IDENTIFICATION OF MOLECULAR MARKERS FOR THE DIAGNOSIS OF LEUKEMIA

KALAIMATHI M
20PBI004

Under the guidance of
Dr.R.Nirmala devi
Assistant Professor

Thesis submitted to

DEPARTMENT OF BIOCHEMISTRY, BIOTECHNOLOGY AND BIOINFORMATICS

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

In Partial Fulfillment of the Requirements for the
Degree of Master of Science in Bioinformatics
May 2022

IDENTIFICATION OF MOLECULAR MARKERS FOR THE DIAGNOSIS OF LEUKEMIA

KALAIMATHI.M
20PBI004

Thesis submitted to

DEPARTMENT OF BIOCHEMISTRY, BIOTECHNOLOGY AND BIOINFORMATICS

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

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Bioinformatics

May 2022

30/5/2022

Signature of the Guide

30/5/2022

Signature of the Head of the Department

ACKNOWLEDGEMENT

I owe a special tribute to **GOD ALMIGHTY** for the opportunity given to me to takeup this work and also for inner strength to complete my work successfully.

I am grateful to **Dr. S. P. THYAGARAJAN,** Chancellor, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore for all the guidance and moral support for the study.

I am thankful to **Dr. V. BHARATHI HARISHANKAR**, Vice Chancellor, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore for the permission granted to complete the dissertation and the course successfully.

I am thankful to **Dr. S. KOWSALYA**, Registrar, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore for providing all the necessary facilities, support and encouragement during the course period.

I am thankful to **Dr. A. VIJAYALAKSHMI**, Dean, School of Biosciences, Avinashilingam Institute for Home Science and Higher Education for Women for the support and guidance during the study period.

I am grateful to **Dr. ANITHA SUBASH**, Professor and Head of the Department of Biochemistry, Biotechnology and Bioinformatics, Avinashilingam Institute for Home Science and Higher Education for Women, for her support and guidance to complete the project successfully.

I am humbled and grateful to **Dr. R. NIRMALADEVI**, Assistant Professor, Department of Biochemistry, Biotechnology and Bioinformatics, Avinashilingam Institute for Home Science and Higher Education for Women, I feel immense pleasure in extending my deep sense of gratitude for her motivation and encouragement throughout the work.

I am immensely grateful to my fellow researchers, **Ranjini B, Divya R and Udayadharshini S** for their help and support given throughout my period of study.

I am greatly indebted to all **Faculty members** of the Department of Biochemistry, Biotechnology and Bioinformatics, School of Biosciences, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore for their help and cooperation throughout the study.

I place my gratitude to foot of my parents for their immense support and guidance during the course of my study.

I express my sincere heart bound thanks to my friends, Department of Biochemistry, Biotechnology and Bioinformatics, for giving an effectionate advice, unconditional love and incredible support for the completion of my project work.

I would like to extend thanks to the many unseen people, who have so generously contributed to the completion of my research work.

KALAIMATHI M

S.No	CONTENTS	PAGE No
	LIST OF FIGURES	
	LIST OF TABLES	
1.0	INTRODUCTION	1
2.0	REVIEW OF LITERATURE	7
3.0	METHODOLOGY	25
4.0	RESULTS	32
5.0	SUMMARY AND CONCLUSION	73
	BIBLIOGRAPHY	76
	WEBOGRAPHY	83

LIST OF FIGURES

FIGURE.NO	TITLE	PAGE. NO
1	Pathway of CML	15
2	Pathway of CML	16
3	Volcano plot	35
4	Venn diagaram	37
5	BiologicalPathway of AML	42
6	BiologicalPathway of CML	43
7	Bar graph of enriched terms across input gene lists of AML	45
8	Summary of enrichment analysis in Transcription Factor Targets of AML	46
9	Bar graph of enriched terms across input gene lists of CML	47
10	Summary of enrichment analysis in Transcription Factor Targets of CML	48

11	**The protein-protein interaction of the DEGs of the AML**	**50**
12	**The protein-protein interaction of the DEGs of the CML**	**55**
13	**The cytoscape analysis of AML**	**59**
14	**The clusters formed using MCODE pluggin in cytoscape analysis of AML sample.**	**60**
15	**The clusters formed using MCODE pluggin in cytoscape analysis of AML sample.**	**61**
16	**The clusters formed using MCODE pluggin in cytoscape analysis of AML sample.**	**63**
17	**The clusters formed using MCODE pluggin in cytoscape analysis of AML sample.**	**64**
18	**The cytoscape analysis of CML**	**66**
19	**The clusters formed using MCODE pluggin in cytoscape analysis of CML sample.**	**66**
20	**The clusters formed using MCODE pluggin in cytoscape analysis of CML sample.**	**68**
21	**The clusters formed using MCODE pluggin in cytoscape analysis of CML sample.**	**69**
22	**The clusters formed using MCODE pluggin in cytoscape analysis of CML sample.**	**71**

LIST OF TABLES

TABLE.NO	TOPIC	PAGE.NO
1	List of genes involved in leukemia	9
2	DEGs of microaarray data	35
3	DIFFERENTIALLY EXPRESSED GENES of AML	38
4	DIFFERENTIALLY EXPRESSED GENES of CML	40
5	Summary of enrichment analysis in METASCAPE	48
6	Enrichment analysis in Transcription Factor Targets.	49
7	The data highlights the top enriched genes in STRING tool of AML	53
8	The data highlights the top enriched genes in STRING tool of CML	56

1

INTRODUTION

Cancer is the world's second largest cause of death. Every year, about 10 million new cancer patients are diagnosed, with over 6 million deaths as a result, accounting for roughly 12% of global mortality. Nearly in the next two decades, the number of new cancer cases is predicted to increase by almost 70%, reaching over 15 million new cases diagnosed yearly by 2020. Carcinogens, infections, genetic mutations, hormones, immunological disorders, and the adoption of behavioural and dietary risk factors, such as smoking, bad food, physical inactivity, and environmental pollution, have all contributed to this rapid increase of cancer (Rady *et al*., 2018).

Normal cells may be mutated as a result of the risk factors acting alone or in combination. Many of these mutations were caused uncontrolled cell division and also the cancer is caused by altering the expression or activity of important gene products. Surgery, radiation-based therapy, chemotherapy, gene therapy, and/or hormone therapy are currently the most common cancer treatment options, which can be used alone or in combination. Antimetabolites, DNA-interacting agents, antitubulin agents, hormones, and molecular targeting agents are the most often used chemotherapy medications. They all seek to eliminate malignant cells or inhibit their proliferation (Rady *et al*., 2018).

Leukemia is the major pediatric tumors accounting for about 30% of diagnoses. Leukemia is a major hematological malignancy that confers mortality and morbidity throughout different ages. It was estimated that there were around 350 000 new cases in 2012 worldwide. Leukemia diagnosis is challenged by different factors including lack of healthcare access and misclassification due to lack of experienced personnel. There are two predominant subtypes; acute lymphoblastic leukemia (ALL) and acute

myeloid leukemia (AML). AML is not as great, accounting for roughly 18% of juvenile leukemia diagnoses (Puumala *et al*., 2020).

Acute lymphocytic leukemia (ALL), or acute lymphoblastic leukemia, is a malignant disorder in which the neoplastic cells are derived from primitive hematopoietic progenitor cells. Over the last two decades, the prognosis for ALL has significantly improved, with current cure rates of 60–70 percent among youngsters. However, there has been less progress in the treatment of adult ALL (Leahy *et al*., 2021). Acute lymphoblastic leukemia (ALL) survival rates for children have grown considerably over the previous four decades, with 5-year survival rates of > 90% in recent trials. Successful treatment of ALL is linked to an increased risk of late death, second neoplasms, chronic health issues, endocrine dysfunction, and psychological function in this young population. Research efforts conducted through large survivor cohorts, such as the Childhood Cancer Survivor Study, are providing new and important insights into the very long-term consequences of ALL therapy, while providing direction for screening recommendations and intervention-based approaches for reducing late morbidity and mortality (Robison *et al*., 2021).

The second most prevalent leukemia in adults is acute myeloid leukemia (AML). Although the typical age of diagnosis is 67 years, with one-third of patients being 75 years or older, the elderly have limited treatment options, with 5-year survival rates of only 5%. A comprehensive evaluation was undertaken to assess the efficacy and safety of treatment regimens in elderly AML patients (60 years old)(Bell *et al*., 2018). Acute myeloid leukemia (AML) is a diverse group of diseases with different clinical, morphological, and molecular characteristics, according to the 2016 World Health Organization classification. A 7 + 3 intensive chemotherapy (IC) induction regimen has been the standard of care for fit patients for decades combining cytarabine and an anthracycline, followed by post-remission consolidation therapy often based on high dose cytarabine (HDAC) with or without allogeneic hematopoietic stem cell transplantation (alloHSCT) or, less frequently in recent years, autologous hematopoietic stem cell transplantation (Treatment of elderly AML patients who are declared unfit for intense therapy due to poor performance or co- morbidities is extremely tough (Cerrano *et al*., 2019).

CLL (chronic lymphocytic leukemia) is a cancerous tumor of the blood that is difficult to treat therapeutic based methods. Chemotherapy, radiation therapy, immunotherapy (such as interferon, anti-CD20 monoclonal antibodies, and anti-CD52 antibodies), and hematopoietic stem cell transplantation are the most common treatments, but relapse is common (Zhang *et al*., 2020). The mechanism of leukemogenic transformation is becoming more well-understood. A recent study of the genomic landscape of CLL in large cohorts found that the disease is frequently triggered by the loss or insertion of large quantities of chromosomal material, followed by further mutations that make the leukemia more aggressive (Hallek *et al*., 2019).

CML (chronic myeloid leukemia) is a myeloproliferative neoplasm characterized by dysregulated granulocyte production and uncontrolled proliferation with relatively normal differentiation. It is an uncommon kind of leukemia that affects adolescents, adults, and children are rarely affected. The average age of diagnosis for CML is 64 years. CML in adolescents and young adults has been the subject of few investigations. A study to look into the cohort and see if there are any differences between them and older CML patients. It accounts for 2% to 3% of all leukemia in children and adolescents. Epidemiologic statistics reveal a 0.6 to 1.2 per million child per year incidence that rises with age: CML is extremely rare in children, with just 0.7 per million children aged 1 to 14 years and 1.2 per million teenagers diagnosed each year. CML in children and teenagers is not the same as CML in adults (Hijiya *et al*., 2019; Yassin *et al*., 2019).

Bioinformatics is the science of combining computers, software tools, and databases to answer biological issues. For huge endeavours that create large data sets, bioinformatics methods are frequently applied. Genomic and proteomic research are two key large-scale bioinformatics applications. Bioinformatics tools aid in the comparison of genetic and genomic data, as well as the comprehension of molecular biology's evolutionary features in general. It aids in the analysis and cataloguing of biological pathways and networks that are important parts of systems biology on a more integrated level. Because genetic data has a context, bioinformatics is critical to genetic research. The setting is biological. The behaviour of living things follows a set

of rules. Tissues and cells, genes and proteins are all affected (https://www.scq.ubc.ca/what-is-bioinformatics).

The use of high-throughput platforms in Gene Expression (GE) has been more beneficial in clinical research in recent years, such as molecular classification, prognosis prediction, and new targeted medication development. Databases are repositories of information, generally store on computers. They contain an organized series of records that can be searched through and displayed in the computer screen, downloaded or emailed to a specific address (Li *et al*., 2019). Network-based modeling approaches offer the ability to better integrate and understand complex disease states and their causes. To date, network analysis has proven to be a reliable and cost-effective method for detecting early disease, predicting co-occurring diseases and relationships, and designing drugs. Despite the fact that integrated genomic analysis of methylation and gene expression in leukemia has been published, effective network analysis integration is still lacking (Sanchez *et al*., 2020).

Many gene expression profiling studies on leukemia revealed hundreds of DEGs implicated in many pathways, biological processes, and molecular functions. The DEG interaction network, on the other hand, has yet to be elucidated. Microarray datasets were evaluated to find DEGs between leukemia and normal cells in this study. The further development of leukemia analysis using DEG functional enrichment and interaction network analysis, as well as survival and mRNA-miRNA interaction analysis. A gene– transcription is built with a factor interaction network to discover important genes in leukemia (Li *et al*., 2019). Although studies have suggested that gene expression-based characteristics could be useful in CML, AML diagnosis and prognosis, due to a lack of validation cohorts, they have yet to be implemented into routine clinical care. Following the validation of large-scale gene expression datasets, a number of novel biomarkers have emerged. The creation of various gene expression profiles allows for the identification of accurate AML and CML biomarkers (Yao *et al*., 2022).

DEG interactions were investigated using the Search Tool for the Retrieval of Interacting Genes (STRING) database, and only validated interactions with a total score of > 0. 4 were considered significant. Then, using the Cytoscape software, the integration of protein-protein interaction (PPI) networks (Li *et al*., 2019). Using the

GEO database of cohorts and the immune score produced from the ESTIMATE algorithm, retrieved a list of microenvironment-related genes that indicated poor prognosis in AML patients and also created a risk rating method to determine the prognostic value of central genes. Furthermore, the relationship between the central gene and immune infiltration was explored (Ni *et al*., 2020). The link between the projected target among DEGs was analyzed using Cytoscape software, which helped to establish the regulatory network in relation to the development of Leukemia (Gao *et al*., 2019). Many GEO microarrays have been utilized in recent years to find differentially expressed genes (DEGs) in AML; a few of these genes are promising indicators for AML diagnosis and prognosis. However, most past research has focused on DEG screening rather than gene interactions (Tan *et al*., 2020). Data mining and integration could be valuable strategies for predicting and understanding the mechanisms of Leukemia disease onset and progression. Identifying new biomarkers that are employed in the diagnosis of leukemia may make early detection and treatment of leukemia easier (Gao *et al*., 2019).

Gene Ontology (GO), a key database for gene set enrichment analysis, can be compared to KEGG PATHWAY. GO is a set of gene function controlled vocabularies arranged into three ontologies: biological process, cellular component, and molecular functions. Many genomes have been annotated with GO terms as a result of community efforts, allowing GO enrichment analysis on a variety of gene sets. An enrichment study, for example, identifies GO terms that are over-represented for a group of human genes, allowing biological interpretation. KEGG pathway mapping, on the other hand, provides more information about how genes or gene products interact in pathways, although the gene coverage is less extensive than GO (Kanehisa *et al*., 2020).

The KEGG database resource was created with the goal of finding cellular and organism-level functions from large-scale molecular-level datasets, particularly entire genome gene sets. Route and MODULE are two databases that provide experimental knowledge of high-level functions extracted from published literature and represented as KEGG pathway maps and KEGG modules, respectively. The KEGG pathway map is a graphical representation of metabolic, signaling, and other molecular interaction/reaction networks that was created by hand (Kanehisa *et al*., 2020).The Molecular Complex Detection (MCODE) was used to filter modules of the PPI

network.GO and KEGG pathway analyses were used to perform functional enrichment and pathway enrichment analysis in the module (Li *et al*., 2019).

The discovery of tumour biomarkers is one of the methods for early detection of cancer, however the regulation of this extremely complex tumour cannot be met solely by a local emphasis on a single or specific gene. Some genes in tumours are abnormally expressed and intimately connected to many other genes, according to the entire regulatory network. Their expression could play a role in the occurrence and progression of malignancies (Xu *et al*., 2020). These genes become hub genes in the regulatory network hub. Large-scale omics data has been generated in recent years as a result of the emergence of high-throughput sequencing technologies such as gene chips and RNA-seq. A comprehensive research of the connection of numerous gene expressions will be made easier with more data analysis and mining (Xu *et al*., 2020). The objectives of the present study are as follows :

- To identify the differentially expressed genes in microarray dataset of AML and CML.
- To identify the molecular makers of leukemia using gene enrichment techniques.
- To analyze PPI (Protein Protein Interaction) using STRING and to construct hub gene clusters using MCODE in CYTOSCAPE.

2

REVIEW OF LITERATURE

Leukemia is the most frequent malignancy in children, accounting for around 30% of all diagnosis. Acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) are the two primary subtypes (AML). AML is a less frequent kind of juvenile leukemia, accounting for about 18% of all diagnosis. Based on cell lineage and epidemiological studies of incidence and risk factors, the aetiology of the two subtypes is likely substantially different. For childhood AML, only a few risk factors have been identified conclusively. AML is caused by the transformation of hematopoietic progenitor cells, which causes a halt in differentiation, resulting in an increase in progenitor cells and a decrease in mature blood cells. Backtracking investigations have shown mutations linked to children AML present at birth, but leukemia development occurs years later (Puumala *et al*., 2020).

2.1 Cancer.

2.2 Biological background of Leukemia.

2. 2.1 List of genes involved in leukemia.

2.3 Types of leukemia.

2.3.1 Acute lymphoid leukemia.

2.3.2 Chronic lymphoid leukemia.

2.3.3 Acute myeloid leukemia.

2.3.4 Chronic myeloid leukemia.

2.4 Pathway analysis of Leukemia.

2.4.1. Pathway of AML.

2.4.2. Pathway of CML.

2.5 Insilico analysis in leukemia.

2.5.1 Differential expression genes.

2.5.2 Affymatrix microarray.

2.5.3 GO and KEGG pathway analysis.

2.5.4 Protein-protein interaction (PPI) using STRING online database.

2.5.5Metascape and cytoscape analysis.

2.6 Diagnostic Biomarkers.

2.1 Cancer:

Cancer is a harmful disease that is characterised by irregular cell deliberation. Because civilization changes are the most common cause of cancer, there is a growing need to discover a better cure for the disease (World Health Organization *et al*., 2019). Non communicable diseases, such as cancer, have been identified as one of the most significant impediments to extending life expectancy in the twenty-first century. It was predicted that there would be an additional 18.1 million cancer cases worldwide in 2018.Leukemia is a malignancy that is caused by the abnormal growth of white blood cells and their precursors. Leukemia affects both children and adults and is caused by changes in the cell regulatory pathway, which leads to uncontrolled multiplication of bone marrow hematopoietic stem cells. Each year, nearly a hundred thousand new cases are recorded in the United States alone. Though white men are more likely to get leukemia, significant rates of leukemia can also be found in children in Asian countries like Sri Lanka (Sung *et al*., 2021).

In every country, cancer is a primary cause of death and a significant impediment to extending life expectancy. Cancer is the primary or second major cause of death before the age of 70 years in 112 of 183 nations, and it ranks third or fourth in another 23 countries, according to WHO projections for 2019. The ascent of cancer as a leading cause of death is partly owing to large reductions in the mortality rates of stroke and coronary heart disease in some countries when compared to cancer. Overall, the global burden of cancer incidence and death is quickly increasing, reflecting both population ageing and growth, as well as changes in the prevalence and distribution of the key cancer risk factors, many of which are linked to socioeconomic development. Comparing the maps, the latter exhibiting the 4-tier Human Development Index (HDI) based on the United Nations' 2019 Human Development Report, reveals the extent to which the position of cancer as a cause of premature death correlates national levels of social and economic development (Sung *et al*., 2021).

2.2 Biological background of Leukemia:

Leukemia is a type of hematologic malignancy that develops from stem cells in the bone marrow and is responsible for a substantial percentage of cancer fatalities. Many patients experience relapse/refractory disease despite treatment options such as standard chemotherapy, targeted therapy, and allogeneic stem cell transplantation (allo-HSCT), and overall survival is poor. Gene mutations create leukemic stem cells (LSCs), which go through an abnormal and poorly controlled proliferation process that plays a role in leukemia evolution, recurrence, and medication resistance (Shi *et al*., 2019).

2.2.1 List of genes involved in leukemia:

The following (Table 1) describes the prominent genes involved in various types of leukemia (http://www.bioinformatics.org/legend/leuk_db.html).

TYPE OF LEUKEMIA	**GENE NAME**
Acute Lymphoblastic Leukemia	MLLT2, MYC, ZNFN1A1,LAF4
Acute Myeloblastic Leukemia	ARNT
Acute Myelogenous Leukemia	IRFI, RGS2, GMPS
Acute Myeloid Leukemia	AF10, CBFB, NUP98, NUP214, HOXA9, REBBP, ARHGEF12, CDX2, LCP1, CEBPA, DEK,FUS, RUNX
Acute Promyelocytic Leukemia	PML, THRA, NPMI
Acute Undifferentiated Leukemia	SET
B-cell Chronic Lymphocytic Leukemia	BCL3, BTGI
B-cell (acute) Leukemia	PBXP1
Chronic Lymphocytic Leukemia	DLEU1, DLEU2
Chronic Myelocytic Leukemia	AXL
Murine Myeloid Leukemia	EVI2A, EVI2B
Myeloid Leukemia	CDC23, CLC
pre B-cell Leukemia	PBX1, PBX2, PBX3
Human Monocytic Leukemia	ETSI

2.3 Types of leukemia:

Pathogenesis, origin, incidence, and prognosis differ among leukemia kinds. The predominant leukemia cells can be mature cells, as in chronic lymphocytic leukemia (CLL), or precursor cells of diverse lineages, as in acute leukemia, or both precursor and mature cells, as in chronic myeloid leukemia (CML).Acute lymphoblastic leukemia (ALL) is the most common acute leukemia type in children and young adults, with an incidence peak between the ages of 2 and 5, whereas acute myeloid leukemia (AML) is the most common acute leukemia type in adults, accounting for 1.3 percent of new cancer cases in the United States. Furthermore, a group of leukemia forms that are infrequently diagnosed (e.g., atypical chronic myeloid leukemia) differ from the four basic varieties in many ways.

Leukemia occurs at varied rates in people of various ages, sexes, and races.The amounts of exposure to environmental and genetic risk factors are primarily responsible for such differences. For example, roughly 10% of people with CLL have a family history of the disease, and ionising radiation is a proven cause of paediatric ALL, as indicated by the tiny but significant increased risk caused by X-ray pelvimetry during pregnancy. The epidemiology of leukemia may alter over time and differ from country to country as a result of earlier efforts to combat the disease (Dong *et al.*, 2020).

2.3.1 Acute lymphoid leukemia:

ALL is defined by a malignant transformation of lymphoid progenitor cells in the bone marrow, blood, and extramedullary locations, as well as uncontrolled multiplication of these cells. ALL is the most common kind of leukemia in children, accounting for 76% of cases and boasting a 91% survival rate after therapy (Godoy *et al*., 2020). Acute lymphoid leukemia (ALL) is a haematological malignancy that affects the progenitor cells of B, T, and NK strains, with a higher frequency in children. Chromosomal abnormalities and genetic changes involved in the differentiation and proliferation of lymphoid progenitor cells characterise the pathophysiology of ALL. Despite a paucity of evidence in the literature, it is thought that leukemogenesis is the result of a complex interaction between environmental and hereditary factors, which results in cellular changes. Environmental factors have been

investigated as potential predisposing factors in the development of ALL, although the results in the literature are still mixed. The purpose of this review is to discuss the important external elements that affect ALL in this setting (Fujita *et al.*, 2021).

Acute leukemia is distinguished from chronic leukemia by the fact that the aberrant blood cells in acute leukemia are immature and unable to carry out normal functions. In chronic leukemia, on the other hand, the aberrant cells may continue to function normally. Leukemia is further divided into lymphoblastic and myeloid leukemia based on the damaged blood cells. The malignant changes in lymphocyte and red blood cell precursor cells, as well as other white blood cell types and platelets, occur in lymphoblastic and myeloid leukemia, respectively. Leukemia can be split into four subtypes based on the above categories: acute and chronic lymphoblastic leukemias (ALL, CLL), and acute and chronic myeloid leukemia (AML, CML).The most prevalent type of leukemia is acute myeloid leukemia, which has the greatest occurrences of new cases and fatalities in leukemia among the four subtypes (Heendeniya *et al*., 2020).

2.3.2 Chronic lymphoid leukemia:

CLL (chronic lymphocytic leukemia) is a type of leukemia that affects the bloodstream and is difficult to treat therapeutically. Chemotherapy, radiation therapy, immunotherapy (such as interferon, anti-CD20 monoclonal antibodies, and anti-CD52 antibodies), and hematopoietic stem cell transplantation are the most common treatments, however relapse is common . As a result, more research into the important molecular mechanisms in CLL development, as well as better focused and customised treatment, is required (Zhang *et al*., 2020).

