

UNDERSTANDING THE DEVELOPMENTAL BIOLOGY OF PEDIATRIC ACUTE MYELOID LEUKEMIA

Submitted in partial fulfillment of the requirements for the award of
Master of Science in Bioinformatics and Data Science

By

K Milind Joseph (41739010)



**DEPARTMENT OF BIOINFORMATICS
SCHOOL OF BIO AND CHEMICAL ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119**

APRIL – 2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC | 12B Status by UGC | Approved by AICTE
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119
www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **K Milind Joseph (41739010)**, who carried out the final year project entitled "***UNDERSTANDING THE DEVELOPMENTAL BIOLOGY OF PEDIATRIC ACUTE MYELOID LEUKEMIA***" under our supervision from January 2023 to April 2023.

Dr. DANIEL ALEX ANAND
Internal Guide

Dr. H. JEMMY CHRISTY
Head of the Department

Submitted for Viva Voce Examination held on

Internal Examiner

External Examiner

DECLARATION

I, **K. MILIND JOSEPH**, hereby declare that the Project Report entitled **UNDERSTANDING THE DEVELOPMENTAL BIOLOGY OF PEDIATRIC ACUTE MYELOID LEUKEMIA** done by me under the guidance of **Dr. ALEX ANAND DANIEL** at Sathyabama Institute of Science and Technology, Jeppiaar Nagar, Rajiv Gandhi Road, Chennai-600119, Tamil Nadu, India, is submitted in partial fulfillment of the requirements for the award of **Master of Science in Bioinformatics and Data Science**.

DATE:

PLACE:



SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENTS

I am delighted to express my heartfelt acknowledgement to the Board of Management of SATHYABAMA for their kind encouragement and support throughout the duration of my project. Their guidance has been invaluable in ensuring the successful completion of this project, and I am sincerely grateful for their contribution.

I sincerely thank Dr. H. Jemmy Christy, Head of the Department, Dept. of Bioinformatics, for his unwavering support and timely provision of necessary details during the progressive reviews. His support has been instrumental in helping me to overcome obstacles and achieve success in this project.

I am deeply indebted to my Project Guide, Dr. Daniel Alex Anand, for his invaluable guidance, suggestions, and constant encouragement throughout the project. His expertise and insights have been instrumental in shaping my work and ensuring its successful completion.

I also wish to extend my gratitude to my family for their support and assistance in various ways, which have contributed to the successful completion of this project.

K Milind Joseph



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC | 12B Status by UGC | Approved by AICTE

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in

CERTIFICATE

This is to certify that the dissertation entitled **“UNDERSTANDING THE DEVELOPMENTAL BIOLOGY OF PEDIATRIC ACUTE MYELOID LEUKEMIA”** was carried out by **Mr. K Milind Joseph (Reg. No. 41739010)**, Department of Bioinformatics and Data Science, Sathyabama Institute of Science and Technology, Chennai in partial fulfillment of the requirements of the award of the **Degree of Master of Science in Bioinformatics and Data Science 2021-2023**, is a bonafide record of the project work done under my guidance and supervision.

This thesis or part, therefore, has not been submitted elsewhere for any other degree or diploma.

Date:

Place:

Dr. DANIEL ALEX ANAND

ABSTRACT

Pediatric acute myeloid leukemia (AML), observed in children under two years, is a type of blood cancer in which bone marrow produces an abnormal number of cells. Through this study, we have attempted to investigate the omics data that contribute significantly to understanding the onset and pathogenesis of the disease. We have analysed several data to identify the genetic and epigenetic changes involved in the development of pediatric AML. In addition, we have studied the role of developmental pathways in the pathogenesis of pediatric AML. We have analysed the GENE DATA SET of 60 genes extracted from the Leukemia Gene Literature Database for Protein-Protein Interactions (PPI), Gene Regulatory Networks (GRN), Diseases, drugs and chemicals using NetworkAnalyst; The hubs in the list of the genes were visualised from Cytoscape (Cytohubba); Protein network was analysed, and the genes involved in the pathway of AML were visualised using STRINGdb, and the mutations data was collected from cBioPortal and DoCM. In addition, the common genes from the three databases (dbLGL, COSMIC and cBioPortal) were analysed for mutation frequency; the highly mutated genes were further analysed for CNVs, SNPs, miRNA and structural variants. We understood that studying genetic changes and mutations that contribute to the development of AML could help understand the disease's biology and identify new targets for therapy, leading to more effective and personalised treatments. Fusion genes are common in pediatric AML and can be helpful in diagnosis and treatment. Fusion genes, including RUNX1-RUNX1T1, CBFB-MYH11, and MLL-rearrangements, can be found in different subtypes of pediatric AML.

Keywords: *Pediatric acute myeloid leukemia, blood cancer, epigenetic changes, developmental pathways, protein-protein interactions, gene regulatory networks.*

Table of Contents

LIST OF ABBREVIATIONS	ix
LIST OF TABLES	x
CHAPTER 1: INTRODUCTION	1
1.1 BACKGROUND	2
1.2 EXISTING PROBLEMS	3
1.3 TREATMENT	3
1.4 BOTTLENECK AREAS	4
CHAPTER 2: REVIEW OF LITERATURE	6
CHAPTER 3: AIM AND SCOPE OF THE STUDY	16
3.1 AIM	16
3.2 SCOPE	17
CHAPTER 4: METHODOLOGY	18
4.1 GENE DATA SETS	18
4.2 DATABASES	18
4.3 FILTERING THE GENES	19
4.4 GENE NETWORK ANALYSIS	20
4.5 IDENTIFICATION OF HUBS PRESENT IN THE NETWORK	20
4.6 VISUALISING GENERIC-PPI NETWORK, GENE-miRNA INTERACTIONS, TF-miRNA CO-REGULATORY GENE NETWORK USING NETWORKANALYST	21
4.7 MUTATION ANALYSIS	21
4.8 PREDICTION OF microRNA TARGETS – TargetScanHuman	22
4.9 ENRICHMENT ANALYSIS OF miRNA AND lncRNA USING NcPath	22
4.10 STATISTICAL ENRICHMENT ANALYSIS	23
4.11 STRUCTURAL VARIANTS VISUALIZED FROM THE ENSEMBL	24
CHAPTER 5: RESULTS AND DISCUSSION	25
5.1 RESULTS	25
5.2 CYTOSCAPE	27
5.3 NETWORKANALYST	29
5.4 COSMIC MUTATION FREQUENCY	32
5.5 MUTATION FREQUENCY FROM cBioPortal	33

5.6 PROTEIN-CODING GENE EXPRESSION PROFILE USING LeukemiaDB	37
5.7 TargetScanHuman	40
5.8. ENRICHMENT ANALYSIS OF miRNA AND lncRNA USING NcPath	43
5.9 STATISTICAL ANALYSIS USING g:Profiler	50
5.10 STRUCTURAL VARIANTS VISUALISED FROM Ensembl	56
5.11 PATHWAY ANALYSIS USING KEGG	60
5.12 STATISCAL ANALYSIS USING GSCALite	61
 CHAPTER 6: SUMMARY AND CONCLUSIONS.....	 77
 REFERENCES	 81

LIST OF ABBREVIATIONS

AML	Acute Myeloid Leukemia
MLL	Mixed Lineage Leukemia
CNV	Copy Number Variation
SNP	Single Nucleotide Polymorphism
TF	Transcription Factor
miRNA	Micro RNA
lncRNA	Long Non-Coding RNA
PPI	Protein-Protein Interactions
NK-AML	Normal Karyotype Acute Myeloid Leukemia
TRRUST	Transcription Regulatory Relationship
COSMIC	Catalogue Of Somatic Mutations in Cancer
dbLGL	Leukemia Gene Literature Database
KEGG	Kyoto Encyclopedia Of Genes And Genomics
TCGA	The Cancer Genome Atlas Program
TARGET	Therapeutically Applicable Research To Generate Effective Treatments
GO	Gene Ontology
GSCALite	Gene Set Cancer Analysis
BLAST	Basic Local Alignment Search Tool
VEP	Variant Effect Predictor
STRING	Search Tool for The Retrieval Of Interacting Genes/Proteins
ORA	Over Representation Analysis
RTK	Receptor Tyrosine Kinase
SCF	Stem Cell Factor
GO:MF	Gene Ontology: Molecular Function
RUNX1	Runt-Related Transcription Factor 1
NUP98	Nucleoporin 98kDa
KMT2A	Lysine (K)-Specific Methyltransferase 2A
NSD1	Nuclear Receptor Binding Set Domain Protein 1
FLT3	FMS-Related Tyrosine Kinase 3

DNMT3A	DNA (Cytosine-5-)-Methyltransferase 3 Alpha
NRAS	Neuroblastoma Ras Viral Oncogene Homolog
KIT	Receptor Tyrosine Kinase Type III
GRN	Gene Regulatory Network

LIST OF TABLES

Table 1: cBioPortal Mutation Data

Table 2: DoCM Mutation Data - 1

Table 3: DoCM Mutation Data - 2

Table 4: Mutation Type T from COSMIC

Table 5: Translocation ETV6

Table 6: Translocation KMT2A

Table 7: Translocation NUP98

Table 8: dbLGL Genes

Table 9: Cosmic Somatic Genes - 1

Table 10: Cosmic Somatic Genes - 2

Table 11: Cosmic Germline Genes

Table 12: cBioPortal Genes

Table 13: Common Genes and Mutations - 1

Table 14: Common Genes and Mutations - 2

CHAPTER 1

INTRODUCTION

Developmental disease genomics is a field of study that aims to understand the genetic causes of developmental disorders and congenital diseases. This involves analysing the genetic changes that occur in an individual's DNA and how these changes contribute to the development of specific diseases. Developmental diseases can arise from genetic mutations, chromosomal abnormalities, or epigenetic changes.

Pediatric acute myeloid leukemia (AML) is a type of blood cancer that affects children. Developmental disease genomics includes analysis of the genetic changes that occur in the DNA of myeloid cells and how these changes contribute to the development of leukemia.

Some of the most studied genes in pediatric AML include FLT3, NPM1, CEBPA, IDH1/2, and DNMT3A. Understanding the genetic basis of pediatric AML can help in the development of new treatments and improve outcomes for children with this disease.

Recurrent genetic mutations occur in acute myeloid leukemia (AML) and have been incorporated into risk stratification to predict the prognoses of AML patients. The bone marrow microenvironment plays a critical role in the development and progression of AML.

Despite improved clinical outcome in recent years, the incidence rate is increasing (Howlader *et al.*, 2020)

Even though the cellular origin of the childhood disease remains unclear, several lines of evidence suggest that its origin is distinct from that of the adult counterpart (Bolouri *et al.*, 2018; Malard and Mohty, 2020).

Acute leukemias are characterized by uncontrolled proliferation of undifferentiated cells, called blasts, which impair normal hematopoiesis, in the bone marrow (BM) and peripheral blood (PB), with secondary infiltration of other tissues. According to morphology and cytochemistry, they are classified into acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL).

On the contrary, the incidence of childhood AML is highest in patients younger than 1 year (infants) (Howlader *et al.*, 2020).

The AML1-ETO translocation (RUNX1-RUNX1T1) derives from the rearrangement of chromosomes 8 and 21 and has been shown to be the most common rearrangement in both children and adults with AML, suggesting its appearance as a postnatal event. Nevertheless, the prenatal occurrence of the translocation is supported by its detection in blood spots collected at birth, which were still available at the time of AML diagnosis (Wiemels *et al.*, 2002).

1.1 BACKGROUND

Cytogenetic data are essential for risk stratification of childhood acute myeloid leukemia (AML); therefore, the impact of chromosomal aberrations is crucial.

Data from a large group of patients younger than 18 years treated according to the study AML-Berlin-Frankfurt-Münster (BFM) 98 (n = 454), including their cytogenetics, were analyzed (*Christine von Neuhoff, Dirk Reinhardt et al.*, (2010)). The prognostic value of rare recurrent chromosomal aberrations still has to be elucidated; therefore, these data will contribute to future risk stratification for treating pediatric AML.

A prenatal origin has been demonstrated for 2 more translocations, PML-RARA and CBFB-MYH11, associated with childhood AML, suggesting that a wide diversity of childhood leukemias originate prenatally (*Cliona M. McHale, Joseph L. Wiemels et al.*).

1.2 EXISTING PROBLEMS

Diagnosis and classification: Pediatric AML is often difficult to diagnose, and classifying different subtypes can be challenging.

Treatment resistance: Some pediatric AML patients may develop resistance to standard treatment regimens, making it difficult to achieve a cure.

Long-term effects of treatment: Children who undergo treatment for AML are at risk of developing long-term side effects, including damage to the heart, lungs, bones, and nervous system.

Relapse: Despite initial treatment success, some pediatric AML patients may experience a relapse of their disease, requiring additional treatment.

Late effects: Survivors of pediatric AML may experience late effects of treatment, such as secondary cancers, growth and development problems, and learning disabilities.

Limited options for high-risk patients: Children with high-risk AML may not respond well to standard treatments and may have limited options for effective therapy.

Lack of personalized medicine: There is a need for more personalized medicine approaches to treatment, as the genetic makeup of each patient's AML may play a role in treatment success.

1.3 TREATMENT

Treating pediatric acute myeloid leukemia (AML) typically involves a combination of

chemotherapy, stem cell transplantation, and targeted therapy. The specific treatment plan will depend on the individual patient and their subtype of AML.

Chemotherapy: This is the mainstay of treatment for pediatric AML and usually involves a combination of several drugs given in a specific order. The goal is to eliminate the leukemia cells and induce remission.

Stem cell transplantation: This procedure involves replacing the patient's bone marrow with healthy stem cells from a donor. This can be done after high-dose chemotherapy to help restore the bone marrow's ability to produce normal blood cells.

Targeted therapy: This type of treatment targets specific genetic mutations that drive the growth of leukemia cells. It is often used in conjunction with chemotherapy and may help improve outcomes for patients with certain subtypes of AML.

Clinical trials: Some pediatric AML patients may be eligible for participation in clinical trials of new and innovative treatments.

1.4 BOTTLENECK AREAS

Lack of knowledge about underlying genetic mutations: There is a lack of knowledge about the genetic mutations that drive AML in children, making it difficult to develop targeted treatments and identify patients at high risk of relapse.

Limited options for high-risk patients: Children with high-risk AML may not respond well to standard treatments and may have limited options for effective therapy.

Side effects: Treatment for pediatric AML can have significant side effects, including damage to normal cells, decreased production of normal blood cells, and increased risk of infection. These side effects can impact the long-term health and quality of life of survivors.

Late effects: Survivors of pediatric AML may experience late effects of treatment, such as secondary cancers, growth and development problems, and learning disabilities.

Access to specialized care: Not all children have access to specialized pediatric cancer centres, which can limit their access to the most advanced treatments and a multidisciplinary team of healthcare providers.

Cost of care: The cost of treatment for pediatric AML can be prohibitively high for some families, limiting access to care and impacting outcomes.

Lack of data on long-term outcomes: There is a lack of data on the long-term outcomes of children treated for AML, making it difficult to determine the best approach to care and to identify areas for improvement.

CHAPTER 2

REVIEW OF LITERATURE

The primary purpose of a literature review is to provide a comprehensive understanding of the existing knowledge on a particular topic or subject. It helps to identify gaps in the current knowledge, highlight areas for further research, and establish a theoretical framework for the study.

A literature review typically involves a thorough search of relevant literature and research databases, followed by a careful analysis of the findings. The review may be conducted as part of a larger research project, or as a standalone piece of work, such as a dissertation or thesis.

The following are the various references reviewed for the purpose of the current study.

- **The genomics of acute myeloid leukemia in children** - Shannon E. Conneely and Rachel E. Rau

Acute myeloid leukemia (AML) is a clinically, morphologically, and genetically heterogeneous disorder.

Core binding factor (CBF) AML is a cytogenetically defined subtype of AML characterized by the presence of either t(8;21)(q22;q22) or inv(16)(p13q22) structural chromosomal aberrations, hereafter referred to as t(8;21) and inv(16), respectively. CBF AML accounts for approximately 20–25% of pediatric AML cases and is classified as a favorable risk subtype.(Harrison CJ, Hills RK *et al.*,) & (Bolouri H, Farrar JE *et al.*,).

Inv(16) and t(8;21) AML are collectively termed CBF AML due to their similar effects on the CBF transcription factor complex as well as their similar outcome profiles. In both inv(16) and t(8;21), the chromosomal rearrangement leads to a fusion gene involving one of the components of the CBF complex and attendant fusion protein expression.(Downing

JR (2003)) & (Speck NA, & Gilliland DG (2002)).

The age distribution also differs, with inv(16) prevalence relatively constant from infancy to adulthood, t(8;21) is far more common in older children and adolescents (Creutzig U, Zimmermann M *et al.*).

Shannon E. Conneely and Rachel E. Rau, The genomics of acute myeloid leukemia in children, *Cancer Metastasis Rev.* 2020 Mar; 39(1): 189–209.

Harrison CJ, Hills RK, Moorman AV, Grimwade DJ, Hann I, Webb DK, Wheatley K, de Graaf SS, van den Berg E, Burnett AK, & Gibson BE (2010). Cytogenetics of childhood acute myeloid leukemia: United Kingdom Medical Research Council treatment trials AML 10 and 12. *Journal of Clinical Oncology*, 28(16), 2674–2681.

Bolouri H, Farrar JE, Triche T Jr, Ries RE, Lim EL, Alonzo TA, Ma Y, Moore R, Mungall AJ, Marra MA, Zhang J, Ma X, Liu Y, Liu Y, Auvil JMG, Davidsen TM, Gesuwan P, Hermida LC, Salhia B, Capone S, Ramsingh G, Zwaan CM, Noort S, Piccolo SR, Kolb EA, Gamis AS, Smith MA, Gerhard DS, & Meshinchi S (2018). The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nature Medicine*, 24(1), 103–112.

Creutzig U, Zimmermann M, Reinhardt D, Rasche M, von Neuhoff C, Alpermann T, Dworzak M, Perglerová K, Zemanova Z, Tchinda J, Bradtke J, Thiede C, & Haferlach C (2016). Changes in cytogenetics and molecular genetics in acute myeloid leukemia from childhood to adult age groups. *Cancer*, 122(24), 3821–3830. PMID: 27529519 DOI: 10.1002/cncr.30220.

James R Downing, The core-binding factor leukemias: lessons learned from murine models, 2003 Feb;13(1):48-54. DOI: 10.1016/s0959-437x(02)00018-7.

Speck NA, & Gilliland DG (2002). Core-binding factors in haematopoiesis and leukemia. *Nature Reviews. Cancer*, 2(7), 502–513. PMID: 12094236 DOI: 10.1038/nrc840

• **Prenatal Origin of Pediatric Leukemia: Lessons From Hematopoietic Development** - Anna Cazzola, Giovanni Cazzaniga, Andrea Biondi, Raffaella Meneveri, Silvia Brunelli and Emanuele Azzoni.

Leukemia is the most frequent childhood malignancy and it is characterized by a

heterogeneous manifestation (Steliarova-Foucher *et al.*, 2017). Despite improved clinical outcome in recent years, the incidence rate is increasing (Howlader *et al.*, 2020).

Even though the cellular origin of the childhood disease remains unclear, several lines of evidence suggest that its origin is distinct from that of the adult counterpart (Bolouri *et al.*, 2018; Malard and Mohty, 2020). Hematopoietic stem cells (HSCs) sustain the production of most blood cells in adults, the ontogeny of vertebrate hematopoiesis is characterized by the presence of HSC-independent hematopoietic cells that originate sequentially during embryo development and in some cases persist to adulthood (Ghosn *et al.*, 2019).

Acute leukemias are characterized by uncontrolled proliferation of undifferentiated cells, called blasts, which impair normal hematopoiesis, in the bone marrow (BM) and peripheral blood (PB), with secondary infiltration of other tissues. According to morphology and cytochemistry, they are classified into acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). Less frequently, they can show an intermediate phenotype, with features of both diseases, and are defined as mixed lineage leukemias. During childhood, lymphoid phenotypes are predominant over myeloid ones, with a highest age-specific incidence between 2 and 5 years. On the contrary, the incidence of childhood AML is highest in patients younger than 1 year (infants) (Howlader *et al.*, 2020).

Cytogenetic and molecular lesions diverge in infant, childhood, and adult acute leukemias in terms of type and recurrence. Chromosomal aberrations and especially translocations involving the gene KMT2A (MLL) are the most common genetic lesions in both AML and ALL infant leukemias but are less frequent in children and adults (Malard and Mohty, 2020; Rice and Roy, 2020).

Focal mutations in the N/KRAS, KIT, and CBL genes are more frequent in children than adults with AML, whereas IDH1, IDH2, RUNX1, NPM1, DNMT3A, and TP53 mutations are almost exclusively found in the latter category (Papaemmanuil *et al.*, 2016; Bolouri *et al.*, 2018).

Steliarova-Foucher, E., Colombet, M., Ries, L. A. G., Moreno, F., Dolya, A., Bray, F., *et al.*, (2017). International incidence of childhood cancer, 2001–10: a population-based registry study. *Lancet Oncol.* 18, 719–731. doi: 10.1016/S1470-2045(17)30186-9.

Howlader, N., Noone, A., Krapcho, M., Miller, D., Brest, A., Yu, M., *et al.*, (2020). SEER Cancer Statistics Review, 1975-2017. Bethesda, MD: National Cancer Institute.

Bolouri, H., Farrar, J. E., Triche, T., Ries, R. E., Lim, E. L., Alonzo, T. A., *et al.*, (2018). The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat. Med.* 24, 103–112. doi: 10.1038/nm.4439.

Malard, F., and Mohty, M. (2020). Acute lymphoblastic leukemia. *Lancet* 395, 1146–1162. doi: 10.1016/S0140-6736(19)33018-1.

Ghosn, E., Yoshimoto, M., Nakauchi, H., Weissman, I. L., and Herzenberg, L. A. (2019). Hematopoietic stem cell-independent hematopoiesis and the origins of innate-like B lymphocytes. *Development* 146: dev170571. doi: 10.1242/dev.170571.

Rice, S., and Roy, A. (2020). MLL-rearranged infant leukemia: a ‘thorn in the side’ of a remarkable success story. *Biochim. Biophys. Acta Gene Regul. Mech.* 1863:194564. doi: 10.1016/j.bbagr.2020.194564.

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V. I., Paschka, P., Roberts, N. D., *et al.*, (2016). Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* 374, 2209–2221. doi: 10.1056/NEJMoa1516192.

Anna Cazzola, Giovanni Cazzaniga, Andrea Biondi, Raffaella Meneveri, Silvia Brunelli and Emanuele Azzoni, Prenatal Origin of Pediatric Leukemia: Lessons From Hematopoietic Development, *Front. Cell Dev. Biol.*, 12 January 2021, Sec. Stem Cell Research, Volume 8 –2020, <https://doi.org/10.3389/fcell.2020.618164>.

- **Pediatric Acute Myeloid Leukemia—Past, Present, and Future** - Dirk Reinhardt, Evangelia Antoniou, and Katharina Waack.

The incidence of pediatric AML is about seven per million, with only minor differences between continents or countries. The malignant blasts originate from early hematopoietic progenitors as an evolution from (pre-)leukemia stem cells. External/environmental factors could explain only a tiny percentage. In addition, predisposing syndromes or germline mutations are associated with less than 10% of pediatric AML. During childhood and adolescence, infants less than two years old and

adolescents have the highest incidence. Whereas MLL-rearranged leukemia dominates during infancy, the frequency of core-binding leukemia (CBL) and AML associated with mutations, such as NPM 1 or FLT3-ITD, increases by age. (Elgarten C.W., Aplenc R.).

Elgarten C.W., Aplenc R. Pediatric acute myeloid leukemia: Updates on biology, risk stratification, and therapy. *Curr. Opin. Pediatr.* 2020;32:57–66. doi: 10.1097/MOP.0000000000000855.

Dirk Reinhardt, Evangelia Antoniou, and Katharina Waack, Pediatric Acute Myeloid Leukemia—Past, Present, and Future, *J Clin Med.* 2022 Feb; 11(3): 504. Published online 2022 Jan 19, doi: 10.3390/jcm11030504, PMCID: PMC8837075, PMID: 35159956.

- **The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions** -Hamid Bolouri, Jason E Farrar *et al.*,

AML is overtaking acute lymphoblastic leukemia (ALL) as the leading cause of childhood leukemic mortality, in part because current prognostic schemas classify many children who will ultimately succumb to their disease as being at low or intermediate risk. Additionally, aside from investigational tyrosine kinase inhibitors for FLT3-activated AML, targeted therapies are not used in pediatric AML. Both problems stem from an inadequate understanding of the biology of childhood AML.

AML is a molecularly heterogeneous group of diseases affecting individuals of all ages (Li S, *et al.*). Recent genome-scale studies have revealed new potentially targetable mutations prevalent in adult de novo AML (Timothy J Ley, Christopher Miller *et al.*), (Patel JP, *et al.*).

