Investigation of genetic variations regulating cell signaling pathways in leukemia

By

Rehana Yasmin

Department of Zoology

Faculty of Biological Sciences

Quaid-i-Azam University, Islamabad, Pakistan

2023

Investigation of genetic variations regulating cell signaling pathways in leukemia

PhD Dissertation

A dissertation submitted in partial fulfilment of requirements for degree of Doctor of Philosophy inHuman Genetics

Supervised by: **Professor Dr. Sajid Malik**

Conducted by: **Rehana Yasmin**

Department of Zoology

Faculty of Biological Sciences

Quaid-i-Azam University, Islamabad, Pakistan

2023

Investigation of genetic variations regulating cell signaling pathways in leukemia

Thesis Submitted by

Rehana Yasmin

Registration No. 03021611003

to

Department of Zoology,

In partial fulfillment of the requirements for degree of

Doctor of Philosophy

in

Zoology (Human Genetics)

Department of Zoology

Faculty of Biological Sciences

Quaid-i-Azam University

Islamabad, Pakistan

2023

DECLARATION

I hereby state that my PhD thesis titled "**Investigation of genetic variations regulating cell signaling pathways in leukemia**" submitted to the Department of Zoology, Faculty of Biological Sciences, Quid-i-Azam University Islamabad, Pakistan for the award of degree of Doctor of Philosophy in Zoology (Human Genetics) is the result of research work carried out by me. I further declare that the results presented in this thesis have not been submitted for award of my other degree from this university or anywhere else in the country/world and the university has right to withdraw my PhD degree, if any statement is found incorrect anytime even after my graduation.

Rehana Yasmin

30th March 2023

Certificate of Approval

To Whom It May Concern

This is to certify that the research work presented in this thesis, entitled " **Investigation of genetic variations regulating cell signaling pathways in leukemia"** was conducted by Ms. Rehana Yasmin under the supervision of Dr. Sajid Malik. No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the Department of Zoology, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Field of Human Genetics, Department of Zoology, Quaid-i-Azam University, Islamabad, Pakistan.

Student Name: Rehana Yasmin Signature: _____

Examination Committee:

a) External Examiner 1:
Name Signature:
(Designation & Office Address)

……..

……..

b) External Examiner 2: Name Signature:_____
(Designation & Office Address)

……..

……..

c) Internal Examiner: Name Signature:_____
(Designation & Office Address)

……..

……..

Supervisor Name: Prof Dr. Sajid Malik Signature:

Name of Dean: Prof. Dr. Sarwat Jahan Signature:

OFFICE OF THE CONTROLLER OF EXAMINATIONS

NOTIFICATION

No_____ Date:

_____

It is notified for the information of all concerned that Ms. Rehana Yasmin, Ph.D. Scholar of Department of Zoology of Quaid-i-Azam University, Islamabad, Pakistan, has completed all the requirements for the award of Ph.D. degree in the discipline Human Genetics as per detail given here:

PhD in Education			Cumulative Result			
Registration No.	Scholar's Name	Father's Name	Credit Hours			Cumulative Grade Point Average CGPA
			Course Work	Research Work	Total	
03021611003	Rehana Yasmin	Muhammad Rafiq				

Research Topic: "Investigation of genetic variations regulating cell signaling pathways in leukemia"

Local Supervisor Name: Prof Dr. Sajid Malik

Foreign/External Examiners:

a) Name:

University:

Address:

b) Name:

University:

Address:

Detail of Research Articles Published based on thesis research work:

1. Paper 1. Yasmin R, Abbasi R, Saeed T, Sadiq M, Yasmeen N, Iqbal M, Alzahrani AK, Kizilbash N, Ugur B, Ahmad N, Malik S. Epidemiological and Clinical Correlates of Leukemia Ascertained in a Multiethnic Cohort of Pakistan. Pakistan J. Zool., pp 1-11, 2023. DOI: https://dx.doi.org/10.17582/journal.pjz/20221215141236

Note: This result is declaration as notice only. Errors and omissions, if any, aresubject to subsequent rectification.

Signed by Controller of Examinations

C
C:
1.
2.

Plagiarism Undertaking

I solemnly declare that the research work presented in the thesis titled **"Investigation of genetic variations regulating cell signaling pathways in leukemia"** is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and University Quaid-i-Azam Islamabad, Pakistan towards plagiarism. Therefore, I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of Ph.D. degree, the University reserves the rights to withdraw/revoke my PhD degree and that HEC and the University have the right to publish my name on the HEC/University Website on which names of students are placed who submitted plagiarized thesis.

Student /Author Signature: _____

Name: **Rehana Yasmin**

Dedicated
to
my beloved parents
and
my family

Acknowledgment

In the name of Allah who is the most Beneficent and the most Merciful. All praises to Almighty Allah, the creator of the universe. I bear witness that the Holy Prophet Muhammad (SAW) is the messenger, whose life is a perfect model for the whole of mankind till the Day of Judgment. Allah enabled me to complete my work. Without the blessings of Allah, I would not had been able to complete this work and be at such a place.

I am especially indebted to my honorable supervisors, **Prof. Dr. Sajid Malik** (Quaid-I-Azam University, Islamabad) and **Dr. Nafees Ahmad** (Principal Scientist, IBGE, Islamabad) for giving me the initiative to undertake this study. I am also thankful to **Dr. Rashda Abbasi** (IBGE, Islamabad) for her strong presence and aid throughout this journey.

I pay my thanks to whole of the faculty of my department, especially the respected teachers, whose valuable knowledge, assistance, cooperation and guidance enabled me to take initiative, develop and furnish my academic career. I pay thanks to my lab fellows, **Ms. Anisa Bibi**, **Ms**. **Farwah Naqvi, Ms. Filza Khurshid**, **Ms. Ayesha Saifullah**, and **Ms. Zainab**, for keeping me active**.**

I especially acknowledge the prayers, endless love, support, and efforts of my parents, family, and friends without which I was unable to reach at this stage. I think this acknowledgment would be incomplete, if I would not express my deepest gratitude to my friends, **Hina, Faiza Fayyaz**, **Saba Mehreen**, **Mehwish**, **Hizran Bibi,** and **Sorat-Un-Nisa** who were always there to support, encourage and appreciate me. I am also grateful to my siblings, **Nisar Jahangir Akhtar**, **Izhar Jahangir Izhar**, **Iqtedar Jahangir Qaiser**, and **Tajdar Jahangir Gohar,** and my dear sister who kept me in their prayers and supported me at every stage of my life.

In the end, I owe my gratitude to my parents whose countless love, prays, and encouragement, due to which I achieved this academic task.

Rehana Yasmin

List of Tables

Table 1.1: WHO classification of leukemia (Arber *et al.*, 2016; Loghavi *et al.*, 2023) 2
Table 1.2: Major cancer hospitals in Pakistan 9
Table 1.3: List of studies focusing on the epidemiology of leukemia reported from Pakistan and other parts of the world in last decade 11
Table 1.4: Leukemia based genetics and cytogenetics studies reported from Pakistan..... 15
Table 1.5: Genetic studies based on the JAK/STAT/SOCS pathway conducted in Pakistan. 16
Table 1.6: Cytokines signaling through JAK/STAT pathway (Morris *et al.*, 2018) 18
Table 1.7: JAK/STAT /SOCS pathway component genes and their biological significance (retrieved from NCBI databases, UCSC, Genebank, SNPedia, and Ensemble) 21
Table 1.8: Downstream genes contributing to crucial biological processes regulated by JAK/STAT pathway, as studied in different models (Carpenter and Lo, 2014). 26
Table 1.9: Genetic alterations in JAK/STAT pathway reported worldwide..... 33
Table 1.10: Single nucleotide polymorphisms studied in JAK/STATS/SOCS pathway across the globe 36
Table 1.12: Different types of nextgeneration sequencing for genomics (Duncavage and Tandon, 2015; Bewicke-Copley *et al.*, 2019)..... 48
Table 2.1A: List of the chemicals used in the study 54
Table 2.1B: List of the equipment used in the current study 54
Table 2.2: Blood cell status reference ranges 62
Table 2.3: The PCR reaction set up for ARMS-PCR 68
Table 2.4: List of tetra ARMS-PCR primers, optimized Tm and product size 69
Table 2.5: The list of IUPAC names and chemical structure and Structure of the selected ligands for STAT5B interaction 79
Table 3.1: Age group-based genderwise categorization of the study population (n=1478), and type-wise distribution of leukemia patients (n=594). ALL = Acute Lymphocytic Leukemia, CLL= Chronic Lymphocytic Leukemia, AML= Acute Myeloid Leukemia, CML = Chronic Myeloid Leukemia 82
Table 3.2: Hematological attributes of the patients (n=616) and controls (n=884)..... 90

Table 3.3A: Overall subtype-wise (mean ± standard deviation) of patients and controls hematological variables 92
Table 3.3B: Overall age and gender-wise (mean ± standard deviation) of patient's hematological variables (n=594) 92
Table 3.4A: Summary of age group and gender-wise distribution of hematological variables (Patients, 594; controls, 884) 94
Table 3.4B: Overall origin and nature-wise distribution of leukemia patients against age, gender, and hematological variables (n=594) 97
Table 3.5: Odd ratio of hematological variables for patients vs controls 112
Table 3.6: General characteristics, demographic and lifestyle-related risk factors details of the study population (leukemia patients, n=616; control samples, n=884) 115
Table 3.7A: Overall age group and gender-wise distribution of demographic and lifestyle-related risk factors of leukemia patients (n=594), in comparison to controls (n=884). 118
Table 3.7B: Overall origin and nature-wise distribution of leukemia patients against age, gender, and risk factors (n=594) 125
Table 3.8: Odds ratios for demographic and lifestyle related variables for patients 163
Table 3.9: Odds ratios for high-risk patient age groups vs control groups (additive model) 171
Table 3.10: Odds ratios comparing high-risk groups in patients cohort (additive model) ... 175
Table 3.11: Demographic details of leukemia patients included in the study 182
Table.3.12: Mutational status of JAK1^{V623A}, JAK2 ^{S473} and STAT5B^{N642H} in leukemia..... 185
Table 3.13: Mutational status of *JAK1*^{V623A}, *JAK2* ^{S473} and *STAT5B*^{N642H} in leukemia subtypes 187
Table 3.14: Clinical findings and lifestyle-related risk factors of the acute lymphoid leukemia patients harboring STAT5B^{N642H} 190
Table 3.15A: Relationship of maternal to paternal age, at birth for leukemia patients in comparison to matched controls (n=338) 192
Table 3.15B: Comparison of parental age at birth for leukemia patients to matched controls (n=338) 192
Table 3.16: The mutations found in Patient I; 2 years old male child with acute lymphoblastic leukemia, through targeted NGS. 195
Table 3.17: The mutations found in patient II; 2 years old male child with acute lymphoblastic

leukemia, through Targeted NGS 197
Table 3.18: Copy number aberrations spanning LOH positions..... 198
Table 3.19: List of STAT5B inhibitors, associated with STAT5B, in resistant leukemia ... 205

List of Figures

Figure 1.1: Map of Pakistan showing the provinces and the major cancer hospitals situated therein 8
Figure 1.2: Structural details of JAKs and STATs proteins (Seif *et al.*, 2017). 20
Figure 1.3: Schematic illustration of the JAK/STAT pathway working and regulation mechanism (Morris *et al.*, 2018). 25
Figure 1.4: Schematic diagram elaborating role of JAK/STAT Pathway in normal hematopoiesis and hematopoiesis during leukemogenesis (Fasouli and Katsantoni, 2021). 32
Figure 1.5: Schematic illustration elucidating the basic steps of fluorescence in situ hybridization (Wippold and Perry, 2007). 40
Figure 1.6: Graphical representation of the comparative genomic hybridization (Dugoff *et al.*, 2016). 42
Table 1.11: Comparison of technical details between conventional cytogenetic and major molecular approaches (Kang, 2018). 43
Figure 1.7: Illustration of next-generation sequencing and its types, and library preparation for the targeted sequencing (Bewicke-Copley *et al.*, 2019). 47
Figure 2.1: Graphical presentation of the study scheme 57
Figure 2.2: Schematic flow chart of targeted next-generation sequencing 74
Figure 2.3: Schematic flow chart of the performed microarray scanning 76
Figure 3.1: Gender-wise distribution of leukemia patients in different age group (n=594) ... 84
Figure 3.2: Gender-wise distribution of leukemia patients against biological status (n=594) 84
Figure 3.3: Distribution of leukemia subtypes in the Pakistani population 85
Figure 3.4: Overall gender-wise distribution (%) of acute/chronic leukemia patients in different age group (n=594). 86
Figure 3.5: Gender-wise distribution of leukemia patients in different age group (years) and disease origin (n=594) 87
Figure 3.6: Comparison of percentages of leukemia subtypes prevalence in both myeloid and lymphoid (n=594) 88
Figure 3.7: Bar graphical representation of hematological parameters in patients and controls

..... 91
Figure 3.8A: Overall age groupwise percentage distribution of leukemia patients against blood WBCs (n=594)..... 100
Figure 3.8B: Overall age groupwise percentage distribution of controls against blood WBCs (n=884) 100
Figure 3.9A: Overall age groupwise percentage distribution of leukemia patients against blood RBCs (n=594)..... 102
Figure 3.9B: Overall age group-wise percentage distribution of controls against blood RBCs (n=884) 102
Figure 3.10A: Overall age groupwise percentage distribution of leukemia patients against hemoglobin (n=594) 104
Figure 3.10B: Overall age groupwise percentage distribution of controls against hemoglobin (n=884) 104
Figure 3.11A: Overall age groupwise percentage distribution of leukemia patients against platelets (n=594)..... 106
Figure 3.11B: Overall age groupwise percentage distribution of controls against platelets (n=884) 106
Figure 3.12A: Overall age groupwise percentage distribution of leukemia patients against lymphocytes (%) (n=594)..... 108
Figure 3.12B: Overall age groupwise percentage distribution of controls against lymphocytes (%) (n=884). 108
Figure 3.13A: Overall age group-wise distribution of leukemia patients against neutrophils % (n=594) 110
Figure 3.13B: Overall age group-wise distribution of controls against neutrophils (%) (n=884) 110
Figure 3.14: Forest plot depicting significant hematological variables associated with leukemia..... 113
Figure 3.15A: Overall age group-wise percentage distribution of leukemia patients against ethnicity (n=594) 133
Figure 3.15B: Overall age groupwise percentage distribution of controls against ethnicity (n=884) 133

Figure 3.16A: Overall age groupwise percentage distribution of leukemia patients against education (n=594) 135
Figure 3.16B: Overall age groupwise percentage distribution of controls against education (n=884) 135
Figure 3.17A. Overall age groupwise percentage distribution of leukemia patients against parental consanguinity (n=594) 137
Figure 3.17B. Overall age grouwise percentage distribution of controls against parental consanguinity (n=884) 137
Figure 3.18: Comparison of parental consanguinity between leukemia patients and controls 138
Figure 3.19A: Overall age group-wise percentage distribution of leukemia patients against family history (n=594) 139
Figure 3.19B: Overall age groupwise percentage distribution of controls against family history (n=884). 139
Figure 3.20: Comparison of family history of cancer between leukemia patients and controls 140
Figure 3.21A: Overall age groupwise percentage distribution of leukemia patients against body weight (n=594) 142
Figure 3.21B: Overall age groupwise percentage distribution of controls against body weight (n=884) 142
Figure 3.22A: Overall age groupwise percentage distribution of leukemia patients against diet (n=594) 144
Figure 3.22B: Overall age groupwise percentage distribution of controls against diet (n=884) 144
Figure 3.23A: Overall age groupwise percentage distribution of leukemia patients against area (n=594) 146
Figure 3.23B: Overall age groupwise percentage distribution of controls against area (n=884) 146
Figure 3.24A: Overall age groupwise percentage distribution of leukemia patients against water type (n=594) 148
Figure 3.24B: Overall age groupwise percentage distribution of controls against water type

(n=884) 148
Figure 3.25A: Overall age groupwise percentage distribution of leukemia patients against fuel (n=594) 150
Figure 3.25B: Overall age group-wise percentage distribution of controls against fuel (n=884) 151
Figure 3.26A: Overall age groupwise percentage distribution of leukemia patients against family type (n=594) 153
Figure 3.26B: Overall age groupwise percentage distribution of controls against family type (n=884) 153
Figure 3.27A: Overall age groupwise percentage distribution of leukemia patients against tobacco use (n=594) 156
Figure 3.27B: Overall age groupwise percentage distribution of controls against tobacco use (n=884) 156
Figure 3.28A: Overall age groupwise percentage distribution of leukemia patients against microwave use (n=594) 158
Figure 3.28B: Overall age groupwise percentage distribution of controls against microwave use (n=884) 158
Figure 3.29A: Overall age groupwise percentage distribution of leukemia patients against perfume use (n=594) 160
Figure 3.29B: Overall age groupwise percentage distribution of controls against perfume use (n=884) 160
161
Figure 3.30: Comparison of environmental and lifestyle style-related risk factors between patients and controls 161
Figure 3.31: Comparison of ethnicity, marital status, and education between patients and controls 165
Figure.3.32: Comparison of parental consanguinity and family history of cancer between patients and controls 166
Figure 3.33: Comparison of diet and body weight of cancer between leukemia patients and controls 167
Figure 3.34: Odds ratios depicting risk factors 170

Figure 3.35: Odds ratios were calculated to identify age group-related risk factors for patients compared to controls 174
Figure 3.36: Odds ratios of high-risk patients' age group within the patient cohort..... 179
Figure 3.37: Percentage distribution of demographic risk factors of the leukemia patients screened for *JAK1*^{V623A}, *JAK2*^{S473,} and *STAT5B*^{N642H} 183
Figure 3.38: Gel picture showing the bands for allele-specific PCR for STATFB^{N642} 185
Figure 3.39: Percentage distribution of demographic risk factors of the leukemia patients screened for *JAK1*^{V623A}, *JAK2*^{S473,} and *STAT5B*^{N642H} 186
Figure 3.40: Percentage distribution of leukemia subtypes in the studied population screened for *JAK1*^{V623A}, *JAK2*^{S473,} and *STAT5B*^{N642H} in leukemia. 188
Figure 3.41: The graphical interface of the ChAS, showing the positions of copy number variation. 200
Figure 3.42: STAT5B interactome with other proteins regulating several transduction pathways. 205
Figure 3.43: Interaction of AC-4-130 and Stafib-2 with wildtype STAT5B and STAT5B^{N642H+} proteins. 206
Figure 3.44: Interaction of Indirubin, and Pimozide with wildtype STAT5B and STAT5B^{N642H+} proteins. 207

List of Acronyms

S#	Abbreviation	Definition
1	ABL	Abelson Leukemia
2	ADGRG6	Adhesion G protein-Coupled Receptor G6
3	AKT	RAC-Alpha Serine/Threonine-Protein Kinase
4	ALK	Anaplastic Lymphoma Kinase
5	ALL	Acute Lymphocytic Leukemia
6	AMKL	Acute Mega Karyoblastic Leukaemia
7	AML	Acute Myelogenous/Myeloid Leukemia
8	AMP	Adenosine Monophosphate
9	APC	Antigen-Presenting cell
10	APL	Acute Promyelocytic Leukemia
11	ARMS	Amplification Refractory Mutation System
12	AYAS	Adolescent and Young Adults
13	BATF	Basic Leucine Zipper ATF-Like Transcription Factor
14	BCL2	B-Cell Leukemia/Lymphoma 2 Protein
15	BCR	Breakpoint Cluster region
16	C2C12	Murine Myoblast Cell Line
17	CADD	Computer-Aided Drug Designing
18	CDC25A	Cell Division Cycle 25A
19	CEBPA	CCAAT Enhancer Binding Protein Alpha
20	CGH	Comparative Genomic Hybridization
21	CI	Confidence Interval
22	CISH	Cytokine Inducible SH2 Containing Protein
23	CLL	Chronic Lymphocytic Leukemia
24	CML	Chronic Myeloid Leukemia
25	CMML	Chronic Myelomonocytic Leukemia
26	CNS	Central Nervous System
27	COX	Cyclooxygenase

28	CRP	C-reactive protein
29	CSF	Cerebrospinal Fluid
30	CTLA	Cytotoxic T-Lymphocyte-Associated Protein 4
31	DEBS	6-Deoxyerythronolide B Synthase
32	EBF1	Early B-cell factor 1
33	EDTA	Ethylenediaminetetraacetic Acid
34	EGF	Epidermal Growth Factor
35	EGFR	Estimated Glomerular Filtration Rate
36	EGFR-AS1	EGFR Antisense RNA 1
37	EMF	Electromagnetic Fields
38	EPO	Erythropoietin
39	ERB	Effective Renal Blood
40	ERBB4	Erb-b2 Receptor Tyrosine Kinase 4
41	ETV6	Translocation-Ets-Leukemia Virus -6
42	FAB	Fragment Antigen-Binding Region
43	FAM30A	Family With Sequence Similarity 30 Member A
44	FERM	Protein 4.1R Ezrin Radixin Moesin
45	FGF	Fibroblast Growth Factors
46	FGFR3	Fibroblast Growth Factor Receptor 3
47	FISH	Fluorescence In Situ Hybridization
48	FLT3	Fms-Related Tyrosine Kinase 3
49	FOXO1	Forkhead Box Protein O1
50	FRS2	Factor Receptor Substrate 2
51	GOF	Gain-of-Function
52	GWAS	Genome-Wide Association Studies
53	GWR	Geographically Weighted Regression
54	HDAC3	Histone Deacetylase 3
55	HGF	Hepatocyte Growth Factor
56	HPLC	High-Performance Liquid Chromatography
57	HRAS	Harvey Rat Sarcoma Virus
58	HSC	Hematopoietic Stem Cells

59	HSPC	Health and Care Professions Council
60	HUVEC	Human Umbilical Vein Endothelial Cells
61	ICAM	Intercellular Adhesion Molecule 1
62	ICF	Intracellular fluid
63	IDH1	Isocitrate Dehydrogenase 1,
64	IFN	Interferons
65	IGF1R	Insulinlike Growth Factor1 Receptor
66	IKZF1	Ikaros Family Zinc Finger Protein 1
67	IL	Interleukin
68	INF	Interferons
69	IRF4	Interferon Regulatory Factor 4
70	ITD	Intertriginous dermatitis
71	JAKS	Janus Kinases
72	JCDR	Journal of Clinical and Diagnostic Research for doctors
73	JHTD	Journal Healthcare Treatment Development
74	KCR	Kentucky Cancer Registry
75	KDR	Kinase Insert Domain Receptor
76	KMT2A	Lysine Methyltransferase 2A
77	LOH	Loss of Heterozygosity
78	LSD1	Lysine-Specific Demethylase 1
79	MAPK	Mitogen-Activated Protein Kinase
80	MB49	MB49 Mouse Bladder Carcinoma Cell Line-49
81	MCF7	Michigan Cancer Foundation-7
82	MET	Metabolic Equivalent
83	MMP	Matrix Metalloproteinases
84	MPD	Myeloproliferative Disorders
85	MPN	Myeloproliferative Neoplasms
86	MRD	Minimal Residual Disease
87	MRI	Magnetic Resonance Imaging
88	MUC	Metastatic Urothelial Carcinoma
89	MYC	Myelocytomatosis

90	NGAL	Neutrophil Gelatinase-Associated Lipocalin
91	NGS	Next Generation Sequencing
92	NOS	Nitric oxide synthases
93	NOTCH1	Neurogenic Locus Notch Homolog Protein 1
94	NPAS2	Neuronal PAS Domain Protein 2
95	OMIM	Online Mendelian Inheritance in Man
96	PAHs	Polyaromatic Hydrocarbons
97	PC3	Prostate Cancer Cell Line
98	PDB	Protein Data Bank
99	PDGF	Platelet-Derived Growth Factor
100	PDGFRA	Platelet-Derived Growth Factor Receptor A
101	PI3K	Phosphoinositide 3-Kinases
102	PIK3CA	PIK3CA Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
103	PLL	Prolymphocytic Leukemias
104	PLT	Platelet Count
105	PMF	Plasma Master File
106	POMC	Proopiomelanocortin
107	PRL	Prolactin
108	PV	Polycythemia Vera
109	RANTES	Regulated Upon Activation, Normal T Cell Expressed and Secreted
110	RBC	Red Blood Cell
111	RET	Rearranged During Transfection
112	ROR	RAR-Related Orphan Receptors
113	RPMI	Roswell Park Memorial Institute
114	RPS15	Ribosomal Protein S15
115	RUNX1	Runt-Related Transcription Factor 1
116	SAA1	Serum Amyloid A1
117	SCK1	Subclinical Ketosis in First Lactation
118	SF3B1	Splicing Factor 3B Subunit 1

119	SH2	Src-Homology 2
120	SKY	Spectral Karyotyping
121	SLL	Small Lymphocytic Lymphoma
122	SMAD4	SMA- and MAD-Related Protein 4
123	SNP	Single Nucleotide Polymorphism
124	SOCS	Suppressors of Cytokine Signaling
125	SPAG9	Sperm Associated Antigen 9
126	STATS	Signal Transducer and Activator of Transcription
127	STK11	Serine/threonine Kinase 11
128	TARSEQ	Targeted Sequencing
129	TBE	Tris-Borate-EDTA
130	TEL	Translocation E26 Leukaemia
131	TGF	Transforming Growth Factor
132	TIMP	Tissue Inhibitors of Metalloproteinases
133	TKD	Tyrosine Kinase Domain
134	TLR4	Toll-Like Receptor 4
135	TP53	Tumor Protein P53
136	TPO	Thyroid Peroxidase
137	TSLP	Thymic Stromal Lymphopoietin
138	TYK2	Tyrosine Kinase 2
139	U87MG	Uppsala 87 Malignant Glioma
140	VEGF	Vascular Endothelial Growth Factors
141	VSM	Vascular Smooth Muscle
142	WAF1	Wild-Type Activating Fragment-1
143	WBC	White Blood Cells
144	WES	Whole Exome Sequencing
145	WGS	Whole Genome Sequencing
146	WNT	Wingless-Related Integration Site

Table of Contents

By 1
Supervised by: 2
Conducted by: 2
To Whom It May Concern 5
Research Topic: “Investigation of genetic variations regulating cell signaling pathways in leukemia” 6
Signed by Controller of Examinations 7
List of Tables i
List of Figures iv
List of Acronyms ix
Abstract xx
CHAPTER 1
1 Leukemia 1
1.1 Types and classification 1
1.1.1 Acute leukemia 3
1.1.1.1 Acute lymphoblastic leukemia 3
1.1.1.2 Acute myeloid leukemia 4
1.1.2 Chronic leukemia 4
1.1.2.1 Chronic lymphocytic leukemia 5
1.1.2.2 Chronic myelogenous leukemia (CML) 5
1.2 Signs and symptoms 6
1.3 Leukemia epidemiology 6
1.3.1 Status in Pakistan 8
1.4 Risk factors 13
1.5 Genetics 13
1.5.1 Role of signaling pathways 17
1.5.2 JAK-STAT pathway 17
1.5.2.1 Components of JAK/STAT pathway 19
1.5.2.2 JAK/STAT mechanism of action 23
1.5.2.3 Role of JAK/STAT pathway in hematopoiesis 29

1.5.2.4 JAK-STAT in hematologic malignancies..... 30
1.6 Diagnosis..... 37
(i). Physical examination: 37
(ii). Complete blood count (CBC): 37
(iii). Blood cell examination: 37
(iv). Bone marrow biopsy/ aspiration: 37
(v). Cytogenetics testing: 37
(vi). Immunophenotyping: 37
(vii). Imaging and other tests: 38
1.6.1 Modern cytogenetics techniques 38
(i). Fluorescence in situ hybridization (FISH) 39
(ii). Comparative genomic hybridization 41
(iii). Array comparative genomic hybridization (aCGH) 41
1.6.2 Modern genetics techniques 44
(i) Whole genome sequencing (WGS) 44
(ii) Whole exome sequencing 45
(iii) Targeted genomic sequencing (TS) 45
1.7 Treatment 49
(i). Chemotherapy: 49
(a) Induction: 50
(b) Consolidation/intensification: 50
(c) Maintenance therapy 50
(ii). Radiation therapy: 50
(iii). Immunotherapy/ biologic therapy: 50
(iv) Hematopoietic cell/ 51
(v) Targeted therapy: 51
1.7.1 In silico molecular docking 51
Aims 53
Objectives: 53
CHAPTER 2
2. Methodology 54

2.1 Materials 54
2.2 Methods 55
2.2.1. Study population 55
2.2.2. Data collection and sampling 56
2.2.3. Study scheme 57
2.2.3.1 Epidemiological study 58
2.2.3.1.1 Sample and study parameters 58
(i). Age group and gender: 58
(ii). Ethnicity: 58
(iii). Education: 59
(iv). Diet: 59
(v). Weight: 59
(vi). Parental consanguinity: 60
(vii). Family history: 60
(viii). Marital status: 60
(ix). Area: 60
(xi). Environmental status: 61
(xii). Water type: 61
(xiii). Domestic fuel type: 61
(xiv). Family type: 61
(xv). Carbonated drinks: 61
(xvi). Tobacco use: 62
(xvii). Perfume use: 62
(xviii). Microwave use 62
(xix). Hematological variables 62
2.2.3.2 Molecular genetic study 64
2.2.3.2.1 Genomic DNA extraction 64
(i) Manual DNA extraction 65
(ii) Genomic DNA extraction via kit 66
2.2.3.2.2 DNA quantification and dilution 67
2.2.3.2.3 PCR amplifications 67

2.2.3.2.3.1 SNP typing (ARMS-PCR) 67
(i). Agarose gel electrophoresis for PCR-amplified DNA 70
(ii). PCR product purification 70
(iii). Purification of the sequencing reaction 71
2.2.3.2.3.2 Targeted next-generation sequencing (Tar-Seq) 71
(i). Amplification of Targets 71
(ii). Partial digestion of the amplicons 72
(iii). Adaptors ligation 72
(iv). Library purification 73
(v). Quantify the unamplified library by qPCR 73
2.2.3.2.3.3 Microarray assay 75
2.2.3.3 In silico drug-protein interaction analysis 77
2.2.3.3.1 Ligand molecules preparation 77
2.2.3.3.2 Protein macromolecules preparation 77
2.2.3.3.3 Ligand-protein docking 78
2.2.4 Statistical analysis 80

CHAPTER 3

3. Results 81
3.1 Epidemiological analysis 81
3.1.1 Clinical and paraclinical variables 81
3.1.1.1 Sampling characteristics and overall percentage distribution of study population 81
3.1.1.2 Age and genderwise distribution 83
3.1.1.3 Leukemia subtypes 85
(i). Disease nature-wise distribution of leukemia patients 86
(i). Disease origin-wise distribution of leukemia patients 87
(iii). Comparison of leukemia subtypes prevalence in both myeloid and lymphoid leukemia 88
3.1.1.4 Paraclinical and hematological variables 89
3.1.1.4.1 Age group and gender-wise distribution of hematological variables 93
3.1.1.4.1.1 White blood cells 99

3.1.1.4.1.2 Red blood cells 101
3.1.1.4.1.3 Hemoglobin status 103
3.1.1.4.1.4 Platelets status 105
3.1.1.4.1.5 Lymphocytes (%) status 107
3.1.1.4.1.6 Neutrophils (%) status 109
3.1.1.4.2 Association of hematological variables to leukemia 111
3.1.2 Demographic and lifestyle-related variables 114
3.1.2.1 Age and gender-wise distribution of demographic and lifestyle among patients and controls 132
3.1.2.1.1 Ethnicity 132
3.1.2.1.2 Education 134
3.1.2.1.3 Consanguinity 136
3.1.2. 1.4 Family history 138
3.1.2. 1.5 Weight status 141
3.1.2. 1.6 Diet 143
3.1.2.1.7 Area 145
3.1.2.1.8 Water type 147
3.1.2.1.9 Fuel 149
3.1.2.1.10 Family type 152
3.1.2.1.11 Carbonated drinks 154
3.1.2.1.12 Tobacco use 155
3.1.2.1.1.13 Microwave use 157
3.1.2.1.14 Perfume use 159
3.1.2.2 Association of demographical and lifestyle factors with leukemia risk 162
3.1.2.2.1 Age, Gender, ethnicity, education, and leukemia risk 165
3.1.2.2.2 Consanguinity and family history as a risk factor for leukemia 166
3.1.2.2.3 Diet, weight, and leukemia risk 167
3.1.2.2.4 Lifestyle and residential risk factors associated with leukemia 168
3.1.2.2.5 Factors reducing leukemia risk 169
3.2 Genetic analysis 180
3.2.1 Mutation Screening through ARMS 181

3.2.1.1 Results 181
3.2.1.1.1 Samples characteristics 181
3.2.1.1.2 Identification of cases with STAT5B mutations 184
3. 2.1.1.3 Clinical features associated with STAT5B^{N642H} 188
3. 2.1.1.4 Association of parental ages with leukemia 191
3.2.2 Mutation screening through targeted NGS (TarSeq) 194
3.2.2.1 Results 194
3.2.2.1.1 Patient 1 194
3.2.2.1.2 Patient II 196
3.2.3 Mutational screening through microarray assay 198
3.3 In silico protein drug interaction analysis 201
3.3.1 Molecular docking analysis 201
3.3.2 Molecular interaction of the ligands with wildtype STAT5B 201
3.3.3 Molecular interaction of the ligands with STAT5B ^{N642H} 202
CHAPTER 4
4. Discussion 208
4.1 Epidemiology 208
4.2. Genetic analysis 218
4.2.1 Mutational screening through amplification refractory system 218
4.2.2. Mutational screening through targeted NGS approach 221
4.2.3 Cytogenetic analysis through microarray assay 228
4.3 In silico drug-protein interaction 230
Strengths and limitations of the study 233
Conclusion and future recommendations 234
REFERENCES 237
Annexure 1: Consent form 297
Annexure 2. Data collection proforma/questionnaire 299
Annexure 3: List of genes on the cancer panel 302
Annexure 4. Plagerism Report 311
Annexure 5. Article Abstract 312

ABSTRACT

Leukemia is a heterogeneous group of hematological cancer responsible for a multitude of morbidities and mortalities worldwide. It is one of the most prevalent cancers among pediatric malignancies in Pakistan and causes a huge economic burden. Epidemiological integrated transdisciplinary research is needed for the estimation of the actual leukemia burden and contributing risk factors in Pakistan. Leukemia incidence, subtypes distribution, and survival variance point towards the strong involvement of population genetics. There are numerous yet diverse numbers of genomic variations, single nucleotide polymorphisms (SNPs), mutations (point, insertions, deletions), and chromosomal aberrations actively elucidated for a role in triggering and increasing susceptibility of leukemia. Screening of the genetic factors contributing to leukemia in diverse and distinct populations is of particular importance. Genetic factors affect personalized drug decisions, treatment outcomes, and patient survival. Revealing the genomic landscape of individual patients, with modern genetic techniques, can help design personalized treatment plans leading to highly effective treatment protocols. In silico interaction studies, involving docking of compounds having inhibitory potential can be helpful in the development of effective, and specific inhibitors. The present study was designed to investigate the epidemiological, environmental, and lifestyle related risk factors, mutational status of selected *JAK-STAT* pathway genes, and screening of mutational hot spots in key genes of crucial signaling pathways in leukemia patients in the Pakistani population. The present study also included a virtual exploration of the molecular interaction of potential inhibitors with their mutated products. A total of 1500 subjects, including 616 patients and 884 controls were recruited through a consented, retrospective cross-sectional sampling design from tertiary care hospitals in Pakistan. DNA was extracted from blood samples obtained from the study participants. For the risk factors identification, the data of 594 patients and 884 controls were processed. Descriptive summaries were generated, and risk factors were analyzed through logistic regression. For allele-specific PCR, the DNA samples of 276 patients, were processed. The genetic status of $JAK1^{V623A}$, $JAK2^{S473}$, and $STAT5B^{N642H}$ was screened through allele-specific PCR. Targeted next-generation sequencing (NGS) was performed for selected patient's samples by using Thermofisher Ion S5™ Systems, Ion Torrent, using Ion AmpliSeq™ Library Kit 2.0, with Ion Ampliseq Cancer hotspot panel 2. The panel comprised 50 crucial human genes involved in cell signaling pathways. The sequencing results were analyzed through Thermofisher Scientific, based Ion Reporter™ Software and interpreted. Microarray was performed on a patient's DNA sample through Cyto-Scan Cytogenetics Suite by Thermofisher as per manual instructions and the data was analyzed through Thermofisher Chromosome Analysis Suite (ChAS) 4.3 software. In-silico analysis was performed through molecular docking of wildtype and mutant protein retrieved from the protein databank. Selected protein inhibitor structures were retrieved from the PubChem database. The ligands and protein were prepared, and docking was performed through Auto Dock Vina 123 and visualized through Discovery Studio Visualizer 2021. The present study observed that acute lymphoblastic leukemia comprised almost 70% of the total patients, with male preponderance. We identified Pathan ethnicity (OR=2.85; 95%CI=2.29-3.54), no formal education (OR=3.36; 95%CI=2.62-4.32), poor diet (OR=2.34; 95%CI=1.79-3.06), lower BMI (OR=1.95; 95%CI=1.50-2.60), parental consanguinity (OR=2.13; 95%CI=1.67-2.71), positive family history (OR=4.24; 95%CI=2.18-8.26), rural residential setup (OR=2.93; 95%CI=2.10-4.10), drinking of groundwater (OR=2.25; 95%CI=1.6479-3.0964), wooden fuel (OR= 3.97;

95%CI=3.14-5.01), carbonated drinks (OR=1.25, 95%CI=1.00-1.57) and tobacco usage (OR=1.57, 95%CI=1.24-1.98) as significant risk factors for leukemia. However, odds ratios were significantly lower for patients using a microwave oven (OR=0.25; 95%CI=0.18-0.35), and perfumes (OR=0.42; 95%CI=0.33-0.53). Males exhibit an increased risk for lymphoid leukemia as compared to myeloid leukemia (OR=1.97; 95%CI=1.38-2.80). Paraclinical parameters indicated that 71% of the cases had >50% of blast cells. Leukocytosis (OR= 9.06; 95% CI=6.46-12.71), anemia (OR= 15.84; 95% CI=11.84-21.21), low hemoglobin (OR=8.11; 95% CI=6.35-10.37), thrombocytopenia (OR=32.40; 95% CI=21.57-48.68), lymphocytosis (OR= 3.41; 95% CI=2.55-4.57), and neutropenia (OR=7.32; 95% CI=5.59-9.60) had significantly higher odd ratio for leukemia patients. The allele-specific amplification showed that all the patients were homozygous normal for *JAK1*^{V623A}, *JAK2* ^{S473} major allele. However, 6 patients (5 male, 1 female) in the age range of 10-11 years, with ALL were *STAT5B*$^{N642H+}$. The NGS data identified the presence of 26 genetic variations on different genomic loci, including 14 exonic mutations; comprised of 3 missense variations (*TP53;* c.215C>G/ p.Pro72Arg**,** *ALK;* c.3551G>A/ p.Gly1184Glu*, SMAD4;* c.767A>T/ p.Gln256Leu), and 11 synonymous variations (*FGFR3*; c.1953G>A/Thr651Thr, *PDGFRA;* c.2472C>T/p.Val824Val, & c.1701A>G/ Pro567Pro, *RET*; c.2307G>T/ Leu769Leu, *APC;* c.4479G>A/p.Thr1493Thr, *PIK3CA*/p.Thr1025Thr; c.3075C>T, *EGFR;* c.2361G>A/p. Gln787Gln, *EGFR-AS1*; c.2361G>A/p.Gln787Gln, *MET*; c.534C>T/ p.Ser178Ser, ABL1; c.1149C>T/ p.Gly383Gly, *HRAS;* c.81T>C/ p.His27His). The remaining variation included 9 intronic variations (rs3738868, rs5030613, rs7692791, rs10006115, rs3729674, rs2491231, rs2075606, rs839541, INDEL rs1412472143/rs67894136) and a downstream UTR_3 multiple nucleotide variation (rs38669350). The microarray analysis identified the presence of 97 loss of heterozygosity (LOH) loci, 2 copy number loss at chromosome 8 (p11.22, q24.23), and a copy number gain on chromosome 14 (q32.33). The molecular docking of the ligands to wildtype and STAT5B^{N642H+}revealed that AC-4-130, Pimozide, Indirubin, and Stafib-2 have higher, but differential docking affinities for both normal and mutated STAT5B. However, AC-4-130 has a higher affinity for wildtype and Stafib-2 has stable molecular interaction with STAT5B^{N642H+}. Risk factors associated with leukemia are mainly related to exposure due to rural residence, poor lifestyle, and family history of the disease. The disease incidence can be minimized by designing and implementing risk mitigation strategies. The aggressive form of pediatric leukemia, carrying STAT5B^{N642H+} mutation has been identified in the studied population. The molecular docking study predicted that AC-14-30 and stafib-2 are potential ligands for inhibition of constitutively active STAT5B if modified and optimized for use in combination therapy. NGS and microarray are efficient platforms, to perform sensitive, reliable, and simultaneous detection of genomic/genetic aberrations in one setting. technologies are valuable for identifying individual genetic profiles and clustering the patients for targeted therapies.