The clonal growth and accumulation of mature, generally CD5-positive B-cells in the blood, bone marrow, lymph nodes, and spleen are hallmarks of chronic lymphocytic leukemia. It was recently shown that the ability to create clonal B cells in CLL can be acquired at the hematopoietic stem cell (HSC) stage, implying that multipotent, self-renewing HSCs are involved in the original leukemogenic event in CLL. The mechanism of leukemogenic transformation is becoming more well-understood. A recent study of the genomic landscape of CLL in large cohorts found that the disease is frequently triggered by the loss or insertion of large

quantities of chromosomal material, followed by further mutations that make the leukemia more aggressive (Hallek *et al*., 2019).

Because chronic lymphocytic leukemia is such a heterogenous disease, researchers are working hard to create increasingly accurate and exact analytics to predict future events including the need for and likelihood of responding to therapy, progression-free survival, and survival. Clinical and laboratory factors are often used to create these analytics. Some of these characteristics are incorporated into scores or staging systems that are prognostic (therapy-independent), while others are predictive (therapy-dependent).Distinct therapy have different predictive characteristics. Predictive biomarkers are particularly essential since responsiveness to therapy is a critical requirement for improving survival (Montserrat *et al*., 2019).

2.3.3 Acute myeloid leukemia:

Acute myeloid leukemia (AML) is a clonal haematopoiesis condition marked by the development of undifferentiated myeloid precursors, which results in dysregulated haematopoiesis and bone marrow failure.AML is the most frequent form of acute leukemia, and it has a wide range of molecular profiles, clinical manifestations, therapeutic responses, and prognosis. Despite the fact that rigorous chemotherapy and haematopoietic stem cell transplantation (HSCT) have improved AML patient outcomes, refractory disease, recurrence, and treatment-induced mortality affect around 50% of younger patients and 80% of individuals over the age of 60. Patients with advanced AML who are unable to get intense chemotherapy due to side effects have a median life of only 5-10 months (Wong *et al*., 2019).

Acute myeloid leukemia (AML) is the most common malignant myeloid disorder in adulthood, still incurable in most patients, with a 35–40% five-year overall survival (OS) rate for non-promyelocytic AML. Many blast clones have been found in samples from a single patient, indicating that the disease is highly diverse both among patients and within the same patient. As a result, AML poses a significant hurdle for tailored therapy. Many biomarkers guiding therapy strategy have been identified as a result of the combination of modern diagnostic technologies such as genome-wide molecular profiling, immunophenotyping, cytogenetic, and clinical characteristics. Despite the

decisive contribution of currents prognostic markers, all these combining approaches fail to tackle disease complexity and outcome(Takam Kamga *et al*., 2019).Acute myeloid leukemia (AML) is a well-known heterogeneous haematological malignancy with a wide prognosis that is influenced by clinical variables, cytogenetics, and molecular features. Recurrent genetic anomalies categorise AML into three risk status categories: low-risk, intermediate-risk, and high-risk, according to the current risk classification system (Du *et al*., 2019).

Acute myeloid leukemia (AML) is the most frequent acute leukemia in adults, with only a 35–40 percent five-year overall survival rate for non-promyelocytic AML. Despite this, accurate biomarkers for early diagnosis and prediction of chemotherapy or allogeneic transplantation response are still missing. Different components of the bone marrow stromal niche enhance the persistence of residual leukemia cells during chemo- or immunotherapy through a redundant panel of molecular pathways, such as Wnt, Notch, and Hedgehog, which support cell survival, proliferation, and chemoresistance. However, the relevance of these pathways in the interaction of AML cells with the bone marrow microenvironment is yet unknown (Takam Kamg *et al*., 2020).

2.3.4 Chronic myeloid leukemia:

Chronic myeloid leukemia (CML) is an uncommon kind of leukemia that affects children and adolescents. It accounts for 2% to 3% of all leukemia in children and adolescents. Epidemiologic statistics reveal a 0.6 to 1.2 per million child/year incidence that rises with age: CML is extremely rare in children, with just 0.7 per million children aged 1 to 14 years and 1.2 per million teenagers diagnosed each year. CML in children and teenagers is not the same as CML in adults. 1,3 Children and adolescents with CML have a higher proportion of mutated cancer driver genes (Hijiya *et al*., 2019).

CML is a myeloproliferative neoplasm characterised by dysregulated granulocyte production and uncontrolled proliferation with relatively normal differentiation. It mostly affects adults, and children are rarely affected. CML is diagnosed at an average age of 64 years. CML in adolescents and young adults has been the subject of few investigations. The biggest cohorts of CML-affected

adolescents and young adults in Qatar. The goal of this study is to look into this cohort and see if there are any differences between them and older CML patients (Yassin *et al*., 2019). Chronic myeloid leukemia (CML) is a myeloproliferative neoplasm that affects 1-2 people out of every 100,000. It is responsible for about 15% of newly diagnosed instances of leukemia in adults. In the United States, around 8450 new CML cases are expected to be diagnosed in 2020, with approximately 1080 individuals dying from CML. Annual mortality in CML has fallen from 10% to 20% to 1% to 2% since the introduction of imatinib in 2000. As a result, the prevalence of CML in the United States (US) has increased by roughly 7000-8800 cases per year since 2000, to an expected 150-180 000+ cases per year in 2020, based on an incidence of 5000 cases per year in 2020 (Jabbour *et al*., 2020).

2.4 Pathway analysis of Leukemia:

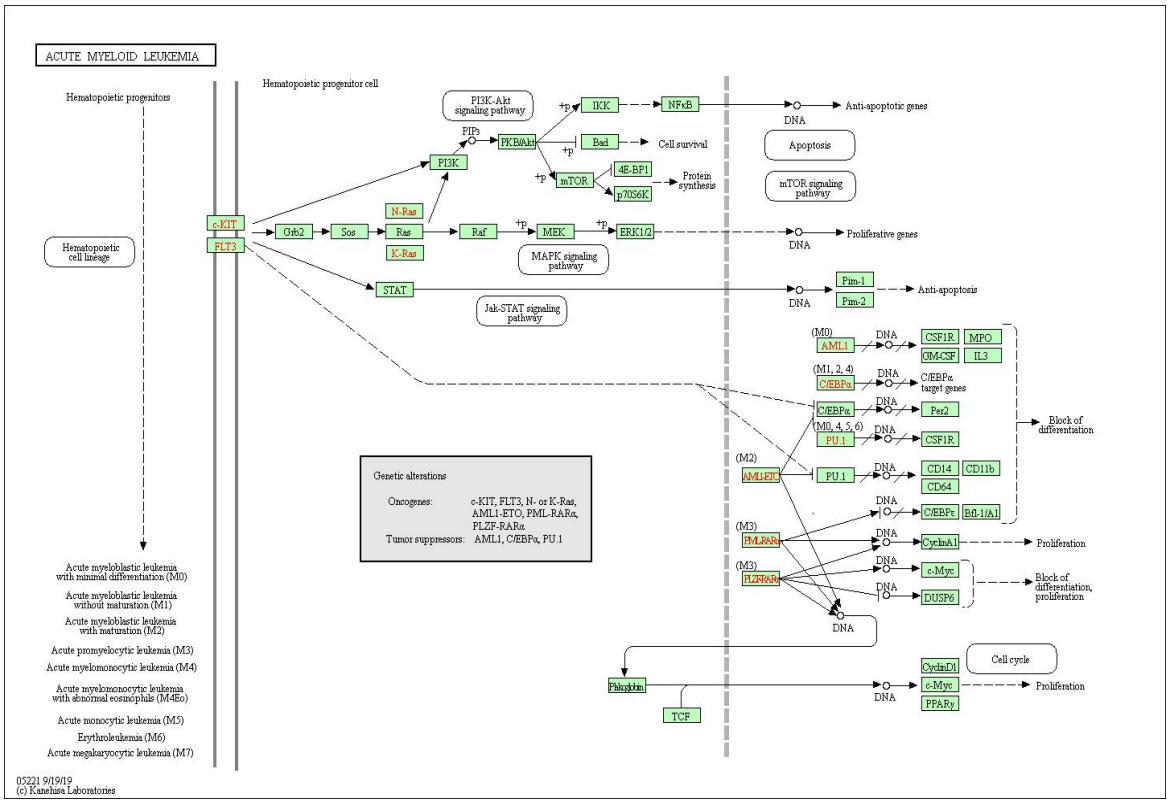
A limited number of self-renewing cells called leukemia stem cells (LSCs) initiate and maintain the leukemia pathway, which share features with hematopoietic stem cells (HSCs), the self-renewing cells that create healthy blood cells (A Pathway to Leukemia 2010). The routes are taken from the KEGG database (Kyoto Encyclopedia of Genes and Genomes Genome Net). KEGG PATHWAY is a set of hand-drawn pathway maps that represent our understanding of molecular interaction, reaction, and relation networks in the areas of Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, Organismal Systems, Human Diseases, and Drug Development. KEGG PATHWAY is a Pathway Mapping reference database (https://www.genome.jp/kegg/pathway.html). The AML and CML pathways are described as follows among the various kinds of leukemia.

2.4.1. Pathway of AML:

AML is a disorder marked by uncontrolled proliferation of clonal neoplastic cells and the buildup of blasts with a faulty differentiation programme in the bone marrow. AML accounts for over 80% of all adult leukemia and is still the leading cause of leukemia death. There are two basic types of genetic events that are required for leukemic transformation. Disordered cell development and overexpression of cell survival genes are thought to be a crucial initial step. These activating events were most frequently found in the RTKs Flt3, N-Ras, and K-Ras, as well as Kit and infrequently in other RTKs. A second process required for leukemogenesis is changes

in myeloid transcription factors that control hematopoietic differentiation.By suppressing target genes, transcription factor fusion proteins like AML-ETO, PML-RAR alpha, or PLZF-RAR alpha prevent myeloid cell development. In some circumstances, the transcription factors have been altered (http://www.kegg.jp/dbget-bin/www_bget?hsa05221).

FIGURE:1Pathway of AML

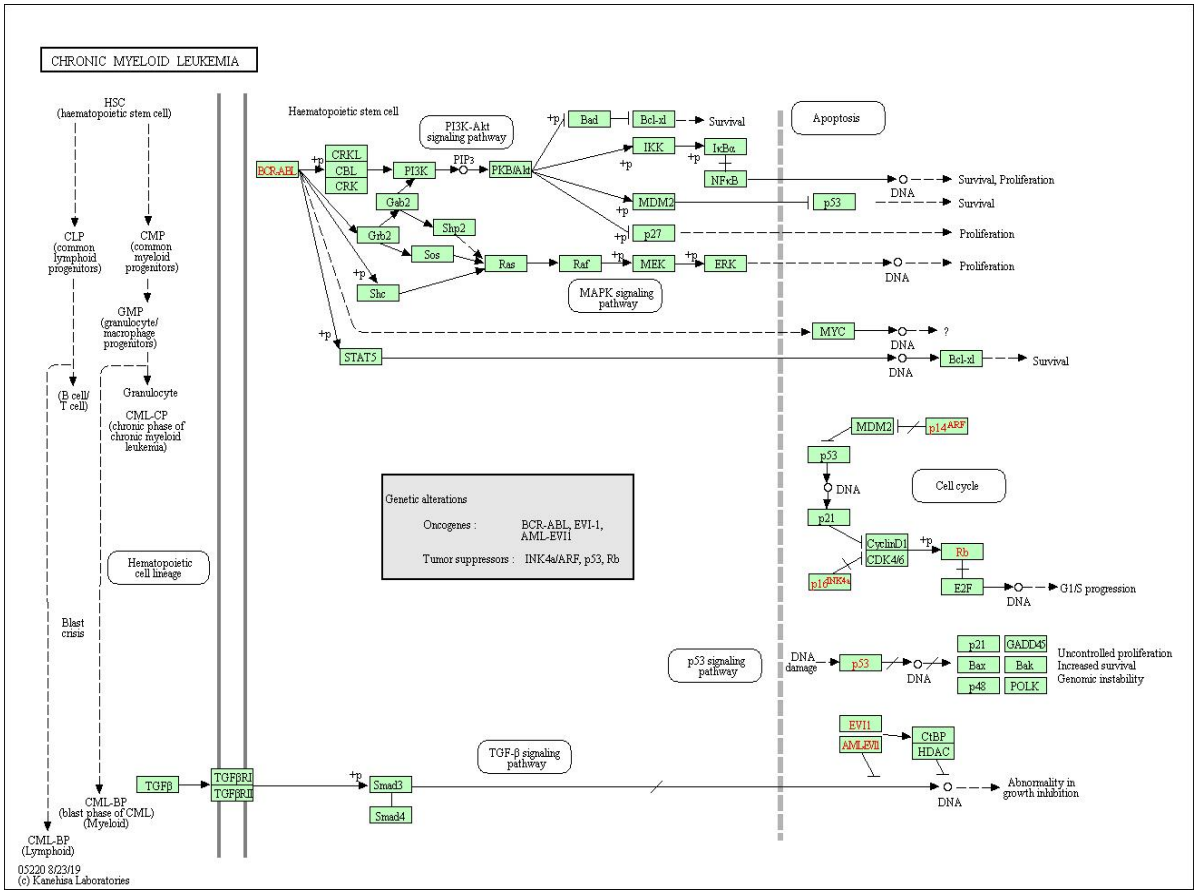


2.4.2. Pathway of CML:

A clonal myeloproliferative disease of a pluripotent stem cell is chronic myeloid leukemia (CML). CML has a three-phase natural history, with an initial chronic phase (CP) marked by the proliferation of functionally normal myeloid cells, followed by an accelerated phase (AP), and lastly a more aggressive blast phase (BP) marked by the loss of terminal differentiation potential. CML is linked to a specific chromosome anomaly, the t(9; 22) reciprocal translocation that gives rise to the Philadelphia (Ph) chromosome, at the cellular level. A molecular rearrangement between the c-ABL proto-oncogene on chromosome 9 and the BCR (breakpoint cluster region) gene on chromosome 22 results in the Ph chromosome. The BCR/ABL

fusion gene produces p210 BCR/ABL, an oncoprotein with constitutive tyrosine kinase activity and cytoplasmic localization, unlike the normal p145 c-Abl. While the fusion of c-ABL and BCR is thought to be the fundamental cause of the chronic phase of CML, other molecular alterations are required to move to blast crisis. Mutations in TP53, RB, and p16/INK4A, as well as overexpression of genes like EVII, are common secondary abnormalities. Additional chromosomal translocations, such as t(3;21)(q26:q22), are also seen, resulting in AMLI-EVII (https://www.kegg.jp/dbget-bin/www.bgethsa05221).

FIGURE:2Pathway of CML



2.5 Insilico analysis in leukemia:

Network-based modelling approaches offer the ability to better integrate and understand complex disease states and their causes. To date, network analysis has proven to be a reliable and cost-effective method for detecting early disease, predicting co-occurring diseases and relationships, and designing drugs. Despite the

fact that integrated genomic analysis of methylation and gene expression in leukemia has been published, effective network analysis integration is still lacking**.** Protein-protein interaction networks (PPI) are a type of network that focuses solely on protein-protein interactions and the resulting cell activities. A PPI network is a (un)directed graph/network with vertices representing proteins (or protein-coding genes) and edges representing their interactions. Two proteins are related by an edge if they are jointly contributing to a shared function, which does not necessarily imply a physical binding connection (Sanchez *et al*., 2020).

Although studies have suggested that gene expression-based characteristics could be useful in CML,AML diagnosis and prognosis, due to a lack of validation cohorts, they have yet to be implemented into routine clinical care. Following the validation of large-scale gene expression datasets, a number of novel biomarkers have emerged. The creation of various gene expression profiles allows for the identification of accurate AML,CML biomarkers (Yao *et al*., 2022).

2.5.1 Differential expression gene:

The use of high-throughput platforms in gene expression (GE) has been more beneficial in clinical research in recent years, such as molecular classification, prognosis prediction, and new targeted medication development. Many gene expression profiling studies on leukemia revealed hundreds of DEGs implicated in many pathways, biological processes, and molecular function. The DEG interaction network, on the other hand, has yet to be elucidated. Microarray datasets were evaluated to find DEGs between leukemia and normal cells in this study. The further development of leukemia analysis using DEG functional enrichment and interaction network analysis, as well as survival and mRNA-miRNA interaction analysis.A gene–transcription is built with a factor interaction network to discover important genes in leukemia (Li *et al*., 2018).

GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r) can be used to compare DEGs between AML and non-cancer patients. GEO2R is a web-based interactive tool that lets users to compare two or more GEO datasets to find DEGs under experimental conditions. The differential genes of each dataset were retrieved first, followed by the intersections, and the differentially expressed genes were discovered in various gene sets (Fu *et al*., 2020).

High-throughput sequencing and microarrays of human disease samples have recently created a large amount of bioinformatics data, which has aided in understanding the molecular pathways involved in the biological process. The shared differentially expressed genes (DEGs) from two gene expression profiles of GC-sensitive and GC-resistant ALL samples were identified and integrated in this work. Following that, we used in vitro cell tests to confirm the bioinformatics findings. Individual experimental data from prior studies are combined in our findings to depict the landscape of biological mechanisms that lead to the acquisition of GC resistance. Novel therapy options and therapeutic targets can be predicted based on these findings (Chen *et al*., 2020).

Many GEO microarrays have been utilised in recent years to find differentially expressed genes (DEGs) in AML; a few of these genes are promising indicators for AML diagnosis and prognosis. However, most past research has focused on DEG screening rather than gene interactions. Using microarray or RNA (ribonucleic acid) sequencing data, weighted gene co-expression network analysis (WGCNA) can be utilised to find correlations between DEGs. WGCNA identifies gene modules that are highly linked and ties them to clinical characteristics. Network-based gene screening of prospective biomarkers or therapeutic targets is made easier by correlation networks (Tan *et al*., 2020).

Microarray data analysis is a relatively recent technology that intends to assist in the correct medical diagnosis of a large number of genes across many experimental circumstances in order to find the right treatment for numerous diseases. Because of the expensive and sophisticated nature of microarray datasets, it is difficult to anticipate and thus necessitates rigorous testing with proper statistical tools in order to produce useful results. Gene expression is a well-known mechanism that translates a gene's DNA sequence to its corresponding mRNA sequences, and then to amino acid sequences in proteins (Panda *et al*., 2020).

Microarray data analysis is a powerful tool with many applications provides gene expression profiling to characterise the expression levels of hundreds of thousands of genes in cells that are connected with the corresponding protein, which aids in the understanding of the cellular mechanism of biological processes. Data mining is used to extract meaningful observations from large and complex microarray gene expression datasets, such as those used in post-genome cancer diagnostics, to learn more about how genes are regulated, how genes influence cancerous cell

mutation, and how performance varies depending on various medical experimental conditions, to name a few. The microarray dataset lists samples or conditions in rows and the genes that they correspond to in columns (Panda *et al*., 2020).

Gene classifications try to figure out which genes are expressed and which are controlled. This technique has the potential to help with disease prognosis as well as the development of algorithms to detect and forecast diseases based on genetic profiles. The method also addresses the problem of a small number of relevant features in groups of similar data. It is best to locate acceptable elements throughout this area. To minimise data dimensionality, decisive stages are required. Because not enough DNA is necessary to identify specific types of malignant cells, it is difficult to do so (Jahwar *et al*., 2021)

The mechanisms and altered gene profiles implicated in AML chemo-resistance are still unknown. Several subgroups of AMLs treated with varied regimens appear to share regulatory genes or pathways relevant to their chemo-resistance phenotype. Down-regulated genes and pathways linked to relapse following treatment were studied in AML samples using a gene set enrichment approach. To find highly differentially expressed genes between chemo-resistant and chemo-sensitive groups, researchers examined five AML libraries generated from GEO and ArrayExpress sources. To examine biological pathways related to AML chemotherapeutic resistance, functional and pathway enrichment analysis of differentially expressed genes was undertaken (Modarres *et al*., 2021).

2.5.2 Affymatrix microarray:

The Affymetrix GeneChip system was developed through microarray studies. The Aflymetrix probes were based on publically available data. The probe sets' sequences were chosen from GenBank, dbEST, and RefSeq. The UniGene database (Build 133, April 20, 2001) was used to build the sequence clusters, which were subsequently refined by analysis. A number of other publicly available databases were used in the comparison procedure, including the Washington University EST trace repository and the University of California, Santa Cruz Golden-Path human genome database (April 2001 release). These databases' sequences were gathered and grouped into groupings of comparable sequences. Photolithography (the use of light to control the creation of several layers of material) is used to make the probes on the chip, a

technique borrowed from the computer chip industry. Each GeneChip has about 1,000,000 characteristics on it. Each probe is found in pairs, with one being a perfect match (PM) and the other having a mismatch (MM) in the middle. Non-specific cross hybridization signals can be quantified and subtracted using these probe pairs. The intensity ratios and discrepancies in hybridization signals between the partners serve as markers of specific target abundance. On the GeneChip, eleven probe pairs describe each gene or transcript.To describe their uniqueness and/or ability to bind distinct genes or splice variants, the probe sets are given different suffixes (Xu *et al*., 2020).

2.5.3 GO and KEGG pathway analysis:

Metascape.org was used to enrich genes for GO biological activities and pathways. All genes in the genome are used as the enrichment background for the gene list provided to metascape.org in pathway and process enrichment analysis. On the basis of membership similarities, terms with a P value less than 0.01, a minimum count of 3, and an enrichment factor larger than 1.5 (defined as the ratio of observed counts to counts expected by chance) are collated and categorised into clusters. STRING was used to look into functional protein-protein interaction networks (Lai *et al*., 2020).

Gene Ontology (GO), a key database for gene set enrichment analysis, can be compared to KEGG PATHWAY. GO is a set of gene function controlled vocabularies arranged into three ontologies: biological process, cellular component, and molecular function. Many genomes have been annotated with GO terms as a result of community efforts, allowing GO enrichment analysis on a variety of gene sets. An enrichment study, for example, identifies GO terms that are over-represented for a group of human genes, allowing biological interpretation. KEGG pathway mapping, on the other hand, provides more information about how genes or gene products interact in pathways, although the gene coverage is less extensive than GO.

The KEGG database resource was created with the goal of finding cellular and organism-level functions from large-scale molecular-level datasets, particularly entire genome gene sets. Route and MODULE are two databases that provide experimental knowledge of high-level functions extracted from published literature and represented as KEGG pathway maps and KEGG modules, respectively. The KEGG pathway map

is a graphical representation of metabolic, signalling, and other molecular interaction/reaction networks that was created by hand (Kanehisa *et al*., 2020).

2.5.4 Protein-protein interaction (PPI) using STRING online database:

Protein–protein interactions (PPIs) are involved in a variety of biological activities, including metabolic and signalling pathways, cellular functions, and the immune system. Protein mutations can impact protein folding and stability, altering the kinetics and thermodynamics of PPIs as a result. Such mutations can be beneficial to the organism through evolution, or they can be harmful and cause a disease phenotype. Understanding the impact of these mutations, particularly the conformation and thermodynamic alterations of the interacting proteins, is crucial for a variety of medicinal applications, such as disease-associated mutation studies, drug design, and therapeutic intervention (Zhou *et al*., 2020).

The huge amount of experimental PPI data being created on a regular basis has necessitated the creation of computer-readable biological databases to organise and process it. The biomolecular interaction network database (BIND), for example, is built on a specification system that allows for a detailed description of how the PPI information was acquired through experimentation, frequently with direct links to the concluding evidence from the literature. The database of interacting proteins (DIP) is another collection of protein-protein binary interactions found via testing. The BioGRID database (Biological General Repository for Interaction Datasets) covers protein and genetic interactions between thirteen different species. Interactions are added to the databases on a regular basis after thorough vetting of the main literature. Other databases, such as BIND and MIPS (Munich Information Center for Protein Sequences), as well as large-scale research, are used to extract interaction data. Hit-Predict is a database of high-confidence protein-protein interactions, from which we may obtain the total number of interactions for a protein in a given species and display all of the interactions with confidence scores (Rao *et al.*, 2014).