However, the relevance of these findings to childhood AML remains unclear, as several of the most common mutations in adults are far less prevalent in pediatric AML (Ho PA, *et al.*), (Farrar JE, *et al.*).

Like adult AML, pediatric AML has one of the lowest rates of mutation among cancers that are molecularly well characterized, with <1 somatic change in a protein-coding region per mega base in most cases.

Bolouri, H., Farrar, J. E., Triche, T. J., Ries, R. E., Lim, E. L., Alonzo, T. A.,

Meshinchi, S. (2018). The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nature Medicine*, 24(1), 103–112. <https://doi.org/10.1038/nm.4439>

Patel, J., Gonen, M., Figueroa, M. E., Kharfan-Dabaja, M. A., Sun, Z., Racevskis, J., Levine, R. L. (2012). Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia. *The New England Journal of Medicine*, 366(12), 1079–1089. <https://doi.org/10.1056/nejmoa1112304>

Ley, T. J., Miller, C. J., Ding, L., Raphael, B. J., Mungall, A. J., Robertson, G., Wilson, R. K. (2013). Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *The New England Journal of Medicine*, 368(22), 2059–2074. <https://doi.org/10.1056/nejmoa1301689>

- **Identification of a prognostic signature based on copy number variations (CNVs) and CNV-modulated gene expression in acute myeloid leukemia**

Prognostic studies have provided information concerning potential outcomes and survival rates that is crucial to understanding the disease and improving its treatment [Wee Y, Liu Y, Lu J, Li X, Zhao M 2018]. Copy number variation (CNV), a potential valuable prognostic marker, is a type of structural variation larger than 1 kilobase pair and involves unbalanced rearrangements that result in gains or losses in the DNA content [Zarrei M, MacDonald JR et al., 2015], (Shlien A, Malkin D. 2009) CNVs drive genome evolution and alter gene expression [Liang L, Fang JY et al., 2016] ,] and thereby influence the progression of diseases such as cancers [Lauer S, Gresham D. 2019].

Niu, C. (2021). Identification of a prognostic signature based on copy number variations (CNVs) and CNV-modulated gene expression in acute myeloid leukemia. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8748127/>

Wee Y, Liu Y, Lu J, Li X, Zhao M. Identification of novel prognosis-related genes associated with cancer using integrative network analysis. *Sci Rep*. 2018;8:3233.

Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet*. 2015;16:172–183.

Liang L, Fang JY, Xu J. Gastric cancer and gene copy number variation: emerging

cancer drivers for targeted therapy. *Oncogene*. 2016;35:1475–1482.

Shlien A, Malkin D. Copy number variations and cancer. *Genome Med*. 2009;1:62.
Lauer S, Gresham D. An evolving view of copy number variants. *Curr Genet*. 2019;65:1287–1295.

- **CEBPA copy number variations in normal karyotype acute myeloid leukemia: Possible role of breakpoint-associated microhomology and chromatin status in CEBPA mutagenesis.**

Copy number variations (CNV) in CEBPA locus represent heterogeneous group of mutations accompanying acute myeloid leukemia (AML).

Double mutant AMLs presented uniform biological features when compared to single CEBPA mutations and accompanied mostly younger patients.

The data suggest that double vs. single CEBPA mutants might present distinct pathways of leukemia transformation conferring different clinical features. Additionally, there are data showing that different mutant CEBPA subgroups might be preceded by distinct mechanisms of pathogenesis, and that various factors might affect frameshift vs. in-frame mutagenesis (*B.J. Wouters et al., (2009)*). However, little is known about interaction between these factors and generation of CNVs within CEBPA locus (Libura, M., Pawelczyk, M., Florek, I., Matiakowska, K., Jaźwiec, B., Borg, K., Haus, O. (2015)). CEBPA copy number variations in normal karyotype acute myeloid leukemia: Possible role of breakpoint-associated microhomology and chromatin status in CEBPA mutagenesis. *Blood Cells Molecules and Diseases*, 55(4), 284–292. <https://doi.org/10.1016/j.bcmd.2015.07.002>

Wouters, B. J., Löwenberg, B., Erpelinck-Verschueren, C. A., Van Putten, W. L., Valk, P. J. M., & Delwel, R. (2009). Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood*, 113(13), 3088–3091. <https://doi.org/10.1182/blood-2008-09-179895>

- **Cytogenetics of Pediatric Acute Myeloid Leukemia: A Review of the Current Knowledge.**

Pediatric acute myeloid leukemia is a rare and heterogeneous disease in relation to morphology, immunophenotyping, germline and somatic cytogenetic and genetic abnormalities. Over recent decades, outcomes have greatly improved, although survival rates remain around 70% and the relapse rate is high, at around 30%. Cytogenetics is an important factor for diagnosis and indication of prognosis. The main cytogenetic abnormalities are referenced in the current WHO classification of acute myeloid leukemia, where there is an indication for risk-adapted therapy.

Quessada, J., Cuccuini, W., Saultier, P., Loosveld, M., Harrison, C. J., & Lafage-Pochitaloff, M. (2021). Cytogenetics of Pediatric Acute Myeloid Leukemia: A Review of the Current Knowledge. *Genes*, 12(6), 924. <https://doi.org/10.3390/genes12060924>

- **Transcriptome analysis offers a comprehensive illustration of the genetic background of pediatric acute myeloid leukemia.**

Recent advances in the genetic understanding of acute myeloid leukemia (AML) have improved clinical outcomes in pediatric patients. However, ~40% of patients with pediatric AML relapse, resulting in a relatively low overall survival rate of ~70%. Using RNA-seq, we identified 54 in-frame gene fusions and 1 RUNX1 out-of-frame fusion in 53 of 139 patients. Moreover, we found at least 258 gene fusions in 369 patients (70%) through reverse transcription polymerase chain reaction and RNA-seq. Five gene rearrangements were newly identified, namely, NPM1-CCDC28A, TRIP12-NPM1, MLLT10-DNAJC1, TBL1XR1-RARB, and RUNX1-FNBP1. In addition, we found rare gene rearrangements, namely, MYB-GATA1, NPM1-MLF1, ETV6-NCOA2, ETV6-MECOM, ETV6-CTNNB1, RUNX1-PRDM16, RUNX1-CBFA2T2, and RUNX1-CBFA2T3. Among the remaining 111 patients, KMT2A-PTD, biallelic CEBPA, and NPM1 gene mutations were found in 11, 23, and 17 patients, respectively. These mutations were completely mutually exclusive with any gene fusions.

Shiba, N., Yoshida, K., Hara, Y., Yamato, G., Shiraishi, Y., Matsuo, H., Hayashi, Y. (2019). Transcriptome analysis offers a comprehensive illustration of the genetic

background of pediatric acute myeloid leukemia. *Blood Advances*, 3(20), 3157–3169.
<https://doi.org/10.1182/bloodadvances.2019000404>

- **Mutational Profiling of Pediatric Myeloid Leukemia Subtypes without Clinically Significant Chromosomal Aberrations.**

The discovery of novel significant molecular and genetic markers is important for the diagnostics, prognosis, and therapy selection in hematological malignancies. Distinct cytogenetic aberrations leading to the formation of fusion genes are found in more than 40% of pediatric cases of acute myeloid leukemia (AML); however, the tumor cells in approximately 20% of these patients display cytogenetically normal karyotype (NK-AML). In 34 pediatric cases of different AML types, the coding regions of 26 genes involved in the AML pathogenesis were analyzed by massive parallel sequencing. Sequencing revealed the somatic mutations in genes that are involved in various intracellular signaling pathways, including the CEBPA, ETV, IDH1, JAK2, and NRAS genes. In addition, rare genetic variants were found in CUX1, FLT3, TET2, PTPN11, and NUP98 genes.

Ghukasyan, L. G., Krasnov, G. S., Muravenko, O. V., Baidun, L. V., Ibragimova, S. Z., & Nasedkina, T. V. (2019). Mutation profile of subtypes of acute myeloid leukemia in children in the absence of known chromosomal rearrangements. *Molecular biologia*.
<https://doi.org/10.1134/s002689841903008x>

- **Genomic Profiling of Pediatric Acute Myeloid Leukemia Reveals a Changing Mutational Landscape from Disease Diagnosis to Relapse**

Recurrent somatic variants [in (greater than or equal to) 2 patients] were identified for 10 genes (FLT3, NRAS, PTPN11, WT1, TET2, DHX15, DHX30, KIT, ETV6, KRAS), with variable persistence at relapse.

Molecular profiling of AML in pediatric and adult studies have started to characterize AML as a disease with distinct age dependent alterations (*Schuback HL, Arceci RJ et al., (2013)*). Many of the novel and clinically impactful alterations identified in adult AML by whole genome sequencing (*Ley TJ, Ding L, Walter MJ, McLellan MD et al.,*

(2010), Mardis ER, Ding L, Dooling DJ et al., (2009)) are not present in childhood AML,(Ley TJ, Ding L, Walter MJ, McLellan MD et al., (2010), Ho PA, Kutny MA, Alonzo TA et al., (2011)) suggesting a distinct age associated biology and highlighting the need for comprehensive genomic profiling of this disease in children. Patients with AML having a high mutant to normal allelic ratio FLT3/ITD had adverse outcomes, while those with AML having a lower FLT3/ITD allelic ratio showed outcomes similar to patients without FLT3 mutations (Meshinchi S, Alonzo TA et al., (2006)).

CHAPTER 3

AIM AND SCOPE OF THE STUDY

3.1 AIM

The main goal of this project is to identify new genetic and epigenetic changes involved in the development of pediatric Acute Myeloid Leukemia (AML). AML is a heterogeneous disease with various genetic and epigenetic alterations that drive disease initiation and progression. Therefore, identifying new genetic and epigenetic changes can improve our understanding of the mechanisms of the disease and help develop more effective targeted therapies.

Further, the aim is to investigate the role of developmental pathways in the pathogenesis of pediatric AML and to identify potential therapeutic targets. AML pathogenesis involves various genetic and epigenetic alterations that affect cellular pathways involved in normal hematopoietic development. Understanding the role of developmental pathways in the pathogenesis of AML can provide valuable insights into the mechanisms underlying disease initiation and progression.

In addition, the study also aims to understand the interaction between the leukemia cells and the microenvironment in pediatric AML, including the interactions between the leukemia cells and the bone marrow, immune cells, and stromal cells. Leukemia cells do not exist in isolation and instead interact with various cells within their microenvironment. These interactions play a crucial role in the initiation, progression, and resistance of leukemia cells to therapy. Therefore, understanding the mechanisms of these interactions is essential for developing targeted therapies that can disrupt the microenvironment and improve treatment efficacy.

In addition, the present study aims to develop new models for studying pediatric AML that more accurately reflect the complexity of the disease and its underlying biology, identify new predictive markers for response to therapy, and develop new prognostic models for pediatric AML. Developing new markers for predicting treatment response can help identify patients likely to benefit from a particular therapy, thereby avoiding unnecessary treatment and minimizing the risk of side effects.

3.2 SCOPE

The scope of this project is to investigate gene clusters related to pediatric Acute Myeloid Leukemia by analyzing various gene databases using bioinformatics tools and techniques. Our primary objective is identifying and analyzing these gene clusters to better understand the disease. Additionally, we will conduct a comprehensive literature review and data analysis to achieve our goals. To achieve these objectives, we will employ bioinformatics tools and techniques and conduct an extensive literature review to analyze the data.

CHAPTER 4

METHODOLOGY

4.1 GENE DATA SETS

Gene data sets were downloaded from multiple databases, and the vast data was filtered according to our parameters using Microsoft Excel.

4.2 DATABASES

LEUKEMIA GENE LITERATURE DATABASE - <http://soft.bioinfo-minzhao.org/lgl/>

We obtained a total of 1805 human leukemia-associated genes. Of these, 1722 genes were derived from adult patients, while only 312 were from childhood leukemia patients. Only 67 human genes were associated with AML (61), CML (4) and CLL (2) in the childhood leukemia cases. ([Yining Liu, et al.](#)).

The cancer types were browsed within the database, and the required Cancer sub-type (Childhood Acute Myeloid Leukemia) was selected. A total of 61 genes were associated with the subtype, and they were analysed for further analysis.

A wide range of analyses can be done within the database, such as the Expression of Leukemia genes, their Interactions, their regulation, and mutations can be analysed.

COSMIC - CATALOGUE OF SOMATIC MUTATIONS IN CANCER DATABASE

(<https://cancer.sanger.ac.uk/cosmic>)

The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website have been developed to store somatic mutation data in a single location and display the data and other information related to human cancer. Currently, the database holds information on 66 634 samples and reports 10 647 mutations. Through the web pages, these data can be queried, displayed as figures or tables, and exported in several formats. COSMIC is

an ongoing project to curate somatic mutation data and release it through the website. (Bamford, S., Dawson, E., Forbes, S. et al.,)

COSMIC held vast data, which gave a greater insight into the progress of the project. The dataset included very informative data regarding the genes, mutations, mutation types, translocations, gene fusions, chromosome band, genome location, molecular genetics, disease type, cell line, role in cancer etc. This data has been widely used in the project for various analyses.

cBioPortal- A Web Platform for Gene-Based Exploration (<https://www.cbioportal.org/>)

This is a database online tool which does analysis based on the data present in the database to the personalised input data. *The cBio Cancer Genomics Portal (<http://cbioportal.org>) is an open-access resource for the interactive exploration of multidimensional cancer genomics data sets, currently providing access to data from more than 5,000 tumour samples from 20 cancer studies. The cBio Cancer Genomics Portal significantly lowers the barriers between complex genomic data and cancer researchers who want rapid, intuitive, and high-quality access to molecular profiles and clinical attributes from large-scale cancer genomics projects and empowers researchers to translate these rich data sets into biologic insights and clinical applications. Cancer Discov; 2(5); 401–4. ©2012 AACR. (Cerami et al., 2012 & Gao et al., 2013)*

It gave insights into the mutation frequency, the structural variation, CN segments etc.

4.3 FILTERING THE GENES

A total of 529 genes were extracted from the databases.

dbLGL- 61 genes

COSMIC 74 genes

cBioPortal 394 genes, out of which only 68 cancer-causing genes were filtered using Excel and taken for analysis.

Using Microsoft Excel, 31 common genes were filtered from 528 genes and were carried out for further analysis.

4.4 GENE NETWORK ANALYSIS

The network analysis was based only on the genes taken from dbLGL (Leukemia Gene Literature db).

STRINGdb, Cytoscape and NetworkAnalyst tools were considered for Network analysis of the genes. Their Networks and interactions have been visualized to better understand the disease development.

STRINGdb: *All protein–protein association evidence in STRING is assessed and quantified. The result of this benchmarking/calibration is a STRING ‘confidence score’ for each association; confidence scores are scaled between zero and one. Confidence scores are first computed separately per evidence type and then integrated into a final, ‘combined’ confidence score. (Damian Szklarczyk et al.,).*

The 60 genes from dbLGL (Leukemia gene literature database) were analysed for Protein network interactions using STRINGdb. The interactions and the genes involved in the development of the disease and their interactions were visualized.

The STRING network was downloaded in a .TSV file and was imported into Cytoscape for Network visualization and identification of the hubs present in the network using Cytohubba.

4.5 IDENTIFICATION OF HUBS PRESENT IN THE NETWORK

Cytohubba is an in-built tool in Cytoscape used to identify hubs present in the network. *Cytoscape plugin cytoHubba for ranking nodes in a network by their network features. CytoHubba provides 11 topological analysis methods, including Degree, Edge Percolated Component, Maximum Neighborhood Component, Density of Maximum Neighborhood Component, Maximal Clique Centrality and six centralities (Bottleneck, EcCentricity, Closeness, Radiality, Betweenness, and Stress) based on shortest paths. (Chia-Hao Chin et al.,).*

4.6 VISUALISING GENERIC-PPI NETWORK, GENE-miRNA INTERACTIONS, TF-miRNA CO-REGULATORY GENE NETWORK USING NETWORKANALYST

NetworkAnalyst – a comprehensive gene expression profiling via network visual analytics.

Sixty-one genes from dbLGL were given as input in NetworkAnalyst to visualise generic-PPI network, Gene-miRNA interactions, and TF-miRNA co-regulatory gene network. The Networks were visualized by looking into the interactions, their nodes, their edges, and the degree of centrality scores were analysed for better understanding.

NetworkAnalyst holds few external tools within the database, which allows us to look into more factors like the regulation expression of miRNA from TRUST and the Functional explorer of Disease Pathway using KEGG were analysed for a better understanding.

4.7 MUTATION ANALYSIS

The common genes from the databases are considered for mutation analysis. The cBio Cancer Genomics Portal is an open-access resource for the interactive exploration of multidimensional cancer genomics data sets. (Ethan Cerami et al.).

The official gene symbols were given as input in cBioPortal (<https://www.cbioportal.org/>). A study related to our work is selected. In our case, we took up the Pediatric Acute Myeloid Leukemia data from the TARGET study, which acts as a background for the queries on the website.

After the “Query by Gene” option is selected, since we have a list of genes, we can also explore the study without giving any input of genes. The genes are analysed with the selected study, and of the options mentioned for analysis and a better understanding of the study, Mutation frequency, Copy number segments, Structural Variation and associated pathway analysis were done.

4.8 PREDICTION OF microRNA TARGETS – TargetScanHuman

The miRNAs belong to a class of small endogenous RNAs that regulate gene expression post-transcriptionally and play a role in gene silencing and translation inhibition by binding to target genes. The miRNAs represent a highly conserved class of tissue-specific genes found in all eukaryotic cells preserved across species since their discovery in 1993. (Vishnoi and Rani, 2017; Lu and Rothenberg, 2018).

TargetScan predicts biological targets of miRNAs by searching for the presence of conserved 8mer, 7mer, and 6mer sites that match the seed region of each miRNA (et al., Sean E McGeary, Kathy S Lin et al.).

The genes based on the mutation frequency identified through cBioPortal have been used to predict miRNA targets. The genes were searched against the database, and the miRNA targets and their binding sites were located using this tool.

4.9 ENRICHMENT ANALYSIS OF miRNA AND lncRNA USING NcPath (<http://ncpath.pianlab.cn/#/Home>)

Non-coding RNAs play important roles in transcriptional processes and participate in the regulation of various biological functions, in particular, miRNAs and lncRNAs. A novel pathway database named NcPath by integrating and visualizing a total of 178308 humans experimentally validated miRNA–target interactions (MTIs), 32282 experimentally verified lncRNA–target interactions (LTIs) and 4837 experimentally validated human ceRNA networks across 222 KEGG pathways (including 27 sub-categories).

Platform that provides annotation and visualization of the pathways on which non-coding RNAs (miRNA and lncRNA) are involved and provide support to multimodal non-coding RNAs enrichment analysis. (Zutan Li, Yuan Zhang et al.)

Since it is associated with KEGG, searching for the disease name gives you a table of miRNAs and lncRNAs involved in the disease pathway. Can also give input on the list of

the genes which will identify the miRNAs and lncRNAs linked to the gene.

Based on the mutation frequency, the miRNA and lncRNAs of NRAS, FLT3 and KIT genes were considered for analysis.

LeukemiaDB: A COMPREHENSIVE HUMAN LEUKEMIA TRANSCRIPTOME DATABASE

LeukemiaDB provides valuable insights into oncogenesis and progression of leukemia (Mei Luo, Yaru Miao et al.,).

LeukemiaDB was used to analyse the protein-coding expression and associated lncRNA expression. The mutated genes were searched to analyse the expression analysis of protein-gene coding and lncRNA expression.

4.10 STATISTICAL ENRICHMENT ANALYSIS

g: Profiler was used to profile the statistical enrichment analysis.

g: Profiler has a simple, user-friendly web interface with powerful visualisation for capturing Gene Ontology (GO), pathway, or transcription factor binding site enrichments down to individual gene levels. (Jüri Reimand, Meelis Kull et al.,).

The common genes from the three datasets were given as input to analyse the statistical enrichment analysis related to the genes.

The statistical data were reviewed to understand the enrichment and their functioning using g:GOST in-built tool for statistical enrichment analysis.

g:SNPense was used to identify the genes related to the SNPs taken from SNPedia, and the expression level was analysed using LeukemiaDB. (Michael Cariaso & Greg Lennon (2011)).

GSCALite - GENE SET CANCER ANALYSIS

A web server for gene set cancer analysis. GSCALite is a user-friendly web server for dynamic analysis and visualization of gene set in cancer and drug sensitivity correlation. To analyze a set of genes in cancers with the following functional modules. (i) Differential expression in tumor versus normal, and the survival analysis; (ii) Genomic variations and their survival analysis; (iii) Gene expression associated cancer pathway activity; (iv) miRNA regulatory network for genes (Chun-Jie Liu, Fei-Fei Hu, et al.).

The common list of genes is given as an input, and an analysis of CNVs and SNPs has been done.

4.11 STRUCTURAL VARIANTS VISUALIZED FROM THE ENSEMBL

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function, and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species. (Fiona Cunningham, James E. Allen et al.).

The Ensembl Genome browser (<https://www.ensembl.org/index.html?redirect=no>) was visited to visualize structural variants present in the genes. The mutated genes were searched, and the structural variation in the genes was analysed.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 RESULTS

5.1.1 STRING - Protein Network

Sixty-one genes from dbLGL were given as input in the spin database to visualise and analyse the protein network. The network was visualised and analysed for the genes/proteins involved in pediatric acute myeloid leukemia and the genes involved in the pathway of AML.

The network analysis:

number of nodes: 57

number of edges: 335

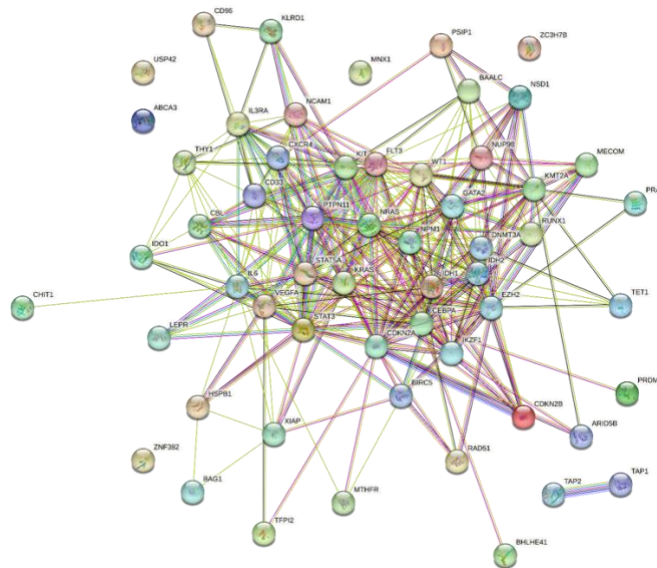
average node degree: 11.8

avg. local clustering coefficient: 0.653

expected number of edges: 77

PPI enrichment p-value: $< 1.0e-16$

PPI enrichment p-value: $< 1.0e-16$ indicates that proteins have more interactions among themselves than expected for a random set of proteins of the same size and degree distribution drawn from the genome. Such enrichment indicates that the proteins are partially biologically connected as a group. It is used to understand the underlying molecular mechanisms of the disease.

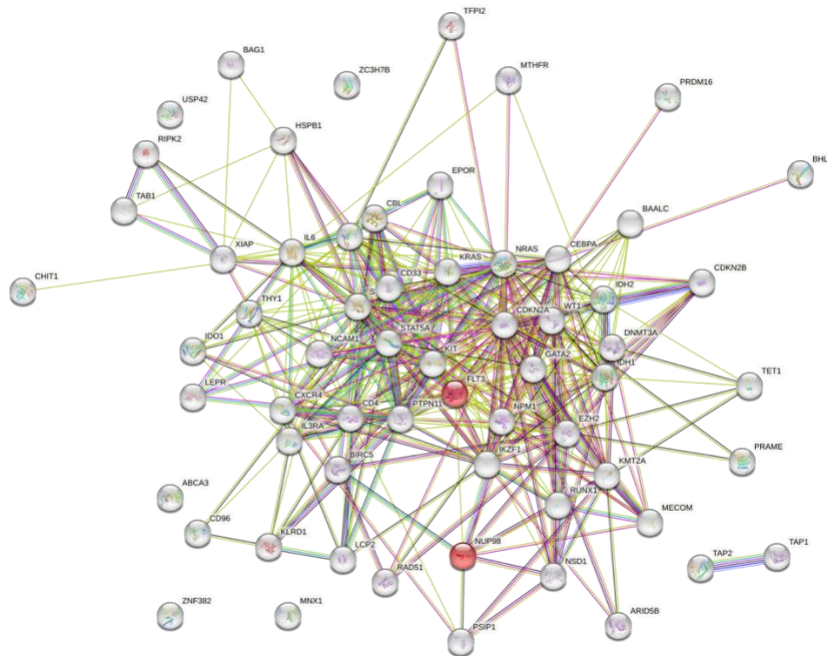


5.1.2 Genes Involved In Pediatric Acute Myeloid Leukemia

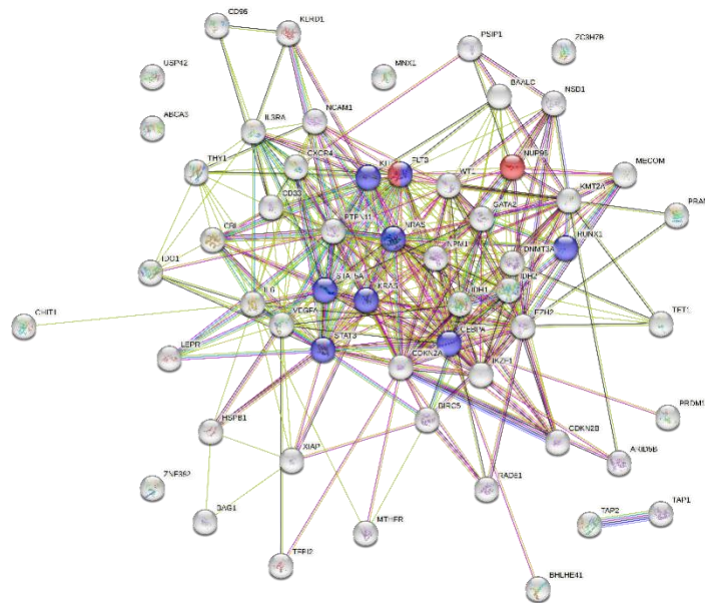
Disease-gene associations (DISEASES)				
disease	description	count in network	strength	false discovery rate
DOID:8632	Kaposi sarcoma	2 of 2	2.5	0.0034
DOID:6367	Acral lentiginous melanoma	2 of 2	2.5	0.0034
DOID:14731	Weaver syndrome	2 of 2	2.5	0.0034
DOID:0070323	Childhood acute myeloid leukemia	2 of 2	2.5	0.0034
DOID:0060221	Maffucci syndrome	2 of 2	2.5	0.0034

(more ...)

