myeloid Leukemia (AML)
- To identify genes that could be involved in alternative pathways linked to the pathogenesis of Acute Myeloid Leukemia (AML)
- To identify potential mechanism, biomarkers and therapeutic targets for Acute Myeloid Leukemia (AML)

REVIEW OF LITERATURE

2

REVIEW OF LITERATURE

The review of literature related to the study "Computational identification of key genes and pathways associated with Acute Myeloid Leukemia" is presented in the following headings:

2.1. Hematopoiesis

Hematopoiesis includes the formation and development of blood cells which is initiated by a small fraction of multipotent HSCs. HSCs are located in the bone marrow and they have a capability for self-renewal. HSC can be distinguished by the lack of lineage markers and presence of CD133 or c-kit or sca-1 (Wognum *et al*., 2003). The self renewal ability of the HSCs maintains the stem cell pool and supplies cells for multilineage hematopoiesis during the entire lifespan of the individual. During the process of hematopoiesis HSCs differentiate into long and short term HSCs and subsequently into common lymphoid and myeloid progenitors (Akashi *et al*., 2000). The common myeloid progenitors differentiates into granulocyte–macrophage progenitors (GMP) which in turn produce monocytes, macrophages and granulocytes, and megakaryocyteerythroid progenitors (MEP) which differentiate into megakaryocytes, platelets and erythrocytes (Akashi *et al*., 2000).

Furthermore, the common lymphoid progenitors differentiate into mature T and B cells (Akashi *et al*., 1999). Hematopoiesis is highly controlled by intrinsic mechanisms such as the expression of transcription factors or epigenetic modifications, and extrinsic mechanisms like growth factors, cytokines and the microenvironment or stem cell niche (Metcalf, 1998). The balance between these mechanisms determines whether cells remain quiescent, proliferate, differentiate, self-renew or undergo apoptosis. In normal conditions, the progenitors mature into several types of specialized cells to execute their function in immunity, coagulation or transport of molecules. During life cycle, cell may acquire genetic damage caused by internal or external foreign substances and these genetically damaged immature blasts may fill up the bone marrow and displace normal hematopoietic cells

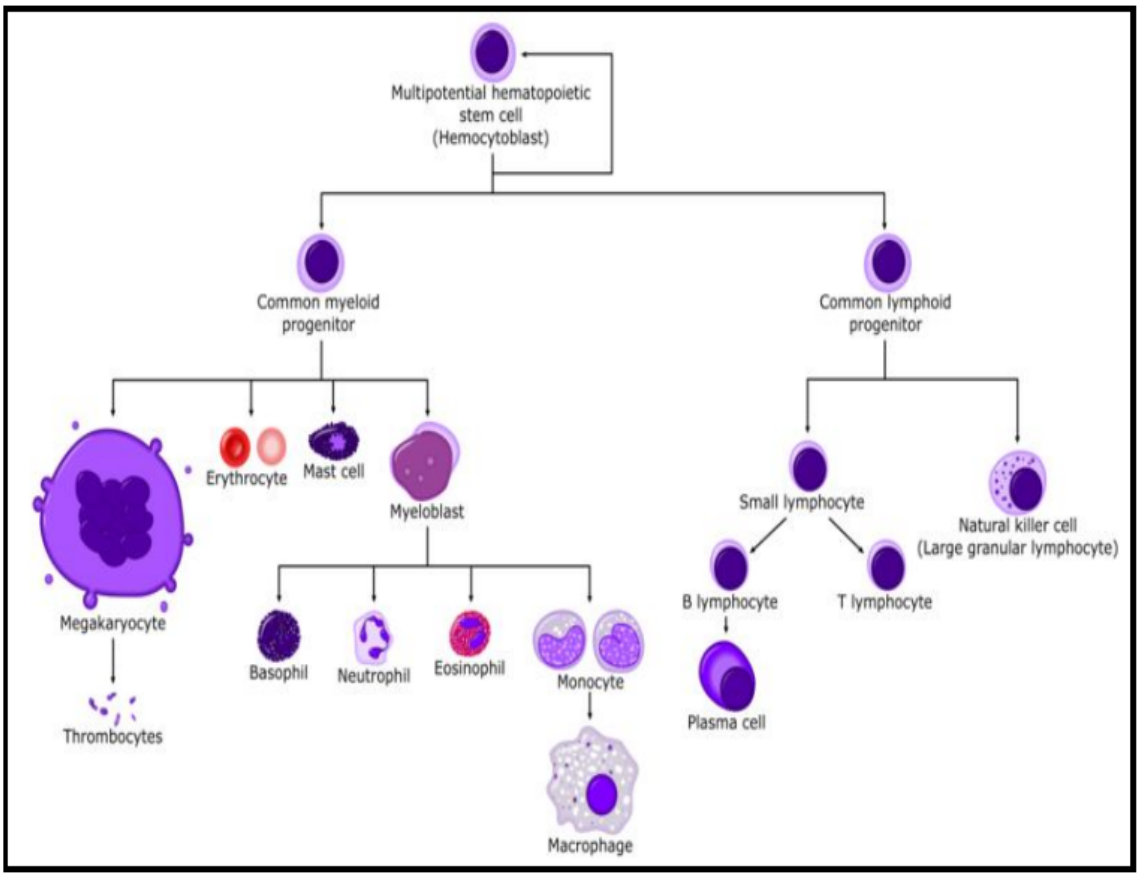


Figure 2.1 Hematopoiesis (Adapted from Adolfsson *et al.*, 2005)

2.2. Leukemia

Leukemia is a progressive, neoplastic disease of the hematopoietic system characterized by unregulated proliferation of uncommitted or partially committed stem cells. It includes a heterogeneous group of neoplasms that differ with respect to aggressiveness, cell of origin, clinical features, and response to therapy. Leukemia's are divided into two broad categories that are based on the cell involved (myeloid or lymphoid) and disease aggressiveness (either acute or chronic). Acute Leukemia (AL) is malignant disorders of bone marrow hematopoietic precursors. AL is broadly divided into AML and ALL based on cell of origin (Haase *et al.*,

1995). In most acute leukemias, the malignant transformation is believed to occur in hematopoietic 'sleepers' or inactive multipotent stem or early progenitor cells with the capacity of lineage differentiation (Helleberg *et al*., 1997). The hallmark of these diseases is the leukemic blast cells, which ultimately interfere with normal bone marrow elements necessary for survival.

2.3. Epidemiology

AML is a grave disease with an incidence of 4 per 100,000 a year. It can present in all ages, but the median age is 70 years. One-third of such patients have secondary AML, that is, AML following chemoradiotherapy or a transformation from previous MDS or myeloproliferative neoplasia. A combination of genetic, epigenetic, and environmental factors may be responsible for the development of most cases of AML (Juliusson *et al*., 2021).

Although AML are infrequent diseases, they are highly malignant neoplasms responsible for a large number of cancer-related deaths. AML is the most common type of leukemia in adults, yet continues to have the lowest survival rate of all leukemias. While results of treatment have improved steadily in younger adults over the past 20 years, there have been limited changes in survival among individuals of age > 60 years (Deschler and Lubbert, 2008).

The incidence rate of AML varies with gender and race. AML is more common in males than females. In 2000, the US age-adjusted incidence rate of AML was 3.9 per 100,000 across both sexes, 4.8 per 100,000 for males and 3.3 per 100,000 for females. Of the estimated 10,500 new cases to be diagnosed in 2003, approximately 5800 (55%) will be in males. Among children, the incidence rate of AML has not been shown to be associated with gender or race. Between 1887 and 1999, AML was more common in whites than in blacks. However, in 2000, the incidence of AML for blacks surpassed that for whites: 3.9 per 100,000 for whites and 4.3 per 100,000 for blacks. The reasons for these gender and racial differences have not been established (Redaelli *et al*., 2003).

2.3.1. India:

In the developing countries, including India, the incidence rate of leukemia are 3 to 4 per 100,000 (Ferlay *et al*. 2008). In India, myeloid leukemias were more common than lymphoid leukemias. For myeloid leukemias, Aizawal district showed the highest Average Annual Return

(AAR) (5.6) followed by Mizoram state (3.2), Imphal (3.0) and Delhi (2.7) in males. For females, Imphal west district (2.7) registered the highest AAR followed by Thiruvanthapurum (2.3) and Ahmadabad urban (2.1) [ICMR 2010]. In Delhi, leukemia accounts for approximately 5% of all cancers. Male to Female ratio was higher in lymphoid leukemias (2.48) than myeloid leukemias (1.61). In patients, below 14 years of age, lymphoid leukemia is the most common childhood cancer accounting for 42% of all cancers. In adult, myeloid leukemia is more common than lymphoid constituting 63% of all leukemias (ICMR 2010).

2.4. Etiology

The exact cause of leukemia is unknown despite many advances in the treatment, but damage to the DNA of HSCs is a key event. Epidemiologic studies of AML have examined possible risk factors, including genetic, infectious and environmental, in an attempt to determine its etiology. The environmental risk factors include ionizing radiation (Infante-Rivard *et al*., 2000; Mahoney *et al*., 2004), non-ionizing radiation, chemical, pesticides (Glass *et al*., 2003), alcohol use, cigarette smoking [Pogoda *et al*., 2004] and illicit drug use. Most environmental risk factors have been found to be weakly and inconsistently associated with either form of AML. Chemotherapeutic drugs (alkylating agents and topoisomerase inhibitors) used to treat cancers such as breast cancer, or lymphomas are also associated with an increased risk of Leukemia. There are a number of well-known risk factors that are established causes while others are presumed causes of leukemia (O'Connor *et al*., 2007).

2.4.1. Pathology

The pathogenesis of AML involves an array of molecular alterations that disrupt almost every facet of cell transformation. These processes include the regulation of cell proliferation, differentiation, self-renewal, survival, cell cycle checkpoint control, DNA repair and chromatin stability, and cell dissemination. Normal regulatory networks are disrupted or usurped by these leukemogenic insults, and the understanding of these alterations is guiding the design of new therapeutic strategies. These include inappropriate proliferation in the absence of normal growth signals, indefinite self-renewal in a manner analogous to a stem cell, escape from programmed cell death, and inhibition of differentiation, aberrant cell cycle checkpoint control, genomic instability and multi-organ dissemination of leukemic cells (Licht and Sternberg, 2005).

The pathogenesis of AML is characterized by the serial acquisition of somatic mutations and several genes are recurrently mutated in AML. Exposures to benzene, cigarette smoking, pesticides, embalming fluids, accidental or professional ionization radiation, therapeutic radiotherapy, and radioactive I-131 therapy can cause AML with or without a preceding MDS phase (Juliusson *et al.*, 2021). Pathologic classifications must change continuously to reflect advances in our understanding of disease. The pathologic classification of AMLs used by most physicians today, however, does not provide the most information possible for these diseases. The French-American-British (FAB) Cooperative Group provided clear and useful criteria for the pathologic classification of AML and the classification was modified to incorporate new disease types that required ancillary testing not used in the original descriptions.

Some of the FAB categories of AML, such as M3 and M4Eo, correlate with prognostically significant cytogenetic abnormalities; however, the lack of reproducibility and clinical significance of some of the disease groups in the FAB system is now generally acknowledged. In addition, the FAB classification does not accommodate the importance of myelodysplasia-associated changes in adult AML. Alternative classification systems have been proposed that incorporate immunophenotyping, cytogenetic, and myelodysplastic changes, and morphologic and immunophenotypic features are now well described for some AMLs that are highly suggestive of recurring cytogenetic abnormalities (Arber, 2001).

2.4.2. Risk factors

A variety of environmental and chemical exposures are assumed to be associated with a variably elevated risk of developing AML in adults. Exposure to ionizing radiation is linked to AML. Among survivors of the atomic bomb explosions in Japan, an increased incidence of AML was observed with a peak at 5–7 years after exposure. Also, therapeutic radiation has been found to increase the risk of secondary AML. Chemotherapeutic agents, such as alkylating agents and topoisomerase II inhibitors, have been reported to increase the incidence of AML and will be discussed in detail below. A number of other substances (therapeutic and occupational) have been linked to an increased risk for the development of AML. Benzene is the best studied and widely used potentially leukemogenic agent. Persons exposed to embalming fluids, ethylene oxides, and herbicides also appear to be at increased risk. Furthermore, smoking has been

discussed to be associated with an increased risk of developing AML (particularly of FAB subtype M2), especially in those aged 60–75 (Deschler *et al*., 2008).

Epidemiologic studies of acute leukemias in children have examined a number of possible risk factors (e.g., environmental, genetic, or infectious) in an effort to determine the etiology of the disease. Only one environmental risk factor has been significantly linked with either ALL or AML; most environmental risk factors [e.g., electromagnetic fields, cigarette smoking] have been weakly or inconsistently associated with either form of childhood leukemia. Because childhood leukemia is a rare occurrence, prospective studies are difficult to conduct, and therefore studies most frequently use a retrospective case–control design (Belson *et al*., 2007).

2.4.3. Prognostic factors in AML

Prognostic factors can be subdivided into those that are related to the patient and those that are related to the disease. Patient-associated factors (e.g., increasing age, coexisting conditions, and poor performance status) commonly predict treatment-related early death, whereas disease-related factors (e.g., white-cell count, prior myelodysplastic syndrome or cytotoxic therapy for another disorder, and leukemic-cell genetic changes) predict resistance to current standard therapy. Because of marked improvements in supportive care in many older patients, the risk of treatment-related death is considerably lower than the risk that the disease will prove to be resistant to treatment. Indeed, treatment-related mortality appears to have decreased substantially in recent years (Othus *et al*., 2013).

The evaluation of molecular genetic lesions as prognostic and predictive markers is an active research area. Currently, three molecular markers (NPM1 and CEBPA mutations and FLT3 internal tandem duplications) are used in clinical practice, as reflected in the European Leukemia Net (ELN) recommendation. It is expected that additional markers (e.g., RUNX1, ASXL1, and TP53) that have consistently been associated with an inferior outcome will soon be included in these recommendations. The prognostic importance of other mutated genes (e.g., DNMT3A, IDH1, IDH2) is less clear (Dohner *et al*., 2015).

Prognostic factors in AML

Prognostic Factors	Favourable prognosis	Unfavourable prognosis
Age	Children	Elderly
Gender	Female	Male
WBC	Low $\leq 20\times10^9$ /L	High ($\geq100\times10^9$ /L
Response after Induction	CR	PR/RD
Time to relapse	Long	Short
Performance scale	Good	Poor
FAB Morphology	M3, M4	M0, M5, M6, M7
Karyotype	Inv(16), t(16;16), t(15;17), t(8;21)	$7q^-$, $5q^-$, t(9;22), 11 q23 with MLL rearrangements
Numerical aberrations	-	-5, -7, +8
Immunophenotype	Panmyeloid marker	CD34+ve, CD13+ve, CD14+ve, CD11b, CD11c, Biophenotypic (>2 lymphoid markers).

2.5. Classification of AML

Because of the increasing recognition of the importance of genetic events to the diagnosis and treatment of the acute leukemias, the proposed new World Health Organization (WHO) classification incorporates genetic aberrations and immunology as major defining features in addition to morphology. In a hierarchal approach, genetic changes have precedence in the acute myeloid leukemias and immunology and genetic changes have precedence in the acute

lymphoblastic leukemias. Four major groups of acute myeloid leukemia are recognized: 1) Acute myeloid leukemias with recurrent genetic abnormalities, 2) Acute myeloid leukemia with multilineage dysplasia, 3) Acute myeloid leukemias, therapy related, and 4) Acute myeloid leukemia not otherwise categorized (Brunning, 2003).

2.5.1. FAB Classification of AML

Prior to the proposals of the FAB cooperative group in 1976, there was a plethora of terminology in regard to the classification of the acute leukemias. Although investigators in different institutions generally understood each other and were able to interrelate the various terms for the different types of acute leukemia, the lack of uniform terminology hindered communication and the objective evaluation of the various therapeutic protocols. As a result, the introduction of the FAB classification was a major and necessary advance in the recognition and understanding of the morphologic types of acute leukemia. The FAB classification is primarily morphologic based with the addition of a small number of cytochemical reactions and reflected the status of morphologic diagnosis at the time it was proposed in 1976. According to the morphological FAB classification, AML can be divided into eight subtypes (M0-M7). The subtypes differ with respect to the myeloid lineage involved (e.g., granulocytic, monocytic, erythrocytic, megakaryocytic) and the degree of leukemic cell differentiation (Hwang, 2020).

2.6. Cytochemistry

Cytochemistry in hematology refers to the staining methods used to identify the chemical composition of cells without significantly altering the cell morphology. Most cellular cytochemical markers are organelle-associated enzymes and other proteins. Cytochemical staining reactions are of two types: 1. Enzymatic: e.g. Myeloperoxidase (MPO), 2. Non enzymatic: e.g. Sudan black (Stains lipids) PAS (stains glycogen).

Cytochemistry and light microscopy (primarily Wright-Giemsa or May-GrünwaldGiemsa staining) is the principle method for the diagnosis and FAB-subclassification of AML (Smith *et al*., 2004). For diagnosis, all cases of AML, except M0 must stain positive for MPO and Sudan Black. MPO is an enzyme present in the primary and secondary granules of granulocytes and their precursors- metamyelocyte, myelocyte promyelocyte and myeloblast. Monoblasts,

monocytes and lymphoblasts are negative. Therefore, MPO positivity is a paramount feature of AML. It is done side by side with a Romanowsky stain. Sudan Black B stains lipid membranes of the granules, which contain enzyme myeloperoxidase. SBB also stains the cells of myeloid series. Other stains that can be briefly mentioned are non-specific esterase (AML-M2, M4 or M5) and chloroacetate esterase (late myeloblast and early promyelocyte stage AML-M3) (Smith *et al*., 2004).

2.7. Molecular alterations and specific genes involved in AML

A third of the AMLs have a mutation of FLT3. FLT3 can be rearranged respectively with ITD or with TKD (in 25% and 7% of AML, respectively). Both mutations lead to receptor activation, but with different prognostic implications: FLT3-ITD is an independent factor that predicts a higher relapse rate and worse survival, while FLT3-TKD is associated with better survival. In contrast to AML with balanced rearrangements, AML with chromosomal aneuploidies represent a significantly more heterogeneous subgroup. The majority (>60%) have a complex karyotype (≥3 chromosomal events).

The most common abnormalities are −5/5q, −7/7q, −17/17p, and −12/12p. Around 50% of patients in this homogeneous group have a mutation of TP53, which is associated with more advanced age. The median age of patients with chromosomal aneuploidies and mutation of TP53 is 58 years, compared to 49 years of patients who only have aneuploidy. In addition, the coexistence of aneuploidy and mutation of TP53 is associated with a very unfavourable prognosis, compared to just having the complex karyotype alone. Due to this, the study of the TP531 mutation is recommended at the time of diagnosis, in addition to the karyotype (Moarri and Papaemmanuil, 2017; Levis, 2013).

Research carried out in patients with normal karyotype has shown that some of these patients have mutations of the genes CEBPA and NMP1. These mutations are present in 10% of AML with a normal karyotype and can be observed in the context of an abnormal karyotype, for example (9q), and are mutually exclusive of balanced rearrangements. Several studies have shown that only cases of biallelic mutation of CEBPA associate with a favourable prognosis in AML, because they are usually associated with mutations of GATA2 and typically lack FLT3-ITD (Wouters *et al*., 2009; Taskesen *et al*., 2011; Green *at al*., 2010).

Mutations of NPM1 can be observed in up to 50% of AML with normal karyotype. The absence of FLT3-ITD rearrangements is associated with a good prognosis, while if both mutations coexist, the risk of relapse increases and the prognosis worsens (Vassiliou *et al*., 2011). RUNX1 is a recurrent somatic mutation in patients with AML with normal karyotype. It is usually exclusive of balanced translocations and mutations of NPM1 and CEBPA and is associated with type M0 AML of the French–American–British classification, trisomy of chromosome 13 and mutations in MLL-PTD, IDH2 and ASLX1. Further, RUNX1 is mutated in 10% of the cases of myelodysplastic syndromes and, in certain cases, could indicate a subgroup of AML secondary to these syndromes. Some studies in young patients with AML have shown that mutations in RUNX1 are an independent factor of adverse clinical prognosis (Tang *et al*., 2009 ; Gaidzik *et al*., 2016).

Within the genetic heterogeneity of AML, the main epigenetic mutations have been described in DNMT3A, TET2, ASXL1, CREBBP and EZH2. Therefore, DNMT3A is mutated in 30% of AML with normal karyotype. Mutations in DNMT3A, NPM1 and FLT3- ITD usually "co-segregate", suggesting their cooperation in the pathogenesis of AML. Similarly, deregulation of the IDH/WT1/TET2 pathway could be operative in a third of AML cases affecting young adults (Shlush *et al*., 2014).

Hematopoiesis is highly regulated through cytokine-induced stimulation of multiple signal transduction pathways in order to mediate appropriate differentiation and proliferation of specific progenitor populations. Ligand-induced stimulation of the FLT3 leads to activation of multiple downstream effector pathways resulting in differentiation and proliferation of specific progenitor cell populations.

Genomic alterations of the FLT3 gene, including FLT3 internal tandem duplication (FLT3/ITD) and FLT3 activation loop mutation (FLT3/ALM) lead to autonomous receptor activation, dysregulation of FLT3 signal transduction pathways, contribute to myeloid pathogenesis, and have been linked to response to therapy and clinical outcome. Exploring the mechanisms by which these FLT3 alterations lead to dysregulated proliferation should provide a better understanding of the molecular pathogenesis of AML and may provide insights into

potential therapeutic interventions. FLT3 inhibitors are under evaluation for their efficacy in AML patients with FLT3 mutations (Meshinchi and Appelbaum, 2009).

The majority of genetic and molecular studies on AML have focused on the characterization of alterations present at the time of diagnosis, yet, as outlined above, a large proportion of AML patients with primarily responsive disease ultimately die due to relapse with refractory leukemia. The survival of stem cells, whose regrowth leads to disease recurrence, is assumed to be due in part to protective effects of the microenvironment and in part to cell autonomous mechanisms elicited by molecular alterations in the stem cells themselves, as has been impressively demonstrated in the case of acute lymphoblastic leukemia. Such molecular changes may already have been present in a subset of stem cells at presentation, or may have emerged during, and even as a consequence of the mutagenic effects of, cytostatic therapy (Hackl *et al*., 2017).

2.7.1. Gene function

Nucleophosmin is a nucleolar phosphoprotein that is more abundant in tumor cells than in normal resting cells. Stimulation of the growth of normal cells, e.g., mitogen activation of B lymphocytes, was accompanied by an increase in nucleophosmin protein level. They stated that nucleophosmin is likely involved in the assembly of ribosomal proteins into ribosomes. Electron microscopic study indicated that nucleophosmin is concentrated in the granular region of the nucleolus, where ribosome assembly occurs (Chan *et al*., 1989).

Nucleophosmin as a substrate of CDK2/cyclin E in centrosome duplication. NPM1 associated with unduplicated centrosomes, and dissociated from centrosomes by CDK2/cyclin E-mediated phosphorylation. An anti-NPM1 antibody, which blocked this phosphorylation, suppressed the initiation of centrosome duplication in vivo. Moreover, expression of a nonphosphorylatable mutant NPM1 in cells effectively blocked centrosome duplication.

NPM1 is a target of CDK2/cyclin E in the initiation of centrosome duplication. Centrosome-bound NPM/B23 dissociates from centrosomes upon phosphorylation by CDK2/cyclin E, which in turn triggers initiation of centriole duplication. When the nuclear membrane breaks down during mitosis, most nuclear proteins disperse throughout the cytoplasm, including NPM/B23, some of which relocalizes to the centrosomes (spindle poles). After mitosis,

each daughter cell receives one centrosome bound by NPM/B23, which dissociates from the centrosome upon exposure to CDK2/cyclin E in the next cell cycle. In this setting, NPM/B23 would constitute a licensing system for centrosome duplication, ensuring the coordination of centrosome and DNA duplication, as well as restricting centrosome duplication to occur once, and only once, within a single cell cycle (Okuda *et al*., 2000).