CHAPTER 1

INTRODUCTION

1 Leukemia

Leukemia is a heterogeneous group of clonal hematologic malignancies characterized by the abnormal proliferation of hematopoietic cells interrupting the normal function of blood and bone marrow. Leukemogenesis is a composite process that manipulates a hematopoietic progenitor cell's sensitivity to inductive stimuli at numerous stages. There has recently been a lot of research done based on genetic, infectious, and environmental risk factors involved in leukemia (Deschler and Lübbert, 2006; Belson *et al.*, 2007). Leukemia can affect anybody at any age, from newborns to the elderly persons, although various types of leukemia have highly distinct age distributions (Juliusson *et al.*, 2012). In addition, compared to the general population, siblings and close relatives of leukemia patients are more likely to acquire leukemia (Sun *et al.*, 2015). The necessity of the hour is to develop primary preventive measures for leukemia patients (Whitehead *et al.*, 2016). The predominant leukemia cells may be mature, precursor cells of multiple lineages, or both. Clinical, morphological, hematological, immune-phenotypic, cytogenetic, and molecular data are used to classify acute leukemia with accuracy and prognosis estimation (Jaffe, 2001; Mrozek *et al.*, 2004). Based on its etiology, leukemia is subdivided into multiple subtypes.

1.1 Types and classification

Leukemia is categorized into two major subgroups i.e., acute, and chronic leukemia, based on the maturation stage of the affected blasts (immature cells). While, based on affected cell lineage, it is divided into lymphoid and myeloid leukemia. Myeloid leukemia originates from myeloid progenitor cells, which subsequently develops into white/red blood cells and platelets. Lymphocytic/lymphoid leukemia originates from lymphoid progenitor cell, which subsequently develops into lymphocytes and natural killer cells. World Health Organization (WHO) has categorized leukemias and lymphomas into various entities based on clinic-morphological features, immunophenotyping, cytogenetics and molecular biology (Vardiman *et al.*, 2009). The WHO revised classification appears in Table 1.1.

Table 1.1: WHO classification of leukemia (Arber *et al.*, 2016; Loghavi *et al.*, 2023)

S#	Tissue Origin	Malignancy type	Subtype
1	Myeloid	Myeloproliferative neoplasms (MPN)	Chronic myeloid leukemia (CML), BCR-ABL11 Chronic neutrophilic leukemia (CNL) Chronic eosinophilic leukemia, not otherwise specified (NOS) Chronic myelomonocytic leukemia (CMML) Atypical chronic myeloid leukemia (CML), BCR-ABL12 Juvenile myelomonocytic leukemia (JMML)
		Acute myeloid leukemia (AML)	AML with recurrent genetic abnormalities Acute myelomonocytic leukemia Acute monoblastic/monocytic leukemia Pure erythroid leukemia Acute megakaryoblastic leukemia Acute basophilic leukemia Acute panmyelosis with myelofibrosis Myeloid leukemia associated with Down syndrome
2	Lymphoid	B-lymphoblastic leukemia	B-lymphoblastic leukemia with recurrent genetic abnormalities B-lymphoblastic leukemia with t (9;22) (q34.1; q11.2); BCR-ABL1 B-lymphoblastic leukemia with t(v;11q23.3); KMT2A rearranged. B-lymphoblastic leukemia with t (12;21) (p13.2; q22.1); ETV6-RUNX1 B-lymphoblastic leukemia with hyper diploidy B-lymphoblastic leukemia with hypodiploidy
		T-lymphoblastic leukemia	Provisional entity: Early T-cell precursor lymphoblastic leukemia Provisional entity: Natural killer (NK) cell lymphoblastic leukemia
		Chronic lymphocytic leukemia	Chronic lymphocytic leukemia Small lymphocytic lymphoma

1.1.1 Acute leukemia

Acute leukemia is a diverse category of hematological cancers characterized by the clonal proliferation of immature myeloid or lymphoid progenitor cells (Kumar *et al.*, 2015). It is a quickly progressing disease, because it arises from rapidly dividing immature blast cells. Its symptoms manifest within weeks. Acute leukemia can be fatal, if left untreated, but it is treatable with the right therapy. The blast cells gradually replace normal hematopoietic tissue and may spread to other bodily organs (Kallen *et al.*, 2019). In the United States, leukemia is responsible for around 20,000 cancer morbidities and over 10,000 fatalities every year (Jemal *et al.*, 2010). Acute leukemia is subdivided into two main types, based on the cell lineage affected: acute lymphocytic leukemia (ALL) and acute myeloid leukemia (AML).

1.1.1.1 Acute lymphoblastic leukemia

The most prevalent cancer in children is ALL. Clinically and morphologically it is diverse (Jemal *et al.*, 2010). It can be of two basic types, i.e. B- ALL or T- ALL, according to the type of lymphocytes proliferating abnormally. ALL has peak incidence in children aged 2 to 5 years and is linked to a worse prognosis in male patients (Haavisto *et al.*, 2016; Steliarova-Foucher *et al.*, 2017). It has a bimodal distribution, with an early peak between the ages of 4-5 years, followed by a second steady increase to a peak around the age of 50 years. ALL is five times more prevalent than Acute myeloid leukemia (AML) and is responsible for around 78% of all pediatric leukemias and 26% of all cancer incidences in children (Charalambous and Vasileiou, 2012; Morra *et al.*, 2017). ALL frequently includes the meninges, which can happen during any phase of development and progression of the disease. The cure rate for childhood ALL is above 80% worldwide (Pui *et al.*, 2008). In children with ALL, chromosomal abnormalities are strong predictors of prognosis. The survival rate of childhood ALL patients has increased globally, including Pakistan, because of better diagnosis and treatment (Idris *et al.*, 2010). Some cases of childhood leukemia could be caused readily by in-utero leukemogenic translocations and gene fusions. It is morphologically categorized into L-1, L-2, and L-3 sub-types using FAB (French, American, and British) criteria, which is clinically repeatable (Bennett *et al.*, 1976).

1.1.1.2 Acute myeloid leukemia

Acute myeloid leukemia (AML) is a category of hematological cancers that develop inside monocyte, megakaryocytic, erythroid, and myeloid, cell lineages in the bone marrow, stamped by the malignant clonal expansion of progenitor cells along with the differentiation arrest (De Kouchkovsky and Abdul-Hay, 2016). AML is the recurrent subtype of leukemia in adults, responsible for 62% of all leukemia fatalities (SEER, 2019). AML is the second most frequent subtype in children, and accounts for upto20% of pediatric leukemia diagnosed and up to 30% of pediatric leukemia-related deaths (Quintana *et al.*, 2005; Margolin, 2011). An increase in age is directly proportionate to an increase in the incidence of AML, that is the reason it is 30 times more common in individuals aging more than 65 years (Bhatia and Neglia, 1995). A maturational arrest of bone marrow cells in the initial stages of development is the underlying pathophysiology of AML. The cause of this halt is unknown, although it is thought to entail the activation of aberrant genes caused by chromosomal translocations and other genetic abnormalities in many cases (Bennett *et al.*, 1985). Based on the kind of cellular origin and the maturation phase of the leukemic cells, the French-American-British (FAB) classification system has classified AML into eight subtypes, i.e., M0, M1, M2, M3, M4, M5 (beginning as undeveloped white blood cells), M6 (begins in immature red blood cells), M7 (begins in platelet-progenitor cells).

1.1.2 Chronic leukemia

Chronic myeloid leukemia (CML) and chronic lymphocytic leukemia (CLL) are two types of chronic leukemia (CL). Chronic leukemia is slowly progressing leukemia, as it involves the transformation of the comparatively mature cell, although some immature blast cells are also involved. The symptoms are slow to manifest and even take years occasionally. Other types of chronic myeloid leukemia, including chronic myelomonocytic leukemia (CMML), also do exist.

1.1.2.1 Chronic lymphocytic leukemia

Chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL) is a hematologic tumor of memory B-cells and recurrent leukemia in people of mostly European ancestry (Yang *et al.*, 2015). CLL is now understood to be the same biological illness as small lymphocytic lymphoma (SLL), but with a distinct clinical appearance (Dores *et al.*, 2007). The two illnesses are currently classified together as CLL/SLL by the WHO. CLL is the most frequent kind of leukemia in the Western world, affecting mostly the elderly people (Tausch *et al.*, 2014). It is marked by an overabundance of mature small B lymphocytes in the blood, bone marrow, lymph nodes, and other lymphoid organs (Zenz *et al.*, 2010). The biological heterogeneity, genomic abnormalities, and mutational profile dictate the diverse clinical presentation of CLL (Zenz *et al.*, 2010; Nadeu *et al.*, 2018). In 2010, 14,990 new cases of CLL were identified, with 4,390 deaths (ACS, 2019). CLL is uncommon among the Asians (Chihara *et al.*, 2014). It is comparatively more frequent than other chronic leukemia with the highest prevalence in 60–80 years old individuals.

1.1.2.2 Chronic myelogenous leukemia (CML)

CML is a myeloproliferative disease marked by the presence of the Philadelphia chromosome, which arises as the result of the fusion of the ABL1 gene on chromosome 9 with the BCR gene on chromosome 22 (Dingli *et al.*, 2010; Granatowicz *et al.*, 2015). The BCR-ABL oncogene induces constitutive expression of an oncoprotein (p210) with tyrosine kinase activity (Lugo *et al.*, 1990). This causes anti-apoptotic pathways to be activated, as well as the uncontrolled proliferation of stem cells, hence increasing myeloid cell mass. The disease progresses through a clonal progression from a reasonably stable chronic phase to an accelerated phase lasting 4–6 months, ending in a blast crisis, which is linked with a poor prognosis and clinical results (Calabretta and Perrotti, 2004; Siegel *et al.*, 2017a). The male-to-female ratio for CML is 1.4:1, and the average age of onset is between 40 and 60 years.

1.2 Signs and symptoms

Leukemia exhibits variation in symptoms at initial diagnosis, which reflects the disease's biological diversity (Lim *et al.*, 2014). The variation may depend on leukemia type, age, and geographical location. Some of the general manifestations of leukemia include fatigue, fever, recurrent infections, chills, weight loss, bone and joint pain, night sweats, hyperoxia, hepatomegaly, splenomegaly, lymphadenopathy, pallor and purpura (Preethi, 2014; Shahab and Raziq, 2014; Sultan *et al.*, 2016b; Jaime-Pérez *et al.*, 2019; Louvigné *et al.*, 2020). Some studies also included rashes and abdominal pain as potential symptoms of leukemia manifestation (Jawaid *et al.*, 2017; Al-Abady, 2021). Anemia is a common clinical consequence associated with acute leukemia that causes a variety of symptoms, including tiredness (Steele and Narendran, 2012). Anemia, infections, bleeding and pallor are all the major signs of bone marrow failure caused by tumor cells infiltrating into the bone marrow, while hepatosplenomegaly is caused by tumor cell infiltration to the liver and spleen (Naeem *et al.*, 2017). The most frequent problems that might lead to mortality in leukemia patients include anemia, infection, and hemorrhages caused by bone marrow loss (Zeidner *et al.*, 2018).

1.3 Leukemia epidemiology

Descriptive epidemiological research can reveal information about leukemia etiology, biology, and at-risk populations. Due to population expansion, urbanization, changing dietary patterns, and increased smoking, the developing nations have higher incidence of hematological malignancies (Rathee *et al.*, 2014). Leukemia accounts for an appreciable number of morbidities and mortalities worldwide. In 2015, there were some 606,000 new cases of leukemia worldwide and 353,000 deaths (Fitzmaurice *et al.*, 2017). In 2020, there were 60,530 morbidities and 23,100 mortalities due to leukemia in the US (Siegel *et al.*, 2021). In most communities, leukemia is the most prevalent kind of cancer in children (Kaatsch, 2010), and accounts for 25–35% of all pediatric cancer cases (Monge *et al.*, 2002; Kaatsch, 2010). Leukemia affects about 8-10 individuals per 100,000 each year in United States (Terwilliger and Abdul-Hay, 2017).

Leukemia incidence is mounting worldwide. According to Global Cancer Observatory (GCO), leukemia ranked 4^{rth} in incidence (153,510) among the top cancers in the general population, in the age range of 0-44 years old across the globe and topmost in mortality (81,033), while it is prevalent in males (87,463) (GLOBOCAN, 2021). The incidence of leukemia subtypes varies according to geography, age, gender, and race, and exhibits different clinical features, treatment response, relapse kinetics, and relapse sites (Belurkar *et al.*, 2013). The exact cause of leukemia is still not known, but certain factors, like host and environment are supposed to play a role in its pathogenesis (ACS, 2016). Each year, 2,000-2,500 new cases of childhood ALL are identified in the United States (Phillips *et al.*, 2015). Every year, some 3,250 children under the age of 20 years are afflicted with leukemia, with approximately 2,400 new cases being diagnosed (Jensen *et al.*, 2004). Until 2017, the incidence of AML was higher than the other three subtypes of leukemia in adults in the US, when it was surpassed by chronic lymphocytic leukemia (CLL) (De Kouchkovsky and Abdul-Hay, 2016; Siegel *et al.*, 2017b; Siegel *et al.*, 2019). In 2019, 21,450 individuals (11,650 men and 9,800 women) were to be diagnosed with AML (Siegel *et al.*, 2019; SEER, 2019).

In the United States, CLL is the most prevalent type of adult leukemia, affecting around 4 people per 100,000. Differences in CLL incidence between Asians and those of predominantly European ancestry have been linked to genetic and environmental variables, such as socioeconomic status, access to health care, and exposure to possible leukemia-causing substances (Gale *et al.*, 2000; Jang *et al.*, 2013). The highest rates of CML are seen in Italy, Australia, and the United States, whereas the lowest rates are reported in the Asia (Linet, 2006). CLL deficit in Asians may be gradually vanishing. Using an age-period-cohort model, data from the Taiwan National Cancer Registry revealed that the incidence of CLL increased with each subsequent birth cohort (Wu *et al.*, 2010). CLL may not be as infrequent among Chinese as previously assumed, according to non chinese-population-based research (Gross *et al.*, 2008).

1.3.1 Status in Pakistan

According to the Global Cancer Observatory 2018, Asia is the world's most leukemia-affected region, with a 48.7% incidence rate and a 53.7% mortality rate (ACS, 2019). Cancer data for Pakistan is fragmentary. There are just two regional cancer registries. The Punjab Cancer Registry has its headquartered in Lahore (Badar *et al.*, 2016). It was founded in 2005, and it primarily serves the ethnic Punjabi community of Lahore and its environs. The Karachi South Cancer Registry is the other regional cancer registry situated in Karachi (Bhurgri *et al.*, 2000). Although both registries cover cancer incidence in their respective jurisdictions, nothing can be inferred with certainty about the nationwide incidence and prevalence status of leukemia in the country. There are a few oncology centers across the country due to limitations of funds (Table 1.2, Figure 1.1).

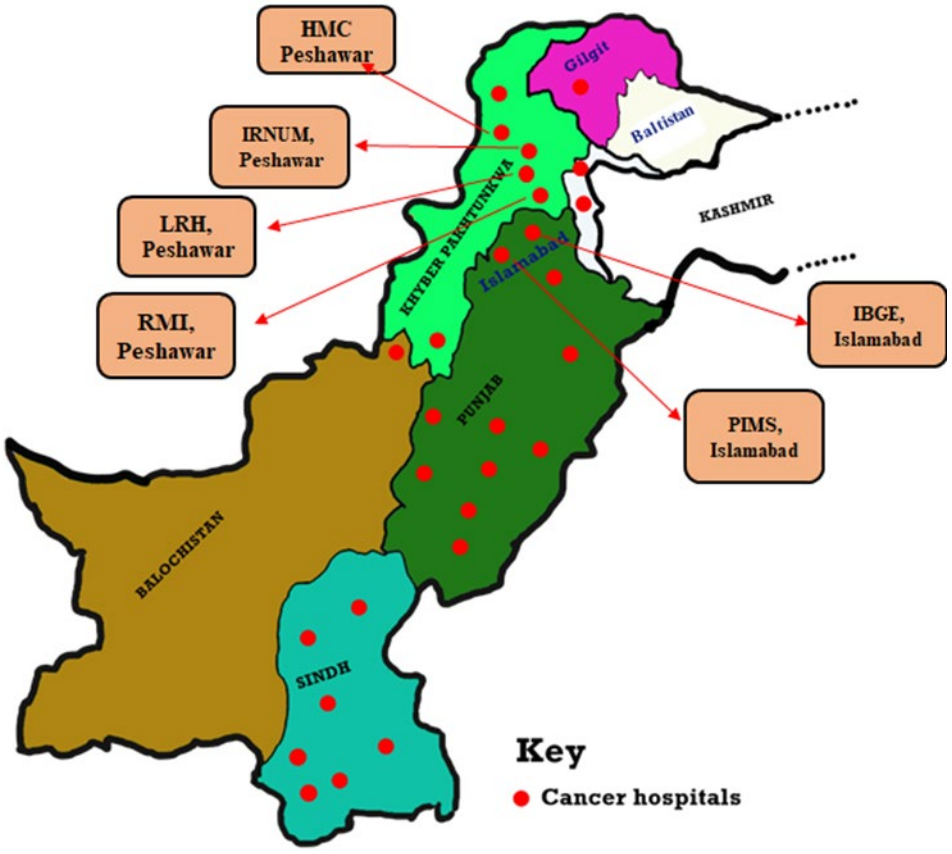


Figure 1.1: Map of Pakistan showing the provinces and the major cancer hospitals situated therein

Table 1.2: Major cancer hospitals in Pakistan

S#	Province	City	Institute	*Acronym*
1	Khyber Pakhtunkhwa	Peshawar	Shaukat Khanum Memorial Cancer Hospital & Research Centre	*SKMCH*
		Peshawar	Institute of Radiotherapy & Nuclear Medicine	*IRNUM*
		Abbottabad	Institute of Nuclear Medicine, Oncology & Radiotherapy	*INOR*
		Swat	Swat Institute of Nuclear Medicine Oncology & Radiotherapy	*SINOR*
		Bannu	Bannu Institute of Nuclear Medicine Oncology and Radiotherapy	*BINOR*
2	Punjab	Lahore	Shaukat Khanum Memorial Cancer Hospital & Research Centre	*SKMCH*
		Lahore	Centre for Nuclear Medicine	*CENUM*
		Lahore	Institute of Nuclear Medicine and Oncology	*INMOL*
		Bahawalpur	Bahawalpur Institute of Nuclear Medicine & Oncology	*BINO*
		Gujranwala	Gujranwala Institute of Nuclear Medicine	*GINUM*
		Faisalabad	Punjab Institute of Nuclear Medicine and Radiotherapy	*PINUM*
		Multan	Multan Institute of Nuclear Medicine and Radiotherapy	*MINAR*
3	Sindh	Karachi	Shaukat Khanum Memorial Cancer Hospital & Research Centre	*SKMCH*
		Karachi	Karachi Institute of Radiotherapy and Nuclear Medicine	*KIRAN*
		Karachi	The Aga Khan University Hospital	
		Karachi	Atomic Energy Medical Centre	*AEMC*
		Nawabshah	Nawabshah Nuclear Medicine Oncology & Radiotherapy Institute	*NORIN*
		Larkana	Larkana Institute of Radiotherapy and Nuclear Medicine	*LINAR*
		Jamshoro	Nuclear Institute of Medicine & Radiotherapy	*NIMRA*
4	Balochistan	Quetta	Centre For Nuclear Medicine & Radiotherapy	*CENAR*
5	Gilgit Baltistan	Gilgit	Gilgit Institute of Nuclear Medicine Oncology	*GINOR*
6	Federal Capital Islamabad	-	Nuclear Medicine, Oncology & Radiotherapy Institute	*NORI*
			Shifa International Hospital Islamabad	
			Pakistan Institute of Medical Sciences	*PIMS*

In Pakistan, the incidence rate of leukemia is rising, yet the local biomedical literature in the disciplines of oncology and leukemia epidemiology is critically lacking. Every year, around 150, 000 new cases of cancer are diagnosed in Pakistan, with a high death rate of 60%–80% (Bray *et al.*, 2018). Leukemia is one of the most common cancers in Pakistani children (3,005) and ranks 9th in terms of prevalence, with a global prevalence of 660,087 cases in 2021(Globocan, 2021; Shahid *et al.*, 2021). According to the Punjab cancer registry 2017 (Afzal, 2020), ALL is the most common childhood cancer in Punjab, Pakistan, with an incidence rate of 20.8% (Afzal, 2020). It accounts for one-third of all recorded cases of childhood cancer (Yasmeen and Ashraf, 2009). Table 1.3 enlists some recent leukemia-based descriptive research undertaken in Pakistan and other regions of the world. The table provides a preview of the studies, focused on epidemiological and clinical aspects, e.g., hematological profile, cytogenetics, risk factors, and prevalence of leukemia in different countries.

Table 1.3: List of studies focusing on the epidemiology of leukemia reported from Pakistan and other parts of the world in last decade (Retrieved from Google scholar and PubMed using the search terms "Leukemia epidemiology", Leukemia research in Pakistan "Leukemia in Asia", "Hematological malignancies", "Leukemia genetics", "JAK/STAT pathway", "*JAK/STAT* mutation in leukemia" and "*JAK/STAT* studies in Pakistan")

S#	Study type	Disease	Sample size	Age	Study Intent	Country	Reference
1	Epidemiological	Pediatric B-ALL	66	1;15	*Nrp-1*	Pakistan	Sahar and Khan, 2019
2	Epidemiological	All types	149	All ages	Risk factors	Pakistan	Nasir *et al.*, 2015
3	Cross-sectional	AML	92	15:55	Cytogenetics	Pakistan	Shaikh *et al.*, 2020
4	Case-control	Leukemia	45	2;12	Diet	Pakistan	Ayub *et al.*, 2020
5	Cross-sectional	Leukemia	400	All ages	Prevalence	Pakistan	Ahmad *et al.*, 2019
6	Cross-sectional	Adult ALL	51	≥15	Demographical, clinic-hematological profiles	Pakistan	Sultan *et al.*, 2016b
7	Cross-sectional	ALL	111	05;55	Tumor Lysis Syndrome in Acute Leukemia	Pakistan	Samad *et al.*, 2019
8	Cross-sectional	B-CLL	101	43-85	Frequency of Zap-70 and CD38	Pakistan	Khaliq *et al.*, 2019
9	Cross-sectional	Overall cancers	789	All	Prevalence of Cancer in District Banu	Pakistan	Rehman *et al.*, 2020
10	Retrospective	Pediatric ALL	153	<15years	Cytogenetic analysis	Pakistan	Shaikh *et al.*, 2014b
11	Cross-sectional	Pediatric AML	54	0:18 years	Clinical and Hematological profiles	Indonesia	Wiraatmadja *et al.*, 2019
12	Cross-sectional	Pediatric ALL	203	0:15	Hematological profiles	Mexican States	Jaime-Pérez *et al.*, 2019
13	Cross-sectional	CML	33	18-69	Hematological profiles	Indonesia	Wiyono *et al.*, 2020
14	Retrospective cohort	ALL	238	0-10	Death causes	Vietnam	Kiem Hao *et al.*, 2020
15	Retrospective	APL (AML	90	>18	Epidemiological and clinical variables	Iraq	Ahmed *et al.*, 2019
16	Retrospective	CML	90	5–72	Clinical, hematological, molecular profile	India	Kumar *et al.*, 2020a
17	Cross-sectional	All HMs*	1,100		Hm pattern	Bangladesh	Sarwar *et al.*
18	Retrospective	Pediatric ALL	177	<15	Epidemiological, clinical characteristics	Yemen	Al-Shehab *et al.*, 2020
20	Prospective	Pediatric ALL	38	0.66 - 15	Clinical presentations, laboratory features,	Vietnam	Hoa, 2020

21	Descriptive	leukemia	100	All	Compared the Quality of Life	Iran	Poorcheraghi *et al.*, 2019
22	Retrospective	Pediatric ALL	69	0:16	Clinical characteristics, Treatment outcome	Palestine	Shawahna, 2020

Nrp 1: Neuropilin-1 gene, Zap-70: Zeta-chain-associated protein kinase-70, CD38: cluster of differentiation 38, APL: Acute promyelocytic leukemia, HMs*: Hematological malignancies, Cytogenetics: studies the number and morphology of chromosomes.

1.4 Risk factors

The exact cause of leukemia remains unknown (Amitay and Keinan-Boker, 2015). Leukemia is caused by a combination of factors, related to both the genetics and the environment. Environmental exposures, particularly those related to agricultural labor, have been linked in certain studies to an increased risk of leukemia, including CML (Van Maele-Fabry *et al.*, 2007; Bonner *et al.*, 2010). Tobacco use, certain pesticides, herbicides, drugs, alcohol, benzene other aromatic hydrocarbons, and ionizing radiations have all been linked to leukemia (Van Maele-Fabry *et al.*, 2007; Cocco *et al.*, 2010; Vlaanderen *et al.*, 2011). Some studies have linked pediatric leukemia to early-life nutrition, increased intake of hotdogs, and cooked meat containing carcinogens (Kwan *et al.*, 2004; Güngör *et al.*, 2019). In the context of leukemia, lifestyle-related risk factors linked with numerous solid malignancies have received minimal attention.

1.5 Genetics

Multiple variables, including environmental and genetic factors, have been linked to the general pathogenesis of leukemia. The existence of recurring genetic markers might indicate leukemia subgroups with different roots. Multiple mutations and translocations have been linked to different leukemia subtypes. Molecular screening can easily detect these translocations. The molecular epidemiology of leukemia in underdeveloped nations is poorly understood, and the few findings that have been published are from single-country populations. In the recent decade, genome-wide association studies (GWAS) have revealed several germline variations that are significantly linked to leukemia susceptibility in children (especially acute lymphoblastic leukemia) (Gutierrez-Camino *et al.*, 2017).

Genetic variations could explain differences in overall prognosis and treatment responsiveness in leukemia patients. Several familial cancer syndromes have been identified to contain the potential for hematological malignancies predisposition (Bannon *et al.*, 2017), providing an opportunity to learn more about the genetic predisposition to cancer genesis. The recent introduction of next-generation sequencing techniques has been contributing to the

known repertoire of genetic variations that lead to hematological malignancies and has aided in the identification of novel familial cancer syndromes linked to leukemia susceptibility. Tables 1.4 and 1.5 highlight some research from Pakistan that address various genetic features of leukemia subtypes. Table 1.4 summarizes the genetic research conducted on different subtypes of leukemia in different age groups during the last decade, showing the presence of several mutations in different genes, while Table 1.5 enlisted the genetic research, particularly confined to mutations in *JAK/STAT* pathway (human and animals) conducted in Pakistan. These studies confirm the presence of polymorphisms and mutation in candidate genes, e.g., JAK1, JAK2, SOCS3, and their association with diseases.

Table 1.4: Leukemia based genetics and cytogenetics studies reported from Pakistan

S#	Article Type	Type of Study	Disease	Sample Size	Population	Target Genes	Reference
1	Research	Genetic	CML	25	Adults	*JAK2* and *BCR/ABL*	Tabassum *et al.*, 2014
2	Research	Genetic	AML	70	Mixed	*FLT3/TKD* and *IDH1*	Ali *et al.*, 2017
3	Research	Cytogenetics	ALL	167	Children	-	Iqbal, 2014
4	Research	Cytogenetics	ALL	166	Children	-	Shaikh *et al.*, 2014a
5	Research	Cytogenetics	CML	51	Adults	-	Tahira *et al.*, 2015
6	Case Report	Genetic	ALL	1	Adults	*RCSD1-ABL1*	Kamran *et al.*, 2015
7	Research	Metabolomics	ALL	186	Adults	Metabolic fingerprinting	Musharraf *et al.*, 2017
8	Research	Genetic	CML	90	Adults	BCR-ABL mutations	Akram *et al.*, 2017
11	Research	Genetic	ALL	104	Adults	Fusion oncogenes	Sabir *et al.*, 2012
12	Research	Genetic	ALL	101	Children	Fusion oncogenes	Awan *et al.*, 2012
13	Research	Genetic	ALL/AML	55	Adults	Flt3 mutations	Ishfaq *et al.*, 2012
14	Research	Genetic	CLL	-	Adults	*NPAS2* (rs2305160)	Rana *et al.*, 2014
16	Research	Genetic	ALL	78	Adults	*BCR-ABL* fusion	Faiz *et al.*, 2011

BCR/ABL+: breakpoint cluster region and Abelson proto-oncogene, *JAK2*: Janus kinase 2, *FLT3:* Fms Related Receptor Tyrosine Kinase 3, TKD: tyrosine kinase domain, and IDH1: Isocitrate dehydrogenase 1, *NPAS2:* Neuronal PAS Domain Protein 2. RCSD 1: RCSD domain-containing.

Table 1.5: Genetic studies based on the JAK/STAT/SOCS pathway conducted in Pakistan

S#	Disease	Gene	Sample size	Polymorphism	Association	Significance	Reference
1	Mastitis (Cattles)	*JAK2, STAT5*	418	g.39645396A>G g.43673888A>G	Yes	Yes	Usman *et al.*, 2015
2	Sepsis	*JAK, STAT, TLR4*	-	-		Formal modeling	Paracha *et al.*, 2014
3	CML	*JAK2*	25	*JAK2* ^{V617F}	Yes	Yes	Tabassum *et al.*, 2014
4	CML	*JAK2*	45	*JAK2* ^{V617F}	Yes	-	Pahore *et al.*, 2011
5	HCV	*SOC3*	250	rs4969170	Yes	rs4969170	Aslam *et al.*, 2016

STAT5: Signal transducer and activator of transcription 5, SOC3: Suppressor Of Cytokine Signaling 3, TLR4: Toll-like receptors 4.

1.5.1 Role of signaling pathways

The proliferation of hematopoietic cells is regulated by signaling pathways activated by growth factors/cytokines and their receptors. Cytokines are released glycoproteins that govern the hematological, immunological and inflammatory response, by acting as intercellular messengers. Several signal transduction pathways serve the hematopoietic stem cells (HSCs) and leukemic stem cells (LSCs) for their survival and maintenance e.g., the PI3K/AKT/mTOR, Notch, WNT/-catenin, TGF, Ras/Raf, and JAK/STAT pathway. The activation of these pathways by molecular abnormalities is thought to be the cause of neoplastic transformation. Several hematopoietic growth factors and cytokines have been shown to have distinct biological functions in various tissues, while also performing redundant roles. During research on interferon function, the JAK/STAT pathway was discovered as a collateral gain (Rane and Reddy, 2002). Aberrations in the JAK/STAT pathway, like other major signaling pathways, play a role in leukemogenesis.

1.5.2 JAK-STAT pathway

The Janus kinase (JAK) signal transducer and activator of transcription (STAT) pathway are one of the most well-studied signal transduction mechanisms adopted for communication between the nucleus and the external environment. It is triggered by a variety of hormones, cytokines, and growth factors. The details of some of the cytokines, which include interleukins, erythropoietin, type ii and type iii interferons, and their biological roles are enlisted in Table 1.6, and it controls cell differentiation, survival, proliferation, and migration based on the cellular environment and external stimulus. JAK-STAT pathway also has a role in hematopoiesis, stem cell maintenance, and immune cell formation, along with other developmental and homeostatic processes. The pathway was initially connected to the survival, proliferation, and differentiation of hematopoietic cells, but later on, it has been linked to growth control, developmental processes, and homeostasis maintenance of various tissues (Levy and Darnell, 2002; Schindler *et al.*, 2007). Each cytokine receptor activates a distinct combination of JAKs and STATs, which is defined by the structure of the receptor chains' intracellular domains (Levy and Darnell, 2002; Schindler *et al.*, 2007).

Table 1.6: Cytokines signaling through JAK/STAT pathway (Morris *et al.*, 2018)

Abbreviation	Name	Biological role
EPO	Erythropoietin	Stimulates the formation of erythrocytes
TPO	Thrombopoietin	Stimulates the formation of megakaryocytes/platelets
GH	Growth Hormone	Growth
PRL	Prolactin	Milk production
LEP	Leptin	Regulates appetite
Others		
IL-12	Interleukin-12	Stimulates T- and NK-cells
IL-13	Interleukin-13	Pleiotropic, airway epithelia, allergic response
IL-23	Interleukin-23	Inflammation
TSLP	Thymic-stromal LymphoPoietin	Inflammatory stimulates T- and B-cells
Class II cytokines		
Type I interferon		
IFNβ	Interferon beta	Anti-viral, ubiquitously expressed
IFNε	Interferon epsilon	Anti-viral, expressed in the female reproductive tract
IFNκ	Interferon kappa	Anti-viral, expressed by keratinocytes
IFNω	Interferon omega	Anti-viral, secreted by leukocytes
Type II interferon		
	Interferon-gamma	activates macrophages/monocytes
Type III interferon		
IFNλ1	Interferon lambda1	Anti-viral, similar to type I but acts on fewer cell-types
IFNλ2	Interferon lambda2	Anti-viral, similar to type I but acts on fewer cell-types
IFNλ3	Interferon lambda3	Anti-viral, similar to type I but acts on fewer cell-types
IL-10 family		
IL-19	Interleukin-19	Inflammatory acts on dermal cells
IL-20	Interleukin-20	Inflammatory acts on dermal cells
IL-22	Interleukin-22	Inflammatory, secreted by Th1 cells, acts on dermal cells
IL-24	Interleukin-24	Inflammatory acts on dermal cells
IL-26	Interleukin-26	Antimicrobial, T_H17 cytokine

1.5.2.1 Components of JAK/STAT pathway

In humans, the JAK family consists of four non-receptor tyrosine kinase proteins designated as JAK1, JAK2, JAK3, and TYK2 (Seavey and Dobrzanski, 2012), each of which has seven homology domains (JH1–JH7) (Ghoreschi *et al.*, 2009). The kinase (JH1) and pseudo kinase (JH2) domains are present in the C-terminal domain of the JAKs product. The activation loop of the JH1 domain contains tyrosine residues, which are required for JAK activation. JH2 is a pseudo kinase domain that is physically identical to JH1 but lacks tyrosine kinase residues, which is important in JH1 activity control (Vainchenker *et al.*, 2008). The SH2-related domain, which facilitates the provision of a docking site to phosphorylated tyrosine residues, is made up of JH3 and half of JH4. The FERM (four-point-one, ezrin, radixin, moesin) region is thought to be involved in JAK's interaction with cytokine receptors (Figure 1.2), is made up of half of the JH4 to JH7 domains (Vainchenker *et al.*, 2008). These tyrosine kinases contain 1,130-1,142 amino acids and have a molecular weight of 120-140 kDa (Kisseleva *et al.*, 2002).

The STAT family consists of seven members, i.e., STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, and STAT6. All these share highly conserved homology domains, i.e., a C-terminal transactivation domain, an N-terminal an SH2 domain, a DNA binding domain, and a spiral domain (Miklossy *et al.*, 2013). These have a molecular weight of 75-95 kDa and 748 - 851 amino acids. Some STATs dimerization and their interactions with other proteins are controlled by their N-terminal region. The spiral coiled-coil domain is involved in a variety of different protein-protein interactions (Kiu and Nicholson, 2012). The STAT dimers' ability to bind to DNA is determined by the DNA-binding domain (Ehret *et al.*, 2001). STATs are directed to tyrosine-phosphorylated peptide sequences inside their dimerization partners by the SH2 domain, which regulates a variety of intracellular signaling processes (Miklossy *et al.*, 2013). The transactivation domain contains two amino acidic residues (tyrosine and serine) these are crucial for STAT activity. Tyrosine is phosphorylated by JAKs to promote STATs dimerization, while the transcriptional activity of STATs is enhanced by the serine phosphorylation achieved through mitogen-activated protein kinases (MAPKs) (Miklossy *et al.*, 2013).

The recruitment of different JAKs and STATs largely depends on the locating tissue specialization and the receptors involved in the signaling transduction (Schindler and Plumlee, 2008). Interferon (IFN) signaling recruits STAT1 and STAT2, while STAT3 is mainly activated byIL 6, IL - 10, epidermal growth factor (EGF), and platelet-derived growth factor (PDGF). STAT4 is involved in IL-12 and IL23 signaling, while STAT5A and STAT5B are involved in the signaling of growth hormone (GH), prolactin (PRL), and the granulocyte-macrophage colony-stimulating factor (GM-CSF). Details of JAK/STAT/SOCS pathway component gene's and their biological importance are given in Table 1.7.

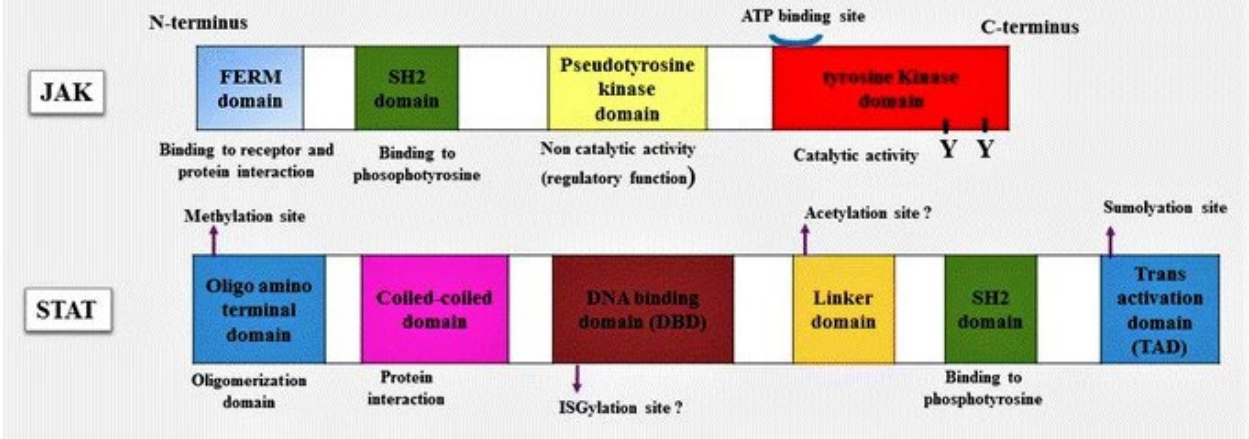


Figure 1.2: Structural details of JAKs and STATs proteins (Seif *et al.*, 2017).