Because a PPI is made up of two or more proteins, the total number of PPIs (interactome) is substantially greater than the human protein target space, which is believed to be around 19,000 proteins. The Interactome Database at the Center for Cancer Systems Biology (CCSB) contains 13,993 human PPIs and 3,787 viral

protein-human protein interactions that were discovered through research. There are 64,006 binary interactions in the Human Reference Interactome (HuRI). The STRING database contains 505,116 human PPIs that have been experimentally determined or predicted with high confidence. The size of the human interactome was then estimated to be around 130,000 by comparing PPIs found using the yeast-two-hybrid approach with known human PPIs. Using a statistical approach that calculates the size of the complete network from sub-network data, the size of the interactome was estimated to be 650,000 (Shin *et al*., 2020).

Using the STRING database and chosen enriched genes, the Protein-Protein Interaction network was created. The STRING database contains information on both known and predicted protein-protein interactions. Using the String database and settings of 0.4 and higher, the selected genes were utilised to build a PPI network (Kumar *et al*., 2020).

2.5.5 Metascape and cytoscape analysis :

Process and pathway enrichment studies of genes co-expressing with SCAMP2/4/5 in leukemia were performed using Metascape (http://metascape.org). Metascape offers multi-platform studies of multiomic data as an effective, efficient, and user-friendly gene-list analysis tool. Only terms with a p value of less than 0.01, a minimum count of three, and an enrichment factor more than 1.5 were considered significant. BioGRID, InWeb IM, and OmniPath were used to build the core protein–protein interaction (PPI) network, and Metascape was used to show the results. Then, using parameters such as degree cutoff = 2, node score cutoff = 0.2, K-score = 2 and MAX depth = 100, MCODE (Minimal Common Oncology Data Elements) cluster analysis was used to discover critical MCODE clusters in the PPI network. Furthermore, Metascape was used to estimate the functions of the most significant modules from the PPI network with a value of p 0.05 (Yue *et al*., 2021). Metascape (http://metascape.org) is a web-based integrated portal that integrates membership search, gene annotation, interactome analysis,Over 40 different knowledge bases are used for functional enrichment, including GO and KEGG analyses. GO and KEGG enrichment studies using molecules and their relevant genes were done in the Metascape database using the selected similar expression genes

strongly associated with each molecule found using the GEO dataset (Zhang *et al.*, 2021) Then, using parameters such as degree cutoff = 2, node score cutoff = 0.2, K-score = 2 and MAX depth = 100, MCODE (Minimal Common Oncology Data Elements) cluster analysis was used to discover critical MCODE clusters in the PPI network. Furthermore, Metascape was used to estimate the functions of the most significant modules from the PPI network with a value of p 0.05 (Yue *et al*., 2021).

2.6 Diagnostic Biomarkers:

Biomarkers were discovered at many levels, including genomes, transcriptomics, and proteomics, may lead to improved treatment for malignancies such chronic lymphocytic leukemia (CLL).CLL is a type of lymphoid cancer that is defined by the aggregation of mature B cells.MiRNAs have emerged as new diagnostic and therapeutic biomarkers in individuals with CLL illness, with other biomarkers such as CD38, chromosomal abnormalities, ZAP-70, TP53, and microRNA [miRNA] (Mirzaei *et al*., 2018).

In the "finding phase" for cancer-related biomarkers, microarray-based expression analysis has been widely used. The goal of exploratory studies in the discovery phase, according to the National Cancer Institute's Early Detection Research Network, is to create a short list of 1 to 10 high-priority candidates. A short list is essential since target validation takes time, money, and effort, and a small set is ideal for the development of low-cost diagnostic tests. Choosing the "best" markers from thousands of genes, on the other hand, is a difficult undertaking. The advantages of combining biomarkers to predict outcome have been well recognised in the literature (Oehler *et al*., 2019).

The three main kinds of extremely diverse haematological malignancies derived from myeloid and lymphoid cell lineages are leukemia, lymphoma, and multiple myeloma .Acute myeloid leukemia (AML) is characterised by the aberrant increase and accumulation of immature myeloid cells in the bone marrow and blood, interfering with normal cellular growth.AML is a cancer that is extremely aggressive and has a terrible prognosis.In adults, it is also the most frequent kind of acute leukemia.Treatment techniques and success rates differ depending on a variety of circumstances, including the patient's AML subtype, prognostic factors, age, and overall health (Turk *et al*., 2020).

The discovery of tumour biomarkers is one of the methods for early detection of cancer, however the regulation of this extremely complex tumour cannot be met solely by a local emphasis on a single or specific gene. Some genes in tumours are abnormally expressed and intimately connected to many other genes, according to the entire regulatory network. Their expression could play a role in the occurrence and progression of malignancies. These genes become hub genes in the regulatory network hub. Large-scale omics data has been generated in recent years as a result of the emergence of high-throughput sequencing technologies such as gene chips and RNA-seq. A comprehensive research of the connection of numerous gene expressions will be made easier with more data analysis and mining (Xu *et al*., 2020).

CYTOSCAPE analysis, which is a scientific approach for describing the pattern of gene association between diverse samples. Based on the intrinsicality of gene sets, it can be utilised to find very synergistically changed gene sets. Candidate biomarker genes or therapeutic targets are identified by association with gene sets and phenotypes. It uses thousands or almost 10,000 of the most variable genes, or all of the genes, to identify the set of genes of interest and perform a meaningful association analysis with a given phenotype, as opposed to genes that just focus on differential expression. As a result, the results acquired using this method a/re more biologically significant and reliable.

3

METHODOLOGY

The present study aims to identify new targets for the diagnosis and treatment of AML (acute myeloid leukemia) and CML (chronic myeloid leukemia).The GSE140315 (GPL11180 [HT_MG-430_PM] Affymetrix HT MG-430 PM Array Plate, GPL16417 Illumina MiSeq (*Mus musculus*) of AML and GSE44589 (GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) of CML were selected from the database of the Gene Expression Omnibus and compared using the GEO2R tool to identify differentially expressed genes (DEGs). GEO2R is an interactive web tool that allows users to compare two or more groups of Samples in a GEO Series in order to identify genes that are differentially expressed across experimental conditions. The GSE140315 and GSE44589 were selected from the database of the Gene Expression Omnibus and compared using the GEO2R tool to identify differentially expressed genes (DEGs). Also the common DEGS of all the samples of leukemia was analyzed for their PPIs. The major modules from the PPI network were screened using the Molecular Complex Detection (MCODE) tool in the cytoscape. The methodology of this study includes the retrieval of microarray data from Gene Expression Omnibus(GEO) (https://www.ncbi.nlm.nih.gov/geo/). *In silico* methods were used to determine the differentially expressed gene (DEG) in the dataset. Metascape database (http://metascape.org) is used to perform GO and KEGG pathway enrichment analyses. STRING (https://string-db.org/) is used to construct a protein-protein interaction (PPI) network. The network was visualized in Cytoscape software. The hub genes from the PPI network were found using CytoHubba, a cytoscape tool. The Molecular Complex Detection (MCODE) is used to identify the merged network's hub genes. Hub genes with a high degree of connectivity in the PPI network are significantly enriched in positive regulation of transcription. These genes play key roles in the characteristics and the development of the disease and they serve as diagnostic markers for leukemia.

3.1 Data repositories

3.2 Gene Expression Omnibus (GEO)

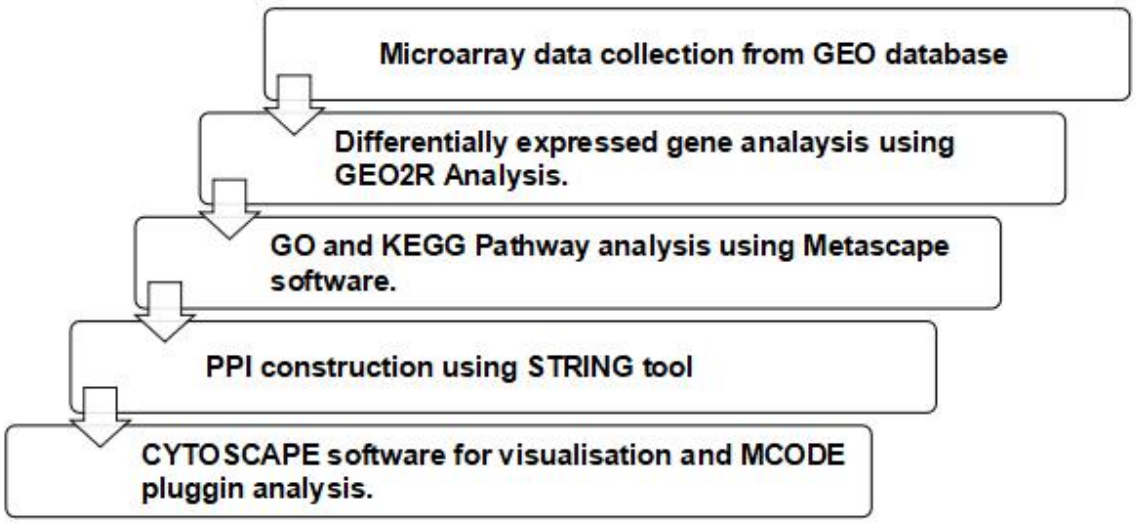
3.3 Microarray data retrieval

3.4 Differentially expressed gene (DEG) using GEO2r analysis

3.5 Metascape analysis

3.6 PPI construction using STRING

3.7 CYTOSCAPE analysis.



3.1 Data repositories:

Databases are repositories of information, generally stored on computers. They contain an organized series of records that can be searched through and displayed in the computer screen , downloaded or emailed to a specific address. A microarray database can be a repository containing microarray gene expression data. Microarray data repositories are large collections of data that are implemented from different array experiments to serve the research community. There are several repositories maintaining huge amount of microarray expression data providing open source access to retrieve and evaluate gene expression. A peer reviewed, public repository that adheres to educational or trade standards and is intended to be utilized by several

analytical applications. A good example of this is the Gene Expression Omnibus (GEO) from NCBI.

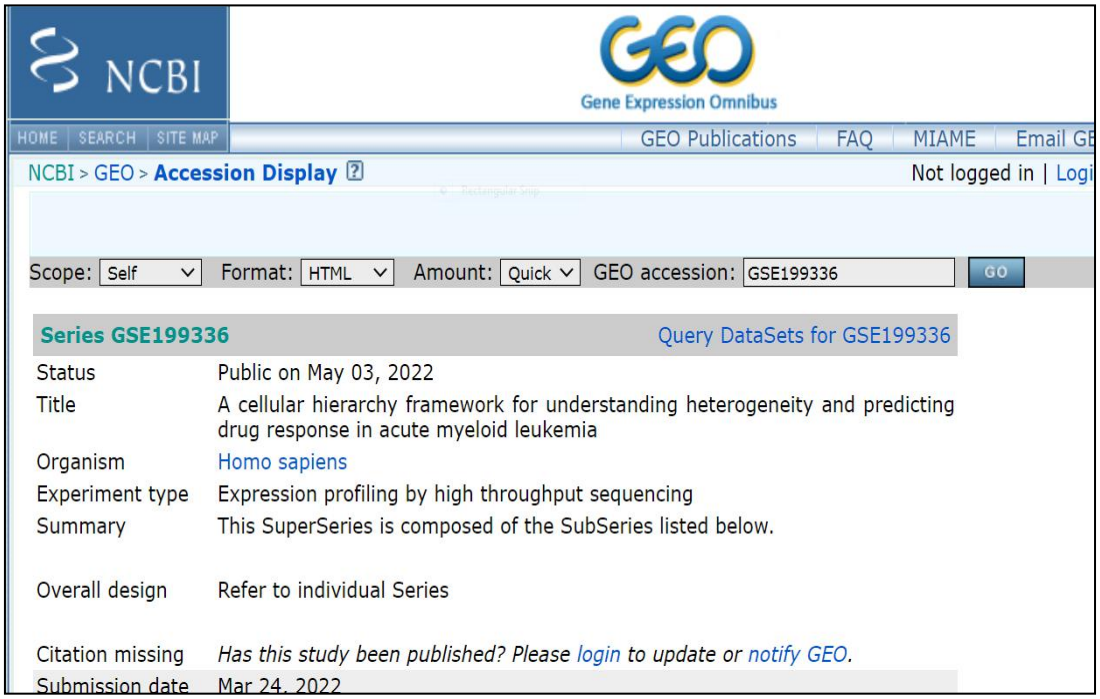
3.2 Gene Expression Omnibus (GEO):

GEO (https://www.ncbi.nlm.nih.gov/geo/) is a public functional genomics data information repository supporting MIAME (Minimum Information About a Microarray Experiment) compliant data information submissions. GEO is an international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high -throughput functional genomics data submitted by the research community. The gene expression omnibus maintained by NCBI-National Center for Biotechnology Information (http://www.ncbi.nih.gov/geo/) is one of the largest and most widely used database. Gene expression omnibus is a tool in NCBI where the data was taken. Each sample has its own accession number. The other format is not accepted by the tool. Gene expression omnibus consist of samples in tab delimited format which either opens in note pad or Microsoft excel. Microarray data helps especially in the identification of single-nucleotide polymorphisms (SNPs) and mutations, classification of tumors, identification of target genes of tumor suppressors, identification of cancer biomarkers, identification of genes associated with chemo-resistance, and drug discovery.

3.3 Microarray data retrieval:

Gene Expression Omnibus (GEO) is an online database which provides comprehensive data on gene profiling and sequencing. A microarray database is a repository containing microarray gene expression data. The key uses of a microarray database are to store the measurement data, manage a searchable index, and make the data available to other applications for analysis and interpretation. In GEO database (https://www.ncbi.nlm.nih.gov/geo/), retrieval of the gene expression profile dataset GSE140315andGSE44589. The GPL11180 [HT_MG-430_PM] Affymetrix HT MG-430 PM Array Plate, GPL16417 Illumina MiSeq (*Mus musculus*), the microarray data set GSE140315containing 8 samples and GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, the microarray data set

GSE44589containing 198 samples. These microarray data sets were used to perform GEO2r analysis.



3.4 Differentially expressed gene (DEG) using GEO2r analysis:

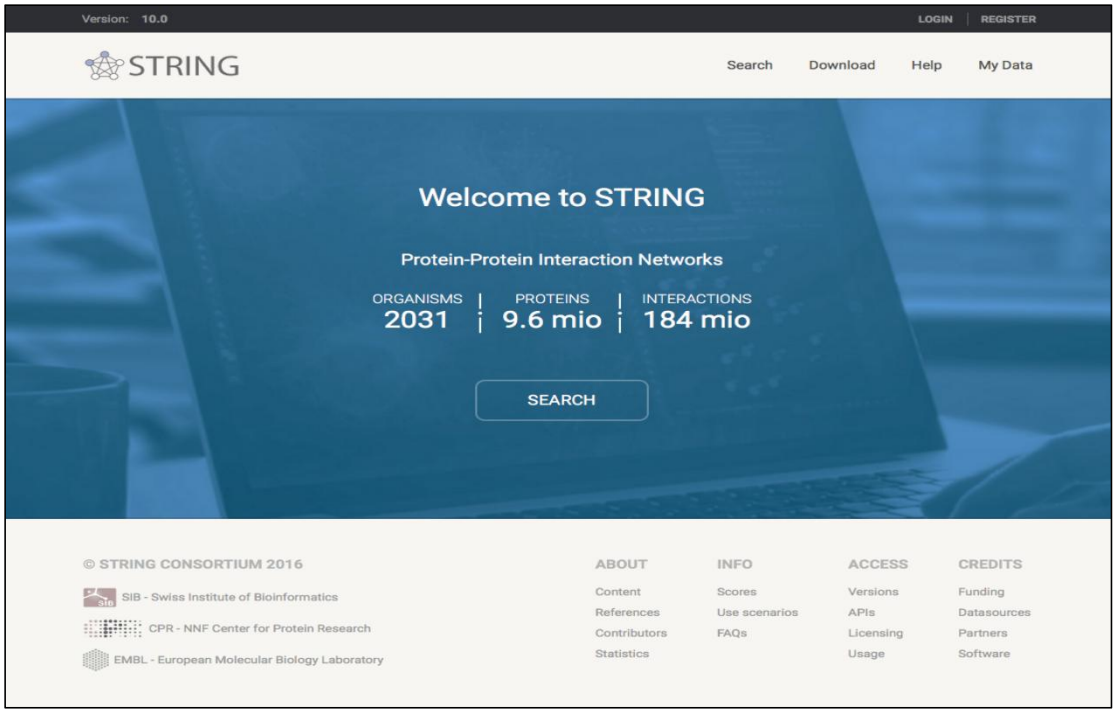
GEO2R is an interactive web tool that allows users to compare two or more groups of Samples in a GEO Series in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance, and as a collection of graphic plots to help visualize differentially expressed genes and assess data set quality. Using GEO2R comparing two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance can be done. The present study aimed to identify new targets for the diagnosis and treatment of AML and CML. The microarray data sets GSE140315and GSE44589 were selected from the AML database of the Gene Expression Omnibus and compared using the GEO2R tool to identify differentially expressed genes (DEGs). Differentially expressed genes (DEGs) from expression profile datasets were screened using GEO2R, and bioinformatics analysis was performed on the DEGs.

3.5 Metascape analysis:

Metascape (http://metascape.org) is a gene annotation and analysis resource that helps biologists make sense of one or multiple gene lists. It provides automated meta-analysis tools to understand either common or unique pathways and protein networks within a group of orthogonal target-discovery studies.Itprovides a significantly simplified user experience through a one-click Express Analysis interface to generate interpretable outputs. Gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed by Metascape.P value < 0.05 was considered statistically significant.

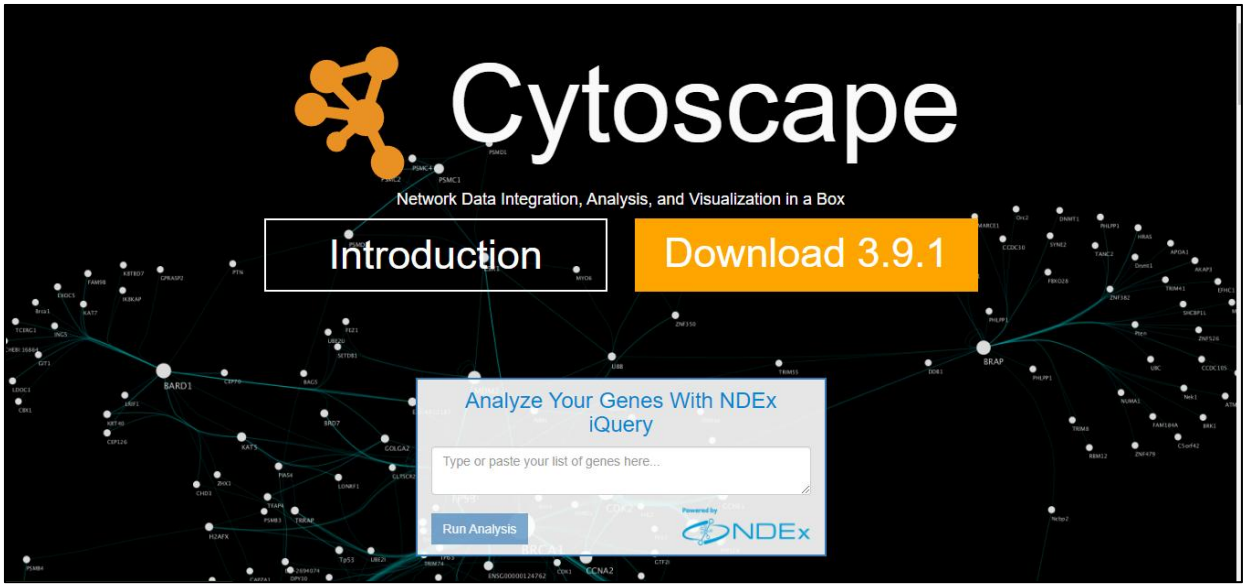
3.6 PPI construction using STRING:

STRING (https://string-db.org/) is a database of known and predicted protein-protein interactions networks designed for protein functional enrichment analysis. The interactions include direct (physical) and indirect (functional) associations. They stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases. The STRING database currently covers 24,584,628 proteins from 5,090 organisms.The Search Tool by choosing Multiple Proteins was used to construct the PPI networks of our gene set.



3.7 CYTOSCAPE analysis:

Cytoscape (https://cytoscape.org/) is an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other data. The cyto-Hubba plugin and Molecular Complex Detection (MCODE) plugin in Cytoscape were used to identify hub genes and screen modules of the PPI network. All parameters of the plugin were left at their default values. GO and KEGG enrichment of hub genes and genes in modules were also analyzed by Metascape. CytoHubba plugin uses 12 algorithms to evaluate the hub genes, and the top hub genes which were found in at least 9/12 algorithms were considered as hub genes. The intersections between the genes in the key module and the hub genes screened by CytoHubba application were considered as hub genes. Hub genes with a high degree of connectivity in the PPI network are significantly enriched in positive regulation of transcription. These genes play key roles in the characteristics and the development of the disease and they serve as diagnosticmarkers for leukemia.



TheCytoscape software was used to visualize the network. The Molecular Complex Detection (MCODE) is used to identify the merged network's hub genes. The clusters were formed with nodes of hub genes present in them. Clusters of hubs integrating PPI subnetworks comprise the backbone of a network. The essentiality of gene hubs in preserving the integrity of the interacting network is quantitatively expressed in network centrality statistics. The results of the *In silico* parameters carried out in the present study are discussed in the next chapter.

4

RESULTS AND DISCUSSION

Cancer remains a major cause of morbidity and mortality around the world. Cancer begins when a cell breaks free from the normal restraints on uncontrolled growth and spread. Leukemias are a group of life threatening malignant disorders of the blood and bone marrow. In the adolescent and young adult (AYA) population, the acute leukemias are most prevalent, with chronic myeloid leukemia being infrequently seen. In the last century, the emergence of *in silico* tools has improved the quality of healthcare studies by providing high quality predictions. This study is to analyze the molecular markers in leukemia sample using *in silico* analysis.

Molecular markers serve as biomarkers for the diagnosis of leukemia. The GSE140315 (GPL11180 [HT_MG-430_PM] Affymetrix HT MG-430 PM Array Plate, GPL16417 Illumina MiSeq (*Mus musculus*) of AML and GSE44589 (GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) of CML were selected from the database of the Gene Expression Omnibus and compared using the GEO2R tool to identify differentially expressed genes (DEGs). GEO2R is an interactive web tool that allows users to compare two or more groups of samples in a GEO Series in order to identify genes that are differentially expressed across experimental conditions.

Results are presented as a table of genes ordered by significance, and as a collection of graphic plots to help visualize differentially expressed genes and assess data set quality. Also the common DEGS of all the samples of Leukemia was analyzed for their PPIs. The major modules from the PPI network were screened using the Molecular Complex Detection (MCODE) tool in the cytoscape. The hub genes from the PPI network were found using CytoHubba, a cytoscape tool. The methodology of this study includes the retrieval of microarray data from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/).

In silico methods were used to determine the differentially expressed genes (DEGs) in the chosen dataset of the present study. Metascape database

(http://metascape.org) is used to perform GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analyses. KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies. STRING (https://string-db.org/) is used to construct a protein-protein interaction (PPI) network . The network was visualized in the Cytoscape software. The Molecular Complex Detection (MCODE) is used to identify the merged network's hub genes. Hub genes with a high degree of connectivity in the PPI network are significantly enriched in positive regulation of transcription. The discovery of tumor biomarkers is one of the methods for early detection of cancer, however the regulation of this extremely complex tumors cannot be met solely by a local emphasis on a single or specific gene. Some genes in tumors are abnormally expressed and intimately connected to many other genes, according to the entire regulatory network.

The molecular interactions were analysed using *in silico* methods and the results are presented as follows.