5.1.3 Genes Involved In Disease-Gene Association of Pediatric AML



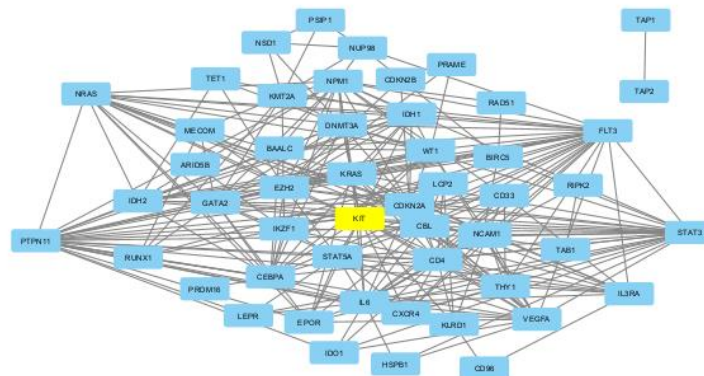
5.1.4 Network Showing The Highlighted Genes Involved In The Pathway Of AML



5.2 CYTOSCAPE

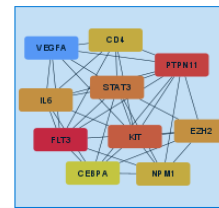
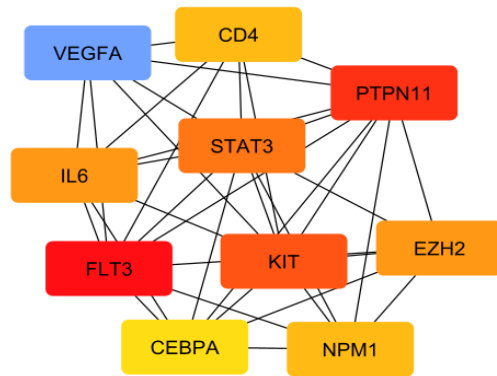
The network was loaded, and basic functions of interactions and node and edge editing, and the layout of the network were modified. Visualised network from the spin database was exported to Cytoscape for visualisation and identification of the hubs present in the network using Cytohubba, a plug-in present in Cytoscape.

The resulting hubs network was taken based on the degree of connection to a node.



5.2.1 Cytohubba Network - Identification Of Hub Genes

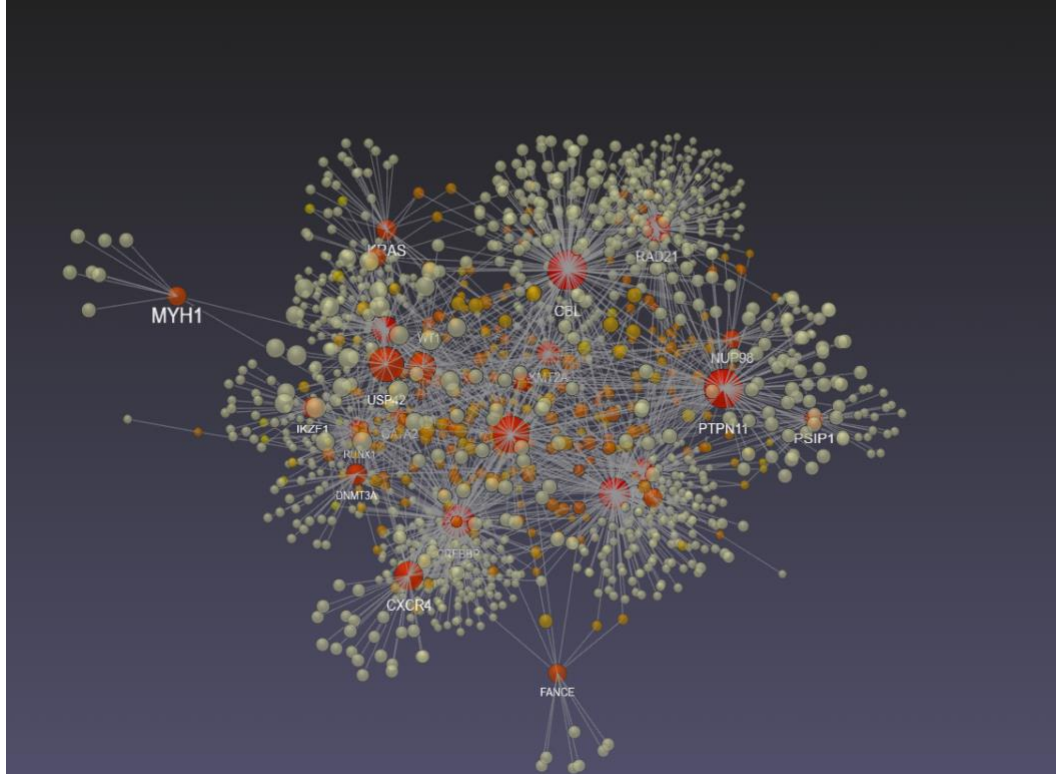
Genes having more interactions with other genes are defined as hub genes. Genes with greater degree score have more interactions with other genes.



Rank	Name	Score
1	FLT3	28
2	PTPN11	26
3	KIT	25
4	STAT3	23
5	EZH2	21
6	IL6	21
7	CD4	19
8	NPM1	19
9	VEGFA	18
10	CEBPA	18

Table 1 - Top 10 Gene Hubs in Network Sheet1 Ranked by Degree Method

5.3 NETWORKANALYST



3D model of generic - PPI network

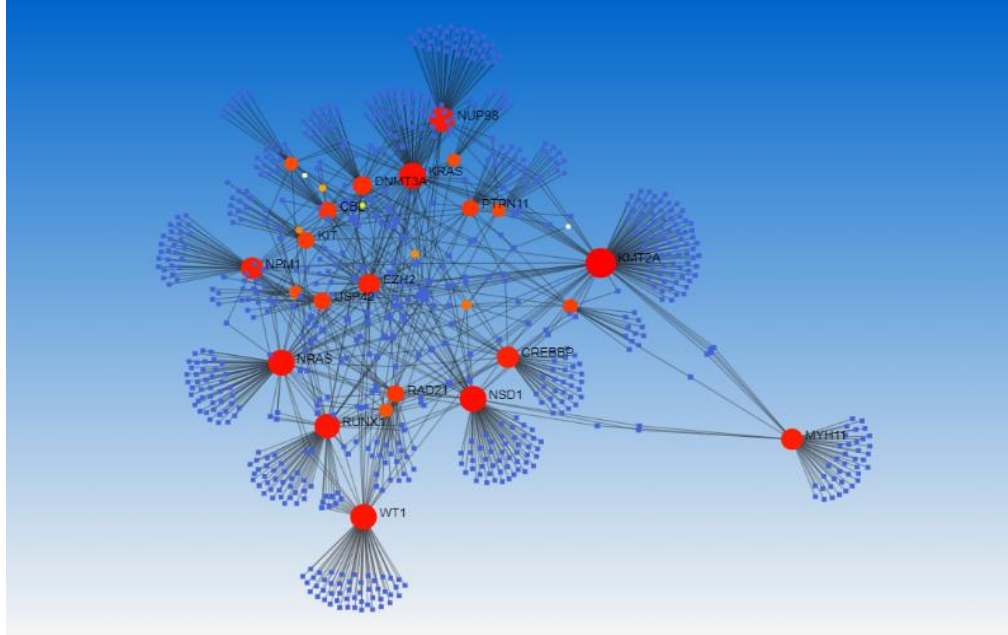
Node Table				
	Search	Modify	Delete	
☐	ID	Name	Degree	Expr.
☐	1387	CREBBP	294	
☐	5885	RAD21	216	
☐	867	CBL	216	
☐	4869	NPM1	203	
☐	2146	EZH2	149	
☐	7150	TOP1	129	

The alignment based on the degree score.

RESULT:

Sixty-one genes from dbLGL were given as input in Network Analyst for visualisation of generic-PPI network, gene-mRNA network, and gene-TF co-expression regulation. Their alignment has been analysed based on their core.

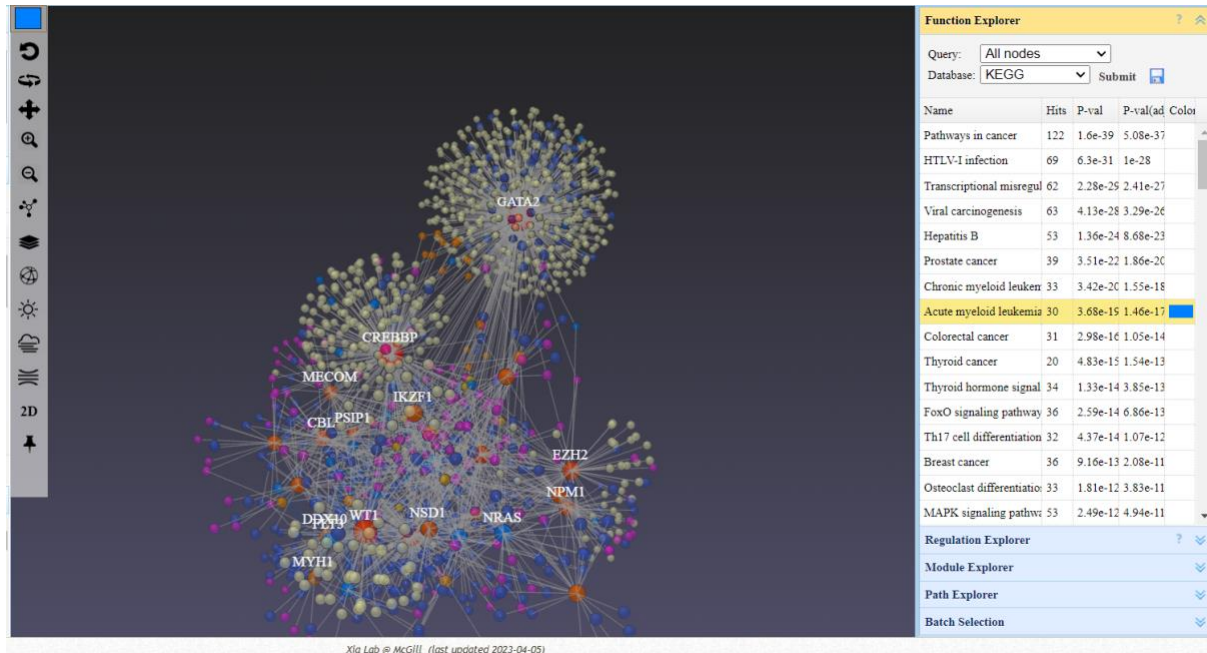
5.3.1 Gene-miRNA Network



Regulation Explorer				
Query: All nodes				
Database: TF[TRRUST]				
Submit				
Add Nodes				
☐	Name	ID	Hits	Color
☐	MYC	4609	4	
☐	SP1	6667	4	
☐	EP300	2033	3	
☐	CTCF	10664	2	
☐	ETS1	2113	2	
☐	GATA1	2623	2	
☐	HDAC3	4790	2	
☐	KLF5	5970	2	
☐	NFKB1	6774	2	
☐	RELA	688	2	
☐	STAT3	7528	2	
☐	YY1	8841	2	
☐	AR	10014	1	

Regulation of the Transcription Factors was analysed using TRRUST

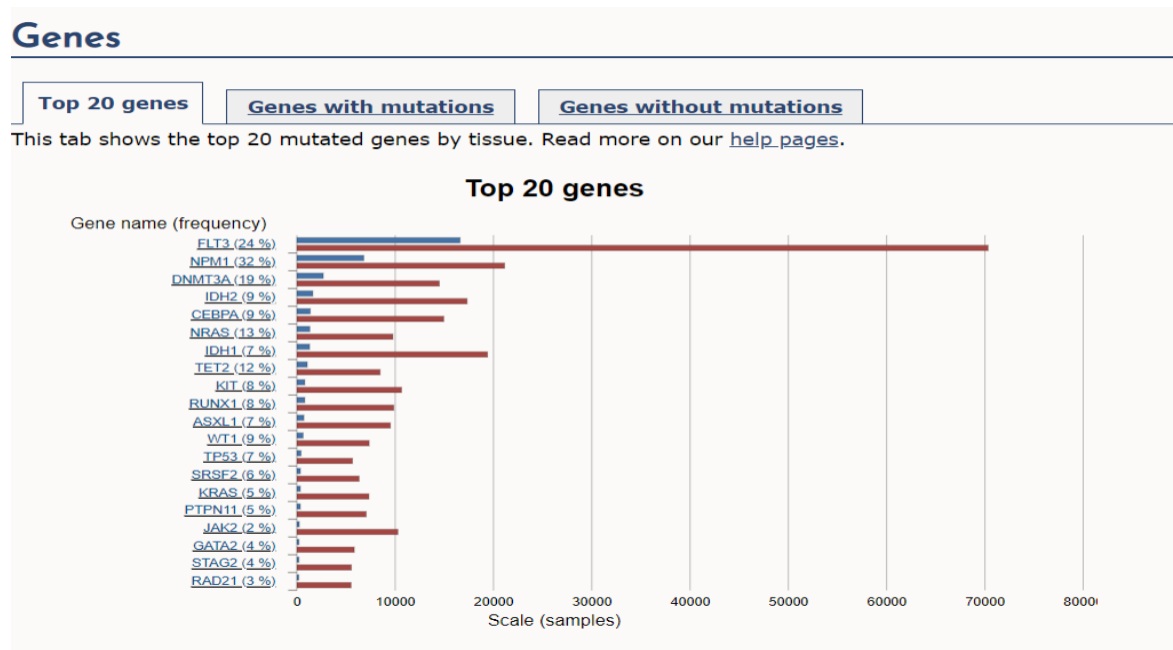
5.3.2 AML Genes Involved in TF Co-expression



MYC	KIT
STAT5B	KRAS
STAT5A	JUP
STAT3	CSF1R
SPI1	FLT3
RUNX1	RUNX1T1
RELA	CSF2
NFKB1	CHUK
CEBPA	AKT1
IKBKG	MAPK1
LEF1	GRB2
PML	CCND1
RARA	
TCF7	
TCF7L1	
TCF7L2	
ZBTB16	
NRAS	

These are the genes involved.

5.4 COSMIC MUTATION FREQUENCY

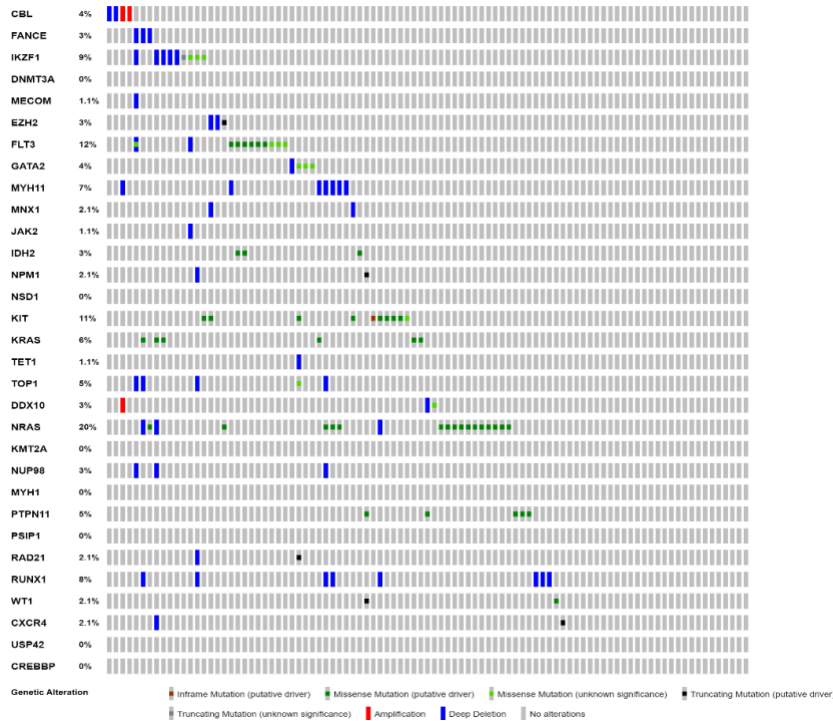


As the graph depicts, the genes FLT3 and NPM1 have more frequency of mutations.

RESULT:

Genes from the COSMIC database were analysed for mutation frequency within the COSMIC database. The frequency is based on the number of samples with their observed mutation to the sample from the database. FLT3 & NPM1 genes have greater mutation frequency compared to others.

5.5 MUTATION FREQUENCY FROM cBioPortal

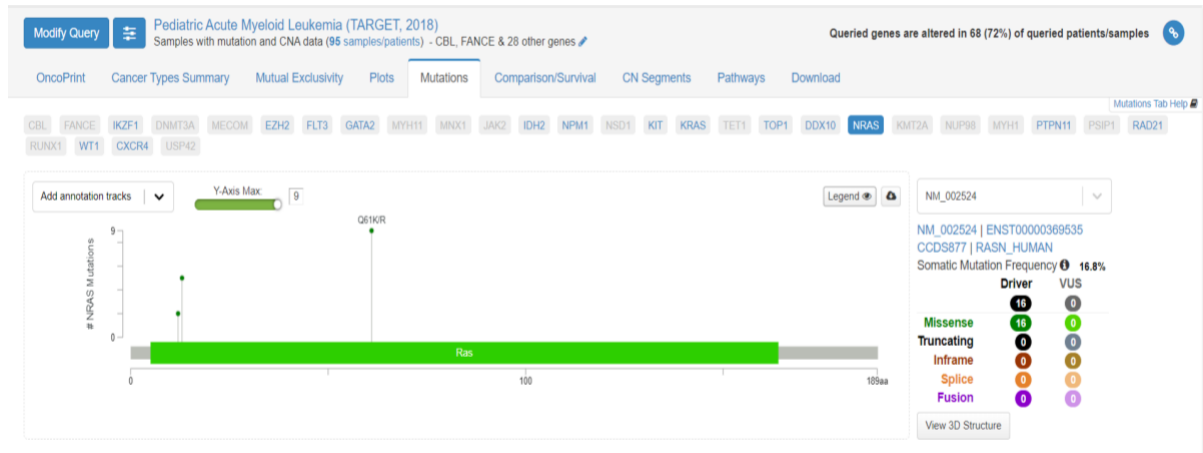


RESULT:

The common genes (31) from the three databases were given as input in the cBioPortal web page relative study has been selected for analysis of the input genes. The mutation frequencies were given based on the chosen relative study (Pediatric acute myeloid leukemia (TARGET)).

The three genes NRAS-20%, FLT3-12%, and KIT-11% have greater mutation frequencies.

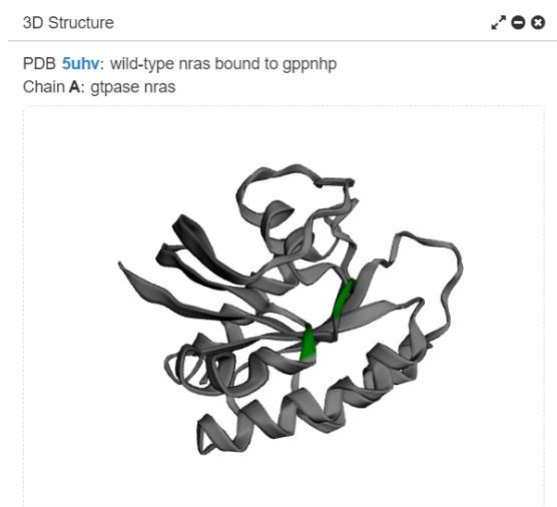
5.5.1 Mutations of NRAS



RESULT:

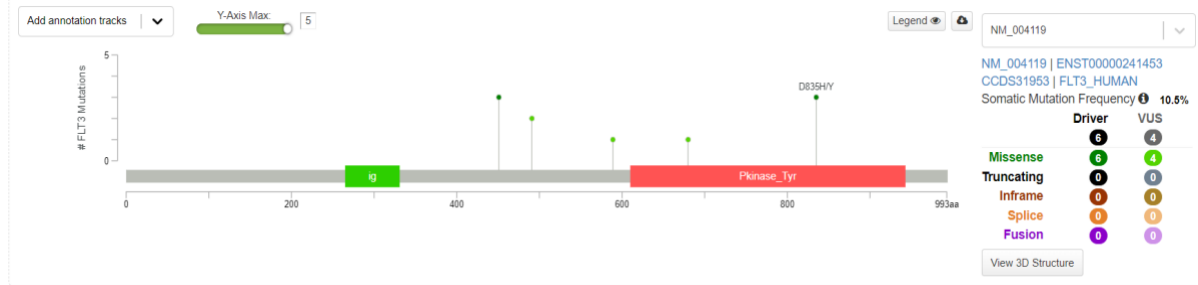
NRAS gene has 16 missense (driver) mutations and stands out with 20% mutation frequency compared with the other genes.

5.5.2 Structural Variation in NRAS



Green colour indicating missense mutations.

5.5.3 Mutation of FLT3



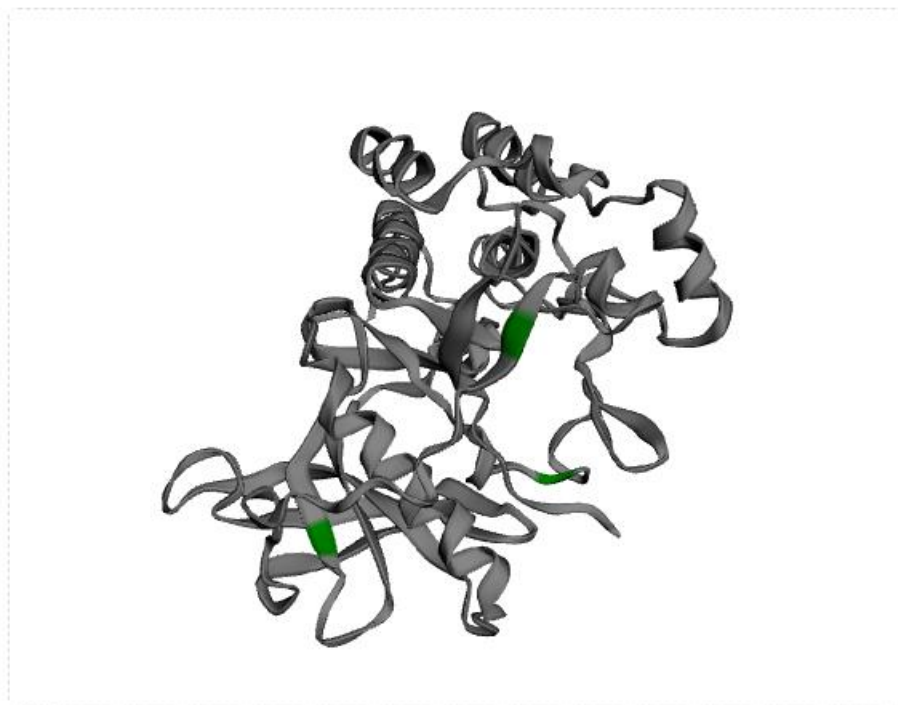
5.5.4 Structural variation in FLT3

3D Structure



PDB [6il3](#): crystal structure of the flt3 kinase bound to a small [...]