Table 1.7: JAK/STAT /SOCS pathway component genes and their biological significance (retrieved from NCBI databases, UCSC, Genebank, SNPedia, and Ensemble)

Gene	ID	Sequence	Exon	Variants	Loci	PDB	Expression	SNPs	Biological role
(a). JAKS (Janus Kinases)									
JAK1	3716	234,524	29	9	1p31.3	NP_002218.2	Lymph Node, Fat	804	INF-signal transduction
JAK2	3717	143,098	26	7	9p.24.1	NP_004963.1	Appendix, Heart	1017	Prolactin receptor, INF gamma
JAK3	3718	23,290	25	4	19p31.11	NP_000206.2	Appendix, Lymph Nodes	1117	Interleukin receptors
TYK2	7297	30,045	29	11	10p13.2	NP_003322.3	Spleen, BM	1249	Interferon signaling pathways
(b). STATS (Signal Transducer and Activator of Transcription)									
STAT1	6772	45,262	26	4	2q32.2	NP_009330.1	Appendix, Ln, Spleen	434	Activated by INFS and Ils Cell viability
STAT2	6773	18,657	25	8	12q13.3	NP_005410.1	Spleen, Appendices	652	Cell signaling, Defense to viruses
STAT3	6774	75,180	24	12	17q21.2	NP_644805.1	Gall Bladder, Lungs	463	Cell growth, apoptosis
STAT4	6775	143,109	30	5	2q32.3	NP_003142.1	Thymus, Myeloid Cells	513	T-cell differentiation
STAT5A	6776	24,397	20	5	17q21.1	XP_005257681.1	Fat, Lymph Nodes, Bm	667	Anti-apoptotic
STATB	6777	89,261	22	3	17q21.2	NP_036580.2	Fat, Ovary, Endometrium, Lymph Nodes, Spleen, Bm	554	Apoptosis, gland development
STAT6	6778	16,010	24	13	12q13.3	NP_003144.3	Spleen, Skin, Lymph Nodes	724	Ant apoptotic, The differentiation Surface marker expression

Gene	ID	Sequence	Exon	Variants	Loci	PDB	Expression	SNPs	Biological role
(C). SOCS (Suppressor of Cytokine Signaling Proteins) (STAT Induced STAT Inhibitors)									
CISH	1154	5,378	4	3	3p21.2	NP_037456.5	Thyroid, Kidney	282	STAT inhibitor
SOCS1	8651	1,766	2	1	16p13.13	NP_003736.1	Appendix, Lymph Nodes	287	Modulator of IFN-gamma, postnatal growth, survival.
SOCS2	8835	56,415	12	19	12q22	XP_016875641.1	Endometrium, Prostrate	195	Regulation of IGF1R mediated cell signaling.
SOCS3	9021	3,303	2	1	17q25.3	NP_003946.3	Gall Bladder, Urinary Bladder	216	Negative regulation of fetal liver hematopoiesis, and placental development.
SOCS4	122809	22,364	3	4	14q22.3	XP_011534728.1	Appendix, Lymph Node	440	Negative regulators of cytokine signaling
SOCS5	9655	63,829	3	2	2p21	NP_659198.1	Ovary, Endometrium	501	Role not determined
SOCS6	9306	41,299	5	4	18q22.2	XP_016881576.1	Small Intestine, Colon	521	Negative regulators of cytokine signaling
SOCS7	30837	53,718	11	3	17q12	-	Testis, Brain	532	-

CISH: Cytokine-inducible SH2-containing protein, BM: Bone marrow.

1.5.2.2 JAK/STAT mechanism of action

More than two decades ago, a series of breakthrough research from the laboratories of James Darnell, George Stark, and Ian Kerr discovered the molecular intricacies of the JAK/STAT pathway (Stark and Darnell Jr, 2012). It is incredibly basic signal transduction, where the messenger molecule requires only three components to induce a response (kinase, receptor, and transcription factor). The cytokine lands on a unique receptor on the cell surface of the target cell. These receptors have intracellular domains that are always linked with tyrosine kinases from the JAK (Janus Kinase) family (Kawamura *et al.*, 1994). Before cytokine exposure, JAKs are inactive; nevertheless, the autoactivation of the binding of the JAKs is achieved through transphosphorylation, as a result of cytokine and receptor binding (Feng *et al.*, 1997). After activation of JAKs, the tyrosine present on the intracellular domain is phosphorylated, which serves as a docking site for the STATs family (Argetsinger *et al.*, 1993). JAK then phosphorylates receptor-localized STATs (Schindler *et al.*, 1992). This causes them to dissociate from the receptor and translocate to the nucleus, where they stimulate the expression of cytokine-responsive genes (Fig.1.3), resulting in proliferation and/or differentiation (Schindler and Darnell Jr, 1995).

A variety of proteins decrease cytokine signaling at different stages of the route to ensure that signaling is turned off correctly. The suppressors of the cytokine signaling (SOCS) family, in particular, are negative feedback inhibitors for the JAK//STATs signal transduction pathway. Each pathway initiated by a particular JAK-STAT combination is shut down by a specific SOCS, through directly binding with JAKs phosphorylation site, and blocking their activation, which targets only that cascade (Figure 1.3). *JAK1, JAK2,* and *TYK2* are all expressed throughout the body, but JAK3 is only found in hematopoietic, vascular smooth muscle, and endothelial cells (Lai and Johnson, 2010). Genes involved in cell survival, proliferation, and differentiation, such as *MYC*, cyclin D1, survivin, and *BCL2*, are transcriptional targets of the JAK-activated STAT family, especially STAT3 and STAT5 (Aoki *et al.*, 2003; Zhang *et al.*, 2003). Aberrant JAK/STAT pathway activation is considered to play a role in the etiology of various cancers, and it is becoming evident that STATs are involved in both the intrinsic and extrinsic cancer-associated inflammatory milieu (Levine *et al.*, 2007;

Vainchenker *et al.*, 2008). Many of the corresponding proteins have long been known to have a role in tumor development and progression, establishing a biological connection between JAK-STAT signaling and cancer. The JAK-STAT system, like other signaling pathways in the cell, is carefully regulated at several levels (Figure 1.3). STAT proteins, for example, may be subjected to a range of post-translational changes, such as serine phosphorylation, which is necessary for the full transcriptional activity (Schindler *et al.*, 2007). Downstream genes, contributing to crucial biological processes regulated by the JAK/STAT pathway studied in different models, are listed in Table 1.8. The table shows the genes whose transcription activation is dependent on transcription factor STAT3 is activated as a result of JAK/STAT signaling in human cell lines and murine cell lines/models. These genes play crucial roles in different biological processes, ranging from cell division to apoptosis, STAT3 binds to binding site, located in the regulatory region of the respective gene, specific for each gene and initiates transcription of the gene.

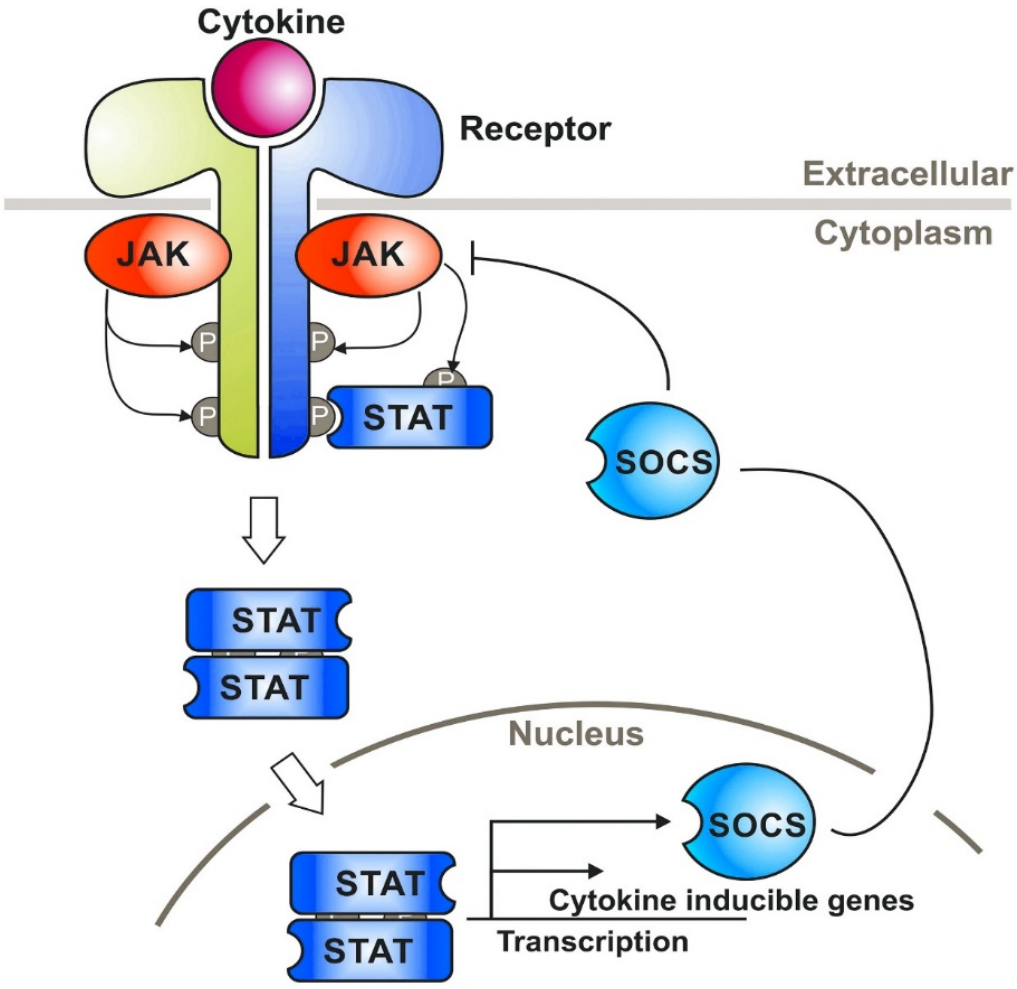


Figure 1.3: Schematic illustration of the JAK/STAT pathway working and regulation mechanism (Morris *et al.*, 2018).

The figure shows the general canonical mechanism of the JAK/STAT signaling pathway and its negative regulation. The cytokine molecule binds to the cytokine receptor (gp130), which undergoes conformational changes and employs JAKs. After attachment JAKs molecules activate each other auto/cross phosphorylation. STAT proteins attach themselves to the activated JAKs and get phosphorylated. After activation, STATs detached from JAKs and form a homo or hetero dimer and translocate to the nucleus. In the nucleus, they attach to the GAS elements of the target genes and start transcription. To inhibit this signaling, SOCS proteins are produced through another JAK/STAT signaling cascade. The SOCS proteins attach themselves to JAKS at the activation site and thus block JAKs activation. SOCS regulates JAK signaling through other mechanisms like binding to STATs binding sites or directing JAKs degradation through ubiquitination.

Table 1.8: Downstream genes contributing to crucial biological processes regulated by JAK/STAT pathway, as studied in different models (Carpenter and Lo, 2014).

STAT3-Regulated Genes	STAT3 Binding Site(s)	Cell/Tissue Type(s)	Species
Transcription Factors			
C-FOS	−348 to −339 bp	HepG2, A431 Cells	Human
HIF-1A	−363 to −355 bp	A2058, v-Src-3T3 Cells, B16 Tumors	Human Murine
C-MYC	+84 to +115 bp	HepG2, BAF-G277, KT-3, CCE ES Cells	Human Murine
SOX2	−5.7 to −3.3 kb −528 to +238 bp	CCE ES Cells	Murine
NANOG	−871 to −585 bp	Mouse Embryonic Cells	Murine
TWIST	−116 to −107 bp −103 to −96 bp	A431 Cells	Human
ZEB1	−310 to −130 bp	SW1116, LoVo Cells	Human
P53	−128 bp	NIH-3T3, MEF Cells	Murine
OCT-1	−3.5 to −2.5 kb	Eca-109 Cells	Human
Transcription Factors			
FOXO1	−515 bp	CD4+ T Cells	Murine
FOXO3A	+196 bp	CD4+ T Cells	Murine
FOXP3	Intron 1	293 Cells	Human
NECDIN	−588 bp	v-Src-3T3	Murine
Survival and Metastasis			
p21CIP1/WAF1	−4183 bp −2540 bp −640 bp	MG63, A431, HT-29, WiDr, HepG2 Cells	Human
PI3K P50A	−276 bp	Mammary Gland	Murine
PI3K P55A	−624 bp	Mammary Gland	Murine
Stat3 role in Tumor Surveillance			
IL-6	−73 to −54 bp	NIH-3T3, CT26	Murine
TNF-A	−1452 bp	Macrophages, SCK1	Murine
IFN-Γ	105 to 542 bp	T Cells	Human Murine
RANTES	−120 to −1 bp	PC3, NIH-3T3 Cells	Human Murine
CRP	−112 to −105 bp	Hep3B Cells	Human
STAT1	−604 to −596 bp −444 to −435 bp −363 to −356 bp −246 to −239 bp	MDA-MB-468 Cells	Human
RORΓT	1st Intron	TH17 Cells	Murine

STAT3-Regulated Genes	STAT3 Binding Site(s)	Cell/Tissue Type(s)	Species
RORα	1st Intron	TH17 Cells	Murine
BATF	2nd Intron	TH17 Cells	Murine
IRF4	Proximal Promoter	TH17 Cells	Murine
IL-6Rα	1st Intron	TH17 Cells	Murine
IL-23R	UD	TH17 Cells	Murine
IL-17A	−144 bp	TH17 Cells	Murine
IL-17F	−309 bp −326 bp	TH17 Cells	Murine
Apoptosis and proliferation			
BCL-2	−1022 to −1002 bp	Hela Cells	Human
MCL-1	−94 to −86 bp	U266, v-Src-3T3 Cells	Human/Murine
BCL-XL	−600 to 0 bp	U266 Myeloma, NIH-3T3 Cells	Human Murine
SURVIVIN	−1174 to −1166 bp −1095 to −1087 bp	MDA-MB-453, NIH-3T3 Cells	Human Murine
FAS	−460 to −240 bp	Myeloma Cells	Human
HSP70	−122 to −90 bp	VSM, HeLa Cells	Human
HSP90A	−1642 to −1485 bp	Jurkat Cells	Human
HSP90B	−643 to −623 bp	VSM Cells	Human
CYCLIN-D1	−984 bp, −568 bp, −475 bp, −239 bp	293T, 3YI, NIH-3T3, 2fTG Cells	Murine
Immune-suppression and inflammation			
IL-10	−120 to −111 bp	RPMI-8226 B Cells	Human
IL-23	−1159 to +160 bp	B16 Tumors	Murine
TGF-β	−3155 to −2515 bp	CD4+ T Cells	Murine
COX-2	−134 to −127 bp	U87MG Cells	Human
Metastasis			
MMP-1	−79 to −42 bp	T24, HT-29 Cells	Human
MMP-2	−1657 to −1620 bp −625 to −601 bp	C4 K1735 Cells	Murine
MMP-3	−410 to −110 bp	HBVE Cells	Human
MMP-9	−942 to −934 bp	MCF7 Cells	Human
FASCIN	−1095 to −1067 bp −975 to −948 bp	4T1, MDA-MB-231 Cells	Human Murine
VIMENTIN	−757 to −749 bp	MDA-MB-231, C2C12 Cells	Human Murine
RHOU	−1067 to −324 bp	MEF Cells	Murine
ICAM-1	−76 to −66 bp −175 to −97 bp	HepG2, BV2 Cells	Human Murine

STAT3-Regulated Genes	STAT3 Binding Site(s)	Cell/Tissue Type(s)	Species
NGAL	−170 bp	Primary Macrophages	Human
POMC	−399 to −374 bp	AtT20 Cells	Murine
SAA1	−226 to +24 bp	HepG2 Cells	Human
Angiogenesis			
VEGF-A	−848 bp	v-Src-3T3 Cells	Murine
bFGF	−997 to −989 bp	HUVEC	Human
HGF	−149 bp, −110 bp	SP1, RINm5F Cells	Murine
Cell Signaling			
AKT	Proximal −2.2 kb	293 Cells	Human
PIM-1	−934 to −905 bp	Microglial Cells	Murine
TNF-R2	−1578 bp, −364 bp	SW480 Cells	Human
S1P-R1	−588 bp	MB49 Cells, B16 Tumors	Murine
MUC-1	−503 to −495 bp	T74D, ZR-75-1 Cells	Human
Others			
TIMP-1	−49 to −41 bp	HepG2, WI38, CD4+ T Cells	Human Murine
JunB	−196 to −91 bp	HepG2 Cells	Human
iNOS	−142 to −130 bp −84 to −60 bp	A431 Cells	Human
CDC25A	−222 to +58 bp	HepG2, Saos Cells	Human

1.5.2.3 Role of JAK/STAT pathway in hematopoiesis

Hematopoietic stem cells (HSCs) are multipotent and self-renewing cells that are found mostly in the bone marrow (Seita and Weissman, 2010). HSCs replenish blood throughout life of an individual (Orkin and Zon, 2008; Dzierzak and Philipsen, 2013). HSCs can be directed to differentiate or multiply based on their cellular programming and extrinsic cues from the microenvironment (Nakamura-Ishizu *et al.*, 2014). Activation of the JAK/STAT pathway activates the STATs that stimulate hematopoiesis, HSC survival, division, and self-renewal. Extracellular cues from the milieu or niche, as well as internal signaling cascades, direct hematopoietic stem cells (HSCs), to create all mature blood cells.

The JAK-STAT pathway is known to be dysregulated in a variety of cancers. Leukemia is characterized by the abnormal proliferation of aberrant blood cells and abnormalities from HSCs. STAT5 regulates proper lymphoid-myeloid differentiation and is involved in leukemia (Wang and Bunting, 2013). Several hematopoietic and other cytokines activate JAK2, which causes phosphorylation of STATs, particularly STAT5, which controls HSC expansion, survival, and self-renewal (Wang and Bunting, 2013; Bousoik and Montazeri Aliabadi, 2018). HSC requires, JAK1 and JAK2 to maintain its stability. *In-vivo* conditional *Jak1* deletion decreased the potency of self-renewal and changed the differentiation pattern of HSCs (Kleppe *et al.*, 2017), but conditional *Jak2* knockout causes increased apoptosis, bone marrow failure, and loss of quiescence, indicating that it plays a crucial role in the HSC maintenance and proliferation (Akada *et al.*, 2014). JAK3 is required for innate lymphoid cell growth (Robinette *et al.*, 2018), while TYK2 is required for B-lymphoid tumor regulation (Stoiber *et al.*, 2004; Robinette *et al.*, 2018).

Similarly, STAT1 is a key player in megakaryopoiesis (Huang *et al.*, 2007). Under the stimulated but not homeostatic conditions, activated STAT3 promotes HSC self-renewal, making STAT3 important for hematopoietic regeneration (Chung *et al.*, 2006). In early pro-B cells, STAT3 phosphorylation is essential for IFN-b-induced death (Gamero *et al.*, 2006). STAT5's function in normal and leukemic stem cell self-renewal has been established by selective activation (Kato *et al.*, 2005). STAT5 maintains the hematopoietic reserve and promotes multilineage hematolymphoid development by having survival effects on HSCs. The

hematopoietic capacity of *STAT5A/5B*-deficient animals is reduced in a variety of blood lineages (Snow *et al.*, 2002). High levels of STAT5A activity in CD34C cells inhibited myelopoiesis and promoted erythropoiesis, whereas moderate levels resulted in maximal proliferation (Wierenga *et al.*, 2008). At the single-cell level, there are differences in cytokine responses in STAT5 phosphorylation between leukemic and normal progenitors (Han *et al.*, 2009). STAT5A and STAT5B have diverse cell-growth-promoting characteristics, which effect the biological activity of hematopoietic stem and progenitor cells (HSPCs) in different ways. STAT5A phosphorylation at Ser779/780 (mouse/human) regulates HSPC proliferation, transformation, and expansion more effectively than STAT5B (Ghanem *et al.*, 2017). In both normal and leukemic hematopoiesis, other STATs play a role (Figure 1.4).

1.5.2.4 JAK-STAT in hematologic malignancies

Genetic investigations have proven the link between hematologic diseases and JAK-STAT mutations since the 1990s (Leonard and O'shea, 1998; Levine *et al.*, 2007; Jatiani *et al.*, 2010). Genetic mutations resulting in constitutive activation of JAK-STAT can occur anywhere in the pathway, including mutations in *JAKs*, *STATs*, upstream oncogenes, transmembrane receptor alterations, and abnormalities in autocrine/paracrine cytokine secretion, all of which result in activation of STATs (O'shea *et al.*, 2015). Hematological malignancies could be caused by translocations/mutation of the *JAK2* gene (Baxter *et al.*, 2005; James *et al.*, 2005; Jones *et al.*, 2005; Levine *et al.*, 2005). *JAK2* activity is known to be enhanced by the activation mutation V617F, which resultantly elevates STAT5 activity (Levine *et al.*, 2005). The mechanisms adopted by *JAK-STAT* or related mutations leading to hematological neoplasms have been studied using mouse models (Dunbar *et al.*, 2017).

Since the 1990s, cells from acute leukemia have been discovered to be constitutively activated by STAT1, STAT3, and STAT5 (Gouilleux-Gruart *et al.*, 1996). STAT1 has been identified as a tumor promoter in the development of leukemia (Kovacic *et al.*, 2006). STAT5 influences myeloid and lymphoid lineages, contributing to the development of cancer. Multilineage leukemias have been produced by a constitutively active STAT5A mutant that forms increased amounts of stable tetramers (Moriggl *et al.*, 2005). The phosphorylation of

STAT5A Ser725 and 779 observed in primary patient samples and human leukemic cell lines have been revealed to be critical for the malignant transformation of the hematopoietic cell (Friedbichler *et al.*, 2010), Furthermore, the N-terminus of STAT5A/B has a role in the transformation of B-lymphoid cells (Hoelbl *et al.*, 2006). The JAK-STAT pathway has been associated with translocations and inversions, as well as fusion oncogenes.

FLT3 gene is known to have frequent alteration associated with AML (Kiyoi *et al.*, 2002; Network, 2013). STAT target genes have been activated by AML-specific Flt3 mutations (Mizuki *et al.*, 2003), and constitutive activation of STAT5 is known to have resulted from the FLT3-D835 mutation (Taketani *et al.*, 2004). STAT5 is essential for AML cell differentiation and proliferation (Sueur *et al.*, 2020). Furthermore, an enhanced inflammatory response in the human AML niche promotes leukemic proliferation by increasing JAK/STAT pathway activation in BM stromal cells and AML blasts (Habbel *et al.*, 2020). In *BCR/ABL*-induced leukemia, STAT5B is more essential than STAT5A, which explains the high incidence of STAT5B alteration in hematopoietic neoplasms (Kollmann *et al.*, 2019). STAT5A and STAT5B variants with high activity levels in the hematological system of transgenic mice can cause a fatal disease that resembles human peripheral T-cell lymphoma (PTCL), and increased STAT5A/B expression has been seen in human PTCL samples (Maurer *et al.*, 2020). Large granular lymphocytic leukemia (LGL) patients had STAT3 (Koskela *et al.*, 2012) and STAT5B gene mutations, with the $STAT5B^{N642H}$ variant linked to poor disease prognosis (Rajala *et al.*, 2013). The same mutation has been related to a greater risk of relapse in T-cell acute lymphoblastic leukemia (T-ALL) in children (Bandapalli *et al.*, 2014). Early B cell factor 1 (EBF1), a crucial contributor to B-cell lymphopoiesis, has recently been demonstrated to have an inhibitory effect in chronic lymphocytic leukemia (CLL) via inactivating the STAT5 pathway (Wang *et al.*, 2021).

All of these findings confirm the practical contribution of mutated JAKs and STATs in hematologic malignancies in one way or the other. Several JAK/STATs pathway-regulated genes performing in normal and leukemic settings (Figure 1.4) have been identified by numerous studies, enlisted in Table 1.9 and Table 1.10.

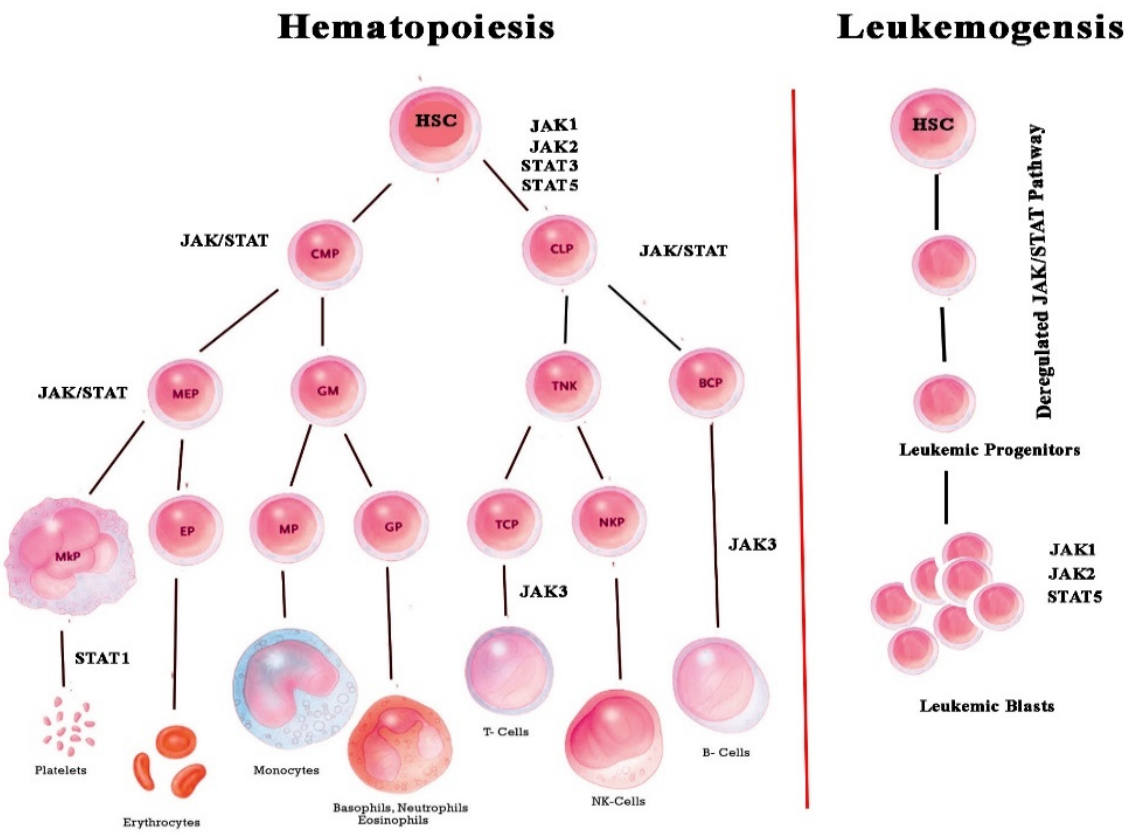


Figure 1.4: Schematic diagram elaborating role of JAK/STAT Pathway in normal hematopoiesis and hematopoiesis during leukemogenesis (Fasouli and Katsantoni, 2021).

The figure explains the role of the JAK/STAT pathway in normal hematopoiesis and the process of leukemogenesis. The hematopoietic stem cell (HSC), differentiates into myeloid progenitor or lymphoid progenitor cell lineage depending on the type of normal JAK/STAT pathway activated in the cell, e.g., JAK1, JAK2, STAT3, and STAT5 differentiate HSC into lymphoid lineage precursor cells, Which under influence of JAK3 can be differentiated and matured into either B and T lymphocytes, depending on the cytokine signal received. The same JAK under the influence of cytokines specific for myeloid origin can differentiate the cell into granulocytes/monocytes or platelets. However, if this pathway is dysregulated due to any mutations and aberrations, the signaling pathway loses its track, and instead of differentiating the stem cells into blood cells, it starts accumulating immature/undifferentiated leukemic cells known as blast cells.

Table 1.9: Genetic alterations in JAK/STAT pathway reported worldwide.

Gene	Mutation	Sample Size	Effect	Disease	Association	Population	Reference
JAK1	S646P A639G P960S	453	Activation	ALL	S646p	Chinese	Li *et al.*, 2017
	L624_R629>W S646F V658F	187	Activation	ALL	Poor prognosis	Colorado	Mullighan *et al.*, 2009
	V623A T478S	94	Activation	AML	Pathogenesis.	Columbus	Tomasson *et al.*, 2008
JAK2	V617F	1/12		10AML, 2MPD	Uncommon	White	Lin *et al.*, 2010
	V617F	25	Activation	CML (ph.+)	Bcr/abl V617f	Pakistan	Tabassum *et al.*, 2014
	R683G	42	Activation	DS-ALL		White	Kearney *et al.*, 2009
	Jak-specific insertion (JSI),	-	-	MPDs	Negative regulation of kinase activity	German	Haan *et al.*, 2009a
	R938Q	1	Activation	B-ALL	Reduced sensitivity to ruxolitinib	Australia	Sadras *et al.*, 2017
	R683G, R683S, I682F QginsR683, R867Q D873N, P933	187	Activation	ALL)	Poor prognosis	Colorado	Mullighan *et al.*, 2009
	SPAG9 JAKFusion	1	-	B- ALL	Aggressive ALL	Japan	Kawamura *et al.*, 2015
	V617F K539L T875N	Cell lines	-	MPD, leukemia	Downregulation of mutant jaks by SOCS	Germany	Haan *et al.*, 2009b
	T875N	Cell lines	Activating	AMKL	Induced MPD	-	Mercher *et al.*, 2006

Gene	Mutation	Sample Size	Effect	Disease	Association	Population	Reference
JAK3	P132A	1/119	Activation	AML	Oncogenic	Caucasian	Riera *et al.*, 2011
	V722I	2/134		AML		Caucasian	Riera *et al.*, 2011
	P132T	1/119		AML		Caucasian	Riera *et al.*, 2011
	V678L Q865E M511I	155	Activation	ALL	Development Of leukemia	French	Vicente *et al.*, 2015
	S789P	187	Activation	ALL	High risk	UK	Zhang *et al.*, 2011
	572V V722I P132T	Cell lines	GOF	AMKL	Transformation	Portland	Walters *et al.*, 2006
	S789P	187	Activation	ALL	Poor prognosis	Colorado	Mullighan *et al.*, 2009
	M511I R657Q V674F V678L	32	Activation	T-PLL	Drugs targeting	Germany	Bergmann *et al.*, 2014
	p.R840C	1	GOF	CTLA-S	Immunodeficiency	US	Sic *et al.*, 2017
	p.Met511Ile,	43	Activating	T-PLL	Therapeutical relevance	Germany	López *et al.*, 2016
TYK2	E957D	Cell lines	GOF	T-ALL	Transformation	USA	Sanda *et al.*, 2013
	p.Pro760Leu, p.Gly761Val	424	Activating (Germline	ALL	Development of ALL.	Netherlands	Waanders *et al.*, 2017
STAT5B	Asp642His (N642H	-	GOF	TCN		Austria	Pham *et al.*, 2018
	p.Asn642His	43	Activating	T-PLL	Therapeutics	Germany	López *et al.*, 2016

ph.+: Philadelphia chromosome +. GOF: Gain of function,

Table 1.10: Single nucleotide polymorphisms studied in JAK/STATS/SOCS pathway across the globe

Gene	SNP	Sample Size	Effect	Disease	Association	Population	Reference
JAK2	rs2230724 rs56118985	260	Activation	AL	Both	Chinese	Zhong *et al.*, 2012b
	rs7046736, rs10815148, rs12342421, rs10758669 rs3808850, rs10974947	179		PV	Yes	White	Pardanani *et al.*, 2007
	A830G	152	-	AML	Ara-C-based chemotherapy	Chinese	Zhong *et al.*, 2010
	46/1 haplotype (rs12343867)	176		AML	adverse prognostic factor in normal karyotype patients	Hungary	Nahajevszky *et al.*, 2011
	rs12343867 46/1	130	-	PMF	inferior survival	USA	Tefferi *et al.*, 2010
	A830G	152	-	AML	Ara-C-based chemotherapy	Chinese	Zhong *et al.*, 2010
	rs12339666	524 JAK2V617F-	-	MPN:	genome-wide significance in JAK2V617F-negative MPN:	UK	Tapper *et al.*, 2015
	rs2230724	260/280	-	AML	Increased susceptibility	Chinese	Zhong *et al.*, 2012b
STAT3	rs17886724	344/ 346	-	Leukemia	prognostic	Chinses	Zhong *et al.*, 2012a
	rs6503691	174	-	CML	Response to IFN	German	Kreil *et al.*, 2010
	rs7211777	441cases/552 controls	-	Breast Cancer	Increased risk	Germany	Vaclavicek *et al.*, 2007
STAT5A	rs2293157	344/ 346	-	Leukemia	significant	Chinses	Zhong *et al.*, 2012b
STAT5B	rs6503691	441cases/552 controls	-	Breast Cancer	Increased risk	Germany	Vaclavicek *et al.*, 2007
	rs2293157	344/366	-	leukemia	prognostic	Chinese	Zhong *et al.*, 2012a

1.6 Diagnosis

Leukemia is declared when a person has >25% of the leukemic blast cells present in the blood. Diagnostic tests and procedures used to diagnose and verify leukemia types are as follows:

(i). Physical examination: During the physical examination, the physical symptoms of the patient are assessed by the oncologist e.g., swollen lymph nodes, enlarged spleen, bruises, etc.

(ii). Complete blood count (CBC): This blood test is performed to evaluate the white blood cell count, which should be greater than normal in leukemia patients.

(iii). Blood cell examination: This test is performed to check the presence of abnormally shaped blood cells.

(iv). Bone marrow biopsy/ aspiration: This test is performed to confirm the diagnosis of leukemia. Bone marrow fluid is drawn through a long needle inserted in patients' pelvic bone (usually) and then observed under a microscope to check the presence of leukemic cells. The typical procedure used for leukemia is bone marrow aspiration with trephine biopsy (Weinkauff *et al.*, 1999). In resource-deficient developing countries, where most patients cannot afford the high expense of cytogenetic testing, bone marrow aspiration is the most used technique for the confirmed diagnosis of leukemia. This also helps to determine whether the disease develops from the B lymphocyte or T lymphocyte, and also helps in the development of a treatment plan.

(v). Cytogenetics testing: This is done to see the presence of any chromosomal aberrations, once the leukemia is diagnosed.

(vi). Immunophenotyping: This test is performed through flow cytometry, to correctly diagnose the subtype of leukemia by checking the expressing cell surface makers. It is critical

to have an accurate diagnosis of acute leukemia since treatment is determined by the subtype (Ullah *et al.*, 2007).

(vii). Imaging and other tests: To confirm the complication and potential damage that happened to other organs, other tests chest X-ray, CT scan, or magnetic resonance imaging (MRI) scan. A lumbar puncture (also called a spinal tap) may also be done, to observe whether cancer cells have spread to the spinal fluid (Rowe, 1999).

The French-American-British (FAB) classification uses only shape and cytochemical labeling of blasts to classify leukemia (ALL /AML). However, according to the World Health Organization's (WHO) recent categorization systems, leukemic blasts must be further evaluated using molecular analysis and flow cytometry. These four techniques of evaluation (morphology, staining, genetic analysis, and flow cytometry) not only distinguish ALL from AML but also define acute leukemia subtypes.

Diagnosis of leukemia is a rapidly evolving. The diagnostic arsenal for leukemia patients has been greatly extended by recent developments in molecular cytogenetic and genetic approaches. The advancements in molecular genetics have shifted the focus of attention away from categorizing every leukemia case according to well-established cytogenetic, cytomorphologic, and PCR-based classification systems and toward the identification of a highly individualized disease profile at diagnosis elucidating clear information for suitable treatment and appropriate follow-up approaches (Wouters *et al.*, 2009).

1.6.1 Modern cytogenetics techniques

Cytogenetics explores the relationship existing between chromosomal abnormalities and genetic disorders in humans (Li and Pinkel, 2006). G and R banding is a common cytogenetic procedure. Not only does the banding analysis take time and effort, but it may miss potentially clinically important submicroscopic chromosomal abnormalities due to its limited resolution power (Simons *et al.*, 2012). Molecular cytogenetic techniques such as fluorescence in situ hybridization (FISH), spectral karyotyping (SKY), and comparative genomic

hybridization (CGH) have surfaced as efficient diagnostic alternatives to address the limitations of banding analysis (Kang and Koo, 2012), as these improve the comprehensive interpretation of numerical and complicated chromosomal anomalies by bridging the gap amongst traditional cytogenetic analysis and molecular genetic investigations (Bejjani and Shaffer, 2008; Russo and Degrassi, 2018). A comparison of technical details between conventional cytogenetic and major molecular approaches is given in Table 1.11.

(i). Fluorescence in situ hybridization (FISH)

Fluorescence in situ hybridization (FISH) is a molecular cytogenetic procedure that employs fluorescent probes that bind uniquely to the section of nuclear DNA with higher complementarity. These probes could be labeled both directly through fluorophore incorporation and indirectly via hapten molecule. FISH has a higher resolution than conventional banding as it can detect smaller genetic variations ranging from 50 Kb to 100 Kb, and also provides their visualization of uncultured cells (Li and Andersson, 2009). Genetic rearrangements have been detected in leukemia by using FISH (Shaffer and Bejjani, 2006). Employing diverse, multicolor FISH assays augments thorough characterization of numerical and complex chromosomal aberrations irrespective of their intricate organization (Chandran *et al.*, 2019). Nevertheless, the fact that only a certain number of probes can be used for FISH, to distinguish these properly, limits the full potential of the staining pattern that could be produced to explore the chromosomal complexity at a higher resolution (Figure 1.5).

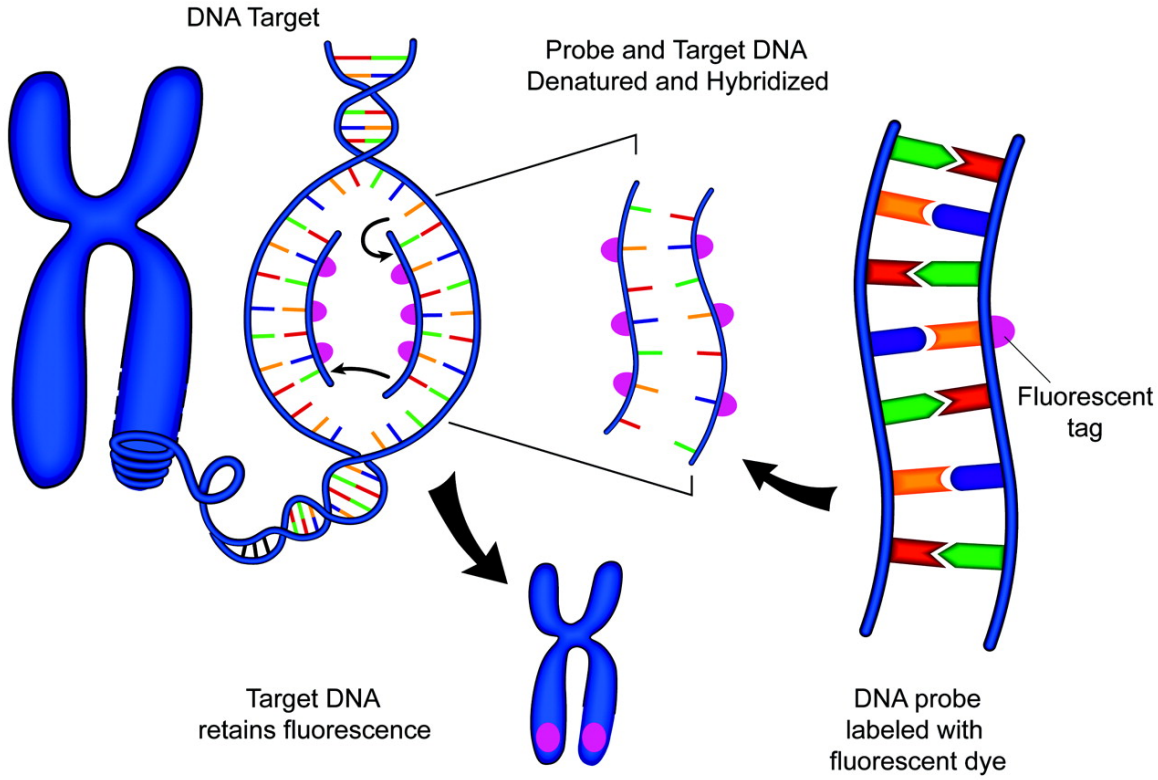


Figure 1.5: Schematic illustration elucidating the basic steps of fluorescence in situ hybridization (Wippold and Perry, 2007).

The figure explains the work mechanism of the FISH technique. Small fragments of florescent single-stranded DNA/RNA, known as probes, are designed for specific regions of the genome. Cells are immobilized on the slide, genomic DNA is denatured, and incubated with the probe molecules. After incubation under stringent conditions, the slide is washed and visualized for any signal through fluorescence microscopy. The fluorescence signals confirm the presence of the respective sequence presence in the genome.

(ii). Comparative genomic hybridization

Comparative genomic hybridization (CGH) is a variation of FISH that can differentiate the difference between a reference and a test genome based on DNA copy variation. This technology can be employed to reveal the genetic anomalies present in disease models, as it screens the complete genome to detect copy number variations (Kang, 2018). Though this technique is time-consuming and labor-intensive, but it can detect multiple abnormalities in the genome like variation in whole genome ploidy, inversions, balanced chromosomal translocations, and mosaicism, can't be detected by CGH due to its limited resolution of 5-10 Mb (metaphase chromosomes) (Gullotta *et al.*, 2007). An advanced technique with higher resolution is required to tackle such alterations (Fig.1.6).