4.1 Microarray data retrieval:

Gene Expression Omnibus (GEO) is an online database which provides a comprehensive data on gene profiling and sequencing. The microarray data taken from GEO database were GSE16722, GSE140315, GSE167221, GSE1791, GSE44589, GSE195471, GSE157636, GSE133986, GSE158964, GSE192618, GSE92969, GSE180351, GSE192628, GSE148094, GSE178068, GSE190931, GSE150963, GSE124390 related to both AML and CML. Among the selected data only GSE140315 and GSE44589 has the samples specific for GEO2r analysis. These two data were more significant due to the possession of differentially expressed genes and upregulated and download genes present in the dataset. GSE140315 and GSE44589 data were retrieved from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). Based on the GPL11180 [HT_MG-430_PM] Affymetrix HT MG-430 PM Array Plate, GPL16417 Illumina MiSeq (*Mus musculus*), GSE140315 contains 8 samples whereas based on the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, GSE44589 contains 198 samples were chosen for the present study. These microarray data sets were used to perform GEO2r analysis.

4.2 Analysis of Differentially expressed genes (DEG) using GEO2r analysis:

The present study aims to identify new targets for the diagnosis and treatment of AML and CML. The GSE140315 and GSE44589 were selected from the database of the Gene Expression Omnibus and compared using the GEO2R tool to identify differentially expressed genes (DEGs). Based on the GPL11180 [HT_MG-430_PM] Affymetrix HT MG-430 PM Array Plate, GPL16417 Illumina MiSeq (*Mus musculus*), the microarray data set GSE140315 had 8 samples (4 normal and 4 leukemic samples). Based on the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, the microarray data set GSE44589 had 198 samples (120 normal and 78 leukemic samples). Differentially expressed genes (DEGs) from expression profile datasets were screened using GEO2R, and bioinformatics analysis was performed on the DEGs. A gene is declared differentially expressed if an observed difference or change in read counts or expression levels between two experimental conditions is statistically significant.

The up-regulated genes are mainly components that control cellular proliferation, whereas the down-regulated genes consist of proteins exposed on or secreted from the cell surface.

TABLE 2: Differentially Expressed Genes of datasets - GSE140315 and GSE44589

SAMPLES TAKEN	TOTAL DEGs	UP-REGULATED	DOWN-REGULATED
AML	227	71	156
CML	253	37	216

Several graphical plots are generated to help users further explore differentially expressed genes and assess dataset quality using GEO 2r analysis.

A Volcano plot displays statistical significance (-log10 P value) versus magnitude of change (log2 fold change) and is useful for visualizing differentially expressed genes. Volcano plot representing the DEGs of samples, satisfying the criteria of p adj<0.05. Blue color represents the downregulated genes and red color represents the upregulated genes in the sample.

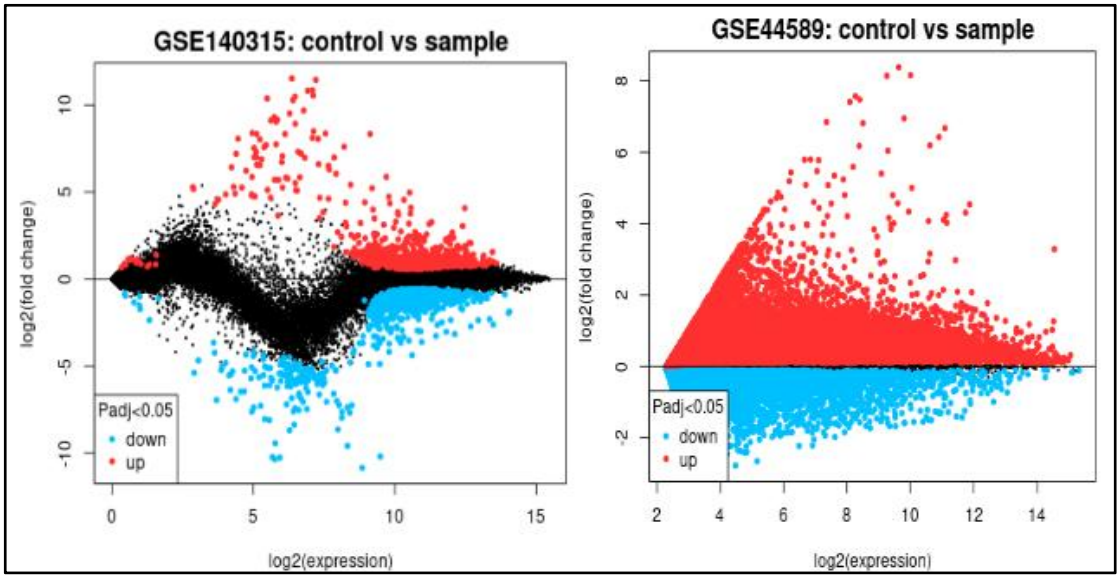


FIGURE 3: Volcano plot

P.Va l ⇨ P-value after adjustment for multiple testing. This column is generally recommended as the primary statistic by which to interpret results. Genes with the smallest P-values will be the most reliable.

P.Value ⇨ Raw P-value.

t ⇨ Moderated t-statistic (only available when two groups of samples are defined).

B ⇨ B-statistic or log-odds that the gene is differentially expressed (only available when two groups of samples are defined).

logFC ⇨ Log2-fold change between two experimental conditions (only available when two groups of samples are defined).

F ⇨ Moderated F-statistic combines the t-statistics for all the pair-wise comparisons into an overall test of significance for that gene (only available when more than two groups of Samples are defined)**.**

In AML, 227 genes were analysed, a total of 71 upregulated and 156 downregulated DEGs were identified. In CML, 253 genes were analysed, a total of 37 upregulated and 216 downregulated DEGs were identified (Table 2 and figure 3).

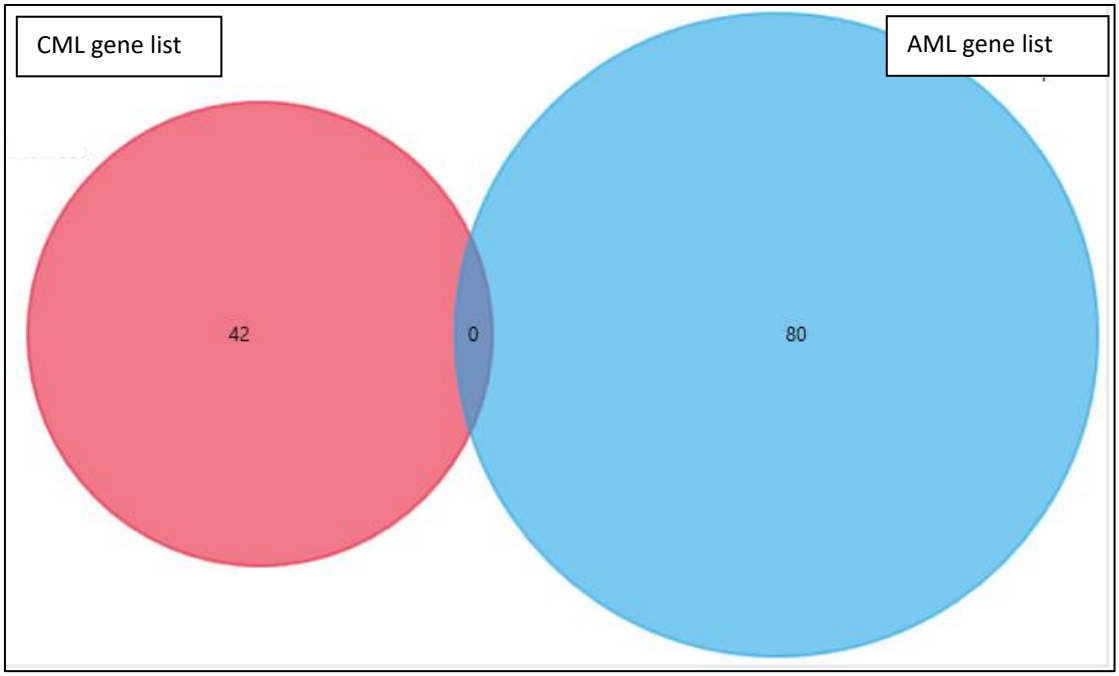


FIGURE 4:. The common DEGs among GSE140315 and GSE44589.

The Venn diagram shown here represents that there is no common genes in both AML and CML datasets.

Narendra *et al.,*(2021) analyzed the GSE7114 microarray dataset downloaded from GEO database using GEO2R tool and predicted DEGs in the form of Venn diagram and volcano plots to give a direct presentation of all the examined genes in the dataset. As shown in Venn diagram, a total of 749 significant DEGs (Adj.P < 0.05) were identified between parental and resistance cell lines. While, in volcano plot, red and blue dots displayed 387 (Adj.P < 0.05, log FC > 1) upregulated and 225 (Adj.P < 0.05, log FC < − 1) downregulated genes, respectively.

4.2.1 DIFFERENTIALLY EXPRESSED GENES IN AML AND CML

The list of differentially expressed genes (upregulated) in AML were given in the table 3

TABLE 3: DIFFERENTIALLY EXPRESSED GENES IN AML

Plxdc2	Pdlim1	Akr1c12	Tsen15
Cth	Arpc5l	Psat1	F630003A18Rik
Stc2	Gpr171	Snora44 Snhg12	Tnfsf11
Igfbp4	St8sia4	Hist1h2bc	Aars
Stc2	Thy1	Chdh	Esf1
Ptgds	Hist1h1c	Lgals3	Spata5
Rgs5	Elane	4930519L02Rik	Wdr74
F7	Akr1c13	Phgdh	Mdfic
Nrip1	Klrb1f	Zfas1	Pck2

Bcat1	Slc7a1	Hist1h1c	Rab38
Mpo	Diaph2	Nupr1	Chac1
Crisp1	Xpot	Snord22 Snhg1	Erbb3
Igfbp4	Plxnd1	Als2cl	Csprs
P2rx3	Trib3	Siglec5	Ffar2
Tusc1	Nupr1	Gsn Mthfd2	Hacd4
Arhgap6	Fam84b	Nebl	5430416N02Rik
Cxxc5	Snord22 Snhg1	Ptgds	Metrnl
Cpa3	Esf1	Gsn Mthfd2	Metrnl
F2r	Atf5	Nkg7	Gpt2
Thy1	Aars	Cd44	Rgs5
Ddit3	Snord22 Snhg1	Tsr1 Srr	Slc15a2
Arhgap6	Socs2	Shmt2	Cyb5r1
Hbb-b2	Snx10	Akr1c12	9530053H05Rik
Cyb5r1	Cd44	Zbtb18	Akr1c12
9530053H05Rik	Hist1h2bp	Snhg5	Aqp9
Arhgap6	Gpt2	Lgals3	Snora44 Snhg12
Hbb-b2	Rgs5	Cyb5r1	Hist1h2bc
Slc15a2	Ddit3	Chdh	F3

4.2.1.1 SIGNIFICANT GENES OF AML AND ITS ROLE:

1. Aminoacyl-tRNA synthetases (aaRSs):

Aminoacyl-tRNA synthetases (aaRSs) have emerged as multifaceted proteins with complex connections to human disease, including cancer. The aaRSs are a group of 20 cytoplasmic and 19 mitochondrial enzymes (with glycyl- and lysyl-tRNA synthetase being shared between the two localities) responsible for aminoacylating tRNAs with their cognate amino acids to support protein synthesis. Acute myeloid leukemia (AML) underexpresses a majority of the aaRSs, with minimal changes in these genes at the DNA level (Wang *et al*., 2020).

2. SNHG15:

The presence of the SNHG15 cytosine-to-thymidine variant was also found to associate with better outcome in an independent dataset of CN-AML patients, despite differences in treatment protocols and RNA sequencing techniques. In order to gain biological insights, we cloned and overexpressed both wild-type and variant versions of the SNHG15 lncRNA (Papaioannou *et al*., 2022).

3. RGS5:

The RGS5 peptides represent, candidates for the development of cancer vaccines designed to target malignant cells and tumor vessels. RGS5 is overexpressed in

pericytes of newly developing tumor vessels, indicating that RGS5 plays an important role during tumor angiogenesis (Kruger *et al*., 2015).

4. FLT3:

The FMS-like tyrosine kinase 3 internal tandem duplication (FLT3-ITD) mutation represents the most frequent genetic alteration in acute myeloid leukemia (AML) and is associated with poor prognosis. The mutation promotes cancer cell survival and proliferation, and shifts their glucose metabolism towards aerobic glycolysis, a frequent alteration in cancer. . Several gene mutations have been identified as leukemia risk factors, with FLT3 the most frequent site of genetic alterations associated with poor disease outcome (George *et al*., 2015).

5. GPT2:

GPT2 plays a role in glutamine metabolism and encodes a glutamic-pyruvic transaminase that catalyzes reverse transamination between alanine and 2-oxoglutarate to generate pyruvate and glutamate. Cancer cells exhibit a metabolic reprogramming that results in increased lactate acid production in the Warburg effect and the use of glutamine to replenish the tricarboxylic acid cycle (TCA) (Honma *et al.*, 2017).

4.2.2 DIFFERENTIALLY EXPRESSED GENES IN CML

The list of differentially expressed genes (upregulated) in CML were given in the table 4.

TABLE 4: DIFFERENTIALLY EXPRESSED GENES IN CML

LOC101927122	HIST1H4F	STX7	AMY2B
SLC34A3	CMTM3	LOC400927-CSN	AMY2A
LAT2	PCSK4	AKAP7	NDUFS7
ALPP	USP35	GTSE1	STXBP2
DBP	MATN4	WFDC12	ZNF524
KRT8	ALPPL2	AMPD3	EHMT1
ADARB1	AMPD2	CD300LG	CPNE7
CACNA1I	POLR3H	L3MBTL1	ZNF821
EZH1	FAM91A1	MAPK15	ABHD1
SPDEF	AMPD1	ARF1	ARF3
ANK1	ANK2	ANK3	RING1

4.2.2.1 SIGNIFICANT GENES OF CML AND ITS ROLE:

1. PTRH1:

PTRH1 was found to be significantly overexpressed in patients with deeper tumor infiltration. PTRH1 expression was not significantly different between the tumor tissues and normal tissues (Zhang *et al*., 2020).

2. FAN1:

FAN1 encodes a DNA repair nuclease. Genetic deficiencies, copy number variants, and single nucleotide variants of FAN1 have been linked to karyomegalic interstitial nephritis, cancer, and most recently repeat expansion diseases. Individuals carrying FAN1-deficiencies do not present with the clinical attributes typical of FA patients (i.e., anemia, bone marrow failure, skin pigmentation anomalies, skeletal abnormalities of the upper limbs, or spontaneous leukemia) (Ionita-Laza *et al*., 2014).

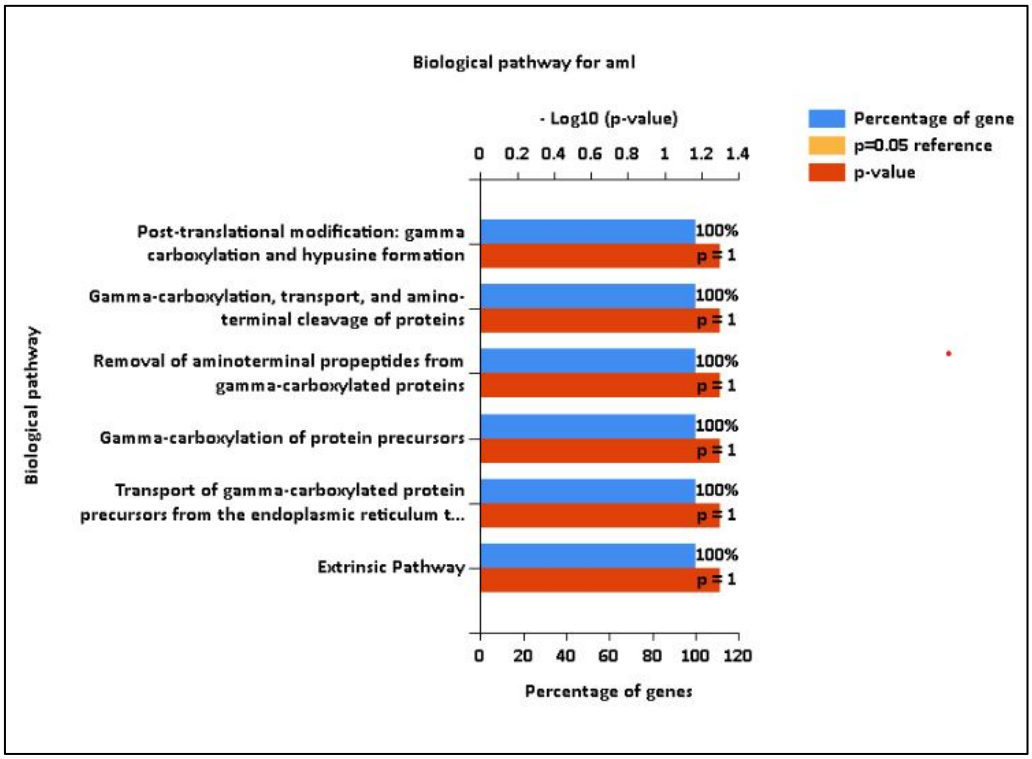
3. USP35:

The new target genes that were upregulated by TPA-mediated leukemia cell differentiation were identified by analyzing ChIP-seq and microarray data. The knockdown of USP3 inhibited TPA-induced leukemia cell differentiation via regulating H2AK119ub. USP3 is reported to be associated in cancer development. For example, USP3 activity is essential for hepatocyte growth factor-induced scattering. Depletion of USP3 leads to loss of cell–cell contacts and motility, indicating its role in adjusting cancer cell activity. USP3 knock out (KO) cells were interrupted to preserve chromosome integrity and USP3 KO mice promote tumor development (Lancini et al. 2014). Moreover, USP3 regulates H2AK119ub levels in γH2AX at K118 and K119. It was found that USP3 decreased H2AK119ub levels in TPA-treated HL-60 cells, resulting in inducing leukemia cell differentiation (Chae *et al.,* 2019).

4.2.3 BIOLOGICAL PATHWAY OF AML:

Pathway enrichment analysis helps researchers gain mechanistic insight into gene lists generated from genome-scale (omics) experiments. This method identifies biological pathways that are enriched in a gene list more than would be expected by chance. A biological pathway is a series of actions among molecules in a cell that leads to a certain product or a change in the cell. It can trigger the assembly of new molecules, such as a fat or protein, turn genes on and off, or spur a cell to move.

FIGURE 5: BiologicalPathway of AML



The biological pathway of AML is represented in **figure 5.** The biological pathways involved in AML include post-translation modification, gamma carboxylation and hypusine formation, gamma-carboxylation, transport, and amino-terminal cleavage of proteins, removal of aminoterminal pro-peptides from gamma-carboxylated proteins, gamma-carboxylation of protein precursors, transport of gamma-carboxylated protein precursors from the endoplasmic reticulum, extrinsic pathway.

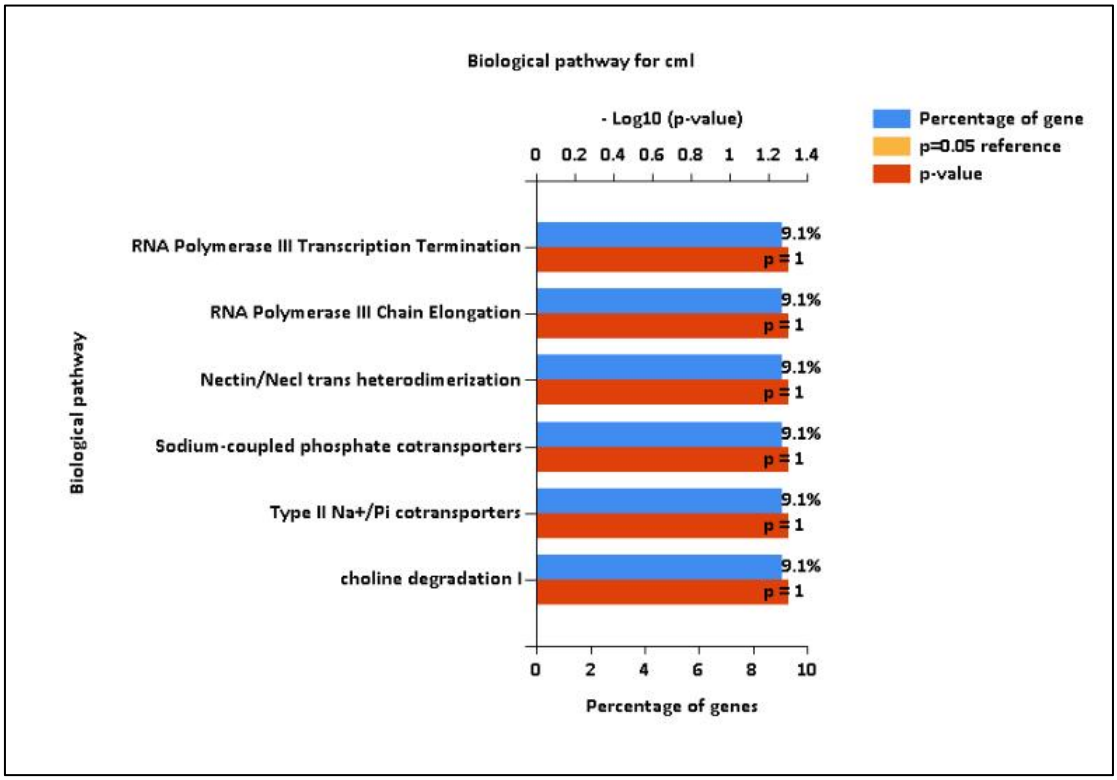
Acute myeloid leukemia (AML) is the result of a multistep transforming process of hematopoietic precursor cells (HPCs) which enables them to proceed through limitless numbers of cell cycles and to become resistant to cell death. Early studies have implicated over-expression of tissue factor (TF), the principal initiator of the extrinsic coagulation protease cascade, in the pathogenesis of AML-associated DIC. The commercial TF ELISA used in our study does not detect alternatively spliced TF, future studies should therefore investigate, if proteolytic degradation of full-length TF by enzymes liberated from dying myeloblasts represents a novel mechanism of

controlling the extrinsic coagulation pathway in newly diagnosed AML which was reported by (Hogstrand *et al.,* 2012).

4.2.4 BIOLOGICAL PATHWAY OF CML:

Researchers can use pathway enrichment analysis to acquire mechanistic insight into gene lists generated by genome-scale (omics) investigations. This approach finds biological pathways that are more enriched in a gene list than is expected by chance. The biochemical mechanism that leads to CML. A biological route is a chain of events involving molecules in a cell that results in the production of a certain product or a change in the cell. It can cause new molecules to form, such as fats or proteins, turn genes on and off, or cause a cell to move.

FIGURE:6 BiologicalPathway of CML



The biological pathways involved in CML **(Figure 6)** include RNA polymerase lll transcription termination, RNA polymerase lllchain elongation, Nectin/Necl trans heterodimerization, Sodium-coupled phosphate cotransporters, Type llNa+/Pi cotransporters, Choline degradation.

The PLC- and PLD mediated pathways of PtdCho breakdown were inactive in HL60 cells suggest that the development of agents which selectively inhibit PtdCho-specific PLC or PLD but no inositol lipid-specific PLC may be useful in assessing the role of PtdCho

breakdown in the regulation of proliferation of leukaemia cells. These agents may find applications in the selective blocking of cell-cycle transit of normal cells whose proliferation is dependent on both inositol lipid- and PtdCho-derived signals while allowing the continued proliferation of leukemia cells which id reported by (Gitendra *et al.,2011)*

4.3 Pathway analysis using Metascape:

When comparing with significant metabolic pathways in GO and KEGG pathways, Acute Myeloid Leukemia (AML) and Chronic Myeloid Leukemia (CML) includes certain genes enriched in these pathways.

4.3.1 Pathway enrichment of Acute Myeloid Leukemia(AML):

Pathway enrichment is a method to identify classes of genes or proteins that are over-represented in a large set of genes or proteins, and may have an association with disease phenotypes**.** An enrichment analysis is an *in silico* method which identifies enriched or over-represented gene sets among a list of ranked genes. Gene sets are groups of genes that are functionally related according to current study. It is used to determine whether a pre-defined set of genes show statistically significant differential expression between two phenotypes. KEGG mapping is the process to map molecular objects (genes, proteins and small molecules) to molecular interaction/reaction/relation networks (KEGG pathway maps and KEGG modules). It is not simply an enrichment process; rather it is a set operation to generate a new set. Gene Ontology (GO) term enrichment is a technique for interpreting sets of genes making use of the Gene Ontology system of classification, in which genes are assigned to a set of predefined bins depending on their functional characteristics.