By immunohistochemistry using antibodies that did not differentiate between NPM1 isoforms, nuclear staining for NPM1 in control HeLa cells and cytoplasmic staining following transfection with FGF2. They concluded that overexpression of FGF2 caused the redistribution of both NPM1 isoforms. By transfection of the C-terminally truncated NPM1 variant (NPM2) into radiosensitive HeLa cells, showed that the radio resistance associated with FGF2 overexpression was mediated by increased expression of this NPM1 isoform (Dalenc *et al*., 2002).

Using immunohistochemical analysis, colocalization of NPM1 with DDX31 in nucleoli in renal cell carcinoma cell lines. Reciprocal co immunoprecipitation analysis showed that full-length DDX31 interacted with NPM1 in RCC cells. Knockdown of either DDX31 or NPM1 attenuated preribosomal RNA biogenesis. Knockdown of DDX31 also reduced cell growth, concomitant with translocation of NPM1 from nucleoli to cytoplasm. Cytoplasmic NPM1 bound HDM2, thereby reducing binding of HDM2 to p53 and causing G1 cell cycle arrest and apoptosis (Fukawa *et al*., 2012).

NPM1 regulated 2-prime-O-methylation of ribosomal RNA (rRNA) in mouse embryonic fibroblasts (MEFs) by interacting with C/D box small nucleolar RNAs (snoRNAs) and the rRNA 2-prime-O-methyltransferase FBL. Microarray analysis showed that Npm1 deletion in MEFs affected 2-prime-O-methylation sites in 28S rRNA and impaired internal ribosome entry site (IRES) translation through 2-prime-O-methylation regulation. Deletion and overexpression experiments demonstrated that NPM1 also controlled cell growth and differentiation through regulation of 2-prime-O-methylation in K562 human erythroleukemia cells (Nachmani *et al*., 2019).

2.7.2. NPM1/ALK fusion protein

ALK tyrosine kinase expression is normally confined to neural cells, but chromosomal translocations involving ALK and various partners, most frequently NPM1, result in ectopic expression of ALK in a subset of T-cell lymphomas (TCLs) . The NPM1/ALK fusion protein contains the NPM1 oligomerization motif and the ALK catalytic domain, is constitutively activated through autophosphorylation, and mediates malignant cell transformation in vitro and in vivo by activating downstream effectors, including STAT3.

TCL cell lines expressing NPM1/ALK expressed STAT5B, but not STAT5A, protein, whereas normal resting and activated T cells from peripheral blood and ALK-negative TCL cell lines expressed STAT5A protein. Activated NPM1/ALK-positive TCL cell lines also did not express STAT5A mRNA, in spite of having an intact STAT5A gene. Analysis of the CpG island in the STAT5A promoter showed that the region was methylated in NPM1/ALK-positive, but not NPM1/ALK-negative, T cells. Chromatin immunoprecipitation analysis revealed that SP1 bound the STAT5A promoter in normal activated T cells, whereas MECP2 bound the promoter of NPM1/ALK-positive TCL cells.

Demethylation of the promoter resulted in STAT5A activation and inhibition of NPM1/ALK expression by binding of STAT5A to the NPM1/ALK fusion gene. Expression of NPM1/ALK in NPM1/ALK-negative TCL cells resulted in silencing of STAT5A in a STAT3-dependent manner, whereas small interfering RNA mediated-depletion of NPM1/ALK resulted in STAT5A expression. NPM1/ALK induces epigenetic silencing of the STAT5A gene and that the STAT5A protein can act as a tumor suppressor by inhibiting NPM1/ALK expression (Zhang *et al*., 2007).

2.7.3. Biochemical features

The structure of an N-terminal domain of Xenopus nucleoplasmin (Np-core), which is related to NPM1, at 2.3-angstrom resolution. The Np-core monomer is an 8-stranded beta barrel that fits snugly within a stable pentamer. In the crystal, 2 pentamers associate to form a decamer. The authors showed that both Np and Np-core are competent to assemble large complexes that

contain the 4 core histones. These complexes each contain 5 histone octamers that dock to a central Np decamer. (Dutta *et al*., 2001)

2.7.4. Molecular genetics

2.7.4.1. Somatic Mutations

NPM, a nucleocytoplasmic shuttling protein with prominent nucleolar localization, regulates the ARF tumor suppressor pathway. Chromosomal translocations involving the NPM gene cause cytoplasmic dislocation of the NPM protein. Immunohistochemical methods to study the subcellular localization of NPM in bone marrow biopsy specimens from 591 patients with primary acute myelogenous leukemia. They then correlated the presence of cytoplasmic NPM with clinical and biologic features of the disease.

Cytoplasmic NPM was detected in 35.2% of the 591 specimens from patients with primary AML but not in 135 secondary AML specimens or in 980 hematopoietic or extrahematopoietic neoplasms other than AML. It was associated with a wide spectrum of morphologic subtypes of the disease, a normal karyotype, and responsiveness to induction chemotherapy, but not with recurrent genetic abnormalities. There was a high frequency of internal tandem duplications of FLT3 and absence of CD34 and CD133 in AML specimens with a normal karyotype and cytoplasmic dislocation of NPM, but not in those in which the protein was restricted to the nucleus.

AML specimens with cytoplasmic NPM carried mutations in the NPM gene this mutant gene caused cytoplasmic localization of NPM in transfected cells. All 6 NPM mutant proteins showed mutations in at least 1 of the tryptophan residues at positions 288 and 290 and shared the same last 5 amino acid residues. Thus, despite genetic heterogeneity, all NPM gene mutations resulted in a distinct sequence in the NPM protein C terminus. Cytoplasmic NPM is a characteristic feature of a large subgroup of patients with AML who have a normal karyotype, NPM gene mutations, and responsiveness to induction chemotherapy (Falini *et al*., 2005)

NPM staining in cases of AML with aberrant cytoplasmic localization of the protein is mostly cytoplasmic, which suggests that the mutant NPM acts dominantly on the product of the remaining wild-type allele, causing its retention in the cytoplasm by heterodimerization. By

microRNA (miRNA) expression profiling, 36 upregulated and 21 downregulated miRNAs in AML patients with NPM1 mutations compared with AML patients without NPM1 mutations. miR10A and miR10B showed the greatest upregulation, with increases of 20- and 16.67-fold, respectively. Mir22 showed greatest downregulation, with a reduction of 0.31-fold (Grisendi and Pandolfi, 2005). AML with NPM1 mutations has a distinctive miRNA signature (Garzon *et al*., 2008).

The genomes of 200 clinically annotated adult cases of *de novo* AML, using either whole-genome sequencing (50 cases) or whole-exome sequencing (150 cases), along with RNA and miRNA sequencing and DNA methylation analysis. The authors identified recurrent mutations in the NPM1 gene in 54 of 200 (27%) samples (The Cancer Genome Atlas Research Network 2013).

The study of the Cancer Genome Atlas Research Network (2013) did not reveal which mutations occurred in the founding clone, as would be expected for an initiator of disease, and which occurred in minor clones, which subsequently drive disease (Brewin *et al*., 2013). The genes mutated almost exclusively in founding clones in their study included NPM1 (3 of 3 mutations in founding clones). They identified several other genes that contained mutations they considered probable initiators, and other genes in which mutations were considered probably cooperating mutations (Miller *et al*., 2013).

Quantitative RT-PCR assays to detect minimal residual disease in 2,569 samples obtained from 346 patients with NPM1-mutated AML who had undergone intensive treatment in the National Cancer Research Institute AML17 trial. The authors used a custom 51-gene panel to perform targeted sequencing of 223 samples obtained at the time of diagnosis and 49 samples obtained at the time of relapse. Mutations associated with preleukemic clones were tracked by means of digital polymerase chain reaction. Molecular profiling highlighted the complexity of NPM1-mutated AML, with segregation of patients into more than 150 subgroups, thus precluding reliable outcome prediction. The determination of minimal residual disease status was more informative.

Persistence of NPM1-mutated transcripts in blood was present in 15% of the patients after the second chemotherapy cycle and was associated with a greater risk of relapse after 3

years of follow-up than was an absence of such transcripts (82% vs 30%; hazard ratio 4.80; 95% CI 2.95-7.80; p less than 0.001) and a lower rate of survival (24% vs 75%; hazard ratio for death, 4.38; 95% CI 2.57-7.47; p less than 0.001). The presence of minimal residual disease was the only independent prognostic factor for death in multivariate analysis (hazard ratio, 4.84; 95% CI 2.57 to 9.15; p less than 0.001). These results were validated in an independent cohort. On sequential monitoring of minimal residual disease, relapse was reliably predicted by a rising level of NPM1-mutated transcripts. Although mutations associated with preleukemic clones remained detectable during ongoing remission after chemotherapy, NPM1 mutations were detected in 69 of 70 patients at the time of relapse and provided a better marker of disease status (Ivey *et al*., 2016).

2.8. Treatment methods in AML

2.8.1. Induction therapy

Cytarabine has been the backbone of treatment for several decades combined with an anthracycline, usually daunorubicin. The combination of 3 days of daunorubicin and 7 days of cytarabine (3+7) has been accepted as the standard of care for induction treatment. Even after more than 40 years, nothing have convincingly displaced this combination; although, there is a general acceptance that more intensification is feasible and may produce more and better qualities of remission. Several attempts have been made to increase the daily dose and the duration of cytarabine without conclusively improving overall survival (OS). So whether cytarabine is given in a daily dose of 200 mg/m^2 by continuous infusion or twice per day bolus, doubling to 400 mg/m^2, extending to 10 days or escalating to a 3g/m^2, has not made a major impact. Several studies have compared alternatives to daunorubicin, again without convincingly establishing benefit (Burnett *et al*., 2011).

During the past 35 years, a series of studies has established an induction regimen that has become a standard of care for patients not participating on a clinical trial. A widely used combination for induction is the cell cycle–specific agent cytarabine 100 mg/m^2 by continuous infusion for 7 days and the non–cell-cycle–specific anthracycline antibiotic daunorubicin 45 to 60 mg/m^2/d intravenously for 3 days. To improve the Complete Remission (CR) rate, studies have tested alternative and higher doses of anthracyclines or the anthracenediones, higher doses

of cytarabine, new agents combined with cytarabine and daunorubicin such as etoposide, the purine analog fludarabine or the camptothecin topotecan, or sequential standard therapy followed by high doses of cytarabine. Despite theoretic advantages, none of these approaches is definitively better than the standard regimen (Tallman *et al*., 2005).

2.8.2. Postinduction therapy

To judge the efficacy of the induction therapy, a bone marrow aspirate and biopsy should be performed 7 to 10 days after completion of induction therapy. In patients who have received standard-dose cytarabine induction and have residual blasts without hypoplasia, additional therapy with standard-dose cytarabine and anthracycline should be considered. For those with significant residual blasts or clear-cut induction failure, escalation to high-dose cytarabine with or without an anthracycline is the most common salvage strategy. Other options include an allogeneic HSCT if a matched sibling or alternative donor has been identified, or participation in a clinical trial. For patients whose clinical condition has deteriorated such that active treatment is no longer appropriate, best supportive care should be continued. If the marrow is hypoplastic (defined as cellularity < 10%-20% and residual blasts < 5%-10%), additional treatment selection may be deferred until marrow recovery, when the remission status can be assessed.

Patients initially treated with high-dose cytarabine and who have significant residual blasts 7 to 10 days after completion of induction chemotherapy are considered to have experienced induction failure. These patients should be considered for a clinical trial, allogeneic HSCT with matched sibling or matched unrelated donor, or best supportive care. Additional high-dose cytarabine at this time is unlikely to induce remission in these cases. If an HLA-matched sibling or matched unrelated donor has been identified, an allogeneic HSCT may salvage 25% to 30% of patients with induction failure. If no donor is immediately available, patients should be considered for a clinical trial. Again, if the patient's clinical condition has deteriorated to a point at which active therapy would be detrimental, best supportive care may be the most appropriate option (Donnell *et al*., 2012).

2.8.3. Remission induction therapy

As with younger AML patients, the backbone of remission induction in older adults consists of an anthracycline (daunorubicin or idarubicin) or anthracenedione (mitoxantrone) and cytosine arabinoside (Ara-C), a regimen that has not changed in over two decades. Typically, daunorubicin is given at a dose of 45 mg/m^2/d × 3 days, or mitoxantrone or idarubicin are given at doses of 12 mg/m^2/d × 3 days, in combination with Ara-C, which is administered as a continuous infusion at 100 or 200 mg/m^2/d × 7 days (frequently referred to as 7+3 chemotherapy).

While studies have compared different anthracyclines and anthracenediones, varied doses and schedules, and added additional agents with some improvement in CR rates, they have not demonstrated an improvement in OS rates. For example, a recent study from the ECOG randomized older AML patients to remission induction therapy with either daunorubicin, idarubicin, or mitoxantrone along with a standard dose of Ara-C. The outcome was not significantly different, with CR rates of 40%, 43%, and 46% and median survivals of 7.7, 7.5, and 7.2 months, respectively. Once a decision has been made to initiate intensive chemotherapy, it should not be delayed, as this may impact outcome (Stone *et al*., 2004).

2.8.4. Post remission therapy

In recent years, an increasing number of studies of postremission therapy have been undertaken in an attempt to learn how to prevent leukemic relapse. These studies have often given apparently conflicting results, primarily because remission duration is dependent on multiple factors, including the type of leukemia, host factors, and the nature of induction therapy, as well as the post remission treatment. To evaluate various studies, post remission treatment can be divided into three types.

Maintenance therapy may be defined as post-remission treatment that is less myelosuppressive and intensive than induction. Consolidation may be used to describe regimens that are essentially the same as those administered for induction, both in terms of drugs and doses used. Intensification may be defined as therapy with the same drugs used in induction at significantly higher doses or with myelosuppressive doses of active agents other than those

administered in induction. Recent studies that use standard daunorubicin-cytarabine type induction treatment allow assessment of the current role of post remission therapy. With new induction regimens, results of current trials may not be applicable (Bloomfield, 1985).

Of course, post-remission therapy in AML is predicated on the success of induction therapy, and advances in remission induction treatment strategies for AML have led to modest but significant improvements in achievement of CR and OS. Age and comorbidities have played an important role in selection of induction regimens, with hypomethylating agents, CPX-351, lomustine, glasdegib, and venetoclax recently serving as important adjuncts that result in improved CR, disease-free survival (DFS), and OS rates in older adults with AML. Though the influence of age on post-remission strategies is an underexplored area, cytogenetics has played a key role in tailoring post-remission therapy. Patients with Core Binding Factor (CBF) AML have typically been treated with post-remission cytarabine while forgoing allogeneic hematopoietic stem cell transplantation (HSCT); those with intermediate or adverse-risk cytogenetics have been treated with intensive multi-agent chemotherapy followed, if feasible, by allogeneic HSCT while in CR1 (Derman and Larson, 2019).

2.8.5. Hematopoietic stem cell transplantation

Allogeneic hematopoietic cell transplantation (HCT) is an effective post-remission consolidation treatment, potentially curative, in patients with AML. Since the first report of a successful bone marrow transplant in 1957, there has been a steady increase in the numbers of patients receiving HCT for AML. Worldwide, over a third of HCTs are performed as therapy for AML, more than any other diagnosis, while autologous HCT for AML accounts for less than 3% of activity. Recent years have witnessed the important role of molecular markers in the management of AML. HCT is curative for many patients with AML and assessment of the potential benefit to an individual patient needs to start at diagnosis of AML so that HCT outcome is not compromised by undue delay (Kassim and Savani, 2017).

2.8.6. Autologous stem cell transplantation

Autologous stem cell transplantation is nowadays a recognized therapeutic option for AML therapy. Numerous retrospective data exist from single institutions and from international

registries on the use of conventional chemotherapy alone or allogeneic or autologous blood or marrow stem cell transplantation (ABMT). Recently, results of randomized studies comparing chemotherapy alone to allogeneic and autologous stem cell transplantation have become available. Prospective randomized studies have shown its superiority over conventional chemotherapy, with the possible reservation that none has tested high-dose ARAC in a parallel arm.

As it presently stands, indications of ABMT for many teams still reside largely in impossibilities of allogeneic transplants and concern primarily good- or standard-risk AML in older patients with no family donor. An important randomized international study evaluating the source of stem cells and possibly purging with the help of the NOD-SCID mouse model for evaluation of minimal residual disease is needed. A foreseen reduction in transplant-related mortality with PB and a hopefully successful search for reduction in relapse incidence with immunomodulation or leukemia vaccines would be the keys for ABMT to become the first rather than the second therapeutic option (Gorin, 1998).

2.9. Targeted therapy in AML

Recent advances in molecular technology and a better understanding of the biology of AML have led to the identification of novel molecular markers in AML. These markers may play a role in defining prognosis, developing targeted therapies and detecting minimal residual disease in patients with AML. A number of molecular markers have been identified, including FMS like tyrosine kinase 3 (FLT3) mutations, RAS/RAF/MEK/ERK pathway (RAS) mutations and mammalian Janus kinase (JAK2) mutations (Daver and Cortes, 2013).

2.9.1. FLT3 Inhibitors

FLT3 is a transmembrane ligand-activated receptor tyrosine kinase (RTK) which plays an important role in the early stages of both myeloid and lymphoid lineage development. FLT3 ligand binds and activates FLT3 through various signaling pathways, such as PI3K, RAS, and STAT5. FLT3 mutations are found in approximately 30–35% of newly diagnosed AML cases with either internal tandem duplications (FLT3-ITD) within the juxtamembrane domain coding region (exons 14 and 15, or missense mutations in the tyrosine kinase domain (FLT3-TKD) in

the activation loop (exon20). FLT3-ITD and FLT3-TKD type mutations occur in about 25% and 7–10% of AML patients, respectively. FLT3-ITD mutation had been considered as a negative prognostic marker, used for AML risk stratification and disease monitoring via MRD, with the clinical importance of early detection at diagnosis and again at relapse (Yu *et al*., 2020).

Patients with mutations in FLT3 have a worse prognosis when treated with conventional chemotherapy, compared to patients with wild-type FLT. Initial studies using small molecule FLT3 inhibitors have offered encouragement that a more selective therapy may improve the outcome of patients. In addition, when FLT3 inhibitors are added to the conventional arsenal of AML treatments patients with AML expressing mutated FLT3 may experience significant clinical benefit (Daver and Cortes, 2013).

2.9.2. Midostaurin

Midostaurin was approved by the US FDA (Food and Drug Administration) for AML induction and consolidation based on the RATIFY trial, which took 13 years to complete (Stein *et al*., 2017). The RATIFY trial was the first large multicenter study investigating the addition of Midostaurin to induction and consolidation and continued as maintenance therapy for 1 year in patients not proceeding to allogeneic transplant (Stone *et al*., 2017 ; Daver *et al*., 2019 ; Stone *et al*., 2018). Patients with FLT3 mutations, either ITD or TKD, had a 4-year survival of 51.4% on Midostaurin versus 44.2% on placebo ($P = .0074$), and the benefit was most pronounced in Nucleophosmin 1 (NPM1) wt and FLT3 patients. In another study, Midostaurin was added to intensive induction chemotherapy, consolidation and continued as maintenance in FLT3-ITD AML patients CR plus complete remission with incomplete hematologic recovery (CRi) after induction therapy was observed in 76.4% patients. Event free survival (EFS) and OS at 2 years were 39 and 34% in younger and 53 and 46% in older patients, respectively. Propensity score-weighted analysis revealed a significant improvement of EFS by Midostaurin overall and in older patients (Schlenk *et al*., 2019).

In a new retrospective exploratory study, multivariate Cox model for OS using allogeneic hematopoietic stem cell transplantation (allo-HSCT) in first complete remission (CR1) as a time-dependent variable revealed treatment with Midostaurin, allo-HSCT, European Leukemia Net (ELN) favorable-risk group, and lower WBC counts as significant favorable factors. There was a

consistent beneficial effect of Midostaurin across ELN risk groups (Dohner *et al*., 2020). Midostaurin has been recommended as frontline therapy for the FLT3 gene mutated AML patients with either FLT3-ITD or FLT3-TKD. It's also been proved cost-effectiveness when Midostaurin was combined with standard chemotherapy in the treatment of newly diagnosed FLT3-mutated AML patients (Shimada, 2019 ; Antar *et al*., 2020 ; Tremblay *et al*., 2020). Midostaurin is among the least potent FLT3 inhibitors. More potent FLT3 inhibitors are Gilteritinib, Quizartinib, and Crenolanib. Early phase trials combining these newer generation FLT3 TKIs with 7 + 3 induction chemotherapy in the frontline setting have been reported recently with meaningfully higher response rate (Larrosa and Garcia, 2017).

2.9.3. Quizartinib

Quizartinib (AC220) is a novel second-generation class 3 RTK inhibitor with potent FLT3 activity *in vitro* and *in vivo*. In addition to FLT3, quizartinib inhibits c-KIT, platelet derived growth factor, and RET. AC220 has been explored in a phase 1 trial in 76-relapsed/refractory or untreated, elderly AML patients. Quizartinib had significant clinical activity, inducing a reduction in blast counts, including full CR in some patients. Ten of 76 (13%) patients had a CR (2 CR, 6 CR with incomplete blood count recovery, and two CR with incomplete platelet recovery) and 13 (17%) had a partial remission (PR). The median duration of response was 14 weeks, with duration of responses up to 67+ weeks. Higher overall response rates and CR rates were observed in FLT3-ITD patients (56 and 28%, respectively) compared with those lacking the mutation (20 and 7%, respectively). The most commonly reported adverse events that were possibly drug-related were reversible prolongation of QTc; others were mainly grade 2 and included peripheral edema, dysgeusia, and nausea. Phase 2/3 studies combining quizartinib with standard chemotherapy agents are currently ongoing (Zarrinkar *et al*., 2009 ; Cortes *et al*., 2009).

2.9.4. RAS

The RAS is critical for proliferation of many human cancers. Normally, this pathway is activated by the binding of extracellular growth factors to membrane receptors which leads to activation of the small guanosine triphosphate-binding protein RAS. As a result, RAS adopts an

activated conformation, which stimulates downstream signaling. One of the critical downstream proteins is RAF, which phosphorylates MEK. Activated MEK then phosphorylates residues on the MAP kinases: ERK1 and ERK2. Phosphorylated ERK dimerizes and translocates to the nucleus, where it is involved in cellular proliferation, nuclear transport, DNA repair, nucleosome assembly, mRNA processing, and translation. Mutated oncogenic forms of RAS are found in approximately 15% of all cancer including approximately 30% of AML and acute lymphocytic leukemia. Because of its central role in the ERK pathway, MEK is thought to be an important therapeutic target in AML (Milella *et al*., 2005 ; Davies *et al*., 2002).

Final and preliminary clinical data are available for three MEK inhibitors in solid tumor phase 1/2 trials (CI-1040, PD-0325901, and AZD6244). To date, these agents have been dosed in a total of approximately 500 patients. In total, there have been 12 PRs and one CR. The most common toxicities in the completed trials include rash, diarrhea, nausea, vomiting, peripheral edema, and fatigue. More concerning are reports of optical disturbances, including transient blurred vision, RVO, or optic neuropathy. A decrease in cardiac ejection fraction and transaminitis have also been reported (Friday and Adjei, 2008).