(iii). Array comparative genomic hybridization (aCGH)

The initiation of microarray technology in clinical cytogenetics has started impacting our existing knowledge about the genetic basis of human disease, e.g. leukemia, with record momentum of acquisition (De Paula Junior *et al.*, 2018; Lin *et al.*, 2018). This technology permits the DNA comparison of the labeled genome (test and reference) by combining the microarray platform with fluorescence techniques, which leads to the subsequent sensitive and accurate identification of any aberrations on multiple locations in a single experimental setup, as the microarray can accommodate the use of thousand individual DNA sequences (Gresham *et al.*, 2008; Kriegsmann *et al.*, 2019).

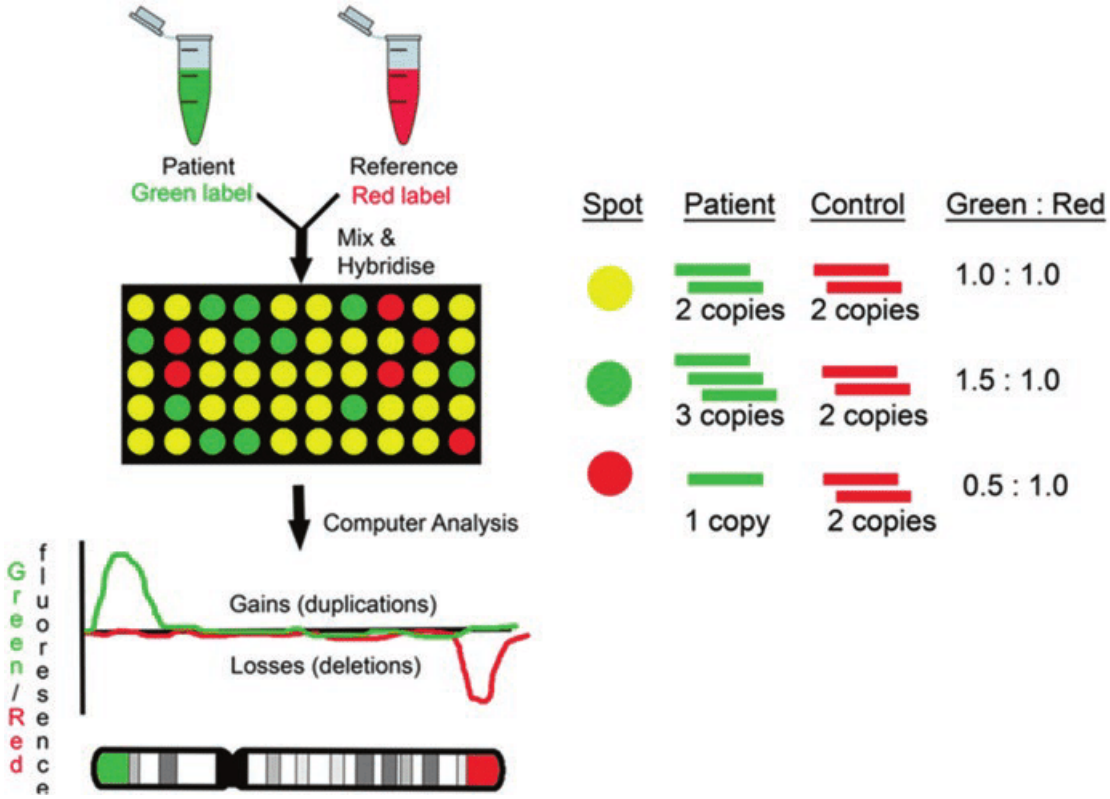


Figure 1.6: Graphical representation of the comparative genomic hybridization (Dugoff *et al.*, 2016).

Figure shows the basic steps involved in comparative genomic hybridization analysis. The patients and normal (reference genome) are labeled with different fluorophores/florescent dyes and hybridized to the array containing fixed probes. The amount of patient and normal DNA is scanned at each probe. The presence of an extra copy number or the absence of a normal copy number is determined by the color of the scan at each point. In the present image, the yellow color of the spot indicates that there is no loss or gain in the patient genome at the respective position. The green color indicates the presence of more than normal copies (patient DNA was labeled green), and the red color indicates an absence of a normal copy from the patient's genome.

Table 1.11: Comparison of technical details between conventional cytogenetic and major molecular approaches (Kang, 2018).

Techniques	Method	Characteristics	Application	Advantages	Limitations
Conventional cytogenetics (G-banding, R-banding)	Cell culturing	Special dye generates a banding pattern for each chromosome	Detection of numerical and structural chromosomal anomalies	Genome-wide screening for chromosome-level abnormalities	Time-consuming and labor-intensive Limited submicroscopic level resolution
Conventional FISH	Molecular technique	Labeled DNA is used as a probe to search for target sequences in chromosome	Detect all types of balanced and unbalanced defects	Interphase cytogenetics possible Simple procedures	Limited complexity of staining pattern due to the limited number of FISH probes
Spectral Karyotyping (SKY)	Arresting cells in metaphase	Chromosome specific probes allows the painting of every chromosome	Detection of rearrangements including complex anomalies	Fast characterization of euchromatic marker chromosome content	Unable to detect intrachromosomal rearrangements
CGH	Molecular technique	Comparative hybridization of differentially labeled total genomic tumor DNA and reference DNA	Identify and assess biomarkers Gene discovery, functional analysis	Whole genome-wide screening Of unbalanced genomic anomalies No need for cell culture	Fail to detect balanced chromosomal translocations, mosaicism, inversions, changes in whole genome ploidy
Array-CGH	Molecular technique	Identification of DNA sequences by specific DNA binding proteins in cells	Identification of cryptic rearrangements (aneuploidy, deletions, duplications or amplifications)	High-resolution target-specific detection of gene amplification, submicroscopic information on imbalances	Higher technical requirements and expensive for routine applications

CGH: Comparative Genomic Hybridization, G-banding: Giemsa banding, R-banding: Reverse banding

1.6.2 Modern genetics techniques

With the new era of advanced genetic techniques, like high-throughput sequencing, it is possible to collect valuable information about possible mutations holding predictive or prognostic significance in leukemia. The application of high-throughput genetic techniques in hematology has recently been extensively reviewed (Merker *et al.*, 2012; Braggio *et al.*, 2013; Black *et al.*, 2015). Among these, next-generation sequencing (NGS) is considered to be the most decorated technique used in the detection of genetic basis leading to the pathogenesis of cancer including leukemia. Next Generation Sequencing (NGS), being the provider of a versatile, accurate, and highly sensitive high-throughput platform for extensive genetic testing, has vividly upgraded cancer research to the next level.

Most specialized hematologic laboratories have initiated employing NGS in recent years. NGS-based detection of the frequently mutated genes in leukemia possesses functional clinical applications for diagnosis, prognosis, prediction of targeted therapy response, and MRD monitoring in leukemia patients (Coccaro *et al.*, 2019; Yang *et al.*, 2020). Reoccurring somatic mutations in leukemia patients, even in those with normal cytogenetics evaluation, can be detected by utilizing comprehensive gene panels via NGS. Next generation sequencing has been divided into main types based on target sequence resolution, e.g. whole genome sequencing (WGS), whole-exome sequencing (WES), and targeted sequencing (TS) (Figure 1.7). The details of all three types are presented in Table 1.12.

(i) Whole genome sequencing (WGS)

Whole-genome sequencing (WGS) is used to perform a comprehensive scan of both the noncoding and coding genome, whole 3 G-bases human genome at the resolution of a single nucleotide. This technique is generally carried out by sidewise paired comparison of the transformed genome to the normal genome to identify the full spectrum of genetic alterations e.g. single nucleotide polymorphisms (SNPs), translocation, copy number alterations, and insertion/ deletion variants (indels). Therefore, it is considered suitable to identify previously unknown variations in cancer subtypes. Despite all its worthiness, WGS is still, a time-consuming, complicated, and non-economical ($5,000 and 10,000/sample) option. Studies

based on the mutational analysis status of leukemia have been recently known to use WGS (Puente *et al.*, 2011; Wang *et al.*, 2011; Burns *et al.*, 2019).

(ii) Whole exome sequencing

The target for whole-exome sequencing (WES) has been narrowed down to only protein-coding regions of approximately 20,000 genes, known as exons, which collectively comprised only 1% of the whole human genome. WES is a better, faster, and comparatively economical ($500–$2000/sample), alternative to WGS, as the exons inhibit about 85% of pathogenic mutations of the genome. WES permits increased depth of coverage (generally $>10^{09}$) and hence facilitates the identification of lower frequency variants. However, it is unable to detect structural variants (comprising larger deletions/insertions) and translocation as they involve intronic regions. Like other high throughput sequencing techniques, WES also utilizes the mechanism of comparison of the cancerous genome to the normal one to identify genomic variations, but here, exons are enriched before the sequencing is initiated. WES has been actively implicated in the detection of somatic and genetic mutations in leukemia patients (Fabbri *et al.*, 2011; Quesada *et al.*, 2012; Landau *et al.*, 2015; Ljungström *et al.*, 2016).

(iii) Targeted genomic sequencing (TS)

Targeted genomic sequencing (TS) selectively targets the panel of strongly disease-associated genes with pathogenesis/clinical relevance. This selective targeting precisely detects low frequency with high confidence, accuracy, and sensitivity as compared to other sequencing approaches. Even low-quality, fragmented DNA can be profiled through targeted genomic sequencing. TS provides in-depth sequence resolution with minimum data complexities and economical cost. The mutational status of genes defined by TS has been used for risk stratification of a patient in clinical trials and cancer research studies (Xu *et al.*, 2017; Oh *et al.*, 2019; Wu *et al.*, 2019a). Targeted sequencing can be performed in either amplicon or capture-based approaches, however, amplicon-based sequencing is more suitable to process low-quality DNA as it requires much less DNA to start with. Targeted sequencing has been recently used to analyze the genetic mutational status of leukemia in some recent studies (Ivey

et al., 2016; Abelson *et al.*, 2018), which suggests its potential role in the accurate diagnosis and personalized therapy of leukemia.

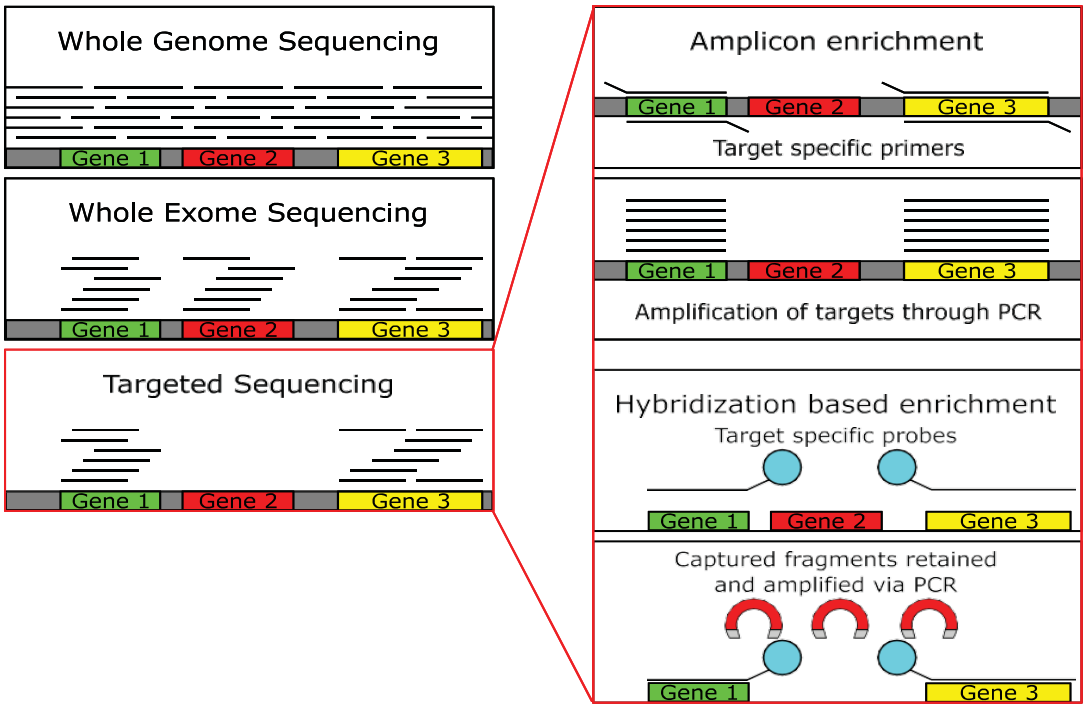


Figure 1.7: Illustration of next-generation sequencing and its types, and library preparation for the targeted sequencing (Bewicke-Copley *et al.*, 2019).

Table 1.12: Different types of next generation sequencing for genomics (Duncavage and Tandon, 2015; Bewicke-Copley *et al.*, 2019).

Platform	Cost/sample	Sites	Region Size	Depth	Data size	Advantages	Disadvantages
WGS	$1000–$3000	All coding and non-coding regions	~3×10^9	30–60	~60 GB–350 GB	Comprehensive and unbiased approach. Capable of identifying single nucleotide variants, indels, and structural variants including translocations and copy number variants	High Cost Substantial data storage requirements Relatively lower coverage impairs sensitivity for variant detection
WES	$500–$2000	Exonic regions	~6×10^7	150–200	~5 GB–20 GB	Sequencing is restricted to exonic regions of the genome where pathogenic mutations are more common	Noncoding genomic sequences are not evaluated. Cannot identify translocations. Difficult to identify copy. Number alterations.
TS	$300–$1000	Specifically targeted regions	~1×10^5 – 1×10^7	200–1000	~100 MB– 5 GB	High sensitivity and specificity of variant detection Low cost Streamlined clinical interpretation	Translocation detection is limited by intronic breakpoints included in target capture space Difficulty in identifying copy number variants.

WGS: Whole-genome sequencing, WES: whole-exome sequencing, TS: targeted genomic sequencing.

At the moment, all these advanced technologies have enabled the accurate and sensitive identification of genomic imbalances comprising rearrangements, base-pair changes, insertions, amplifications, duplications, and deletions. The streamlining of genomic characterization will play a special role in the diagnostics of leukemia since it will provide important information regarding the role of genetic alterations in the pathogenesis, classification, diagnosis, prognosis, and therapy of leukemia. Major shifts in genetic diagnostics will be observed with the use of NGS applications in routine diagnosis, due to an increase in computing power and a decrease in net cost. It has been strongly assumed that complete mapping of the transformed genome for cancer patients will be possible shortly, as massive whole-genome parallel sequencing takes over (Wilch and Morton, 2018). Combining NGS with cytogenetic and clinical data can make a more accurate and precise classification of leukemia subtypes possible.

1.7 Treatment

Most acute lymphoblastic leukemia patients can have different types of treatments which depend upon the subtype of malignancy they are suffering from, and the extent of the disease spread. However, there are five major treatment categories.

(i). Chemotherapy: Chemotherapy is the initial treatment of choice involving cytotoxic chemotherapy for leukemia combined with various leukemic drugs. During chemotherapy, drugs (steroids, methotrexate) are given in the form of pills, shots under the skin (subcutaneously), or administered through an IV into a vein or a central line. To minimize infection risks, many patients have an intravenous catheter put into a wide vein through a cone-shaped port with a silicone nose that is surgically implanted beneath the skin, generally around the collarbone (Messinger *et al.*, 2012). Usually, combination chemotherapy is done. Chemo-treatment usually comprises of cycles. The duration of chemo may be 6 months or more. Chemotherapy treatments can be extensive and intense.

In leukemia, cranial irradiation is integrated into possible etiological factors for osteoporosis. For better tolerance in pediatric patient's combination of prednisolone or

dexamethasone, vincristine, and asparaginase, is used to induce remission. By irradiation of cytarabine and methotrexate central nervous system prophylaxis can be achieved (Cortes *et al.*, 2010).

Chemotherapy consists of three phases, according to their specific goals.

(a) Induction: It is the starting phase of chemotherapy. It is done to completely eradicate every possible leukemic cell from the blood and bone marrow (remission). It usually takes 6 weeks to complete this phase.

(b) Consolidation/intensification: It is the second phase, which begins leukemia remission. Its goal is to eliminate any remaining leukemic blasts that escaped the induction phase. Consolidation therapy cycles take four to six months to end.

(c) Maintenance therapy: It is the third and final phase. Its goal is to destroy any leukemic survivor cells of the first two phases, to avoid potential relapse that may have survived the first two treatment phases. Maintenance therapy for about two years.

(ii). Radiation therapy: It uses radiation energy (cancerous site/ whole body as part of a hematopoietic cell transplant) to kill leukemia cells/block their proliferation. For the prevention of the recurrence of leukemia in the central nervous system prophylaxis, radiation can be used for the treatment. Radiation therapy is used as a common method in the treatment of whole-brain prophylaxis of children with acute lymphoblastic leukemia. Central nervous system (CNS) chemotherapy provided favorable results with less development of side effects. Most adult leukemia specialists employ intrathecal chemotherapy and have abandoned the use of radiation treatment for CNS prophylaxis.

(iii). Immunotherapy/ biologic therapy: This therapy is used to boost the natural defensive immunity of the patient's body with the help of mediators like CAR-T cell therapy, interferon, and interleukins.

(iv) Hematopoietic cell/bone marrow transplant: This procedure replenishes the healthy blood cells, by introducing HSCs to the patient's body, taken either from the patient before treatment or from matching donors. Healthy hematopoietic cells flourish and divide producing healthy bone marrow and blood cells.

(v) Targeted therapy: This treatment uses features-specific drugs against leukemia cells. It works by halting the leukemic cell's ability to proliferate/induce apoptosis or deprive of blood supply, without harming normal cells e.g., monoclonal antibodies (such as inotuzumab rituximab, ofatumumab, etc) and tyrosine kinase inhibitors (imatinib, ruxolitinib, etc).

Most types of targeted therapy help treat cancer by interfering with specific proteins that help tumors growth and spread throughout the body. Targeted drug designing for altered proteins is of crucial importance for this therapy. Both conventional trial-based techniques and computer-aided drug designing have proved their worth in the successful design of targeted therapy.

1.7.1 In silico molecular docking

Computational structure-based drug discovery is one of the most recent approaches, employed to explore the activity of a drug, through computational screening by simulating nature through several techniques. Molecular docking is one of the most common techniques in computer-aided drug designing (CADD) for novel drug lead detection (Hughes et al., 2011; Raj et al., 2019). Molecular docking has proved its role in the *in silico* screening of compounds against their macromolecular targets by employing three-dimensional models. In the present era, CADD is being actively utilized for quick, energy, and cost-effective annotation and analysis of big drug libraries (Chen et al., 2016). Insilco drug designing methodology has emerged as an inventive substitute to complement the process of drug development either through the identification of potential inhibitors for perpetrator proteins (Hall Jr and Ji, 2020; Kumar *et al.*, 2020b; Rahman *et al.*, 2021) or drug repurposing; reuse drug, already existing for the treatment of new disease (Alnajjar et al., 2020; Bhattacharya et al., 2022). Drug repurposing has the potential to save the time and energy required for de novo drug

development (Corsello *et al.*, 2017; Eliaa *et al.*, 2020; Ghanem *et al.*, 2020). CADD has led to the sightings of potential drug candidates for multiple morbidities (Barrows *et al.*, 2016). To confirm the in vivo bindings and effectiveness of the compounds, observed to have strong, stable interactions and lower binding energies, screened through molecular docking, are then assessed through experimentation on animal models (Parks and Smith, 2020).

STAT phosphorylation inhibition is accepted as a feasible curative approach for tumorigenesis disruption. The STA5B-SH2 domain is required both for its dimerization and recruitment to tyrosine-phosphorylated receptor complexes. STAT-SH2 domain inhibitors can inhibit kinase-mediated STAT phosphorylation by blocking the cognate binding site. The main aim of this virtual molecular interaction study was to identify the highly interacting drug with the STAT5B-SH2 domain and to propose an effective drug for potential personalized use for STAT5B inhibition in aggressive/resistant STAT5B $^{N642H+}$ leukemia.

In conclusion, the absence of central population-based cancer registries in the country makes the true elucidation of leukemia prevalence challenging. Further, the selection bias of the study population and partially recorded epidemiological data limits the value of the published data. Population-specific epidemiology can directly contribute to a better understanding of specific disease subtypes and a better prognosis for patients. In this context, the present descriptive-epidemiological study is designed to investigate the bio-demographic attributes of leukemia patients in a multiethnic population, to document its histopathological differentials, and to identify potential risk factors associated with leukemia subtypes.

Corresponding to other cancers, leukemia is also a product of aberrated interaction between genetics and the environment. Irregulated JAK/STAT pathway is known to play a role in leukemogenesis, genes targeted by STAT factors can be useful in stratification strategies, management of leukemia, and provision of novel therapeutic targets. But its extensive role, especially in genetics from the underdeveloped world is still to be unraveled.

Aims

The rationale of the present study is "To study the epidemiology registry of different types of leukemia among Pakistani patients and investigate the crucial signaling pathways in the leukemic cells and their potential inhibitors".

Objectives:

- Epidemiological study of leukemia patients for the determining socio-demographic correlates in Pakistan
- Investigation of the risk factors associated with leukemia incidence.
- Identification of rare polymorphisms and mutations in *JAK/STAT* pathway inside leukemia patients and elucidate their prognostic value.
- Screening of mutational hot spots in key genes of crucial signaling pathways in leukemia patients
- *In silico* analysis of the molecular interaction of crucial mutated proteins identified with their potential inhibitors.

CHAPTER 2

METHODOLOGY

2. Methodology

2.1 Materials

The chemicals and types of equipment used to conduct this study are listed in Tables 2.1A and 2.1 B.

Table 2.1A: List of the chemicals used in the study

S. No	Chemical	Manufacturer
1	TRIS-EDTA	Invitrogen
2	NaCl	Invitrogen
3	SDS	Invitrogen
4	DNA Extraction kit	QIAwave DNA Blood & Tissue Kit (Qiagen)
5	Agarose gel	Invitrogen
6	Primers	Fermentas
7	Mgcl2	Thermofisher Scientific
8	Taq polymerase	Applied Biosciences
9	Big dye (Sanger Sequencing)	Applied Biosciences

Table 2.1B: Llist of the equipment used in the current study

S. No	Equipment	Manufacturer	Model
1	Balance	Ohaus	Harvard Trip
2	Centrifuge-Benchtop	Mistral	Mistral 3000i
3	Centrifuge-Microfuge	Eppendorf	5415C
4	Centrifuge-Minifuge	Scilogex	D1008
5	Gel electrophoresis system horizontal	EC Apparatus Corp.	Maxicell EC360M
6	Nanodrop	Applied Biosciences	
7	PCR machine	BioRad	T100
8	Pipetting Aid-Electronic	Biohit (Midiplus)	71093X
9	Pipette electronic/variable volume	Eppendorf	P10-100
10	Pipette manual/variable volume	Anachem	P05-50
11	Power supply	BioRad	3000
12	Vortex	VWR (Genie2)	G-560E
13	Water Bath	Grant Instruments Ltd.	LSB12
14	Ion Torrent	Thermofisher Scientific	
15	Microarray	Thermofisher Scientific	

2.2 Methods

The present study included basic, clinical, epidemiological, and genetic analyses of the study population consisted of patients with leukemia and their respective controls. The study protocol was approved by the Ethical Review Committees of Quaid-i-Azam University (Letter No. DEBS/2016-619), Islamabad, and the Institute of Biomedical & Genetic Engineering (IB&GE) Islamabad. The research project was allowed by the Ethical Review Committees of Shaheed Zulfiqar Ali Bhutto Medical University (Letter No. 1-1/2015/ERB/SZAMBU), Islamabad, Rehman Medical Institute (RMI), Lady Reading Hospital (LRH), and Institute of Radiotherapy and Nuclear Medicine (IRNUM), Peshawar, for data and samples provision.

2.2.1. Study population

For this study, the data of 1,500 samples, including616 patients (411 males and 205 females) and 884 control subjects (633 males and 262 females) was obtained. However, for extensive analysis, 22 subjects with insufficient data were excluded and 1478 subjects were processed for detailed analysis. Registered leukemia patients were enrolled from tertiary care hospitals of Islamabad and Peshawar, with informed written consent (Annexure 1) of the patient/legal guardian, according to the Helsinki II declaration, from March 2017 to January 2020. The sampling technique used was non-probability convenience sampling. The sample size was calculated by using the Daniel equation (a variant for the disease of unknown prevalence)., Arround90% of the eligible patients participated in the study. Patients were recruited irrespective of their gender, age group, and ethnicity from the Pakistani population, suffering from some type of leukemia. Both newly diagnosed and those on active treatment were included in the study without bias, while. patients with other closely related cancers, i.e., lymphoma, myeloproliferative neoplasms (CML exempted), and off-treatment leukemia patients, were excluded from the study. The patient coming to the target cancer centers and their controls were mainly coming from four regions of Pakistan, viz., Punjab, Khyber Pakhtunkhwa (KPK), Kashmir, and the Federal capital, Islamabad.

For comparison, age, gender, and ethnicity-matched healthy controls, belonging to similar areas, environmental conditions, and socio-economic status, but biologically unrelated, were randomly collected from the source population (produced patients), without exposure bias . About 70% of the eligible controls consented to participate in the study, the hesitant controls were mainly comprised of children.

2.2.2. Data collection and sampling

A team of trained systematically interviewed both leukemia patients/ guardians (children) and controls, using a mixed questionnaire, containing both open-ended and closed-ended questions based on the type of variable in question (Annexure 2). Detailed medical records of the patients were acquired and data on clinical diagnosis, demography, lifestyle, and risk factors were obtained. Complete blood picture profiling was done for both patients and controls, using services of the qualified nurses and medical technicians. The definitions of the demographic variables were obtained from the Pakistan Demographic and Health Survey (Nips, 2019).

Both patients and controls were grouped into different classes according to the requirement of downstream analysis. Different anthropometric, demographic, and clinical parameters were analyzed and compared. The blood samples, collected by qualified nurses, were kept in an EDTA vacutainer and stored at -4°C in the Human Genetics Laboratory (Department of Animal Sciences), Quaid-i-Azam University, Islamabad, and were transported to the Institute of Biomedical and Genetic Engineering (IBGE), Islamabad, later on, for genetic analysis. No financial liability on the part of the subjects was involved.

2.2.3. Study scheme

The study was designed to cover both epidemiological and genetic aspects of leukemia affecting the Pakistani population. Therefore, it was divided into three sections, each addressing a certain objective. The detailed study scheme is presented in Figure 2.1.

The study was completed in three sections.

- Epidemiology
- Genetic Analysis
- *In silico* Interaction Analysis

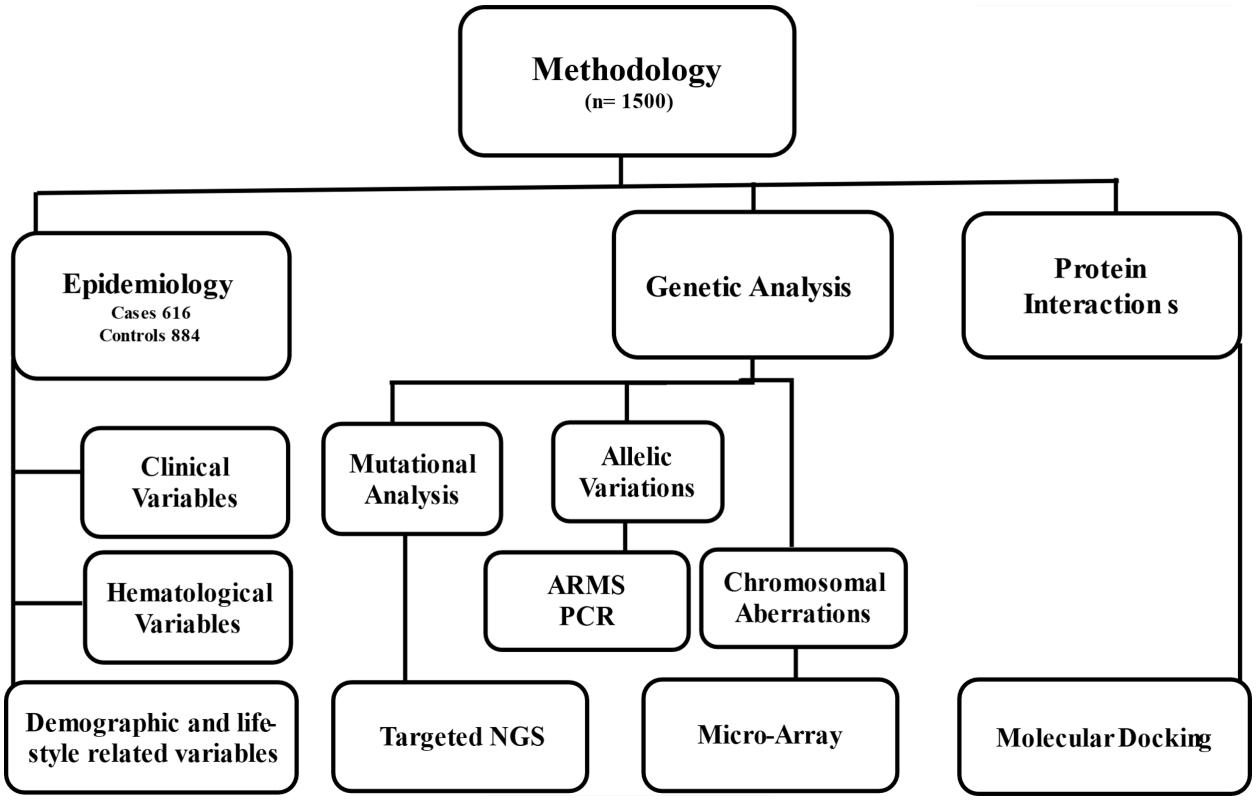


Figure 2.1: Graphical presentation of the study scheme

2.2.3.1 Epidemiological study

The samples for the epidemiological study were collected to determine the prevalence of various leukemia-related risk factors in the study population. The data from 616 leukemia samples, used for samples for the epidemiological study, were collected. For comprehension, the samples were divided based on the disease-originating tissue type as lymphoid and myeloid leukemia, and the nature of progression as acute and chronic leukemia.

Anthropometry was done using standardized techniques. Height and weight were determined, and BMI was calculated by the formula: weight (kg)/height (m^2). According to the revised criteria for the Asian population, the category of overweight was defined as BMI ≥25 Kg/m^2 (WHO, 2004). Later on, 22 samples with more deficient clinical data were excluded from the study, and further evaluations were done using the remaining 594 samples.

The clinical, biochemical, physical, demographic, and risk factors related attributes were investigated in dimensions of origin and nature in all magnitudes.

2.2.3.1.1 Sample and study parameters

The variables considered for this study were weight, diet, education, parental consanguinity, family history of cancer, area, drinking water source, fuel type used in the household (wood or natural gas), use of carbonated drink, perfume use, and tobacco use (Table 3.6).

(i). Age group and gender:

The patients and controls were categorized into five groups based on their age; infants (<1 year), preschoolers (1.1-5years), children (5.1-15 years), adolescents, and young adults (AYAs; 15.1-39 years), and aged (>39.1. All of these age groups were subdivided into male and female patients.

(ii). Ethnicity:

Pakistani population is multiethnic. Therefore, the source population consisted of multiple ethnicities. The major ethnic classes represented by study participants were Pathans, Punjabi, and Kashmiris and a small number of participants were representatives of other ethnicities, i.e., Sindhi, Balochi, Baltians, etc. Since the number of such representations was fewer, therefore all of such groups were added up and represented by group names and other ethnicities.

(iii). Education:

Individuals who obtained some level (primary; 5 years, secondary; 10 years, graduation; 14 years, masters; 16 years, and so on) of formal education were considered educated, children below the age of 5 years were placed in the inadmissible category, and individuals who have not obtained any form of school education were placed in no formal education category. The education level was assessed by asking "What is the highest grade of school completed".

(iv). Diet:

Diet is categorized into standard and poor diet. A standard diet includes an intake of fruit, vegetables, legumes, nuts, whole grains, and meat. A poor diet is a diet lacking one or more of the food groups (WHO, 2019). Participants in the study, who were unable to consume any major group of bio-nutrients like proteins or carbohydrates, continuously for longer than WHO-recommended durations were considered to have a poor diet. The diet type was assessed by asking "Enlist the dietary item you on consume a weekly basis?".

(v). Weight:

Body mass index (BMI) was derived from the height and weight (patients' weights at the time of diagnosis were considered) taken during the sampling. Age and gender-specific

indices were used for BMI/percentile calculation. According to international standards, a ch'ld's weight below the 5th percentile, 6-85 percentile, and >85 percentile were considered as underweight, normal, and overweight respectively. In adults, individuals with a BMI <18.5kg/m2 were considered underweight, with a BMI of 18.5kg/m2-24.9kg/m2 as normal, while a BMI >24.9kg/m2 was considered overweight**.**

(vi). Parental consanguinity:

The marriage/reproductive relationship between two closely related individuals is defined as consanguinity. It is a concept of both social and genetic importance. Information regarding parental consanguinity was collected for study participants. For each patient, a detailed pedigree was constructed. Parental marriage types up to the second cousin with an inbreeding coefficient of F=0.0156 were considered consanguineous.

(vii). Family history:

The information about family history of cancer was obtained from all the participants. The individuals having any family member, alive or dead suffered from any form of cancer, solid or hematological were included in the "Yes" category, while the remaining were added into the "No" category.

(viii). Marital status:

Marital status was recorded in a dichotomous pattern, as married or unmarried, only for patients above the legal marital age, of 18 years. Patients who were younger than 18 years of age were placed in the category "Below marital age" (the legal age for marriage in Pakistan is 18 years old; Child Marriage Restraint Act, 1929, Muslim Family Laws Ordinance, 1961).

(ix). Area:

The information regarding the area where study participants were living for more than two years was recorded. The area where participants lived in villages and countryside areas

with diluted populations and fewer infrastructures and services were considered rural areas, while the area with big towns and cities with dense populations and advanced infrastructure was considered urban areas.

(xi). Environmental status:

Residential surroundings with no noise, traffic, and physical form of pollutants were considered a clean environment while surroundings with daily exposure to noise, heavy traffic, and physical pollutants were considered a polluted environment.

(xii). Water type:

Water from underground sources, like well, bores, and springs was grouped as groundwater. Water from dams/streams and taps was considered surface water, while water treated with water purified by physical (filtration) or chemical techniques was considered treated water.

(xiii). Domestic fuel type:

The fuel used in homes for cooking food and heating purposes was considered a domestic fuel source. Information regarding domestic fuel sources was recorded in a dichotomous manner, i.e., wood/biomass and natural gas.

(xiv). Family type:

The Family consisting of parents and their children, typically living in one home residence was recorded as a nuclear family, while the family set up consisting of parents like father, mother, and their children, aunts, uncles, grandparents, and cousins, all living in the same household were considered as extended/joint family.

(xv). Carbonated drinks:

Participants consuming form of carbonated drinks in their routine were considered using carbonated drinks, while individuals consuming no carbonated drinks or consuming very rarely, like once a month were considered not using carbonated drinks.

(xvi). Tobacco use:

Participants using tobacco in the form of smoking e.g., cigarettes or smokeless in the form of naswar bheeri, huqqa either active or passive (parental; in the case of children) all were considered using tobacco, while individuals away from any form of tobacco use and continuous exposure were considered no using tobacco.

(xvii). Perfume use:

Participants wearing perfume in any form, liquid (perfume, body spray, attar), or powdered perfume were considered in the perfume use category, while individuals not using any of them were added to the no perfume use category.

(xviii). Microwave use

Participants using a microwave oven for cooking or heating food were considered using the microwave, while individuals not using any microwave were included in the no-use group.

(xix). Hematological variables

Hematological fluctuations of blood cellular count for all study participants were also compared in this study. All the parameters were categorized into increment, normal, and decrement categories according to the employed standard reference level for children, adults, males, and females, in the country (Table 2.2).

Table 2.2: Blood cell status reference ranges

Blood Constituent	**Categories**	**Ranges**			
		Male		Female	
		Child	Adult	Child	Adult
	Decrement	< 5	< 4	< 5	< 4
WBCs (10^9/L)	Normal	= 5-13	= 4-11	= 5-13	= 4-10
	Increment	> 13	> 11	> 13	> 10
	Decrement	< 4	< 3.5	< 4	< 4.5
RBCs (10^6/µL)	Normal	= 4-5.2	= 3.5-5.6	= 4-5.2	=4.5-5.5
	Increment	> 5.2	> 5.6	> 5.2	> 5.5
	Decrement	< 11.5	< 11	< 11.5	< 13
Hemoglobin (g/dL)	Normal	= 11.5-15.5	= 11-18	= 11.5-15.5	= 13-17
	Increment	> 15.5	> 18	> 15.5	> 17
	Decrement	< 170	< 150	< 170	< 150
Platelets (10^3/µL)	Normal	= 170-450	= 150-400	= 170-450	= 150-410
	Increment	> 450	> 400	> 450	> 410
	Decrement	< 30	< 12	< 30	< 25
Lymphocytes %	Normal	= 30-40	= 12-50	= 30-40	= 25-40
	Increment	> 40	> 50	> 40	> 40
	Decrement	< 60	< 37	< 60	< 45
Neutrophils %	Normal	= 60 - 70	= 37-75	= 60-70	= 45-70
	Increment	> 70	> 75	> 70	> 70

2.2.3.2 Molecular genetic study

This study was based on finding the genetic alteration in JAK/STAT pathway, contributing to the process of leukemogenesis. The portion was completed with three distinct approaches.

i. SNP Typing (ARMS-PCR)
ii. Next Generation Sequencing (NGS)
iii. Microarray

2.2.3.2.1 Genomic DNA extraction

Genomic DNA was extracted from the fresh and stored frozen blood by two methods.

i. Manual DNA Extraction
ii. DNA Extraction via Kit (QIAwave DNA Blood & Tissue Kit)

The blood samples obtained from study participants were divided into two categories based on the sample volume collected, (i). >800 μL considered as suitable for manual extraction, (ii). <800 μL, considered minute quantity, and selected for kit-based extraction. Manual extraction is suitable for samples collected in good quantity because there are low chances of sample wastage, and even after repeated washing steps included in the manual protocol, still good-quality genomic DNA can be obtained for further processing. About 90% of the samples were extracted through manual DNA extraction, mainly comprised of samples from AYAs and adult individuals. Kit-based DNA extraction is a safer method for samples having minute/lesser volume because due to its high optimization and fewer steps involved in the extraction, there are smaller chances of losing the sample during the DNA extraction process, and hence high-quality genomic DNA is extracted, though the sample is minute. DNA from about 50 pediatric blood samples was extracted through a kit-based method using QIAwave DNA Blood & Tissue Kit.

(i) Manual DNA extraction

A modified standard salting-out extraction technique was used for the extraction of genomic DNA from whole venous blood samples with a volume greater than 800 µL (Miller *et al.*, 1988).

Solution Composition:

	Composition	Concentration	pH
i	**TNE Buffer**		
	Tris-Cl*	10 mM	pH: 8.0
	EDTA*	02 mM	pH: 8.0
	NaCl*	400 mM	
ii	**TE/Tris-EDTA**		
	Tris-Cl*	10 mM	pH: 8.0
	EDTA*	02 mM	pH: 8.0

iii Tris-Cl* 1M

Dissolved 121 g of Tris base in 800 ml d.H_2O, adjusted the pH to 8.0, brought the total volume to 1 liter with dH_2O, autoclaved, and stored at room temperature.

iv EDTA, 0.5M

Dissolved 186.1g of disodium EDTA dehydrate in 800ml d.H_2O, adjusted the pH to 8.0, brought the total volume to 1 liter with dH_2O, autoclaved, and stored at room temperature.

v NaCl, 1M

Dissolved 58.34g of NaCl in 900 ml d.H_2O, adjusted the total volume to 1 liter with dH_2O.

vi SDS (10%)

Dissolved 10g of SDS in 80 ml d.H_2O, adjusted the volume to 100 ml with d.H_2O.