An enrichment analysis is a bioinformatics method which identifies enriched or over-represented gene sets among a list of ranked genes. Gene sets are groups of genes that are functionally related according to current knowledge. Pathway enrichment analysis is an essential step for interpreting high-throughput (omics) data that uses current knowledge of genes and biological processes. A common application determines statistical enrichment of molecular pathways, biological processes and other functional annotations in long lists of candidate genes. Genomic, transcriptomic, proteomic and epigenomic experiments emphasize complementary aspects of underlying biology and are best analyzed integratively, as is now routinely done in large-scale projects such as The Cancer Genome Atlas (TCGA), Clinical Proteome Tumor Analysis Consortium (CPTAC), International Cancer Genome Consortium (ICGC), Genotype-Tissue Expression (GTEx), and others.

Thus, simultaneous analysis of multiple candidate gene lists for characteristic pathways is increasingly needed.

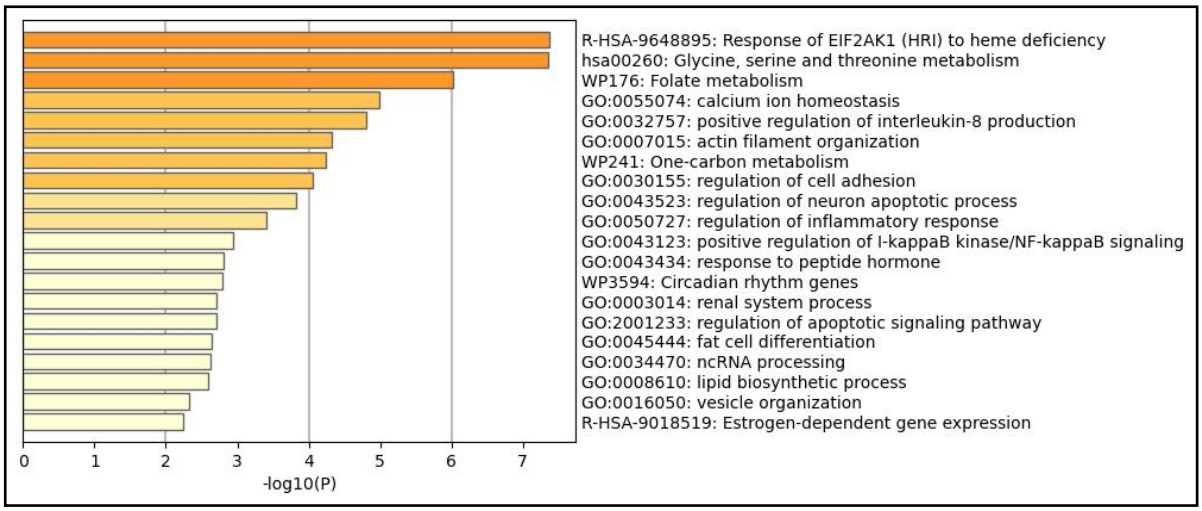


FIGURE:7 Bar graph of enriched terms across input gene lists, colored by p-values.

The regulation of cell adhesion is under control of multiple signaling pathways and mechanisms that allow fine-tuning of the process. Continuous formation and disassembly is necessary during embryogenesis and development. Some of the processes involved at this stage, such as EMT, are hijacked during carcinogenesis. Any apoptotic process in a neuron, the basic cellular unit of nervous tissue. Each neuron consists of a body, an axon, and dendrites. Their purpose is to receive, conduct, and transmit impulses in the nervous system. Transcription factor NF-kappaB plays a key regulatory role in the cellular response to pro-inflammatory cytokines such as tumor necrosis factor-alpha (TNF). In the absence of TNF, NF-kappaB is sequestered in the cytoplasm by inhibitory IkappaB proteins. Phosphorylation of IkappaBby the beta-catalytic subunit of IKK, a multicomponent IkappaB kinase, targets the inhibitor for proteolytic destruction and facilitates nuclear translocation of NF-kappaB. This pathway is initiated by TNF-dependent phosphorylation of T loop serines in IKKbeta, which greatly stimulates IkappaB kinase activity. Nuclear condensation, cell shrinkage, membrane blebbing, and DNA fragmentation are all hallmarks of apoptosis, a regulated cellular suicide mechanism. The central regulators of apoptosis are caspases, a family of cysteine proteases.

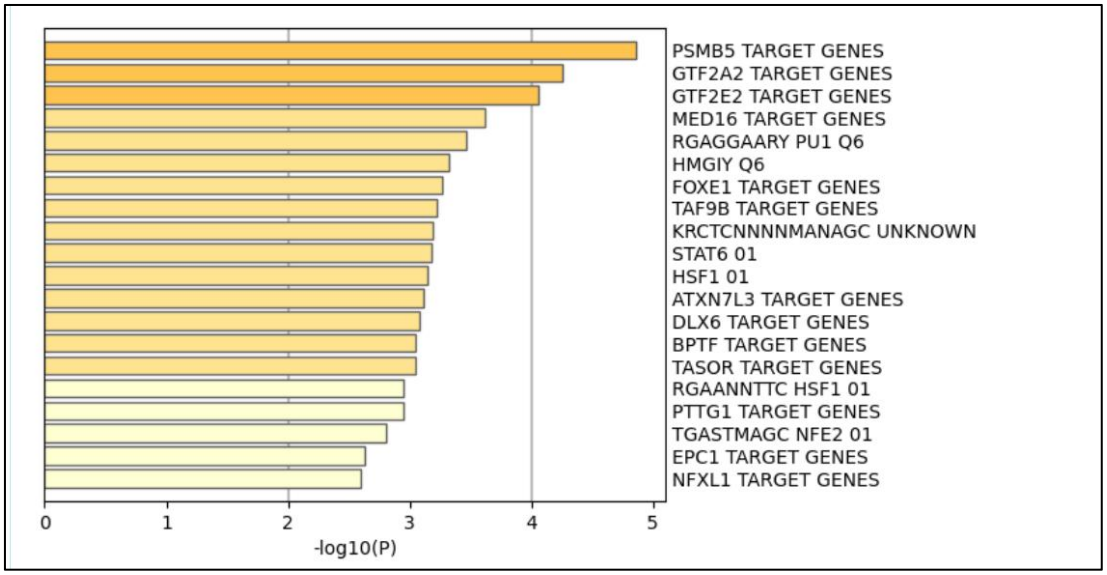


FIGURE:8 Summary of enrichment analysis in Transcription Factor Targets.

In molecular biology, a transcription factor (TF) (or sequence-specific DNA-binding factor) is a protein that controls the rate of transcription of genetic information from DNA to messenger RNA, by binding to a specific DNA sequence. The function of TFs is to regulate—turn on and off—genes in order to make sure that they are expressed in the desired cells at the right time and in the right amount throughout the life of the cell and the organism. Transcription factors are of clinical significance for at least two reasons: mutations can be associated with specific diseases, and they can be targets of medications. The proteasome is a multicatalytic proteinase complex with a highly ordered ring-shaped 20S core structure. The core structure is composed of 4 rings of 28 non-identical subunits; 2 rings are composed of 7 alpha subunits and 2 rings are composed of 7 beta subunits. Proteasomes are distributed throughout eukaryotic cells at a high concentration and cleave peptides in an ATP/ubiquitin-dependent process in a non-lysosomal pathway. An essential function of a modified proteasome, the immunoproteasome, is the processing of class I MHC peptides. This gene encodes a member of the proteasome B-type family, also known as the T1B family, that is a 20S core beta subunit in the proteasome. BPTF is the largest subunit of the NURF (Nucleosome Remodeling Factor) complex.

4.3.2 Pathway enrichment of Chronic Myeloid Leukemia(CML):

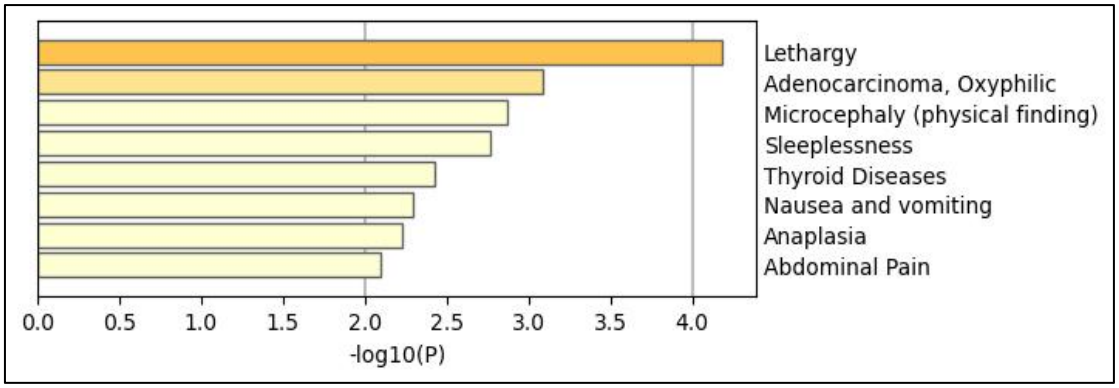


FIGURE:9 Bar graph of enriched terms across input gene lists, colored by p-values.

METASCAPE ANALYSIS:

Metascape is a free gene annotation and analysis resource that helps biologists make sense of one or multiple gene lists. Metascape provides automated meta-analysis tools to understand either common or unique pathways and protein networks within a group of orthogonal target-discovery studies. Metascape implements a CAME analysis workflow:

- Conversion: Convert gene identifiers from popular types (such as Symbol, RefSeq, Ensembl, UniProt, UCSC) into human Entrez gene IDs and vice versa.
- Annotation: Extract from dozens of function-relevant gene annotations, including protein families, transmembrane/secreted predictions, disease associations, compound associations, etc.
- Membership: Flag gene memberships based on a custom keyword search within selected ontologies, e.g., highlight known "invasion" genes.
- Enrichment: Identify enriched biological themes, particularly GO terms, KEGG, Reactome, BioCarta, WikiPathways as well as other pathways and data sets collected in MSigDB, etc. In addition, enriched ontology terms are automatically clustered to reduce redundancy for easier interpretation. Protein-protein interaction networks are constructed based on STRING, BioGRID, OmniPath, etc. Dense components are identified and biologically interpreted.

GO	Category	Description	Count	%	Log10(P)
GO:0098662	GO Biological Processes	inorganic cat ion transmembrane transport	4	9.76	-2.12
GO:0002703	GO Biological Processes	regulation of leukocyte mediated immunity	4	9.76	-3.53
GO:0051640	GO Biological Processes	organelle localization	4	9.76	-2.33
GO:0030855	GO Biological Processes	epithelial cell differentiation	4	9.76	-2.12

TABLE:5 Summary of enrichment analysis in METASCAPE

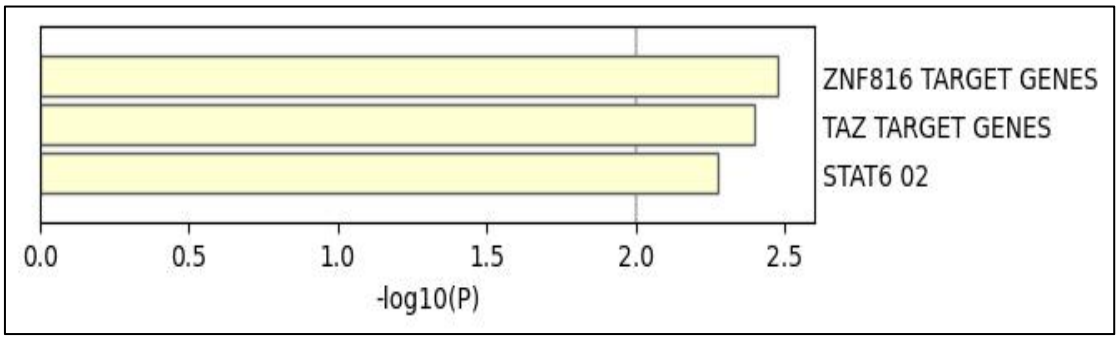


FIGURE:10 Summary of enrichment analysis in Transcription Factor Targets.

TABLE:6 Enrichment analysis in Transcription Factor Targets.

GO	Description	Count	%	Log10(P)
M40760	ZNF816 TARGET GENES	4	9.8	-2.5
M30192	TAZ TARGET GENES	4	9.8	-2.4
M7994	STAT6 02	3	7.3	-2.3

4.4 STRING analysis:

The common DEGS of all the samples of Leukemia was analyzed for their PPIs. Within these DEGs, Further these genes were shortlisted under the condition of those which has its scores more than 0.9 with other genes. The DEGs were found to be

active participants of the ways of ABC transporters. Local STRING network clusters or simply STRING clusters are precomputed protein clusters derived by hierarchically clustering the full STRING network using an average linkage algorithm. The smallest clusters generated consist of 5 proteins and largest used in the enrichment analysis contain 200 proteins. In order to reduce redundancy all clusters with a size difference of less than 5 proteins towards their child cluster were removed (next smallest cluster in the hierarchy). The names are derived automatically based on a cluster's consensus protein annotations taken from GO, KEGG, Reactome, UniProt, Pfam, SMART, and InterPro. All clusters and their hierarchical tree are available for download in the download section of STRING.

Comparative analysis of protein protein interactions of individual DEGs and common protein protein interactions, the process cell cycle, DNA replication, Hematopoietic cell age, signaling pathway and Progesterone-mediated oocyte maturation. According to the comparative study of DEGS in Leukemia, the most of the DEGs are significantly involved in the ways of Cell cycle and signaling pathway. The network view summarizes the network of predicted associations for a particular group of proteins. The network nodes are proteins. The edges represent the predicted functional associations. Nodes are modeled as masses and edges as springs; the final position of the nodes in the image is computed by minimizing the 'energy' of the system.

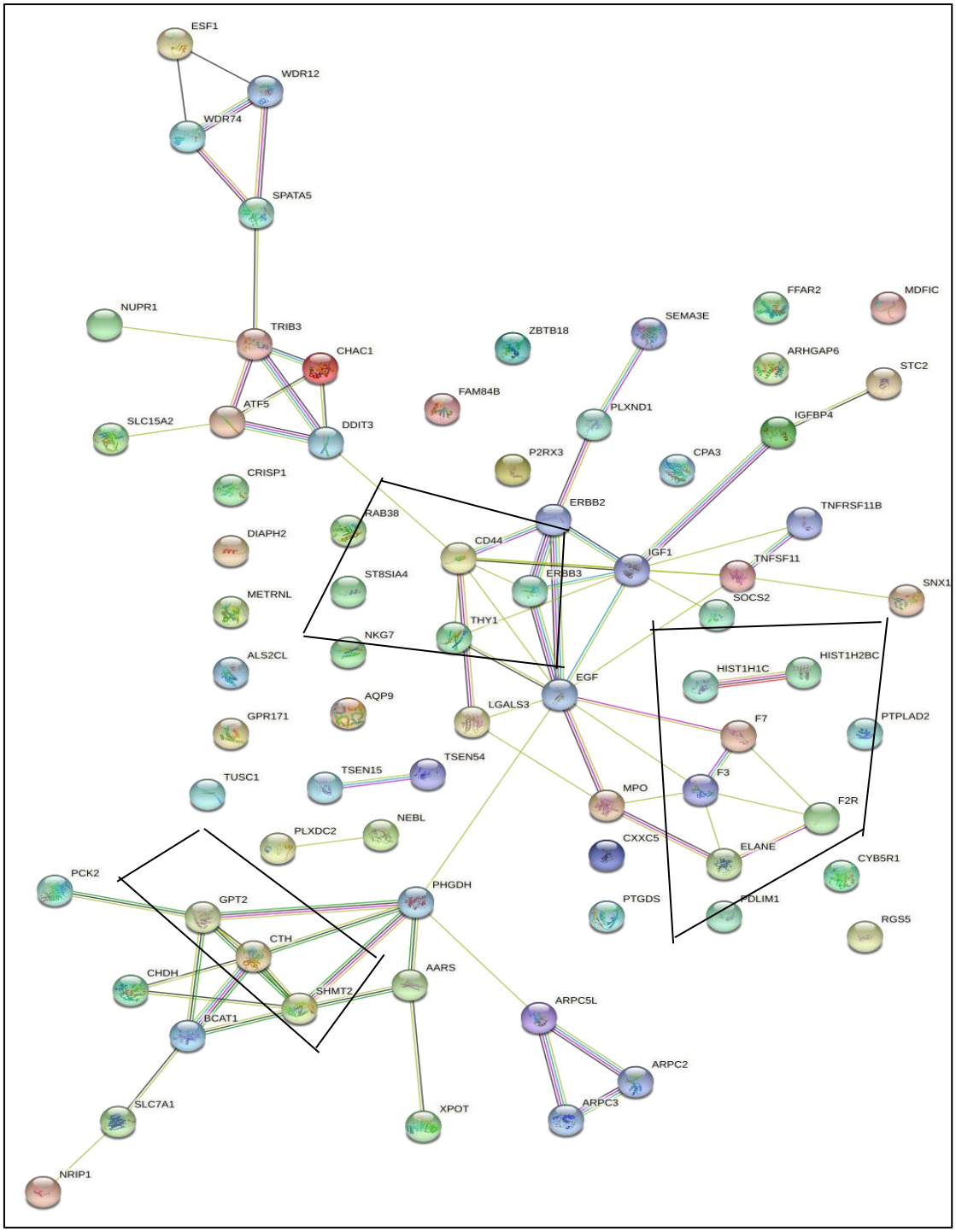


FIGURE:11 The protein-protein interaction of the DEGs of the AML using STRING tool.

AML:

The combined score is computed by combining the probabilities from the different evidence channels and corrected for the probability of randomly observing an interaction. The entire database content of STRING is pre-computed, stored in a relational database and available for separate download. All interaction evidence that

contributes to a given network is benchmarked and scored , and the scores are integrated into a final 'combined score'. This score is scaled between zero and one and provides an estimate of STRING's confidence on whether a proposed association is biologically meaningful given all the contributing evidence. Each association is provided as a connection between two non-identical proteins, each from a different protein-coding gene locus.

The STRING database aims to integrate all known and predicted associations between proteins, including both physical interactions as well as functional associations. To achieve this, STRING collects and scores evidence from a number of sources:

- Automated text mining of the scientific literature.
- Databases of interaction experiments and annotated complexes/pathways.
- Computational interaction predictions from co-expression and from conserved genomic context.
- Systematic transfers of interaction evidence from one organism to another.
- The experiments channel: Here, evidence comes from actual experiments in the lab (including biochemical, biophysical, as well as genetic experiments). This channel is populated mainly from the primary interaction databases organized in the IMEx consortium, plus BioGRID.
- The database channel: In this channel, STRING collects evidence that has been asserted by a human expert curator; this information is imported from pathway databases.
- The textmining channel: STRING searches for mentions of protein names in all PubMed abstracts, in an in-house collection of more than three million fulltext articles, and in other text collections. Pairs of proteins are given an association score when they are frequently mentioned together in the same paper, abstract or even sentence (relative to how often they are mentioned separately). This score is raised further when it has been possible when one or more sentences through Natural Language Processing, and a concept connecting the two proteins was encountered (such as 'binding' or 'phosphorylation by').
- The co expression channel: For this channel, gene expression data originating from a variety of expression experiments are normalized, pruned and then

correlated. Pairs of proteins that are consistently similar in their expression patterns, under a variety of conditions, will receive a high association score. In addition to large-scale microarray data, in version 10.5 of STRING, RNA seq expression data are now also processed; this results in the inclusion of 16 previously non-covered organisms into this channel.

- The neighborhood channel: This channel, and the next two, are genome-based prediction channels, whose functionality is generally most relevant for Bacteria and Archaea. In the neighborhood channel, genes are given an association score where they are consistently observed in each other's genome neighborhood (such as in the case of conserved, co-transcribed ‘operons’).
- The fusion channel: Pairs of proteins are given an association score when there is at least one organism where their respective orthologs have fused into a single, protein-coding gene.
- The co-occurrence channel: In this channel, STRING evaluates the phylogenetic distribution of orthologs of all proteins in a given organism. If two proteins show a high similarity in this distribution, i.e. if their orthologs tend to be observed as ‘present’ or ‘absent’ in the same subsets of organisms, then an association score is assigned. For this channel, the details of the STRING implementation have recently been described, separately .

TABLE:7 The data highlights the top enriched genes in STRING tool with node1, node2, coexpression, combined Score.

node1	node2	coexpression	combined score
Aars	Psat1	0.146	0.657
Aars	Xpot	0.151	0.499
Akr1c12	Akr1c13	0.697	0.701
Akr1c13	Akr1c12	0.697	0.701
Arpc2	Pdlim1	0	0.54
Arpc2	Arpc5l	0.317	0.999
Arpc5l	Arpc2	0.317	0.999
Atf5	Slc15a2	0	0.636
Atf5	Trib3	0.149	0.664
Bcat1	Psat1	0.136	0.5
Bcat1	Shmt2	0.146	0.49
Bcat1	Cth	0.065	0.937
Cd44	Tnfsf11	0	0.463
Chac1	Ddit3	0.122	0.656
Chdh	Cth	0.085	0.425
Cpa3	Elane	0.123	0.416
Cth	Psat1	0.103	0.574
Cth	Chdh	0.085	0.425
Cyb5r1	Hbb-bs	0.063	0.906
Ddit3	Nupr1	0.146	0.462
Ddit3	Trib3	0.232	0.985
Elane	Cpa3	0.123	0.416
Elane	Mpo	0.597	0.936
Elane	F3	0	0.604
Erbb3	Cd44	0.048	0.702
Esf1	Wdr74	0.793	0.793
F2	Cpa3	0.08	0.571
F2	F2r	0	0.999
F2r	Cpa3	0	0.486
F2r	F7	0	0.478
F2r	Elane	0	0.451
F3	F2	0	0.969
F3	F7	0	0.999
F7	F2	0.239	0.481
F7	F3	0	0.999
F7	F2r	0	0.478
Gpt2	Psat1	0.066	0.658
Gpt2	Phgdh	0.067	0.837
Hbb-bs	Cyb5r1	0.063	0.906
Hist1h1c	Hist1h2bc	0.606	0.812
Hist1h2bc	Hist1h2bp	0.406	0.406
Hist1h2bc	Hist1h1c	0.606	0.812

Hist1h2bp	Hist1h2bc	0.406	0.406
Lgals3	Cd44	0.232	0.745
Mpo	F2	0	0.583
Mpo	Elane	0.597	0.936
Nebl	Plxdc2	0	0.577
Nupr1	Ddit3	0.146	0.462
Nupr1	Trib3	0.127	0.538
Pck2	Gpt2	0	0.52
Pck2	Phgdh	0.099	0.4
Pdlim1	Arpc2	0	0.54
Phgdh	Psat1	0.905	0.999
Phgdh	Shmt2	0.201	0.838
Phgdh	Gpt2	0.067	0.837
Phgdh	Aars	0.117	0.561
Phgdh	Pck2	0.099	0.4
Phgdh	Cth	0.078	0.574
Plxdc2	Nebl	0	0.577
Psat1	Trib3	0.133	0.416
Psat1	Xpot	0.124	0.425
Psat1	Slc7a1	0.103	0.456
Psat1	Cth	0.103	0.574
Psat1	Aars	0.146	0.657
Psat1	Gpt2	0.066	0.658
Psat1	Shmt2	0.226	0.875
Psat1	Phgdh	0.905	0.999
Shmt2	Psat1	0.226	0.875
Shmt2	Phgdh	0.201	0.838
Siglec5	Cd44	0	0.522
Siglec5	Thy1	0	0.552
Slc15a2	Atf5	0	0.636
Slc7a1	Psat1	0.103	0.456
Slc7a1	Bcat1	0.062	0.523
Spata5	Trib3	0.064	0.418
Spata5	Wdr74	0.085	0.495
Thy1	Cd44	0	0.902
Thy1	Siglec5	0	0.552
Tnfrsf11a	Tnfsf11	0	0.999
Tnfsf11	Cd44	0	0.463
Trib3	Psat1	0.133	0.416
Trib3	Ddit3	0.232	0.985
Wdr74	Spata5	0.085	0.495
Xpot	Psat1	0.124	0.425
Xpot	Aars	0.151	0.499

The coexpression view shows the genes that are co-expressed in the same or in other species (transferred by homology). Co-expression is shown by a red square: more intense color of the square represent a higher association score of the expression

data. The combined score is computed by combining the probabilities from the different evidence channels and corrected for the probability of randomly observing an interaction.