2.9.5. JAKs

The mammalian JAKs protein family consists of four cytoplasmic tyrosine kinases (JAK1, JAK2, JAK3, and TYK2) that play a role in hematopoiesis. Upon activation by an associated receptor, JAKs phosphorylate cytoplasmic signal transducer and activator of transcriptions (STATs), resulting in altered expression of target genes. Aberrant activation of JAKs has been associated with increased malignant cell proliferation and survival in patients with Philadelphia chromosome negative myeloproliferative disease. Although JAK2 mutations are extremely rare in AML, STAT3 and/or STAT5 are activated in a majority of AML samples. This may occur via mutations in costimulatory molecules. For example, it has been shown that mutations in FLT3 domain can constitutively activate this receptor kinase without ligand binding, resulting in the activation of downstream prosurvival signals including JAK/STAT5. Thus, there may be a role for JAK/STAT5 inhibitors in AML (James *et al*., 2005 ; Steensma *et al*., 2006).

Ruxolitinib is a JAK2 inhibitor that has been extensively studied in patients with myeloproliferative neoplasms. In a phase 2 study of ruxolitinib in patients with relapsed/refractory leukemias, 18 patients with relapsed and refractory leukemias nine *de novo* AML, three secondary AML (sAML), two acute lymphocytic leukemia, one MDS, two chronic myelomonocytic leukemia (CMML), and one chronic myeloid leukemia] were treated with ruxolitinib. Five patients (one with AML, two with sAML, and three with CMML) had the JAK2 V617F mutation. Three patients (including two with sAML and one with CMML, all with JAK2 mutation) had significant declines in their bone marrow blasts (to <5%) and clinical improvement (Ravandi *et al*., 2009).

Other drugs acting on the JAK2 pathway include AZ 960 (potent and selective ATP competitive inhibitor of the JAK2 kinase), Ki11502 (a novel multitargeted RTK inhibitor), CYT387, SAR302503 (formerly TG101348, a novel JAK2-FLT3 inhibitor), and SB1518 (also JAK2-FLT3 inhibitor). These are currently in clinical trials (Daver and Cortes, 2013).

2.9.6. Monoclonal antibodies and immunoconjugates

Conventional chemotherapeutic treatments of acute leukemias are often life threatening due to a lack of specificity for hematopoietic cells. It is possible that approaches that specifically target leukemic blast cells might be safer, and possibly more effective, than the use of current nonspecific chemotherapeutic agents.

Leukemia-specific antigens are uncommon, however. Therefore most monoclonal antibody targeting approaches have been directed against normal hematopoietic cell surface antigens that are also expressed by leukemic blast cells. Nonspecific binding to most normal tissues is typically avoided with this approach. Both unconjugated and conjugated antibodies have shown promise in early clinical trials, and may represent appealing therapeutic alternatives for patients with AML (Sievers, 2000).

2.10. System biology approach

AML is a hematological cancer comprising of cancer stem cells (CSCs) that are responsible for the disease progression, drug resistance and post treatment relapses. Advances in genomic technologies have identified AML as a genetically heterogenous disease with dysregulated gene expression networks. Furthermore, observation of intracellular signaling in

individual CSCs by mass cytometry has demonstrated the dysregulation of the mitogen associated protein kinase (MAPK) pathways. It has been envisaged that the future treatment for AML would entail upon formulating individualized treatment plans leading to decreased drug related toxicities for patients. However the emerging role of signaling pathways as dynamic molecular switches influencing the cell cycle process, thereby leading to varying stages of cell differentiation, is making community rethink about the current strategies used for the treatment of AML.

This kind of systems biology approach to resolve such a complex problem would ultimately lead us to some kind of hypotheses that could be tested on some other cell types to see whether the features that are being observed in selected cellular systems are universal or very specific related to a cell type. In future, such systems level understanding of the R-point control by signaling pathways would be significant in employing novel therapeutic strategies for treating AML patients (Vaidya, 2014).

MATERIALS AND METHODS

3

MATERIALS AND METHODS

The materials used and methods adopted in the present study entitled "Computational identification of key genes and pathways associated with Acute Myeloid Leukemia" is furnished below.

3.1. RNA – sequencing data

RNA-Seq (named as an abbreviation of RNA sequencing) is a sequencing technique which uses next-generation sequencing (NGS) to reveal the presence and quantity of RNA in a biological sample at a given moment, analyzing the continuously changing cellular transcriptome. RNA-Seq data analyses typically consist of (1) accurate mapping of millions of short sequencing reads to a reference genome, including the identification of splicing events; (2) quantifying expression levels of genes, transcripts, and exons; (3) differential analysis of gene expression among different biological conditions; and (4) biological interpretation of differentially expressed genes (Zhao *et al*., 2016).

An RNA-Seq dataset of adult AML, which included human AML patients whole transcriptome sequencing dataset and corresponding survival profiles, was download from The Cancer Genome Atlas (TCGA) database (https://gdc-portal.nci.nih.gov/) (Cancer Genome Atlas Research Network, 2013) to compare the genes and mRNA expression between NPM1 mutation and wild-type adult AML patients.

3.2. Identification of differentially expressed genes (DEGs)

It is expected that emerging digital gene expression (DGE) technologies will overtake microarray technologies in the near future for many functional genomics applications. EdgeR is a Bioconductor software package for examining differential expression of replicated count data. The schematic representation of applied methods and the resulted outcomes are shown in Figure.2.

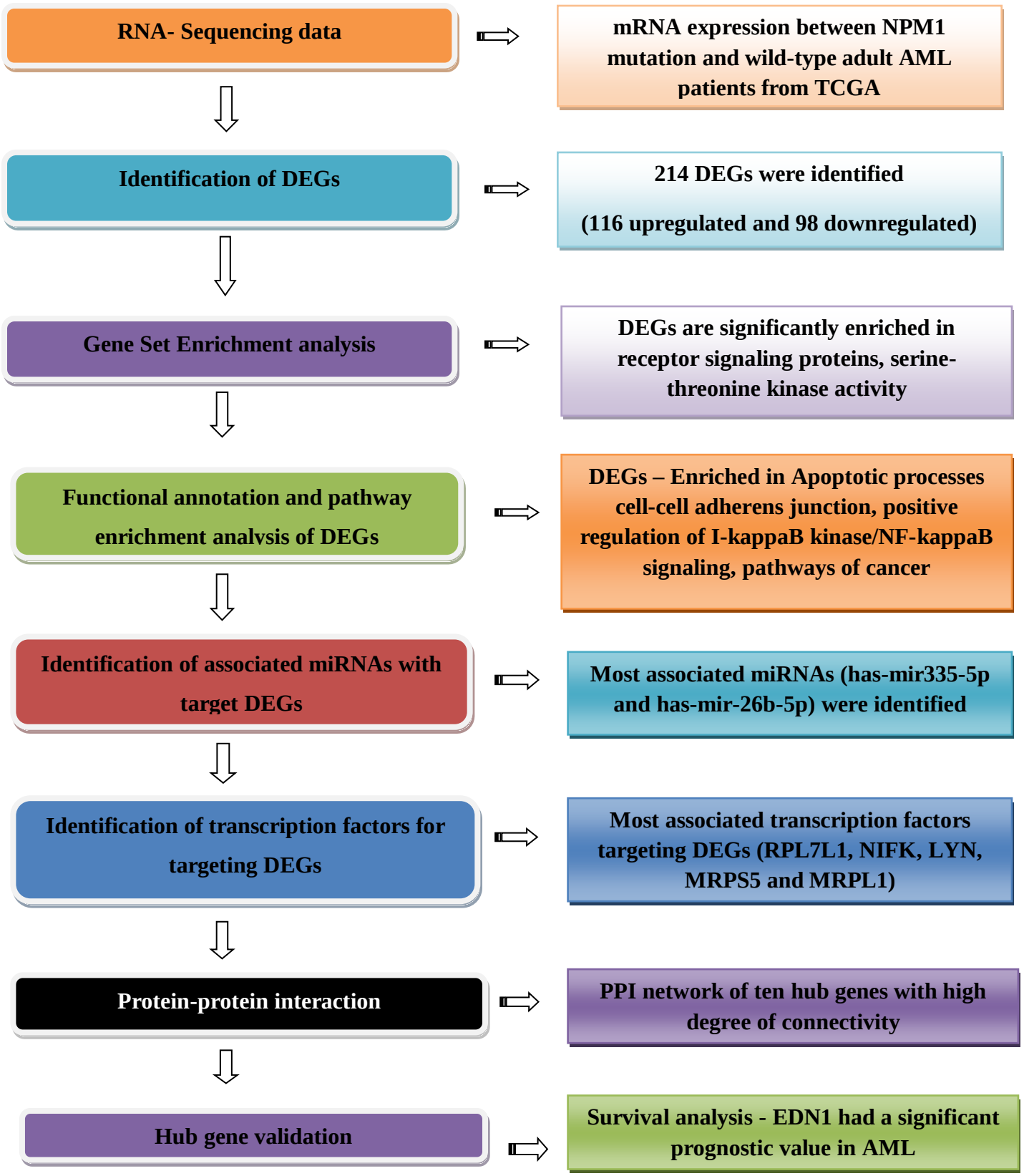


Fig.2. Gene level differential expression analysis of workflow

EdgeR was used to screen DEGs between NPM1 mutation and wild-type AML patients according to the user's guide (Robinson *et al*., 2010). DEGs were identified with the cut-off value of log2|fold change (FC)| ≥ 1 and P-Value 0.05. A heat map and volcano plot of the DEGs were drawn via the "ggplot2" package in the R platform.

EdgeR may also be useful in other experiments that generate counts, such as ChIP-seq, in proteomics experiments where spectral counts are used to summarize the peptide abundance (Wong *et al., 2008)* or in barcoding experiments where several species are counted (Andersson *et al.*, 2008). The software is designed for finding changes between two or more groups when at least one of the groups has replicated measurements.

3.3. Gene set enrichment analysis (GSEA)

GSEA (https://www.gsea-msigdb.org) is a powerful tool for the interpretation of high-throughput expression studies such as mass spectrometry-based proteomics or Next-Generation Sequencing, in order to identify insights into biological processes or pathways underlying a given phenotype (Zito *et al*., 2021). GSEA considers experiments with genome wide expression profiles from samples belonging to two classes, labeled 1 or 2. Genes are ranked based on the correlation between their expression and the class distinction by using any suitable metric (Subramanian *et al*., 2005). There are three key elements of the GSEA method:

- Step 1: Calculation of an Enrichment Score.
- Step 2: Estimation of Significance Level of ES.
- Step 3: Adjustment for Multiple Hypothesis Testing.

To investigate the effect of NPM1 mutations on various biological function gene sets in adult AML patients, differences in gene mRNA expression levels of biological functional annotation and pathways between NPM1 mutation and wild-type patients were analyzed by GSEA v2-2.2.3 (http://software.broadinstitute.org/gsea/downloads.jsp). Reference gene sets from the Molecular Signatures Database (MSigDB) of c2 (c2.cp.kegg.v5.2.symbols.gmt) and c5 (c5.bp.v5.2.symbols.gmt; c5.mf.v5.2.symbols.gmt; c5.cc.v5.2.symbols.gmt; consist of genes annotated by the same GO terms) respectively. The MSigDB of c2 is a pathway gene set, which was curated from publications and extracted from canonical pathways and experimental

signatures, whereas the MSigDB of c5 was constructed on genes annotated by the same GO terms. The number of permutations was set at 1,000. Enrichment results satisfying a nominal P-value cutoff of, 0.05 with a false discovery rate (FDR), 0.25 were considered statistically significant.

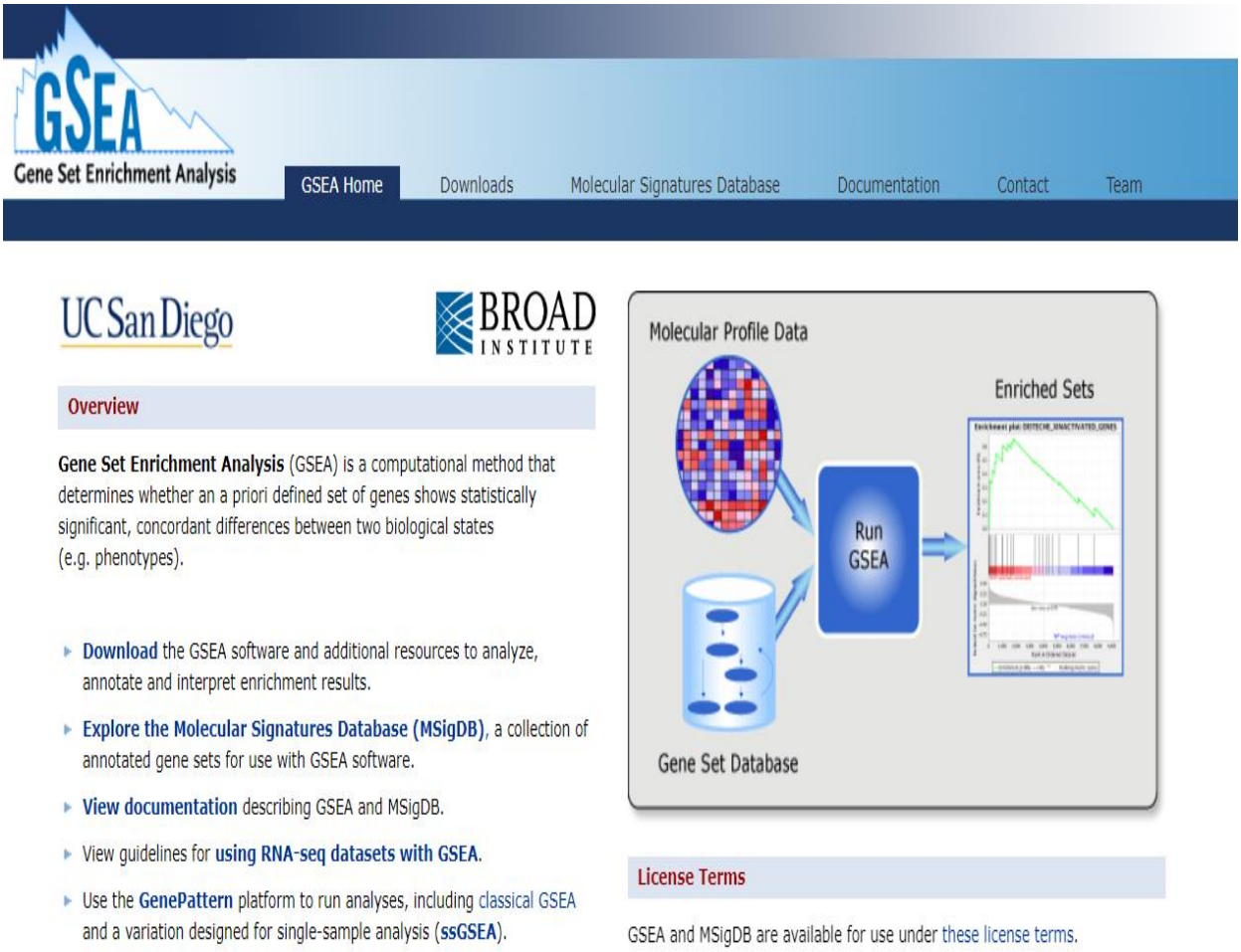


Fig.3. The homepage for GSEA

3.4. Functional annotation and pathway enrichment analysis of DEGs

In the post-genomic era, biological interpretation of large gene lists derived from high-throughput experiments, such as genes from microarray experiments, is a challenging task. The first version of DAVID (the Database for Annotation, Visualization and Integration Discovery), released in 2003 as well as a number of other similar publicly available high-throughput functional annotation tools, partially address the challenge by systematically mapping a large

number of interesting genes in a list to associated GO terms, and then statistically highlighting the most over-represented GO terms out of a list of hundreds or thousands of terms. This increases the likelihood that the investigator will identify the biological processes most pertinent to the biological phenomena under study. While this tool is extremely useful and has been cited in hundreds of publications during the past three years, the development of other effective data mining algorithms, as additional components to the DAVID Bioinformatics Resources, will improve the power of investigators to analyze their gene lists from different biological angles.

Fig.4. The homepage for Database for Annotation, Visualization and Integrated Discovery (DAVID)

The DAVID Gene Functional Classification Tool (http://david.abcc.ncifcrf.gov) uses a novel agglomeration algorithm to condense a list of genes or associated biological terms into organized classes of related genes or biology, called biological modules. This organization is accomplished by mining the complex biological co-occurrences found in multiple sources of functional annotation. It is a powerful method to group functionally related genes and terms into a manageable number of biological modules for efficient interpretation of gene lists in a network context (Huang *et al*., 2007).

DAVID assists in the interpretation of genome-scale datasets by facilitating the transition from data collection to biological meaning. DAVID is composed of four main modules: Annotation Tool, GO (Gene Ontology) charts and KEGG Kyoto Encyclopedia of Genes and Genomes charts. The Annotation Tool is an automated method for the functional annotation of gene lists. Any combination of annotation data can be chosen from 10 options by selecting the appropriate checkboxes. The annotations are added to the submitted gene list by selecting the upload button, which returns an Hypertext Markup Language (HTML) table containing the user's original list of identifiers appended with the chosen functional annotations. Unannotated genes are included in the output with no appended data for tracking purposes (Dennis *et al*., 2003).

The GOCharts module graphically displays the distribution of differentially expressed genes among functional categories using the controlled vocabulary of the GO Consortium, which provides a structured language that can be applied to the functions of genes and proteins in all organisms even as knowledge continues to accumulate and change. The structure of GO starts with three main categories, biological process, molecular function, and cellular component. Biological Process includes broad biological goals, such as mitosis or purine metabolism, that are accomplished by ordered assemblies of molecular functions.

Classification data is displayed as a bar chart, where the length of the bar represents the number of gene identifiers in each category. The user can set visualization parameters for sorting output data and displaying categories that contain at least a minimum number of genes. Selecting an individual bar opens a new HTML table displaying the gene identifier, LocusLink number, gene name, the current classification, and other classifications for each gene in that category. A 'Show All' button opens a new HTML table displaying all classification data and a 'Show Chart Data' button opens an HTML table containing the underlying chart data, thus allowing users to

recreate customized chart graphics in a spreadsheet program. A new chart can be displayed for any subset of genes by selecting the classification type and level using the checkboxes and radio buttons available within the user's current page that allow for drill-down capabilities. A count of the number of genes annotated is included in the output, and unannotated genes are binned into the 'unclassified' category, thus providing users with an automated tracking system for genes not annotated (Ashburner, 2000).

KEGG Charts graphically display the distribution of differentially expressed genes among KEGG biochemical pathways. Each pathway is linked to the KEGG pathway map, wherein differentially expressed genes from the original list are highlighted in red. In this view genes are further linked to additional annotations available through KEGG's Integrated database retrieval system (DBGET). As with GoCharts, the user can set visualization parameters for sorting output data and displaying categories that contain at least a minimum number of genes and the KEGG Charts visualization inherits all of the dynamic features of GoCharts (Kanehisa and Goto, 2000).

DAVID is used to annotate input genes, classify gene functions, identify gene conversions, and carry out GO term analysis (Jiao *et al*., 2012). To identify the DEG's functional annotation, we analyzed GO terms and KEGG pathway enrichment with DAVID, while specifying a P-value < 0.05 for statistical significance.

3.5. Identification of associated miRNAs with target DEGs

miRNet (https:// www.mirnet.ca/) is an easy-to-use, web-based platform designed to help elucidate miRNA functions by integrating users' data with existing knowledge via network-based visual analytics. Since its first release in 2016, miRNet has been accessed by >20, 000 researchers worldwide, with ~100 users on a daily basis. While version 1.0 was focused primarily on miRNA-target gene interactions, it has become clear that in order to obtain a global view of miRNA functions, it is necessary to bring other important players into the context during analysis (Chang *et al*., 2020).

By integrating multiple high-quality miRNA-target data sources and advanced statistical methods into a powerful network visualization system, miRNet allows researchers to easily navigate the complex landscape of miRNA–target interactions to obtain deep biological insights

(Fan and Xia, 2018). The associated miRNAs of differentially expressed target genes were collected and also graphically visualized through miRNet.

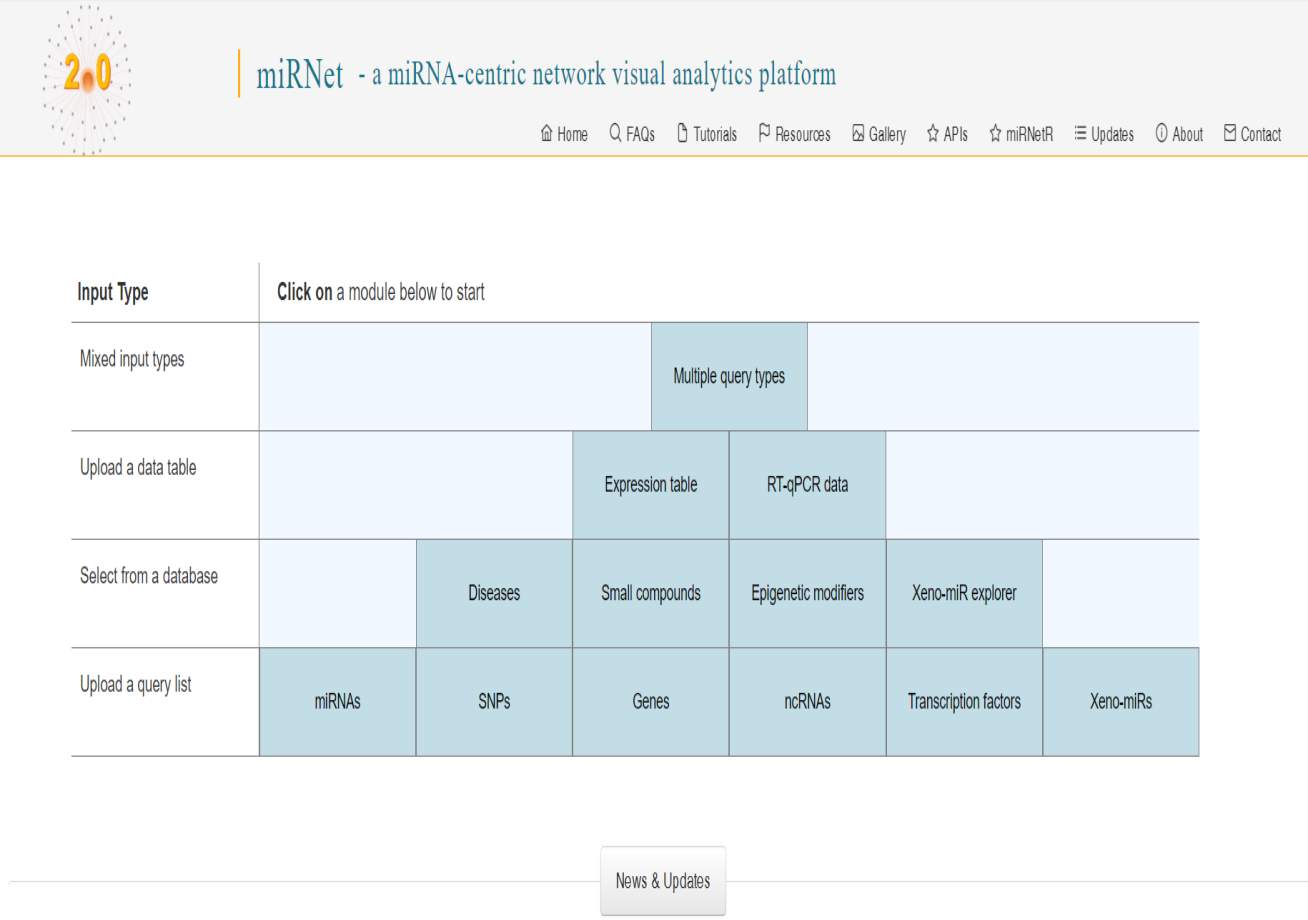


Fig.5. The homepage for miRNet

3.6. Identification of transcription factors for targeting DEGs

NetworkAnalyst (www.networkanalyst.ca) was first released in 2014 to address the key need for interpreting gene expression data within the context of Transcription Factor (TF)-gene interaction networks. It was soon updated for gene expression meta-analysis with improved workflow and performance. Over the years, NetworkAnalyst has been continuously updated based on community feedback and technology progresses (Zhou *et al*., 2019).