Procedure:

Blood was transferred from the vacutainer to 15mL falcon tubes, added up to 14 mL of TE buffer., mixed them thoroughly, and spun at 12000 rpm at 4°C. The supernatant was discarded into a glass discarder and repeated step 1. After getting the white pellet was resuspended in 3 ml TNE buffer and 15 µL proteinase k, along with 100 µL of 10% SDS was added. The samples were incubated overnight at 55°C in a water bath shaker. The next morning 1 mL of saturated 5M (NaCl) was added, shaken vigorously, and incubated on ice for 10 minutes. After incubation, the mixture was centrifuged at 3000 rpm for 10 minutes at 4°C. The supernatant was transferred into a new 15ml falcon tube and 2 volumes of chilled absolute ethanol were added. The tubes were mixed until the DNA threads were visible and incubated at -20°C overnight. After incubation, the tubes were centrifuged at 3000 rpm for 60 minutes at 4°C. The supernatant was discarded, 2 ml of 70% chilled ethanol was added to the tubes, mixed gently, centrifuged at 3000 rpm for 60 minutes at 4°C, and discarded the supernatant. The pellets were air-dried and resuspended in 100 µL of TE buffer.

(ii) Genomic DNA extraction via kit

The genomic DNA from the samples containing <800 µL was extracted via "Qiagen Genomic DNA Extraction Kit".

Procedure:

Pipetted 40 µL of proteinase K into the 1.5 ml centrifuge tube and added 400 µL of blood samples and 400 µL of AL Buffer into it. The tubes were vortexed for 15 seconds and incubated for 10 minutes at 56°C in a water bath shaker. After incubation, the mixture was centrifuged briefly to remove liquid from the lid corners of the tube, and added 100% ethanol to it. Mixing was done through vortexing for 15 seconds and then recentrifuged the tubes briefly. In the next step, the mixture was transferred to spin columns provided with the extraction kit without wetting the rim of the lid and placed the spin column in the collection tube, centrifuged at 8000 rpm for 1 minute, and discarded the elute containing the filtrate. After filtration, 500 µL of AW1 buffer was carefully added and centrifuged at 8000 rpm for 1 minute and discarded the elute. The spin columns were placed into a clean collection tube, 500 µL of

AW2 buffer was added carefully to each column and centrifuged at 14000 rpm for 3 minutes. After centrifugation, the elute-containing collection tube was discarded. Placed the spin column into a 1.5ml centrifuge tube and 200 µL of buffer AE was added. The samples were incubated at room temperature for 5 minutes and then centrifuged at 8000 rpm for 2 minutes to elute the DNA into the centrifuge tube. The spin columns were discarded and the DNA in the centrifuged tubes was stored at -4°C.

2.2.3.2.2 DNA quantification and dilution

The quantification of DNA was measured by spectrophotometer (nanodrop). The results of spectrophotometry of these samples showed that they contain an appropriate amount of DNA. The samples with bigger pellets, containing a very high concentration of DNA were further diluted and finally, samples with a DNA concentration >50 ng/µL were selected for PCR amplification.

2.2.3.2.3 PCR amplifications

2.2.3.2.3.1 SNP typing (ARMS-PCR)

Polymorphisms in *JAK1, JAK2*, and *STAT5B* genes were detected by ARMS-PCR. Genomic DNA of 276 leukemia patients, randomly selected and having good concentration, were amplified through ARMS PCR to detect the genetic variations at the selected positions. These samples were from different leukemia subtypes, age groups, and gender. The details of the technique performed are as follows. The reaction was performed in a total volume of 20 µL, including 2 µL of 50 ng of DNA, 2 µL of 10 x PCR buffer, 1 µL of 25mM MgCl2, 1 µL from 10 pmol of each primer, 1 µL of dNTPs (10 mmol/L), 0.25 µL of Taq DNA Polymerase (5 U/µL), and 12.75 µL of PCR grade water in 200 µL PCR tubes (Axygen, USA) per reaction. The mixture was centrifuged at 8,000 rpm for 20 seconds for complete mixing. The procedure was performed in the Biorad PCR System (T100). The reaction mixture composition and PCR conditions are detailed in Table 2.3. The details of each primer set sequence, annealing temperature, and amplicon size are listed in Table 2.4.

Table 2.3: The PCR reaction set up for ARMS-PCR

ARMS-PCR Reaction Set-Up					
PCR reaction mixture					
Reagent	**Final conc.**	**Stock**	**Final**	**/Rxn**	**/276Rxn**
Auto dH2O	Fill vol			12.75	3519
PCR buffer (NH4+)	1x	10	1	2	552
MgCl2 (25 mM)	1.5 mM	25	2.5	1	**276**
dNTPs (2.5 mM)	0.2 mM	2.5	0.125	1	**276**
Primers (25 pM)	0.5 µM	10	0.5	1	**276**
Primers (25 pM)	0.5 µM	10	0.5	1	**276**
Taq Poly. (5 U/µL)	0.05 U/µL	5	0.0625	0.25	69
Template DNA	2 µL, 50 ng/µL	0.05	1.5	1	**276**
	Volume/rxn (µL)			20	-

PCR Conditions

S. No	**Primers**	**Step I**	**Step II**		**Step III**	**Storage**
		Denaturation (45sec)	Annealing (30sec)	Extension (45sec)	Final Extension (7min °C)	∞
1	$JK1^{V623}$	95°C	62°C	72°C	72°C	-
2	$JK2^{S473}$	95°C	60°C	72°C	72°C	-
3	$STAT5B^{N642}$	95°C	62°C	72°C	72°C	-

Table 2.4: List of tetra ARMS-PCR primers, optimized Tm and product size

S#	Gene	Primer ID	Primer Sequence	T_M	Product size (bp)
1	*JAK1*	JK1-V623-F	GGAGCCGTTTAGGGAGTAGG	62°C	591
		JK1-V623-R	GGCTGAGTGTGAGGAGCTGT		
		JK1-V623-Ft	AGAAGAAGATAAAAGTGATCCTCAAGGT	62°C	268
		JK1-V623-R	GGCTGAGTGTGAGGAGCTGT		
		JK1-V623-F	GGAGCCGTTTAGGGAGTAGG	62°C	368
		JK1-V623-Rc	TGTGGCTGGGGTCTATGG		
2	*JAK2*	JK2-473-F	TGGCAGGAATACATCAAAATC	60°C	677
		JK2-473-R	CTTGGGAGTTGCGATATAGG		
		JK2-473-Ftt	GAACCAAATGGTGTTTCAGTT	60°C	341
		JK2-473-R	CTTGGGAGTTGCGATATAGG		
		JK2-473-F	TGGCAGGAATACATCAAAATC	60°C	382
		JK2-473-Raa	AAATATCAAATCTTCATTTCTGACTTT		
3	*STAT5*	STAT5B-642-F	AGATCGCACCACTGCACTC	62°C	578
		STAT5B-642-R	TCCAGCCCAAGGTTTACATC		
		STAT5B-642-Fc	GATCTAGAGGAAAGAATGTTTTCGC	62°C	351
		STAT5B-642-R	TCCAGCCCAAGGTTTACATC		
		STAT5B-642-F	AGATCGCACCACTGCACTC	62°C	281
		STAT5B-642-Ra	GGTGGTAAAAGGCATCAGCTT		

(i). Agarose gel electrophoresis for PCR-amplified DNA

Amplified DNA products were confirmed by 2% agarose gel electrophoresis using the 100 bps DNA ladder (MBI-Fermentas, UK). Two percent agarose gel (2% w/v) was prepared by melting 7.2 gm of agarose (Sigma-Aldrich Mo, USA) in 360mL of 1 X TBE buffer (Tris 89.1 mM, EDTA 2.5 mM, Borate 88.9 mM) in a microwave oven. Ethidium bromide (15 µL) was added to stain the gel; sample combs were placed in the gel tray before pouring the agarose solution and the gel was left to set at room temperature. Once polymerized, the gel combs were removed, and the 6 µL PCR product was mixed with 2 µL of bromophenol blue and loaded into the wells to confirm amplification.

(ii). PCR product purification

After confirmation of the presence of the desired bands by visualization on a gel, the PCR product was set for purification for the sequencing to be performed.

Procedure:

A total of 30 µL of the PCR product was mixed with 3 µL of 10 M sodium acetate ($C_2H_7NO_2$) and 60 µL of chilled absolute (100%) ethanol. The mixture was thoroughly mixed and left on ice overnight. The next day the mixture was spun at 14000 rpm for 15 minutes, the supernatant was discarded, and the pellet was washed with 70% chilled ethanol by spinning again at 14000 rpm for 15 minutes, discarding the supernatant and air drying the pellet. After drying the pellet was resuspended in 15 µL of dH_2O and Spectrophotometry was done for the purified product to ensure enough concentration of the product for the subsequent reactions.

(iii). Purification of the sequencing reaction

After completion of the sequencing reaction, the reaction product was cleaned to remove excess dye and labeled nucleotide which could lead to the generation of suboptimal sequencing data.

Procedure:

The sequencing reaction product (10 µL) was mixed with 2.5 µL (2.5 mM) EDTA and 30 µL absolute ethanol. The mixture was mixed well, left at room temperature for 15 minutes, and then spun at 14000 rpm for 20 minutes. The supernatant was discarded, and the pellet was washed with 60 µL of 70% ethanol and centrifuged at 14000 rpm for 10 minutes. After washing, discarded the supernatant, the pellet was air dried, resuspended in 10 µL H1D1 formamide, and processed to obtain sequencing data through sanger sequencing, using Big Dye (Applied Biosciences) for verification of the PCR results.

2.2.3.2.3.2 Targeted next-generation sequencing (Tar-Seq)

Targeted next-generation sequencing was performed with DNA of two male pediatric ALL patients to explore the mutational status of crucial candidate genes playing role in vital signaling pathways. The reactions were performed through Thermo-fisher Ion S5TM Systems, Ion Torrent, using Ion AmpliSeq™ Library Kit 2.0 with Ion Ampliseq Cancer Hot Spot Panel 2, consisting of 50 genes (Annexure 3). The protocol provided with the panel was followed (Figure 2.2), details are given as follows.

(i). Amplification of Targets

The master mix was prepared for a 2X Single-primer pool panel in a well of 96-well PCR plate by adding, 4 µL of 5X Ion AmpliSeq™ HiFi Mix, and 10 µL of 2X Ion AmpliSeq™ Primer Pool, per each sample. Transferred 14 µL of the mix into individual wells and added 6 µL of DNA (1–100 ng). Added nuclease-free water to bring reaction volume to 20

µL where required and sealed the plate with a MicroAmp™ Clear Adhesive Film. After sealing, MicroAmp™ Compression Pad was placed on the plate, then loaded the plate into the thermal cycler. The amplification was carried out by holding the reaction at 99°C for 2 minutes to activate the enzyme, followed by denaturation at 99°C for 15 seconds annealing/extension for 4 minutes at 60°C for each cycle, and then holding the product at 10°C. After the completion of amplification, the plate containing the reaction mixture was centrifuged briefly to collect the contents at the bottom of the wells and removed the seal off the plate.

(ii). Partial digestion of the amplicons

Added 2 µL of FuPa Reagent to each amplified sample to bring the reaction volume to ~22 µL and sealed the plate with a clear adhesive film, vortexed thoroughly, then centrifuged to collect droplets. Placed a compression pad on the plate, loaded in the thermal cycler, then ran the samples at 50°C for ten minutes, 55°C for 10 minutes, 60°C for 20 minutes followed by a hold at 10°C for an hour. The plate was recentrifuged to collect the contents.

(iii). Adaptors ligation

Carefully removed the plate seal, and added 4 µL of switch solution, 2 µL of Adapters (Ion Torrent™ Dual Barcode Adapters), and 2 µL of DNA ligase in the listed order to each well containing the digested product. The reaction volume at this stage is ~30 µL in each well. The plate was sealed with a new MicroAmp™ Clear Adhesive Film, vortexed thoroughly, and centrifuged briefly. After centrifugation, a MicroAmp™ Compression pad was placed on the plate, loaded to the thermal cycler, and ran on then run the following program 22°C for 30 minutes, 68°C for 5 minutes, 72°C for 5 minutes followed by a hold at 10°C for 24 hours.

(iv). Library purification

After the adapter ligation, the plate was briefly centrifuged, and removed the plate seal, then added 45 µL (1.5 X sample volume) of Agencourt™ AMPure™ XP Reagent was to each library and mixed the bead suspension with the DNA thoroughly by pipetting. Incubated the mixture for 5 minutes at room temperature and placed the plate in a magnetic rack. After the solution became clear, removed the plate, and carefully discarded the supernatant without disturbing the pellet. Added 150 µL of freshly prepared 70% ethanol and carefully discarded the supernatant without disturbing the pellet. Repeated the same for a second wash. After washing, ethanol droplets were removed from the well, and beads were air-dried at room temperature for 5 minutes while keeping the plate in the magnet.

(v). Quantify the unamplified library by qPCR

Removed the plate with purified libraries from the plate magnet, then added 50 µL of Low TE to the pellet to disperse the beads, incubated at room temperature for at least 2 minutes, and then placed the plate on the magnet for at least 2 minutes. After incubation, taken 2 µL of supernatant, containing the library, and added to 198 µL of nuclease-free water to prepare a 100-fold dilution for quantification. For each sample, control, and standard, combined 20 µL of 2X Ion Library qPCR Master Mix and 2 µL of Ion Library TaqMan® Quantitation Assay, 20X to prepare reaction mixtures, and then mix thoroughly. After preparation of the reaction mixture, 11-µL aliquots were dispensed into the wells of a PCR plate, and added 9 µL of the diluted (1:100) Ion AmpliSeq™ library and 9 µL of each control dilution was to each well (two wells per sample as noted before), for a total reaction volume of 20 µL. The average concentration of the undiluted Ion AmpliSeq™ library was calculated by multiplying the concentration that is determined with qPCR by 100. Based on the calculated library concentration, determined the dilution that results in a concentration of ~100 pM and diluted the libraries to ~100 pM, then proceeded to template preparation. After template preparation, added 25 µL of each diluted library to the bottom of the appropriate Ion Chef™ Library sample tube and load the sample to the sequencer. Sequencing data was read through Thermofisher scientific-based Ion Reporter™ Software and interpreted.

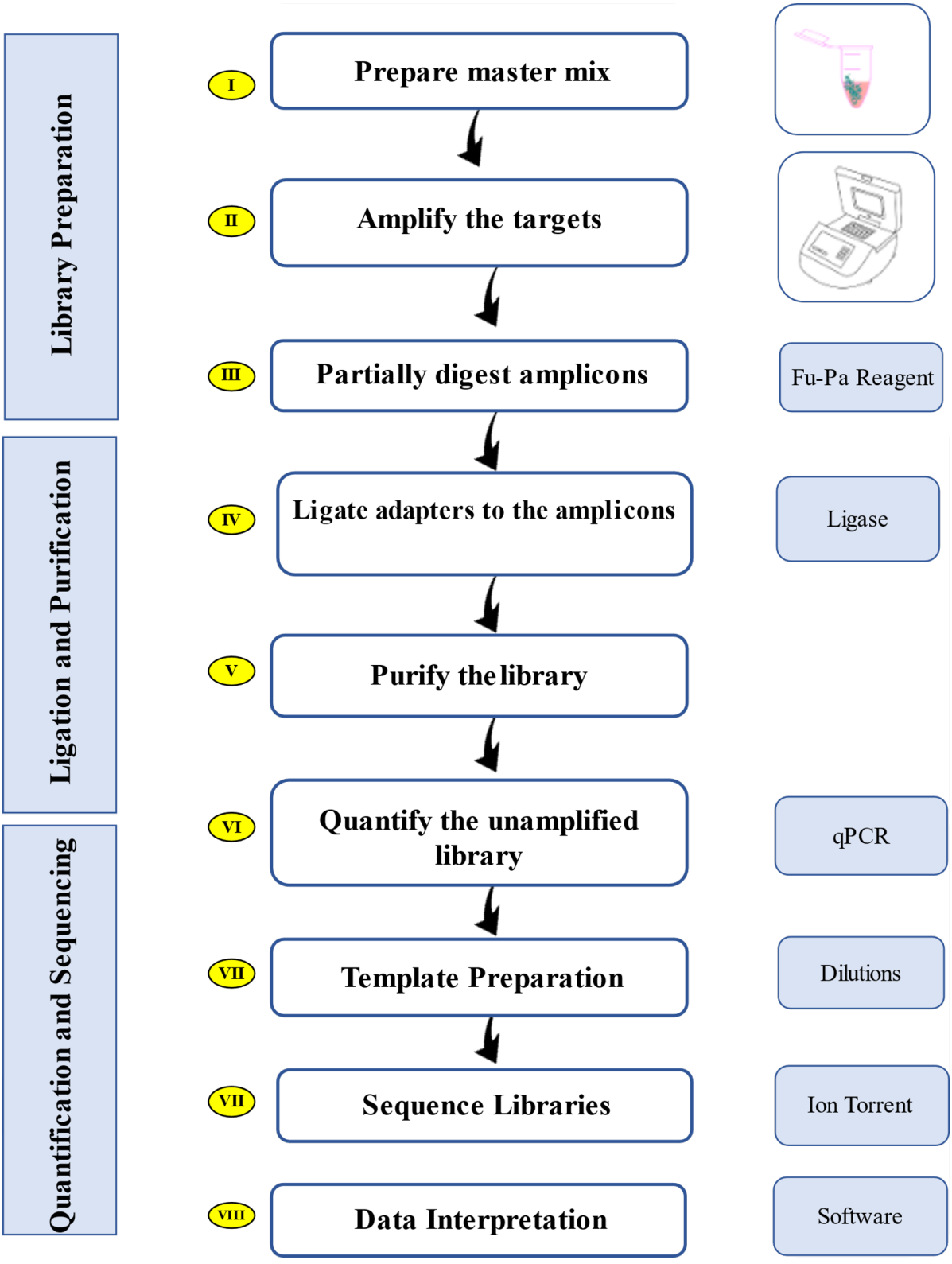


Figure 2.2: Schematic flow chart of targeted next-generation sequencing

2.2.3.2.3.3 Microarray assay

Microarray was performed on a male pediatric DNA, by utilizing Cyto-Scan Cytogenetics Suite by Thermo-fisher as per manual instructions. The major steps are labeled in the following flowchart. The data was analyzed through Chromosome Analysis Suite (ChAS) 4.3 software. Figure 2.3 is summarizing the overall steps performed during cytoscan microarray analysis.

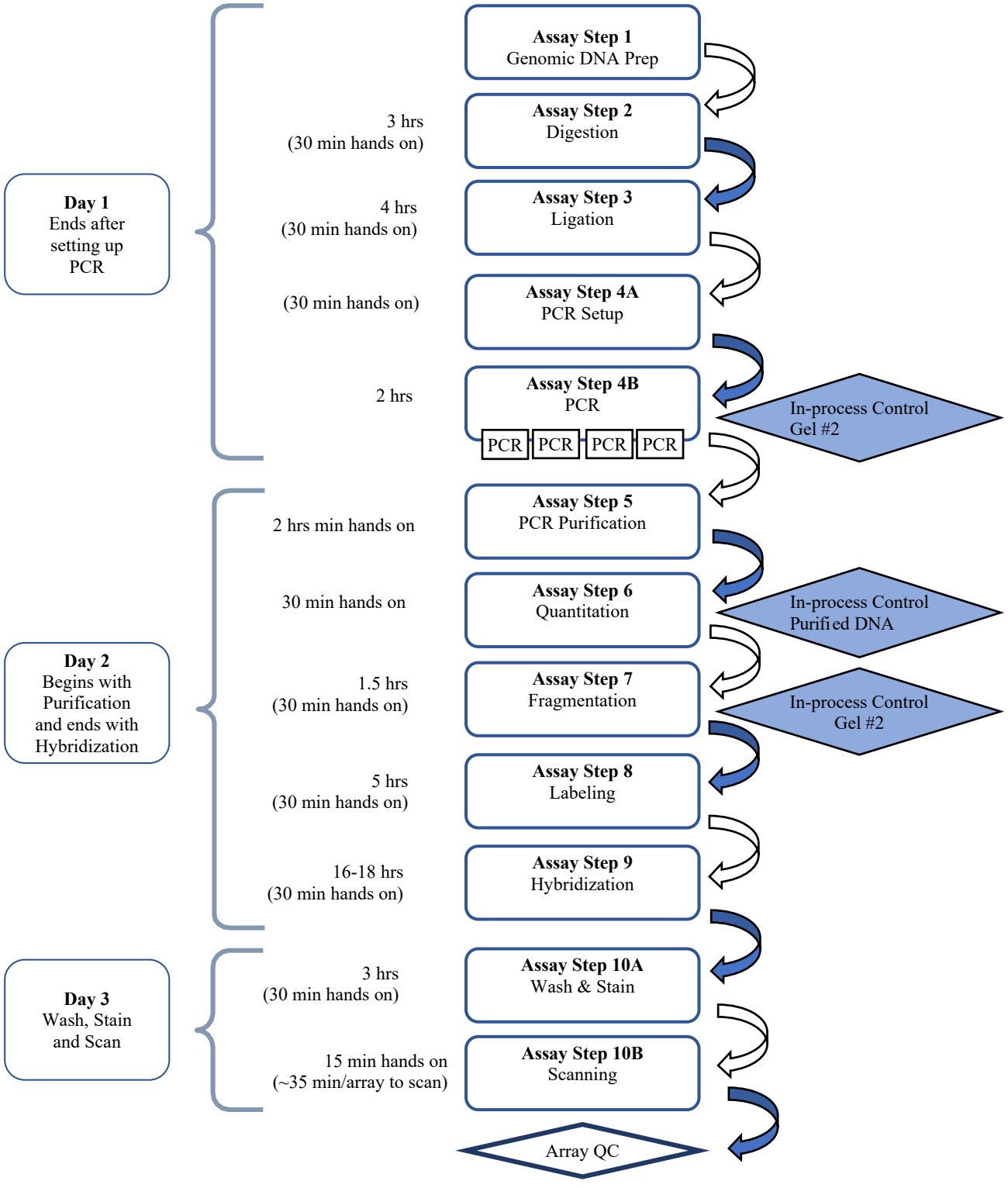


Figure 2.3: Schematic flow chart of the performed microarray scanning

2.2.3.3 *In silico* drug-protein interaction analysis

2.2.3.3.1 Ligand molecules preparation

The structural information of the eight selected compounds, 17F, AC-019, AC-4-130, Indirubin, Piceatannol, Pimozide, Stafib-2, Sulforaphane (Table 2.5), viz; described to pose an anti-anticancer potential, and are being used in preclinical studies/clinical trials for the treatment of onco-hematological disorders, were retrieved through literature survey (Su and David, 2000; Nam et al., 2012; Mistry et al., 2013; Cumaraswamy et al., 2014; Pinz et al., 2014; Elumalai et al., 2017; Wingelhofer et al., 2018; Brachet-Botineau et al., 2019). The 3D structures of the ligands were obtained from the PubChem database (https:// pubchem.ncbi.nlm.nih.gov/) in .sdf format or created through Chem draw Professional 15.0 where required. All the ligands with .sdf file extensions were converted via Discovery Studio Visualizer 2021(Biovia, 2021) into a .pdb file format. Energy minimization was done, and all ligands were converted and saved in.pdbqt format. The ligands structures optimized through the procedure were then used to perform molecular docking.

2.2.3.3.2 Protein macromolecules preparation

A recently elucidated, three-dimensional X-ray crystal structure structures of both normal STAT5B, and STAT5B ^{N642H} (RCSB-PDB. ID: 6MBW, 6MBZ, respectively), selected as molecular targets, were retrieved from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB-PDB) (www.rcsb.org) in pdb format. The crystal structure of STAT5B demonstrated that the protein exists in dimer form, composed of two identical monomers of the same conformation (chains A and B). Therefore, Chain A was selected for docking.

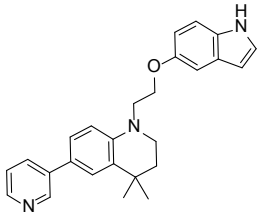
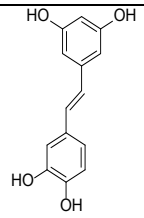
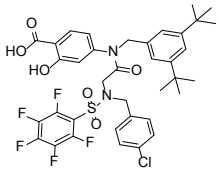
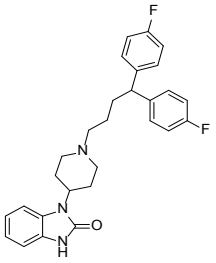
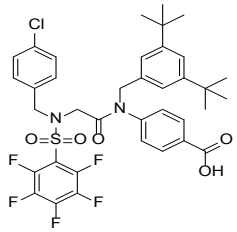
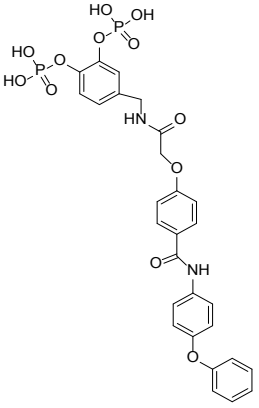
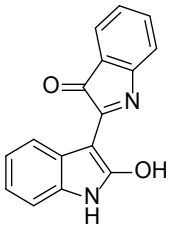
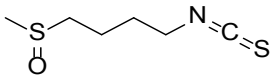
The water molecules and heteroatoms were removed from both protein structures through Discovery Studio Visualizer 2021(Biovia, 2021), and the proteins were saved in pdb format. Afterward, these pdb files were treated in MGL tools 1.5.7, polar hydrogen and Kollmann charges were added. Grid boxes were fixed around the protein (mainly covering SH2 domains), to define the active sites. The grid box center values for both targets 6MBW,

6MBZ molecules were kept at 60 Å × 60 Å × 60 Å centered at X:20.991, Y:1.055, Z:89.603 Å, separately. The prepared structures were exported in pdbqt format, the final docking form.

2.2.3.3.3 Ligand‑protein docking

The eight selected compounds were docked against the wildtype STAT5B and the mutated STAT5B^{N642H+}, as molecular targets. Docking was performed between ligand and protein molecules by using Auto Dock vina 1.23 (Eberhardt et al., 2021). The exhaustiveness value was set to 8, therefore each ligand was docked in 9 different ligand molecular poses. The accuracy of the docking was assured by re-docking to the co-crystal bound ligand. A cut-off of 2.0 Å was considered as the standard Root mean square deviation (RMSD) value, and the most favorable conformation was represented by the lowest free energy of binding (ΔG), with a higher number of interactions. After completion of docking, the ligand conformation, presenting the minimum binding energy (Kcal/mol), and RMSD value in the minimum range were selected, and all of the visualizations were again done through Discovery Studio Visualizer 2021 (Biovia, 2021).

Table 2.5: The list of IUPAC names and chemical structure and Structure of the selected ligands for STAT5B interaction

S#	Inhibitor	Structure	S#	STAT Inhibitor	Structure
1	**17F** CID*: NA 1-(2-((1H-Indol-5-yl)oxy)ethyl)-4,4-dimethyl-6-(pyridin-3-yl)- 1,2,3,4-tetrahydroquinoline (17f)		5	**Piceatannol** CID: 667639 4-[(*E*)-2-(3,5-dihydroxyphenyl) ethenyl] benzene-1,2-diol	
2	**AC-3-019** CID: NA 4-(2-((*N*-(4-chlorobenzyl)-2,3,4,5,6-pentafluorophenyl) sulfonamido)-*N*-(3,5-di-*tert*-butylbenzyl) acetamido)-2-hydroxybenzoic acid		6	**Pimozide** CID: 16362 3-[1-[4,4-bis(4-fluorophenyl)butyl] piperidin-4-yl]-1*H*-benzimidazol-2-one	
3	**AC-4-130** CID:154701370 4-[[2-[(4-chlorophenyl) methyl-(2,3,4,5,6-pentafluorophenyl) sulfonyl amino] acetyl]-[(3,5-ditert butylphenyl) methyl] amino]benzoic acid		7	**Stafib-2** CID: 129896883 [4-[[[2-[4-[(4phenoxyphenyl) carbamoyl] phenoxy]acetyl]amino] methyl] -2-phosphonooxyphenyl] dihydrogen phosphate	
4	**Indirubin** CID: 10177 2-(2-hydroxy-1*H*-indol-3-yl)indol-3-one		8	**Sulforaphane** CID: 5350 1-isothiocyanato-4-methylsulfinylbutane	

CID*: PubChem Identifier

2.2.4 Statistical analysis

The results for epidemiological and genetic analysis were individually analyzed. Descriptive statistical methods were used to summarize the clinical characteristics, demographic, and lifestyle-related factors of the study participants. We determined potential demographic, lifestyle, and clinical risk factors by using univariate logistic regression, comparing each parameter between patients and controls.

The comparison was performed on three distinct levels: (i). Compared overall numbers of patients and controls against each variable, (ii). Categorizing both patients and controls in age groups (<1, 1.1-5, 5.1-15, 15.1-39, >39 years; Table 1) and then performed a group-to-group comparison of patients and controls against each variable, (iii). To identify the high-risk age group within the patients, the age groups within the patients' cohort were compared with one another. In age, group-wise comparison, a group vs a group and a group vs all other groups assessment was implied, separately for every variable, in both (ii) and (iii) levels of comparison. The percentage of the missing data was excluded in the respective statistical analysis to avoid its interference with the assessment. Odds ratios with 95% confidence intervals were calculated to identify the presence and strength of the association. A p-value of < 0.05 was considered statistically significant for all of the tests.

The blood picture variables of the patients (WBCs, RBCs, hemoglobin, platelets, lymphocytes %, and neutrophils%) are mentioned as mean ± SD. Two-tailed T-Test for independent variables, for the sample with unequal variance, was used to compare the means of the patients and controls. Non-significant variables were removed in a stepwise approach and significant predictors were thoroughly explained. The demographic and data clinical data of the samples used for the genetic study were analyzed through descriptive statistics with groupwise logistic regression. The descriptive statistics and T-Test statistics were calculated in Excel 365. The calculation of the odd ratios and logistic regression analysis was done through STATA 15.0 software and the online Medicalc calculator, available at https://www.medcalc.org/calc/odds_ratio.php.

CHAPTER 3

RESULTS

3. Results

The result section is divided into two major divisions: (i) epidemiological study, and (ii) molecular genetic analysis

3.1 Epidemiological analysis

In this division, the epidemiology of leukemia patients has been analyzed in three dimensions, viz., physical and demographic characteristics, clinic-pathological attributes, and lifestyle-related risk factors. This division is addressed in two major subsections: (i) Clinical and paraclinical variables, and (ii) Demographic and lifestyle-related variables.

3.1.1 Clinical and paraclinical variables

This section deals with the incidence of leukemia, and its major subtypes, along with a detailed evaluation of paraclinical variables like hematological variables via thorough comparison with the same variables in control samples.

3.1.1.1 Sampling characteristics and overall percentage distribution of study population

A total of 1500 samples were collected for this study, comprised of 616 leukemia patients and 884 age, gender, and ethnicity-matched controls. The 616 leukemic samples were collected from the patients, visiting tertiary care hospitals of Pakistan, located in Islamabad and Peshawar, with proper consent (Appendix A). The demographic, anthropometric, and clinical data were obtained on a proper questionnaire (Appendix B). Questionaries were filled out from all patients via direct interview. Detailed clinical and paraclinical reports were obtained for 92% of the patients remaining who were not available in the facilities. Among all, the data of 594 patients were included in the analysis and the remaining 22 patients were excluded due to data unavailability.

Table 3.1: Age group-based gender-wise categorization of the study population (n=1478), and type-wise distribution of leukemia patients (n=594). ALL = Acute Lymphocytic Leukemia, CLL= Chronic Lymphocytic Leukemia, AML= Acute Myeloid Leukemia, CML = Chronic Myeloid Leukemia.

Age Group (Years)	Gender	Patients									Controls	Net Total
		Leukemia subtype				Disease origin		Disease type		Total	Total	
		ALL	CLL	AML	CML	Lymphoid	Myeloid	Acute	Chronic			
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	
<1 Infants	Male	4 (0.67)	0 (0)	0 (0)	1 (0.17)	4 (80)	1 (20)	4 (80)	1 (20)	5 (0.84)	3 (0.34)	8
	Female	2 (0.34)	0 (0)	0 (0)	0 (0)	2 (100)	0 (0)	2 (100)	0 (0)	2 (0.34)	7 (0.79)	9
1.1-5 Pre-school	Male	75 (12.63)	0 (0)	4 (0.67)	1 (0.17)	75 (93.75)	5 (6.25)	79 (98.75)	1 (1.25)	80 (13.47)	53 (6)	133
	Female	33 (5.56)	0 (0)	2 (0.34)	1 (0.17)	33 (91.67)	3 (8.33)	35 (97.22)	1 (2.78)	36 (6.06)	21 (2.38)	57
5.1-15 Children	Male	103 (17.34)	0 (0)	12 (2.02)	2 (0.34)	103 (88.03)	14 (11.97)	115 (98.29)	2 (1.71)	117 (19.7)	70 (7.92)	187
	Female	47 (7.91)	0 (0)	6 (1.01)	1 (0.17)	47 (87.04)	7 (12.96)	53 (98.15)	1 (1.85)	54 (9.09)	47 (5.32)	101
15.1-39 AYAs*	Male	58 (9.76)	1 (0.17)	15 (2.53)	23 (3.87)	59 (60.82)	38 (39.18)	73 (75.26)	24 (24.74)	97 (16.33)	385 (43.55)	482
	Female	14 (2.36)	0 (0)	16 (2.69)	21 (3.54)	14 (27.45)	37 (72.55)	30 (58.82)	21 (41.18)	51 (8.59)	118 (13.35)	169
39.1 & > Aged	Male	15 (2.53)	22 (3.7)	12 (2.02)	52 (8.75)	37 (36.63)	64 (63.37)	27 (26.73)	74 (73.27)	101 (17)	112 (12.67)	213
	Female	1 (0.17)	7 (1.18)	12 (2.02)	31 (5.22)	8 (15.69)	43 (84.31)	13 (25.49)	38 (74.51)	51 (8.59)	68 (7.69)	119
Total		352 (59.26)	30 (5.05)	79 (13.3)	133 (22.39)	382 (64.31)	212 (35.69)	431 (72.56)	163 (27.44)	594 (100)	884 (100)	1478

*AYAs: Adolescents and young adults, # Out of the total 616 patients, the data 22 were excluded in the subsequent analysis

3.1.1.2 Age and gender-wise distribution

Among 616 patients, 422 (67%) were males and 203 (33%) were females, with a male; female ratio of 2:1. The patients were divided into five age groups, perceived as concerned with cancer etiology, named infants (<1 year), preschoolers (1.1-5 years), children (5.1-15 years), adolescen The same male dominance is observed in each leukemia subtype in every age group, except for AYAs and aged in AML, where there is no male dominance in our data. Also, in CML AYAs have negligible male dominance. AYAs show a unique pattern in gender distribution (Table 3.1).ts and young adults (15.1-39 years), and aged (>39. The highest occurrence of leukemia is seen in children of 5 to 15 years of age (28.79%), closely followed by participants >39 years of age (25.59%) (Table 3.1, Figure 3.1). Overall, the highest percentage of leukemia is recorded in children from age of a single month to 15 years (50%), which is almost double that of any other group (Table 3.1, Figure 3.2).

The control group was comprised of 1.13% of infants below 1 year, 8.37% were in the age range 1.1 - 5 years old, and 13.4% of the controls were from the age group 5.1-15 years old. Among all, 56.9% of the controls were in the age range 15.1-39 years, and 20.36 % of the patients were from the age group 39.1 & older. About 70% of the controls were male and 30% were female (Table 3.1).

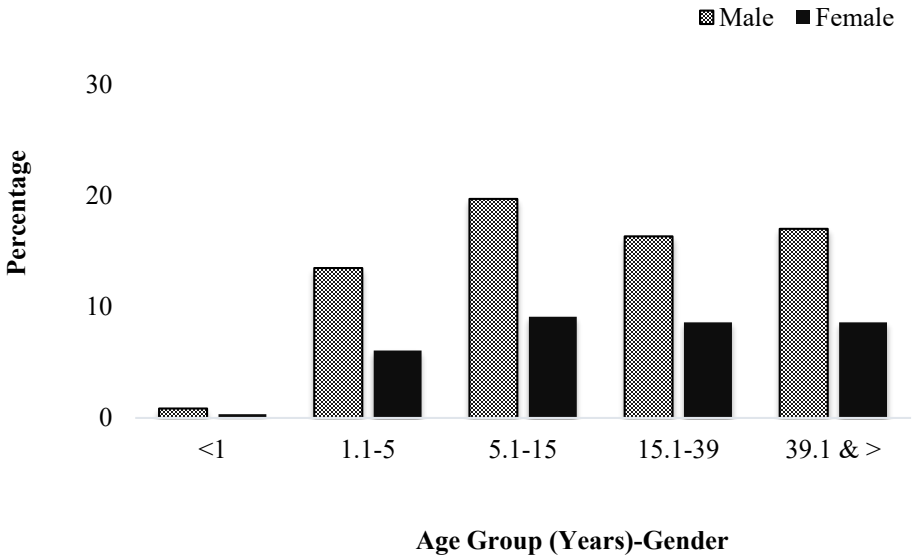


Figure 3.1: Gender-wise distribution of leukemia patients in different age group (n=594)

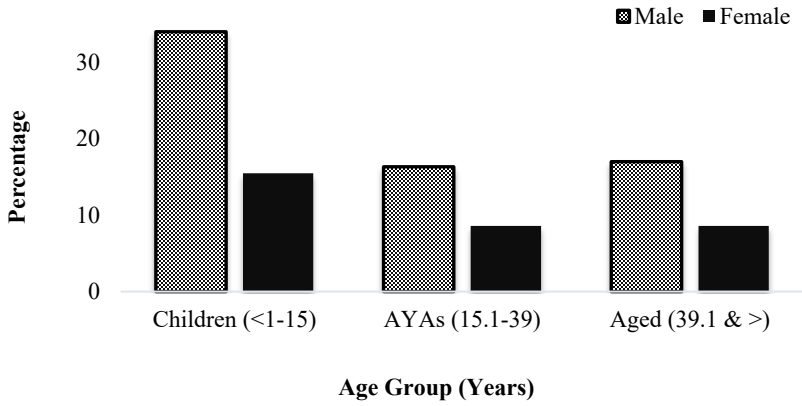


Figure 3.2: Gender-wise distribution of leukemia patients against biological status (n=594)

3.1.1.3 Leukemia subtypes

ALL is the most prevalent (59%) , followed by CML (22%), AML (14%) and CLL (5%) the least prevalent.. AML (13%) is more prevalent in AYAs and aged, while CLL (5%) is prevalent only in aged individuals, with a male-to-female ratio of 3:1. ALL is the highest affecting type in AYAS in comparison to other leukemia types. In myeloid leukemia particularly in AML AYAs, there is no male dominance, while in CML it is negligible (Table 3.1, Figures. 3.3 & 3.5).

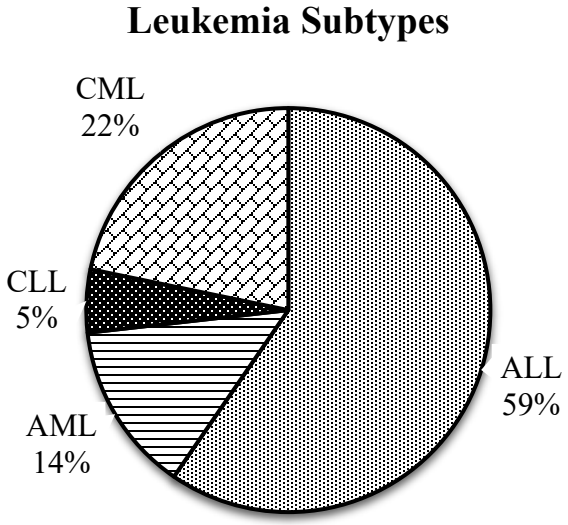


Figure 3.3: Distribution of leukemia subtypes in the Pakistani population

The prevalence of leukemia subtypes in the Pakistani population varied in percentage distribution. Acute lymphoblastic leukemia (ALL) emerged as the most prevalent subtype, notably affecting children aged 1 to 15 years. Myeloid leukemias, including chronic myeloid leukemia (CML) and acute myeloid leukemia (AML), were more commonly observed in adults aged 15.1 to 39 years and those older than 39 years. Chronic lymphocytic leukemia (CLL) constituted the least common subtype, predominantly occurring in adults over the age of 39 in the study population.

.