Chain **A**: receptor-type tyrosine-protein kinase flt3

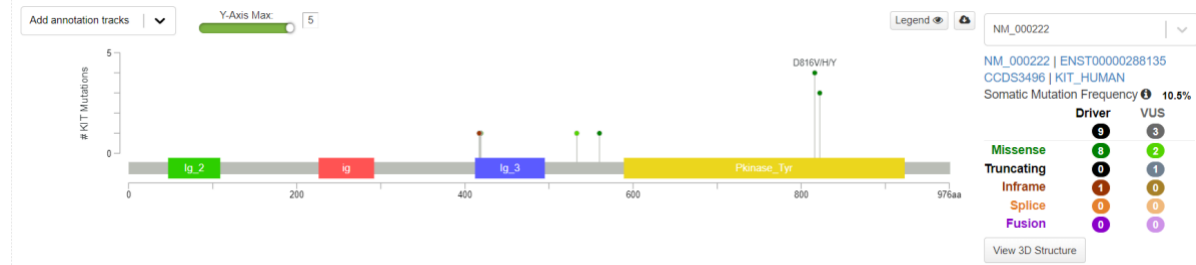


RESULT:

FLT3 has six missense (driver) mutations and four variant uncertain significant mutations that may be involved in developing the disease.

FLT3 12% mutation frequency

5.5.5 Mutation in KIT

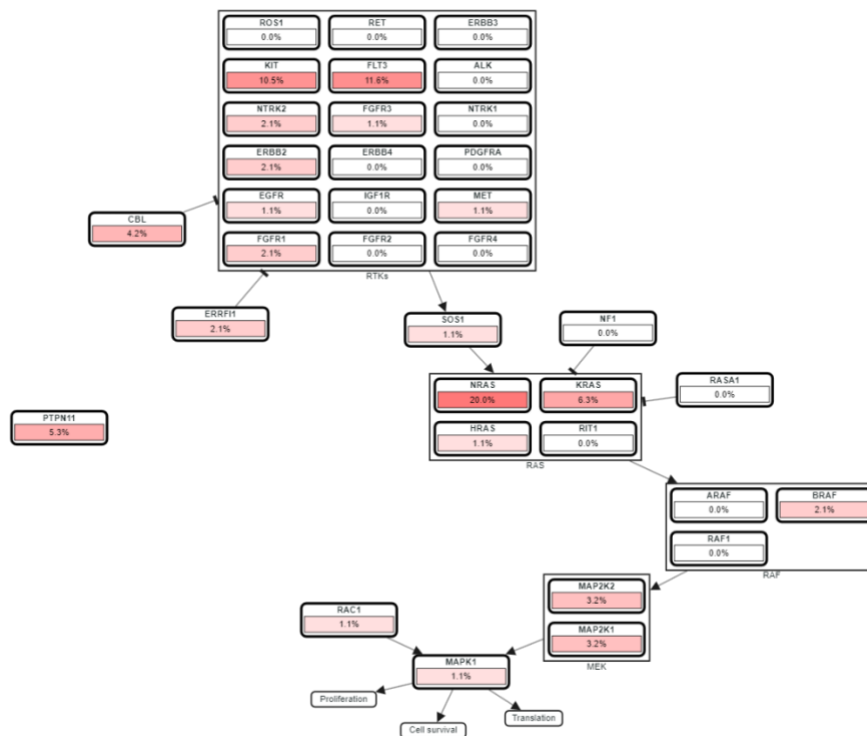


RESULT:

KIT gene has eight missense, one in-frame deletion (driver) mutation, and three variant uncertain significant mutations.

KIT 11% mutation frequency.

5.5.6 Associated Pathway: RTK-RAS



The associated pathway is based on the mutation frequency percentage.

5.5.7 Pathway Score

RESULT:

Pathway name	Score ▾	Genes matched
☒ RTK-RAS	35.00	CBL FLT3 KIT KRAS ...
☐ NOTCH	1.00	CREBBP
☐ MYC	0.00	
☐ Cell Cycle	0.00	
☐ TGF-Beta	0.00	
☐ TP53	0.00	
☐ HIPPO	0.00	
☐ PI3K	0.00	
☐ WNT	0.00	
☐ NRF2	0.00	

RTK-RAS has the highest score according to the input.

5.6 PROTEIN-CODING GENE EXPRESSION PROFILE USING LeukemiaDB

LeukemiaDB -- A comprehensive human leukemia transcriptome database has been visited for analysis of protein-coding genes. Based on the mutation frequency from cBioPortal, NRAS, FLT3, and KIT were taken for further analysis. (<http://bioinfo.life.hust.edu.cn/LeukemiaDB/#!/>)

5.6.1 NRAS

Protein-coding gene expression profile in Acute myeloid leukemia|AML

All protein-coding genes

Specially expressed protein-coding genes

<<-

<

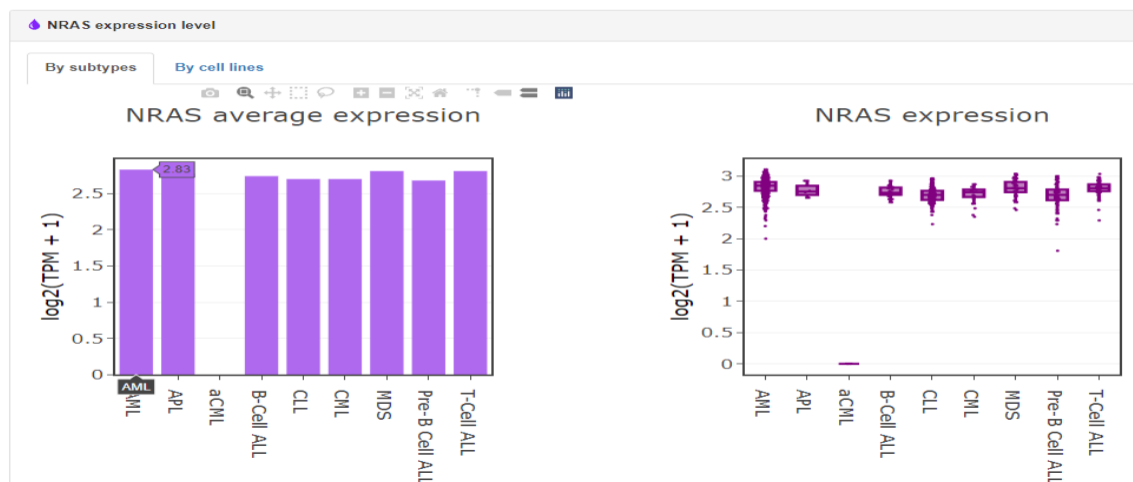
>

->>

Current page: 1; Records per page: 15

Gene name	Ensembl id	Chr	Start	End	Strand	Mean expression	Detail
NRAS	-	-	-	-	-	-	-
NRAS	ENSG00000213281	1	114704469	114716894	-	6.1	Display

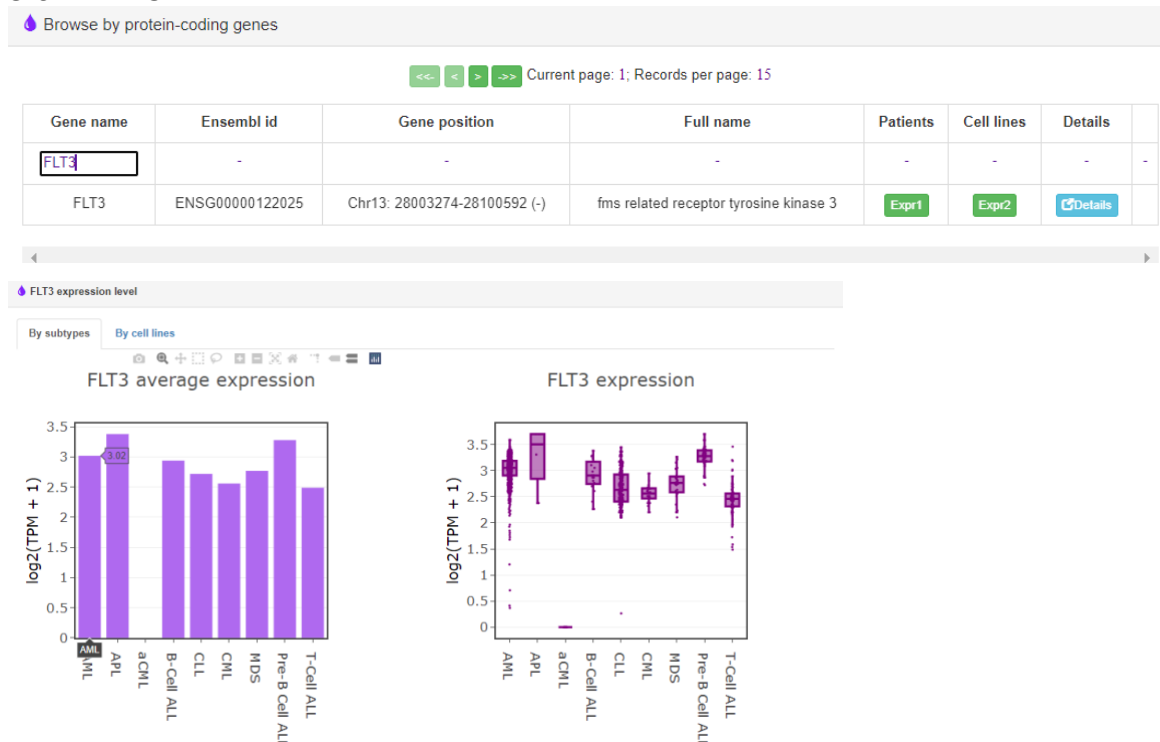
NRAS gene has a mean expression of 6.1 in protein coding for acute myeloid leukemia.



RESULT:

NRAS gene average expression in protein-coding for AML has a log score of 2.83, which is the highest expression among other subtypes.

5.6.2 FLT3



RESULT:

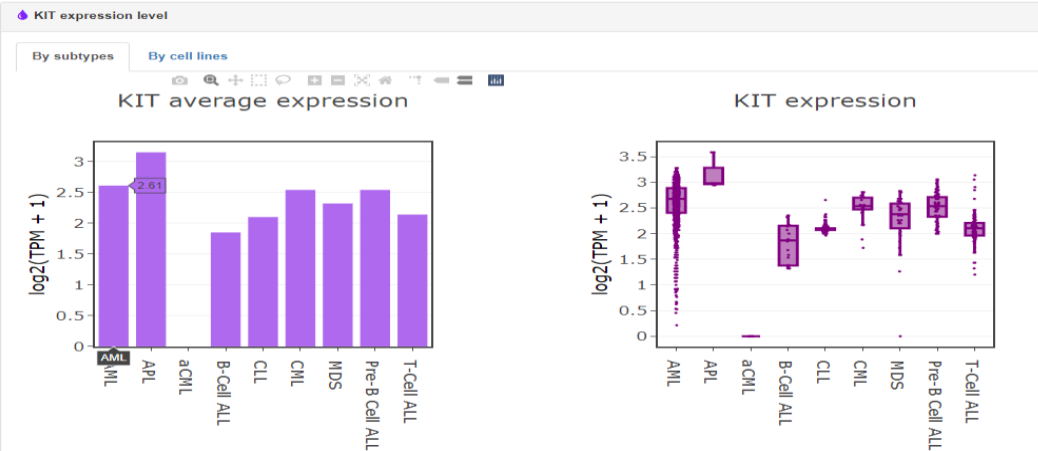
FLT3 gene average expression in protein-coding for AML has a log score of 3.02. Maximum expression 3.584963.

5.6.3 KIT

🔍 Browse by protein-coding genes

<< < > >> Current page: 1; Records per page: 15

Gene name	Ensembl id	Gene position	Full name	Patients	Cell lines	Details	
KIT	-	-	-	-	-	-	-
KIT	ENSG00000157404	Chr4: 54657918-54740715 (+)	KIT proto-oncogene, receptor tyrosine kinase	Expr1	Expr2	Details	



RESULT:

KIT gene average expression in protein-coding for AML has a log score of 2.61. Maximum expression 3.277985.

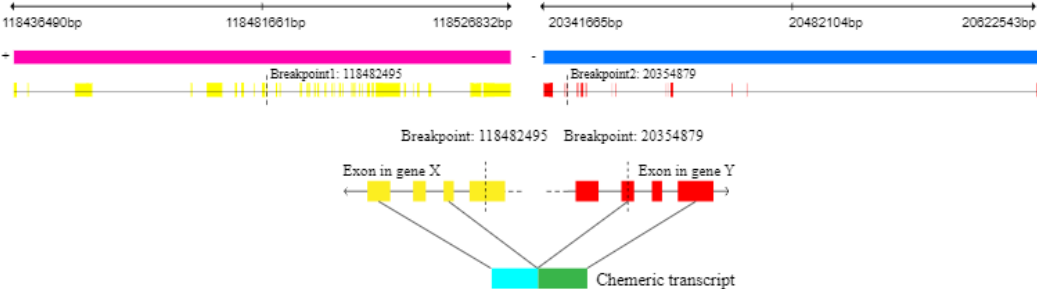
5.6.4 Gene Fusion of KMT2A with MLLT3 Involved in Development of AML

Fusion Name: KMT2A_MLLT3

INTER-CHROMOSOMAL Fusion Transcript

GeneX: KMT2A
Length: 90343
Chromosome: 11

GeneY: MLLT3
Length: 280879
Chromosome: 9



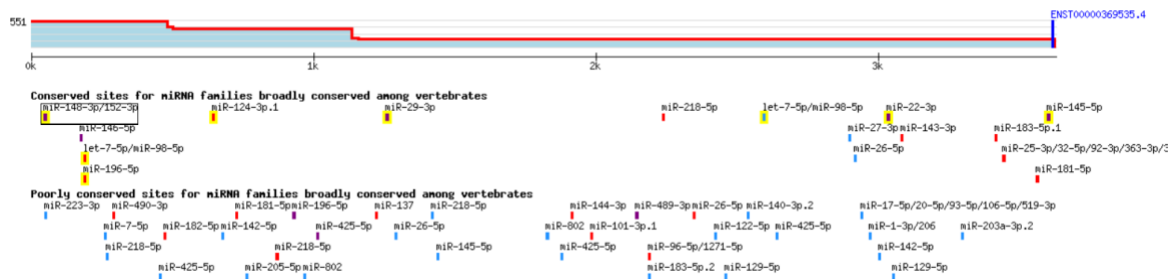
5.7 TargetScanHuman

TargetScanHuman prediction of human v 8.0(https://www.targetscan.org/vert_80/)

Based on the mutation frequency, NRAS, FLT3, and KIT were analysed for micro-RNA targets.

5.7.1 NRAS-Target Scan Human

Human NRAS ENST00000369535.4 3' UTR length: 3630



The predicted miRNA target sites in the 3' UTR of the NRAS gene.

Each entry shows the miRNA name, its seed sequence, its conservation score, and the predicted efficacy of the targeting. The seed sequence is the region of the miRNA that binds to the target site in the 3' UTR of the gene. The conservation score indicates how well-conserved the target site is across different species, with higher scores indicating greater conservation. The predicted efficacy score reflects how strongly the miRNA is predicted to regulate the expression of the gene.

The graph is used to visualise the predicted miRNA target sites in the NRAS gene and to compare the relative strengths of the targeting by different miRNAs.

The graph shows that miRNAs, such as hsa-miR-584-3p, hsa-miR-106b-3p, and hsa-miR-93-3p, are predicted to have relatively strong targeting efficacy. In contrast, others, such as hsa-miR-548h-5p and hsa-miR-376c-3p, are predicted to have weaker targeting effectiveness.

From the graph, we can see that miRNAs, such as hsa-miR-584-3p, hsa-miR-106b-3p,

and hsa-miR-93-3p, are predicted to have relatively strong targeting efficacy, while others, such as hsa-miR-548h-5p and hsa-miR-376c-3p, are predicted to have weaker targeting efficacy.

Conserved

	Predicted consequential pairing of target region (top) and miRNA (bottom)	Site type	Context++ score	Context++ score percentile	Weighted context++ score	Conserved branch length	P _{CT}	Predicted relative K _D
Position 45-52 of NRAS 3' UTR	5' ...GAGCCACUUUCAAGC-UGCACUGA...	8mer	-0.50	99	-0.50	5.673	0.88	-5.640
hsa-miR-152-3p	3' GGUUCAAGACAGUACGUGACU							
Position 45-52 of NRAS 3' UTR	5' ...GAGCCACUUUCAAGC-UGCACUGA...	8mer	-0.50	99	-0.50	5.673	0.88	-5.558
hsa-miR-148b-3p	3' UGUUUCAAGACACUACGUGACU							
Position 45-52 of NRAS 3' UTR	5' ...GAGCCACUUUCAAGC-UGCACUGA...	8mer	-0.50	99	-0.50	5.673	0.88	-5.089
hsa-miR-148a-3p	3' UGUUUCAAGACAUACGUGACU							

Conserved target sites are more likely to be under selective pressure to maintain their function, while non-conserved target sites may be less functional or have arisen more recently in evolution.

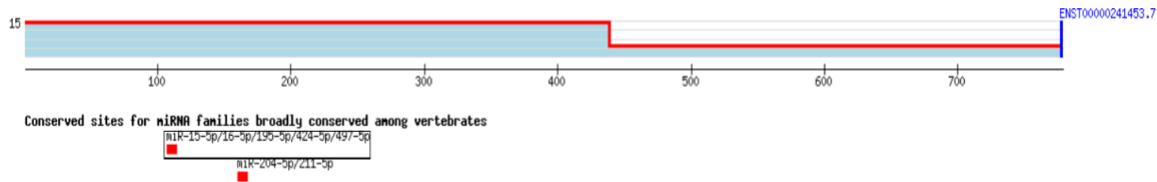
The conserved region table provides important information about the evolutionary conservation of the predicted miRNA target sites in the NRAS gene. The "Context ++ score" predicts the regulatory impact of microRNA (miRNA) binding to the 3' untranslated region (UTR) of the gene.

The predictive K_D value is given in units of nM (nanomolar) and is intended to provide a quantitative estimate of the strength of the miRNA-target interaction.

The P_{CT} measures the density of miRNA target sites in the 3' UTR of the gene and can indicate the potential regulatory impact of miRNAs on the expression of the gene. A higher P_{CT} indicates a higher density of predicted miRNA target sites in the 3' UTR of the gene and suggests that the gene may be more susceptible to miRNA-mediated regulation.

5.7.2. FLT3-TargetScanHuman

Human FLT3 ENST00000241453.7 3' UTR length: 778



Each red bar represents a predicted microRNA target site. The position of the red bar indicates the location of the predicted target site relative to the start of the 3' UTR, and the height of the bar represents the context ++ score for the target site. Higher bars indicate higher context ++ scores, which suggest a greater likelihood that the predicted target site is functional.

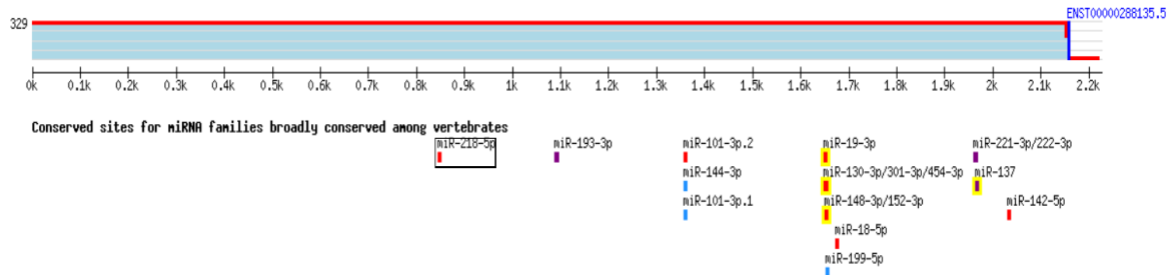
This transcript is associated with AML and is predicted to be targeted by several microRNAs, including hsa-miR-21-5p, hsa-miR-10b-5p, and hsa-miR-125a-5p, among others. Dysregulation of these microRNAs could potentially contribute to the development or progression of AML by affecting the expression of ENST00000241453.7 or other genes involved in AML pathogenesis.

Conserved

	Predicted consequential pairing of target region (top) and miRNA (bottom)	Site type	Context++ score	Context++ score percentile	Weighted context++ score	Conserved branch length	P _{CT}	Predicted relative K _D
Position 107-113 of FLT3 3' UTR	5' ...AAAUCAAUUAUCAAACUGCUGCUU... 3' GCGGUUAUAAUAGC--ACGACGAU	7mer-m8	-0.38	98	-0.38	4.737	0.69	-3.963
Position 107-113 of FLT3 3' UTR	5' ...AAAUCAAUUAUCAAACUGCUGCUU... 3' CCGGUUAUAAAGAC--ACGACGAU	7mer-m8	-0.38	98	-0.38	4.737	0.69	-3.935
Position 107-113 of FLT3 3' UTR	5' ...AAAUCAAUUAUCAAACUGCUGCUU... 3' GUGUUUGGUAAUAC--ACGACGAU	7mer-m8	-0.35	97	-0.35	4.737	0.69	-3.935
Position 107-113 of FLT3 3' UTR	5' ...AAAUCAAUUAUCAAACUGCUGCUU... 3' ACAUUUGGUACUACACGACGAU	7mer-m8	-0.33	97	-0.33	4.737	0.69	-3.935
Position 107-113 of FLT3 3' UTR	5' ...AAAUCAAUUAUCAAACUGCUGCUU... 3' UGUUUUGGUGUCACACGACGAC	7mer-m8	-0.31	96	-0.31	4.737	0.69	-4.366
Position 107-113 of FLT3 3' UTR	5' ...AAAUCAAUUAUCAAACUGCUGCUU... 3' AAGUUUUUUAUUAACGACGAC	7mer-m8	-0.32	96	-0.32	4.737	0.69	-4.897
Position 107-113 of FLT3 3' UTR	5' ...AAAUCAAUUAUCAAACUGCUGCUU... 3' UCCUCAGAACGGUGACGACGAA	7mer-m8	-0.32	96	-0.32	4.737	0.69	-5.060

5.7.3 KIT-TargetScanHuman

Human KIT ENST00000288135.5 3' UTR length: 2226



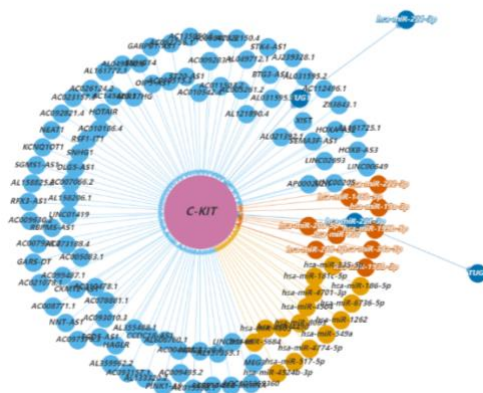
Multiple microRNAs, including hsa-miR-218-5p, have predicted target sites within its 3' UTR. These target sites may be important regulators of expression, and dysregulation of these microRNAs could potentially contribute to the development or progression of AML.

Conserved

	Predicted consequential pairing of target region (top) and miRNA (bottom)	Site type	Context++ score	Context++ score percentile	Weighted context++ score	Conserved branch length	P _{CT}	Predicted relative K _D
Position 844-850 of KIT 3' UTR	5' ...AACAAUUAACCAAAAGCACAG... 3' UGUACCAAUUCUAGUUCGUGUU	7mer-m8	-0.20	89	-0.20	2.995	0.39	-2.790
hsa-miR-218-5p								

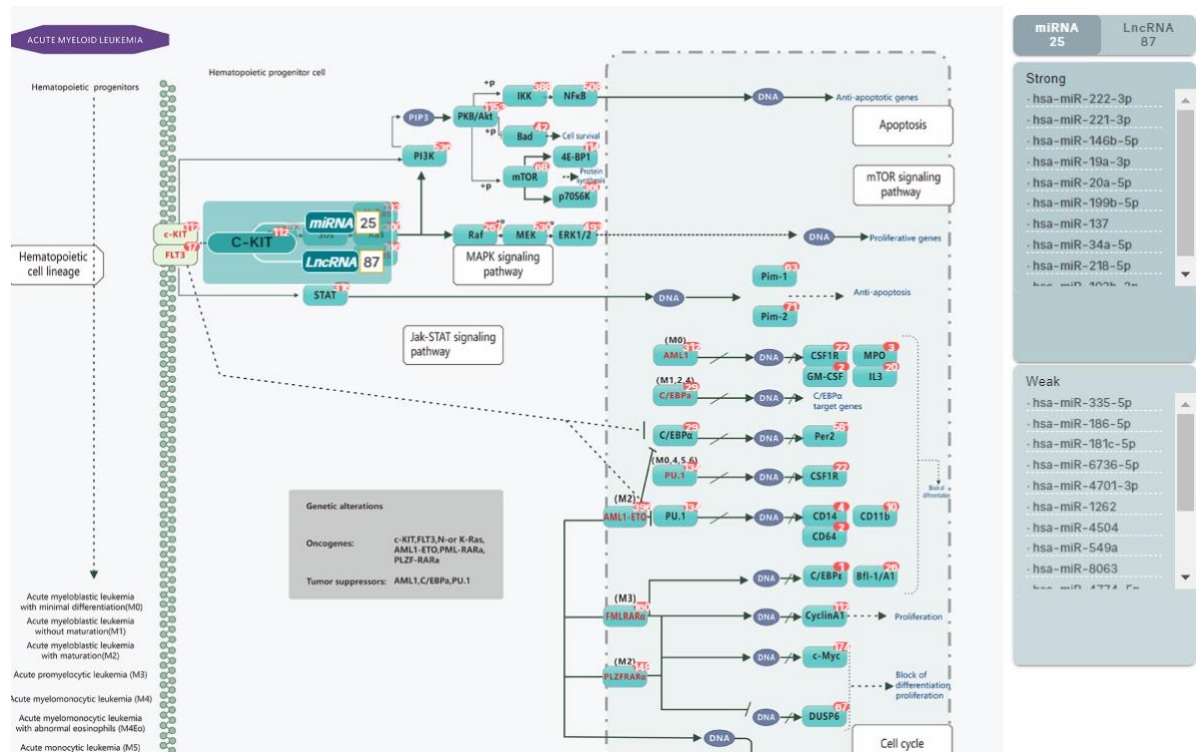
5.8. ENRICHMENT ANALYSIS OF miRNA AND lncRNA USING NcPath

5.8.1 C-KIT



Network miRNAs and lncRNAs linked to the KIT gene

The involvement of the C-KIT gene in the pathway of Acute Myeloid Leukemia with its miRNAs and lncRNAs



C-KIT (also known as CD117) is a gene that encodes a protein called a receptor tyrosine kinase, which is involved in the regulation of cell growth, differentiation, and survival. Mutations in the C-KIT gene have been associated with the development of various types of cancer, including acute myeloid leukemia (AML).