To predict the transcription factors, the functional enrichment analysis tool was performed, which can reveal the enriched transcription factors targeting significant DEGs.

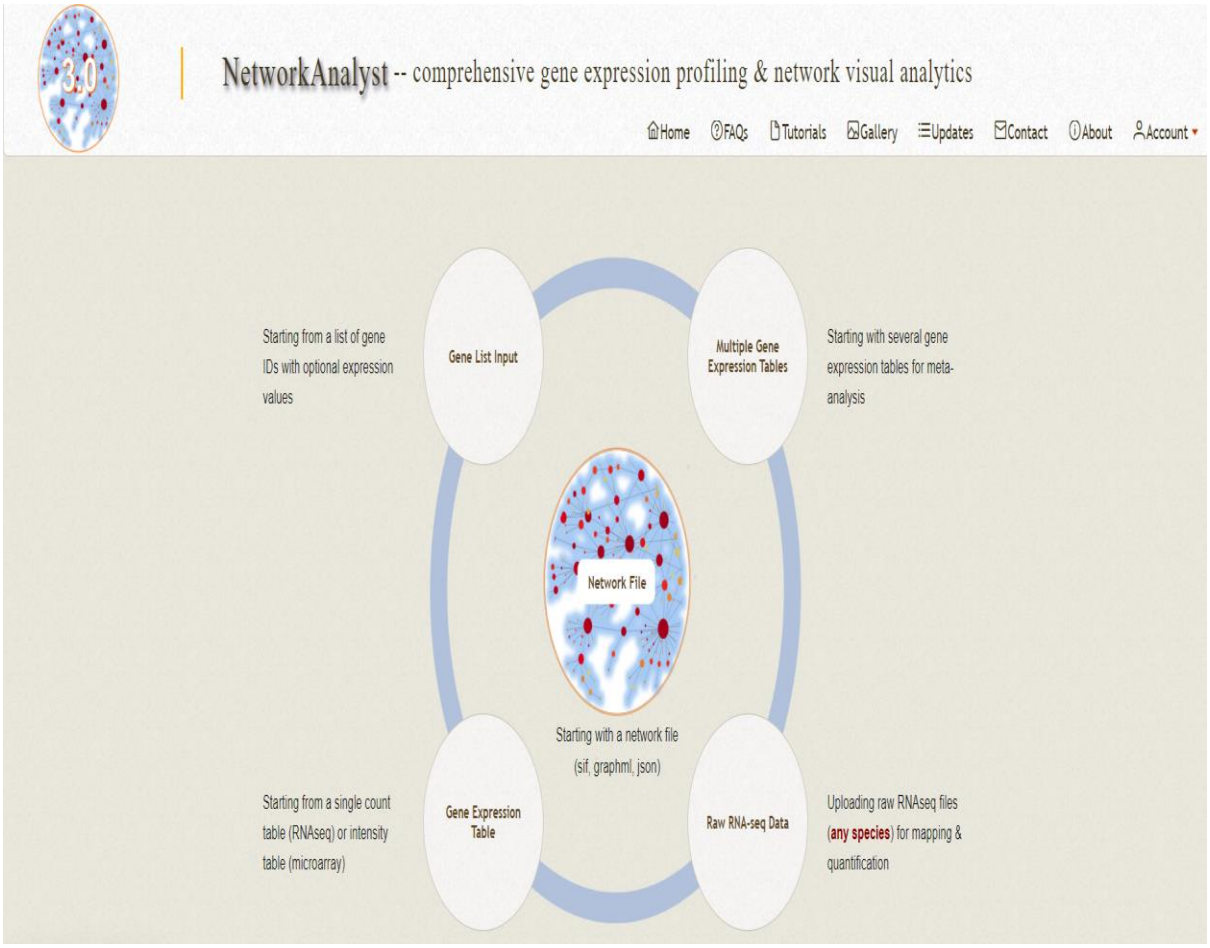


Fig.6. The homepage for Network Analyst

3.7. Integration of protein-protein interaction (PPI)

Protein-protein interaction plays key role in predicting the protein function of target protein and drug ability of molecules. The majority of genes and proteins realize resulting phenotype functions as a set of interactions. The in vitro and in vivo methods like affinity purification, Y2H (yeast 2 hybrid), TAP (tandem affinity purification), and so forth have their own limitations like cost, time, and so forth, and the resultant data sets are noisy and have more false positives to annotate the function of drug molecules.

Thus, *in silico* methods which include sequence-based approaches, structure-based approaches, chromosome proximity, gene fusion, in silico 2 hybrid, phylogenetic tree,

phylogenetic profile, and gene expression-based approaches were developed. Elucidation of protein interaction networks also contributes greatly to the analysis of signal transduction pathways. Recent developments have also led to the construction of networks having all the protein-protein interactions using computational methods for signaling pathways and protein complex identification in specific diseases (Rao *et al*., 2014).

The database Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is a precomputed global resource for the exploration and analysis of these associations. Since the three types of evidence differ conceptually, and the number of predicted interactions is very large, it is essential to be able to assess and compare the significance of individual predictions. Thus, STRING contains a unique scoring-framework based on benchmarks of the different types of associations against a common reference set, integrated in a single confidence score per prediction. The graphical representation of the network of inferred, weighted protein interactions provides a high-level view of functional linkage, facilitating the analysis of modularity in biological processes. STRING is updated continuously, and currently contains 261 033 orthologs in 89 fully sequenced genomes (Mering *et al*., 2003).

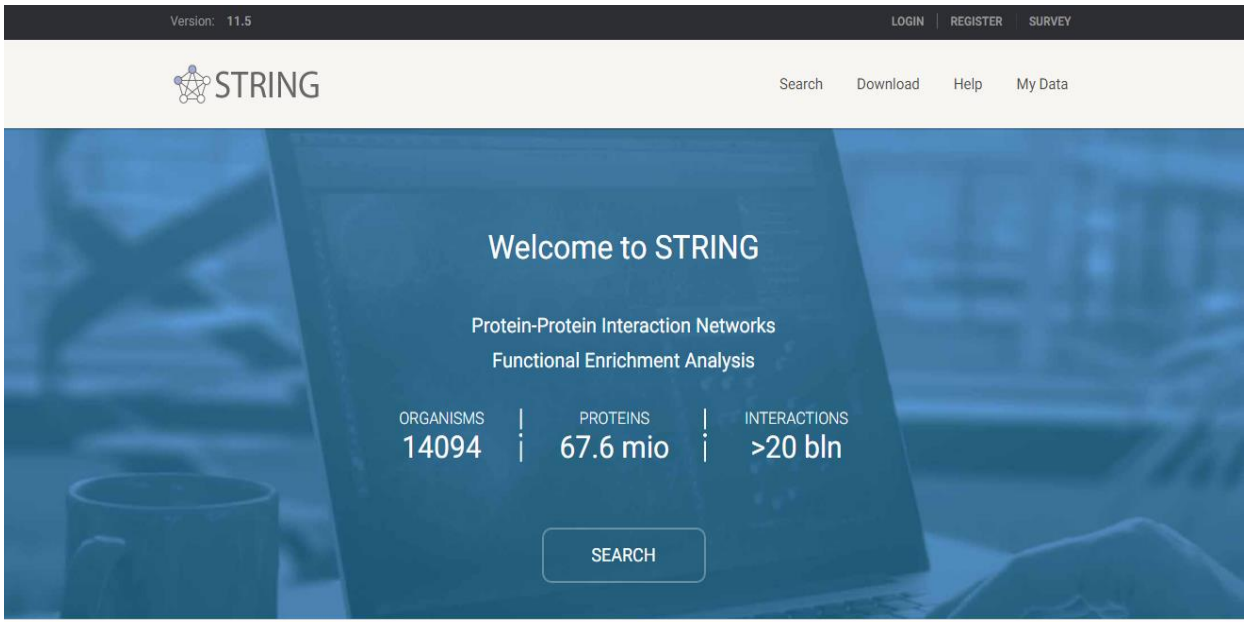


Fig.7. The homepage for STRING

The STRING database (http://string.embl.de/) is an online tool designed to evaluate the PPI information (Szklarczyk *et al*., 2015). To evaluate the PPI relationships among DEGs, we mapped the DEGs to STRING to evaluate the validated interactive relationships among DEGs. Experimentally validated interactions with a combined score > 0.4 were selected as significant.

Using the PPI networks, the cytoHubba plugin in Cytoscape (Shannon *et al*., 2003) was performed to identify hub genes of the PPI network with defaults.

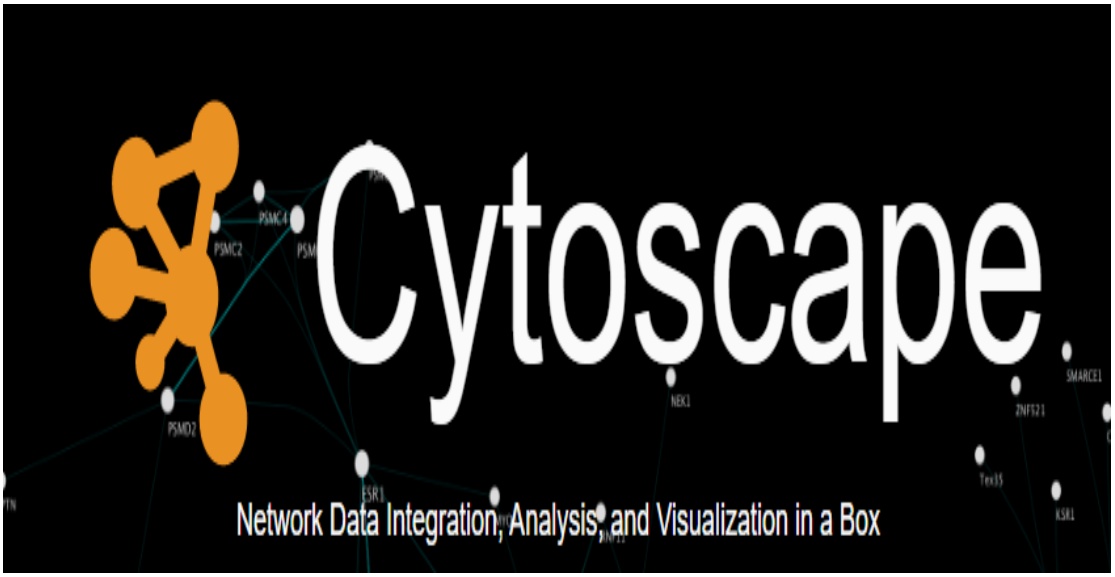


Fig.8. Cytoscape: A software environment for integrated models of bio-molecular interaction networks

Metascape (https://metascape.org) is a web-based portal designed to provide a comprehensive gene list annotation and analysis resource for experimental biologists. In terms of design features, Metascape combines functional enrichment, interactome analysis, gene annotation, and membership search to leverage over 40 independent knowledge bases within one integrated portal. Metascape provides a significantly simplified user experience through a one-click Express Analysis interface to generate interpretable outputs. Taken together, Metascape is an effective and efficient tool for experimental biologists to comprehensively analyze and interpret OMICs-based studies in the big data era. GO enrichment terms of hub genes were analyzed by Metascape (Zhou *et al*., 2019).

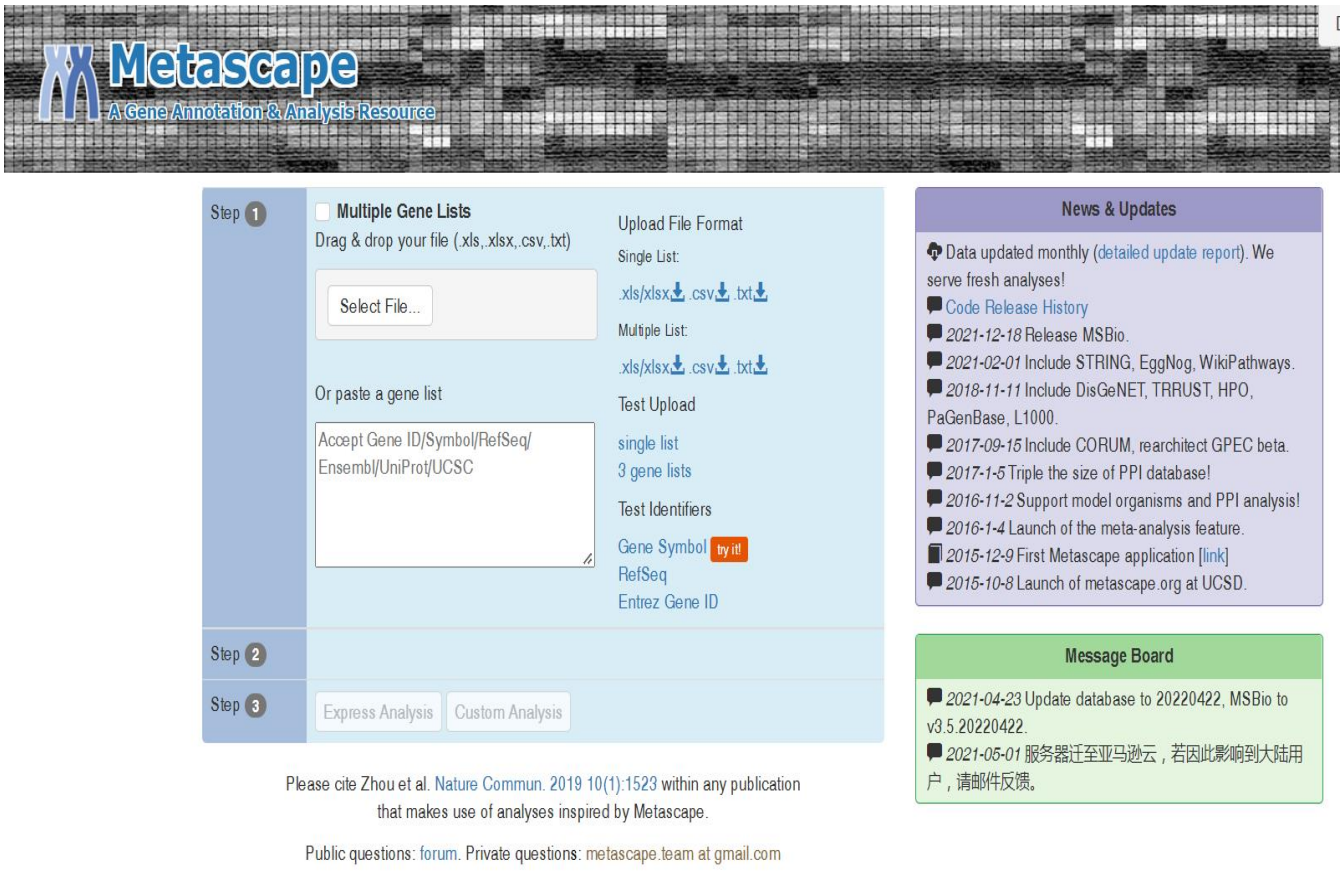


Fig.9. Metascape: A biology oriented resource for the analysis of systems level datasets

3.8. Hub gene validation

Validation of genes is accomplished by UALCAN (http://ualcan,path.uab.edu/) using TCGA RNA-seq data. UALCAN uses TCGA RNA-sequencing and patients clinical data from 33 different cancer types, including several metastatic tumors (Chandrashekar *et al.*, 2017). The web-based platforms user friendly features facilitate:

- relative expression analysis of a query genes across tumor and normal samples, as well as in various tumor sub-groups based on individual cancer stages, tumor grade, race, body weight or other clinico-pathologic features.
- understanding the combined impact of gene expression level and clinico-pathologic features on patient survival.

- identification of the top over- and under-expressed genes in individual cancer types.

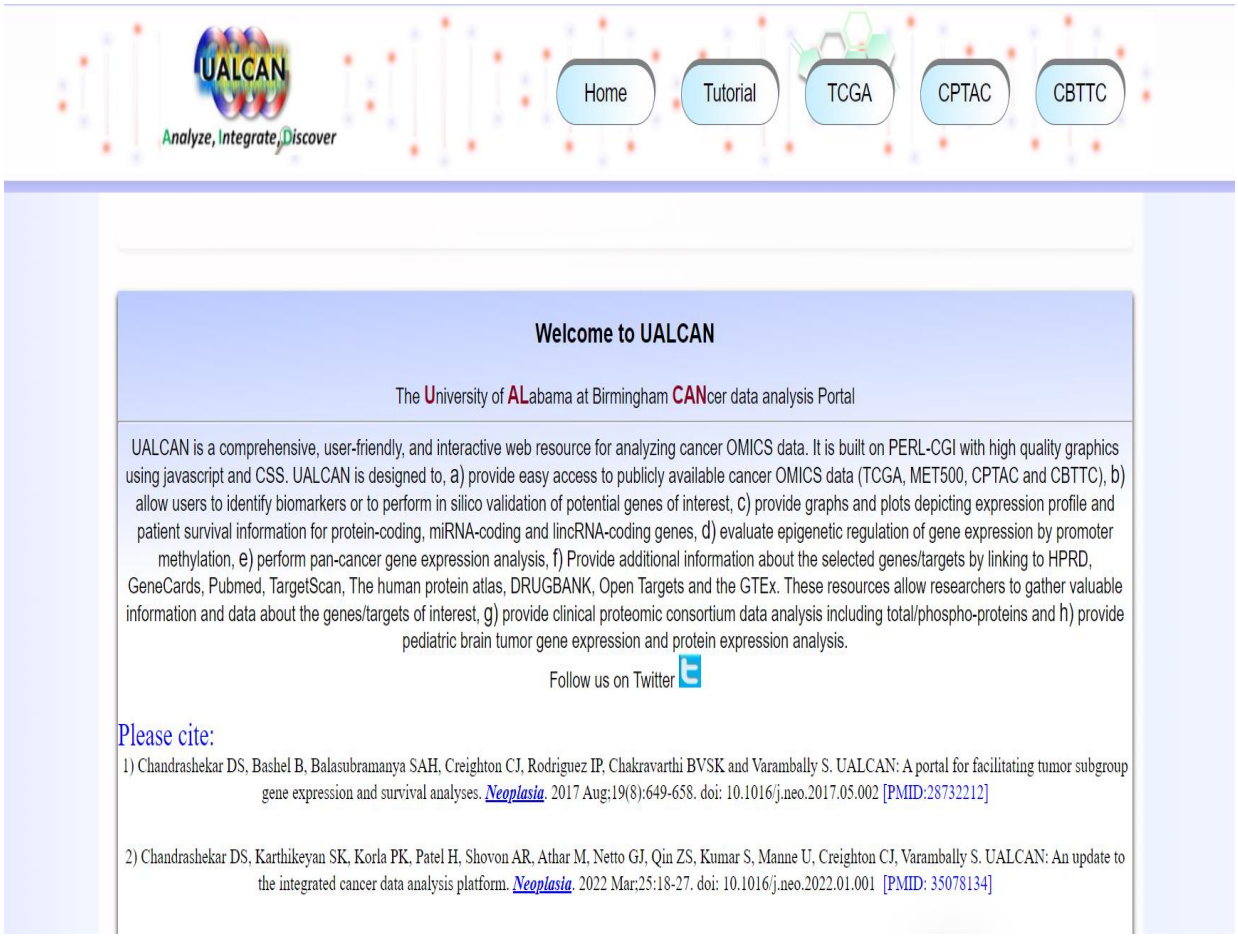


Fig.10. UALCAN: An integrated data mining platform for cancer transcriptome analysis

Using UALCAN, it is possible to explore/validate the pan cancer expression pattern of hundreds of use-defined gene via “Scan by gene classes” option. UALCAN serves as a one-stop-shop by providing easy access to external resources such as GeneCards, Human Protein Reference Database (HPRD) (to explore relevant protein interactions), PubMed, TargetScan (to find predicted microRNA that potentially regulate the gene of interest) and Human Protein Atlas. Here, we applied Kaplan-Meier plots to validate the expression pattern of hub genes among AML patients with a principle scores < 0.05.


RESULT

4

RESULTS

The results pertaining to the study entitled **"**Computational identification of key genes and pathways associated with Acute Myeloid Leukemia**"** are presented in the following headings.

4.1. Gene expression dataset

Information for 170 patients with adult de novo AML and corresponding bone marrow RNA-seq datasets were obtained from the TCGA database. There were 17 AML patients with NPM1 mutation. To understand the relevant process and pathways affected by NPM1 mutation we screened the DEGs between NPM1 mutation and wild-type AML. A total of 214 DEGs (116 upregulated and 98 downregulated) were identified (Figure.11). Top ten up and down regulated genes were represented in Table.I.

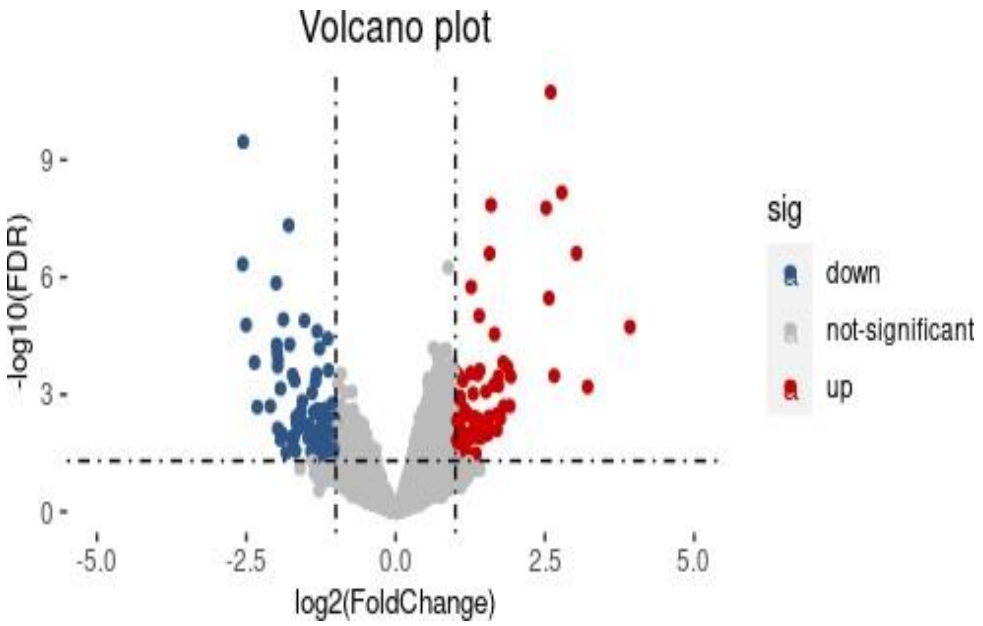


Fig.11. Volcano plot of comparing all of the significant DEGs from NPM1 mutation and NPM1 wild-type samples.

Significance	Gene ID
Up	HOXA4, EWF1A2, SDSL, DOCK3, PTGIS, IL12A, ADRB2,SCARA3,TSPA N32, C3orF80
Down	SETBP1, NAV1, FABN5, IL12RB2, PTPRM, GYPC, SLC4A3, RNY1P12,RNU6-437P, TIFAB

Table I: Top ten up and down regulated genes between NPM1 mutation and wild-type

4.2. GSEA

To investigate the effect of NPM1 mutations on the prognosis of AML patients, the effects of NPM1 mutations on various biological functional gene sets were analyzed by the GSEA approach. The GSEA results are shown in Figure 2. In the GSEA analysis of GO enrichment (Figure.2. A-J) biological processes pertaining to positive regulation of cell differentiation, regulation of myeloid cell differentiation, negative regulation of cell proliferation, leukocyte differentiation and activation, and cell adhesion, were significantly enriched.

This suggests that NPM1 mutations may contribute to disease progression and affect prognosis by influencing cell differentiation, proliferation, and cell adhesion in AML patients. However, the GO enrichment analysis of molecular function was significantly enriched in receptor signaling proteins, serine-threonine kinase activity, and protein tyrosine kinase activity. Furthermore, the cellular component was enriched for adherens junction and cell fraction. In the GSEA analysis of KEGG pathways, the NPM1 mutation group was associated with the MAPK signaling pathway and pathways in cancer (Figure.12. K-L).