CML:

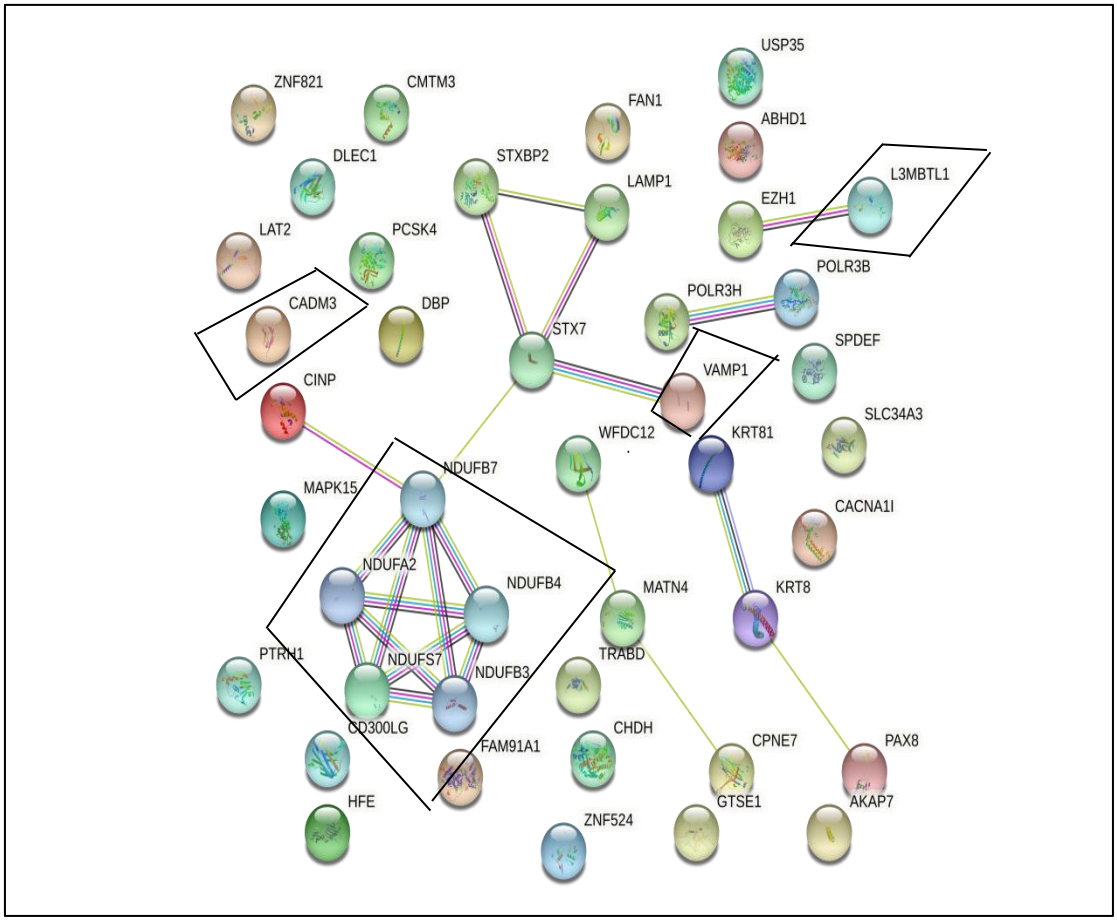


FIGURE:12 The protein-protein interaction of the DEGs of the CML using STRING tool.

TABLE:8 The data highlights the top enriched genes in STRING tool with node1, node2, co expression and combined Score.

node1	node2	Co expression	combined score
A1BG	A2M	0.056	0.475
ADARB1	AMBP	0.111	0.855
ARF1	SHROOM2	0	0.54
ANK1	SHROOM2	0	0.54
ARF3	A2M	0.056	0.475
A2M	AMBP	0.111	0.855
AIRE	NUDT10	0	0.705
AIRE	HIST1H3G	0	0.829
ALPP	HIST1H4F	0.21	0.994
ALPPL2	MRGPRD	0	0.489
AMELX	SHROOM2	0	0.54
AMELX	SLC25A6	0	0.576
AMPD1	IQCF1	0	0.411
AMPD2	ADARB1	0	0.74
AMPD3	ADARB1	0	0.74
AMPH	SRGAP1	0.062	0.448
AMY2B	AMY2B	0	0.802
AMY2A	AMY1B	0	0.803
ANK2	SRC	0.082	0.405
ANK3	AMPH	0.055	0.948
ARHGAP1	ARF1	0.062	0.939
ASB10	EFEMP1	0	0.455
BIN1	SRC	0.082	0.405
BIN1	AMPH	0.055	0.948
C12orf55	CFAP46	0.233	0.768
DYNLRB2	DYNLT1	0.083	0.945
EHMT1	AMPD2	0	0
GPRIN2	HYDIN	0	0.848
GRIK5	GPR6	0.143	0.847
GTSE1	MKI67	0.756	0.835
GTSE1	CDKN1A	0.062	0.855
GTSE1	PLK5	0.084	0.573
HIST1H4F	HIST1H3G	0.21	0.994
IL31RA	MRGPRD	0	0.489
IQCF1	AMPD1	0	0.411
KIAA1731	TRMT6	0	0.544
KY	MYO16	0.063	0.417
L3MBTL1	HIST1H4F	0	0.948
LAMA3	SRC	0.085	0.911
LAMP1	LMAN1	0.062	0.594
LAMP1	AMELX	0	0.939
LINC00116	NDUFS7	0.444	0.444
MANBA	ACP5	0	0.413
MAP3K10	APBA1	0.076	0.463
MAPK15	SRC	0	0.429
MAPK15	MFAP4	0	0.56

NCS1	DTX2	0	0.429
NCS1	ARF1	0.062	0.821
NDUFB4	NDUFS8	0.283	0.992
NDUFS8	NDUFB4	0.283	0.992
NDUFS8	NDUFS7	0.962	0.999
NTN1	SRC	0.062	0.938
NTRK3	SRC	0.062	0.653
NUDT10	AIRE	0	0.705
PAX8	KRT8	0	0.455
PAX8	WNT4	0	0.521
PCDHGC5	SLC46A1	0	0.412
PCSK4	AMH	0.152	0.496
PHIP	XKR6	0	0.461
PKP3	KRT8	0.369	0.737
PLA2G4D	CHDH	0	0.6
RHOB	ARF3	0.107	0.497
RING1	EZH1	0.062	0.975
SHROOM2	APEX1	0	0.438
SHROOM2	MCHR1	0.069	0.599
SHROOM2	AMELX	0	0.54
SHROOM3	APEX1	0	0.536
SLC1A6	GRIA4	0.108	0.433
SLC1A6	MCHR1	0.062	0.592
SLC1A6	NCS1	0.062	0.624
SLC25A6	AMELX	0	0.576
SLC34A3	CLCNKB	0.062	0.404
STXBP2	STX7	0.063	0.554
TAAR9	STX7	0	0.501
TAF1C	POLR1A	0.062	0.854
TEX261	TM4SF20	0	0.575
TIAM2	RHOB	0	0.481
TM4SF20	TEX261	0	0.575
TM4SF20	CMTM3	0	0.413
TMED5	LMAN1	0.186	0.575
TMED5	WNT4	0	0.906
TMEM252	CNDP2	0.052	0.429
VAMP1	STX1B	0.116	0.976
VAPA	VAMP1	0.055	0.85
WBSCR22	TRMT6	0.11	0.641
WBSCR22	TRMT61A	0.299	0.833
XKR6	PHIP	0	0.461
ZNF296	ESRRB	0.063	0.403
ZSCAN10	ARAF	0.055	0.467

4.5 Gene enrichment using Cytoscape:

Cytoscape (https://cytoscape.org/) is an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other data. Additional features are available as plugins. Plugins are available for network and molecular profiling analyses, new layouts, additional file format support and connection with databases and searching in large networks. Plugins may be developed using the Cytoscape open Java software architecture by anyone and plugin community development is encouraged.

4.5.1 Identification of hub genes in AML using CYTOSCAPE :

The major modules from the PPI network were screened using the Molecular Complex Detection (MCODE) tool in the cytoscape, with the degree cut-off set to 5. The hub genes from the PPI network were found using CytoHubba, a cytoscape tool. The top HUB genes reported in at least 9/12 algorithms were considered hub genes by the CytoHubba plugin, which uses 12 algorithms to evaluate hub genes. Hub genes were defined as intersections between the genes in the key module and the hub genes screened using the Cytoscape tool. MCODE finds clusters (highly interconnected regions) in a network. Clusters mean different things in different types of networks. For instance, clusters in a protein-protein interaction network are often protein complexes and parts of pathways, while clusters in a protein similarity network represent protein families. Finally 16 HUB GENES were identified, they are Trib3, Atf5, Chac1, Ddit3, F3, F7, F2r, Cpa3, Mpo, Elane, Phgdh, Psat1, Shmt2, Cd44, Thy1, Siglec5.

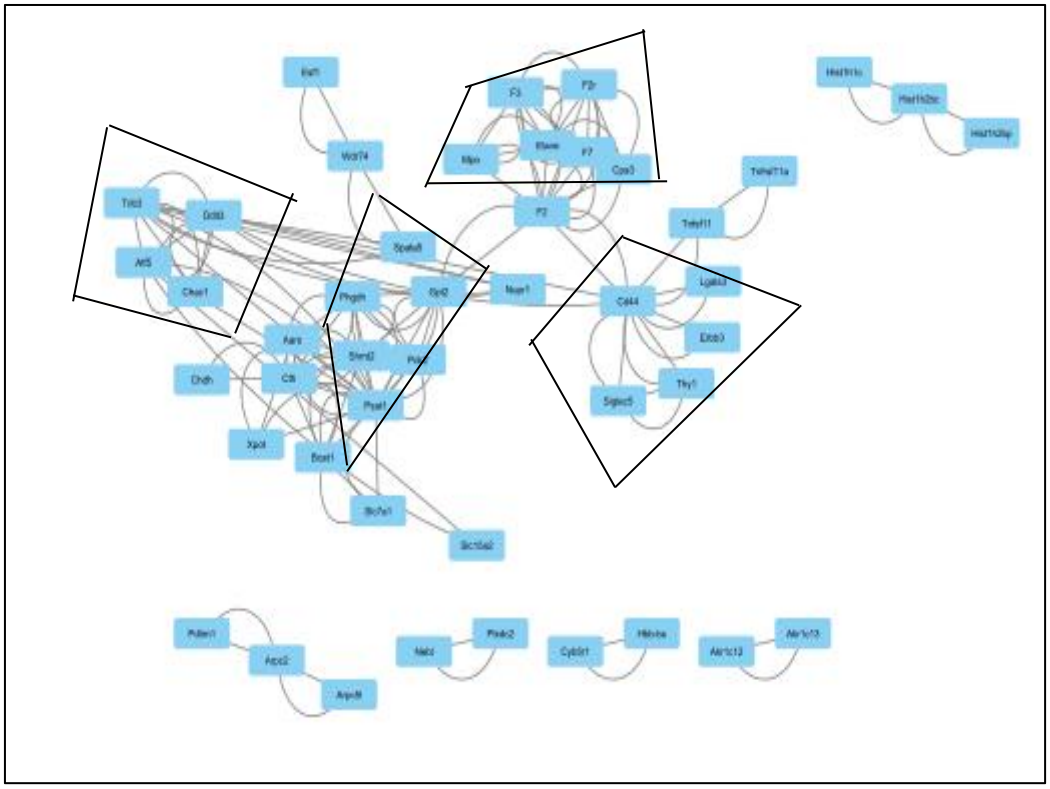


FIGURE:13 The cytoscape analysis of 86 upregulated genes of AML with four clusters

The Molecular Complex Detection (MCODE) is used to identify the merged network's hub genes. The clusters were formed with nodes of hub genes present in them. Clusters of hubs integrating PPI subnetworks comprise the backbone of a network. The essentiality of gene hubs in preserving the integrity of the interacting network is quantitatively expressed in network centrality statistics. Clusters of hubs integrating PPI subnetworks comprise the backbone of a network. The essentiality of gene hubs in preserving the integrity of the interacting network is quantitatively expressed in network centrality statistics. For sub-networks of hubs higher centrality values and linear relationships between the centrality statistics of the network hubs refects a higher number of reported biologically meaningful associations between the hubs and the other genes in the sub-networks and main network. There was a strong correspondence in the network enrichment analyses derived from PPI networks in DEGs supporting the non-random nature of methylation signals within protein-coding regions in signaling pathways linked to cancer development.

CLUSTER :1

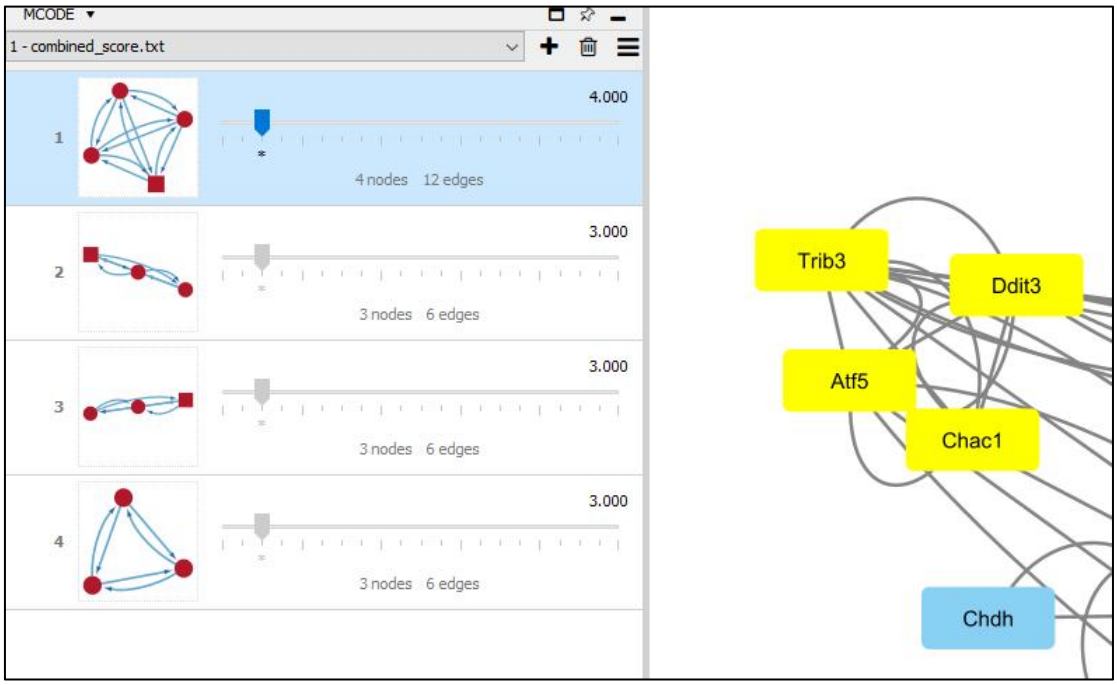


FIGURE:14 The clusters formed using MCODE pluggin in CYTOSCAPE analysis of AML sample.

GENE:1-Trib3:

TRIB3 (Tribbles Pseudokinase 3) is a Protein Coding gene. Diseases associated with TRIB3 include Type 2 Diabetes Mellitus and Acute Promyelocytic Leukemia. Gene Ontology (GO) annotations related to this gene include transferase activity, transferring phosphorus-containing groups and protein kinase binding. Expression of TRIB3 in acute myeloid leukemia.

GENE:2-Atf5:

Activating transcription factor 5 (ATF5) is a cellular prosurvival transcription factor within the basic leucine zipper (bZip) family that is involved in cellular differentiation and promotes cellular adaptation to stress. The importance of ATF5 in the development of leukemia first became evident during analysis of its role in the survival of myeloid progenitor cells. It was found that upon IL-3 deprivation, myeloid progenitor cells underwent apoptosis and subsequent gene expression profiling revealed ATF5 as the most highly downregulated gene.

GENE:3-Chac1:

CHAC1 is involved in the regulation of inflammation in bronchial cells during *Pseudomonas aeruginosa* infection and may explain the excessive inflammation present in the respiratory tracts of patients with cystic fibrosis.CHAC1 was one of the genes most dramatically induced by farnesol in Molt4 cells. Although not well-studied, a recent report demonstrated that CHAC1 functions as a pro-apoptotic protein in endothelial cells and its expression was found to be directly regulated by ATF4 and ATF3 through a bipartite ATF/CRE site in its promoter.

GENE:4-Ddit3:

The protein functions as a dominant-negative inhibitor by forming heterodimers with other C/EBP members, such as C/EBP and LAP (liver activator protein), and preventing their DNA binding activity. The overexpression of DDIT3 transcript has been found to induce increased apoptosis of myeloid cells and block cells in the progression from G1 to S phase . The level of DDIT3 transcript has been revealed down-regulated in myeloid malignancies in our previous study.

CLUSTER :2

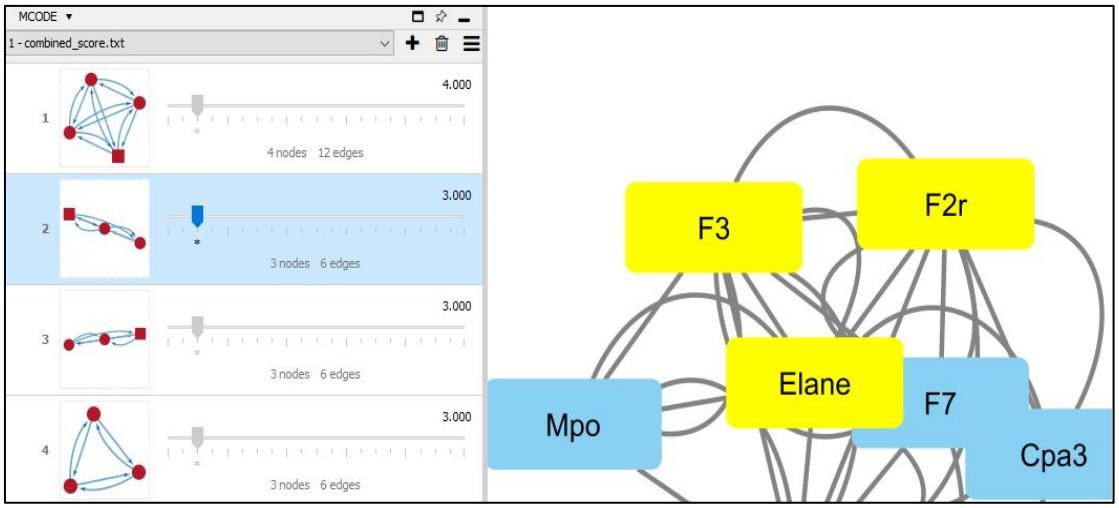


FIGURE:15 The clusters formed using MCODE pluggin in CYTOSCAPE analysis of AML sample.

GENE:1-F3:

This gene encodes coagulation factor III which is a cell surface glycoprotein. This factor enables cells to initiate the blood coagulation cascades, and it functions as the

high-affinity receptor for the coagulation factor VII.F3 expression levels seem to be highly relevant as marker for the prediction of ATRA response in non-APL AML patients.

GENE:2-F2r:

F2R Gene - Coagulation Factor II Thrombin Receptor Coagulation factor II receptor is a 7-transmembrane receptor involved in the regulation of thrombotic response. Proteolytic cleavage leads to the activation of the receptor.F2r knockdown not only suppresses β-catenin but also activates JNK signaling. Consistently, our Western blot analysis revealed activation of JNK in response to inhibition of F2r. Sustained JNK activation has been reported in many types of AML cells and promotes survival signals during leukemia development .

GENE:3-Elane:

The ELANE gene provides instructions for making a protein called neutrophil elastase. This protein is found in neutrophils, a type of white blood cell that plays a role in inflammation and in fighting infection. When the body starts an immune response to fight an infection, neutrophils release neutrophil elastase. The role of ELANE in leukemia progression, the expression level of ELANE, as well as correlation and survival analyses, was performed in patients with leukemia. Gene expression of ELANE on the normal monocytes and myeloid leukemia samples was obtained and compared via Oncomine

CLUSTER :3

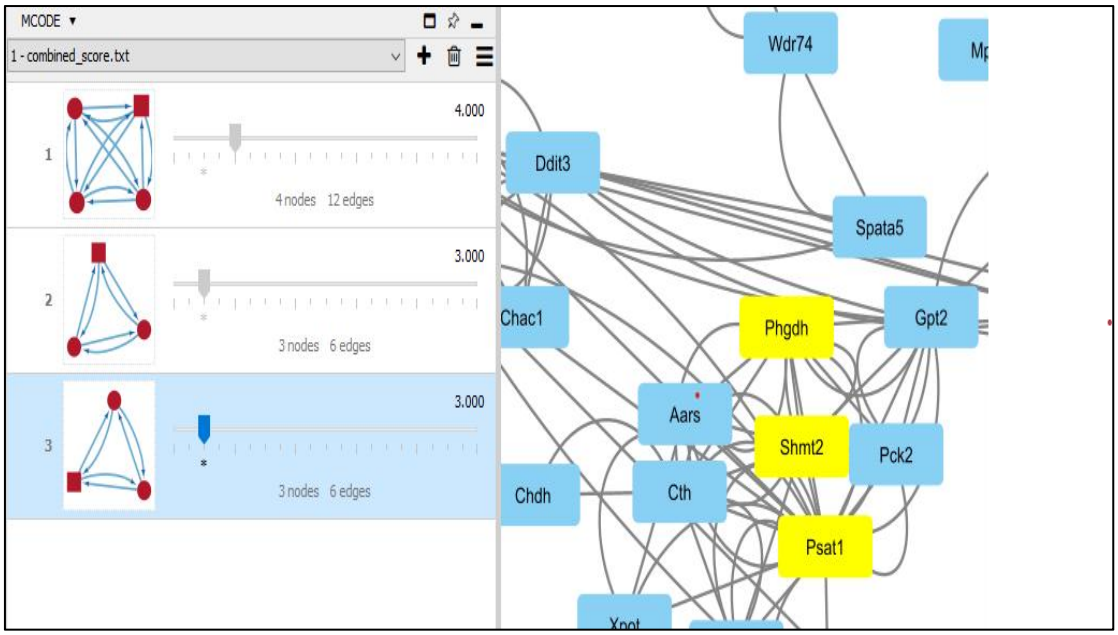


FIGURE:16 The clusters formed using MCODE pluggin in CYTOSCAPE analysis of AML sample.

GENE:1-Phgdh:

The PHGDH gene provides instructions for making the parts (subunits) that make up the phosphoglycerate dehydrogenase enzyme. Four PHGDH subunits combine to form the enzyme. This enzyme is involved in the production (synthesis) of the protein building block (amino acid) serine. Phgdh is a downstream target gene of P2X7, overexpressed Phgdh in P2x7-null AML cells and transplanted these cells into lethally irradiated recipient mice. The overexpression of Phgdh in the P2x7-null AML cells fully reversed the extended survival of the recipient mice to levels comparable to those of the recipient mice transplanted with WT leukemia cells.

GENE:2-Shmt2:

Loss of SHMT2 was necessary for impaired growth and cell cycle arrest, with suppression of both SHMT1 and SHMT2 inhibiting leukemia progression in vivo. This gene encodes the mitochondrial form of a pyridoxal phosphate-dependent enzyme that catalyzes the reversible reaction of serine and tetrahydrofolate to glycine

and 5,10-methylene tetrahydrofolate. The encoded product is primarily responsible for glycine synthesis.