The NCPATH webpage provides information on the involvement of C-KIT in the AML pathway. C-KIT is activated by stem cell factor (SCF) and plays a role in the proliferation and survival of leukemic cells in AML. MiRNAs are small non-coding RNAs that regulate gene expression by binding to mRNA and causing its degradation or repression. The NCPATH webpage on the AML pathway shows that several miRNAs are involved in AML. These miRNAs play a role in the regulation of cell proliferation, differentiation, and apoptosis, and their dysregulation has been implicated in the development of AML.

LncRNAs are long, non-coding RNAs involved in various biological processes, including gene expression regulation, chromatin remodelling, and epigenetic modification. These lncRNAs are involved in the regulation of cell proliferation, differentiation, and apoptosis, and their dysregulation has been implicated in the development of AML.

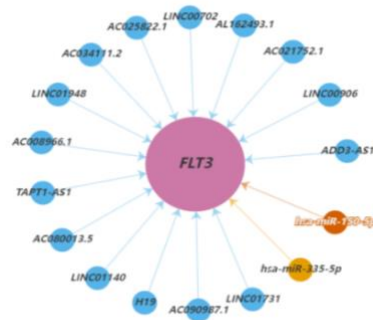
5.8.2 FLT3

FLT3 (Fms-like tyrosine kinase 3) is a gene that encodes a protein called a receptor tyrosine kinase, which is involved in the regulation of cell growth, differentiation, and survival. Mutations in the FLT3 gene have been associated with the development of various types of leukemia, including acute myeloid leukemia (AML).

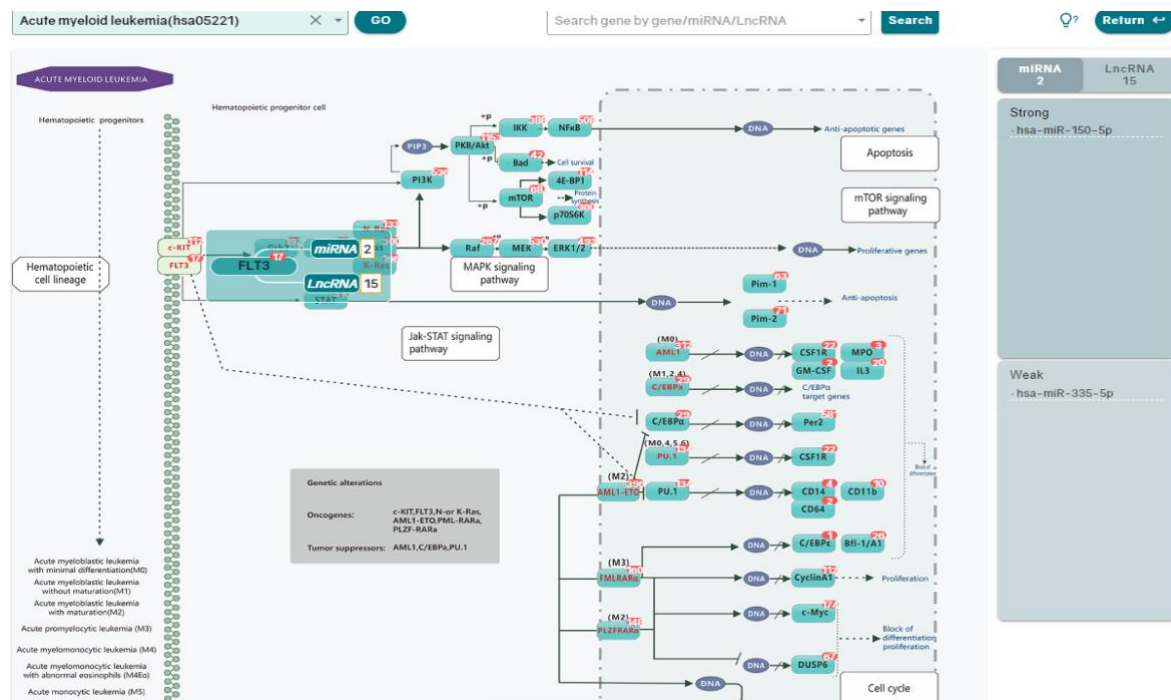
The NCPATH webpage provides information on the involvement of FLT3 in the AML pathway. FLT3 is frequently mutated in AML, and its activation leads to the activation of downstream signaling pathways, such as the PI3K-Akt and RAS-MAPK pathways. The activation of FLT3 promotes the proliferation and survival of leukemic cells in AML.

The NCPATH webpage on the AML pathway shows miRNAs involved in AML, which have been shown to target FLT3 expression. Dysregulation of these miRNAs has been implicated in the development of AML, and targeting them could be a potential therapeutic strategy for AML.

The NCPATH webpage on the AML pathway provides information on the involvement of lncRNAs in AML. These lncRNAs are involved in the regulation of cell proliferation, differentiation, and apoptosis, and their dysregulation has been implicated in the development of AML.



Network of miRNA & lncRNA linked with gene FLT3



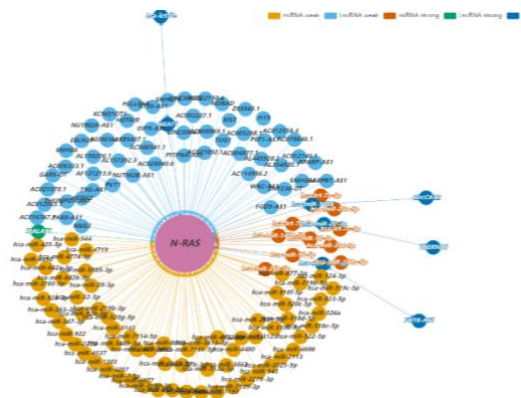
5.8.3 NRAS

N-RAS is a gene that encodes a protein called a small GTPase, which is involved in the regulation of cell growth, differentiation, and survival. Mutations in the N-RAS gene have been associated with the development of various types of cancer, including acute myeloid leukemia (AML).

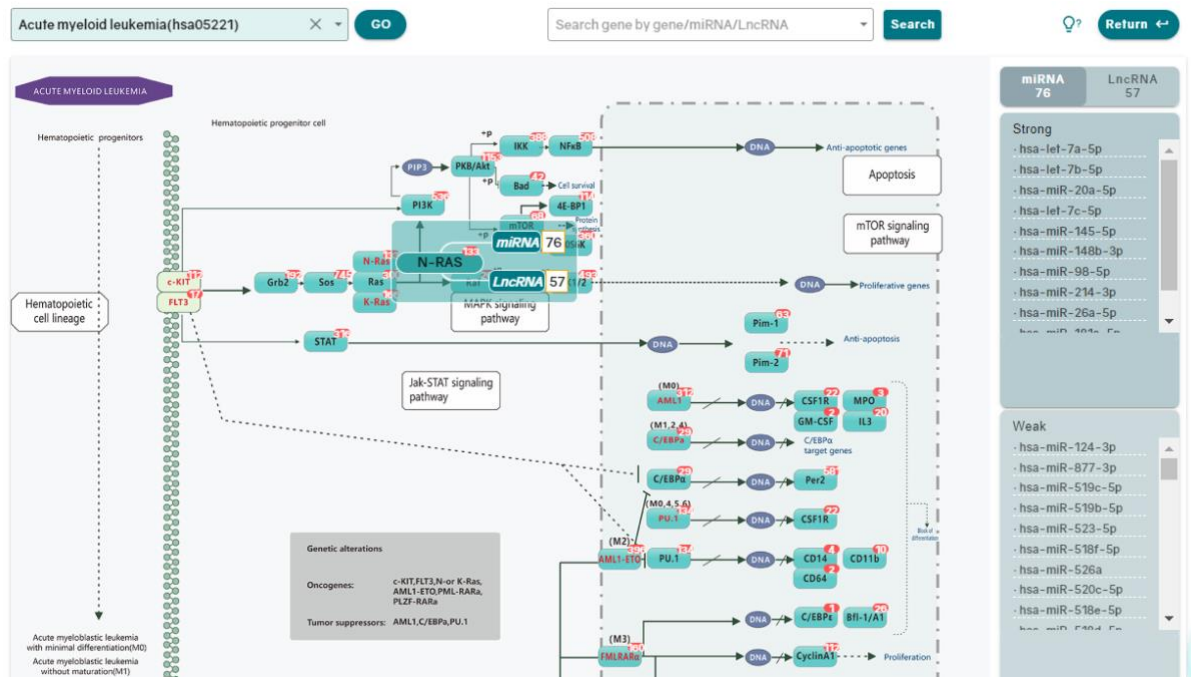
The NCPATH webpage provides information on the involvement of N-RAS in the AML pathway. According to the webpage, mutations in the N-RAS gene are found in

approximately 15-20% of AML cases. Activation of mutations in N-RAS leads to the activation of downstream signaling pathways, such as the RAS-MAPK and PI3K-Akt pathways, which promote the proliferation and survival of leukemic cells in AML.

The involvement of miRNAs and lncRNAs in the AML pathway suggests that they may play a role in the regulation of key genes and signalling pathways that are important for the development and progression of pediatric AML.



Network of miRNA & lncRNA linked with NRAS gene.



KMT2A-MLLT3 fusion gene visualised using the Leukemia database.

Acute myeloid leukemia(hsa05221) × GO						
MiRNA-Target gene						
miRNA	Target Gene	Target Gene (Entrez ID)	References (PMID)	DataSource	Support Type	Experiments
hsa-miR-122-5p	AKT3	10000	19296470	miRTarBase	Strong	Luciferase reporter assay//qRT-PCR
hsa-miR-122-5p	AKT3	10000	24244539	miRTarBase	Strong	Immunoblot//Luciferase reporter assay//qRT-PCR
hsa-miR-16-5p	AKT3	10000	23233752	miRTarBase	Strong	Luciferase reporter assay
hsa-miR-16-5p	AKT3	10000	23592263	miRTarBase	Weak	PAR-CLIP
hsa-miR-16-5p	AKT3	10000	21572407	miRTarBase	Weak	PAR-CLIP
hsa-miR-15a-5p	AKT3	10000	23233752	miRTarBase	Strong	Luciferase reporter assay
hsa-miR-15a-5p	AKT3	10000	23592263	miRTarBase	Weak	PAR-CLIP
hsa-miR-15a-5p	AKT3	10000	21572407	miRTarBase	Weak	PAR-CLIP
hsa-miR-519d-3p	AKT3	10000	22262409	miRTarBase	Strong	Luciferase reporter assay//Western blot
hsa-miR-29c-3p	AKT3	10000	20371350	miRTarBase	Weak	Sequencing
Rows per page: 10 1-10 of 3501 Page: 1 < >						

The miRNA's associated with AML

LncRNA-Target gene						
LncRNA	Target Gene	Target Gene (Entrez ID)	References	DataSource	Support Type	Supporting evidence
NEAT1	AKT1	207	25155612	High-throughput data	Strong	CHART-seq
AC009654.1	AKT1	207	RD00118961	RNAinter	Weak	pxMARGI
TSIX	AKT1	207	RD00133136	RNAinter	Weak	diMARGI
LINC02298	AKT1	207	RD00060295	RNAinter	Weak	pxMARGI
MIR29B2CHG	AKT1	207	-	NcPath	Weak	ceRNA
SLC9A3-AS1	AKT1	207	-	NcPath	Weak	ceRNA
HCG18	AKT1	207	-	NcPath	Weak	ceRNA
EBLN3P	AKT1	207	-	NcPath	Weak	ceRNA
H19	AKT1	207	-	NcPath	Weak	ceRNA
KCNQ1OT1	AKT1	207	-	NcPath	Weak	ceRNA
Rows per page: 10 1-10 of 7720 Page: 1 < >						

The lncRNA's associated with AML.

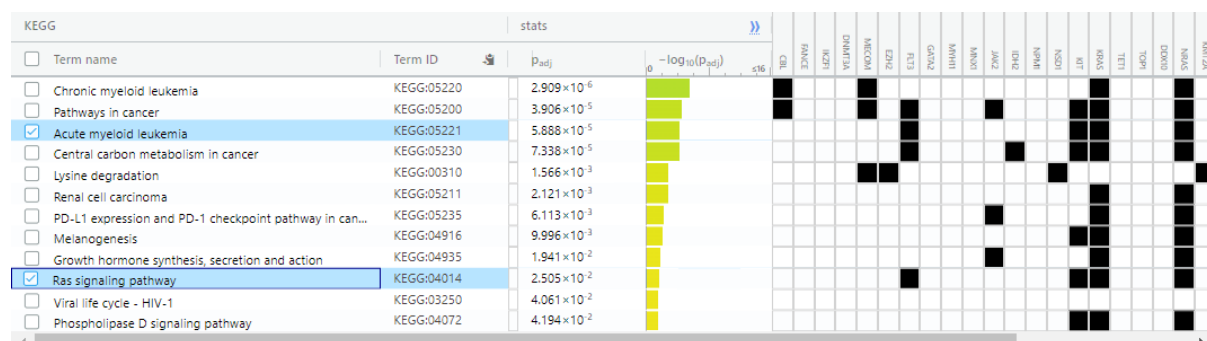
5.9 STATISTICAL ANALYSIS USING g:Profiler

g:Profiler is used to interpret gene data through statistical enrichment analysis (<https://biit.cs.ut.ee/gprofiler/gost>).

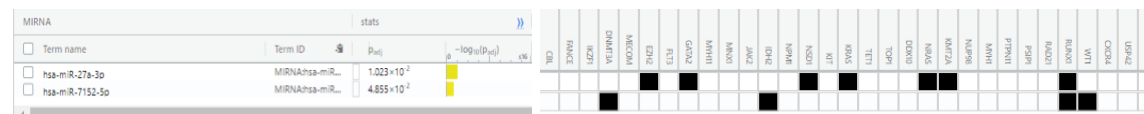
The common genes from three databases are given as input for a general overview and statistical analysis of genes and their associated functions. g:GOST is used to perform functional enrichment analysis, also known as over-representation analysis (ORA) or gene set enrichment analysis, on the input gene list. It maps genes to known functional information sources and detects statistically significantly enriched terms.

5.9.1 KEGG Pathway Statistical Analysis

KEGG Pathway Statistical analysis for disease (AML) and the pathway involved have P adjusted score of 5.88×10^{-5} and 2.505×10^{-2} and their respective genes involved.

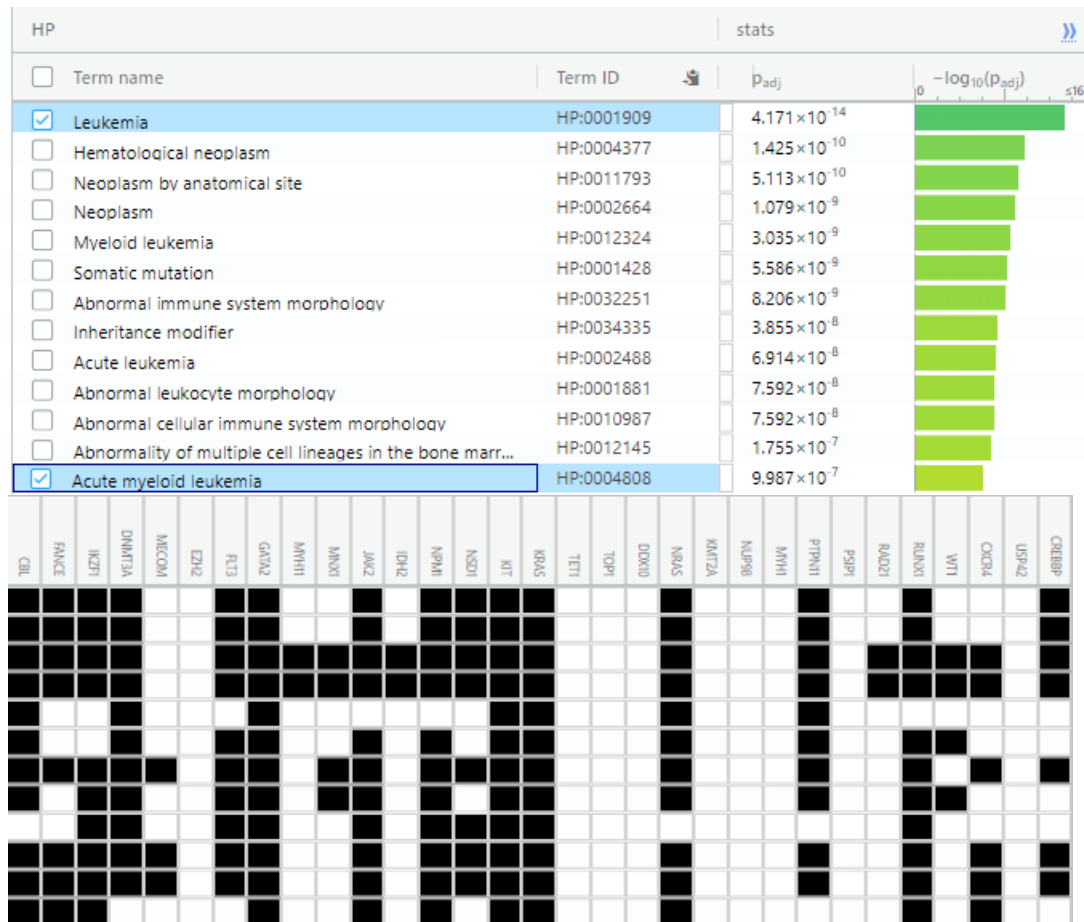


5.9.2 Statistical Analysis of miRNA



Statistical analysis of miRNA to the input genes as Padj score of 1.023×10^{-2} & 4.855×10^{-2} .

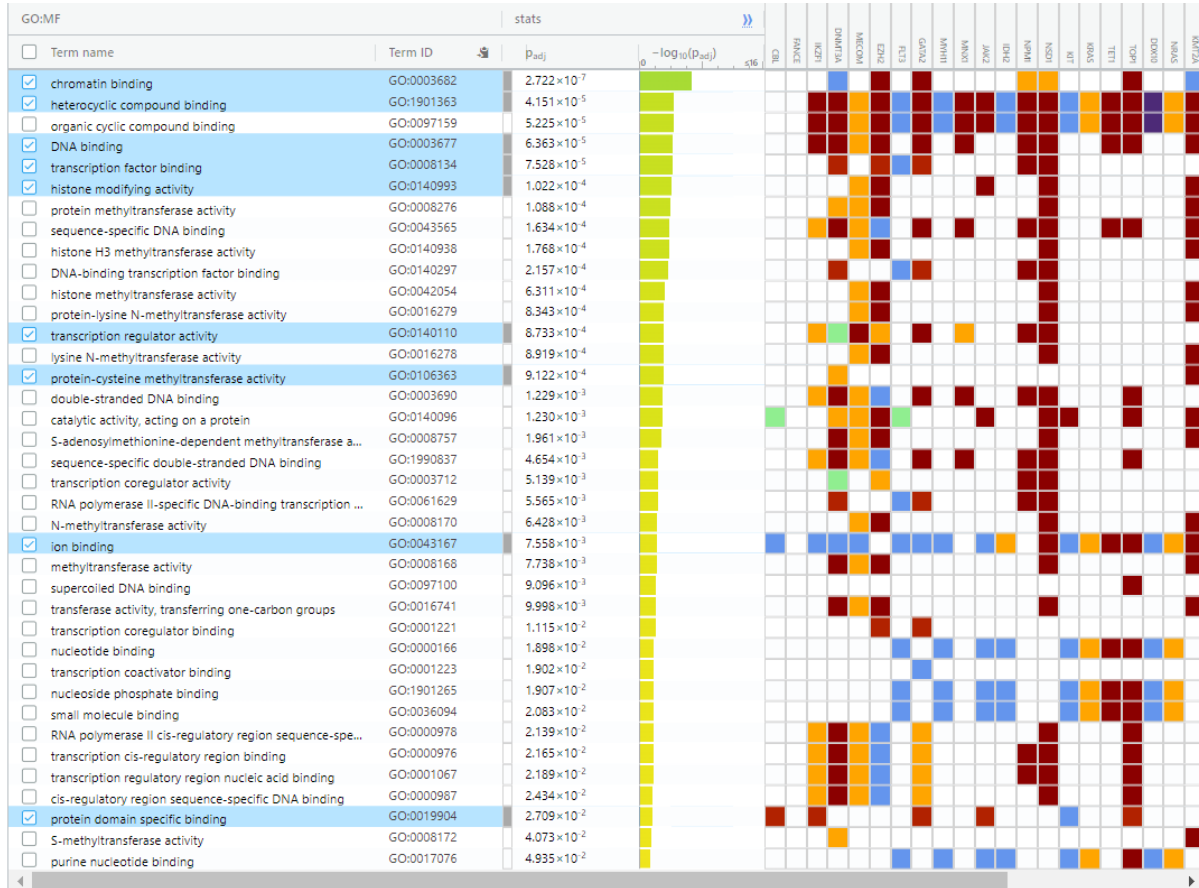
5.9.3 Statistical Analysis of Disease



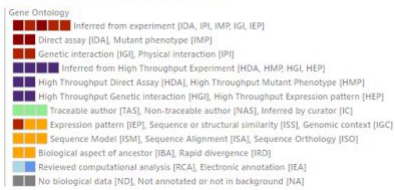
Statistical analysis of disease to an input gene has a Padj score of 4.171×10^{-14} for leukemia and 9.987×10^{-7} for acute myeloid leukemia.

5.9.4 Statistical Analysis of Gene Ontology: Molecular Function

General overview and statistical analysis of genes and their associated functions.