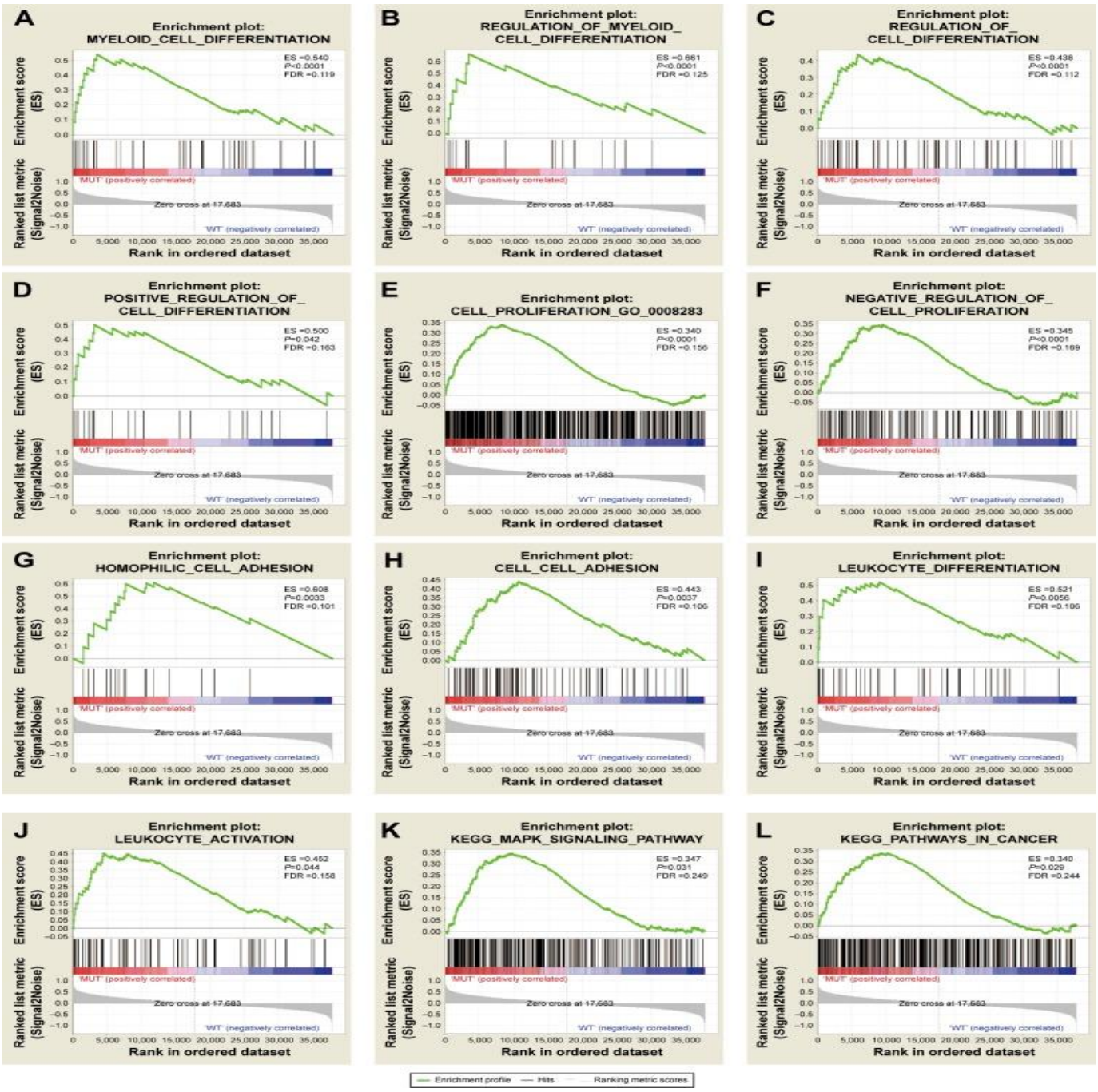


Fig.12. GSEA results of NPM1 mutations in AML patients

4.3. GO and KEGG enrichment analyses of DEGs

In total, 214 DEGs were submitted for GO and KEGG pathway analyses with DAVID, respectively. For biological processes, DEGs suggested significant enrichment in apoptotic processes. For cellular components and molecular function, DEGs were enriched in the cell-cell adherens junction and positive regulation of I-kappaB kinase/NF-kappaB signaling is shown in Table.II.

Category	Term	Count	%	P-value
BP_DIRECT	Translation	17	2.28	0.00179
BP_DIRECT	rRNA prosessing	12	1.35	1.90E-05
BP_DIRECT	Apoptotic process	39	3.86	7.30E-07
BP_DIRECT	Mitotic nuclear division	15	1.03	0.010349
BP_DIRECT	positive regulation of I-kappaBkinase/NF-kappaB signaling	20	2.32	0.004496
BP_DIRECT	Cell cycle	14	1.73	0.010349
BP_DIRECT	Cellular response to DNA damage stimulus	8	1.02	0.020261
BP_DIRECT	Protein folding	3	0.97	0.047991
CC_DIRECT	Cytosol	22	17.34	1.34E-15
CC_DIRECT	Cell-cell adherens junction	34	22.41	2.11E-04
CC_DIRECT	Protein complex	11	12.89	4.32E-04
CC_DIRECT	Centrosome	18	14.82	2.90E-03
CC_DIRECT	Lysosome	10	11.09	2.62E-02
MF_DIRECT	mRNA binding	4	1.06	3.47E-03
MF_DIRECT	Protein serine/threonine kinaseactivity	8	2.47	1.43E-02
MF_DIRECT	Ligase activity	4	1.8	2.89E-02
MF_DIRECT	Identical protein binding	6	4.43	3.96E-02
MF_DIRECT	Trancription factor binding	4	1.83	4.81E-02

Table II. Gene Ontologies of DEGs involved in NPM1 mutation associated AML

KEGG pathway analysis was also conducted for DEGs. The result showed DEGs were highly enriched in pathways of cancer Table.III. Consequently, KEGG Mapper v3.2 was used, it is an efficient tool for pathway mapping, which were provided to highlight the specified genes of selected pathways is shown in Figure.13.

Pathway ID	Term	Count	%	P-value
hsa00190	Oxidative phosphorylation	18	17.36	2.15E-04
hsa05200	Pathways in cancer	42	25.62	5.14E-05
hsa04110	Cell cycle	12	11.43	7.55E-03

Table III. Enriched pathways of DEGs involved in NPM1 mutation associated AML

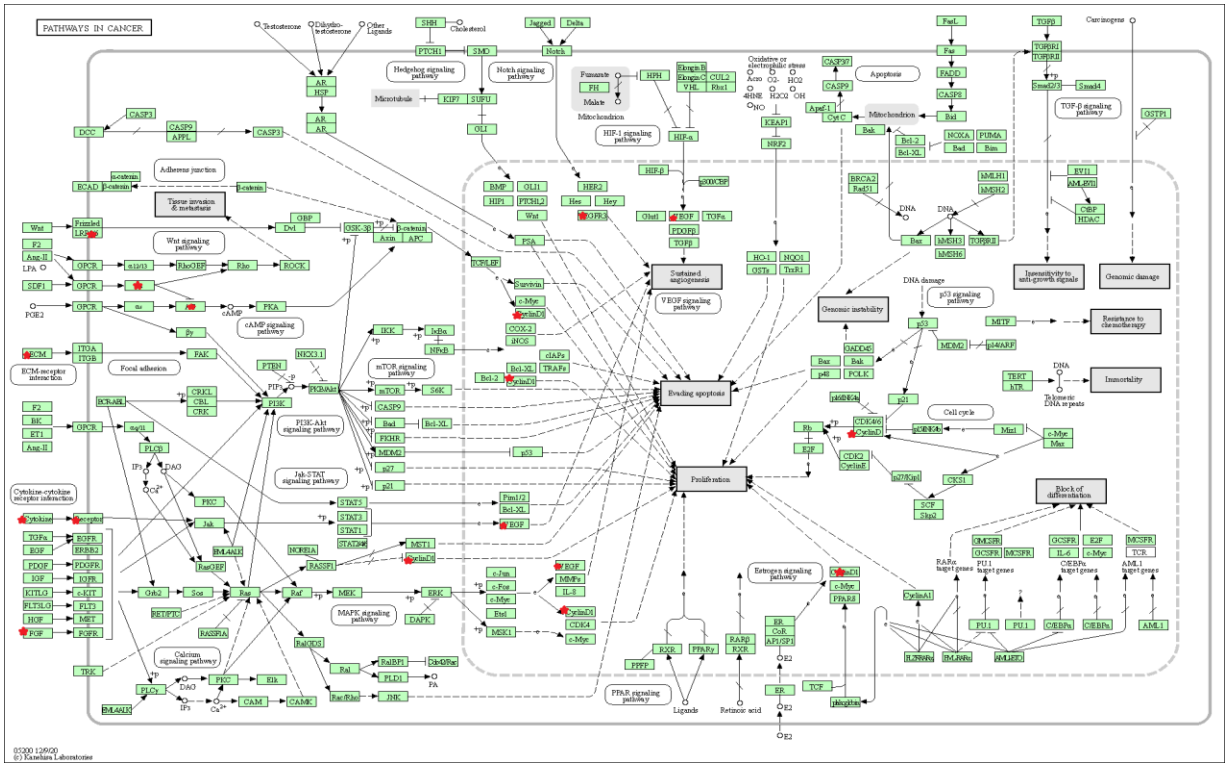


Fig.13. Reconstructed KEGG pathway of cancer signaling by using the KEGG mapper tool. The leading significant network system related to DEGs where highlighted with the red colour.

4.4. Identification of miRNA-target gene network

As it is known that the miRNA chips of microarray analysis indicated the involvement of several miRNAs, this study was identified the associated miRNAs for all the significant up and down regulated genes from miRNet, but here before the miRNA prediction, the removal of duplicated genes was done to get only unique genes, which resulted total 1108 numbers of miRNAs for DEGs respectively. The resulted data were visualized through gene-miRNA interaction network graph with highly associated genes and miRNAs i.e. the highly associated genes were PPM1L, MRO, RASSF9, GLUL, GNAS, SLC25A33 and the miRNAs were has-mir335-5p and has-mir-26b-5p (Figure.14).

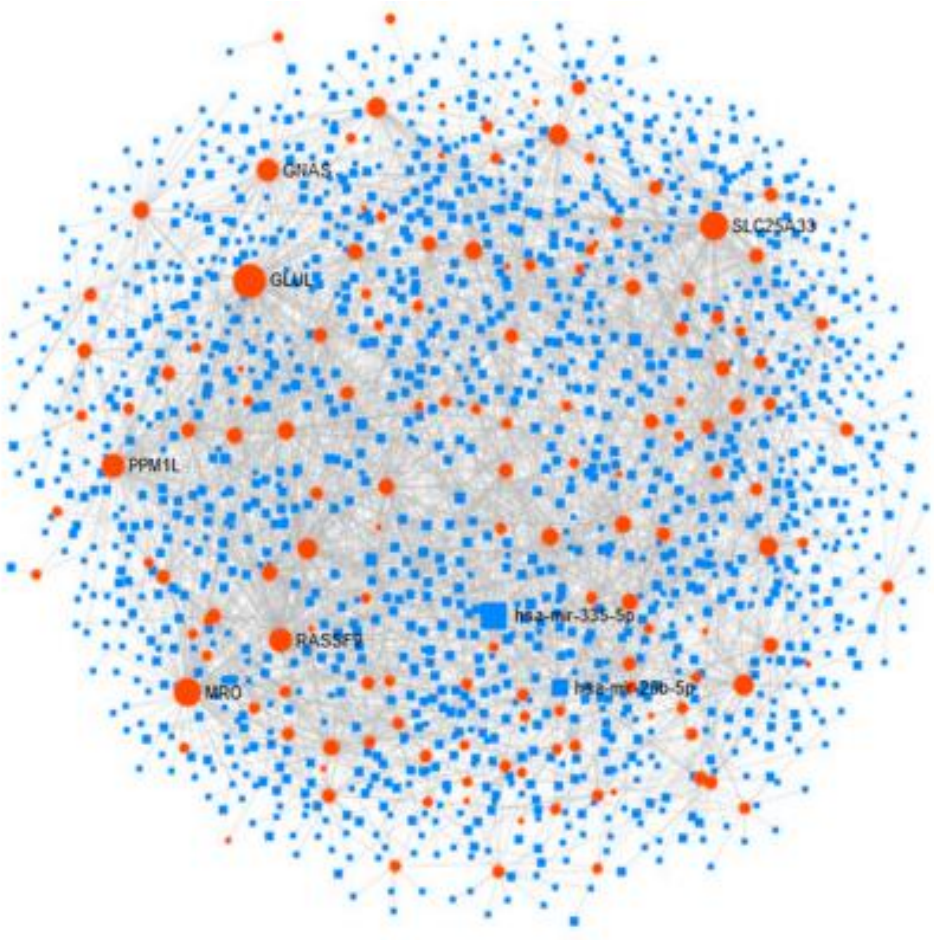


Fig.14. Most associated miRNAs with DEGs

4.5. Transcription factors and survival analysis

Analysis of the transcription factors targeting potential DEGs may be useful in understanding the mechanisms of AML progression. Consequently, the Network analyst was then employed to distinguish the probable transcription factors targeting significant DEGs, and eventually, RPL7L1, NIFK, LYN, MRPS5 and MRPL1 were recognized to be top five factors is shown in Figure.15. Interestingly, these transcription factors have been represented to be functional in the driving numerous cancer types especially AML and drug resistance.

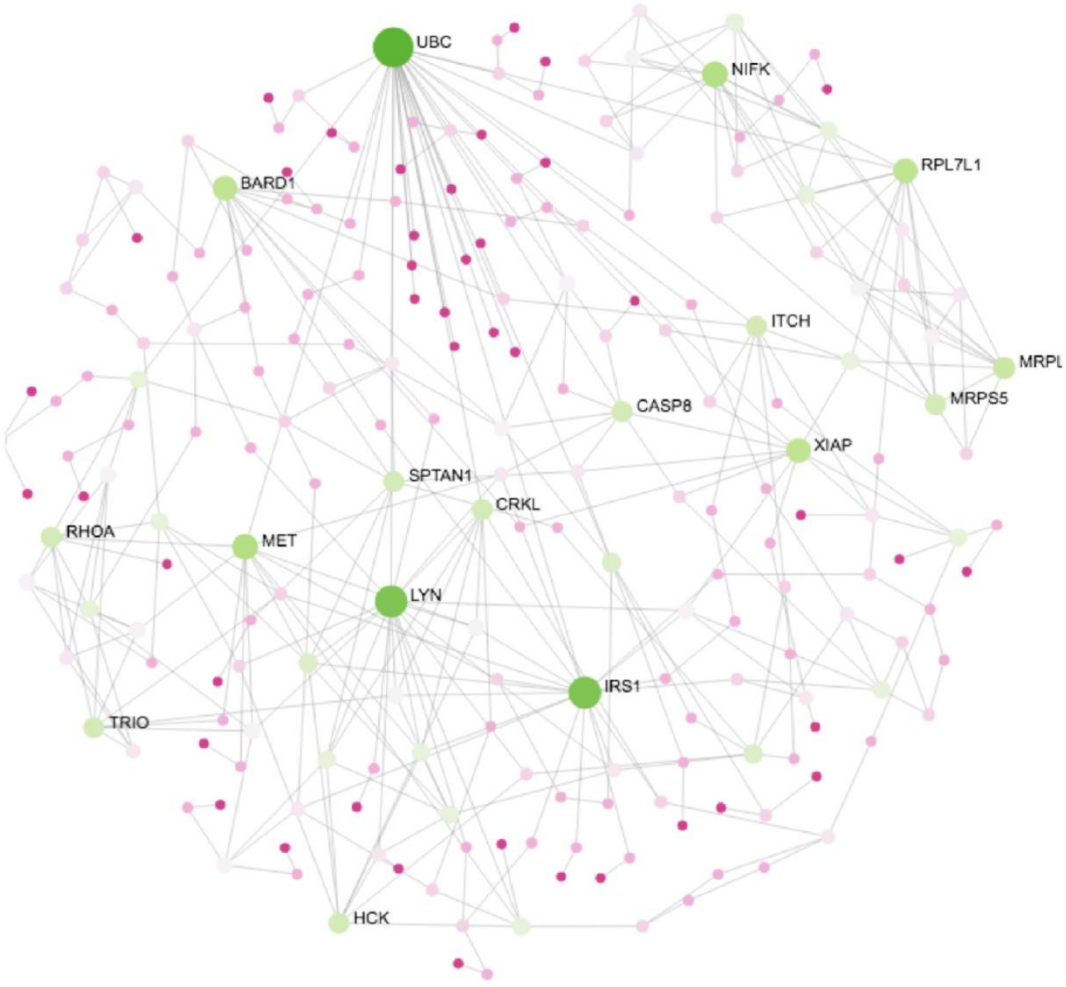


Fig.15. Most associated transcription factors with DEGs

4.6. PPI network and module analysis

We constructed the protein-protein interactome networks and identified some NPM1 mutation associated hub genes. The network was examined by using cytoscape software (Figure.16). The top ten genes ranked by degree were identified as hub genes, including AKT1, SRC, EDN1, ADRB2, PRKACB, MAPK3, GNAS, PRKACG, suggests that hub genes are significantly enriched in biological processes, developmental, and metabolic which have been associated with cancer (Figure.17). PRKAR2B and PRKAR1A had the highest degree of nodes.

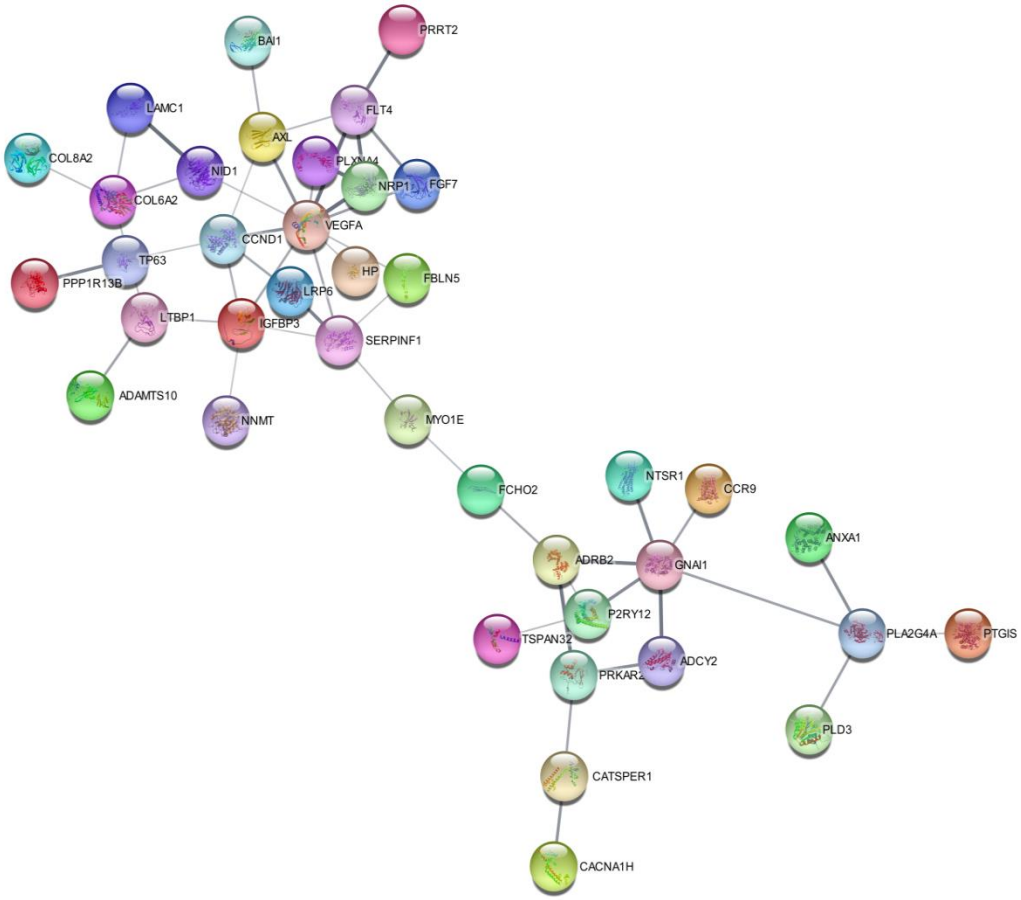


Fig.16. Confidence view of protein-protein interaction network of DEGs predicted using Cytoscape Software

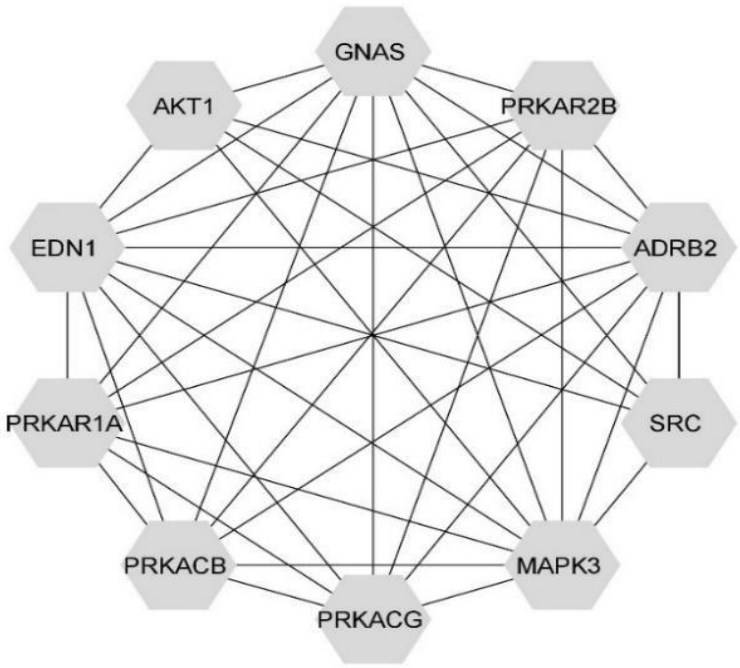


Fig.17. PPI network of ten hub genes with high degree of connectivity

Top ten GO enrichment terms and pathways were reported, and DEGs were observed to be mostly enriched in positive regulation of cellular processes, protein binding (Figure.18).

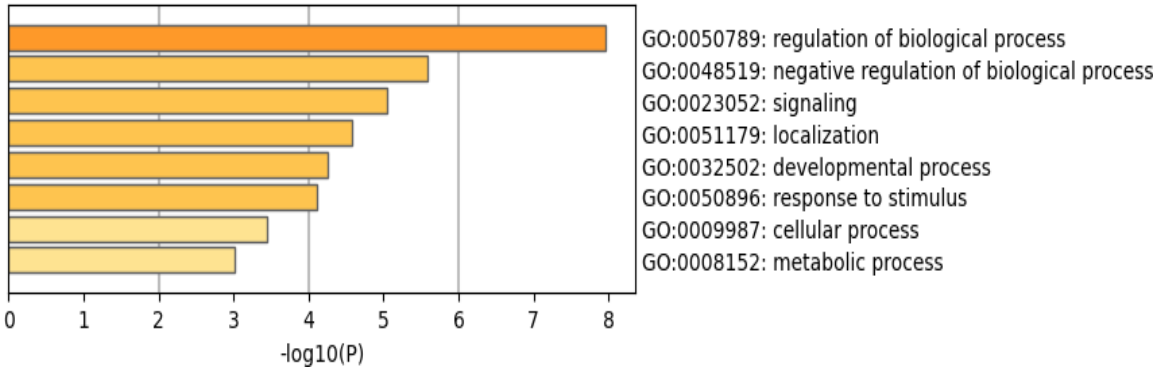


Fig.18. GO term analysis of top ten hub genes by Metascape

4.7. Expression level and prognostic value of hub genes

The expression level of ten hub genes (ADRB2, AKT1, EDN1, GNAS , MAPK3, PRKACB, PRKACG, PRKAR1A, PRKAR2B, SRC) in NPM1 mutation and wild-type AML patients is shown in Figure.19. (A-J).