GENE:3-Psat1:

PSAT1 is the protein-coding gene of phosphoserine aminotransferase, which catalyzes serine biosynthesis. PSAT1 is overexpressed in colon cancers where it contributes to cell proliferation and chemoresistance, which result in a poor prognosis. Psat1 mutations have been identified in a number of cancer types, especially in gliomas and acute myelogenous leukemia (AML).

CLUSTER :4

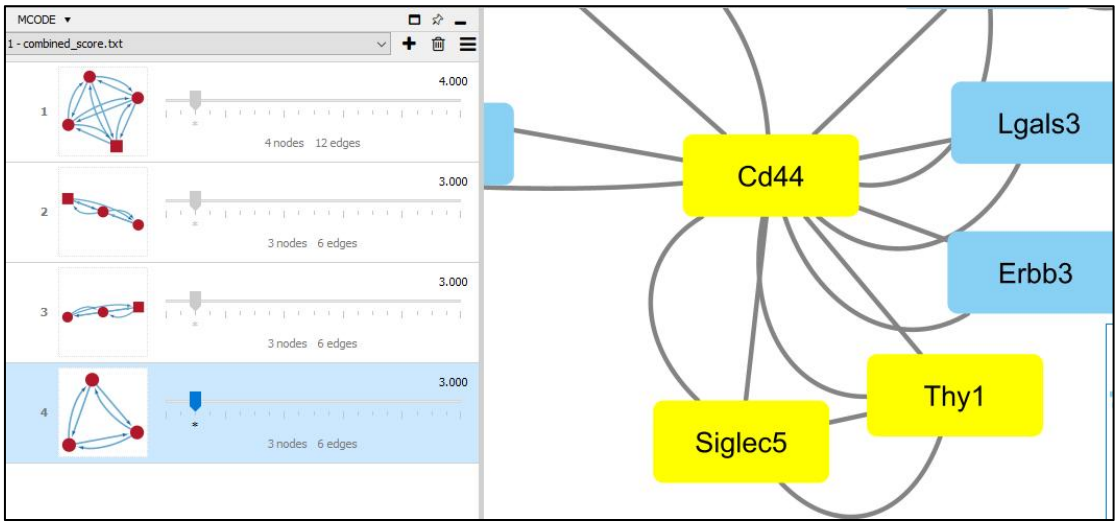


FIGURE:17 The clusters formed using MCODE pluggin in CYTOSCAPE analysis of AML sample.

GENE:1-Cd44:

Known functions of CD44 are cellular adhesion (aggregation and migration), hyaluronate degradation, lymphocyte activation, lymph node homing, myelopoiesis and lymphopoiesis, angiogenesis, and release of cytokines. Patients of the CD44-high group presented with an increased incidence of diffuse BM infiltration, which is a negative prognostic marker itself .

GENE:2-Thy1:

The encoded protein is involved in cell adhesion and cell communication in numerous cell types, but particularly in cells of the immune and nervous systems. In most cases AML progenitor cells which have long-term proliferative capacity in vitro and the ability to repopulate sublethally irradiated NOD/SCID mice have the phenotype $CD34^{+}$/Thy-1^{-}.

GENE:3-Siglec5:

SIGLEC5 (Sialic Acid Binding Ig Like Lectin 5) is a Protein Coding gene. Diseases associated with SIGLEC5 include Ophthalmia Neonatorum and Ovarian Cystic Teratoma. Among its related pathways are Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell and Innate Immune System.Siglec-5 expression can be used to detect low concentrations of AML cells in human bone marrow specimens, and functions as a potential target for leukemic therapy. The developed probes and identified Siglec-5 protein may potentially be used for leukemic cell detection and therapy in our future clinical practice.

4.5.2 Identification of hub genes in CML using CYTOSCAPE:

The Molecular Complex Detection (MCODE) tool in cytoscape was used to filter the major modules from the PPI network, with the degree cut-off set at 5. CytoHubba, a cytoscape programme, was used to find the hub genes from the PPI network. The CytoHubba plugin, which uses 12 algorithms to evaluate hub genes, considered the top HUB genes reported in at least 9/12 algorithms to be hub genes. The intersections between the genes in the key module and the hub genes screened with the Cytoscape programme were classified as hub genes. The backbone of a network is made up of clusters of hubs that connect PPI subnetworks. Gene hubs' importance in maintaining the integrity of the interacting network is quantified in network centrality statistics. For sub-networks of hubs higher centrality values and linear relationships between the centrality statistics of the network hubs refects a higher number of reported biologically meaningful associations between the hubs and the other genes in the sub-networks and main network. There was strong correspondence in the network enrichment analyses derived from PPI networks in DEGs supporting the non-random nature of methylation signals within protein-coding regions in signaling pathways

linked to cancer development. Finally 17 HUB GENES were identified ,they are ARF1, ARF3, ANK1, ANK2, ANK3, RING1, L3MBTL1, EHMT1, HIST1H4F, ALPP, ALPPL2, AMY2B, AMY2A, AMPD3, ADARB1, AMPD1, AMPD2.

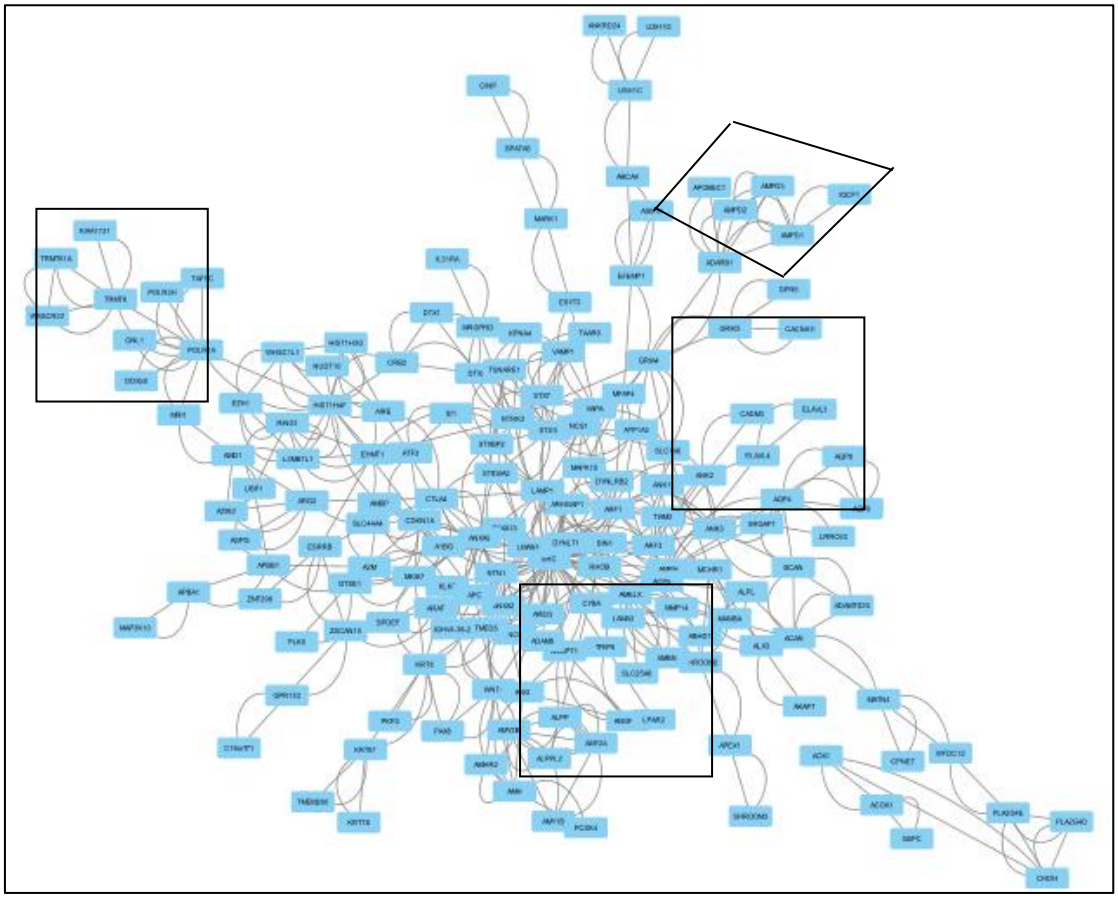


FIGURE:18 It is the cytoscape analysis of 37 upregulated genes of CML with four clusters

The Molecular Complex Detection (MCODE) is used to identify the merged network's hub genes. The clusters were formed with nodes of hub genes present in them. Clusters of hubs integrating PPI subnetworks comprise the backbone of a network. The essentiality of gene hubs in preserving the integrity of the interacting network is quantitatively expressed in network centrality statistics.

CLUSTER :1

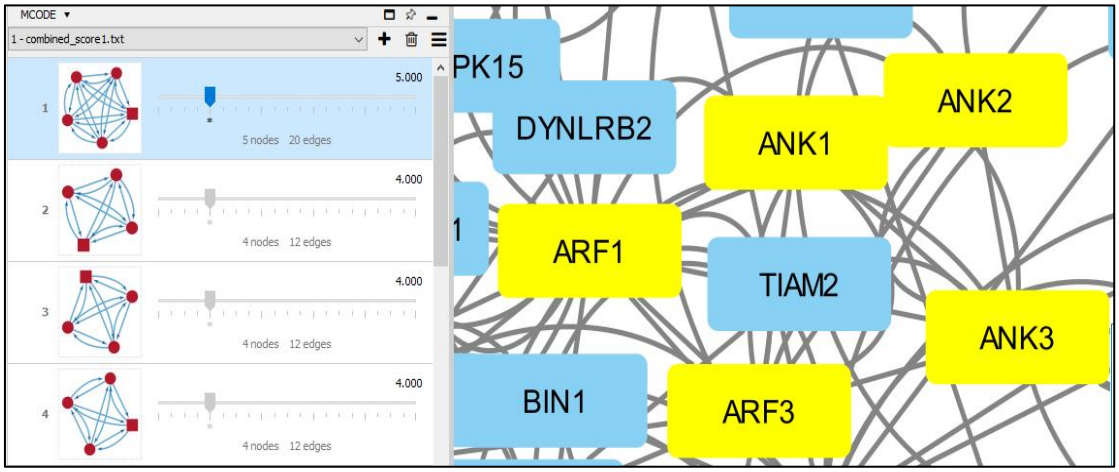


FIGURE:19 The clusters formed using MCODE pluggin in CYTOSCAPE analysis of CML sample.

GENE:1-ARF1:

ADP-ribosylation factor 1 (ARF1) is a member of the human ARF gene family. The family members encode small guanine nucleotide-binding proteins that stimulate the ADP-ribosyltransferase activity of cholera toxin and play a role in vesicular trafficking as activators of phospholipase D.Bcr-Abl(+), $Arf^{+/-}$ pre-B cells induced lympholeukemias in which the WT Arf allele, but not the closely linked Ink4a gene, was inactivated.

GENE:2-ARF3:

ADP-ribosylation factor 3 (ARF3) is a member of the human ARF gene family. These genes encode small guanine nucleotide-binding proteins that stimulate the ADP-ribosyltransferase activity of cholera toxin and play a role in vesicular trafficking and as activators of phospholipase D. *ARF3* encodes a guanine nucleotide-binding protein functioning in vesicular trafficking, phospholipase D activation and intracellular transport.

GENE:3-ANK1:

The ANK1 gene provides instruction for making a protein called ankyrin-1.This protein is primarily active (expressed) in red blood cells, but it is also found in muscle and brain cells. In red blood cells, ankyrin-1 is located at the cell membrane, where it

attaches (binds) to other membrane proteins. The down-regulated genes were leukemia- associated *GATA3*, *EPOR*, *ANK1*, and *HEMGN.*

GENE:4-ANK2:

The ANK2 gene provides instructions for making a protein called ankyrin-B. Ankyrin-B is part of a family of ankyrin proteins, which interact with many other types of proteins in cells throughout the body. The Ank2 protein is selectively targeted to premyelinated axons, and is possibly involved in neuronal migration.

GENE:5-ANK3:

ANK3 Gene - Ankyrin 3Ankyrins are a family of proteins that are believed to link the integral membrane proteins to the underlying spectrin-actin cytoskeleton and play key roles in activities such as cell motility, activation, proliferation, contact, and the maintenance of specialized membrane domains. The ANK3 gene encodes a protein that is required to direct the localization of potassium channels to the axons and nodes of Ranvier. Potassium channels can also regulate cancer cell migration via noncanonical functions. The best-illustrated example is the HERG channel, which can form a macromolecular complex with VEGFR-1 and β1 integrin. The assembly of this protein complex confers a pro-migratory phenotype in leukemia, an effect that can be inhibited by application of a HERG channel blocker.

CLUSTER :2

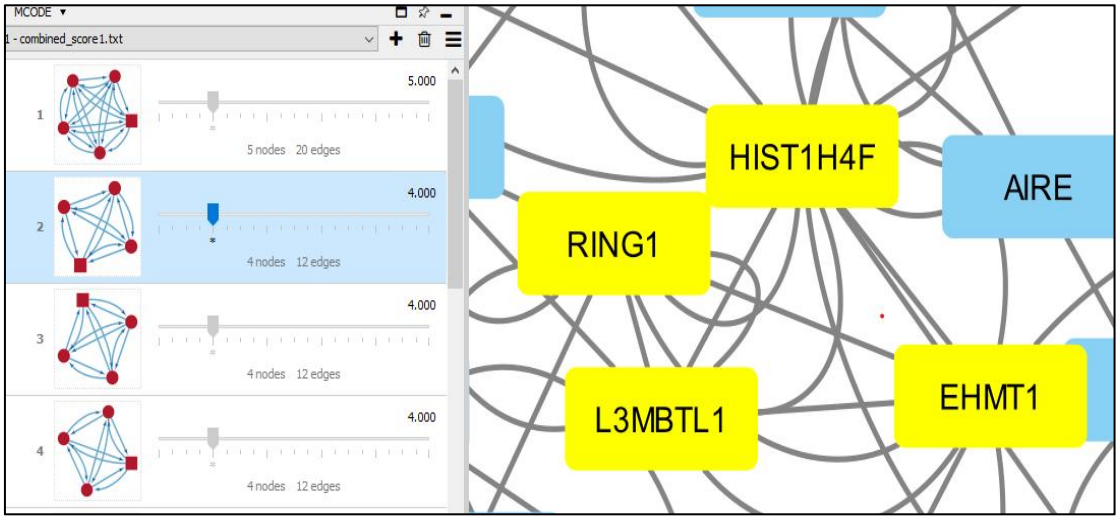


FIGURE:20 The clusters formed using MCODE pluggin in CYTOSCAPE analysis of AML sample.

GENE:1-RING1:

This gene belongs to the RING finger family, members of which encode proteins characterized by a RING domain, a zinc-binding motif related to the zinc finger domain. The RING1 expression was positively correlated with the proportion of bone marrow cells, reflecting the degree of tumor load in MDS patients.

GENE:2-L3MBTL1:

This gene represents a polycomb group gene. The encoded protein functions to regulate gene activity, likely via chromatin modification. The encoded protein may also be necessary for mitosis.L3MBTL1 loss can induce erythroid differentiation in leukemia cells as well as in normal human HSPCs.

GENE:3-EHMT1:

The EHMT1 gene provides instructions for making an enzyme called euchromatic histone methyltransferase 1. Histone methyl transferases are enzymes that modify proteins called histones. Histones are structural proteins that attach (bind) to DNA and give chromosomes their shape. Phosphorylation of EHMT2/EHMT1 by AURKB prevents CBX3 recruitment, reduces death gene activation by GC, and reduces leukemia cell death.

GENE:4-HIST1H4F:

HIST1H4F cDNA / gene is a gene with protein product which located on 6p22.2. The HIST1H4F gene is conserved in chimpanzee, Rhesus monkey, cow, mouse, rat, zebrafish, fruit fly, and frog. Organisms have orthologs with human gene HIST1H4F.The difference inH3K27me3 enrichment at the 5 “variable” HIST1 regions (HIST1H2BG-2AE, HIST1H1D,HIST1H4F, HIST1H4G and HIST1H3F-2BH) but not in the two flanking regions HIST1H1Eand HIST1H4E, in two “high H3K27me3” and in two patients.

CLUSTER :3

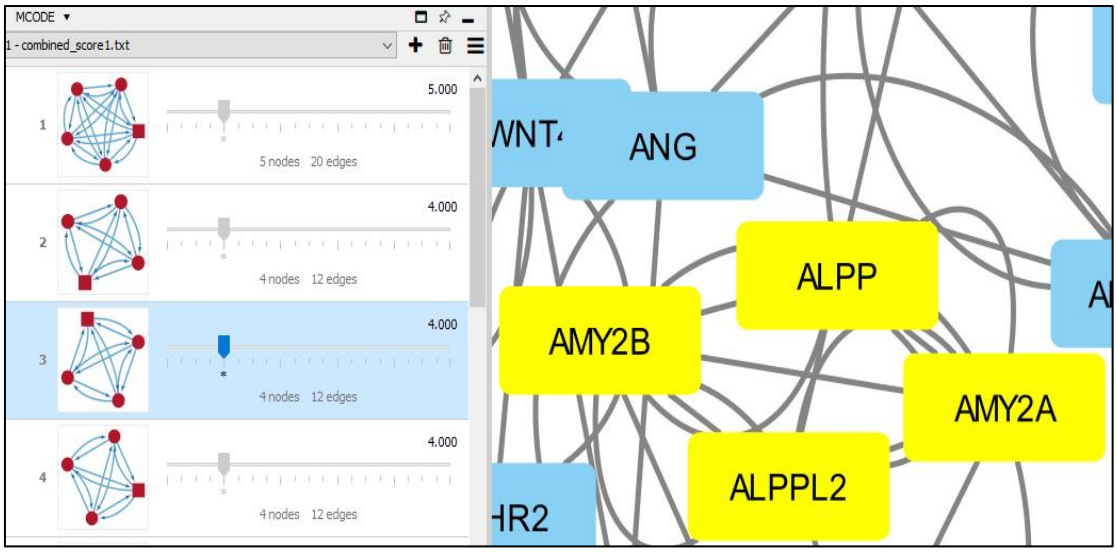


FIGURE:21 The clusters formed using MCODE pluggin in CYTOSCAPE analysis of CML sample.

GENE:1-ALPP:

The protein encoded by this gene is an alkaline phosphatase, a metalloenzyme that catalyzes the hydrolysis of phosphoric acid monoesters. It belongs to a multigene family composed of four alkaline phosphatase isoenzymes. The expression of APP mRNA among AML subtypes was significantly different(*P*=0.019), M2 with t(8;21) was the highest expression subtype and M5b was the lowest. APP expression had no significant effect on AML clinical characteristic excepting AML subtypes.

GENE:2-ALPPL2:

Alkaline phosphatase, placental-like 2 is a protein that in humans is encoded by the ALPPL2gene. It contains HAAO gene and the second one includes ALPPL2, ALPI, ECEL1, CHRND, CHRNG, TIGD1, and EIF4E2 genes. None of these genes was noted as the component of cancerogenesis.

GENE:3-AMY2B:

Alpha-amylase 2B is an enzyme that in humans is encoded by the AMY2B gene. On 8 chronic myeloid leukemia patients who developed serum lipase/amylase elevation during treatment with nilotinib. The cause of pancreatitis was suggested to be AML invasion because induction chemotherapy improved both pancreatitis and leukemia.

GENE:4-AMY2A:

This gene encodes a member of the alpha-amylase family of proteins. Amylases are secreted proteins that hydrolyze 1,4-alpha-glucoside bonds.AMY2A gene has not been previously documented as being deleted in GC, lack of the AMY2Agene has been described in connection with human cancers.

CLUSTER :4

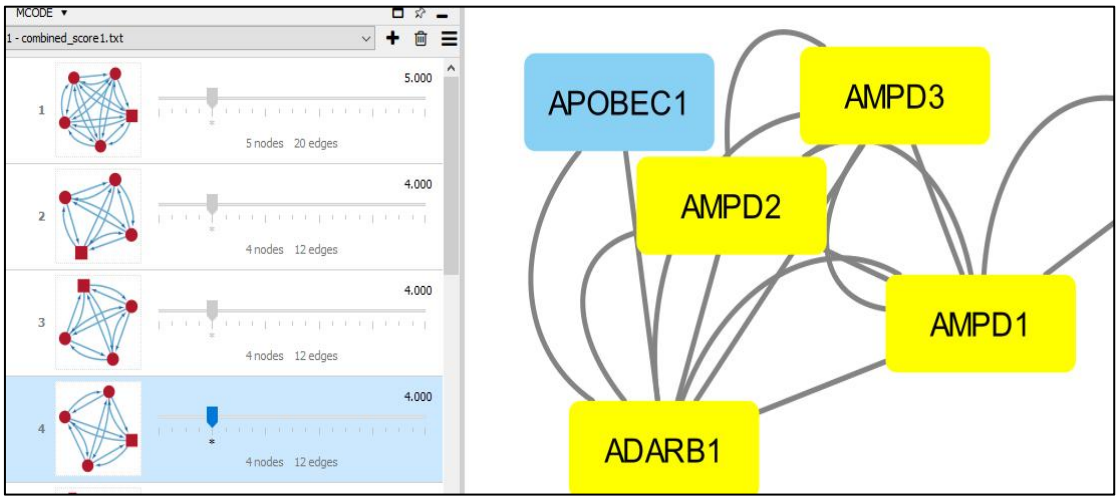


FIGURE:22 The clusters formed using MCODE pluggin in CYTOSCAPE analysis of CML sample.

GENE:1-AMPD3:

AMP deaminase 3 is an enzyme that in humans is encoded by the *AMPD3* gene. This gene encodes a member of the AMP deaminase gene family. In AML-AK patients, besides lgf2, h19 and impact genes, ampd3 and gabrb3were also significantly upregulated than in healthy individuals ($p< 0.0001$).

GENE:2-ADARB1:

This gene encodes the enzyme responsible for pre-mRNA editing of the glutamate receptor subunit B by site-specific deamination of adenosines.The RNA-specific deaminase ADARB1 leads to reduced pri-miRNA processing, but full processing efficiency is recovered upon deletion of the RNA-binding domains or nuclear localization of ADARB1. Thus, we show that, apart from genomic deletion or transcriptional downregulation, aberrant processing of miRNA leads to specific

reduction of miRNAs in leukemic cells. This represents a novel oncogenic mechanism in the pathogenesis of CML.

GENE:3-AMPD1:

The AMPD1 gene provides instructions for producing an enzyme called adenosine monophosphate (AMP) deaminase. The expression of AMPD1 was markedly associated with disease outcome and tumor-infiltrating immune cells (TICs). The expression of AMPD1 was markedly associated with disease outcome and tumor-infiltrating immune cells (TICs).