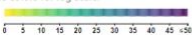
The colors for different evidence codes in the table:



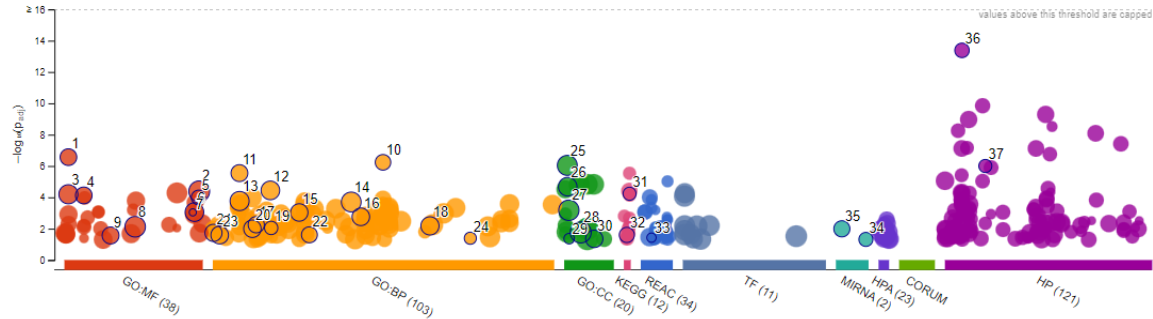
Biological pathways

- KEGG, Reactome
- Regulatory motifs in DNA
- TRANSFAC TFBS, miRtarBase
- Human Protein Atlas
- [Enhanced, Supported, Approved]
- CORUM protein complexes
- CORUM
- Human Phenotype Ontology
- Human Phenotype Ontology (sequence homologs in other species)

The colors for log scale:



5.9.5 Overview of Statistical Analysis



ID	Source	Term ID	Term Name	Padj (query_1)
1	GO:MF	GO:0003682	chromatin binding	2.722×10^{-7}
2	GO:MF	GO:1901363	heterocyclic compound binding	4.151×10^{-5}
3	GO:MF	GO:0003677	DNA binding	6.363×10^{-5}
4	GO:MF	GO:0008134	transcription factor binding	7.528×10^{-5}
5	GO:MF	GO:0140993	histone modifying activity	1.022×10^{-4}
6	GO:MF	GO:0140110	transcription regulator activity	8.733×10^{-4}
7	GO:MF	GO:0106363	protein-cysteine methyltransferase activity	9.122×10^{-4}
8	GO:MF	GO:0043167	ion binding	7.558×10^{-3}
9	GO:MF	GO:0019904	protein domain specific binding	2.709×10^{-2}
10	GO:BP	GO:0048872	homeostasis of number of cells	5.828×10^{-7}
11	GO:BP	GO:0006325	chromatin organization	2.842×10^{-6}
12	GO:BP	GO:0012501	programmed cell death	3.609×10^{-5}
13	GO:BP	GO:0006357	regulation of transcription by RNA polymerase II	1.717×10^{-4}
14	GO:BP	GO:0043412	macromolecule modification	1.982×10^{-4}
15	GO:BP	GO:0030097	hemopoiesis	9.250×10^{-4}
16	GO:BP	GO:0045321	leukocyte activation	1.746×10^{-3}
17	GO:BP	GO:0009314	response to radiation	5.593×10^{-3}
18	GO:BP	GO:0071495	cellular response to endogenous stimulus	6.640×10^{-3}
19	GO:BP	GO:0014068	positive regulation of phosphatidylinositol 3-kinas...	8.977×10^{-3}
20	GO:BP	GO:0008284	positive regulation of cell population proliferation	1.010×10^{-2}
21	GO:BP	GO:0000165	MAPK cascade	1.858×10^{-2}
22	GO:BP	GO:0032259	methylation	2.462×10^{-2}
23	GO:BP	GO:0001932	regulation of protein phosphorylation	2.561×10^{-2}
24	GO:BP	GO:0140718	facultative heterochromatin formation	4.078×10^{-2}
25	GO:CC	GO:0005654	nucleoplasm	9.131×10^{-7}
26	GO:CC	GO:0005694	chromosome	2.092×10^{-5}
27	GO:CC	GO:0005829	cytosol	6.677×10^{-4}
28	GO:CC	GO:0032991	protein-containing complex	1.739×10^{-2}
29	GO:CC	GO:0005859	muscle myosin complex	4.327×10^{-2}
30	GO:CC	GO:0070161	anchoring junction	4.473×10^{-2}
31	KEGG	KEGG:05221	Acute myeloid leukemia	5.888×10^{-5}
32	KEGG	KEGG:04014	Ras signaling pathway	2.505×10^{-2}
33	REAC	REAC:R-HSA-97...	FLT3 signaling by CBL mutants	3.754×10^{-2}
34	MIRNA	MIRNA:hsa-miR...	hsa-miR-7152-5p	4.855×10^{-2}
35	MIRNA	MIRNA:hsa-miR...	hsa-miR-27a-3p	1.023×10^{-2}
36	HP	HP:0001909	Leukemia	4.171×10^{-14}
37	HP	HP:0004808	Acute myeloid leukemia	9.987×10^{-7}

5.9.6 SNP Statistical Analysis Through g:Profiler

SNP data was taken from SNPedia by browsing for SNPs related to acute myeloid leukemia. (https://www.snpedia.com/index.php/Acute_myeloid_leukemia).

The resulting SNPs were taken for analysis using g: SNPense available in g:profiler

rs2302254

rs11554137

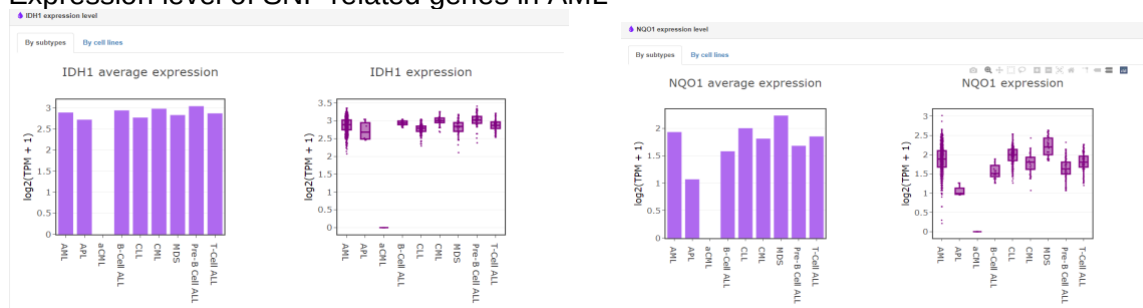
rs1131341

rs532545

A general statistical analysis of the variants was done.

id	chr	start	end	strand	gene ids	gene names	intron variant	missense variant	non-coding transcript exon variant	splice region variant	synonymous variant
rs2302254											
rs11554137	2	208248468	208248468	+	ENSG00000138413	IDH1				1	1
rs1131341	16	69714966	69714966	+	ENSG00000181019	NQO1	1	1	1	1	1
rs532545											

Expression level of SNP related genes in AML



5.9.7 SNPs Related to NRAS Gene

id	chr	start	end	strand	gene ids	gene names	missense variant	stop gained
rs121913237	1	114716126	114716126	+	ENSG00000213261	NRAS	6	
rs121913250	1	114716127	114716127	+	ENSG00000213261	NRAS	6	
rs121434596	1	114716123	114716123	+	ENSG00000213261	NRAS	6	
rs121434595	1	114716124	114716124	+	ENSG00000213261	NRAS	6	
rs121913255	1	114713907	114713907	+	ENSG00000213261	NRAS	4	
rs121913254	1	114713909	114713909	+	ENSG00000213261	NRAS	4	
rs11554290	1	114713908	114713908	+	ENSG00000213261	NRAS	6	

SNPs of NRAS were collected from the Uniport database and submitted for analysis using g: SNPense available in g:Profiler

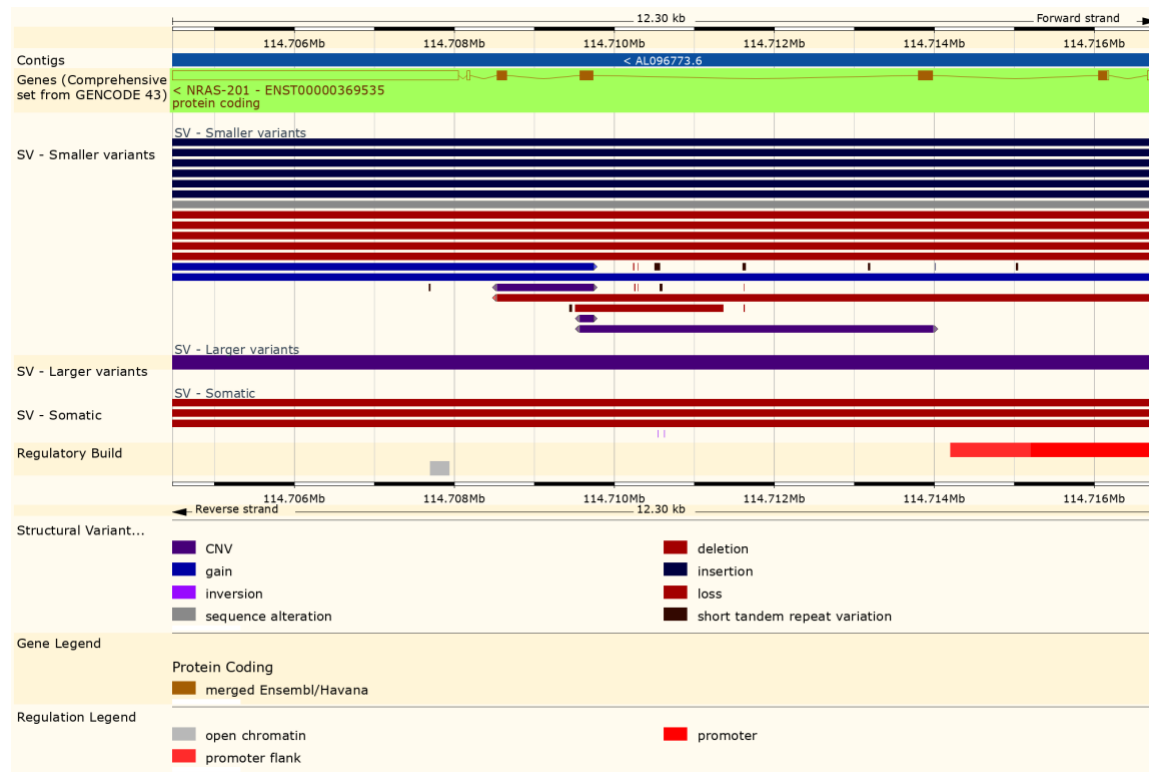
rs121913237 G>A
 rs121913250 G>C
 rs121913237 G>D
 rs121913250 G>R
 rs121913250 G>S
 rs121913237 G>V
 rs121434596 G>D
 rs121434595 G>R
 rs121434596 G>V
 rs121913255 Q>H
 rs121913255 Q>H
 rs121913254 Q>K
 rs11554290 Q>L
 rs11554290 Q>P

Based on the analysis, missense variants are more in number.

5.10 STRUCTURAL VARIANTS VISUALISED FROM Ensembl

The Ensembl Genome browser was visited to visualise structural variants present in the genes.

5.10.1 NRAS



Ensembl genome browser showing structural variants (SVs) in the human gene ENSG00000213281, specifically in the region between nucleotide positions 114,704,469 and 114,716,771 on chromosome 1, as well as the transcript ENST00000369535 that corresponds to this gene.

SVs that have been identified in this gene region and transcript, along with their type, location, and size.

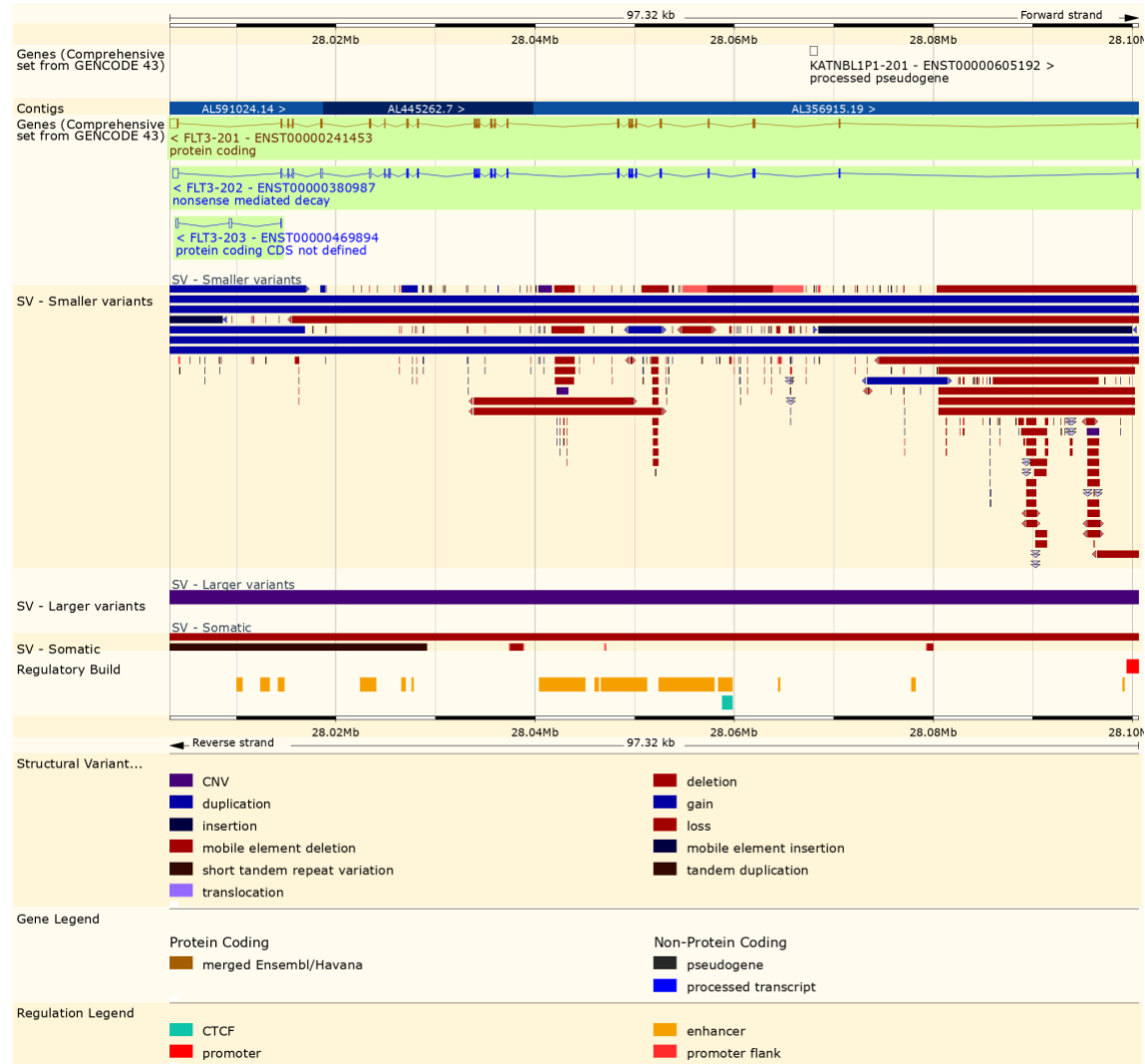
A deletion of 141 bp at position 114,704,617, which means that a segment of DNA has been removed from this location.

A deletion of 2.3 kb at position 114,708,122, which means that a larger segment of DNA has been removed from this location.

A duplication of 1.1 kb at position 114,708,201, which means that a segment of DNA has been copied and inserted into this location.

A translocation of 1.8 kb between positions 114,709,452 and 114,711,259, which means that a segment of DNA has been moved from one location to another on a different chromosome.

5.10.2 FLT3



Structural variants are genomic variations that involve larger segments of DNA, including deletions, duplications, inversions, and translocations, that can affect gene function and potentially lead to disease. SVs that have been identified in ENSG00000122025, along with their type, location, and size.

A deletion of 3.3 kb at position 28,030,868, which means that a segment of DNA has been removed from this location.

A duplication of 1.8 kb at position 28,072,464, which means that a segment of DNA has been copied and inserted into this location.

An inversion of 31.7 kb at position 28,076,435, which means that a segment of DNA has been flipped around and reinserted in the opposite orientation.

A translocation of 103.3 kb between positions 28,043,540 and 28,146,846, which means that a segment of DNA has been moved from one location to another on a different chromosome.

5.10.3 KIT

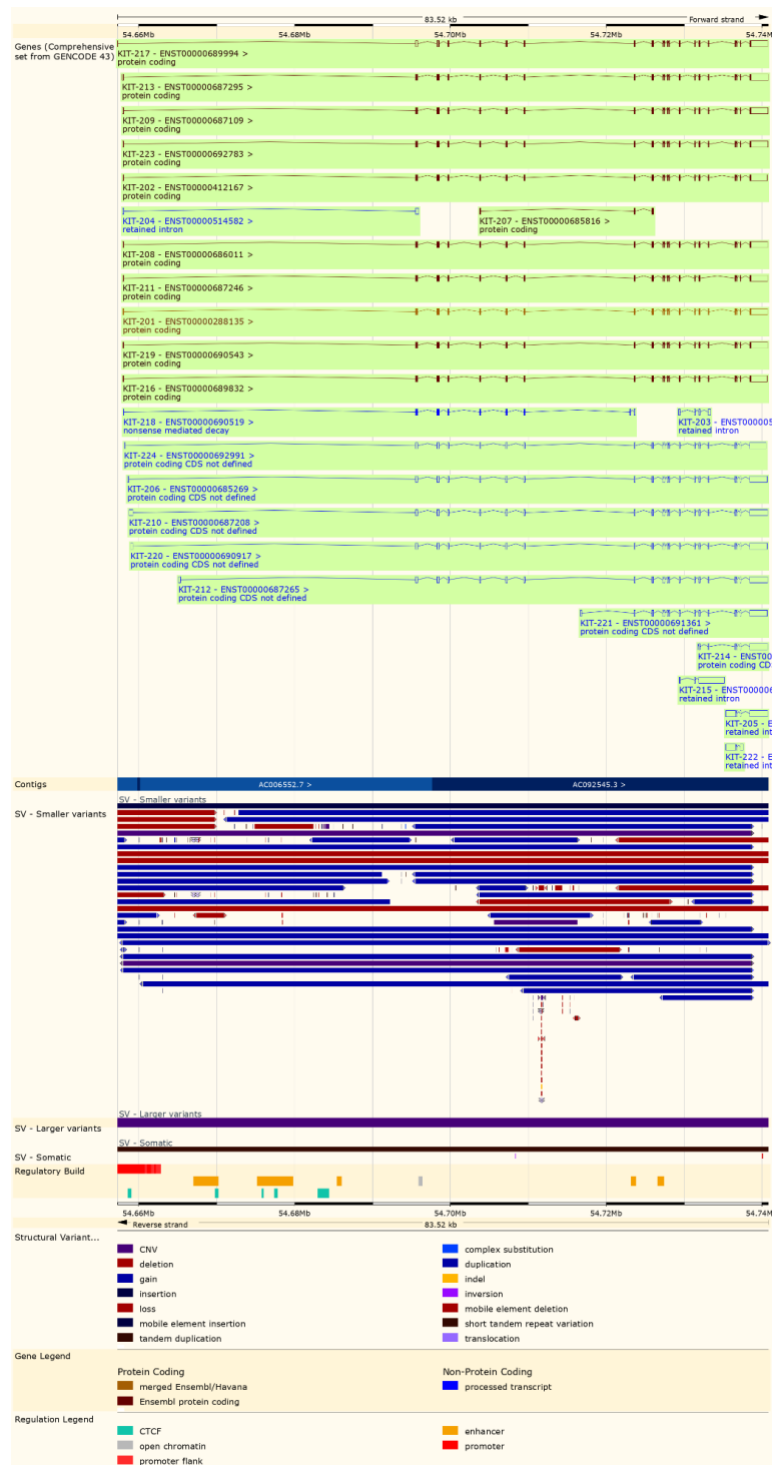
Ensembl genome browser showing structural variants (SVs) in the human gene ENSG00000157404, specifically in the region between nucleotide positions 54,657,267 and 54,740,783 on chromosome 4.

On this page, you can see a list of SVs that have been identified in this gene region, along with their type, location, and size. Some of the SVs listed include:

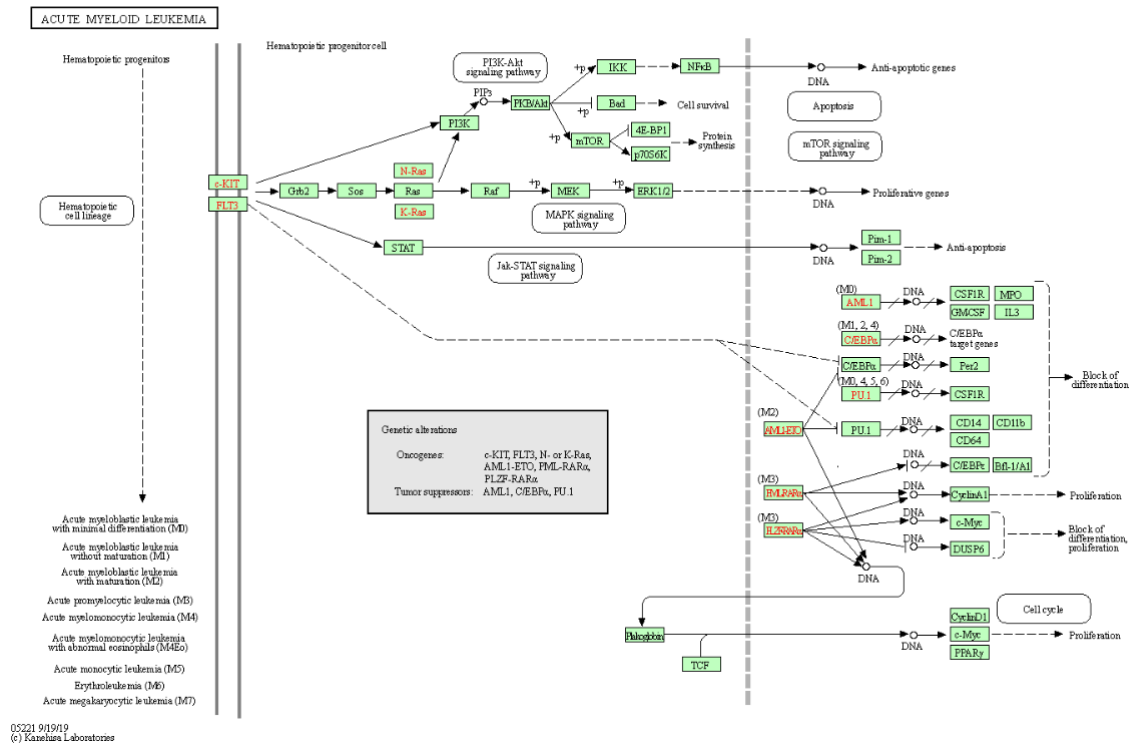
A deletion of 77.5 kb at position 54,659,215, which means that a large segment of DNA has been removed from this location.

A duplication of 5.5 kb at position 54,691,118, which means that a segment of DNA has been copied and inserted into this location.

An inversion of 20.5 kb at position 54,697,555, which means that a segment of DNA has been flipped around and reinserted in the opposite orientation.



5.11 PATHWAY ANALYSIS USING KEGG



The pathway shows the molecular interactions and signalling, pathways that are involved in the development and progression of AML.

The pathway includes genes and proteins that are mutated or dysregulated in AML, such as FLT3, NPM1, KIT, and DNMT3A.

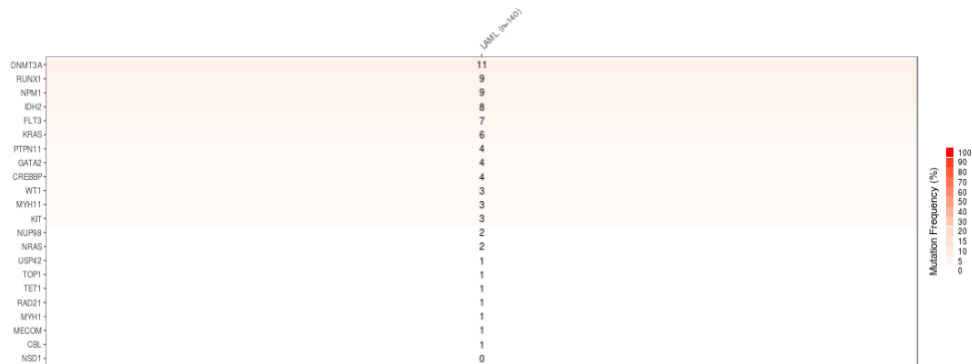
It highlights several key cellular processes that are altered in AML, including abnormal cell proliferation, differentiation, and apoptosis (programmed cell death).

It shows how these alterations lead to the accumulation of immature myeloid cells in the bone marrow and blood, which is a hallmark of AML. It also includes drugs that are used to treat AML, such as tyrosine kinase inhibitors and chemotherapy agents. The pathway provides a framework for understanding the molecular mechanisms underlying AML and for developing new therapeutic strategies to treat this disease.

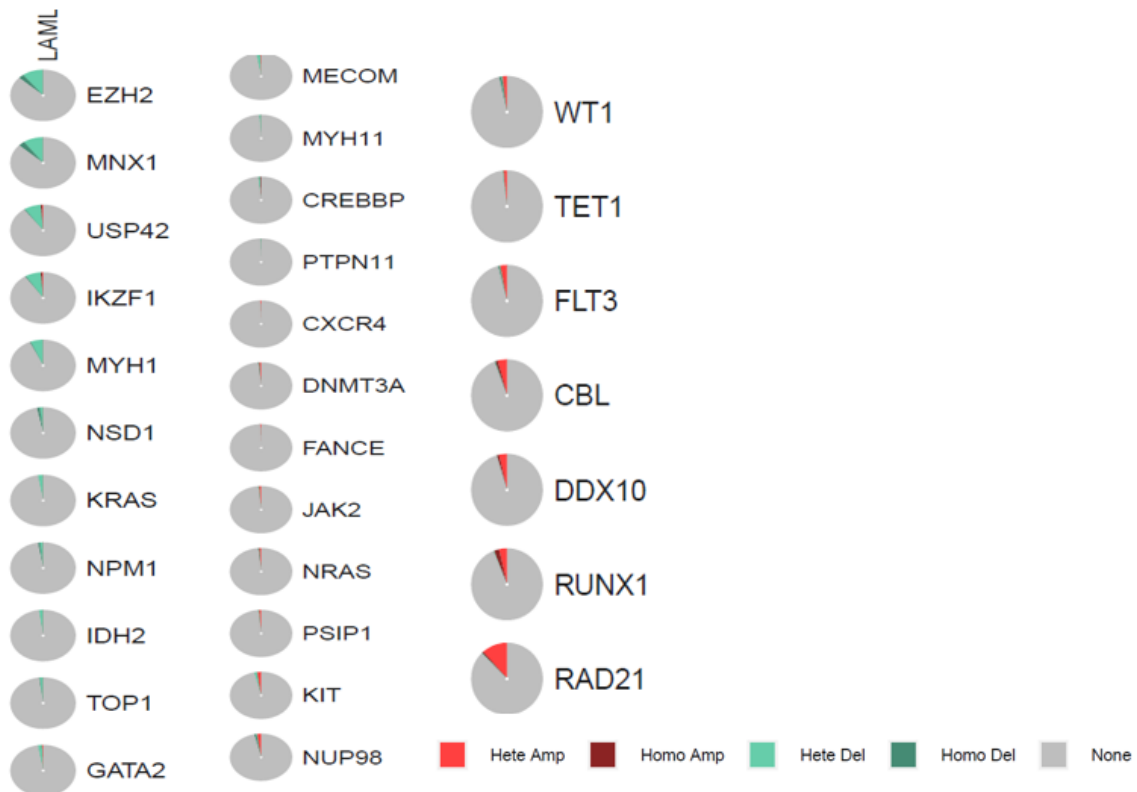
5.12 STATISCAL ANALYSIS USING GSCALite

5.12.1 Single Nucleotide Polymorphism Statistical Analysis Using GSCALite

SNV Percentage Profile – It shows the number of SNPs associated with the gene and their frequency of mutation.