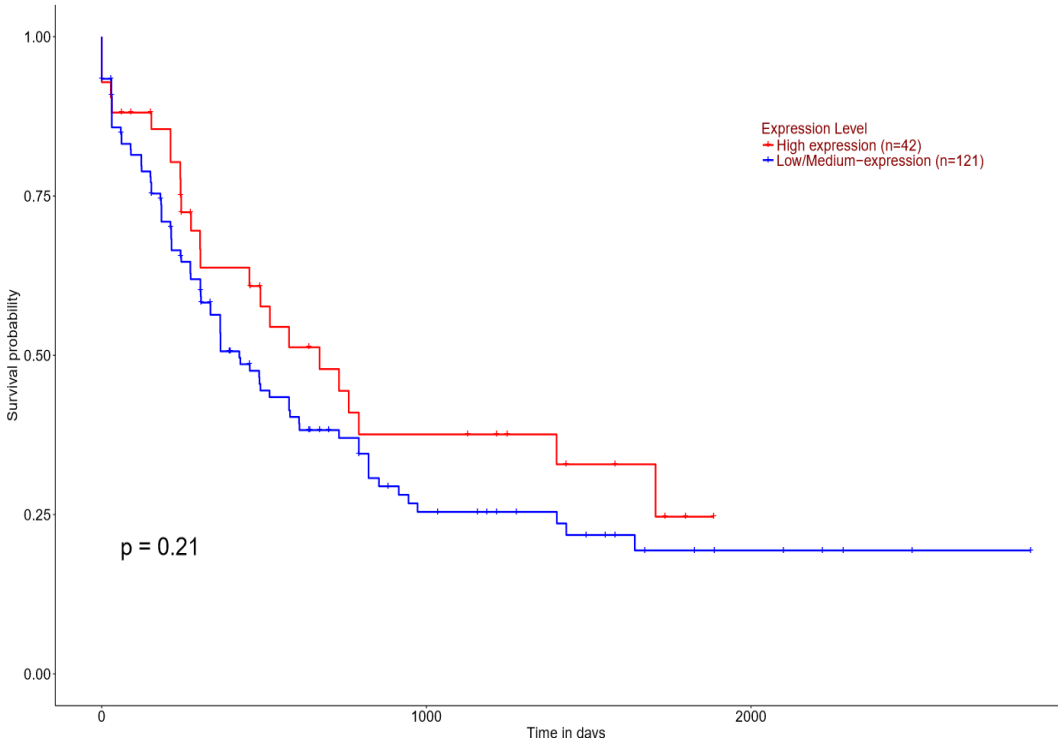
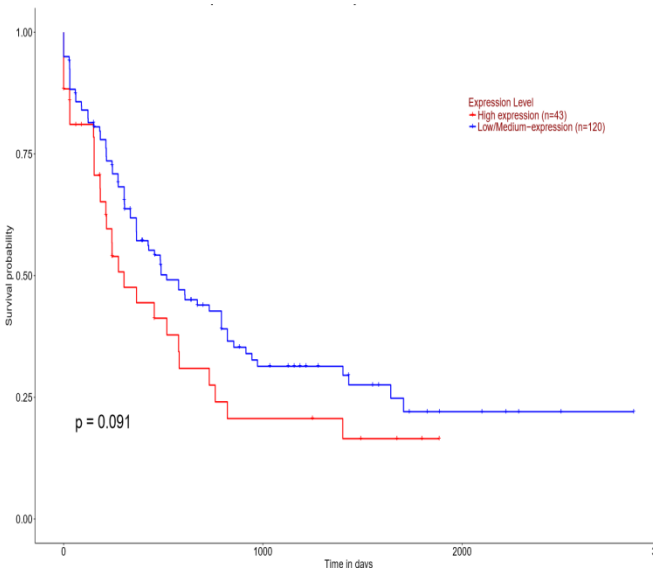
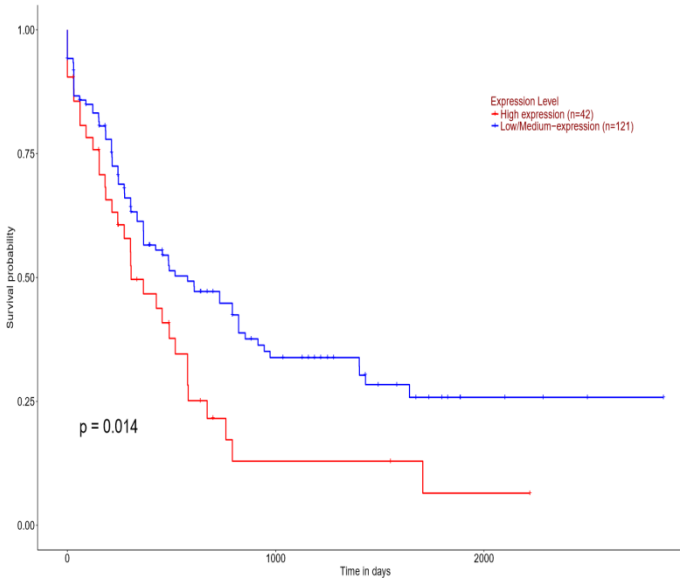


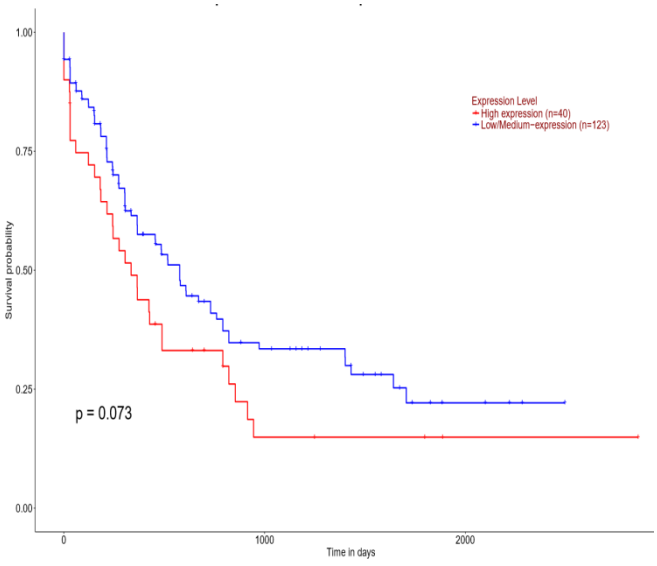
Fig.19. (A) Expression level and prognostic value of ADRB2



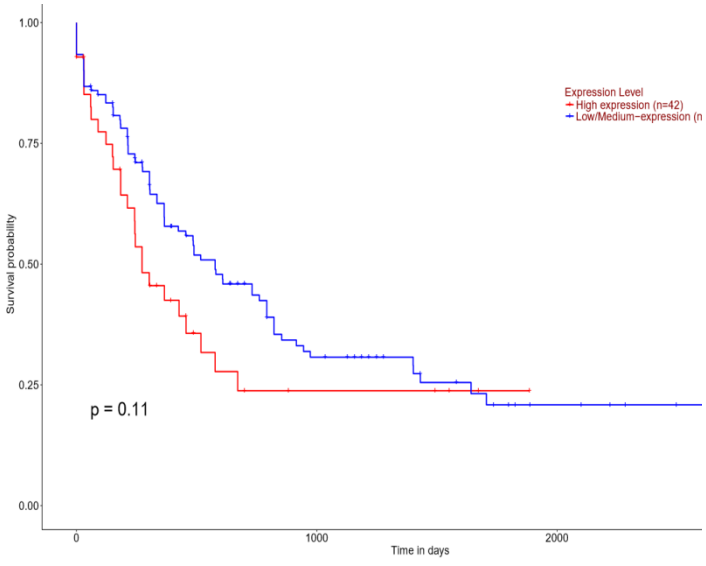
(B) Expression level and prognostic value of AKT1



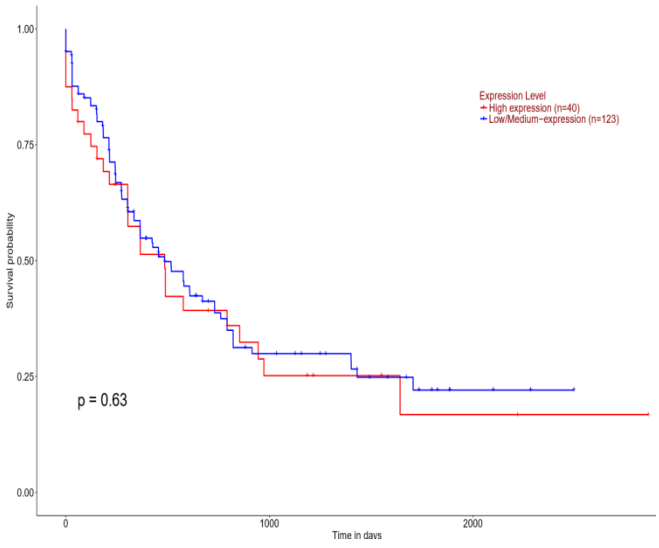
(C) Expression level and prognostic value of EDN1



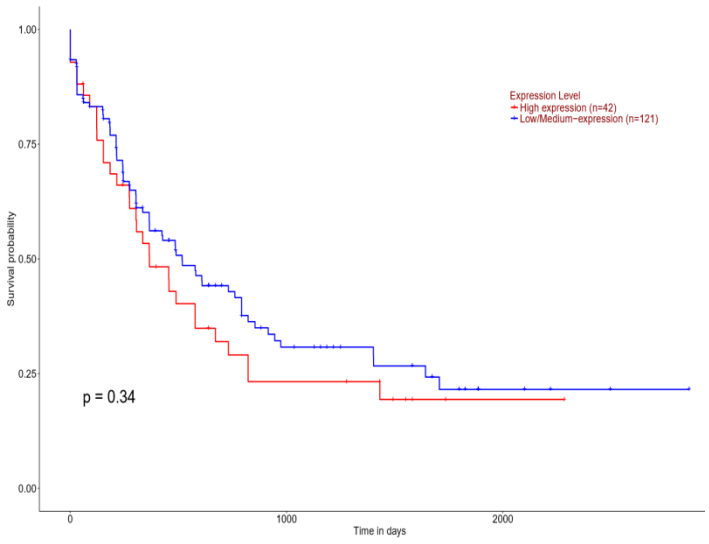
(D) Expression level and prognostic value of GNAS



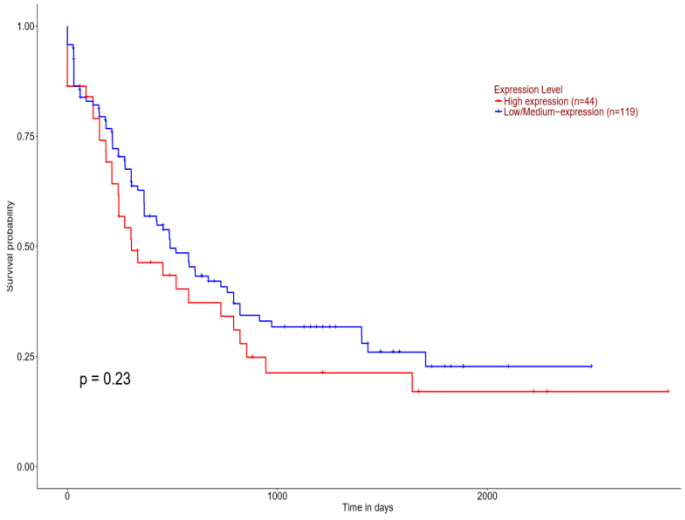
(E) Expression level and prognostic value of MAPK3



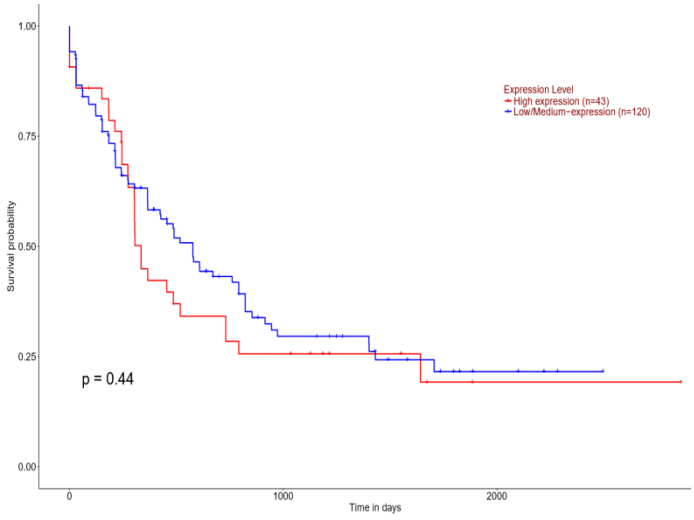
(F) Expression level and prognostic value of PRKACB



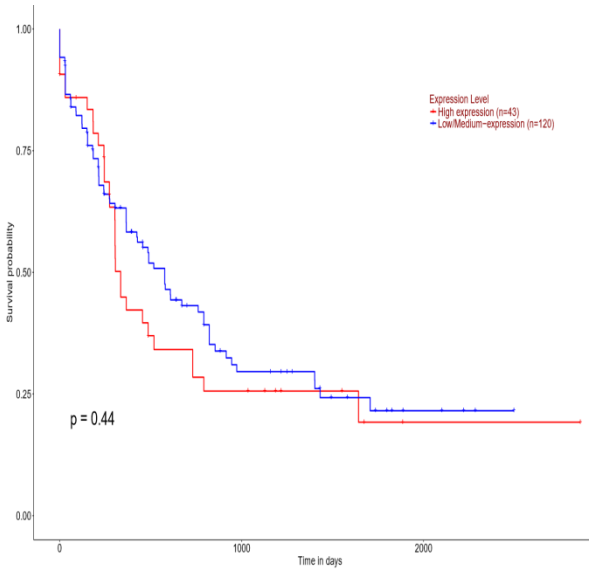
(G) Expression level and prognostic value of PRKACG



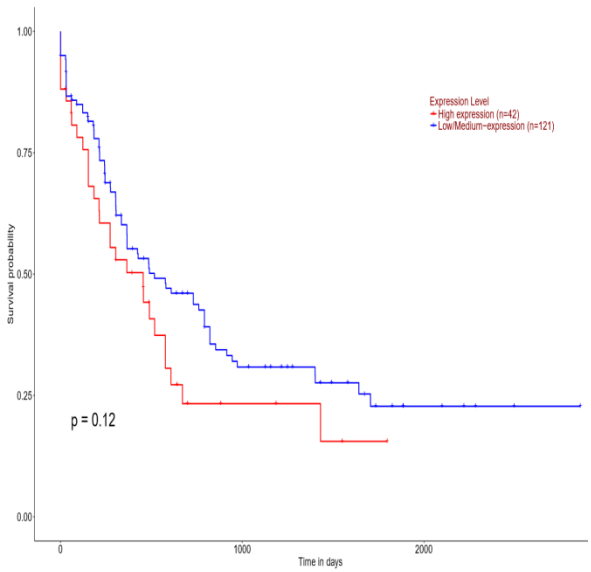
(H) Expression level and prognostic value of PRKAR1A



(I) Expression level and prognostic value of PRKAR2B



(I) Expression level and prognostic value of PRKAR2B



(J) Expression level and prognostic value of SRC

The expression level and prognostic value of ten hub genes are ADRB2 (P = 0.21), AKT1 (P = 0.091), EDN1 (P = 0.014), GNAS (P = 0.073), MAPK3 (P = 0.11), PRKACB (P = 0.63), PRKACG (P = 0.34), PRKAR1A (P = 0.23), PRKAR2B (P = 0.44), SRC (P = 0.12). Corresponding survival analyses suggested that that EDN1 had a significant prognostic value in AML.

DISCUSSION

5

DISCUSSION

AML is characterized by proliferative, clonal, abnormally differentiated, and occasionally poorly differentiated cells of the hematopoietic system, which infiltrate the bone marrow, blood, and other tissues. It is well accepted that genetic variants can be used as independent prognostic biomarkers for AML due to their potential efficacy of chemotherapy. NPM1 mutations in exon 12 represent the most frequent molecular aberrations in adult patients with AML Molecular detection of NPM1 mutation A could be a useful marker for routine monitoring of minimal residual disease (MRD) (Papadaki et al., 2009). NPM1 mutations have been reported as a dangerous element in AML. NPM1 might be a novel biomarker for the early diagnosis and treatment of AML.

However, the changes in biological process and signaling pathway NPM1 mutation cause had not been reported. Therefore, further investigation is essential for better understanding of the biological roles of NPM1 mutations in AML. Hepatocyte growth factor promotes proliferation, invasion, and metastasis of AML cells by regulating the PI3K-Akt and MAPK/ERK signaling pathway. In addition, drug resistance of AML stem cells can be modulated by surviving expression via the MAPK signaling pathway.

Herein, we used an RNAseq dataset of adult AML from the TCGA database to identify the key genes and pathways associated with NPM1 mutation via bioinformatic analysis.

Stirewalt et al. (2008) studied identification of genes with abnormal expression changes in AML.

In the present study, altogether 214 differentially expressed genes were identified, of which 116 were upregulated and 98 were downregulated. GSEA analysis in the present study suggests that NPM1 mutations were significantly associated with the regulation of cell differentiation, especially for myeloid cell differentiation, negative regulation of cell proliferation, cell adhesion, and leukocyte differentiation and activation. The GO term analysis indicated that downregulated DEGs in NPM1 mutation AML patients were related to biological

processes of cell differentiation and development. The GO term enrichment analysis suggests that upregulated DEGs were related to complement activation, and cell adhesion biological processes.

GO analysis showed that DEGs were notably abundant in apoptosis. It was suggested that NPM1 mutations may contribute to disease progression and affect prognosis by influencing apoptotic process and protein serine/threonine kinase activity in AML patients. DEGs were enriched in various pathways in cancer and oxidative phosphorylation. Consistent with previous studies, metabolic pathways have been reported that affect the pathogenesis and prognosis of AML.

The deregulation of miRNAs can affect most of the normal biological processes such as development, metabolism, immunity and leading to cause complex diseases. So this study was subjected to reveal the link between target genes and miRNAs and identified the mostly linked miRNAs (has-mir-335-5p, has-mir-26b-5p) and target genes through miRNet, which is a robust tool for micro RNA target information and also for visualization (Fan et al., 2016). In the present study, the highly associated genes were PPM1L, MRO, RASSF9, GLUL, GNAS, SLC25A33. This study falls in line with the reports of Cammarata et al. (2010) who reported Specific miRNAs with consolidated role on cell proliferation and differentiation such as miR-155, miR-221, let-7, miR-126 and miR-196b as assosciated miRNAs.

Multiple aberrant microRNA expression has been reported in gastric cancer. Among them, microRNA-335-5p (miR-335), a microRNA regulated by DNA methylation, has been reported to possess both tumor suppressor and tumor promoter activities (Sandoval-Bórquez et al., 2017). The ability of miR-26b-5p to regulate T cell factor 4 (TCF-4) expression and thereby control human adipose-derived mesenchymal stem cell (hADMSC) adipogenic differentiation (Luo et al., 2020).

PPM1L acts as a suppressor of the SAPK signaling pathways by associating with and dephosphorylating MAP3K7/TAK1 and MAP3K5, and by attenuating the association between MAP3K7/TAK1 and MAP2K4 or MAP2K6. The protein encoded by this gene is a magnesium or manganese-requiring phosphatase that is involved in several signaling pathways. The encoded

protein down regulates apoptosis signal-regulating kinase 1, a protein that initiates a signaling cascade that leads to apoptosis when cells are subjected to cytotoxic stresses (Lu et al., 2013).

MRO is a protein coding gene. This gene is specifically transcribed in males before and after differentiation of testis, and the encoded protein may play an important role in a mammalian sex determination. Multiple transcript variants encoding different isoforms have been found for this gene. An important paralog of this gene is MROH2B (Smith et al., 2003).

Diseases associated with RASSF9 include Isolated Growth Hormone Deficiency, Type Ii. The protein encoded by this gene localizes to perinuclear endosomes. This protein associates with peptidylglycine alpha-amidating monooxygenase, and may be involved with the trafficking of this enzyme through secretory or endosomal pathways. However, its roles in tumorigenesis, particularly in non-small cell lung cancer (NSCLC), are still not understood well (Yuan et al., 2021).

Glutamate-ammonia ligase (GLUL) belongs to the glutamine synthetase family. It catalyzes the synthesis of glutamine from glutamate and ammonia in an ATP-dependent reaction. Higher expression of GLUL in the breast cancer patients was associated with larger tumor size and higher level of HER2 expression. In addition, GLUL was heterogeneously expressed in various breast cancer cells. Among its related pathways are Neuroinflammation and glutamatergic signaling and glutamate dependent acid resistance (Wang et al., 2017).

GNAS is a complex imprinted gene that uses multiple promoters to generate several gene products, including the G protein α-subunit (Gsα) that couples seven-transmembrane receptors to the cAMP-generating enzyme adenylyl cyclase. GNAS mutations cause developmental delay, short stature, and skeletal abnormalities in a syndrome called Albright's hereditary osteodystrophy. Because of imprinting, mutations on the maternal allele also cause obesity and hormone resistance (Mendes de Oliveira et al., 2021).

SLC25A33 belongs to the SLC25 family of mitochondrial carrier proteins. Diseases associated with SLC25A33 include Orofaciodigital syndrome VIII and Schnyder Corneal Dystrophy (Haitina et al., 2006).

Moreover, in the present study several epigenetic transcription factors that regulate common DEGs were screened by employing Network Analyst database. Top five factors, including RPL7L1, NIFK, LYN, MRPS5 and MRPL1 were observed. Similar studies were conducted by Takei and Kobayashi (2019). Their results corroborates with our results.

RPL7L1 enables RNA binding activity. Predicted to be involved in maturation of LSU-rRNA from tricistronic rRNA transcript. Diseases associated with RPL7L1 include Charcot-Marie-Tooth Disease, Axonal, and Type 2N. GO annotations related to this gene include RNA binding (Maserati et al., 2014).

NIFK gene encodes a protein that interacts with the forkhead-associated domain of the Ki-67 antigen. The encoded protein may bind RNA and may play a role in mitosis and cell cycle progression. Multiple pseudogenes exist on chromosomes 5, 10, 12, 15, and 19. NIFK expression is clinically associated with poor prognosis and metastasis (Lin et al., 2016).

Src family kinases such as Lyn are important signaling intermediaries, relaying and modulating different inputs to regulate various outputs, such as proliferation, differentiation, apoptosis, migration and metabolism. Intriguingly, Lyn can mediate both positive and negative signaling processes within the same or different cellular contexts. Lyn has also been found to be important for maintaining the leukemic phenotype of many different liquid cancers including acute myeloid leukaemia (AML), chronic myeloid leukaemia (CML) and B-cell lymphocytic leukemia (BCLL) (Ingley, 2012).

Mammalian mitochondrial ribosomal proteins are encoded by nuclear genes and help in protein synthesis within the mitochondrion. Mitochondrial ribosomes (mitoribosomes) consist of a small 28S subunit and a large 39S subunit. Diseases associated with MRPS5 include Raynaud-Claes Syndrome and Hypotonia-Cystinuria Syndrome. Among its related pathways are mitochondrial translation and Metabolism of proteins (Akbergenov et al., 2018).

MRPL1 (Mitochondrial Ribosomal Protein L1) is a Protein Coding gene. Among its related pathways are mitochondrial translation and Metabolism of proteins. GO annotations related to this gene include RNA binding and structural constituent of ribosome (Islam et al., 2019).