(i). Disease nature-wise distribution of leukemia patients

Overall, 73% of the patients were suffering from acute leukemia, mainly consisting of children and AYAs. While chronic leukemia comprised 27% of the total patients and was common in aged individuals (Table 3.1). In AYAs, acute leukemia is a highly dominant disease form, this dominancy is due to the male patients, which are three times more affected by acute than chronic leukemia, while females showed the same prevalence for both acute and chronic leukemias. Concerning gender dominance in each class, male dominance is observed overall, however, a unique pattern has been observed in AYAs in chronic leukemia, where there is no evident male (4%) to female (3.54%) dominance (Table 3.1, Figure 3.4).

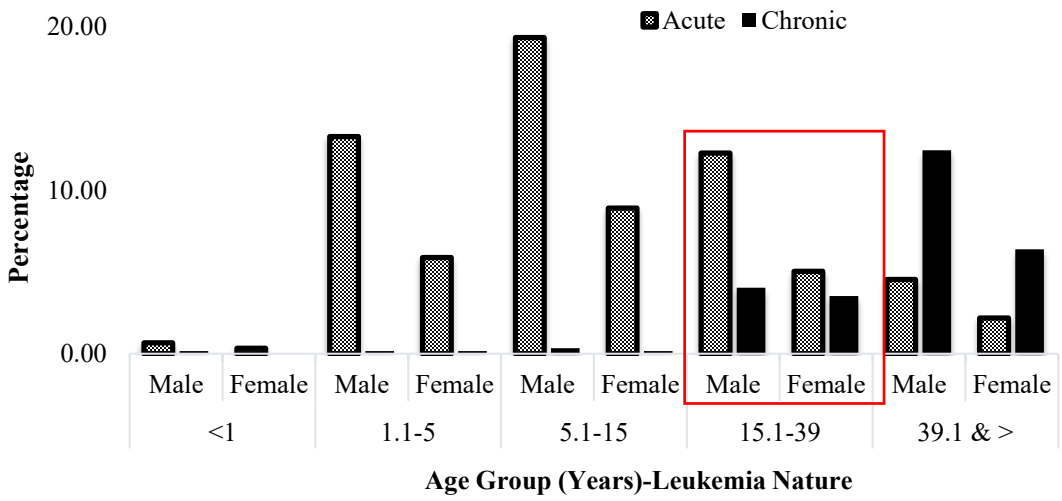


Figure 3.4: Overall gender-wise distribution (%) of acute/chronic leukemia patients in different age group (n=594).

(i). Disease origin-wise distribution of leukemia patients

A very high proportion (64.3%) of the patients were suffering from lymphoid leukemia compared to myeloid leukemia (35.69%). (Table 3.1). The largest portion (25%) of the children and preschoolers (18.19%) were effected from lymphoid leukemia; while myeloid leukemia was more prevalent (18.1%) in aged patients. In lymphoids, 92% are ALL, and 7.8% CLL. In myeloid leukemia, 62.76% are CML and 37.26% are AML. Lymphoid leukemia is most prevalent in 5.1-15 years age group. CLL was only seen in individuals with >39 years of age, 73% of which are male patients. In AYAs, the lymphoid and myeloid prevalence is almost the same, but in males, lymphoid leukemia is 1.5 times more prevalent, while in females' myeloid leukemia is 2.6 times more common than lymphoid. Myeloid leukemia is most prevalent in aged individuals (18%). The highest frequency of AML was observed in AYAs, and of CML in aged ones. With the increase in age, there is a gradual decrease in lymphoid and an increase in myeloid leukemia in both males and females (Table 3.1, Figure 3.5).

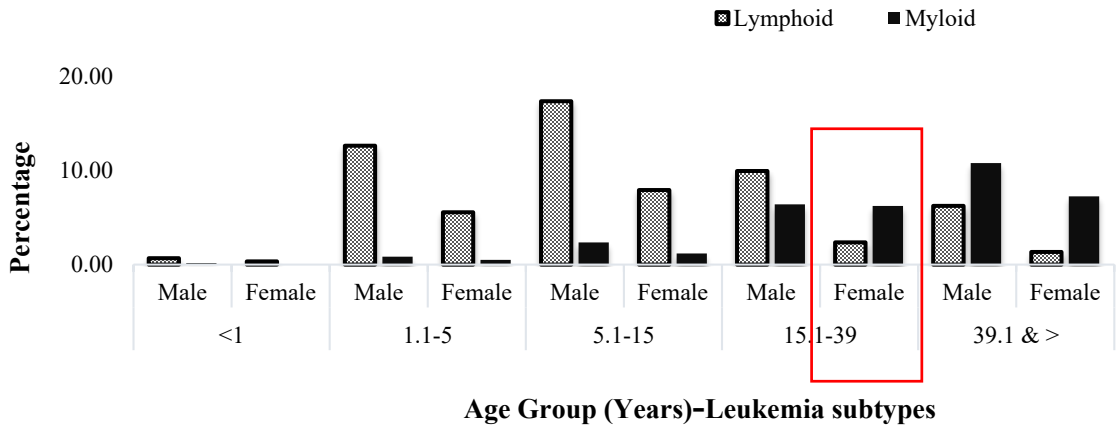


Figure 3.5: Gender-wise distribution of leukemia patients in different age group (years) and disease origin (n=594)

(iii). Comparison of leukemia subtypes prevalence in both myeloid and lymphoid leukemia

ALL has a direct relation with growing age in children as it increases in prevalence as the age advances, reaching its peak at the age group of 5 to 15 years, after that age, the effect is reversed, and ALL prevalence fell and reached its minimum in the aged group. AML also has a direct increase in peak with growing age, reaching its maximum in AYAs (14.62%), and then a slight drop in the aged class. CML is highly frequent in AYAs but its maximum peak is in aged individuals, However, CLL (7.59%) appears only in the aged group (Table 3.1; Figure 3.6).

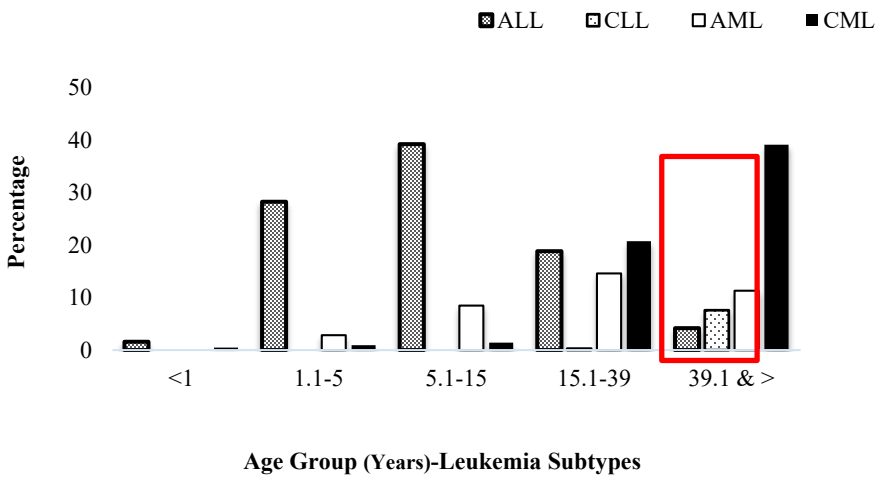


Figure 3.6: Comparison of percentages of leukemia subtypes prevalence in both myeloid and lymphoid (n=594)

3.1.1.4 Paraclinical and hematological variables

Blood biochemistry reports provided the number of key cells circulating in the human blood, along with blast cells. Blast cells are poorly differentiated, immature blood cells, which can differentiate into any cell type. Accumulation of abnormal blast cells, also known as leukemic blasts (>20% blast cells in bone marrow or peripheral blood), is a characteristic of leukemia. Blast cell percentages obtained from patients' suggested that 71% of the cases had raised blast cell percentage of >50%, while 29 % had blast cell frequency of <50%, in which most of the cases have more than 20% of blast cells (Table 3.2).

The average of hematological findings was almost the same across all of the leukemia subtypes, however, there were significant differences in comparison to controls ($P^* \leq 0.0016$; Table 3.3A). The mean value for blast in patients was 63.8±27.96% (SD). The mean value for white blood cells (WBCs) was 41.95±85.03, red blood cells (RBCs) was 3.45 ±1.14, and hemoglobin was 9.47 ±3.08, while for controls the values were 8.44 ±6.36, 5.02 ±0.72, and 13.78 ±2.43, respectively. The mean value of platelet count for cases was 180.86 ±204.59, for lymphocytes was 37.64±28.22%, and for neutrophils the mean value was 31.23±25.99%. In controls, the mean for lymphocytes was 33.24 ±10.46% and for neutrophils, % was 55.58 ±12.28 (Table 3.3A & B)

Table 3.2: Hematological attributes of the patients (n=616) and controls (n=884)

Variable*	Patients	Controls	Total	Variable	Patients	Controls	Total
	N (%)	N (%)			N (%)	N (%)	
WBC Status (10^9/L)				**Lymphocytes (%)**			
Normal	182 (31.99)	618 (69.91)	800	Normal	135 (21.92)	544 (61.54)	679
Increment	228 (40.07)	84 (9.5)	312	Increment	152 (24.68)	96 (10.86)	248
Decrement	159 (27.94)	33 (3.73)	192	Decrement	160 (25.97)	93 (10.52)	253
Unavailable	0 (0)	149 (16.86)	196	Unavailable	169 (27.44)	151 (17.08)	320
Total	569 (100)	884 (100)	1500	Total	616 (100)	884 (100)	1500
RBC Status (10^9/L)				**Neutrophils (%) Status**			
Normal	175 (28.41)	510 (57.69)	685	Normal	119 (19.32)	533 (60.29)	652
Increment	12 (1.95)	145 (16.4)	157	Increment	33 (5.36)	57 (6.45)	90
Decrement	362 (58.77)	80 (9.05)	442	Decrement	270 (43.83)	143 (16.18)	413
Unavailable	67 (10.88)	149 (16.86)	216	Unavailable	194 (31.49)	151 (17.08)	345
Total	616 (100)	884 (100)	1500	Total	616 (100)	884 (100)	1500
Hemoglobin Status (g/dL)				**Blast cell ratio**			
Normal	143 (23.21)	571 (64.59)	714	≤50%	86 (13.96)		86
Increment	2 (0.32)	4 (0.45)	6	>50%	210 (34.09)		210
Decrement	425 (68.99)	161 (18.21)	586	Unavailable	320 (51.95)		320
Unavailable	46 (7.47)	148 (16.74)	194	Total	616 (100)		616
Total	616 (100)	884 (100)	1500	**Alkaline Phosphates**			
Platelets Status (1000/µL)				Normal	19 (3.08)		19
Normal	185 (30.03)	645 (72.96)	830	Increment	63 (10.23)		63
Increment	58 (9.42)	62 (7.01)	120	Decrement	0 (0)		0
Decrement	323 (52.44)	29 (3.28)	352	Unavailable	534 (86.69)		534
Unavailable	50 (8.12)	148 (16.74)	198	Total	616 (100)		616
Total	616 (100)	884 (100)	1500				

Out of the total 616 patients, the data 22 were excluded in the subsequent analysis

*Only noteworthy attributes are depicted, WBCs, White blood cells; RBCs, Red blood cells.

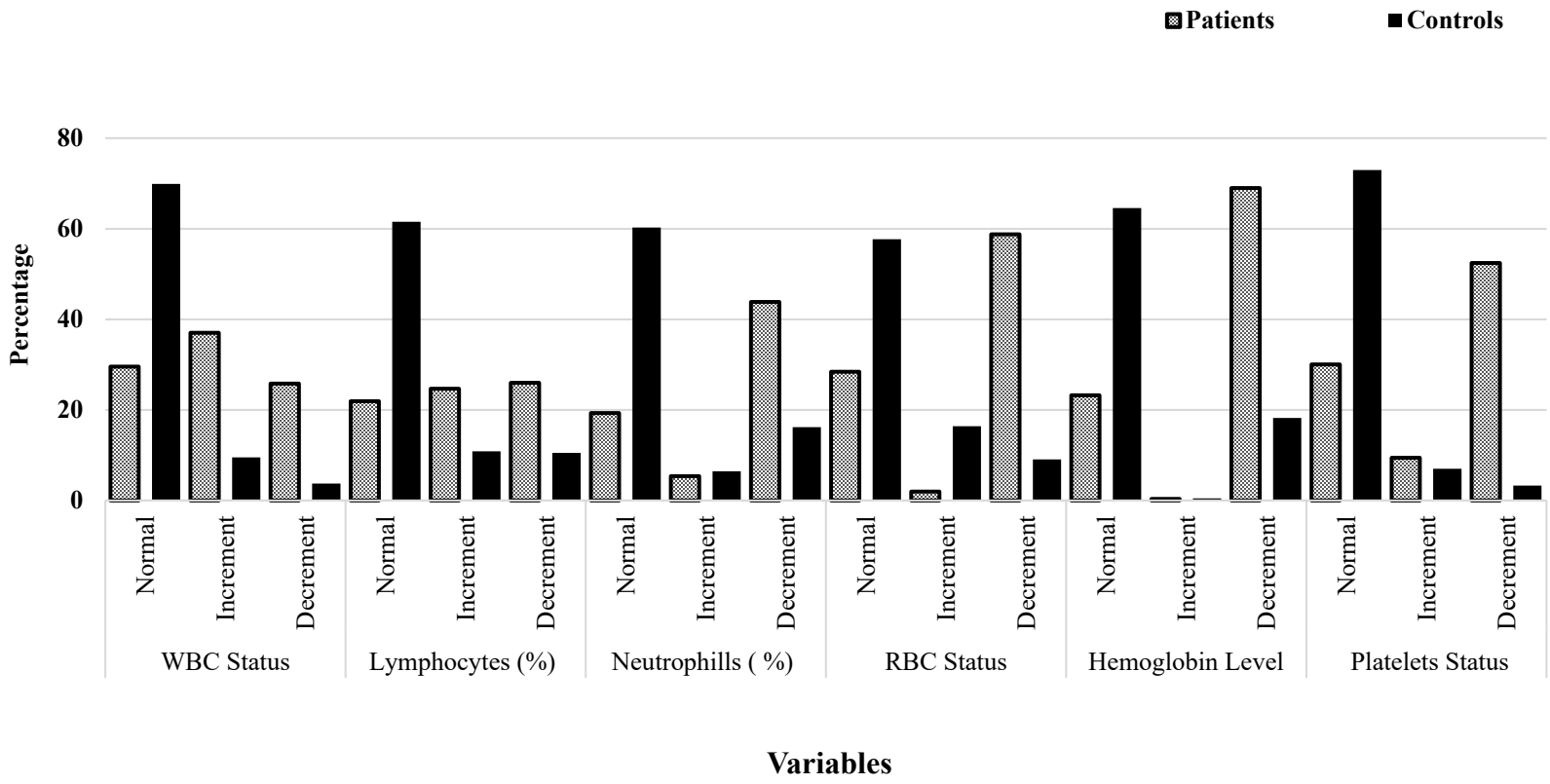


Figure 3.7: Bar graphical representation of hematological parameters in patients and controls

Table 3.3A: Overall subtype-wise (mean ± standard deviation) of patients and controls hematological variables

Characteristics	ALL (n=367) (Mean+ SD)	CLL (n=31) (Mean+ SD)	Lymphoid (n=31)	AML (n=84) (Mean+ SD)	CML (n=31) (Mean+ SD)	Myeloid (n=134) (Mean+ SD)	Total (n=594) (Mean+ SD)	Controls (n=884) (Mean+ SD)	P-value
WBCs (10^9/L)	41.95± 85.03	42.34 ±87.28	41.94±85.6	41.75 ±85.12	44.59 ±88.23	42.61±85.8	41.95±85.03	8.44 ±6.36	< 0.000
RBCs (10^6/µL)	3.45 ±1.14	3.52 ±1.17	3.44±1.13	3.46 ±1.15	3.49 ±1.14	3.45±1.14	3.45 ±1.14	5.02 ±0.72	< 0.000
Hemoglobin (g/dL)	9.47 ± 3.08	9.59 ±3.06	9.46±3.08	9.48 ±3.10	9.57 ±3.06	9.48±3.09	9.47 ±3.08	13.78 ±2.43	< 0.000
Platelets (1000/µL)	180.86 ±204.59	185.31 ± 209.17	180±203	180.10 ±203.46	185.77 ±207.43	180±203	180.86 ±204.59	293.03 ±96.42	< 0.000
Lymphocytes %	37.64 ± 28.22	35.46 ±26.60	37.61±28.2	37.17 ±27.89	35.78 ±27.15	36.90±27.8	37.64±28.22	33.24 ±10.46	< 0.0016
Neutrophils %	31.23 ±25.99	31.15 ±25.50	31.24±25.9	31.37 ±26.07	32.04 ±26.07	31.54±26.03	31.23±25.99	55.58 ±12.28	< 0.000
Blast cells ratio	-	-	-	-	-	-	63.8±27.96	-	-

WBCs, White blood cells; RBCs, Red blood cells.

Table 3.3B: Overall age and gender-wise (mean ± standard deviation) of patient's hematological variables (n=594)

Characteristics	Children (n=255)		Adults (n=339)		Total
	Male (n=174) (Mean+ SD)	Female (n=81) (Mean+ SD)	Male (n=226) (Mean+ SD)	Female (n=113) (Mean+ SD)	Patients (n=594) (Mean+ SD)
WBCs (10^9/L)	40.28±83.74	39.66 ±83.12	42.23 ±85.40	42.86±85.89	41.95±85.03
RBCs (10^6/µL)	3.45 ±1.14	3.46 ±1.15	3.46 ±1.15	3.45 ±1.15	3.45 ±1.14
Hemoglobin (g/dL)	9.48 ±3.07	9.48 ± 3.09	9.48 ±3.10	9.47 ±3.11	9.47 ±3.08
Platelets (1000/µL)	180.60 ± 205.46	181.32 ± 206.16	180.68 ±203.58	180.67 ± 203.48	180.86 ±204.59
Lymphocytes %	37.98 ±28.20	37.61 ±27.95	37.04 ±27.85	36.78 ± 27.84	37.64±28.22
Neutrophils %	31.16 ±26.02	31.57 ±26.09	31.54 ± 26.05	31.55 ± 26.06	31.23±25.99
Blast cells ratio	-	-	-	-	63.8±27.96

WBCs, White blood cells; RBCs, Red blood cell

3.1.1.4.1 Age group and gender-wise distribution of hematological variables

The hematological variables were subcategorized based on age groups and gender to obtain a further detailed comparison of patients and controls. A detailed comparison is given in Table 3.4 (A & B).

Table 3.4A: Summary of age group and gender-wise distribution of hematological variables (Patients, 594; controls, 884)

Variables		Patients					Controls					Net Total
Age Group	Gender	Normal	Increment	Decrement	No info	Total	Normal	Increment	Decrement	No info	Total	
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
WBC Status												
<1	Male	0 (0)	5 (0.84)	0 (0)	0 (0)	5 (0.84)	1 (0.11)	1 (0.11)	0 (0)	1 (0.11)	3 (0.34)	8
	Female	0 (0)	0 (0)	1 (0.17)	1 (0.17)	2 (0.34)	3 (0.34)	2 (0.23)	1 (0.11)	1 (0.11)	7 (0.79)	9
1.1-5	Male	22 (3.7)	27 (4.55)	30 (5.05)	1 (0.17)	80 (13.47)	25 (2.83)	4 (0.45)	0 (0)	24 (2.71)	53 (6)	133
	Female	10 (1.68)	10 (1.68)	15 (2.53)	1 (0.17)	36 (6.06)	13 (1.47)	2 (0.23)	1 (0.11)	5 (0.57)	21 (2.38)	57
5.1-15	Male	36 (6.06)	31 (5.22)	47 (7.91)	3 (0.51)	117 (19.7)	43 (4.86)	3 (0.34)	5 (0.57)	19 (2.15)	70 (7.92)	187
	Female	10 (1.68)	18 (3.03)	21 (3.54)	5 (0.84)	54 (9.09)	33 (3.73)	2 (0.23)	0 (0)	12 (1.36)	47 (5.32)	101
15.1-39	Male	32 (5.39)	37 (6.23)	20 (3.37)	8 (1.35)	97 (16.33)	290 (32.81)	15 (1.7)	11 (1.24)	69 (7.81)	385 (43.55)	482
	Female	17 (2.86)	21 (3.54)	10 (1.68)	3 (0.51)	51 (8.59)	72 (8.14)	35 (3.96)	4 (0.45)	7 (0.79)	118 (13.35)	169
39.1 & >	Male	40 (6.73)	49 (8.25)	10 (1.68)	2 (0.34)	101 (17)	89 (10.07)	8 (0.9)	8 (0.9)	7 (0.79)	112 (12.67)	213
	Female	15 (2.53)	30 (5.05)	5 (0.84)	1 (0.17)	51 (8.59)	49 (5.54)	12 (1.36)	3 (0.34)	4 (0.45)	68 (7.69)	119
Total		182 (30.64)	228 (38.38)	159 (26.7)	25 (4.21)	594 (100)	618 (69.91)	84 (9.5)	33 (3.73)	149 (16.8)	884 (100)	1478
RBC Status												
<1	Male	1 (0.17)	0 (0)	4 (0.67)	0 (0)	5 (0.84)	2 (0.23)	0 (0)	0 (0)	1 (0.11)	3 (0.34)	8
	Female	1 (0.17)	0 (0)	0 (0)	1 (0.17)	2 (0.34)	4 (0.45)	0 (0)	2 (0.23)	1 (0.11)	7 (0.79)	9
1.1-5	Male	18 (3.03)	1 (0.17)	59 (9.93)	2 (0.34)	80 (13.47)	24 (2.71)	0 (0)	5 (0.57)	24 (2.71)	53 (6)	133
	Female	3 (0.51)	2 (0.34)	30 (5.05)	1 (0.17)	36 (6.06)	8 (0.9)	6 (0.68)	2 (0.23)	5 (0.57)	21 (2.38)	57
5.1-15	Male	34 (5.72)	3 (0.51)	70 (11.78)	10 (1.68)	117 (19.7)	39 (4.41)	10 (1.13)	2 (0.23)	19 (2.15)	70 (7.92)	187
	Female	10 (1.68)	1 (0.17)	36 (6.06)	7 (1.18)	54 (9.09)	32 (3.62)	0 (0)	3 (0.34)	12 (1.36)	47 (5.32)	101
15.1-39	Male	41 (6.9)	1 (0.17)	42 (7.07)	13 (2.19)	97 (16.33)	218 (24.66)	97	1 (0.11)	69 (7.81)	385 (43.55)	482
	Female	8 (1.35)	0 (0)	39 (6.57)	4 (0.67)	51 (8.59)	59 (6.67)	4 (0.45)	48 (5.43)	7 (0.79)	118 (13.35)	169
39.1 & >	Male	47 (7.91)	4 (0.67)	44 (7.41)	6 (1.01)	101 (17)	81 (9.16)	21 (2.38)	3 (0.34)	7 (0.79)	112 (12.67)	213
	Female	12 (2.02)	0 (0)	38 (6.4)	1 (0.17)	51 (8.59)	43 (4.86)	7 (0.79)	14 (1.58)	4 (0.45)	68 (7.69)	119
Total		175 (29.46)	12 (2.02)	362 (60.9)	45 (7.58)	594 (100)	510 (57.69)	145 (16.4)	80 (9.05)	149 (16.8)	884 (100)	1478

Hemoglobin Level												
<1	Male	0 (0)	0 (0)	5 (0.84)	0 (0)	5 (0.84)	2 (0.23)	0 (0)	0 (0)	1 (0.11)	3 (0.34)	8
	Female	0 (0)	0 (0)	1 (0.17)	1 (0.17)	2 (0.34)	1 (0.11)	0 (0)	5 (0.57)	1 (0.11)	7 (0.79)	9
1.1-5	Male	11 (1.85)	0 (0)	68 (11.45)	1 (0.17)	80 (13.47)	7 (0.79)	0 (0)	22 (2.49)	24 (2.71)	53 (6)	133
	Female	5 (0.84)	1 (0.17)	29 (4.88)	1 (0.17)	36 (6.06)	6 (0.68)	0 (0)	10 (1.13)	5 (0.57)	21 (2.38)	57
5.1-15	Male	27 (4.55)	1 (0.17)	87 (14.65)	2 (0.34)	117 (19.7)	34 (3.85)	1 (0.11)	16 (1.81)	19 (2.15)	70 (7.92)	187
	Female	11 (1.85)	0 (0)	39 (6.57)	4 (0.67)	54 (9.09)	26 (2.94)	0 (0)	9 (1.02)	12 (1.36)	47 (5.32)	101
15.1-39	Male	29 (4.88)	0 (0)	60 (10.1)	8 (1.35)	97 (16.33)	309 (34.95)	2 (0.23)	5 (0.57)	69 (7.81)	385 (43.55)	482
	Female	5 (0.84)	0 (0)	42 (7.07)	4 (0.67)	51 (8.59)	47 (5.32)	0 (0)	64 (7.24)	7 (0.79)	118 (13.35)	169
39.1 & >	Male	46 (7.74)	0 (0)	53 (8.92)	2 (0.34)	101 (17)	99 (11.2)	1 (0.11)	5 (0.57)	7 (0.79)	112 (12.67)	213
	Female	9 (1.52)	0 (0)	41 (6.9)	1 (0.17)	51 (8.59)	40 (4.52)	0 (0)	25 (2.83)	3 (0.34)	68 (7.69)	119
Total		143 (24.07)	2 (0.34)	425 (71.5)	24 (4.04)	594 (100)	571 (64.59)	4 (0.45)	161 (18.21)	148 (16.7)	884 (100)	1478
Platelets Status												
<1	Male	2 (0.34)	1 (0.17)	2 (0.34)	0 (0)	5 (0.84)	2 (0.23)	0 (0)	0 (0)	1 (0.11)	3 (0.34)	8
	Female	0 (0)	0 (0)	1 (0.17)	1 (0.17)	2 (0.34)	3 (0.34)	2 (0.23)	1 (0.11)	1 (0.11)	7 (0.79)	9
1.1-5	Male	18 (3.03)	5 (0.84)	55 (9.26)	2 (0.34)	80 (13.47)	17 (1.92)	11 (1.24)	1 (0.11)	24 (2.71)	53 (6)	133
	Female	8 (1.35)	1 (0.17)	26 (4.38)	1 (0.17)	36 (6.06)	10 (1.13)	5 (0.57)	1 (0.11)	5 (0.57)	21 (2.38)	57
5.1-15	Male	30 (5.05)	9 (1.52)	75 (12.63)	3 (0.51)	117 (19.7)	39 (4.41)	9 (1.02)	3 (0.34)	19 (2.15)	70 (7.92)	187
	Female	9 (1.52)	5 (0.84)	36 (6.06)	4 (0.67)	54 (9.09)	29 (3.28)	6 (0.68)	0 (0)	12 (1.36)	47 (5.32)	101
15.1-39	Male	27 (4.55)	11 (1.85)	49 (8.25)	10 (1.68)	97 (16.33)	294 (33.26)	13 (1.47)	9 (1.02)	69 (7.81)	385 (43.55)	482
	Female	18 (3.03)	8 (1.35)	21 (3.54)	4 (0.67)	51 (8.59)	100 (11.31)	6 (0.68)	5 (0.57)	7 (0.79)	118 (13.35)	169
39.1 & >	Male	45 (7.58)	10 (1.68)	44 (7.41)	2 (0.34)	101 (17)	95 (10.75)	4 (0.45)	6 (0.68)	7 (0.79)	112 (12.67)	213
	Female	28 (4.71)	8 (1.35)	14 (2.36)	1 (0.17)	51 (8.59)	56 (6.33)	6 (0.68)	3 (0.34)	3 (0.34)	68 (7.69)	119
Total		185 (31.14)	58 (9.76)	323 (54.4)	28 (4.71)	594 (100)	645 (72.96)	62 (7.01)	29 (3.28)	148 (16.7)	884 (100)	1478
Lymphocytes Status												
<1	Male	0 (0)	2 (0.34)	3 (0.51)	0 (0)	5 (0.84)	1 (0.11)	1 (0.11)	0 (0)	1 (0.11)	3 (0.34)	8
	Female	0 (0)	0 (0)	0 (0)	2 (0.34)	2 (0.34)	0 (0)	5 (0.57)	1 (0.11)	1 (0.11)	7 (0.79)	9
1.1-5	Male	6 (1.01)	33 (5.56)	28 (4.71)	13 (2.19)	80 (13.47)	7 (0.79)	16 (1.81)	6 (0.68)	24 (2.71)	53 (6)	133
	Female	2 (0.34)	18 (3.03)	10 (1.68)	6 (1.01)	36 (6.06)	4 (0.45)	10 (1.13)	2 (0.23)	5 (0.57)	21 (2.38)	57
5.1-15	Male	18 (3.03)	39 (6.57)	26 (4.38)	34 (5.72)	117 (19.7)	24 (2.71)	20 (2.26)	7 (0.79)	19 (2.15)	70 (7.92)	187

	Female	5 (0.84)	14 (2.36)	20 (3.37)	15 (2.53)	54 (9.09)	19 (2.15)	9 (1.02)	7 (0.79)	12 (1.36)	47 (5.32)	101
15.1-39	Male	34 (5.72)	9 (1.52)	17 (2.86)	37 (6.23)	97 (16.33)	306 (34.62)	7 (0.79)	0 (0)	72 (8.14)	385 (43.55)	482
	Female	10 (1.68)	7 (1.18)	21 (3.54)	13 (2.19)	51 (8.59)	49 (5.54)	10 (1.13)	52 (5.88)	7 (0.79)	118 (13.35)	169
39.1 & >	Male	50 (8.42)	18 (3.03)	14 (2.36)	19 (3.2)	101 (17)	95 (10.75)	6 (0.68)	4 (0.45)	7 (0.79)	112 (12.67)	213
	Female	10 (1.68)	12 (2.02)	21 (3.54)	8 (1.35)	51 (8.59)	39 (4.41)	12 (1.36)	14 (1.58)	3 (0.34)	68 (7.69)	119
Total		135 (22.73)	152 (25.59)	160 (26.9)	147 (24.7)	594 (100)	544 (61.54)	96 (10.86)	93 (10.52)	151 (17.0)	884 (100)	1478
Neutrophils Status												
<1	Male	0 (0)	0 (0)	5 (0.84)	0 (0)	5 (0.84)	0 (0)	0 (0)	1 (0.11)	2 (0.23)	3 (0.34)	8
	Female	0 (0)	0 (0)	0 (0)	2 (0.34)	2 (0.34)	1 (0.11)	0 (0)	5 (0.57)	1 (0.11)	7 (0.79)	9
1.1-5	Male	4 (0.67)	1 (0.17)	57 (9.6)	18 (3.03)	80 (13.47)	1 (0.11)	3 (0.34)	25 (2.83)	24 (2.71)	53 (6)	133
	Female	0 (0)	4 (0.67)	22 (3.7)	10 (1.68)	36 (6.06)	1 (0.11)	1 (0.11)	14 (1.58)	5 (0.57)	21 (2.38)	57
5.1-15	Male	9 (1.52)	8 (1.35)	61 (10.27)	39 (6.57)	117 (19.7)	12 (1.36)	1 (0.11)	38 (4.3)	19 (2.15)	70 (7.92)	187
	Female	1 (0.17)	3 (0.51)	32 (5.39)	18 (3.03)	54 (9.09)	6 (0.68)	3 (0.34)	26 (2.94)	12 (1.36)	47 (5.32)	101
15.1-39	Male	24 (4.04)	3 (0.51)	29 (4.88)	41 (6.9)	97 (16.33)	294 (33.26)	4 (0.45)	17 (1.92)	70 (7.92)	385 (43.55)	482
	Female	17 (2.86)	5 (0.84)	15 (2.53)	14 (2.36)	51 (8.59)	75 (8.48)	32 (3.62)	4 (0.45)	7 (0.79)	118 (13.35)	169
39.1 & >	Male	45 (7.58)	2 (0.34)	33 (5.56)	21 (3.54)	101 (17)	90 (10.18)	7 (0.79)	8 (0.9)	7 (0.79)	112 (12.67)	213
	Female	19 (3.2)	7 (1.18)	16 (2.69)	9 (1.52)	51 (8.59)	53 (6)	6 (0.68)	5 (0.57)	4 (0.45)	68 (7.69)	119
Total		119 (20.03)	33 (5.56)	270 (45.4)	172 (28.9)	594 (100)	533 (60.29)	57 (6.45)	143 (16.18)	151 (17.0)	884 (100)	1478

Table 3.4B: Overall origin and nature-wise distribution of leukemia patients against age, gender, and hematological variables (n=594)

Age Group	Status	Gender	Disease Origin		Disease Nature		Total
			Lymphoid	Myeloid	Acute	Chronic	
			No. (%)	No. (%)	No. (%)	No. (%)	No. (%)
RBC Status							
<1	Normal	Male	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	Increment	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Decrement	Male	3 (0.51)	1 (0.17)	3 (0.51)	1 (0.17)	4 (0.67)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Unavailable	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
1.1-5	Normal	Male	18 (3.03)	0 (0)	18 (3.03)	0 (0)	18 (3.03)
		Female	3 (0.51)	0 (0)	3 (0.51)	0 (0)	3 (0.51)
	Increment	Male	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
		Female	1 (0.17)	1 (0.17)	2 (0.34)	0 (0)	2 (0.34)
	Decrement	Male	54 (9.09)	5 (0.84)	58 (9.76)	1 (0.17)	59 (9.93)
		Female	28 (4.71)	2 (0.34)	29 (4.88)	1 (0.17)	30 (5.05)
	Unavailable	Male	2 (0.34)	0 (0)	2 (0.34)	0 (0)	2 (0.34)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
5.1-15	Normal	Male	32 (5.39)	2 (0.34)	34 (5.72)	0 (0)	34 (5.72)
		Female	9 (1.52)	1 (0.17)	10 (1.68)	0 (0)	10 (1.68)
	Increment	Male	3 (0.51)	0 (0)	3 (0.51)	0 (0)	3 (0.51)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	Decrement	Male	60 (10.1)	10 (1.68)	69 (11.62)	1 (0.17)	70
		Female	30 (5.05)	6 (1.01)	35 (5.89)	1 (0.17)	36 (6.06)
	Unavailable	Male	8 (1.35)	2 (0.34)	9 (1.52)	1 (0.17)	10 (1.68)
		Female	7 (1.18)	0 (0)	7 (1.18)	0 (0)	7 (1.18)
15.1-39	Normal	Male	21 (3.54)	20 (3.37)	27 (4.55)	14 (2.36)	41 (6.9)
		Female	2 (0.34)	6 (1.01)	2 (0.34)	6 (1.01)	8 (1.35)
	Increment	Male	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Decrement	Male	26 (4.38)	16 (2.69)	34 (5.72)	8 (1.35)	42 (7.07)
		Female	11 (1.85)	28 (4.71)	25 (4.21)	14 (2.36)	39 (6.57)
	Unavailable	Male	11 (1.85)	2 (0.34)	11 (1.85)	2 (0.34)	13 (2.19)
		Female	1 (0.17)	3 (0.51)	3 (0.51)	1 (0.17)	4 (0.67)
39.1 & >	Normal	Male	18 (3.03)	29 (4.88)	4 (0.67)	43 (7.24)	47 (7.91)
		Female	0 (0)	12 (2.02)	1 (0.17)	11 (1.85)	12 (2.02)
	Increment	Male	1 (0.17)	3 (0.51)	0 (0)	4 (0.67)	4 (0.67)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Decrement	Male	17 (2.86)	27 (4.55)	21 (3.54)	23 (3.87)	44 (7.41)
		Female	8 (1.35)	30 (5.05)	11 (1.85)	27 (4.55)	38 (6.4)

	Unavailable	Male	1 (0.17)	5 (0.84)	2 (0.34)	4 (0.67)	6 (1.01)
		Female	0 (0)	1 (0.17)	1 (0.17)	0 (0)	1 (0.17)
Total			382 (64.31)	212 (35.69)	431 (72.56)	163	594 (100)
Platelets Status							
<1	Normal	Male	2 (0.34)	0 (0)	2 (0.34)	0 (0)	2 (0.34)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Increment	Male	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Decrement	Male	1 (0.17)	1 (0.17)	1 (0.17)	1 (0.17)	2 (0.34)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	Unavailable	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
1.1-5	Normal	Male	16 (2.69)	2 (0.34)	17 (2.86)	1 (0.17)	18 (3.03)
		Female	7 (1.18)	1 (0.17)	8 (1.35)	0 (0)	8 (1.35)
	Increment	Male	5 (0.84)	0 (0)	5 (0.84)	0 (0)	5 (0.84)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	Decrement	Male	52 (8.75)	3 (0.51)	55 (9.26)	0 (0)	55 (9.26)
		Female	24 (4.04)	2 (0.34)	25 (4.21)	1 (0.17)	26 (4.38)
	Unavailable	Male	2 (0.34)	0 (0)	2 (0.34)	0 (0)	2 (0.34)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
5.1-15	Normal	Male	29 (4.88)	1 (0.17)	29 (4.88)	1 (0.17)	30 (5.05)
		Female	8 (1.35)	1 (0.17)	9 (1.52)	0 (0)	9 (1.52)
	Increment	Male	9 (1.52)	0 (0)	9 (1.52)	0 (0)	9 (1.52)
		Female	3 (0.51)	2 (0.34)	4 (0.67)	1 (0.17)	5 (0.84)
	Decrement	Male	64 (10.77)	11 (1.85)	75 (12.63)	0 (0)	75
		Female	32 (5.39)	4 (0.67)	36 (6.06)	0 (0)	36 (6.06)
	Unavailable	Male	1 (0.17)	2 (0.34)	2 (0.34)	1 (0.17)	3 (0.51)
		Female	4 (0.67)	0 (0)	4 (0.67)	0 (0)	4 (0.67)
15.1-39	Normal	Male	14 (2.36)	13 (2.19)	14 (2.36)	13 (2.19)	27 (4.55)
		Female	2 (0.34)	16 (2.69)	4 (0.67)	14 (2.36)	18 (3.03)
	Increment	Male	1 (0.17)	10 (1.68)	3 (0.51)	8 (1.35)	11 (1.85)
		Female	2 (0.34)	6 (1.01)	4 (0.67)	4 (0.67)	8 (1.35)
	Decrement	Male	36 (6.06)	13 (2.19)	48 (8.08)	1 (0.17)	49 (8.25)
		Female	9 (1.52)	12 (2.02)	19 (3.2)	2 (0.34)	21 (3.54)
	Unavailable	Male	8 (1.35)	2 (0.34)	8 (1.35)	2 (0.34)	10 (1.68)
		Female	1 (0.17)	3 (0.51)	3 (0.51)	1 (0.17)	4 (0.67)
39.1 & >	Normal	Male	16 (2.69)	29 (4.88)	4 (0.67)	41 (6.9)	45 (7.58)
		Female	3 (0.51)	25 (4.21)	5 (0.84)	23 (3.87)	28 (4.71)
	Increment	Male	2 (0.34)	8 (1.35)	1 (0.17)	9 (1.52)	10 (1.68)
		Female	1 (0.17)	7 (1.18)	0 (0)	8 (1.35)	8 (1.35)
	Decrement	Male	19 (3.2)	25 (4.21)	21 (3.54)	23 (3.87)	44 (7.41)
		Female	4 (0.67)	10 (1.68)	7 (1.18)	7 (1.18)	14 (2.36)
	Unavailable	Male	0 (0)	2 (0.34)	1 (0.17)	1 (0.17)	2 (0.34)
		Female	0 (0)	1 (0.17)	1 (0.17)	0 (0)	1 (0.17)
Total			382 (64.31)	212 (35.69)	431 (72.56)	163	594 (100)

3.1.1.4.1.1 White blood cells

Overall, 32% of the patients had normal white blood cell (WBCs) levels, 40% had increment to normal and 28% had a decrement (Table 3.2, Figure 3.7). There is an obvious decrement in WBCs level to normal in preschoolers (39%) and children (40%), while in AYAs (39%) and aged (52%) there is an overall increment in each of the groups, however, this increment is more evident in aged group. Although both genders contributed to the decrement observed in preschoolers and children, there is slightly more decrement employed by the male participants. Females in aged groups exhibit more increment in WBCs levels in comparison to males in the same group (Table 3.4A, Figure 3.8A).