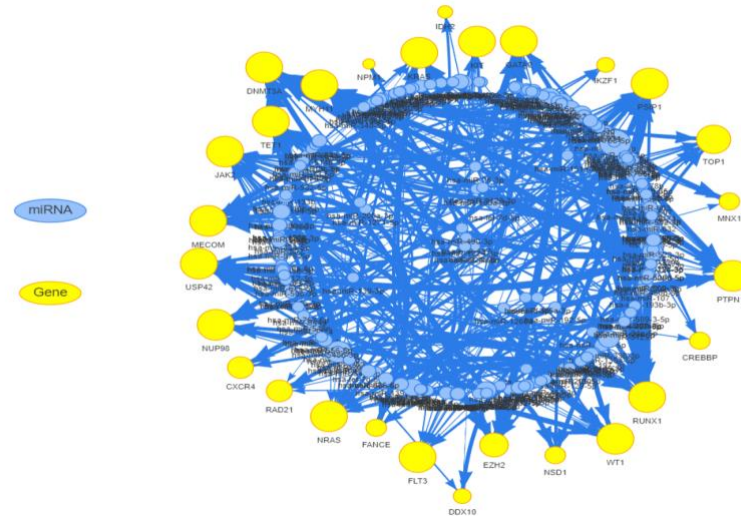
5.12.2 Statistical Analysis of CNV Using GSCALite



Statistical analysis of COPY NUMBER variation observed in the common genes taken from 3 gene data sets.

Respective deletions and amplifications are observed within the gene.

5.12.3 Visualization of miRNA and Gene Network Using GSCALite



Network of the Genes and miRNA.

TABLES

Mutations and Translocations Data analysis is presented in Tables 1 to 7.

Table 1: cBioPortal Mutation Data

Gene	Frequency	Is Cancer Gene		Gene	Frequency	Is Cancer Gene
MYC	1.30%	Yes		ASMTL	0.70%	Yes
CHD2	0.70%	Yes		LRP1B	0.70%	Yes
EP400	0.70%	Yes		KDM6A	0.70%	Yes
NOTCH2	0.70%	Yes		SETD2	0.70%	Yes
MYH11	0.70%	Yes		MKI67	0.70%	Yes
FLT3	6.70%	Yes		U2AF1	0.70%	Yes
IDH2	2.70%	Yes		IKZF1	2.70%	Yes
NPM1	0.70%	Yes		SMARCB1	0.70%	Yes
NRAS	14.70%	Yes		DOT1L	0.70%	Yes
ETV6	0.70%	Yes		ABL1	0.70%	Yes
BCL7A	0.70%	Yes		RIT1	0.70%	Yes
NFE2	0.70%	Yes		WT1	3.30%	Yes
RAC2	0.70%	Yes		DAZAP1	0.70%	Yes
RAD17	0.70%	Yes		ZBTB7A	0.70%	Yes
HLA-C	1.30%	Yes		PTPRT	0.70%	Yes
RAD21	0.70%	Yes		STAG1	0.70%	Yes
PTPN11	6.00%	Yes		BRD4	0.70%	Yes
EP300	1.30%	Yes		TOP1	0.70%	Yes
ZNF384	0.70%	Yes		PPFIBP1	0.70%	Yes
EZH2	1.30%	Yes		PPM1D	0.70%	Yes
PTPN6	0.70%	Yes		TYK2	0.70%	Yes
MEF2C	0.70%	Yes		MLLT10	0.70%	Yes
CXCR4	0.70%	Yes		SMC3	2.00%	Yes
EPHA5	0.70%	Yes		KIT	10.00%	Yes
BCOR	0.70%	Yes		GATA2	2.70%	Yes
SF3B1	1.30%	Yes		GATA1	0.70%	Yes
CCND2	2.70%	Yes		JAK2	1.30%	Yes
ERBB2	0.70%	Yes		JAK3	1.30%	Yes
U2AF2	0.70%	Yes		LRIG3	0.70%	Yes
FANCE	0.70%	Yes		DDX10	0.70%	Yes
PRKD1	0.70%	Yes		KRAS	6.00%	Yes
ERF	0.70%	Yes		RARA	0.70%	Yes
TET2	3.30%	Yes		KMT2D	1.30%	Yes
ERCC3	0.70%	Yes		CSF3R	1.30%	Yes

Table describing the frequency of mutation.

NRAS,KIT,FLT3 and PTPN11 have frequency < 6 %

Out of 394 mutated genes, only 68 are cancer-causing genes. Based on the frequency of the mutation of genes, NRAS gene appears to be the gene with the highest frequency of mutated genes.

Table 2: DoCM Mutation Data-1

Chromosome	Start	Stop	Read	Variant	Reference _version	Gene	Mutation type	Amino acid	Diseases
2	25457242	25457242	C	A	GRCh37	DNMT3A	missense	p.R882L	AML
2	25457243	25457243	G	C	GRCh37	DNMT3A	missense	p.R882G	AML
2	25457243	25457243	G	T	GRCh37	DNMT3A	missense	p.R882S	AML
2	198266832	198266832	T	A	GRCh37	SF3B1	missense	p.K700N	AML
2	198266832	198266832	T	G	GRCh37	SF3B1	missense	p.K700N	AML
2	198266833	198266833	T	A	GRCh37	SF3B1	missense	p.K700I	AML
2	198266833	198266833	T	C	GRCh37	SF3B1	missense	p.K700R	AML
2	198266833	198266833	T	G	GRCh37	SF3B1	missense	p.K700T	AML
2	198266834	198266834	T	C	GRCh37	SF3B1	missense	p.K700E	AML
2	198266834	198266834	T	A	GRCh37	SF3B1	stop_lost	p.K700*	AML
2	198266834	198266834	T	G	GRCh37	SF3B1	missense	p.K700Q	AML
2	198267359	198267359	C	G	GRCh37	SF3B1	missense	p.K666N	AML
2	198267360	198267360	T	C	GRCh37	SF3B1	missense	p.K666R	AML
2	198267360	198267360	T	G	GRCh37	SF3B1	missense	p.K666T	AML
2	209113112	209113112	C	G	GRCh37	IDH1	missense	p.R132P	AML
2	209113113	209113113	G	A	GRCh37	IDH1	missense	p.R132C	AML
3	128200113	128200113	G	A	GRCh37	GATA2	missense	p.R398W	AML
3	128200744	128200744	G	A	GRCh37	GATA2	missense	p.T354M	AML
4	55599320	55599321	GA	AT	GRCh37	KIT	missense	p.D816I	AML
4	55602765	55602765	G	C	GRCh37	KIT	synonymou	p.L862	AML
5	170837546	170837547	-	CATG	GRCh37	NPM1	frameshift	p.W288HX	AML
5	170837546	170837547	-	CCTG	GRCh37	NPM1	frameshift	p.W288PX	AML
5	170837546	170837547	-	TCAG	GRCh37	NPM1	frameshift	p.W288SX	AML
5	170837546	170837547	-	TCTG	GRCh37	NPM1	frameshift	p.W288SX	AML
9	27204927	27204927	G	C	GRCh37	TEK	missense	p.G743A	AML
11	32417909	32417910	-	ACCGTACA	GRCh37	WT1	protein_alt	p.S381LYG	AML
13	28592620	28592620	T	C	GRCh37	FLT3	missense	p.Y842C	AML
13	28592621	28592621	A	G	GRCh37	FLT3	missense	p.Y842H	AML
13	28592622	28592622	G	T	GRCh37	FLT3	missense	p.N841K	AML
13	28592624	28592624	T	G	GRCh37	FLT3	missense	p.N841H	AML
13	28592629	28592629	T	C	GRCh37	FLT3	missense	p.D839G	AML
13	28592629	28592629	T	C	GRCh37	FLT3	missense	p.D839G	AML
13	28592637	28592637	G	C	GRCh37	FLT3	missense	p.I836M	AML
13	28592637	28592639	GAT	-	GRCh37	FLT3	inframe	p.I836-	AML
13	28592638	28592638	A	C	GRCh37	FLT3	missense	p.I836S	AML
13	28592639	28592639	T	G	GRCh37	FLT3	missense	p.I836L	AML
13	28592639	28592639	T	C	GRCh37	FLT3	missense	p.I836V	AML

Table 3: DoCM Mutation Data-2

Chromosome	Start	Stop	Read	Variant	Reference_v ersion	Gene	Mutation type	Amino acid	Diseases
13	28592639	28592639	T	A	GRCh37	FLT3	missense	p.I836F	AML
13	28592640	28592640	A	C	GRCh37	FLT3	missense	p.D835E	AML
13	28592640	28592640	A	T	GRCh37	FLT3	missense	p.D835E	AML
13	28592640	28592642	ATC	-	GRCh37	FLT3	inframe	p.D835-	AML
13	28592640	28592640	A	C	GRCh37	FLT3	missense	p.D835E	AML
13	28592640	28592640	A	T	GRCh37	FLT3	missense	p.D835E	AML
13	28592641	28592641	T	A	GRCh37	FLT3	missense	p.D835V	AML
13	28592641	28592641	T	G	GRCh37	FLT3	missense	p.D835A	AML
13	28592641	28592642	TC	AA	GRCh37	FLT3	missense	p.D835F	AML
13	28592641	28592641	T	A	GRCh37	FLT3	missense	p.D835V	AML
13	28592641	28592641	T	G	GRCh37	FLT3	missense	p.D835A	AML
13	28592642	28592642	C	A	GRCh37	FLT3	missense	p.D835Y	AML
13	28592642	28592642	C	G	GRCh37	FLT3	missense	p.D835H	AML
13	28592642	28592642	C	T	GRCh37	FLT3	missense	p.D835N	AML
13	28592642	28592642	C	A	GRCh37	FLT3	missense	p.D835Y	AML
13	28592642	28592642	C	G	GRCh37	FLT3	missense	p.D835H	AML
13	28592642	28592642	C	T	GRCh37	FLT3	missense	p.D835N	AML
13	28601359	28601359	A	C	GRCh37	FLT3	missense	p.F691L	AML
13	28601359	28601359	A	T	GRCh37	FLT3	missense	p.F691L	AML
13	28601373	28601373	T	A	GRCh37	FLT3	missense	p.I687F	AML
13	28602340	28602340	G	C	GRCh37	FLT3	missense	p.N676K	AML
13	28602340	28602340	G	T	GRCh37	FLT3	missense	p.N676K	AML
13	28602340	28602340	G	T	GRCh37	FLT3	missense	p.N676K	AML
13	28602342	28602342	T	C	GRCh37	FLT3	missense	p.N676D	AML
13	28602381	28602381	T	G	GRCh37	FLT3	missense	p.K663Q	AML
13	28602416	28602416	T	C	GRCh37	FLT3	missense	p.D651G	AML
13	28608111	28608111	C	A	GRCh37	FLT3	missense	p.G619C	AML
13	28608276	28608276	A	G	GRCh37	FLT3	missense	p.F594L	AML
13	28608281	28608281	A	G	GRCh37	FLT3	missense	p.V592A	AML
13	28608284	28608284	T	C	GRCh37	FLT3	missense	p.Y591C	AML
13	28608285	28608285	A	C	GRCh37	FLT3	missense	p.Y591D	AML
13	28608320	28608320	A	G	GRCh37	FLT3	missense	p.V579A	AML
13	28608341	28608341	T	C	GRCh37	FLT3	missense	p.Y572C	AML
15	90631837	90631837	C	G	GRCh37	IDH2	missense	p.R172S	AML
17	7577536	7577536	T	A	GRCh37	TP53	missense	p.R249W	AML
21	36231783	36231783	G	A	GRCh37	RUNX1	stop_lost	p.R201*	AML
21	36252876	36252876	C	A	GRCh37	RUNX1	missense	p.R162S	AML
21	36252876	36252876	C	G	GRCh37	RUNX1	missense	p.R162S	AML
21	36252877	36252877	C	T	GRCh37	RUNX1	missense	p.R162K	AML
21	36252878	36252878	T	C	GRCh37	RUNX1	missense	p.R162G	AML
21	44514777	44514777	T	C	GRCh37	U2AF1	missense	p.Q157R	AML

Mutation data from Database Of Curated Mutations

The Table describes about the Chromosome on which the mutation is taking place, the start and stop positions, the change in the reads, the gene present, mutation type and respective amino acid involved in AML.

Table 4: Mutation Type T from COSMIC

Gene Symbol	Genome Location	Chr Band	Molecular Genetics	Mutation Types	Translocation Partner
DEK	6:18223868-18264823	22.3	Dom	T	NUP214
MN1	22:27748277-27801498	12.1	Dom	T	ETV6
ABI1	10:26746593-26860935	12.1	Dom	T	KMT2A
ABL2	1:179099327-179229601	25.2	Dom	T	ETV6
ARNT	1:150809705-150876768	21.3	Dom	T	ETV6
CBFB	16:67029245-67101022	22.1	Dom	T	MYH11
CDX2	13:27962137-27971139	12.2	Dom	T	ETV6
CLP1	11:57657158-57661868	12.1	Dom	T	KMT2A
ELF4	X:130064928-130110495	26.1	Dom	T	ERG
GAS7	17:9913850-10198551	13.1	Dom	T	KMT2A
GMPS	3:155870536-155944026	25.31	Dom	T	KMT2A
KNL1	15:40594020-40664342	15.1	Dom	T	KMT2A
MLF1	3:158582835-158606459	25.32	Dom	T	NPM1
MNX1	7:157004853-157010651	36.3	Dom	T	ETV6
NSD1	5:177133925-177300215	35.3	Dom	T	NUP98
PER1	17:8140472-8152435	13.1	Dom	T	ETV6
RPN1	3:128619970-128650849	21.3	Dom	T	MECOM
TET1	10:68560656-68694482	21.3	Dom	T	KMT2A
TOP1	20:41028818-41124487	12	Dom	T	NUP98
ACSL6	5:131949973-132011662	31.1	Dom	T	ETV6
CHIC2	4:54009789-54064690	12	Dom	T	ETV6
DDX10	11:108665025-108940930	22.3	Dom	T	NUP98
FNBP1	9:129887187-130043162	34.11	Dom	T	KMT2A
HOXA9	7:27162435-27165530	15.2	Dom	T	NUP98, MSI2
KAT6B	10:74826625-75032622	22.2	Dom	T	CREBBP, KANSL1
KDMSA	12:280129-389454	13.33	Dom	T	NUP98
LASP1	17:38869859-38921770	12	Dom	T	KMT2A
MECOM	3:169083499-169146305	26.2	Dom	T	RUNX1, ETV6, PRDM16, RPN1
MYH11	16:15703172-15857011	13.11	Dom	T	CBFB
NCOA2	8:70109762-70403805	13.3	Dom	T	KAT6A, HEY1, NCOA2
NUP98	11:3675010-3797792	15.4	Dom	T	HOXA9, NSD1, WHSC1L1, DDX10, TOP1, HOXD13, PRRX1, HOXA13, HOXD11, HOXA11, RAP1GDS1, HOXC11
PRRX1	1:170663906-170739419	24.2	Dom	T	NUP98
PSIP1	9:15464066-15511019	22.3	Dom	T	NUP98
RPL22	1:6185020-6199612	36.31	Dom	T	RUNX1
RUNX1	21:34787801-35049344	22.12	Dom	T	RPL22, MECOM, CBFA2T3, RUNX1T1, ETV6, AFF3
SEPTIN5	22:19714464-19723319	11.21	Dom	T	KMT2A
SEPTIN6	X:119617639-119693370	24	Dom	T	KMT2A
SEPTIN9	17:77281410-77500592	25.3	Dom	T	KMT2A
HOXA13	7:27194364-27200106	15.2	Dom	T	NUP98
HOXC11	12:53973126-53977643	13.13	Dom	T	NUP98
HOXC13	12:53938765-53946544	13.13	Dom	T	NUP98
HOXD11	2:176107286-176109588	31.1	Dom	T	NUP98
HOXD13	2:176092891-176095938	31.1	Dom	T	NUP98
NUP214	9:131125561-131234670	34.13	Dom	T	DEK, SET, ABL1
TRIP11	14:91965991-92040059	32.12	Dom	T	PDGFRB
CBFA2T3	16:88874858-88977204	24.3	Dom	T	RUNX1, GLIS2
RUNX1T1	8:91954967-92095654	21.3	Dom	T	KMT2A, RUNX1
ARHGEF12	11:120337078-120489936	23.3	Dom	T	KMT2A

Table describing about translocation mutation taken from COSMIC.

TRANSLOCATION

Translocations are a type of chromosomal abnormality that occur when a segment of one chromosome breaks off and attaches to another chromosome.

This results in the movement of genetic material from one chromosome to another and leads to alterations in gene expression and function.

The most common translocation in pediatric AML involves the fusion of the genes RUNX1 and CBFA2T1, resulting in a fusion gene, RUNX1-RUNX1T1. This translocation is also known as t(8;21) and is associated with a relatively favorable prognosis in pediatric AML.

Table 5: Translocation ETV6

Gene Symbol	Genome Location	Chr Band	Role in Cancer	Mutation Types	Translocation Partner
MN1	22:27748277-27801498	12.1	oncogene, fusion	T	ETV6
ABL2	1:179099327-179229601	25.2	oncogene, fusion	T	ETV6
ARNT	1:150809705-150876768	21.3	oncogene, TSG, fusion	T	ETV6
CDX2	13:27962137-27971139	12.2	TSG, fusion	T	ETV6
MNX1	7:157004853-157010651	36.3	fusion	T	ETV6
PER1	17:8140472-8152435	13.1	TSG, fusion	T	ETV6
ACSL6	5:131949973-132011662	31.1	fusion	T	ETV6
CHIC2	4:54009789-54064690	12	fusion	T	ETV6

ETV6-RUNX1: This fusion gene is found in approximately 20% of pediatric AML cases and is the result of a translocation between the ETV6 gene on chromosome 12 and the RUNX1 gene on chromosome 21.

Table 6: Translocation KMT2A

Gene Symbol	Genome Location	Chr Band	Mutation Types	Translocation Partner
CBL	11:119206276-11930	23.3	T, Mis, S, O	KMT2A
ABI1	10:26746593-268609	12.1	T	KMT2A
CLP1	11:57657158-576618	12.1	T	KMT2A
GAS7	17:9913850-1019855	13.1	T	KMT2A
GMPS	3:155870536-155944	25.31	T	KMT2A
KNL1	15:40594020-406643	15.1	T	KMT2A
TET1	10:68560656-686944	21.3	T	KMT2A
FNBP1	9:129887187-130043	34.11	T	KMT2A
LASP1	17:38869859-389217	12	T	KMT2A
SEPTIN5	22:19714464-197233	11.21	T	KMT2A
SEPTIN6	X:119617639-119693	24	T	KMT2A
SEPTIN9	17:77281410-775005	25.3	T	KMT2A
ARHGAP26	5:142770377-143229	31.3	T, F, S	KMT2A
ARHGEF12	11:120337078-12048	23.3	T	KMT2A
KMT2A	11:118436490-1185268	23.3	T, O	KMT2A, MLLT1, AFF1, MLLT3, MLLT4, FOXO4, MLLT10, MLLT6, ELL, EPS15, MLLT11, CREBBP, SH3GL1, FNBP1, SEPT5, SEPT9, GPHN, GMPS, ABI1, ARHGEF12, GAS7, FOXO3, AFF3, TET1, SEPT6, LPP, RUNX1T1, ARHGAP26, EP300, PICALM, CLP1, PRRC1, DAB2IP, MAPRE1, KIAA1524, SEPT2, SARNP, ZFYVE19, TOP3A, CASC5, EEFSEC, FRYL, SORBS2, ABI2, LASP1, CASP8AP2, AFF4, NRIP3, NCKIPSD, CBL, ACTN4, PDS5A, CT45A2, MYO1F, KIAA0284, BTBD18

- KMT2A-MLLT3 (also known as MLL-AF9): This fusion gene is the most common KMT2A fusion gene in AML, accounting for up to 40% of cases.
- It is the result of a translocation between the KMT2A gene on chromosome 11 and the MLLT3 gene on chromosome 9.

Table 7: Translocation NUP98

Gene Symbol	Genome Location	Chr Band	Mutation Types	Translocation Partner
NSD1	5:177133925-1	35.3	T	NUP98
NSD3	8:38269697-38	11.23	T, A	NUP98
TOP1	20:41028818-4	12	T	NUP98
DDX10	11:108665025-	22.3	T	NUP98
KDM5A	12:280129-389	13.33	T	NUP98
PRRX1	1:170663906-1	24.2	T	NUP98
PSIP1	9:15464066-15	22.3	T	NUP98
HOXA13	7:27194364-27	15.2	T	NUP98
HOXC11	12:53973126-5	13.13	T	NUP98
HOXC13	12:53938765-5	13.13	T	NUP98
HOXD11	2:176107286-1	31.1	T	NUP98
HOXD13	2:176092891-1	31.1	T	NUP98

NUP98-NSD1: This fusion gene results from a translocation between the NUP98 gene on chromosome 11 and the NSD1 gene on chromosome 5. It is found in approximately 5-10% of AML cases and is associated with a poor prognosis.

Genes Data analysis is presented in Tables 8 to 14.

Table 8: dbLGL genes

Gene ID	Gene Symbol	Cytoband	Gene Type		Gene ID	Gene Symbol	Cytoband	Gene Type
21	ABCA3	16p13.3	protein-coding		2122	MECOM	3q26.2	protein-coding
84159	ARID5B	10q21.2	protein-coding		406948	MIR15A	13q14.2	ncRNA
79870	BAALC	8q22.3	protein-coding		407021	MIR29A	7q32.3	ncRNA
573	BAG1	9p13.3	protein-coding		494324	MIR375	2q35	ncRNA
79365	BHLHE41	12p12.1	protein-coding		724033	MIR663A	20p11.1	ncRNA
332	BIRC5	17q25.3	protein-coding		3110	MNX1	7q36.3	protein-coding
867	CBL	11q23.3	protein-coding		4524	MTHFR	1p36.22	protein-coding
945	CD33	19q13.41	protein-coding		4684	NCAM1	11q23.2	protein-coding
10225	CD96	3q13.13-q	protein-coding		4869	NPM1	5q35.1	protein-coding
1029	CDKN2A	9p21.3	protein-coding		4893	NRAS	1p13.2	protein-coding
1050	CEBPA	19q13.11	protein-coding		64324	NSD1	5q35.3	protein-coding
1118	CHIT1	1q32.1	protein-coding		4928	NUP98	11p15.4	protein-coding
7852	CXCR4	2q22.1	protein-coding		23532	PRAME	22q11.22	protein-coding
1788	DNMT3A	2p23.3	protein-coding		63976	PRDM16	1p36.32	protein-coding
2146	EZH2	7q36.1	protein-coding		11168	PSIP1	9p22.3	protein-coding
2322	FLT3	13q12.2	protein-coding		5781	PTPN11	12q24.13	protein-coding
2624	GATA2	3q21.3	protein-coding		5888	RAD51	15q15.1	protein-coding
3315	HSPB1	7q11.23	protein-coding		861	RUNX1	21q22.12	protein-coding
3417	IDH1	2q34	protein-coding		6774	STAT3	17q21.2	protein-coding
3418	IDH2	15q26.1	protein-coding		6776	STAT5A	17q21.2	protein-coding
3620	IDO1	8p11.21	protein-coding		6890	TAP1	6p21.32	protein-coding
10320	IKZF1	7p12.2	protein-coding		6891	TAP2	6p21.32	protein-coding
3563	IL3RA	Xp22.33 ar	protein-coding		80312	TET1	10q21.3	protein-coding
3569	IL6	7p15.3	protein-coding		7980	TFPI2	7q21.3	protein-coding
3815	KIT	4q12	protein-coding		7070	THY1	11q23.3	protein-coding
3824	KLRD1	12p13.2	protein-coding		84132	USP42	7p22.1	protein-coding
4297	KMT2A	11q23.3	protein-coding		7422	VEGFA	6p21.1	protein-coding
3845	KRAS	12p12.1	protein-coding		7490	WT1	11p13	protein-coding
3953	LEPR	1p31.3	protein-coding		331	XIAP	Xq25	protein-coding
					84911	ZNF382	19q13.12	protein-coding

List of genes taken from Leukemia Gene Literature Database.