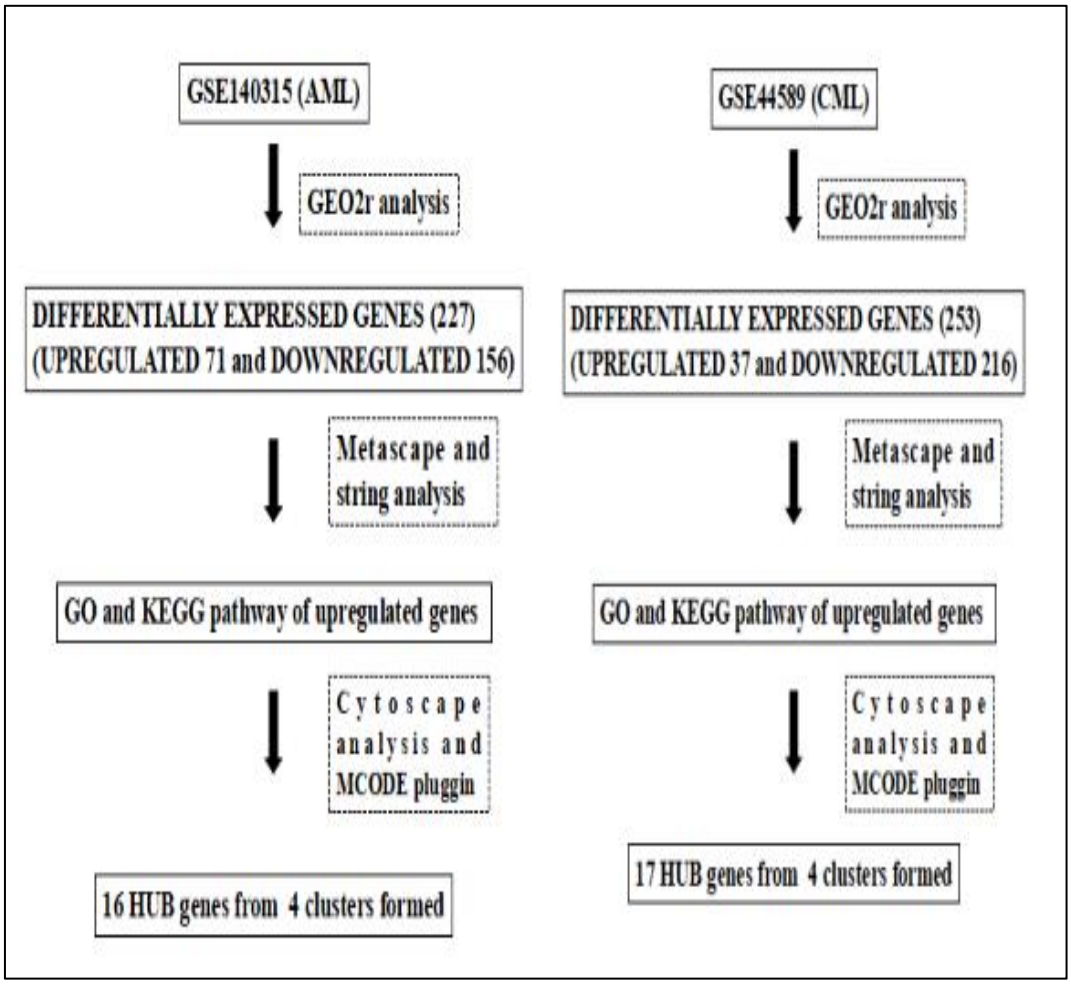
GENE:4-AMPD2:

The AMPD2 gene encodes adenosine monophosphate deaminase-2 (EC 3.5. 4.6), an enzyme that catalyzes the deamination of AMP to IMP and plays an important role in the purine nucleotide cycle. Early amplification of the *AMPD2* gene is based on unequal gene segregation at mitosis, rather than local over-replication causes mutations. Mutations in genes can cause cancer by accelerating cell division rates or inhibiting normal controls on the system.The discovery of tumour biomarkers is one of the methods for early detection of cancer, however the regulation of this extremely complex tumour cannot be met solely by a local emphasis on a single or specific gene. Some genes in tumours are abnormally expressed and intimately connected to many other genes, according to the entire regulatory network. The hub genes identified can be served as a molecular markers for the diagnosis of leukemia.

KEGG enrichment analysis of IRGs revealed that CML involved Th17 cell differentiation, the NF-kappa B signaling pathway, and cytokine-cytokine receptor interaction. A total of 10 hub genes were selected using the PPI network. GSEA showed that these hub genes were related to the gamma-interferon immune response, inflammatory response, and allograft rejection. ROC curve analysis suggested that six hub genes may be potential biomarkers for CML diagnosis. Further analysis indicated that immune cells were associated with the pathogenesis of CML which was reported by Fangyi Yao et al. Finally 16 HUB GENES were identified, they are Trib3, Atf5, Chac1, Ddit3, F3, F7, F2r, Cpa3, Mpo, Elane, Phgdh, Psat1, Shmt2, Cd44, Thy1, Siglec5 in AML and 17 HUB GENES were identified ,they are ARF1, ARF3, ANK1, ANK2, ANK3, RING1, L3MBTL1, EHMT1, HIST1H4F, ALPP, ALPPL2, AMY2B,

AMY2A, AMPD3, ADARB1, AMPD1, AMPD2 in CML.These hub genes can act as a diagnostic markers for leukemia.

SCHEMATIC REPRESENTATION



5

SUMMARY AND CONCLUSION

Cancer is the second leading cause of mortality worldwide. Every year, around 10 million new cancer patients are diagnosed, resulting in over 6 million deaths, accounting for approximately 12% of world mortality. The number of new cancer cases is expected to rise by over 70% over the next two decades, reaching over 15 million new cases diagnosed year by 2020. This study is to analyze the molecular markers in leukemia sample using in silico analysis. Molecular markers serve as biomarkers for the diagnosis of leukemia . The GSE140315 (GPL11180 [HT_MG-430_PM] Affymetrix HT MG-430 PM Array Plate, GPL16417 Illumina MiSeq (Mus musculus) of AML and GSE44589 (GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) of CML were selected from the database of the Gene Expression Omnibus and compared using the GEO2R tool to identify differentially expressed genes (DEGs). Differentially expressed genes (DEGs) from expression profile datasets were screened using GEO2R, and bioinformatics analysis was performed on the DEGs.

A gene is declared differentially expressed if an observed difference or change in read counts or expression levels between two experimental conditions is statistically significant. In AML 227 genes were analysed, a total of 71 upregulated and 156 downregulated DEGs were identified. In CML 253 genes were analysed, a total of 37 upregulated and 216 downregulated DEGs were identified. The up-regulated genes are mainly components that control cellular proliferation, whereas the down-regulated genes consist of proteins exposed on or secreted from the cell surface. The common DEGS of all the samples of Leukemia was analyzed for their PPIs. Within these DEGs, Further these genes were shortlisted under the condition of those which has its scores more than 0.9 with other genes. In the whole network, there are genes as were having higher and significant interactions with other DEGS.

The biological pathways involved in AML include post-translation modification gamma carboxylation and hypusine formation, gamma-carboxylation, transport, and amino-terminal cleavage of proteins, removal of aminoterminal pro-peptides from gamma-carboxylated proteins, gamma-carboxylation of protein precursors, Transport of gamma-carboxylated protein precursors from the endoplasmic reticulum, extrinsic pathway. Acute myeloid leukemia (AML) is the result of a multistep transforming process of hematopoietic precursor cells (HPCs) which enables them to proceed through limitless numbers of cell cycles and to become resistant to cell death. Early studies have implicated over-expression of tissue factor (TF), the principal initiator of the extrinsic coagulation protease cascade, in the pathogenesis of AML-associated DIC. The commercial TF ELISA used in our study does not detect alternatively spliced TF, future studies should therefore investigate, if proteolytic degradation of full-length TF by enzymes liberated from dying myeloblasts represents a novel mechanism of controlling the extrinsic coagulation pathway in newly diagnosed AML. The biological pathways involved in CML include RNA polymerase lll transcription termination, RNA polymerase lll chain elongation, Nectin/Necl trans heterodimerization, Sodium-coupled phosphate cotransporters, Type llNa+/Pi cotransporters, Choline degradation. The PLC- and PLD mediated pathways of PtdCho breakdown were inactive in HL60 cells suggest that the development of agents which selectively inhibit Ptd Cho-specific PLC or PLD but no inositol lipid-specific PLC may be useful in assessing the role of Ptd Cho breakdown in the regulation of proliferation of leukaemia cells. These agents may find applications in the selective blocking of cell-cycle transit of normal cells whose proliferation is dependent on both inositol lipid- and Ptd Cho-derived signals while allowing the continued proliferation of leukemia cells.

Transcription factor NF-kappaB plays a key regulatory role in the cellular response to pro-inflammatory cytokines such as tumor necrosis factor-alpha (TNF). In the absence of TNF, NF-kappaB is sequestered in the cytoplasm by inhibitory IkappaB proteins. Phosphorylation of IkappaBby the beta-catalytic subunit of IKK, a multicomponent IkappaB kinase, targets the inhibitor for proteolytic destruction and facilitates nuclear translocation of NF-kappaB. This pathway is initiated by TNF-dependent phosphorylation of T loop serines in IKKbeta, which greatly stimulates IkappaB kinase activity. Nuclear condensation, cell shrinkage, membrane blebbing, and DNA fragmentation are all hallmarks of apoptosis, a regulated cellular

suicide mechanism. The central regulators of apoptosis are caspases, a family of cysteine proteases. Clusters of hubs integrating PPI subnetworks comprise the backbone of a network. The essentiality of gene hubs in preserving the integrity of the interacting network is quantitatively expressed in network centrality statistics.

The HUB GENES were identified based on the 4 different clusters formed in the Cytoscape software. In AML, there 16 HUB GENES - Trib3, Atf5, Chac1, Ddit3, F3, F7, F2r, Cpa3, Mpo, Elane, Phgdh, Psat1, Shmt2, Cd44, Thy1, Siglec5. In CML, 17 HUB GENES were identified, they are ARF1, ARF3, ANK1, ANK2, ANK3, RING1, L3MBTL1, EHMT1, HIST1H4F, ALPP, ALPPL2, AMY2B, AMY2A, AMPD3, ADARB1, AMPD1, AMPD2 in CML. These hub genes can act as a diagnostic markers for leukemia.

BIBLIOGRAPHY

Bell, J. A., Galaznik, A., Huelin, R., Stokes, M., Guo, Y., Fram, R. J., and Faller, D,V, (2018), Effectiveness and safety of therapeutic regimens for elderly patients with acute myeloid leukemia: a systematic literature review. Clinical Lymphoma Myeloma and Leukemia, 18(7), e303-e314.

Cerrano, M. and Itzykson, R. (2019), New treatment options for acute myeloid leukemia in 2019, Current oncology reports, 21(2), 1-12.

Chae, Y. C., Jung, H., Kim, J. Y., Lee, D. H., and Seo, S. B. (2019). Ubiquitin-specific peptidase 3 induces TPA-mediated leukemia cell differentiation via regulating H2AK119ub. Animal cells and systems, 23(5), 311-317.

Chen, Y., Jiang, P., Wen, J., Wu, Z., Li, J., Chen, Y., and Hu, J. (2020), Integrated bioinformatics analysis of the crucial candidate genes and pathways associated with glucocorticoid resistance in acute lymphoblastic leukemia, Cancer medicine, 9(8), 2918-2929.

Dong, Y., Shi, O., Zeng, Q., Lu, X., Wang, W., Li, Y., and Wang, Q. (2020), Leukemia incidence trends at the global, regional, and national level between 1990 and 2017, Experimental hematology and oncology, 9(1), 1-11.

Du, W., He, J., Zhou, W., Shu, S., Li, J., Liu, W., and Hu, D, (2019), High IL2RA mRNA expression is an independent adverse prognostic biomarker in core binding factor and intermediate-risk acute myeloid leukemia. Journal of translational medicine, 17(1), 1-14.

Fu, W., Cheng, G. B., Ding, Y., Deng, Y. J. and Guo, P. X. (2020), Identification of hub genes and its correlation with the prognosis of acute myeloid leukemia based on high-throughput data analysis, Precision Radiation Oncology, 4(2), 49-56.

Fujita, T. C., Sousa-Pereira, N., Amarante, M. K., and Watanabe, M. A. E. (2021), Acute lymphoid leukemia etiopathogenesis, Molecular Biology Reports, 48(1), 817-822.

Gao, C., Zhou, C., Zhuang, J., Liu, L., Wei, J., Liu, C., and Sun, C. (2019), Identification of key candidate genes and miRNA‐mRNA target pairs in chronic lymphocytic leukemia by integrated bioinformatics analysis. Molecular Medicine Reports, 19(1) 362-374.

George, B., Cortes, J., Cabrero, M., Jain, N., Konopleva, M., Andreeff, M., and Ravandi, F. (2015), Pretreatment FMS like Tyrosine kinase-3 Internal Tandem duplication (FLT3-ITD) Mutant Allele Burden Does Not Predict Response to the Combination of Sorafenib and Azacytidine in Patients with Relapsed or High Risk Elderly Untreated Acute Myeloid Leukemia, Clinical Lymphoma, Myeloma and Leukemia, 15, S196-S197.

Godoy, P. B. G., Simionato, N. M., de Mello, C. B., and Suchecki, D. (2020), Assessment of executive functions after treatment of childhood acute lymphoid leukemia: A systematic review, Neuropsychology Review, 30(3), 386-406.

Gu, S., Zi, J., Han, Q., Song, C., and Ge, Z. (2020), Elevated TNFRSF4 gene expression is a predictor of poor prognosis in non-M3 acute myeloid leukemia, Cancer cell international, 20(1), 1-13.

Hallek, M. (2019), Chronic lymphocytic leukemia: 2020 update on diagnosis, risk stratification and treatment, American journal of hematology, 94(11), 1266-1287.

Heendeniya, S. N., Keerthirathna, L., Manawadu, C. K., Dissanayake, I. H., Ali, R., Mashhour, A., and Peiris, D. C. (2020), Therapeutic efficacy of Nyctanthes arbor-tristis flowers to inhibit proliferation of acute and chronic primary human leukemia cells, with adipocyte differentiation and in silico analysis of interactions between survivin protein and selected secondary metabolites, Biomolecules, 10(2), 165.

Hijiya, N. and Suttorp, M. (2019), How I treat chronic myeloid leukemia in children and adolescents, Blood, The Journal of the American Society of Hematology, 133(22), 2374-2384.

Honma, K., Kamikubo, M., Mochizuki, K., and Goda, T. (2017), Insulin-induced inhibition of gluconeogenesis genes, including glutamic pyruvic transaminase 2, is associated with reduced histone acetylation in a human liver cell line, Metabolism, 71, 118-124.

Ionita-Laza, I., Xu, B., Makarov, V., Buxbaum, J. D., Roos, J. L., Gogos, J. A., and Karayiorgou, M. (2014), Scan statistic-based analysis of exome sequencing data identifies FAN1 at 15q13. 3 as a susceptibility gene for schizophrenia and autism, Proceedings of the National Academy of Sciences, 111(1), 343-348.

Jabbour, E. and Kantarjian, H. (2020), Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring, American journal of hematology, 95(6), 691-709.

Jahwar, A. and Ahmed, N. (2021), Swarm intelligence algorithms in gene selection profile based on classification of microarray data: a review, Journal of Applied Science and Technology Trends, 2(01), 01-09.

Kanehisa, M. and Sato, Y. (2020), KEGG Mapper for inferring cellular functions from protein sequences, Protein Science, 29(1), 28-35.

Kruger, T., Schoor, O., Lemmel, C., Kraemer, B., Reichle, C., Dengjel, J., ... and Stevanović, S. (2015), Lessons to be learned from primary renal cell carcinomas: novel tumor antigens and HLA ligands for immunotherapy, Cancer Immunology, Immunotherapy, 54(9), 826-836.

Kumar, S. (2020), Covid-19: A drug repurposing and biomarker identification by using comprehensive gene-disease associations through protein-protein interaction network analysis, 15, S196-S197.

Lai, C., Vadakekolathu, J., Reeder, S., Church, S. E., Hood, T., Aldoss, I., and Rutella, S. (2020), TP53 abnormalities correlate with immune infiltration and are associated with response to flotetuzumab, an investigational immunotherapy, in acute myeloid leukemia, bioRxiv 25, 106-117.

Leahy, A. B., Newman, H., Li, Y., Liu, H., Myers, R., DiNofia, A., and Maude, S. L. (2021), CD19-targeted chimeric antigen receptor T-cell therapy for CNS relapsed or refractory acute lymphocytic leukaemia: a post-hoc analysis of pooled data from five clinical trials, The Lancet Haematology, 8(10), e711-e722.

Li, T., Gao, X., Han, L., Yu, J. and Li, H. (2018), Identification of hub genes with prognostic values in gastric cancer by bioinformatics analysis, World journal of surgical oncology, 16(1), 1-12.

Mirzaei, H., Fathullahzadeh, S., Khanmohammadi, R., Darijani, M., Momeni, F., Masoudifar, A., and Mirzaei, H. R. (2018), State of the art in microRNA as diagnostic and therapeutic biomarkers in chronic lymphocytic leukemia, Journal of Cellular Physiology, 233(2), 888-900.

Modarres, P., Mohamadi Farsani, F., Nekouie, A. A. Vallian, S. (2021), Meta-analysis of gene signatures and key pathways indicates suppression of JNK pathway as a regulator of chemo-resistance in AML, Scientific reports, 11(1), 1-16.

Montserrat, E. and Gale, R. P. (2019), Predicting the outcome of patients with chronic lymphocytic leukemia: Progress and uncertainty, Cancer, 125(21), 3699-3705.

Narendra, G., Raju, B., Verma, H., and Silakari, O. (2021), Identification of potential genes associated with ALDH1A1 overexpression and cyclophosphamide resistance in chronic myelogenous leukemia using network analysis, Medical Oncology, 38(10), 1-10.

Ni, J., Wu, Y., Qi, F., Li, X., Yu, S., Liu, S., and Zheng, Y. (2020), Screening the cancer genome atlas database for genes of prognostic value in acute myeloid leukemia, Frontiers in oncology, 9, 1509.

Oehler, V. G., Yeung, K. Y., Choi, Y. E., Bumgarner, R. E., Raftery, A. E., and Radich, J. P. (2019), The derivation of diagnostic markers of chronic myeloid leukemia progression from microarray data, Blood, The Journal of the American Society of Hematology, 114(15), 3292-3298.

Panda, M. (2020), Elephant search optimization combined with deep neural network for microarray data analysis, Journal of King Saud University-Computer and Information Sciences, 32(8), 940-948.

Papaioannou, D., Ozer, H. G., Nicolet, D., Urs, A. P., Herold, T., Mrózek, K., ... and Garzon, R. (2022), Clinical and molecular relevance of genetic variants in the non-coding transcriptome of patients with cytogenetically normal acute myeloid leukemia. Haematologica, 107(5), 1034.

Puumala, S. E., Ross, J. A., Aplenc, R., and Spector, L. G. (2020), Epidemiology of childhood acute myeloid leukemia, Pediatric blood and cancer, 60(5), 728-733.

Rady, I., Bloch, M. B., Chamcheu, R. C. N., Banang Mbeumi, S., Anwar, M. R.,Mohamed, H., and Chamcheu, J. C. (2018), Anticancer properties of Graviola (*Annona muricata*): a comprehensive mechanistic review, Oxidative medicine and cellular longevity, 205(8), 560-654.

Rao, V. S., Srinivas, K., Sujini, G. N., and Kumar, G. N. and (2014), Protein-protein interaction detection: methods and analysis, International journal of proteomics, 12(9), 2696.

Robison, L. L. (2021), Late effects of acute lymphoblastic leukemia therapy in patients diagnosed at 0-20 years of age, Hematology 2010, the American Society of Hematology Education Program Book, (1) 2011, 238-242.

Sanchez, R., and Mackenzie, S. A. (2020), Integrative network analysis of differentially methylated and expressed genes for biomarker identification in leukemia, Scientific reports, 10(1), 1-16.

Shi, M., Su, R. J., Parmar, K. P., Chaudhry, R., Sun, K., Rao, J. and Chen, M. (2019), CD123: a novel biomarker for diagnosis and treatment of leukemia. Cardiovascular and Haematological Disorders-Drug Targets (Formerly Current Drug Targets-Cardiovascular and Hematological Disorders), 19(3), 195-204.

Shin, W. H., Kumazawa, K., Imai, K., Hirokawa, T., and Kihara, D. (2020), Current challenges and opportunities in designing protein–protein interaction targeted drugs, Advances and Applications in Bioinformatics and Chemistry: AABC, 13, 11.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021), Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: a cancer journal for clinicians, 71(3), 209-249.

Takam Kamga, P., Collo, G. D., Resci, F., Bazzoni, R., Mercuri, A., Quaglia, F. M., and Krampera, M. (2019), Notch signaling molecules as prognostic biomarkers for acute myeloid leukemia, Cancers, 11(12), 1958.

Takam Kamga, P., Dal Collo, G., Cassaro, A., Bazzoni, R., Delfino, P., Adamo, A., and Krampera, M. (2020), Small molecule inhibitors of microenvironmental Wnt/β-catenin signaling enhance the chemosensitivity of acute myeloid leukemia Cancers, 12(9), 2696.

Tan, Y., Zheng, L., Du, Y., Zhong, Q., Zhu, Y., Liu, Z., and Zhang, Q. (2020), Identification of the hub genes and pathways involved in acute myeloid leukemia using bioinformatics analysis, Medicine, 99(35).

Turk, S., Turk, C., Akbar, M. W., Kucukkaraduman, B., Isbilen, M., Demirkol Canli, S., and Gure, A. O. (2020), Renin angiotensin system genes are biomarkers for personalized treatment of acute myeloid leukemia with Doxorubicin as well as etoposide, PloS one, 15(11), e0242497.

Wang, J., Vallee, I., Dutta, A., Wang, Y., Mo, Z., Liu, Z., Yang, X. L. (2020), Multi-Omics Database Analysis of Aminoacyl-tRNA Synthetases in Cancer, Genes, 11(11), 1384.

Wang, Y., Hu, F., Li, J. Y., Nie, R. C., Chen, S. L., Cai, Y. Y., Liang, Y. (2020), Systematic construction and validation of a metabolic risk model for prognostic prediction in acute myelogenous leukemia, Frontiers in oncology, 10, 540.

Wong, K. K., Lawrie, C. H., and Green, T. M. (2019), Oncogenic roles and inhibitors of DNMT1, DNMT3A, and DNMT3B in acute myeloid leukaemia, Biomarker insights, 14, 1177271919846454.

World Health Organization. (2019), WHO global report on traditional and complementary medicine 2019, World Health Organization.

Xu, X., Xu, L., Huang, H., Li, J., Dong, S., Jin, L., and Li, L. (2020), Identification of hub genes as biomarkers correlated with the proliferation and prognosis in lung cancer: a weighted gene co-expression network analysis, BioMed research international, 2020.

Xu, Z. J., Gu, Y., Wang, C. Z., Jin, Y., Wen, X. M., Ma, J. C., and Lin, J. (2020), The M2 macrophage marker CD206: a novel prognostic indicator for acute myeloid leukemia, Oncoimmunology, 9(1), 1683347.

Yao, F., Zhao, C., Zhong, F., Qin, T., Li, S., Liu, J., ... and Wang, X. (2022), Bioinformatics analysis and identification of hub genes and immune-related molecular mechanisms in chronic myeloid leukemia, 10, e12616.

Yassin, M. A., Abdulla, M. A. J., Chandra, P., Aldapt, M. B., Hamad, Y., Akiki, S. J., ... and Nur, U. (2019), Chronic myeloid leukemia in adolescents and young adults: a single institute experience.

Yue, C., Xie, S., Zhong, J., Zhao, H., Lin, Z., Zhang, L., and Luo, Y. (2021), SCAMP2/5 as diagnostic and prognostic markers for acute myeloid leukemia, Scientific Reports, 11(1), 1-16.

Zhang, W. J., and Zhu, Z. M. (2020), Association between the rs3751143 polymorphism of P2RX7 gene and chronic lymphocytic leukemia: A meta-analysis. Purinergic Signalling, 16(4), 479-484.

Zhang, W., Zhang, W., Gui, L., Yan, X., Zhou, X., Ma, Y., and Shi, J. (2021), Expression and prognosis of the B7 family in acute myeloid leukemia, Annals of Translational Medicine, 9(20).

Zhang, Z., Wang, L., Wang, Q., Zhang, M., Wang, B., Jiang, K., and Shen, Z. (2020), Molecular characterization and clinical relevance of RNA binding proteins in colorectal cancer, Frontiers in Genetics, 1229.

Zhou, G., Chen, M., Ju, C. J., Wang, Z., Jiang, J. Y., and Wang, W. (2020), Mutation effect estimation on protein–protein interactions using deep contextualized representation learning, NAR genomics and bioinformatics, 2(2), lqaa015.

WEBOGRAPHY

- https://www.ncbi.nlm.nih.gov/geo/
- https://www.ncbi.nlm.nih.gov/geo/geo2r/
- http://metascape.org
- http://www.funrich.org
- http://geneontology.org
- https://www.genome.jp/kegg/
- http://www.kegg.jp/dbget-bin/www_bget?hsa05221
- https://www.kegg.jp/dbget-bin/www.bgethsa05221
- https://string-db.org/
- https://cytoscape.org/