Overall, 69.95% of the controls had normal WBCs count, 9.50% had increments, and only 3.74 % had decrements in the count. The WBCs level was normal in a higher number of controls in all age groups and both genders, however, AYA females have shown an increased number of individuals with increment in WBCs count as compared to females in other age groups (Table 3.4A, Figure 3.8B).

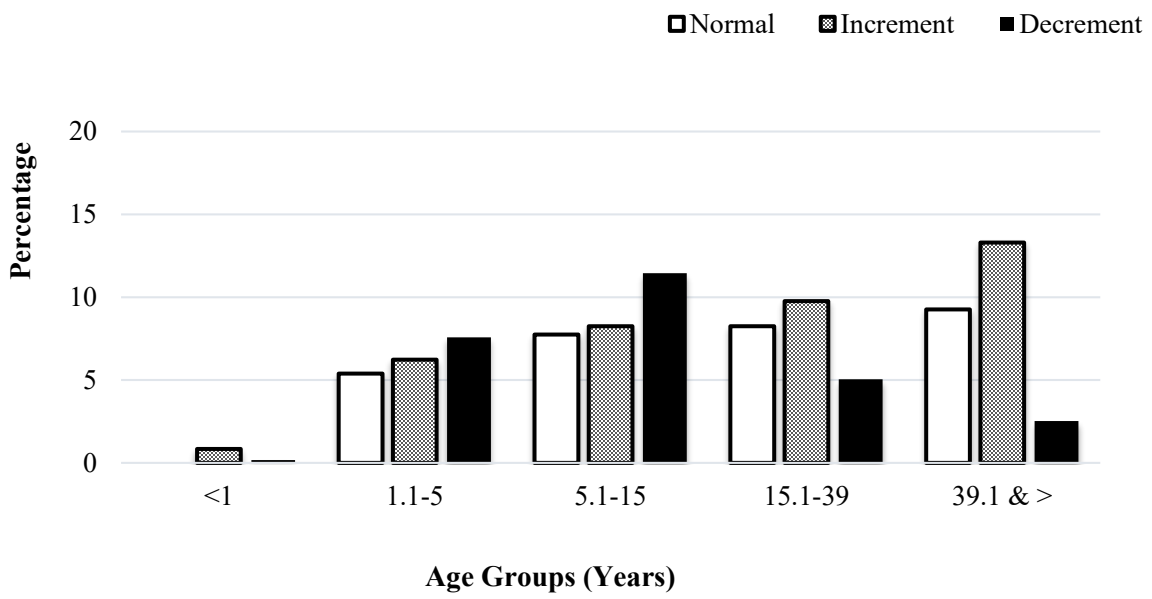


Figure 3.8A: Overall age groupwise percentage distribution of leukemia patients against blood WBCs (n=594)

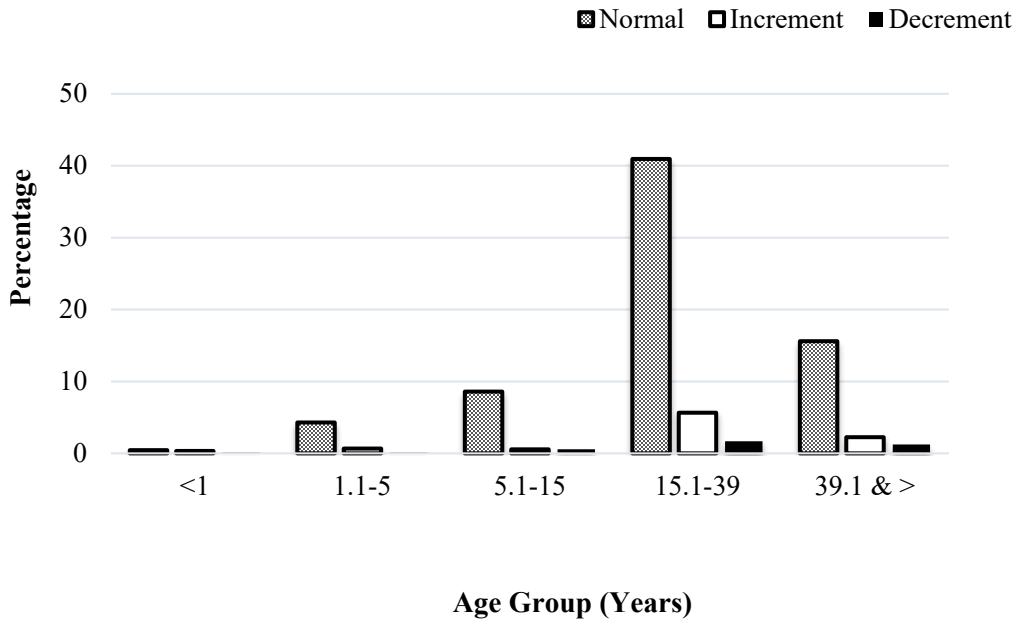


Figure 3.8B: Overall age groupwise percentage distribution of controls against blood WBCs (n=884)

3.1.1.4.1.2 Red blood cells

Overall, 29% of the patients had normal red blood cells (RBC) levels, 2% had increment to normal and 61% had decrement (Table 3.2). In preschoolers, 77%, in children 62%, in AYAs 55% and aged 54% of individuals were dealing with RBCs decrement. In general, overall decrement is reported in all age groups, the highest decrement to normal ratio is observed in preschooler (4.2:1) and children (2.4:1), which decreases as the age increases. Females are here observed to have a much higher number of RBCs decrement as compared to male, though male also suffers from RBCs decrement across all groups (Table 3.4A, Figure 3.9A).

Females in myeloid AYAs (55%) and aged (59%) have the highest tendency towards RBCs decrement. In myeloid AYAs and aged more males had normal than decreased levels. However, lymphoid male participants in AYAs have the same percentages of normal and decrement. In the group aged, a higher number of male participants (42.5%) with chronic leukemia were normal in RBC level. This observation is against the general overall trend for RBCs, in which decrements are decisive findings in all groups. Almost all the acute members, both males and females of the aged group had an RBCs decrement. In general, in acute leukemia, both males and females had higher percentages for individuals with RBC decrements. A similar pattern has been seen in female participants in AYAs and Aged, but the male participants of these two groups (42%, 46.5%) show a higher number of normal RBC levels (Table 3.4B).

Among all 884 controls, 57.69% were normal, 16.4 % were having increments, and 9.05% were having decrements in RBCs count. The RBC count of the remaining 16.8% of controls was not available (Table 3.2). The RBC status was normal across each age group and gender. About 11.43% of the controls, all situated in the AYAs group had having increment in RBCs level, and a majority of them were male controls (Table 3.4A, Figure 3.9B).

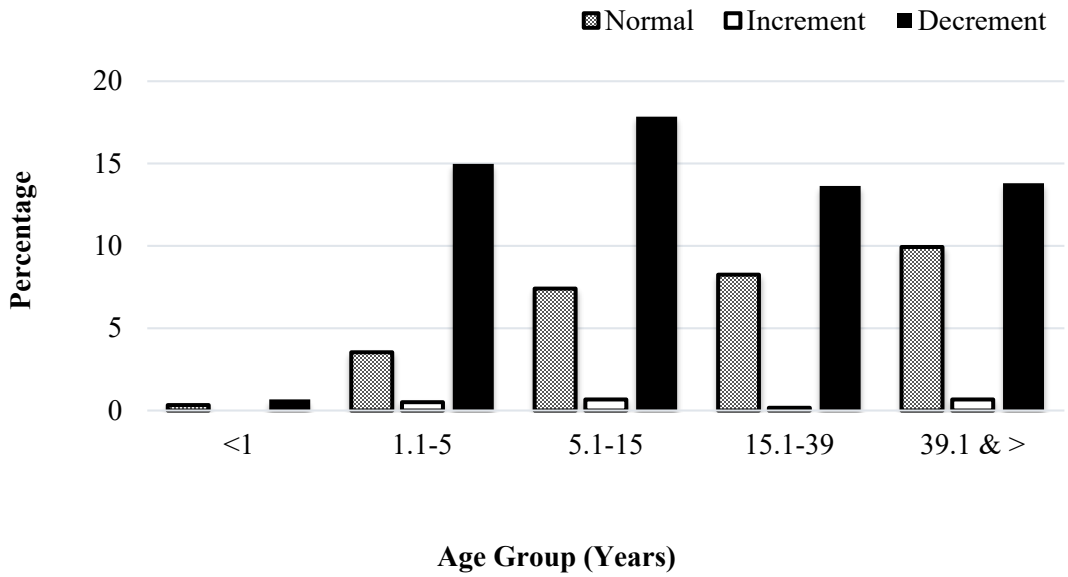


Figure 3.9A: Overall age groupwise percentage distribution of leukemia patients against blood RBCs (n=594)

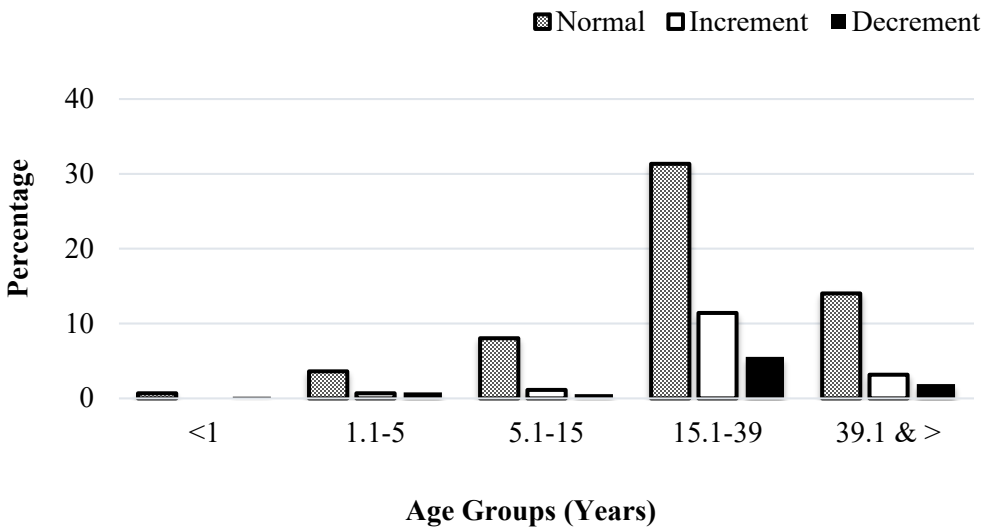


Figure 3.9B: Overall age group-wise percentage distribution of controls against blood RBCs (n=884)

3.1.1.4.1.3 Hemoglobin status

Overall, 23% of the patients had a normal Hemoglobin level, 0.32% had an increment to normal and 68% had a decrement (Table 3.2, Figure 3.7). Hemoglobin is highly decreased across all age groups. Among the groups, among all 84% of patients in preschoolers, 74% in children, 69% in AYAs, and 61% in aged were having decrement hemoglobin levels than normal. In preschoolers and children, the number of males (85% and 74% respectively) with hemoglobin decrement was higher than the males with normal levels and their females' counterparts (80% and 72% respectively). In AYAs, decrement is twice as common as normal in males, while females (82%) in the group numbered highest in hemoglobin decrement, sharing similarity with females in age (80%) (Table 3.4A, Figure 3.10A).

Among all 884 controls, 64.59% were normal, 0.45 % were elevated, and 18.21 % were having lowered hemoglobin levels. The hemoglobin levels of the remaining 16.74% of controls were not available (Table 3.2). The majority of the individual, both males and females in each age group were having normal Hb levels, except for AYAs females, in which more females, 7.25% had lowered Hb levels than those, 5.3% with normal (Table 3.4A, Figure 3.10B)

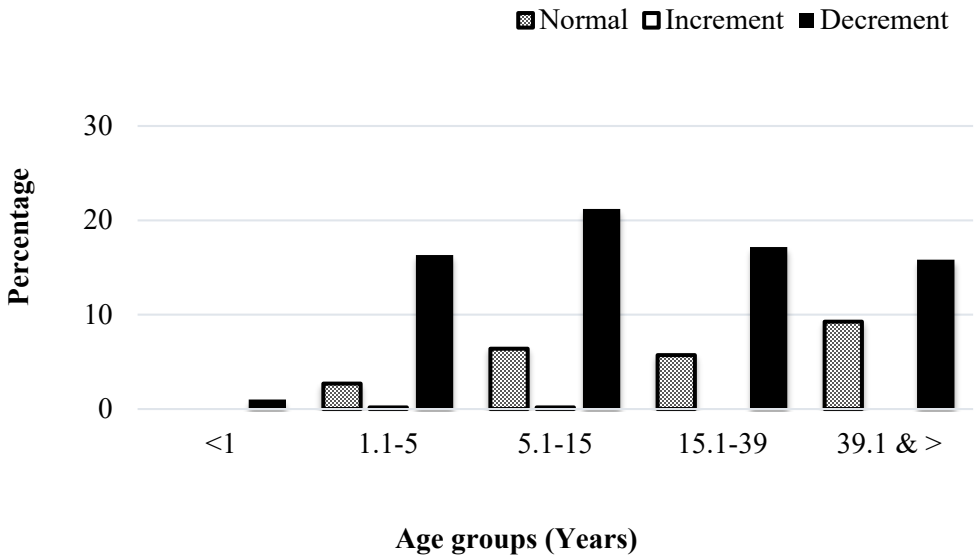


Figure 3.10A: Overall age groupwise percentage distribution of leukemia patients against hemoglobin (n=594)

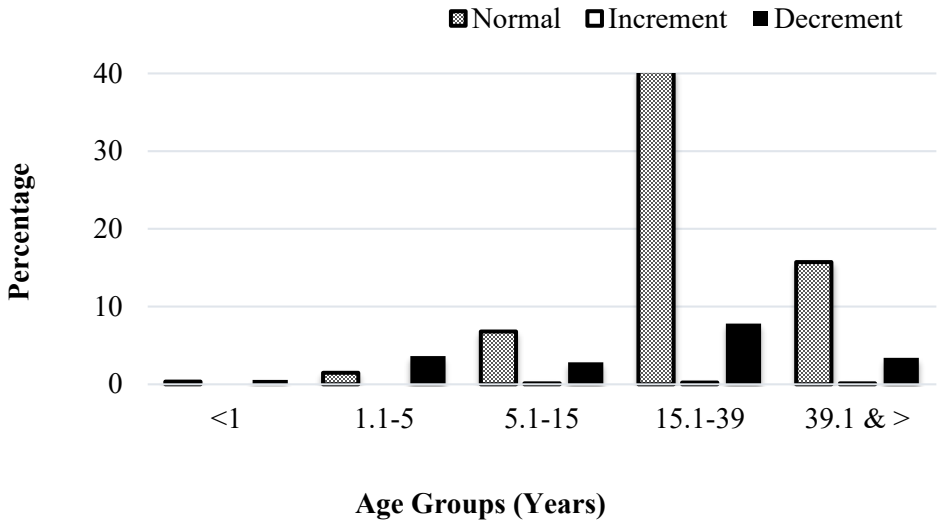


Figure 3.10B: Overall age groupwise percentage distribution of controls against hemoglobin (n=884)

3.1.1.4.1.4 Platelets status

Overall, 30% of the patients had normal platelet levels, 9.4% had increments to normal and 52.4% had decrement (Table 3.2, Figure 3.7). Platelet levels were decreased among all age groups including preschoolers (70%), children (65%), and AYAs (47%) except aged, where it is only 38 %, lesser than normal (48%). In preschoolers, 68% of males and 72% percent females had a decrement in platelet status normal. In children, 43% of the total males and 66% of the females had decrement than normal. However, in AYAs, 50.5% of the males and 41% of the females were in decrement. In the age, 43% of males and 27% of females had decrement in platelets status. In preschoolers and children, there is more decrement in females than males, however, in AYAs and the aged male has higher percentages of decrement than females in the same groups (Table 3.4A, Figure 3.11A).

In lymphoid leukemia, there is more decrement in all age groups in both genders, while in myeloid, there is an increasing shift from decrement to normal, particularly females in AYAs and age groups had more normal platelet levels than males in the same groups. In AYAs, the percentage of male patients with normal platelets level was the same in both lymphoid (52%) and myeloid (48%). In acute leukemia, like lymphoid, there is a decrement in platelets level in both genders across each group. In the chronic aged group, males with normal platelet levels are higher than decrement. Similar reading is observed in the aged female. In aged, the percentage of patients with the decrement of platelets for both acute and chronic males and females are the same (Table 3.4B).

Among all 884 controls, 72.96% were normal, 7.01 % were elevated, and 3.28 % were having lowered platelet levels. The platelet levels of the remaining 16.74% of controls were not available (Table 3.2). Platelet levels of controls were normal across each group and both genders in control participants (Table 3.4A, Figure 3.11B).

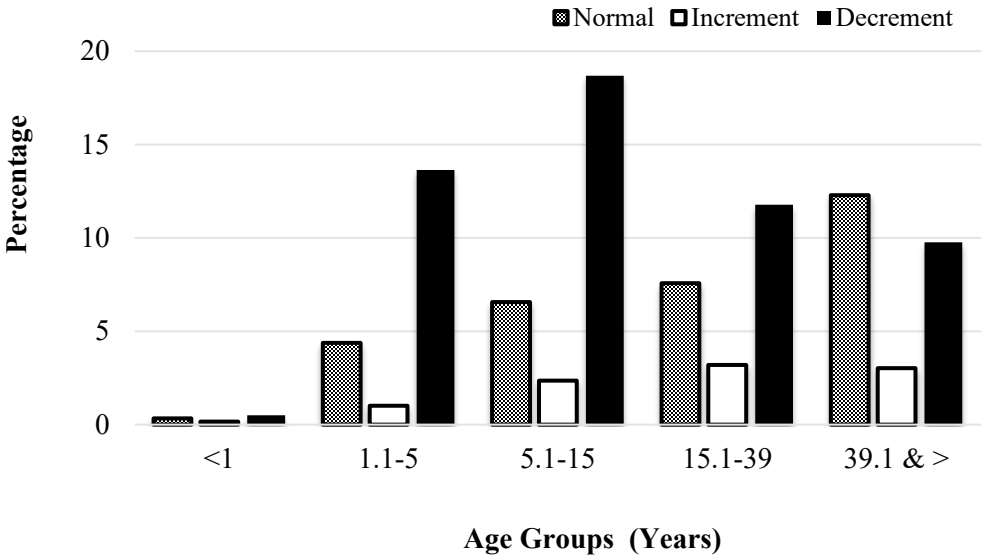


Figure 3.11A: Overall age groupwise percentage distribution of leukemia patients against platelets (n=594)

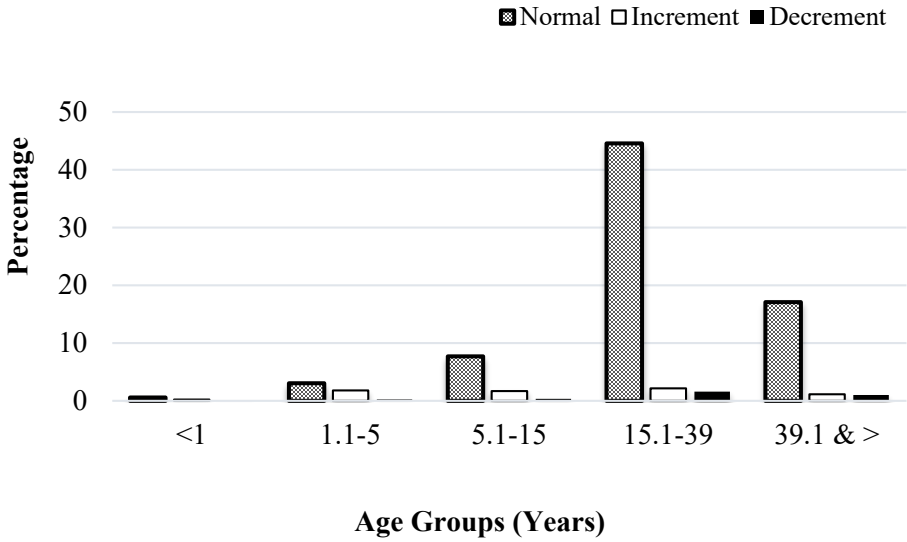


Figure 3.11B: Overall age groupwise percentage distribution of controls against platelets (n=884)

3.1.1.4.1.5 Lymphocytes (%) status

Overall, 22% of the patients had normal lymphocytes %, 25% had increment to normal and 26% had decrement (Table 3.2, Figure 3.7). In general, there is an increment in lymphocyte %, in preschoolers (43%) and children (31%), however in AYAs (30%) and aged (39%) the number of patients with normal lymphocytes predominates. In preschoolers, lymphocyte % is increased in both males (41%) and females (50%). In children, males (33%) had increased lymphocytes while in females (37%) there was a decrement in lymphocytes %. In AYAs (35%) and aged males (49.5%), the number of normal lymphocytes is more prominent, while in females (41%), the higher number of participants had a decrement in lymphocytes in both groups. In general males in preschoolers and children had an increment, in AYAs and aged had normal, while females had overall decrements in lymphocytes %, throughout except in preschoolers, where there is an increment (Table 3.4A, Figure 3.12A).

Among all 884 controls, 61.54% were having normal, 10.86 % were having increment, and 10.52 % were having decrement in lymphocytes %. The lymphocyte % of the remaining controls was not available (Table 3.2). The majority of the control participants, including both males and females in each group, were having normal levels of lymphocytes, with slight and negligible numeral variations, except for AYAs female, who has a higher number, 5.8% for lowered lymphocytes % than the normal 5.2% (Table 3.4A, Figure 3.12B).

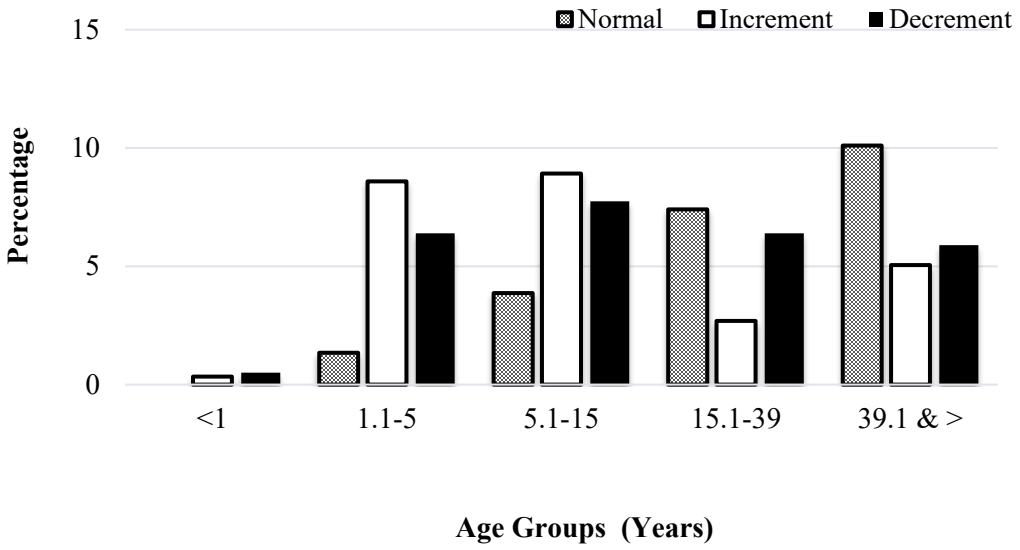


Figure 3.12A: Overall age groupwise percentage distribution of leukemia patients against lymphocytes (%) (n=594)

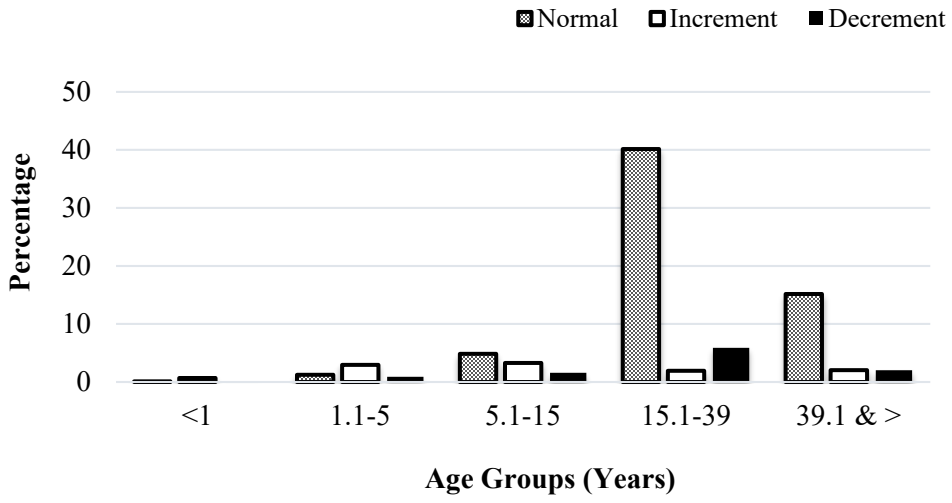


Figure 3.12B: Overall age groupwise percentage distribution of controls against lymphocytes (%) (n=884).

3.1.1.4.1.6 Neutrophils (%) status

Overall, 19% of the patients had normal neutrophils %, 5.3% had increment to normal and 44% had decrement (Table 3.2, Figure 3.7). There is an overall decrement in neutrophil status across each group except the aged. In preschoolers and children, the highest number of individuals with decreased neutrophils levels is 68% and 54% respectively. In AYAs, the normal and decrement percentages were the same (28%), while in aged the normal neutrophils percentages were more frequent (42%). In preschoolers, 71% of males and 61% of females had a decrement in neutrophils. In children, 52% of males and 59% of females had a decrement in neutrophils. In AYAs males (30%) have decrement but females (33%) were normal in this regard and 61% of the females had decrement in neutrophils. In aged both males (44.5%) and females (37%) have the highest percentages of individuals with normal neutrophils (Table 3.4A, Figure 3.13A).

Among all 884 controls, 60.29 % had normal, 6.45 % had having increment, and 16.18 % had having decrement in neutrophils %. The neutrophil % of the remaining controls was not available (Table 3.2). In age groupwise contribution, children, both male and female, up to the age of 15 years were having slight decrement in neutrophil%, although a much smaller number than patients, while AYAs and adults the neutrophils % status is normal across both genders (Table 3.4A, Figure 3.13B).

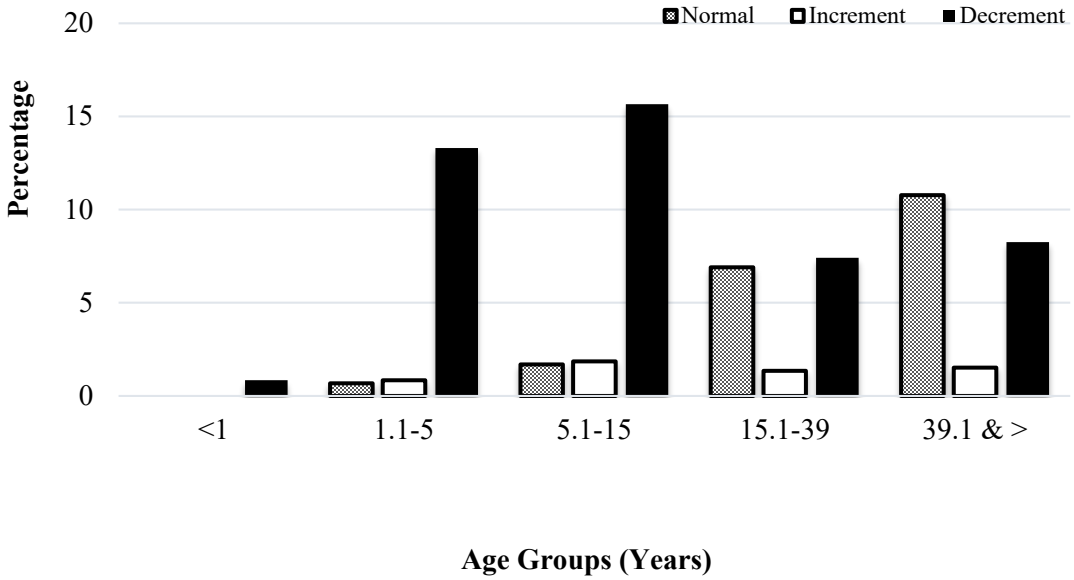


Figure 3.13A: Overall age group-wise distribution of leukemia patients against neutrophils % (n=594)

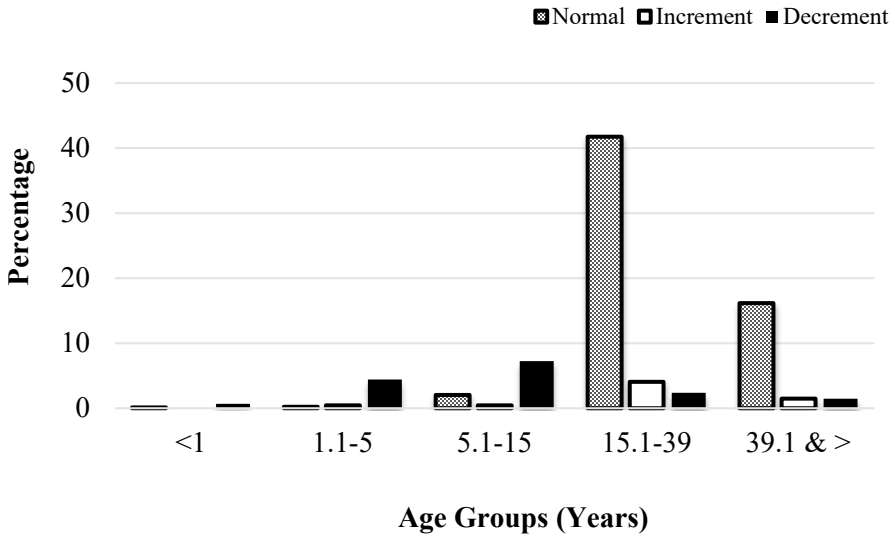


Figure 3.13B: Overall age group-wise distribution of controls against neutrophils (%) (n=884)

3.1.1.4.2 Association of hematological variables to leukemia

In the current study, blast cell percentage data were obtained for patients. About 71% of the cases had >50% blast cells while 29 % had ≤50% blast cells in which most of the cases have more than 20% of blast cells (Table 3.2). In myeloid leukemia, age group 39.1 & > has a significant odd ratio of 5.71 for blast cell % of ≤50%, in comparison to lymphoid leukemia when compared to all other age groups (OR= 5.71; 95% CI=2.01-16.23).

In chronic leukemia, blast cell ≤50% for age group >39.1 is significantly higher than acute leukemia when compared to all other groups (OR= 27.6857; 95% CI= 7.7824-98.4913; $p < 0.0001$; Table 3.10, Figure 3.36). Hyperleukocytosis (OR= 9.06; 95% CI=6.46-12.71), anemia (OR=15.84; 95% CI=11.84-21.21), lower hemoglobin level (OR= 8.11; 95% CI=6.35-10.37), thrombocytopenia (OR=32.40; 95% CI=21.57-48.68), lymphocytosis (OR=3.41; 95% CI=2.55-4.57), and neutropenia (OR=7.32; 95% CI=5.59-9.60) were the significant hematological and paraclinical risk factors associated to leukemia in the studied population (Table 3.5, Figure 3.14)

Table 3.5: Odd ratio of hematological variables for patients vs controls

Categories	Variables	Patients	Controls	Odds Ratios	95 % CI	P-Values
		N (%)	N (%)			
WBCs Count*	Leukocytosis	228 (40.07)	84 (9.5)	9.07	6.47-12.71	P < 0.0001
	Normal, Leukocytopenia	341 (59.93)	651 (73.64)	1.00		
RBCs Count*	Anemia	362 (58.76)	80 (9.04)	15.85	11.84-21.22	P < 0.0001
	Normal, Erythrocytosis	187 (30.36)	655 (74.1)	1.00		
Hemoglobin Status	Lower Hb	425 (68.99)	161 (18.21)	8.12	6.35-10.37	P < 0.0001
	Normal	145 (23.54)	575 (65.05)	1.00		
Platelets Count	Thrombocytopenia	323 (52.43)	29 (3.28)	32.41	21.57-48.68	P < 0.0001
	Normal, Thrombocytosis	243 (39.45)	707 (79.98)	1.00		
Lymphocytes %	Lymphocytosis	152 (24.67)	96 (10.85)	3.42	2.56-4.57	P < 0.0001
	Normal, Lymphocytopenia	295 (47.89)	637 (72.06)	1.00		
Neutrophils %	Neutropenia	270 (43.83)	143 (16.17)	7.33	5.59-9.6	P < 0.0001
	Normal, Neutrocytosis	152 (24.6)	590 (66.7)	1.00		

WBCs: White blood cells, RBCs: Red blood cells, Hb: Hemoglobin

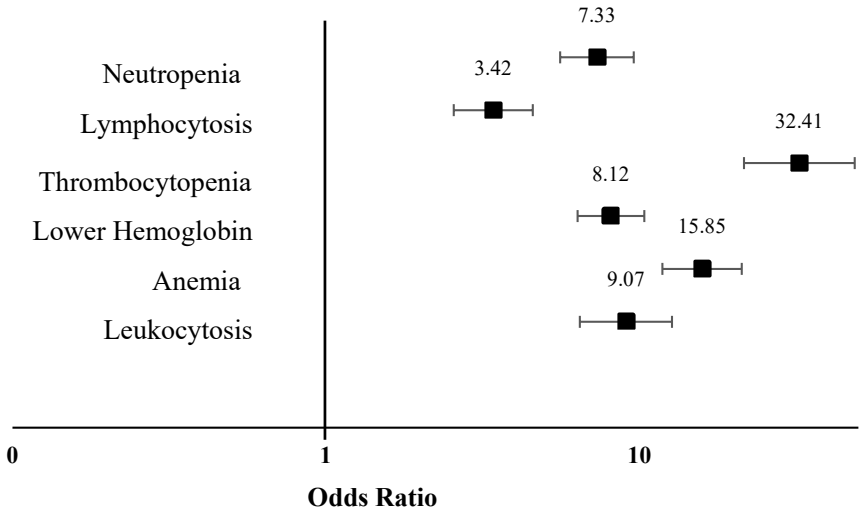


Figure 3.14: Forest plot depicting significant hematological variables associated with leukemia.

Leukocytosis (ref. additive model; normal count, leukocytopenia), and lymphocytosis (ref. additive model; normal count, lymphocytopenia), are significantly associated with leukemia patients in comparison to age and gender-matched controls (CI 95%, p* = 0.05). Also, on the contrary, anemia (ref. additive model; normal count, erythrocytosis), hemoglobin (ref. additive model; normal and higher level), and thrombocytopenia (ref. additive model; normal count, thrombocytosis), were highly significant in leukemia patients.

3.1.2 Demographic and lifestyle-related variables

Demographic data, including gender, age, blood group, weight, ethnicity, marital status, family type, parity, parental consanguinity, family history, patient's other diseases, and education details were recorded for the patients. A risk factor or determinant is a variable associated with an increased risk of disease. In this study, several lifestyle-related risk factors were patterned against leukemia risk, including nutrition, use of microwave, fuel, water, drinks, perfumes, and tobacco in major (Table 3.6).

Table 3.6: General characteristics, demographic and lifestyle-related risk factors details of the study population (leukemia patients, n=616; control samples, n=884)

Variable	Patients	Controls	Total	Variable	Patients	Controls	Total
	N (%)	N (%)			N (%)	N (%)	
Data present				**Family History**			
Yes	566 (91.88)	830 (93.89)	1396	Yes	34 (5.52)	12 (1.36)	46
No	50 (8.12)	54 (6.11)	104	No	582 (94.48)	872 (98.64)	1454
Total	616 (100)	884 (100)	1500	Total	616 (100)	884 (100)	1500
Gender				**Family Type**			
Male	411 (66.72)	623 (70.48)	1034	Nuclear	109 (17.69)	162 (18.33)	271
Female	205 (33.28)	261 (29.52)	466	Extended	214 (34.74)	271 (30.66)	485
Total	616 (100)	884 (100)	1500	Unavailable	293 (47.56)	451 (51.02)	744
Blood Group				Total	616 (100)	884 (100)	1500
A	59 (9.58)	49 (5.54)	108	**Ethnicity**			
B	83 (13.47)	63 (7.13)	146	Pathan	354 (57.47)	289 (32.69)	643
AB	30 (4.87)	18 (2.04)	48	Punjabi	191 (31.01)	494 (55.88)	685
O	54 (8.77)	54 (6.11)	108	Others	45 (7.31)	56 (6.33)	101
Unavailable	390 (63.31)	700 (79.19)	1090	Unavailable	26 (4.22)	45 (5.09)	71
Total	616 (100)	884 (100)	1500	Total	616 (100)	884 (100)	1500
Rh- Factor				**Marital Status**			
+ve	206 (33.44)	166 (18.78)	372	Married	151 (24.51)	462 (52.26)	613
-ve	20 (3.25)	18 (2.04)	38	Unmarried	27 (4.38)	162 (18.33)	189
Unavailable	390 (63.31)	700 (79.19)	1090	Below Marital Age	330 (53.57)	257 (29.07)	587
Total	616 (100)	884 (100)	1500	Unavailable	108 (17.53)	3 (0.34)	111
Age Group				Total	616 (100)	884 (100)	1500
<1	7 (1.14)	10 (1.13)	17	**Education** [b]			
1.1-5	116 (18.83)	74 (8.37)	190	Educated	275 (44.64)	694 (78.51)	969
5.1-15	171 (27.76)	117 (13.24)	288	No Formal Education	105 (17.05)	98 (11.09)	203

15.1-39	148 (24.03)	503 (56.9)	651	Inadmissible	103 (16.72)	58 (6.56)	161
39.1 & >	152 (24.68)	180 (20.36)	332	Unavailable	133 (21.59)	34 (3.85)	167
Unavailable	22 (3.57)	0 (0)	22	Total	616 (100)	884 (100)	1500
Total	616 (100)	884 (100)	1500	**Level of Education**			
BMI				Primary	139 (22.56)	155 (17.53)	294
Normal	128 (20.78)	210 (23.76)	338	Secondary	114 (18.51)	297 (33.60)	411
Overweight	152 (24.68)	467 (52.83)	619	Graduation	22 (3.57)	242 (27.37)	264
Underweight	123 (19.97)	152 (17.19)	275	No Formal Education	105 (17.05)	98 (11.09)	203
Unavailable	213 (34.58)	55 (6.22)	268	Inadmissible	103 (16.72)	58 (6.56)	161
				Unavailable	133 (21.59)	34 (3.85)	167
Total	616 (100)	884 (100)	1500	Total	616 (100)	884 (100)	1500
Diet [c]				**Tobacco use**			
Balanced	367 (59.58)	722 (81.67)	1089	Yes	189 (30.68)	233 (26.36)	422
Poor	150 (24.35)	126 (14.25)	276	No	329 (53.41)	639 (72.29)	968
Unavailable	99 (16.07)	36 (4.07)	135	Unavailable	98 (15.91)	12 (1.36)	110
Total	616 (100)	884 (100)	1500	Total	616 (100)	884 (100)	1500
Microwave Use				**Area**			
Yes	53 (8.6)	294 (33.26)	347	Rural	265 (43.02)	252 (28.51)	517
No	406 (65.91)	573 (64.82)	979	Urban	63 (10.23)	176 (19.91)	239
Unavailable	157 (25.49)	17 (1.92)	174	Unavailable	288 (46.75)	456 (51.58)	744
Total	616 (100)	884 (100)	1500	Total	616 (100)	884 (100)	1500
Water Type				**Environme**nt			
Groundwater	240 (38.96)	241 (27.26)	481	Clean	124 (20.13)	302 (34.16)	426
Surface water	71 (11.53)	139 (15.72)	210	Polluted	45 (7.31)	127 (14.37)	172
Treated Water	11 (1.79)	47 (5.32)	58	Unavailable	447 (72.56)	455 (51.47)	902
Unavailable	294 (47.73)	457 (51.7)	751	Total	616 (100)	884 (100)	1500
Total	616 (100)	884 (100)	1500	**Parental Consanguinity**			
Fuel				CU*	202 (32.79)	193 (21.83)	395
Wood	287 (46.59)	206 (23.3)	493	NCU	299 (48.54)	610 (69)	909

Natural Gas	233 (37.82)	664 (75.11)	897	Unavailable	115 (18.67)	81 (9.16)	196
Unavailable	96 (15.58)	14 (1.58)	110	Total	616 (100)	884 (100)	1500
Total	616 (100)	884 (100)	1500	**Parity**			
Carbonated Drinks				1	87 (14.12)	125 (14.14)	212
Yes	336 (54.55)	519 (58.71)	855	2	53 (8.6)	99 (11.2)	152
No	180 (29.22)	350 (39.59)	530	3	35 (5.68)	68 (7.69)	103
Unavailable	100 (16.23)	15 (1.7)	115	4	34 (5.52)	43 (4.86)	77
Total	616 (100)	884 (100)	1500	5	23 (3.73)	37 (4.19)	60
Perfume Use				6->6	39 (6.33)	40 (4.52)	79
Yes	264 (42.86)	620 (70.14)	884	Unavailable	345 (56.01)	472 (53.39)	817
No	249 (40.42)	248 (28.05)	497	Total	616 (100)	884 (100)	1500
Unavailable	103 (16.72)	16 (1.81)	119				
Total	616 (100)	884 (100)	1500				

Out of the total 616 patients, the data 22 were excluded in the subsequent analysis

BMI = body mass index,

[a] Weight status is derived from the BMI

[b]Assessed by asking “What is the highest grade of school completed”.