Furthermore in the present study, we built the protein-protein interactome networks and selected some hub genes with high connectivity involved in NPM1 mutation AML. The results of GO enrichment analysis of ten hub genes were similar to previous analysis of DEGs by Zhu et al., (2021) and so wherein they described a total of 15 hub genes were identified to be crucial for AML progression. Finally, survival analysis revealed that one of the hub genes, CEACAM5, was significantly associated with AML prognosis and could serve as a potential target for AML treatment.

There are accumulating evidence that ADRB2 signaling contributes to the progression and therapy resistance of prostate cancer (Kulik, 2019). ADRB2 is a protein coding gene. This gene encodes beta-2-adrenergic receptor which is a member of the G protein-coupled receptor superfamily. This gene is intronless. Among its related pathways are Neuroscience and Peptide ligand-binding receptors. GO annotations related to this gene include G protein-coupled receptor activity and enzyme binding. An important paralog of this gene is ADRB1.

AKT1 mediated tumor cell-vasculature reciprocity during the advanced stages of cancers and it differentially regulates cancer metastasis through mechanisms distinct from its protumorigenic effects (Alwhaibi et al., 2019). The AKT1 gene provides instructions for making a protein called AKT1 kinase. This protein is found in various cell types throughout the body, where it plays a critical role in many signaling pathways. Signaling involving AKT1 kinase appears to be essential for the normal development and function of the nervous system. Studies have suggested a role for AKT1 kinase in cell-to-cell communication among nerve cells (neurons), neuronal survival, and the formation of memories. The AKT1 gene belongs to a class of genes known as oncogenes. When mutated, oncogenes have the potential to cause normal cells to become cancerous.

EDN1 is a growth factor that is frequently produced by cancer cells and plays a critical role in tumorigenesis (Kim et al., 2005). This gene encodes a preproprotein that is proteolytically processed to generate a secreted peptide that belongs to the endothelin/sarafotoxin family. Aberrant expression of this gene may promote tumorigenesis. Alternative splicing results in multiple transcript variants. Probable ligand for G-protein coupled receptors EDNRA and EDNRB which activates PTK2B, BCAR1, BCAR3 and, GTPases RAP1 and RHOA cascade in glomerular mesangial cells.

GNAS is a complex imprinted gene that uses multiple promoters to generate several gene products, including the G protein α-subunit (Gsα) that couples seven-transmembrane receptors to the cAMP-generating enzyme adenylyl cyclase. Somatic activating Gsα mutations, which alter key residues required for the GTPase turn-off reaction, are present in various endocrine tumors and fibrous dysplasia of bone, and in a more widespread distribution in patients with McCune-Albright syndrome (Weinstein et al., 2004).

The expression of MAPK3 was highly enhanced in all breast cancer cells and over expression blocked the expression of MAPK3 (Du et al., 2020). MAPK3 is a protein coding gene. The protein encoded by this gene is a member of the MAP kinase family. MAP kinases, also known as extracellular signal-regulated kinases (ERKs), act in a signaling cascade that regulates various cellular processes such as proliferation, differentiation, and cell cycle progression in response to a variety of extracellular signals. This kinase is activated by upstream kinases, resulting in its translocation to the nucleus where it phosphorylates nuclear targets. Alternatively spliced transcript variants encoding different protein isoforms have been described. GO annotations related to this gene include transferase activity, transferring phosphorus-containing groups and protein tyrosine kinase activity. An important paralog of this gene is MAPK1.

PRKACB is a protein coding gene. The protein encoded by this gene is a member of the serine/threonine protein kinase family. The encoded protein is a catalytic subunit of cAMP (cyclic AMP)-dependent protein kinase, which mediates signalling though cAMP. cAMP signaling is important to a number of processes, including cell proliferaton and differentiation. Multiple alternatively spliced transcript variants encoding distinct isoforms have been observed (Espiard et al., 2018).

PRKACG is a Protein Coding gene. Diseases associated with PRKACG include Bleeding Disorder, Platelet-Type, 19 and Primary Pigmented Nodular Adrenocortical Disease. Among its related pathways are Apoptotic Pathways in Synovial Fibroblasts and Regulation of activated PAK-2p34 by proteasome mediated degradation. Gene Ontology (GO) annotations related to this gene include transferase activity, transferring phosphorus-containing groups and protein tyrosine kinase activity. An important paralog of this gene is PRKACA (Manchev, 2015).

PRKAR1A is a bone tumor suppressor gene mainly associated with subclass development and driving RANKL over expression during tumorigenesis. The PRKAR1A gene provides instructions for making one part (subunit) of an enzyme called protein kinase A. This enzyme promotes cell growth and division (proliferation). Protein kinase A is made up of four protein subunits, two of which are called regulatory subunits because they control whether this enzyme is turned on or off. The PRKAR1A gene provides instructions for making one of these regulatory subunits, called type 1 alpha. Protein kinase A remains turned off when the regulatory subunits are attached to the other two subunits of the enzyme. In order to turn on protein kinase A, the regulatory subunits must break away from the enzyme (Bossis and Stratakis, 2004).

PRKAR2B is highly expressed in castration resistant prostate cancer and contributes to aerobic glycolysis and tumor growth in prostate cancer (Xia et al., 2020). cAMP is a signaling molecule important for a variety of cellular functions. cAMP exerts its effects by activating the cAMP-dependent protein kinase, which transduces the signal through phosphorylation of different target proteins. The inactive kinase holoenzyme is a tetramer composed of two regulatory and two catalytic subunits. cAMP causes the dissociation of the inactive holoenzyme into a dimer of regulatory subunits bound to four cAMP and two free monomeric catalytic subunits. Four different regulatory subunits and three catalytic subunits have been identified in humans. The protein encoded by this gene is one of the regulatory subunits. This subunit can be phosphorylated by the activated catalytic subunit. This subunit has been shown to interact with and suppress the transcriptional activity of the cAMP responsive element binding protein 1 (CREB1) in activated T cells. Among its related pathways are DAG and IP3 signaling and Aquaporin mediated transport.

Several studies reported that SRC is a non-receptor tyrosine kinase that is deregulated in many cancer types and in diabetes promoted carcinogenesis (Irby and Yeatman, 2000 ; Wheeler et al., 2009). This gene is highly similar to the v-src gene of Rous sarcoma virus. This proto-oncogene may play a role in the regulation of embryonic development and cell growth. The protein encoded by this gene is a tyrosine-protein kinase whose activity can be inhibited by phosphorylation by c-SRC kinase. Two transcript variants encoding the same protein have been found for this gene. Among its related pathways are cytokine Signaling in Immune

system and PEDF induced Signaling. GO annotations related to this gene include transferase activity, transferring phosphorus-containing groups and protein tyrosine kinase activity.

Our study has shown that NPM1 mutations were associated with poor prognosis, Therefore, with this investigation, we recognize and validate that common DEGs of NPM1 mutation patients may contain the disease progression mechanisms through significant alternative pathways in AML. However, no study has reported the association of EDN1 with AML or other cancers. Thus, it is important to further explore the role of EDN1 in the pathogenesis and prognosis of AML. This may provide a basis to develop molecular markers for improved patient stratification.

SUMMARY AND CONCLUSION

6

CONCLUSION

The present study provides a comprehensive analysis of the prognostic signature of NPM1 mutation and its underlying gene expression profiles in Acute Myeloid Leukemia (AML). Multiple bioinformatics analysis results revealed that cell adhesion, differentiation, and proliferation may play a crucial role in Nucleophosmin (NPM1) mutation of AML. By exploring the TCGA (The Cancer Genome Atlas) database, we have identified 214 differentially expressed genes (DEGs) that are associated with AML progression and cancer. Our comparative genomic analysis of 170 (153 wild type and 17 NPM1 mutation) AML patients indicates that the mutation of the NPM1 gene is related to poor overall survival in patients.

Gene Set Enrichment Analysis (GSEA) analysis in the present study suggests that NPM1 mutations were significantly associated with the regulation of cell differentiation, especially for myeloid cell differentiation, negative regulation of cell proliferation, cell adhesion, and leukocyte differentiation and activation.

The Gene Ontology (GO) term analysis indicated that downregulated differentially expressed genes (DEGs) in NPM1 mutation of AML patients were related to biological processes of cell differentiation and development. The Gene Ontology (GO) term enrichment analysis suggests that upregulated DEGs were related to apoptotic process and protein serine/threonine kinase activity. The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis implies that cancer signaling pathway is significantly enriched in upregulated DEGs' pathways.

This study was carried out to reveal the link between target genes and microRNAs (miRNAs). The present study identified the most associated miRNA (has-mir-335-5p, has-mir-26b-5p) and target genes. Then the Transcription Factors (TFs) (RPL7L1, NIFK, LYN, MRPS5 and MRPL1) that regulate common DEGs were screened.

We have also identified ten potent biomarkers (ADRB2, AKT1, EDN1, GNAS, MAPK3, GNAS, PRKACB, PRKARCG PRKAR1A, PRKAR2B, SRC) based on computational screening. The significant Gene Ontology (GO) term enrichments of the top ten modules were

associated with a variety of G-protein coupled receptor signaling pathways, a family of cell-surface proteins that play critical roles in regulating a variety of pathophysiological processes.

Our study has shown that Nucleophosmin (NPM1) mutations are associated with poor prognosis, in consistence with previous studies. The important functions, pathways, and biomarkers which were identified might be helpful for innovation in medical research for easy diagnosis and treatment for target-based therapeutics. However, further molecular biological experiments are required to confirm our findings.

In summary, we have proved the existence of common differentially expressed genes (DEGs) including alternative pathways in AML by using comprehensive *in silico* methods and computational systems biology examinations. Our findings provide a new insight into better understanding of the mechanisms of hub genes regulation and the pathogenesis of alternative pathways in tumorigenesis and AML progression.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abraham, J. and Staffurth, J., 2016. Hormonal therapy for cancer. *Medicine*, *44*(1), pp.30-33.

Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.J., Thoren, L.A. and Anderson, K., 2005. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: a revised road map for adult blood lineage commitment. *Cell*, *121*(2), pp.295-306.

Akashi, K., Traver, D., Miyamoto, T. and Weissman, I.L., 2000. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, *404*(6774), pp.193-197.

Akbergenov, R., Duscha, S., Fritz, A.K., Juskeviciene, R., Oishi, N., Schmitt, K., Shcherbakov, D., Teo, Y., Boukari, H., Freihofer, P. and Isnard-Petit, P., 2018. Mutant MRPS 5 affects mitoribosomal accuracy and confers stress-related behavioral alterations. *EMBO reports*, *19*(11), p.e46193.

Ali, I., Wani, W.A. and Saleem, K., 2011. Cancer scenario in India with future perspectives. *Cancer therapy*, *8*.

Alwhaibi, A., Verma, A., Adil, M.S. and Somanath, P.R., 2019. The unconventional role of Akt1 in the advanced cancers and in diabetes-promoted carcinogenesis. *Pharmacological research*, *145*, p.104270.

Andersson, A.F., Lindberg, M., Jakobsson, H., Bäckhed, F., Nyrén, P. and Engstrand, L., 2008. Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PloS one*, *3*(7), p.e2836.

Antar, A.I., Otrock, Z.K., Jabbour, E., Mohty, M. and Bazarbachi, A., 2020. FLT3 inhibitors in acute myeloid leukemia: ten frequently asked questions. *Leukemia*, *34*(3), pp.682-696.

Arber, D.A., 2001. Realistic pathologic classification of acute myeloid leukemias. *American journal of clinical pathology*, *115*(4), pp.552-560.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. and Harris, M.A., 2000. Gene ontology: tool for the unification of biology. *Nature genetics*, *25*(1), pp.25-29.

Baskar, R., Lee, K.A., Yeo, R. and Yeoh, K.W., 2012. Cancer and radiation therapy: current advances and future directions. *International journal of medical sciences*, *9*(3), p.193.

Belson, M., Kingsley, B. and Holmes, A., 2007. Risk factors for acute leukemia in children: a review. *Environmental health perspectives*, *115*(1), pp.138-145.

Blackadar, C.B., 2016. Historical review of the causes of cancer. *World journal of clinical oncology*, *7*(1), p.54.

Bloomfield, C.D., 1985. Post-remission therapy in acute myeloid leukemia. *Journal of Clinical Oncology*, *3*(12), pp.1570-1572.

Borghaei, H., Smith, M.R. and Campbell, K.S., 2009. Immunotherapy of cancer. *European journal of pharmacology*, *625*(1-3), pp.41-54.

Bossis, I. and Stratakis, C.A., 2004. Minireview: PRKAR1A: normal and abnormal functions. *Endocrinology*, *145*(12), pp.5452-5458.

Brunning, R.D., 2003, August. Classification of acute leukemias. In *Seminars in Diagnostic Pathology* (Vol. 20, No. 3, pp. 142-153). WB Saunders.

Burnett, A., Wetzler, M. and Lowenberg, B., 2011. Therapeutic advances in acute myeloid leukemia. *J Clin Oncol*, *29*(5), pp.487-494.

Cammarata, G., Augugliaro, L., Salemi, D., Agueli, C., Rosa, M.L., Dagnino, L., Civiletto, G., Messana, F., Marfia, A., Bica, M.G. and Cascio, L., 2010. Differential expression of specific microRNA and their targets in acute myeloid leukemia. *American journal of hematology*, *85*(5), pp.331-339

Cancer Genome Atlas Research Network, 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *New England Journal of Medicine*, *368*(22), pp.2059-2074.

Chabner, B.A. and Roberts, T.G., 2005. Chemotherapy and the war on cancer. *Nature Reviews Cancer*, *5*(1), pp.65-72.

Chan, W.Y., Liu, Q.R., Borjigin, J., Busch, H., Rennert, O.M., Tease, L.A. and Chan, P.K., 1989. Characterization of the cDNA encoding human nucleophosmin and studies of its role in normal and abnormal growth. *Biochemistry*, *28*(3), pp.1033-1039.

Chandrashekar, D.S., Bashel, B., Balasubramanya, S.A.H., Creighton, C.J., Ponce-Rodriguez, I., Chakravarthi, B.V. and Varambally, S., 2017. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*, *19*(8), pp.649-658.

Chang, L., Zhou, G., Soufan, O. and Xia, J., 2020. miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology. *Nucleic Acids Research*, *48*(W1), pp.W244-W251.

Chapla, Ugandhar. (2015). LEUKEMIA – BRIEF REVIEW ON RECENT ADVANCEMENTS IN THERAPY AND MANAGEMENT. Asian Journal of Research in Pharmaceutical Sciences and Biotechnology. 3. 12-26.

Chatterjee, S.K. and Zetter, B.R., 2005. Cancer biomarkers: knowing the present and predicting the future.

Ciavarella, S., Milano, A., Dammacco, F. and Silvestris, F., 2010. Targeted therapies in cancer. *BioDrugs*, *24*(2), pp.77-88.

Cortes, J., Foran, J., Ghirdaladze, D., DeVetten, M.P., Zodelava, M., Holman, P., Levis, M.J., Kantarjian, H.M., Borthakur, G., James, J. and Zarringkar, P.P., 2009. AC220, a potent, selective, second generation FLT3 receptor tyrosine kinase (RTK) inhibitor, in a first-in-human (FIH) phase 1 AML study. *Blood*, *114*(22), p.636.

Dalenc, F., Drouet, J., Ader, I., Delmas, C., Rochaix, P., Favre, G., Cohen-Jonathan, E. and Toulas, C., 2002. Increased expression of a COOH-truncated nucleophosmin resulting from alternative splicing is associated with cellular resistance to ionizing radiation in HeLa cells. *International journal of cancer*, *100*(6), pp.662-668.

Daver, N. and Cortes, J., 2012. Molecular targeted therapy in acute myeloid leukemia. *Hematology*, *17*(sup1), pp.s59-s62.

Daver, N., Schlenk, R.F., Russell, N.H. and Levis, M.J., 2019. Targeting FLT3 mutations in AML: review of current knowledge and evidence. *Leukemia*, *33*(2), pp.299-312.

Davies, H., Bignell, G.R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M.J., Bottomley, W. and Davis, N., 2002. Mutations of the BRAF gene in human cancer. *Nature*, *417*(6892), pp.949-954.

De Kouchkovsky, I. and Abdul-Hay, M., 2016. Acute myeloid leukemia: a comprehensive review and 2016 update. *Blood cancer journal*, *6*(7), pp.e441-e441.

Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A., 2003. DAVID: database for annotation, visualization, and integrated discovery. *Genome biology*, *4*(9), pp.1-11.

Derman, B.A. and Larson, R.A., 2019. Post-remission therapy in acute myeloid leukemia: Are we ready for an individualized approach?. *Best Practice & Research Clinical Haematology*, *32*(4), p.101102.

Deschler, B. and Lübbert, M., 2008. Acute myeloid leukemia: epidemiology and etiology. *Acute Leukemias*, pp.47-56.

Döhner, H., Weisdorf, D.J. and Bloomfield, C.D., 2015. Acute myeloid leukemia. *New England Journal of Medicine*, *373*(12), pp.1136-1152.

Döhner, K., Thiede, C., Jahn, N., Panina, E., Gambietz, A., Larson, R.A., Prior, T.W., Marcucci, G., Jones, D., Krauter, J. and Heuser, M., 2020. Impact of NPM1/FLT3-ITD genotypes defined by the 2017 European LeukemiaNet in patients with acute myeloid leukemia. *Blood*, *135*(5), pp.371-380.

Du, Y., Zhang, J., Meng, Y., Huang, M., Yan, W. and Wu, Z., 2020. MicroRNA-143 targets MAPK3 to regulate the proliferation and bone metastasis of human breast cancer cells. *Amb Express*, *10*(1), pp.1-8.

Dutta, S., Akey, I.V., Dingwall, C., Hartman, K.L., Laue, T., Nolte, R.T., Head, J.F. and Akey, C.W., 2001. The crystal structure of nucleoplasmin-core: implications for histone binding and nucleosome assembly. *Molecular cell*, *8*(4), pp.841-853.

Espiard, S., Knape, M.J., Bathon, K., Assié, G., Rizk-Rabin, M., Faillot, S., Luscap-Rondof, W., Abid, D., Guignat, L., Calebiro, D. and Herberg, F.W., 2018. Activating PRKACB somatic mutation in cortisol-producing adenomas. *JCI insight*, *3*(8).

Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La Starza, R., Diverio, D., Colombo, E., Santucci, A. and Bigerna, B., 2005. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *New England Journal of Medicine*, *352*(3), pp.254-266.

Fan, Y. and Xia, J., 2018. miRNet—functional analysis and visual exploration of miRNA–target interactions in a network context. In *Computational cell biology* (pp. 215-233). Humana Press, New York, NY.

Fan, Y., Siklenka, K., Arora, S.K., Ribeiro, P., Kimmins, S. and Xia, J., 2016. miRNet-dissecting miRNA-target interactions and functional associations through network-based visual analysis. *Nucleic acids research*, *44*(W1), pp.W135-W141.

Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C. and Parkin, D.M., 2008. GLOBOCAN 2008, Cancer incidence and mortality worldwide: IARC CancerBase No. 10. *Lyon, France: International Agency for Research on Cancer*.

Friday, B.B. and Adjei, A.A., 2008. Advances in targeting the Ras/Raf/MEK/Erk mitogen-activated protein kinase cascade with MEK inhibitors for cancer therapy. *Clinical cancer research*, *14*(2), pp.342-346.

Fukawa, T., Ono, M., Matsuo, T., Uehara, H., Miki, T., Nakamura, Y., Kanayama, H.O. and Katagiri, T., 2012. DDX31 regulates the p53-HDM2 pathway and rRNA gene transcription through its interaction with NPM1 in renal cell carcinomas. *Cancer research*, *72*(22), pp.5867-5877.

Gaidzik, V.I., Teleanu, V., Papaemmanuil, E., Weber, D., Paschka, P., Hahn, J., Wallrabenstein, T., Kolbinger, B., Köhne, C.H., Horst, H.A. and Brossart, P., 2016. RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features. *Leukemia*, *30*(11), pp.2160-2168.

Garzon, R., Garofalo, M., Martelli, M.P., Briesewitz, R., Wang, L., Fernandez-Cymering, C., Volinia, S., Liu, C.G., Schnittger, S., Haferlach, T. and Liso, A., 2008. Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin. *Proceedings of the National Academy of Sciences*, *105*(10), pp.3945-3950.

Gekas, C., Rhodes, K.E., Van Handel, B., Chhabra, A., Ueno, M. and Mikkola, H.K., 2010. Hematopoietic stem cell development in the placenta. *The International journal of developmental biology*, *54*, p.1089.

Glass, D.C., Gray, C.N., Jolley, D.J., Gibbons, C., Sim, M.R., Fritschi, L., Adams, G.G., Bisby, J.A. and Manuell, R., 2003. Leukemia risk associated with low-level benzene exposure. *Epidemiology*, pp.569-577.

Gorin, N.C., 1998. Autologous stem cell transplantation in acute myelocytic leukemia. *Blood, The Journal of the American Society of Hematology*, *92*(4), pp.1073-1090.

Green CL, Koo KK, Hills RK, Burnett AK, Linch DC, Gale RE. Prognostic significance of CEBPA mutations in a large cohort of younger adult patients with acute myeloid leukemia: impact of double CEBPA mutations and the interaction with FLT3 and NPM1 mutations. J Clin Oncol. 2010;28:2739–47.

Grisendi, S., Bernardi, R., Rossi, M., Cheng, K., Khandker, L., Manova, K. and Pandolfi, P.P., 2005. Role of nucleophosmin in embryonic development and tumorigenesis. *Nature*, *437*(7055), pp.147-153.

Haase, D., Feuring-Buske, M., Konemann, S., Fonatsch, C., Troff, C., Verbeek, W., Pekrun, A., Hiddemann, W. and Wormann, B., 1995. Evidence for malignant transformation in acute myeloid leukemia at the level of early hematopoietic stem cells by cytogenetic analysis of CD34+ subpopulations.

Hackl, H., Astanina, K. and Wieser, R., 2017. Molecular and genetic alterations associated with therapy resistance and relapse of acute myeloid leukemia. *Journal of hematology & oncology*, *10*(1), pp.1-16.

Haitina, T., Lindblom, J., Renström, T. and Fredriksson, R., 2006. Fourteen novel human members of mitochondrial solute carrier family 25 (SLC25) widely expressed in the central nervous system. *Genomics*, *88*(6), pp.779-790.

Hassanpour, S.H. and Dehghani, M., 2017. Review of cancer from perspective of molecular. *Journal of Cancer Research and Practice*, *4*(4), pp.127-129.

Helleberg, C., Knudsen, H., Hansen, P.B., Nikolajsen, K., Kjaersgaard, E., Ralfkiaer, E. and Johnsen, H.E., 1997. CD34+ megakaryoblastic leukaemic cells are CD38−, but CD61+ and glycophorin A+: improved criteria for diagnosis of AML-M7?. *Leukemia*, *11*(6), pp.830-834.

Hoang, T., 2004. The origin of hematopoietic cell type diversity. *Oncogene*, *23*(43), pp.7188-7198.

Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C. and Lempicki, R.A., 2007. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome biology*, *8*(9), pp.1-16.

Hwang, S.M., 2020. Classification of acute myeloid leukemia. *Blood research*, *55*(S1), pp.S1-S4.

Indian Council of Medical Research. National Cancer Registry Programme, three year report of population based cancer registries 2006-2008: National Cancer Registry Programme (ICMR), Bangalore, 2010.

Infante-Rivard, C., Mathonnet, G. and Sinnett, D., 2000. Risk of childhood leukemia associated with diagnostic irradiation and polymorphisms in DNA repair genes. *Environmental health perspectives*, *108*(6), pp.495-498.

Ingley, E., 2012. Functions of the Lyn tyrosine kinase in health and disease. *Cell Communication and signaling*, *10*(1), pp.1-11.