Table 9: Cosmic Somatic Genes-1

Gene Symbol	Mutation Types	Translocation Partner	Chr Band	Genome Location
CBL	T, Mis, S, O	KMT2A	23.3	11:119206276-119308149
DEK	T	NUP214	22.3	6:18223868-18264823
MN1	T	ETV6	12.1	22:27748277-27801498
ABI1	T	KMT2A	12.1	10:26746593-26860935
ABL2	T	ETV6	25.2	1:179099327-179229601
ARNT	T	ETV6	21.3	1:150809705-150876768
CBFB	T	MYH11	22.1	16:67029245-67101022
CBLB	Mis, S		13.11	3:105655461-105869043
CBLC	M		13.32	19:44777869-44800634
CDX2	T	ETV6	12.2	13:27962137-27971139
CLP1	T	KMT2A	12.1	11:57657158-57661868
ELF4	T	ERG	26.1	X:130064928-130110495
FLT3	Mis, O		12.2	13:28003274-28100592
GAS7	T	KMT2A	13.1	17:9913850-10198551
GMPS	T	KMT2A	25.31	3:155870536-155944026
JAK2	T, Mis, O	ETV6, PCM1, BCL2	24.1	9:4985245-5128183
KNL1	T	KMT2A	15.1	15:40594020-40664342
MLF1	T	NPM1	25.32	3:158582835-158606459
MNX1	T	ETV6	36.3	7:157004853-157010651
NSD1	T	NUP98	35.3	5:177133925-177300215
NSD3	T, A	NUP98	11.23	8:38269697-38382272
PER1	T	ETV6	13.1	17:8140472-8152435
RPN1	T	MECOM	21.3	3:128619970-128650849
TET1	T	KMT2A	21.3	10:68560656-68694482
TOP1	T	NUP98	12	20:41028818-41124487
ACSL6	T	ETV6	31.1	5:131949973-132011662
CEBPA	Mis, N, F		13.11	19:33299934-33302564
CHIC2	T	ETV6	12	4:54009789-54064690
DDX10	T	NUP98	22.3	11:108665025-108940930
FNBP1	T	KMT2A	34.11	9:129887187-130043162
GATA2	Mis		21.3	3:128479427-128493185
HOXA9	T	NUP98, MSI2	15.2	7:27162435-27165530
KAT6A	T, A	CREBBP, NCOA2	11.21	8:41929479-42051987

Table 10: Cosmic Somatic Genes-2

Gene Symbol	Mutation Types	Translocation Partner	Chr Band	Genome Location
KAT6A	T, A	CREBBP, NCOA2	11.21	8:41929479-42051987
KAT6B	T	CREBBP, KANSL1	22.2	10:74826625-75032622
KDM5A	T	NUP98	13.33	12:280129-389454
KMT2A	T, O	KMT2A, MLLT1	23.3	11:118436490-118526832
LASP1	T	KMT2A	12	17:38869859-38921770
MECOM	T	RUNX1, ETV6, P	26.2	3:169083499-169146305
MYH11	T	CBFB	13.11	16:15703172-15857011
NCOA2	T	KAT6A, HEY1, N	13.3	8:70109762-70403805
NUP98	T	HOXA9, NSD1,	15.4	11:3675010-3797792
PRRX1	T	NUP98	24.2	1:170663906-170739419
PSIP1	T	NUP98	22.3	9:15464066-15511019
RAD21	Mis, N, F		24.11	8:116845935-116874866
RPL22	T	RUNX1	36.31	1:6185020-6199612
RUNX1	T	RPL22, MECOM	22.12	21:34787801-35049344
SEPTIN5	T	KMT2A	11.21	22:19714464-19723319
SEPTIN6	T	KMT2A	24	X:119617639-119693370
SEPTIN9	T	KMT2A	25.3	17:77281410-77500592
SMC1A	Mis		11.22	X:53374149-53422728
BCORL1	Mis, N, F		26.1	X:130005188-130058083
CREBBP	T, N, F, Mis, C	KMT2A, MOREL	13.3	16:3725054-3880726
CTNNA1	D, Mis, N, T	RAF1	31.2	5:138753423-138935034
DNMT3A	Mis, F, N, S		23.3	2:25227855-25342590
HOXA13	T	NUP98	15.2	7:27194364-27200106
HOXC11	T	NUP98	13.13	12:53973126-53977643
HOXC13	T	NUP98	13.13	12:53938765-53946544
HOXD11	T	NUP98	31.1	2:176107286-176109588
HOXD13	T	NUP98	31.1	2:176092891-176095938
NUP214	T	DEK, SET, ABL1	34.13	9:131125561-131234670
TRIP11	T	PDGFRB	32.12	14:91965991-92040059
CBFA2T3	T	RUNX1, GLIS2	24.3	16:88874858-88977204
RUNX1T1	T	KMT2A, RUNX1	21.3	8:91954967-92095654
ARHGAP26	T, F, S	KMT2A	31.3	5:142770377-143229011
ARHGEF12	T	KMT2A	23.3	11:120337078-120489936

Total of 65 somatic genes have been extracted from COSMIC database.

These somatic genes occur spontaneously, and do not reside in the germ cells. They are not inherited but acquired due to epigenetic changes or errors made during replication. They are usually dominant as they require only one copy of gene to express.

Table 11: Cosmic Germline Genes

Gene Symbol	Mutation Types	Chr Band	Genome Location
SBDS	F, N, O	11.21	7:66987677-66995601
BRIP1	F, N, Mis	23.2	17:61681266-61863521
FANCA	D, Mis, N, F, S	24.3	16:89737549-89816646
FANCC	D, Mis, N, F, S	22.32	9:95099054-95317702
FANCE	N, F, S	21.31	6:35452361-35467103
FANCF	N, F	14.3	11:22622519-22626787
FANCG	Mis, N, F, S	13.3	9:35073838-35080016
FANCD2	D, Mis, N, F	25.3	3:10026414-10099660

They occur in germlines where cells reproduce and are inherited. They have recessive gene expression, where two copies of genes are needed to express. If only one copy of gene is expressed the subject becomes a carrier. They do not possess translocation genes as they usually involve point mutations, insertions, and deletions.

Table 12: cBioPortal Genes

Genes	
MYC	ASMTL
CHD2	LRP1B
EP400	KDM6A
NOTCH2	SETD2
MYH11	MKI67
FLT3	U2AF1
IDH2	IKZF1
NPM1	SMARCB1
NRAS	DOT1L
ETV6	ABL1
BCL7A	RIT1
NFE2	WT1
RAC2	DAZAP1
RAD17	ZBTB7A
HLA-C	PTPRT
RAD21	STAG1
PTPN11	BRD4
EP300	TOP1
ZNF384	PPFIBP1
EZH2	PPM1D
PTPN6	TYK2
MEF2C	MLLT10
CXCR4	SMC3
EPHA5	KIT
BCOR	GATA2
SF3B1	GATA1
CCND2	JAK2
ERBB2	JAK3
U2AF2	LRIG3
FANCE	DDX10
PRKD1	KRAS
ERF	RARA
TET2	KMT2D
ERCC3	CSF3R

Gene list taken from cBioPortal.

Table 13: Common Genes and Mutations-1

Gene Symbol	Genome Location	Chr Band	Molecular Genetics	Role in Cancer	Mutation Types	Translocation Partner
CBL	11:119206276-119308149	23.3	Dom/Rec	oncogene,	T, Mis, S, O	KMT2A
FANCE	6:35452361-35467103	21.31	Rec	TSG	N, F, S	
IKZF1						
DNMT3A	2:25227855-25342590	23.3	Dom	TSG	Mis, F, N, S	
MECOM	3:169083499-169146305	26.2	Dom	oncogene,	T	RUNX1, ETV6, PRDM16, RPN1
EZH2						
FLT3	13:28003274-28100592	12.2	Dom	oncogene	Mis, O	
GATA2						
MYH11	16:15703172-15857011	13.11	Dom	fusion	T	CBFB
MNX1	7:157004853-157010651	36.3	Dom	fusion	T	ETV6
JAK2	9:4985245-5128183	24.1	Dom	oncogene,	T, Mis, O	ETV6, PCM1, BCR
IDH2						
NPM1						
NSD1	5:177133925-177300215	35.3	Dom	fusion	T	NUP98
KIT						

Table 14: Common Genes and Mutations-2

KRAS						
TET1	10:68560656-68694482	21.3	Dom	oncogene, T		KMT2A
TOP1	20:41028818-41124487	12	Dom	fusion	T	NUP98
DDX10	11:108665025-108940930	22.3	Dom	TSG, fusion	T	NUP98
NRAS						
						MLLT1, AFF1, MLLT3, MLLT4, FOXO4, MLLT10, MLLT6, ELL, EPS15, MLLT11, CREBBP, SH3GL1, FNBP1, SEPT5, SEPT9, GPHN, GMP5, ABI1, ARHGEF12, GAS7, FOXO3, AFF3, TET1, SEPT6, LPP, RUNX1T1, ARHGAP2
KMT2A	11:118436490-118526832	23.3	Dom	oncogene, T, O		
						HOXA9, NSD1, WHSC1L1, DDX10, TOP1, HOXD13, PRRX1, HOXA13, HOXD11, HOXA11, RAP1GDS1, HOXC11
NUP98	11:3675010-3797792	15.4	Dom	oncogene, T		
MYH1						
PTPN11						
PSIP1	9:15464066-15511019	22.3	Dom	oncogene, T		NUP98
RAD21	8:116845935-116874866	24.11	Rec	oncogene, Mis, N, F		
						RPL22, MECOM, CBFA2T3, RUNX1T1, ETV6, AFF3
RUNX1	21:34787801-35049344	22.12	Dom	oncogene, T		
WT1						
CXCR4						
USP42						
						KMT2A, MORF, KAT6A
CREBBP	16:3725054-3880726	13.3	Dom/Rec	oncogene, T, N, F, Mis		

CHAPTER 6

SUMMARY AND CONCLUSIONS

Developmental disease genomics is a rapidly growing field of study that focuses on understanding the genetic causes of developmental disorders and congenital diseases. This field involves analyzing genetic changes that occur in an individual's DNA and how these changes contribute to the development of specific diseases. One example of this is pediatric acute myeloid leukemia (AML), which is a type of blood cancer that affects children. In developmental disease genomics, researchers analyze the genetic changes that occur in the DNA of myeloid cells and how these changes contribute to the development of leukemia. Some of the most studied genes in pediatric AML include FLT3, NPM1, CEBPA, IDH1/2, and DNMT3A.

Understanding the genetic basis of pediatric AML is crucial for the development of new treatments and improving outcomes for children with this disease. Recurrent genetic mutations are often found in AML, and they are used to predict the prognoses of AML patients. The bone marrow microenvironment also plays a crucial role in the development and progression of AML. Overall, developmental disease genomics provides important insights into the genetic causes of developmental disorders and congenital diseases, which can inform the development of new treatments and improve patient outcomes.

The main goal of this project is to identify new genetic and epigenetic changes involved in the development of pediatric acute myeloid leukemia. AML is a heterogeneous disease with various genetic and epigenetic alterations that drive disease initiation and progression. Therefore, identifying new genetic and epigenetic changes can improve our understanding of the mechanisms of the disease and help develop more effective targeted therapies. The scope of this project is to investigate gene clusters related to pediatric acute myeloid leukemia by analyzing various gene databases using bioinformatics tools and techniques.

The study was done using multiple gene data sets from different databases and analysing them to get a wholistic view on the development of the disease by looking at Protein-protein Interactions (PPI), Gene Regulatory Networks (GRN), TF regulation, networks, mutations, translocations, fusion genes, mutation types, proteins involved, pathway analysis and observing the structural variants.

In conclusion, fusion genes play a critical role in the diagnosis and treatment of pediatric acute myeloid leukemia. Different fusion genes are associated with different subtypes of pediatric AML, and their presence can provide valuable information about the prognosis and treatment options for affected patients. The RUNX1-RUNX1T1 fusion gene is relatively common in pediatric AML and is generally associated with a better prognosis. In contrast, the CBFB-MYH11 fusion gene is associated with a higher risk of relapse, while MLL-rearranged AML is generally considered to have a poor prognosis and requires intensive chemotherapy. As our understanding of the genetics of pediatric AML continues to grow, we can hope to develop more targeted and effective treatments for this challenging disease. The ongoing research in this field highlights the importance of fusion genes in improving our understanding and management of pediatric AML.

Due to vast data, better analysis couldn't be done. However, it remains a challenge to understand and find out the exact cause of pediatric AML due to heterogeneity and its complex interactions.

We holistically investigated mutations of the genes and their frequency levels, the structural (deletions, insertions, duplications) and genetic variations (chromosomal aberrations) and concluded that the NRAS gene has the highest mutation frequency along with FLT3 AND KIT genes associated with RTK-RAS pathway.

NRAS gene showed the highest protein coding expression among other subtypes of leukemia. miRNA, their target sites, and their efficacy were analysed for the genes mentioned above.

Enrichment analysis of the non-coding RNAs showed that these genes code for proteins involved in the regulation, dis-regulation, and proliferation involved in the development and progression of pediatric AML. miRNA and lncRNAs associated with the disease were identified.

Overall statistical data of the genes involved and their associated functions in the development of the disease were analysed. Morphological and genotypical variations caused due to mutations, translocation & fusion of genes were analysed.

To summarize:

- The PPI network visualized in string db had an enrichment value of greater than 1.0×10^{-16} , indicating that proteins interact more and are biologically connected as a group.
- The pathway associated with the network had a false discovery rate of 6.64×10^{-9} and a strength of 1.62.
- The Hubbs involved in the network showed the FLT3 gene had more interconnections with a degree score of 28.
- COSMIC mutation frequency showed NPM1 and FLT3 genes having the highest mutation frequencies.
- NRAS, FLT3 and KIT had greater mutation frequencies.
- NRAS has 16 missense driver mutations. G12D is the protein change.
- FLT3 6 missense mutations and 4 VUS mutation. D835Y is a protein change.
- KIT 8 missense, one in-frame deletion and 3 VUS. V560D is a major protein change.
- RTK-RAS is the associated pathway and has a score of 35.
- The mean expression of the NRAS gene in protein-coding – 6.81.
- miR-152-3p is the conserved region for the NRAS gene.
- miR-16-5p is a conserved region for FLT3, and miR-218-5p is a conserved region for KIT.

- KIT gene has 25 miRNA and 87lncRNA involved in the pathway of AML,
- FLT3 gene has two miRNA and 15lncRNA involved in the pathway of AML.
- NRAS gene has 76 miRNA and 57 lncRNA involved in the pathway of AML.
- miR-122-5p is the strong miRNA involved in AML.
- NEAT1 is the strong lncRNA.
- rs121913254 Q>K is the SNP involved in mutation for the NRAS gene.

REFERENCES

- Bateman, A., Martin, M., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Zhang, J. (2022). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1), D523–D531. <https://doi.org/10.1093/nar/gkac1052>
- Bolouri, H., Farrar, J. E., Triche, T. J., Ries, R. E., Lim, E. L., Alonzo, T. A., Meshinchi, S. (2018). The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nature Medicine*, 24(1), 103–112. <https://doi.org/10.1038/nm.4439>
- Cariaso, M., & Lennon, G. (2012). SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Research*, 40(D1), D1308–D1312. <https://doi.org/10.1093/nar/gkr798>
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B., Sumer, S. O., Aksoy, B. A., . . . Schultz, N. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery*, 2(5), 401–404. <https://doi.org/10.1158/2159-8290.cd-12-0095>
- Chin, C. Y., Chen, S., Wu, H., Ho, C., Ko, M., & Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, 8(S4). <https://doi.org/10.1186/1752-0509-8-s4-s11>
- Conneely, S. E., & Rau, R. E. (2020). The genomics of acute myeloid leukemia in children. *Cancer and Metastasis Reviews*, 39(1), 189–209. <https://doi.org/10.1007/s10555-020-09846-1>

Creutzig U, Zimmermann M, Reinhardt D, Rasche M, von Neuhoff C, Alpermann T, Dworzak M, Perglerová K, Zemanova Z, Tchinda J, Bradtke J, Thiede C, & Haferlach C (2016). Changes in cytogenetics and molecular genetics in acute myeloid leukemia from childhood to adult age groups. *Cancer*, 122(24), 3821–3830. PMID: 27529519 DOI: 10.1002/cncr.30220

Cunningham, F., Allen, J. F., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Flicek, P. (2021). Ensembl 2022. *Nucleic Acids Research*, 50(D1), D988–D995. <https://doi.org/10.1093/nar/gkab1049>

Elgarten C.W., Aplenc R. Pediatric acute myeloid leukemia: Updates on biology, risk stratification, and therapy. *Curr. Opin. Pediatr.* 2020;32:57–66. doi: 10.1097/MOP.0000000000000855.

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Schultz, N. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling*, 6(269). <https://doi.org/10.1126/scisignal.2004088>

Ghosn, E., Yoshimoto, M., Nakauchi, H., Weissman, I. L., & Herzenberg, L. A. (2019). Hematopoietic stem cell-independent hematopoiesis and the origins of innate-like B lymphocytes. *Development*, 146(15). <https://doi.org/10.1242/dev.170571>

Ghukasyan, L. G., Krasnov, G. S., Muravenko, O. V., Baidun, L. V., Ibragimova, S. Z., & Nasedkina, T. V. (2019). Mutation profile of subtypes of acute myeloid leukemia in children in the absence of known chromosomal rearrangements. *Molecular biology*. <https://doi.org/10.1134/s002689841903008x>

Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27–30. <https://doi.org/10.1093/nar/28.1.27>

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2022). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*, 51(D1), D587–D592. <https://doi.org/10.1093/nar/gkac963>

Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. W., . . . Wilson, R. K. (2010). DNMT3A Mutations in Acute Myeloid Leukemia. *The New England Journal of Medicine*, 363(25), 2424–2433. <https://doi.org/10.1056/nejmoa1005143>

Li, Z., Zhang, Y., Fang, J., Xu, Z., Zhang, H., Mao, M., . . . Pian, C. (2022). NcPath: a novel platform for visualization and enrichment analysis of human non-coding RNA and KEGG signaling pathways. *Bioinformatics*, 39(1). <https://doi.org/10.1093/bioinformatics/btac812>

Libura, M., Pawelczyk, M., Florek, I., Matiakowska, K., Jaźwiec, B., Borg, K., . . . Haus, O. (2015). CEBPA copy number variations in normal karyotype acute myeloid leukemia: Possible role of breakpoint-associated microhomology and chromatin status in CEBPA mutagenesis. *Blood Cells Molecules and Diseases*, 55(4), 284–292. <https://doi.org/10.1016/j.bcmd.2015.07.002>

Liu, C., Hu, F., Xia, M., Han, L., Zhang, Q., & Guo, A. (2018). GSCALite: a web server for gene set cancer analysis. *Bioinformatics*, 34(21), 3771–3772. <https://doi.org/10.1093/bioinformatics/bty411>

Liu, C., Hu, F., Xie, G., Miao, Y., Li, X., Zeng, Y., & Guo, A. (2022). GSCA: an integrated platform for gene set cancer analysis at genomic, pharmacogenomic and immunogenomic levels. *Briefings in Bioinformatics*, 24(1). <https://doi.org/10.1093/bib/bbac558>

Liu, Y., Luo, M., Jin, Z., Zhao, M., & Qu, H. (2018). dbLGL: an online leukemia gene and literature database for the retrospective comparison of adult and childhood leukemia genetics with literature evidence. *Database*, 2018.

<https://doi.org/10.1093/database/bay062>

Luo, M., Miao, Y., Ke, Y., Guo, A., & Zhang, Q. (2023). A comprehensive landscape of transcription profiles and data resource for human leukemia. *Blood Advances*. <https://doi.org/10.1182/bloodadvances.2022008410>

Malard, F., and Mohty, M. (2020). Acute lymphoblastic leukemia. *Lancet* 395, 1146–1162. doi: 10.1016/S0140-6736(19)33018-1

Mardis, E. R., Ding, L., Dooling, D. J., Larson, D. W., McLellan, M. D., Chen, K., Ley, T. J. (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. *The New England Journal of Medicine*.

<https://doi.org/10.1056/nejmoa0903840>

McGeary, S. E., Lin, K. J., Shi, C. Y., Pham, T., Bisaria, N., Kelley, G. M., & Bartel, D. P. (2019). The biochemical basis of microRNA targeting efficacy. *Science*, 366(6472).

<https://doi.org/10.1126/science.aav1741>

McHale, C. M., Wiemels, J. L., Zhang, L., Ma, X., Buffler, P. A., Feusner, J. H., Smith, M. T. (2003). Prenatal origin of childhood acute myeloid leukemias harboring chromosomal rearrangements t(15;17) and inv(16). *Blood*, 101(11), 4640–4641.

<https://doi.org/10.1182/blood-2003-01-0313>

Meshinchi, S., Alonzo, T. A., Stirewalt, D. L., Zwaan, M., Zimmerman, M. B., Reinhardt, D., Radich, J. P. (2006). Clinical implications of FLT3 mutations in pediatric AML. *Blood*, 108(12), 3654–3661. <https://doi.org/10.1182/blood-2006-03-009233>

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V. I., Paschka, P., Roberts, N. D., *et al.*, (2016). Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* 374, 2209–2221. doi: 10.1056/NEJMoa1516192.

Quessada, J., Cuccuini, W., Saultier, P., Loosveld, M., Harrison, C. J., & Lafage-Pochitaloff, M. (2021). Cytogenetics of Pediatric Acute Myeloid Leukemia: A Review of the Current Knowledge. *Genes*, 12(6), 924. <https://doi.org/10.3390/genes12060924>

Raudvere, U., Kolberg, L., Kuzmin, I. V., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, 47(W1), W191–W198.

<https://doi.org/10.1093/nar/gkz369>

Rice, S., and Roy, A. (2020). MLL-rearranged infant leukemia: a ‘ thorn in the side ’ of a remarkable success story. *Biochim. Biophys. Acta Gene Regul. Mech.* 1863:194564. doi: 10.1016/j.bbagr.2020.194564

Reinhardt, D., Antoniou, E., & Waack, K. (2022). Pediatric Acute Myeloid Leukemia—Past, Present, and Future. *Journal of Clinical Medicine*, 11(3), 504. <https://doi.org/10.3390/jcm11030504>

Schuback, H. L., Arceci, R. J., & Meshinchi, S. (2013). Somatic Characterization of Pediatric Acute Myeloid Leukemia Using Next-Generation Sequencing. *Seminars in Hematology*, 50(4), 325–332. <https://doi.org/10.1053/j.seminhematol.2013.09.003>

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. M., Ramage, D., Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11), 2498–2504.

<https://doi.org/10.1101/gr.1239303>

Shiba, N., Yoshida, K., Hara, Y., Yamato, G., Shiraishi, Y., Matsuo, H., Hayashi, Y. (2019). Transcriptome analysis offers a comprehensive illustration of the genetic background of pediatric acute myeloid leukemia. *Blood Advances*, 3(20), 3157–3169. <https://doi.org/10.1182/bloodadvances.2019000404>

Shlien A, Malkin D. Copy number variations and cancer. *Genome Med.* 2009;1:62.
Speck NA, & Gilliland DG (2002). Core-binding factors in haematopoiesis and leukemia. *Nature Reviews. Cancer*, 2(7), 502–513. PMID: 12094236 DOI: 10.1038/nrc840

Steliarova-Foucher, E., Colombet, M., Ries, L. A. G., Moreno, F., Dolya, A., Bray, F, *et al.*, (2017). International incidence of childhood cancer, 2001–10: a population-based registry study. *Lancet Oncol.* 18, 719–731. doi: 10.1016/S1470-2045(17)30186-9.

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K. C., Mehryary, F., Hachilif, R., Von Mering, C. (2022). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1), D638–D646. <https://doi.org/10.1093/nar/gkac1000>

Wiemels, J. L., Xiao, Z., Buffler, P. A., Maia, A. L., Ma, X., Dicks, B., . . . Greaves, M. (2002). In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood*, 99(10), 3801–3805. <https://doi.org/10.1182/blood.v99.10.3801>.

Wouters, B. J., Löwenberg, B., Erpelinck-Verschueren, C. A., Van Putten, W. L., Valk, P. J. M., & Delwel, R. (2009). Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood*, 113(13), 3088–3091. <https://doi.org/10.1182/blood-2008-09-179895>.