Irby, R.B. and Yeatman, T.J., 2000. Role of Src expression and activation in human cancer. *Oncogene*, *19*(49), pp.5636-5642.

Islam, T., Rahman, M.R., Shuvo, M.A.H., Shahjaman, M., Islam, M.R. and Karim, M.R., 2019. Drug repositioning and biomarkers in low-grade glioma via bioinformatics approach. *Informatics in Medicine Unlocked*, *17*, p.100250.

Ivey, A., Hills, R.K., Simpson, M.A., Jovanovic, J.V., Gilkes, A., Grech, A., Patel, Y., Bhudia, N., Farah, H., Mason, J. and Wall, K., 2016. Assessment of minimal residual disease in standard-risk AML. *New England Journal of Medicine*, *374*(5), pp.422-433.

Jaffray, D.A. and Gospodarowicz, M.K., 2015. Radiation therapy for cancer. *Disease Control Priorities,*, *3*, pp.239-247.

James, C., Ugo, V., Le Couédic, J.P., Staerk, J., Delhommeau, F., Lacout, C., Garçon, L., Raslova, H., Berger, R., Bennaceur-Griscelli, A. and Villeval, J.L., 2005. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *nature*, *434*(7037), pp.1144-1148.

Jiao, X., Sherman, B.T., Huang, D.W., Stephens, R., Baseler, M.W., Lane, H.C. and Lempicki, R.A., 2012. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics*, *28*(13), pp.1805-1806.

Juliusson, G., Lehmann, S. and Lazarevic, V., 2021. Epidemiology and Etiology of AML. In *Acute Myeloid Leukemia* (pp. 1-22). Springer, Cham.

Kanehisa, M. and Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), pp.27-30.

Kassim, A.A. and Savani, B.N., 2017. Hematopoietic stem cell transplantation for acute myeloid leukemia: A review. *Hematology/oncology and stem cell therapy*, *10*(4), pp.245-251.

Kim, T.H., Xiong, H., Zhang, Z. and Ren, B., 2005. β-Catenin activates the growth factor endothelin-1 in colon cancer cells. *Oncogene*, *24*(4), pp.597-604.

Kulik, G., 2019. ADRB2-Targeting therapies for prostate cancer. *Cancers*, *11*(3), p.358.

Larrosa-Garcia, M. and Baer, M.R., 2017. FLT3 inhibitors in acute myeloid leukemia: current status and future directions. *Molecular cancer therapeutics*, *16*(6), pp.991-1001.

Levis, M., 2013. FLT3 mutations in acute myeloid leukemia: what is the best approach in 2013?. *Hematology 2013, the American Society of Hematology Education Program Book*, *2013*(1), pp.220-226.

Licht, J.D. and Sternberg, D.W., 2005. The molecular pathology of acute myeloid leukemia. *ASH Education Program Book*, *2005*(1), pp.137-142.

Lin, T.C., Su, C.Y., Wu, P.Y., Lai, T.C., Pan, W.A., Jan, Y.H., Chang, Y.C., Yeh, C.T., Chen, C.L., Ger, L.P. and Chang, H.T., 2016. The nucleolar protein NIFK promotes cancer progression via CK1α/β-catenin in metastasis and Ki-67-dependent cell proliferation. *Elife*, *5*, p.e11288.

Lin, T.C., Su, C.Y., Wu, P.Y., Lai, T.C., Pan, W.A., Jan, Y.H., Chang, Y.C., Yeh, C.T., Chen, C.L., Ger, L.P. and Chang, H.T., 2016. The nucleolar protein NIFK promotes cancer progression via CK1α/β-catenin in metastasis and Ki-67-dependent cell proliferation. *Elife*, *5*, p.e11288.

Lu, G., Ota, A., Ren, S., Franklin, S., Rau, C.D., Ping, P., Lane, T.F., Zhou, Z.H., Reue, K., Lusis, A.J. and Vondriska, T., 2013. PPM1l encodes an inositol requiring-protein 1 (IRE1) specific phosphatase that regulates the functional outcome of the ER stress response. *Molecular metabolism*, *2*(4), pp.405-416.

Luo, Y., Ji, H., Cao, Y., Ding, X., Li, M., Song, H., Li, S., WaTableng, C., Wu, H., Meng, J. and Du, H., 2020. miR-26b-5p/TCF-4 controls the adipogenic differentiation of human adipose-derived mesenchymal stem cells. *Cell transplantation*, *29*, p.0963689720934418.

Mahoney, M.C., Moysich, K.B., McCarthy, P.L., McDonald, R.C., Stepanenko, V.F., Day, R.W. and Michalek, A.M., 2004. The Chernobyl childhood leukemia study: background & lessons learned. *Environmental Health*, *3*(1), pp.1-7.

Manchev, V., 2015. *Pathogenesis in two inherited thrombocytopenias: PRKACG-related disease and FPD/AML* (Doctoral dissertation, Sorbonne Paris Cité).

Mardis, E.R., 2012. Genome sequencing and cancer. *Current opinion in genetics & development*, *22*(3), pp.245-250.

Marian, A.J., 2014. Sequencing your genome: what does it mean?. *Methodist DeBakey cardiovascular journal*, *10*(1), p.3.

Martelli, A.M., Nyåkern, M., Tabellini, G., Bortul, R., Tazzari, P.L., Evangelisti, C. and Cocco, L., 2006. Phosphoinositide 3-kinase/Akt signaling pathway and its therapeutical implications for human acute myeloid leukemia. *Leukemia*, *20*(6), pp.911-928.

Maserati, M., Dai, X., Walentuk, M. and Mager, J., 2014. Identification of four genes required for mammalian blastocyst formation. *Zygote*, *22*(3), pp.331-339.

Matthews, W., Jordan, C.T., Wiegand, G.W., Pardoll, D. and Lemischka, I.R., 1991. A receptor tyrosine kinase specific to hematopoietic stem and progenitor cell-enriched populations. *Cell*, *65*(7), pp.1143-1152.

Mendes de Oliveira, E., Keogh, J.M., Talbot, F., Henning, E., Ahmed, R., Perdikari, A., Bounds, R., Wasiluk, N., Ayinampudi, V., Barroso, I. and Mokrosiński, J., 2021. Obesity-associated GNAS mutations and the melanocortin pathway. *New England Journal of Medicine*, *385*(17), pp.1581-1592.

Mering, C.V., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. and Snel, B., 2003. STRING: a database of predicted functional associations between proteins. *Nucleic acids research*, *31*(1), pp.258-261.

Meshinchi, S. and Appelbaum, F.R., 2009. Structural and functional alterations of FLT3 in acute myeloid leukemia. *Clinical Cancer Research*, *15*(13), pp.4263-4269.

Metcalf, D., 1998. Lineage commitment and maturation in hematopoietic cells: the case for extrinsic regulation. *Blood, The Journal of the American Society of Hematology*, *92*(2), pp.345-347.

Milella, M., Precupanu, C.M., Gregorj, C., Ricciardi, M.R., Petrucci, M.T., Kornblau, S.M., Tafuri, A. and Andreeff, M., 2005. Beyond single pathway inhibition: MEK inhibitors as a platform for the development of pharmacological combinations with synergistic anti-leukemic effects. *Current pharmaceutical design*, *11*(21), pp.2779-2795.

Miller, C.A., Wilson, R.K. and Ley, T.J., 2013. Genomic landscapes and clonality of de novo AML. *The New England journal of medicine*, *369*(15), p.1473.

Moarii, M. and Papaemmanuil, E., 2017. Classification and risk assessment in AML: integrating cytogenetics and molecular profiling. *Hematology 2014, the American Society of Hematology Education Program Book*, *2017*(1), pp.37-44.

Mudhumita, J.B. and Zon, L.I., 2013. Hematopoiesis. *Development*, *140*(12), pp.2463-2467.

Mwenifumbo, J.C. and Marra, M.A., 2013. Cancer genome-sequencing study design. *Nature Reviews Genetics*, *14*(5), pp.321-332.

Nachmani, D., Bothmer, A.H., Grisendi, S., Mele, A., Bothmer, D., Lee, J.D., Monteleone, E., Cheng, K., Zhang, Y., Bester, A.C. and Guzzetti, A., 2019. Germline NPM1 mutations lead to altered rRNA 2′-O-methylation and cause dyskeratosis congenita. *Nature genetics*, *51*(10), pp.1518-1529.

Nickoloff, J.A., De Haro, L.P., Wray, J. and Hromas, R., 2008. Mechanisms of leukemia translocations. *Current opinion in hematology*, *15*(4), p.338.

O'Connor, S.M. and Boneva, R.S., 2007. Infectious etiologies of childhood leukemia: plausibility and challenges to proof. *Environmental health perspectives*, *115*(1), pp.146-150.

O'Donnell, M.R., Abboud, C.N., Altman, J., Appelbaum, F.R., Arber, D.A., Attar, E., Borate, U., Coutre, S.E., Damon, L.E., Goorha, S. and Lancet, J., 2012. Acute myeloid leukemia. *Journal of the National Comprehensive Cancer Network*, *10*(8), pp.984-1021.

Okuda, M., Horn, H.F., Tarapore, P., Tokuyama, Y., Smulian, A.G., Chan, P.K., Knudsen, E.S., Hofmann, I.A., Snyder, J.D., Bove, K.E. and Fukasawa, K., 2000. Nucleophosmin/B23 is a target of CDK2/cyclin E in centrosome duplication. *Cell*, *103*(1), pp.127-140.

Othus, M., Kantarjian, H., Petersdorf, S., Ravandi, F., Godwin, J., Cortes, J., Pierce, S., Erba, H., Faderl, S., Appelbaum, F.R. and Estey, E., 2014. Declining rates of treatment-related mortality in patients with newly diagnosed AML given 'intense'induction regimens: a report from SWOG and MD Anderson. *Leukemia*, *28*(2), pp.289-292.

Papadaki, C., Dufour, A., Seibl, M., Schneider, S., Bohlander, S.K., Zellmeier, E., Mellert, G., Hiddemann, W. and Spiekermann, K., 2009. Monitoring minimal residual disease in acute

myeloid leukaemia with NPM1 mutations by quantitative PCR: clonal evolution is a limiting factor. *British journal of haematology*, *144*(4), pp.517-523.

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., Potter, N.E., Heuser, M., Thol, F., Bolli, N. and Gundem, G., 2016. Genomic classification and prognosis in acute myeloid leukemia. *New England Journal of Medicine*, *374*(23), pp.2209-2221.

Parcells, B.W., Ikeda, A.K., Simms-Waldrip, T., Moore, T.B. and Sakamoto, K.M., 2006. FMS-like tyrosine kinase 3 in normal hematopoiesis and acute myeloid leukemia. *Stem cells*, *24*(5), pp.1174-1184.

Pérez-Herrero, E. and Fernández-Medarde, A., 2015. Advanced targeted therapies in cancer: Drug nanocarriers, the future of chemotherapy. *European journal of pharmaceutics and biopharmaceutics*, *93*, pp.52-79.

Platanias, L.C., 2003. Map kinase signaling pathways and hematologic malignancies. *Blood*, *101*(12), pp.4667-4679.

Pogoda, J.M., Nichols, P.W. and Preston-Martin, S., 2004. Alcohol consumption and risk of adult-onset acute myeloid leukemia: results from a Los Angeles County case-control study. *Leukemia research*, *28*(9), pp.927-931.

Rao, V.S., Srinivas, K., Sujini, G.N. and Kumar, G.N., 2014. Protein-protein interaction detection: methods and analysis. *International journal of proteomics*, *2014*.

Ravandi, F., Verstovsek, S., Estrov, Z., Burger, J.A., George, S., Bivins, C., Cortes-Franco, J., Garrett, W.M., Newton, R.C. and Kantarjian, H., 2009. Significant Activity of the JAK2 Inhibitor, INCB018424 in Patients with Secondary, Post-Myeloproliferative Disorder (MPD) Acute Myeloid Leukemia (sAML): Results of An Exploratory Phase II Study. *Blood*, *114*(22), p.631.

Redaelli, A., Lee, J.M., Stephens, J.M. and Pashos, C.L., 2003. Epidemiology and clinical burden of acute myeloid leukemia. *Expert Review of Anticancer Therapy*, *3*(5), pp.695-710.

Renneville, A., Roumier, C., Biggio, V., Nibourel, O., Boissel, N., Fenaux, P. and Preudhomme, C., 2008. Cooperating gene mutations in acute myeloid leukemia: a review of the literature. *leukemia*, *22*(5), pp.915-931.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), pp.139-140.

Rosenblatt, E. and Zubizarreta, E. eds., 2017. *Radiotherapy in cancer care: facing the global challenge* (pp. 80112-6). Vienna: International Atomic Energy Agency.

Sandoval-Bórquez, A., Polakovicova, I., Carrasco-Véliz, N., Lobos-González, L., Riquelme, I., Carrasco-Avino, G., Bizama, C., Norero, E., Owen, G.I., Roa, J.C. and Corvalán, A.H., 2017. MicroRNA-335-5p is a potential suppressor of metastasis and invasion in gastric cancer. *Clinical epigenetics*, *9*(1), pp.1-16.

Schlenk, R.F., Weber, D., Fiedler, W., Salih, H.R., Wulf, G., Salwender, H., Schroeder, T., Kindler, T., Lübbert, M., Wolf, D. and Westermann, J., 2019. Midostaurin added to chemotherapy and continued single-agent maintenance therapy in acute myeloid leukemia with FLT3-ITD. *Blood, The Journal of the American Society of Hematology*, *133*(8), pp.840-851.

Seyfried, T.N. and Shelton, L.M., 2010. Cancer as a metabolic disease. *Nutrition & metabolism*, *7*(1), pp.1-22.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, *13*(11), pp.2498-2504.

Shaw, R.J. and Cantley, L.C., 2006. Ras, PI (3) K and mTOR signalling controls tumour cell growth. *Nature*, *441*(7092), pp.424-430.

Shewach, D.S. and Kuchta, R.D., 2009. Introduction to cancer chemotherapeutics. *Chemical reviews*, *109*(7), pp.2859-2861.

Shimada, A., 2019. Hematological malignancies and molecular targeting therapy. *European Journal of Pharmacology*, *862*, p.172641.

Shlush, L.I., Zandi, S., Mitchell, A., Chen, W.C., Brandwein, J.M., Gupta, V., Kennedy, J.A., Schimmer, A.D., Schuh, A.C., Yee, K.W. and McLeod, J.L., 2014. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature*, *506*(7488), pp.328-333.

Sievers, E.L., 2000. Targeted therapy of acute myeloid leukemia with monoclonal antibodies and immunoconjugates. *Cancer chemotherapy and pharmacology*, *46*(1), pp.S18-S22.

Smith, L., Van Hateren, N., Willan, J., Romero, R., Blanco, G., Siggers, P., Walsh, J., Banerjee, R., Denny, P., Ponting, C. and Greenfield, A., 2003. Candidate testis-determining gene, Maestro (Mro), encodes a novel HEAT repeat protein. *Developmental dynamics: an official publication of the American Association of Anatomists*, *227*(4), pp.600-607.

Smith, M., Barnett, M., Bassan, R., Gatta, G., Tondini, C. and Kern, W., 2004. Adult acute myeloid leukaemia. *Critical reviews in oncology/hematology*, *50*(3), pp.197-222.

Steensma, D.P., McClure, R.F., Karp, J.E., Tefferi, A., Lasho, T.L., Powell, H.L., DeWald, G.W. and Kaufmann, S.H., 2006. JAK2 V617F is a rare finding in de novo acute myeloid leukemia, but STAT3 activation is common and remains unexplained. *Leukemia*, *20*(6), pp.971-978.

Stein, E.M., DiNardo, C.D., Pollyea, D.A., Fathi, A.T., Roboz, G.J., Altman, J.K., Stone, R.M., DeAngelo, D.J., Levine, R.L., Flinn, I.W. and Kantarjian, H.M., 2017. Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. *Blood, The Journal of the American Society of Hematology*, *130*(6), pp.722-731.

Stirewalt, D.L., Meshinchi, S., Kopecky, K.J., Fan, W., Pogosova-Agadjanyan, E.L., Engel, J.H., Cronk, M.R., Dorcy, K.S., McQuary, A.R., Hockenbery, D. and Wood, B., 2008. Identification of genes with abnormal expression changes in acute myeloid leukemia. *Genes, Chromosomes and Cancer*, *47*(1), pp.8-20.

Stone, R.M., Mandrekar, S.J., Sanford, B.L., Laumann, K., Geyer, S., Bloomfield, C.D., Thiede, C., Prior, T.W., Döhner, K., Marcucci, G. and Lo-Coco, F., 2017. Midostaurin plus chemotherapy for acute myeloid leukemia with a FLT3 mutation. *New England Journal of Medicine*, *377*(5), pp.454-464.

Stone, R.M., Manley, P.W., Larson, R.A. and Capdeville, R., 2018. Midostaurin: its odyssey from discovery to approval for treating acute myeloid leukemia and advanced systemic mastocytosis. *Blood advances*, *2*(4), pp.444-453.

Stone, R.M., O'Donnell, M.R. and Sekeres, M.A., 2004. Acute myeloid leukemia. *ASH Education Program Book*, *2004*(1), pp.98-117.

Stratton, M.R., Campbell, P.J. and Futreal, P.A., 2009. The cancer genome. *Nature*, *458*(7239), pp.719-724.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), pp.15545-15550.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P. and Kuhn, M., 2015. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*, *43*(D1), pp.D447-D452.

Takei, H. and Kobayashi, S.S., 2019. Targeting transcription factors in acute myeloid leukemia. *International journal of hematology*, *109*(1), pp.28-34.

Tallman, M.S., Gilliland, D.G. and Rowe, J.M., 2005. Drug therapy for acute myeloid leukemia. *Blood*, *106*(4), pp.1154-1163.

Tang, J.L., Hou, H.A., Chen, C.Y., Liu, C.Y., Chou, W.C., Tseng, M.H., Huang, C.F., Lee, F.Y., Liu, M.C., Yao, M. and Huang, S.Y., 2009. AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations. *Blood, The Journal of the American Society of Hematology*, *114*(26), pp.5352-5361.

Taskesen, E., Bullinger, L., Corbacioglu, A., Sanders, M.A., Erpelinck, C.A., Wouters, B.J., van der Poel-van de Luytgaarde, S.C., Damm, F., Krauter, J., Ganser, A. and Schlenk, R.F., 2011. Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity. *Blood, The Journal of the American Society of Hematology*, *117*(8), pp.2469-2475.

Tremblay, G., Cariou, C., Recher, C., Dolph, M., Brandt, P., Blanc, A.S. and Forsythe, A., 2020. Cost-effectiveness of midostaurin in the treatment of newly diagnosed FLT3-mutated acute myeloid leukemia in France. *The European Journal of Health Economics*, *21*(4), pp.543-555.

Vaidya, A., 2014. Can systems biology approach help in finding more effective treatment for acute myeloid leukemia?. *Systems and synthetic biology*, *8*(2), pp.165-167.

Vassiliou GS, Cooper JL, Rad R, Li J, Rice S, Uren A, et al. Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice. Nat Genet. 2011;43:470–5.

Verma, M., Kulshrestha, S. and Puri, A., 2017. Genome sequencing. *Bioinformatics*, pp.3-33.

Voso, M.T., De Bellis, E. and Ottone, T., 2021. Diagnosis and Classification of AML: WHO 2016. In *Acute Myeloid Leukemia* (pp. 23-54). Springer, Cham.

Waldmann, T.A., 2003. Immunotherapy: past, present and future. *Nature medicine*, *9*(3), pp.269-277.

Wang, Y., Fan, S., Lu, J., Zhang, Z., Wu, D., Wu, Z. and Zheng, Y., 2017. GLUL promotes cell proliferation in breast cancer. *Journal of cellular biochemistry*, *118*(8), pp.2018-2025.

Warner, J.K., Wang, J.C., Hope, K.J., Jin, L. and Dick, J.E., 2004. Concepts of human leukemic development. *Oncogene*, *23*(43), pp.7164-7177.

Weinstein, L.S., Liu, J., Sakamoto, A., Xie, T. and Chen, M., 2004. Minireview: GNAS: normal and abnormal functions. *Endocrinology*, *145*(12), pp.5459-5464.

Wheeler, D.L., Iida, M. and Dunn, E.F., 2009. The role of Src in solid tumors. *The oncologist*, *14*(7), pp.667-678.

Wognum, A.W., Eaves, A.C. and Thomas, T.E., 2003. Identification and isolation of hematopoietic stem cells. *Archives of medical research*, *34*(6), pp.461-475.

Wong, J.W., Sullivan, M.J. and Cagney, G., 2008. Computational methods for the comparative quantification of proteins in label-free LCn-MS experiments. *Briefings in bioinformatics*, *9*(2), pp.156-165.

Wouters, B.J., Löwenberg, B., Erpelinck-Verschueren, C.A., van Putten, W.L., Valk, P.J. and Delwel, R., 2009. Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood, The Journal of the American Society of Hematology*, *113*(13), pp.3088-3091.

Xia, L., Sun, J., Xie, S., Chi, C., Zhu, Y., Pan, J., Dong, B., Huang, Y., Xia, W., Sha, J. and Xue, W., 2020. PRKAR2B-HIF-1α loop promotes aerobic glycolysis and tumour growth in prostate cancer. *Cell proliferation*, *53*(11), p.e12918.

Yu, J., Jiang, P.Y., Sun, H., Zhang, X., Jiang, Z., Li, Y. and Song, Y., 2020. Advances in targeted therapy for acute myeloid leukemia. *Biomarker research*, *8*(1), pp.1-11.

Yuan, J., Ju, Q., Zhu, J., Jiang, Y., Yang, X., Liu, X., Ma, J., Sun, C. and Shi, J., 2021. RASSF9 promotes NSCLC cell proliferation by activating the MEK/ERK axis. *Cell Death Discovery*, *7*(1), pp.1-11.

Zarrinkar, P.P., Gunawardane, R.N., Cramer, M.D., Gardner, M.F., Brigham, D., Belli, B., Karaman, M.W., Pratz, K.W., Pallares, G., Chao, Q. and Sprankle, K.G., 2009. AC220 is a uniquely potent and selective inhibitor of FLT3 for the treatment of acute myeloid leukemia (AML). *Blood, The Journal of the American Society of Hematology*, *114*(14), pp.2984-2992.

Zhang, Q., Wang, H.Y., Liu, X. and Wasik, M.A., 2007. STAT5A is epigenetically silenced by the tyrosine kinase NPM1-ALK and acts as a tumor suppressor by reciprocally inhibiting NPM1-ALK expression. *Nature medicine*, *13*(11), pp.1341-1348.

Zhang, Y. and Rowley, J.D., 2006. Chromatin structural elements and chromosomal translocations in leukemia. *DNA repair*, *5*(9-10), pp.1282-1297.

Zhao, S., Zhang, B., Zhang, Y., Gordon, W., Du, S., Paradis, T., Vincent, M. and von Schack, D., 2016. Bioinformatics for RNA-seq data analysis. *Bioinformatics—Updated Features and Applications: InTech*, pp.125-49.

Zhou, G., Soufan, O., Ewald, J., Hancock, R.E., Basu, N. and Xia, J., 2019. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic acids research*, *47*(W1), pp.W234-W241.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C. and Chanda, S.K., 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications*, *10*(1), pp.1-10.

Zhu, R., Lin, W., Tang, L. and Hu, Y., 2021. Identification of hub genes associated with adult acute myeloid leukemia progression through weighted gene co-expression network analysis. *Aging (Albany NY)*, *13*(4), p.5686.

Zito, A., Lualdi, M., Granata, P., Cocciadiferro, D., Novelli, A., Alberio, T., Casalone, R. and Fasano, M., 2021. Gene Set Enrichment Analysis of Interaction Networks Weighted by Node Centrality. *Frontiers in genetics*, *12*, p.221.