Only obviously important attributes are explained in the text.

[c] Assessed by asking “Enlist the dietary item you on consume a weekly basis?”.

*CU: Consanguineous union, NCU: Non-consanguineous Union

Table 3.7A: Overall age group and gender-wise distribution of demographic and lifestyle-related risk factors of leukemia patients (n=594), in comparison to controls (n=884).

Variables		Patients					Controls					Net Total
Ethnicity												
Age Group	Gender	Pathan	Punjabi	Others	No info	Total	Pathan	Punjabi	Others	No info	Total	
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	3 (0.5)	2 (0.3)	0 (0)	0 (0)	5 (0.8)	0 (0)	2 (0.2)	1 (0.1)	0 (0)	3 (0.3)	8
	Female	2 (0.3)	0 (0)	0 (0)	0 (0)	2 (0.3)	1 (0.1)	5 (0.6)	0 (0)	1 (0.1)	7 (0.8)	9
1.1-5	Male	51 (8.6)	24 (4)	4 (0.7)	1 (0.2)	80 (13.5)	8 (0.9)	40 (4.5)	5 (0.6)	0 (0)	53 (6)	133
	Female	17 (2.9)	15 (2.5)	3 (0.5)	1 (0.2)	36 (6.1)	2 (0.2)	16 (1.8)	3 (0.3)	0 (0)	21 (2.4)	57
5.1-15	Male	75 (12.6)	31 (5.2)	10 (1.7)	1 (0.2)	117	13 (1.5)	44 (5)	11 (1.2)	2 (0.2)	70 (7.9)	187
	Female	32 (5.4)	17 (2.9)	4 (0.7)	1 (0.2)	54 (9.1)	9 (1)	34 (3.8)	3 (0.3)	1 (0.1)	47 (5.3)	101
15.1-39	Male	56 (9.4)	30 (5.1)	4 (0.7)	7 (1.2)	97 (16.3)	174	166 (18.8)	17 (1.9)	28 (3.2)	385 (43.6)	482
	Female	21 (3.5)	21 (3.5)	6 (1)	3 (0.5)	51 (8.6)	22 (2.5)	82 (9.3)	7 (0.8)	7 (0.8)	118 (13.3)	169
39.1 &	Male	57 (9.6)	31 (5.2)	7 (1.2)	6 (1)	101 (17)	47 (5.3)	54 (6.1)	6 (0.7)	5 (0.6)	112 (12.7)	213
	Female	23 (3.9)	20 (3.4)	7 (1.2)	1 (0.2)	51 (8.6)	13 (1.5)	51 (5.8)	3 (0.3)	1 (0.1)	68 (7.7)	119
Total		337	191 (32.2)	45 (7.6)	21 (3.5)	594	289	494 (55.9)	56 (6.3)	45 (5.1)	884 (100)	1478
Education												
Age Group	Gender	Educated	No Formal	Inadmissible	No info	Total	Educated	No Formal Education	Inadmissible	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	0 (0)	0 (0)	5 (0.8)	0 (0)	5 (0.8)	0 (0)	0 (0)	3 (0.3)	0 (0)	3 (0.3)	8
	Female	0 (0)	0 (0)	2 (0.3)	0 (0)	2 (0.3)	1 (0.1)	0 (0)	6 (0.7)	0 (0)	7 (0.8)	9
1.1-5	Male	13 (2.2)	0 (0)	64 (10.8)	3 (0.5)	80 (13.5)	17 (1.9)	0 (0)	36 (4.1)	0 (0)	53 (6)	133
	Female	5 (0.8)	0 (0)	29 (4.9)	2 (0.3)	36 (6.1)	8 (0.9)	0 (0)	13 (1.5)	0 (0)	21 (2.4)	57
5.1-15	Male	82 (13.8)	17 (2.9)	2 (0.3)	16 (2.7)	117	65 (7.4)	2 (0.2)	0 (0)	3 (0.3)	70 (7.9)	187
	Female	35 (5.9)	10 (1.7)	1 (0.2)	8 (1.3)	54 (9.1)	41 (4.6)	6 (0.7)	0 (0)	0 (0)	47 (5.3)	101
15.1-39	Male	61 (10.3)	12 (2)	0 (0)	24 (4)	97 (16.3)	342	32 (3.6)	0 (0)	11 (1.2)	385 (43.6)	482
	Female	25 (4.2)	13 (2.2)	0 (0)	13 (2.2)	51 (8.6)	95 (10.7)	16 (1.8)	0 (0)	7 (0.8)	118 (13.3)	169
39.1 &	Male	42 (7.1)	24 (4)	0 (0)	35 (5.9)	101 (17)	78 (8.8)	29 (3.3)	0 (0)	5 (0.6)	112 (12.7)	213

	Female	12 (2)	23 (3.9)	0 (0)	16 (2.7)	51 (8.6)	47 (5.3)	13 (1.5)	0 (0)	8 (0.9)	68 (7.7)	119
Total		275	99 (16.7)	103	117 (19.7)	594	694	98 (11.1)	58 (6.6)	34 (3.8)	884 (100)	1478

Consanguinity

Age Group	Gender	CU	NCU	No info	Total	CU	NCU	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	2 (0.3)	2 (0.3)	1 (0.2)	5 (0.8)	2 (0.2)	0 (0)	1 (0.1)	3 (0.3)	8
	Female	1 (0.2)	1 (0.2)	0 (0)	2 (0.3)	4 (0.5)	1 (0.1)	2 (0.2)	7 (0.8)	9
1.1-5	Male	33 (5.6)	33 (5.6)	14 (2.4)	80 (13.5)	34 (3.8)	5 (0.6)	14 (1.6)	53 (6)	133
	Female	15 (2.5)	17 (2.9)	4 (0.7)	36 (6.1)	11 (1.2)	2 (0.2)	8 (0.9)	21 (2.4)	57
5.1-15	Male	57 (9.6)	46 (7.7)	14 (2.4)	117 (19.7)	36 (4.1)	14 (1.6)	20 (2.3)	70 (7.9)	187
	Female	23 (3.9)	25 (4.2)	6 (1)	54 (9.1)	21 (2.4)	8 (0.9)	18 (2)	47 (5.3)	101
15.1-39	Male	23 (3.9)	47 (7.9)	27 (4.5)	97 (16.3)	46 (5.2)	326 (36.9)	13 (1.5)	385 (43.6)	482
	Female	12 (2)	32 (5.4)	7 (1.2)	51 (8.6)	18 (2)	96 (10.9)	4 (0.5)	118 (13.3)	169
39.1 &	Male	25 (4.2)	55 (9.3)	21 (3.5)	101 (17)	15 (1.7)	97 (11)	0 (0)	112 (12.7)	213
	Female	10 (1.7)	31 (5.2)	10 (1.7)	51 (8.6)	6 (0.7)	61 (6.9)	1 (0.1)	68 (7.7)	119
Total		201	289 (48.7)	104	594 (100)	193	610 (69)	81 (9.2)	884 (100)	1478

Family History

Age Group	Gender	Yes	No	No info	Total	Yes	No	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	0 (0)	5 (0.8)		5 (0.8)	0 (0)	3 (0.3)		3 (0.3)	8
	Female	0 (0)	2 (0.3)		2 (0.3)	0 (0)	7 (0.8)		7 (0.8)	9
1.1-5	Male	4 (0.7)	76 (12.8)		80 (13.5)	5 (0.6)	48 (5.4)		53 (6)	133
	Female	2 (0.3)	34 (5.7)		36 (6.1)	1 (0.1)	20 (2.3)		21 (2.4)	57
5.1-15	Male	12 (2)	105 (17.7)		117 (19.7)	2 (0.2)	68 (7.7)		70 (7.9)	187
	Female	3 (0.5)	51 (8.6)		54 (9.1)	0 (0)	47 (5.3)		47 (5.3)	101
15.1-39	Male	4 (0.7)	93 (15.7)		97 (16.3)	1 (0.1)	384 (43.4)		385 (43.6)	482
	Female	1 (0.2)	50 (8.4)		51 (8.6)	3 (0.3)	115 (13)		118 (13.3)	169
39.1 &	Male	3 (0.5)	98 (16.5)		101 (17)	0 (0)	112 (12.7)		112 (12.7)	213
	Female	3 (0.5)	48 (8.1)		51 (8.6)	0 (0)	68 (7.7)		68 (7.7)	119

Total		32 (5.4)	562 (94.6)		594 (100)		12 (1.4)	872 (98.6)		884 (100)		1478

Weight Status

Age Group	Gender	Normal	Overweight	Underweight	No info	Total	Normal	Overweight	Underweight	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	0 (0)	0 (0)	2 (0.3)	3 (0.5)	5 (0.8)	0 (0)	0 (0)	3 (0.3)	0 (0)	3 (0.3)	8
	Female	0 (0)	0 (0)	0 (0)	2 (0.3)	2 (0.3)	0 (0)	0 (0)	5 (0.6)	2 (0.2)	7 (0.8)	9
1.1-5	Male	6 (1)	2 (0.3)	40 (6.7)	32 (5.4)	80 (13.5)	2 (0.2)	0 (0)	43 (4.9)	8 (0.9)	53 (6)	133
	Female	2 (0.3)	0 (0)	22 (3.7)	12 (2)	36 (6.1)	1 (0.1)	1 (0.1)	19 (2.1)	0 (0)	21 (2.4)	57
5.1-15	Male	39 (6.6)	8 (1.3)	37 (6.2)	33 (5.6)	117	13 (1.5)	6 (0.7)	41 (4.6)	10 (1.1)	70 (7.9)	187
	Female	18 (3)	4 (0.7)	17 (2.9)	15 (2.5)	54 (9.1)	16 (1.8)	4 (0.5)	23 (2.6)	4 (0.5)	47 (5.3)	101
15.1-39	Male	26 (4.4)	29 (4.9)	3 (0.5)	39 (6.6)	97 (16.3)	104	258 (29.2)	8 (0.9)	15 (1.7)	385 (43.6)	482
	Female	5 (0.8)	29 (4.9)	0 (0)	17 (2.9)	51 (8.6)	29 (3.3)	79 (8.9)	4 (0.5)	6 (0.7)	118 (13.3)	169
39.1 & >	Male	25 (4.2)	48 (8.1)	2 (0.3)	26 (4.4)	101 (17)	33 (3.7)	66 (7.5)	6 (0.7)	7 (0.8)	112 (12.7)	213
	Female	7 (1.2)	32 (5.4)	0 (0)	12 (2)	51 (8.6)	12 (1.4)	53 (6)	0 (0)	3 (0.3)	68 (7.7)	119
Total		128 (21.5)	152 (25.6)	123 (20.7)	191 (32.2)	594 (100)	210 (23.8)	467 (52.8)	152 (17.2)	55 (6.2)	884 (100)	1478

Diet

Age Group	Gender	Standard	Poor	No info	Total	Standard	Poor	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	0 (0)	4 (0.7)	1 (0.2)	5 (0.8)	2 (0.2)	1 (0.1)	0 (0)	3 (0.3)	8
	Female	2 (0.3)	0 (0)	0 (0)	2 (0.3)	3 (0.3)	4 (0.5)	0 (0)	7 (0.8)	9
1.1-5	Male	48 (8.1)	19 (3.2)	13 (2.2)	80 (13.5)	29 (3.3)	23 (2.6)	1 (0.1)	53 (6)	133
	Female	25 (4.2)	6 (1)	5 (0.8)	36 (6.1)	13 (1.5)	7 (0.8)	1 (0.1)	21 (2.4)	57
5.1-15	Male	63 (10.6)	46 (7.7)	8 (1.3)	117 (19.7)	49 (5.5)	18 (2)	3 (0.3)	70 (7.9)	187
	Female	38 (6.4)	13 (2.2)	3 (0.5)	54 (9.1)	35 (4)	9 (1)	3 (0.3)	47 (5.3)	101
15.1-39	Male	50 (8.4)	22 (3.7)	25 (4.2)	97 (16.3)	357 (40.4)	19 (2.1)	9 (1)	385 (43.6)	482
	Female	37 (6.2)	5 (0.8)	9 (1.5)	51 (8.6)	86 (9.7)	20 (2.3)	12 (1.4)	118 (13.3)	169
39.1 &	Male	66 (11.1)	16 (2.7)	19 (3.2)	101 (17)	90 (10.2)	20 (2.3)	2 (0.2)	112 (12.7)	213
	Female	35 (5.9)	9 (1.5)	7 (1.2)	51 (8.6)	58 (6.6)	5 (0.6)	5 (0.6)	68 (7.7)	119
Total		364 (61.3)	140 (23.6)	90 (15.2)	594 (100)	722 (81.7)	126 (14.3)	36 (4.1)	884 (100)	1478

Area

Age Group	Gender	Rural	Urban	No info	Total	Rural	Urban	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	0 (0)	0 (0)	5 (0.8)	5 (0.8)	2 (0.2)	1 (0.1)	0 (0)	3 (0.3)	8
	Female	1 (0.2)	0 (0)	1 (0.2)	2 (0.3)	6 (0.7)	0 (0)	1 (0.1)	7 (0.8)	9
1.1-5	Male	47 (7.9)	9 (1.5)	24 (4)	80 (13.5)	26 (2.9)	26 (2.9)	1 (0.1)	53 (6)	133
	Female	19 (3.2)	7 (1.2)	10 (1.7)	36 (6.1)	13 (1.5)	7 (0.8)	1 (0.1)	21 (2.4)	57
5.1-15	Male	74 (12.5)	20 (3.4)	23 (3.9)	117 (19.7)	30 (3.4)	34 (3.8)	6 (0.7)	70 (7.9)	187
	Female	29 (4.9)	12 (2)	13 (2.2)	54 (9.1)	21 (2.4)	21 (2.4)	5 (0.6)	47 (5.3)	101
15.1-39	Male	31 (5.2)	5 (0.8)	61 (10.3)	97 (16.3)	96 (10.9)	52 (5.9)	237	385 (43.6)	482
	Female	11 (1.9)	3 (0.5)	37 (6.2)	51 (8.6)	15 (1.7)	20 (2.3)	83 (9.4)	118 (13.3)	169
39.1 &	Male	28 (4.7)	6 (1)	67 (11.3)	101 (17)	36 (4.1)	9 (1)	67 (7.6)	112 (12.7)	213
	Female	13 (2.2)	1 (0.2)	37 (6.2)	51 (8.6)	7 (0.8)	6 (0.7)	55 (6.2)	68 (7.7)	119
Total		253 (42.6)	63 (10.6)	278 (46.8)	594 (100)	252 (28.5)	176 (19.9)	456 (51.6)	884 (100)	1478

Water Type

Age Group	Gender	Ground water	Surface water	Treated Water	No info	Total	Ground water	Surface water	Treated Water	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	0 (0)	0 (0)	0 (0)	5 (0.8)	5 (0.8)	2 (0.2)	1 (0.1)	0 (0)	0 (0)	3 (0.3)	8
	Female	0 (0)	1 (0.2)	0 (0)	1 (0.2)	2 (0.3)	2 (0.2)	4 (0.5)	0 (0)	1 (0.1)	7 (0.8)	9
1.1-5	Male	44 (7.4)	9 (1.5)	1 (0.2)	26 (4.4)	80 (13.5)	30 (3.4)	17 (1.9)	4 (0.5)	2 (0.2)	53 (6)	133
	Female	18 (3)	5 (0.8)	3 (0.5)	10 (1.7)	36 (6.1)	12 (1.4)	5 (0.6)	3 (0.3)	1 (0.1)	21 (2.4)	57
5.1-15	Male	65 (10.9)	22 (3.7)	5 (0.8)	25 (4.2)	117 (19.7)	35 (4)	22 (2.5)	7 (0.8)	6 (0.7)	70 (7.9)	187
	Female	33 (5.6)	7 (1.2)	1 (0.2)	13 (2.2)	54 (9.1)	27 (3.1)	7 (0.8)	8 (0.9)	5 (0.6)	47 (5.3)	101
15.1-39	Male	25 (4.2)	7 (1.2)	1 (0.2)	64 (10.8)	97 (16.3)	70 (7.9)	62 (7)	17 (1.9)	236 (26.7)	385 (43.6)	482
	Female	10 (1.7)	4 (0.7)	0 (0)	37 (6.2)	51 (8.6)	25 (2.8)	6 (0.7)	5 (0.6)	82 (9.3)	118 (13.3)	169
39.1 &	Male	25 (4.2)	10 (1.7)	0 (0)	66 (11.1)	101 (17)	31 (3.5)	11 (1.2)	2 (0.2)	68 (7.7)	112 (12.7)	213
	Female	8 (1.3)	5 (0.8)	0 (0)	38 (6.4)	51 (8.6)	7 (0.8)	4 (0.5)	1 (0.1)	56 (6.3)	68 (7.7)	119
Total		228 (38.4)	70 (11.8)	11 (1.9)	285 (48)	594 (100)	241 (27.3)	139 (15.7)	47 (5.3)	457 (51.7)	884 (100)	1478

Fuel

Age Group	Gender	Wood	Gas	No info	Total	Wood	Gas	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	0 (0)	4 (0.7)	1 (0.2)	5 (0.8)	1 (0.1)	2 (0.2)	0 (0)	3 (0.3)	8
	Female	1 (0.2)	1 (0.2)	0 (0)	2 (0.3)	4 (0.5)	3 (0.3)	0 (0)	7 (0.8)	9
1.1-5	Male	38 (6.4)	29 (4.9)	13 (2.2)	80 (13.5)	20 (2.3)	32 (3.6)	1 (0.1)	53 (6)	133
	Female	16 (2.7)	15 (2.5)	5 (0.8)	36 (6.1)	9 (1)	12 (1.4)	0 (0)	21 (2.4)	57
5.1-15	Male	67 (11.3)	42 (7.1)	8 (1.3)	117 (19.7)	30 (3.4)	38 (4.3)	2 (0.2)	70 (7.9)	187
	Female	27 (4.5)	24 (4)	3 (0.5)	54 (9.1)	12 (1.4)	33 (3.7)	2 (0.2)	47 (5.3)	101
15.1-39	Male	45 (7.6)	28 (4.7)	24 (4)	97 (16.3)	78 (8.8)	300 (33.9)	7 (0.8)	385 (43.6)	482
	Female	18 (3)	26 (4.4)	7 (1.2)	51 (8.6)	13 (1.5)	103 (11.7)	2 (0.2)	118 (13.3)	169
39.1 &	Male	43 (7.2)	39 (6.6)	19 (3.2)	101 (17)	32 (3.6)	80 (9)	0 (0)	112 (12.7)	213
	Female	21 (3.5)	23 (3.9)	7 (1.2)	51 (8.6)	7 (0.8)	61 (6.9)	0 (0)	68 (7.7)	119
Total		276 (46.5)	231 (38.9)	87 (14.6)	594 (100)	206 (23.3)	664 (75.1)	14 (1.6)	884 (100)	1478

Family Type

Age Group	Gender	Nuclear	Extended	No info	Total	Nuclear	Extended	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	0 (0)	0 (0)	5 (0.8)	5 (0.8)	1 (0.1)	2 (0.2)	0 (0)	3 (0.3)	8
	Female	0 (0)	1 (0.2)	1 (0.2)	2 (0.3)	1 (0.1)	5 (0.6)	1 (0.1)	7 (0.8)	9
1.1-5	Male	16 (2.7)	39 (6.6)	25 (4.2)	80 (13.5)	25 (2.8)	28 (3.2)	0 (0)	53 (6)	133
	Female	8 (1.3)	17 (2.9)	11 (1.9)	36 (6.1)	9 (1)	11 (1.2)	1 (0.1)	21 (2.4)	57
5.1-15	Male	37 (6.2)	55 (9.3)	25 (4.2)	117 (19.7)	33 (3.7)	31 (3.5)	6 (0.7)	70 (7.9)	187
	Female	14 (2.4)	27 (4.5)	13 (2.2)	54 (9.1)	27 (3.1)	16 (1.8)	4 (0.5)	47 (5.3)	101
15.1-39	Male	15 (2.5)	19 (3.2)	63 (10.6)	97 (16.3)	30 (3.4)	120 (13.6)	235 (26.6)	385 (43.6)	482
	Female	2 (0.3)	12 (2)	37 (6.2)	51 (8.6)	14 (1.6)	22 (2.5)	82 (9.3)	118 (13.3)	169
39.1 &	Male	14 (2.4)	20 (3.4)	67 (11.3)	101 (17)	17 (1.9)	28 (3.2)	67 (7.6)	112 (12.7)	213
	Female	1 (0.2)	13 (2.2)	37 (6.2)	51 (8.6)	5 (0.6)	8 (0.9)	55 (6.2)	68 (7.7)	119
Total		107 (18)	203 (34.2)	284 (47.8)	594 (100)	162 (18.3)	271 (30.7)	451 (51)	884 (100)	1478

Carbonated Drinks

Age Group	Gender	Yes	No	No info	Total	Yes	No	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	0 (0)	4 (0.7)	1 (0.2)	5 (0.8)	2 (0.2)	1 (0.1)	0 (0)	3 (0.3)	8
	Female	1 (0.2)	1 (0.2)	0 (0)	2 (0.3)	2 (0.2)	5 (0.6)	0 (0)	7 (0.8)	9
1.1-5	Male	47 (7.9)	20 (3.4)	13 (2.2)	80 (13.5)	25 (2.8)	26 (2.9)	2 (0.2)	53 (6)	133
	Female	21 (3.5)	10 (1.7)	5 (0.8)	36 (6.1)	8 (0.9)	13 (1.5)	0 (0)	21 (2.4)	57
5.1-15	Male	78 (13.1)	30 (5.1)	9 (1.5)	117 (19.7)	50 (5.7)	18 (2)	2 (0.2)	70 (7.9)	187
	Female	40 (6.7)	11 (1.9)	3 (0.5)	54 (9.1)	36 (4.1)	9 (1)	2 (0.2)	47 (5.3)	101
15.1-39	Male	49 (8.2)	24 (4)	24 (4)	97 (16.3)	245 (27.7)	134 (15.2)	6 (0.7)	385 (43.6)	482
	Female	28 (4.7)	14 (2.4)	9 (1.5)	51 (8.6)	72 (8.1)	44 (5)	2 (0.2)	118 (13.3)	169
39.1 &	Male	44 (7.4)	37 (6.2)	20 (3.4)	101 (17)	53 (6)	58 (6.6)	1 (0.1)	112 (12.7)	213
	Female	19 (3.2)	25 (4.2)	7 (1.2)	51 (8.6)	26 (2.9)	42 (4.8)	0 (0)	68 (7.7)	119
Total		327 (55.1)	176 (29.6)	91 (15.3)	594 (100)	519 (58.7)	350 (39.6)	15 (1.7)	884 (100)	1478

Tobacco Use

Age Group	Gender	Yes	No	No info	Total	Yes	No	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	0 (0)	4 (0.7)	1 (0.2)	5 (0.8)	1 (0.1)	2 (0.2)	0 (0)	3 (0.3)	8
	Female	1 (0.2)	1 (0.2)	0 (0)	2 (0.3)	2 (0.2)	5 (0.6)	0 (0)	7 (0.8)	9
1.1-5	Male	22 (3.7)	44 (7.4)	14 (2.4)	80 (13.5)	27 (3.1)	26 (2.9)	0 (0)	53 (6)	133
	Female	12 (2)	19 (3.2)	5 (0.8)	36 (6.1)	6 (0.7)	15 (1.7)	0 (0)	21 (2.4)	57
5.1-15	Male	42 (7.1)	66 (11.1)	9 (1.5)	117 (19.7)	41 (4.6)	27 (3.1)	2 (0.2)	70 (7.9)	187
	Female	20 (3.4)	31 (5.2)	3 (0.5)	54 (9.1)	20 (2.3)	25 (2.8)	2 (0.2)	47 (5.3)	101
15.1-39	Male	22 (3.7)	51 (8.6)	24 (4)	97 (16.3)	90 (10.2)	289 (32.7)	6 (0.7)	385 (43.6)	482
	Female	10 (1.7)	34 (5.7)	7 (1.2)	51 (8.6)	11 (1.2)	105 (11.9)	2 (0.2)	118 (13.3)	169
39.1 &	Male	41 (6.9)	41 (6.9)	19 (3.2)	101 (17)	29 (3.3)	83 (9.4)	0 (0)	112 (12.7)	213
	Female	10 (1.7)	34 (5.7)	7 (1.2)	51 (8.6)	6 (0.7)	62 (7)	0 (0)	68 (7.7)	119
Total		180 (30.3)	325 (54.7)	89 (15)	594 (100)	233 (26.4)	639 (72.3)	12 (1.4)	884 (100)	1478

Microwave Use

Age Group	Gender	Yes	No	No info	Total	Yes	No	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	2 (0.3)	3 (0.5)	0 (0)	5 (0.8)	1 (0.1)	1 (0.1)	1 (0.1)	3 (0.3)	8
	Female	0 (0)	1 (0.2)	1 (0.2)	2 (0.3)	0 (0)	7 (0.8)	0 (0)	7 (0.8)	9
1.1-5	Male	3 (0.5)	57 (9.6)	20 (3.4)	80 (13.5)	9 (1)	42 (4.8)	2 (0.2)	53 (6)	133
	Female	2 (0.3)	28 (4.7)	6 (1)	36 (6.1)	7 (0.8)	13 (1.5)	1 (0.1)	21 (2.4)	57
5.1-15	Male	9 (1.5)	83 (14)	25 (4.2)	117 (19.7)	14 (1.6)	54 (6.1)	2 (0.2)	70 (7.9)	187
	Female	8 (1.3)	36 (6.1)	10 (1.7)	54 (9.1)	7 (0.8)	38 (4.3)	2 (0.2)	47 (5.3)	101
15.1-39	Male	5 (0.8)	62 (10.4)	30 (5.1)	97 (16.3)	133 (15)	245 (27.7)	7 (0.8)	385	482
	Female	9 (1.5)	34 (5.7)	8 (1.3)	51 (8.6)	53 (6)	63 (7.1)	2 (0.2)	118	169
39.1 &	Male	10 (1.7)	64 (10.8)	27 (4.5)	101 (17)	40 (4.5)	72 (8.1)	0 (0)	112	213
	Female	5 (0.8)	37 (6.2)	9 (1.5)	51 (8.6)	30 (3.4)	38 (4.3)	0 (0)	68 (7.7)	119
Total		53 (8.9)	405 (68.2)	136 (22.9)	594 (100)	294 (33.3)	573 (64.8)	17 (1.9)	884 (100)	1478

Perfume Use

Age Group	Gender	Yes	No	No info	Total	Yes	No	No info	Total	Net Total
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N
<1	Male	0 (0)	4 (0.7)	1 (0.2)	5 (0.8)	3 (0.3)	0 (0)	0 (0)	3 (0.3)	8
	Female	2 (0.3)	0 (0)	0 (0)	2 (0.3)	3 (0.3)	4 (0.5)	0 (0)	7 (0.8)	9
1.1-5	Male	43 (7.2)	23 (3.9)	14 (2.4)	80 (13.5)	40 (4.5)	11 (1.2)	2 (0.2)	53 (6)	133
	Female	19 (3.2)	12 (2)	5 (0.8)	36 (6.1)	14 (1.6)	6 (0.7)	1 (0.1)	21 (2.4)	57
5.1-15	Male	63 (10.6)	45 (7.6)	9 (1.5)	117 (19.7)	49 (5.5)	19 (2.1)	2 (0.2)	70 (7.9)	187
	Female	28 (4.7)	23 (3.9)	3 (0.5)	54 (9.1)	31 (3.5)	14 (1.6)	2 (0.2)	47 (5.3)	101
15.1-39	Male	38 (6.4)	34 (5.7)	25 (4.2)	97 (16.3)	298 (33.7)	81 (9.2)	6 (0.7)	385 (43.6)	482
	Female	16 (2.7)	26 (4.4)	9 (1.5)	51 (8.6)	85 (9.6)	31 (3.5)	2 (0.2)	118 (13.3)	169
39.1 &	Male	35 (5.9)	46 (7.7)	20 (3.4)	101 (17)	68 (7.7)	43 (4.9)	1 (0.1)	112 (12.7)	213
	Female	11 (1.9)	33 (5.6)	7 (1.2)	51 (8.6)	29 (3.3)	39 (4.4)	0 (0)	68 (7.7)	119
Total		255 (42.9)	246 (41.4)	93 (15.7)	594 (100)	620 (70.1)	248 (28.1)	16 (1.8)	884 (100)	1478

Table 3.7B: Overall origin and nature-wise distribution of leukemia patients against age, gender, and risk factors (n=594)

Age Group	Status	Gender	Disease Origin		Disease Nature		Total
			Lymphoid	Myeloid	Acute	Chronic	
-			No. (%)	No. (%)	No. (%)	No. (%)	No. (%)
Ethnicity							
<1	Pathan	Male	2 (0.34)	1 (0.17)	2 (0.34)	1 (0.17)	3 (0.51)
		Female	2 (0.34)	0 (0)	2 (0.34)	0 (0)	2 (0.34)
	Punjabi	Male	2 (0.34)	0 (0)	2 (0.34)	0 (0)	2 (0.34)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Others	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Unavailable	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
1.1-5	Pathan	Male	46 (7.74)	5 (0.84)	50 (8.42)	1 (0.17)	51 (8.59)
		Female	16 (2.69)	1 (0.17)	17 (2.86)	0 (0)	17 (2.86)
	Punjabi	Male	24 (4.04)	0 (0)	24 (4.04)	0 (0)	24 (4.04)
		Female	13 (2.19)	2 (0.34)	14 (2.36)	1 (0.17)	15 (2.53)
	Others	Male	4 (0.67)	0 (0)	4 (0.67)	0 (0)	4 (0.67)
		Female	3 (0.51)	0 (0)	3 (0.51)	0 (0)	3 (0.51)
	Unavailable	Male	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
5.1-15	Pathan	Male	69 (11.62)	6 (1.01)	74 (12.46)	1 (0.17)	75 (12.63)
		Female	26 (4.38)	6 (1.01)	31 (5.22)	1 (0.17)	32 (5.39)
	Punjabi	Male	24 (4.04)	7 (1.18)	30 (5.05)	1 (0.17)	31 (5.22)
		Female	17 (2.86)	0 (0)	17 (2.86)	0 (0)	17 (2.86)
	Others	Male	9 (1.52)	1 (0.17)	10 (1.68)	0 (0)	10 (1.68)
		Female	4 (0.67)	0 (0)	4 (0.67)	0 (0)	4 (0.67)
	Unavailable	Male	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
		Female	0 (0)	1 (0.17)	1 (0.17)	0 (0)	1 (0.17)
15.1-39	Pathan	Male	39 (6.57)	17 (2.86)	46 (7.74)	10 (1.68)	56 (9.43)

		Female	9 (1.52)	12 (2.02)	16 (2.69)	5 (0.84)	21 (3.54)
	Punjabi	Male	12 (2.02)	18 (3.03)	18 (3.03)	12 (2.02)	30 (5.05)
		Female	2 (0.34)	19 (3.2)	6 (1.01)	15 (2.53)	21 (3.54)
	Others	Male	3 (0.51)	1 (0.17)	3 (0.51)	1 (0.17)	4 (0.67)
		Female	2 (0.34)	4 (0.67)	5 (0.84)	1 (0.17)	6 (1.01)
	Unavailable	Male	5 (0.84)	2 (0.34)	6 (1.01)	1 (0.17)	7 (1.18)
		Female	1 (0.17)	2 (0.34)	3 (0.51)	0 (0)	3 (0.51)
39.1 & >	Pathan	Male	29 (4.88)	28 (4.71)	15 (2.53)	42 (7.07)	57 (9.6)
		Female	7 (1.18)	16 (2.69)	9 (1.52)	14 (2.36)	23 (3.87)
	Punjabi	Male	6 (1.01)	25 (4.21)	7 (1.18)	24 (4.04)	31 (5.22)
		Female	0 (0)	20 (3.37)	0 (0)	20 (3.37)	20 (3.37)
	Others	Male	2 (0.34)	5 (0.84)	2 (0.34)	5 (0.84)	7 (1.18)
		Female	1 (0.17)	6 (1.01)	3 (0.51)	4 (0.67)	7 (1.18)
	Unavailable	Male	0 (0)	6 (1.01)	3 (0.51)	3 (0.51)	6 (1.01)
		Female	0 (0)	1 (0.17)	1 (0.17)	0 (0)	1 (0.17)
Total			382 (64.31)	212 (35.69)	431 (72.56)	163 (27.44)	594 (100)
Area			No. (%)	No. (%)	No. (%)	No. (%)	No. (%)
<1	Rural	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	Urban	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Unavailable	Male	4 (0.67)	1 (0.17)	4 (0.67)	1 (0.17)	5 (0.84)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
1.1-5	Rural	Male	44 (7.41)	3 (0.51)	47 (7.91)	0 (0)	47 (7.91)
		Female	18 (3.03)	1 (0.17)	18 (3.03)	1 (0.17)	19 (3.2)
	Urban	Male	9 (1.52)	0 (0)	9 (1.52)	0 (0)	9 (1.52)
		Female	6 (1.01)	1 (0.17)	7 (1.18)	0 (0)	7 (1.18)
	Unavailable	Male	22 (3.7)	2 (0.34)	23 (3.87)	1 (0.17)	24 (4.04)
		Female	9 (1.52)	1 (0.17)	10 (1.68)	0 (0)	10 (1.68)
5.1-15	Rural	Male	65 (10.94)	9 (1.52)	73 (12.29)	1 (0.17)	74 (12.46)
		Female	25 (4.21)	4 (0.67)	28 (4.71)	1 (0.17)	29 (4.88)

	Urban	Male	19 (3.2)	1 (0.17)	20 (3.37)	0 (0)	20 (3.37)
		Female	12 (2.02)	0 (0)	12 (2.02)	0 (0)	12 (2.02)
	Unavailable	Male	19 (3.2)	4 (0.67)	22 (3.7)	1 (0.17)	23 (3.87)
		Female	10 (1.68)	3 (0.51)	13 (2.19)	0 (0)	13 (2.19)
15.1-39	Rural	Male	23 (3.87)	8 (1.35)	25 (4.21)	6 (1.01)	31 (5.22)
		Female	5 (0.84)	6 (1.01)	9 (1.52)	2 (0.34)	11 (1.85)
	Urban	Male	5 (0.84)	0 (0)	5 (0.84)	0 (0)	5 (0.84)
		Female	1 (0.17)	2 (0.34)	2 (0.34)	1 (0.17)	3 (0.51)
	Unavailable	Male	31 (5.22)	30 (5.05)	43 (7.24)	18 (3.03)	61 (10.27)
		Female	8 (1.35)	29 (4.88)	19 (3.2)	18 (3.03)	37 (6.23)
39.1 & >	Rural	Male	16 (2.69)	12 (2.02)	5 (0.84)	23 (3.87)	28 (4.71)
		Female	5 (0.84)	8 (1.35)	4 (0.67)	9 (1.52)	13 (2.19)
	Urban	Male	3 (0.51)	3 (0.51)	3 (0.51)	3 (0.51)	6 (1.01)
		Female	0 (0)	1 (0.17)	1 (0.17)	0 (0)	1 (0.17)
	Unavailable	Male	18 (3.03)	49 (8.25)	19 (3.2)	48 (8.08)	67 (11.28)
		Female	3 (0.51)	34 (5.72)	8 (1.35)	29 (4.88)	37 (6.23)
Total			382 (64.31)	212 (35.69)	431 (72.56)	163 (27.44)	594 (100)
Water Type							
<1	Groundwater	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Surface water	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	Treated Water	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Unavailable	Male	4 (0.67)	1 (0.17)	4 (0.67)	1 (0.17)	5 (0.84)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
1.1-5	Groundwater	Male	42 (7.07)	2 (0.34)	44 (7.41)	0 (0)	44 (7.41)
		Female	16 (2.69)	2 (0.34)	17 (2.86)	1 (0.17)	18 (3.03)
	Surface water	Male	8 (1.35)	1 (0.17)	9 (1.52)	0 (0)	9 (1.52)
		Female	5 (0.84)	0 (0)	5 (0.84)	0 (0)	5 (0.84)
	Treated Water	Male	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)

		Female	3 (0.51)	0 (0)	3 (0.51)	0 (0)	3 (0.51)
	Unavailable	Male	24 (4.04)	2 (0.34)	25 (4.21)	1 (0.17)	26 (4.38)
		Female	9 (1.52)	1 (0.17)	10 (1.68)	0 (0)	10 (1.68)
5.1-15	Groundwater	Male	56 (9.43)	9 (1.52)	64 (10.77)	1 (0.17)	65 (10.94)
		Female	29 (4.88)	4 (0.67)	32 (5.39)	1 (0.17)	33 (5.56)
	Surface water	Male	21 (3.54)	1 (0.17)	22 (3.7)	0 (0)	22 (3.7)
		Female	7 (1.18)	0 (0)	7 (1.18)	0 (0)	7 (1.18)
	Treated Water	Male	5 (0.84)	0 (0)	5 (0.84)	0 (0)	5 (0.84)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	Unavailable	Male	21 (3.54)	4 (0.67)	24 (4.04)	1 (0.17)	25 (4.21)
		Female	10 (1.68)	3 (0.51)	13 (2.19)	0 (0)	13 (2.19)
15.1-39	Groundwater	Male	21 (3.54)	4 (0.67)	20 (3.37)	5 (0.84)	25 (4.21)
		Female	5 (0.84)	5 (0.84)	7 (1.18)	3 (0.51)	10 (1.68)
	Surface-water	Male	5 (0.84)	2 (0.34)	7 (1.18)	0 (0)	7 (1.18)
		Female	1 (0.17)	3 (0.51)	4 (0.67)	0 (0)	4 (0.67)
	Treated Water	Male	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Unavailable	Male	32 (5.39)	32 (5.39)	45 (7.58)	19 (3.2)	64 (10.77)
		Female	8 (1.35)	29 (4.88)	19 (3.2)	18 (3.03)	37 (6.23)
39.1 & >	Ground water	Male	14 (2.36)	11 (1.85)	7 (1.18)	18 (3.03)	25 (4.21)
		Female	3 (0.51)	5 (0.84)	2 (0.34)	6 (1.01)	8 (1.35)
	Surface water	Male	6 (1.01)	4 (0.67)	1 (0.17)	9 (1.52)	10 (1.68)
		Female	2 (0.34)	3 (0.51)	2 (0.34)	3 (0.51)	5 (0.84)
	Treated Water	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Unavailable	Male	17 (2.86)	49 (8.25)	19 (3.2)	47 (7.91)	66 (11.11)
		Female	3 (0.51)	35 (5.89)	9 (1.52)	29 (4.88)	38 (6.4)
Fuel							
<1	Wood	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	Gas	Male	3 (0.51)	1 (0.17)	3 (0.51)	1 (0.17)	4 (0.67)

		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	Unavailable	Male	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
1.1-5	Wood	Male	36 (6.06)	2 (0.34)	38 (6.4)	0 (0)	38 (6.4)
		Female	15 (2.53)	1 (0.17)	15 (2.53)	1 (0.17)	16 (2.69)
	Gas	Male	27 (4.55)	2 (0.34)	29 (4.88)	0 (0)	29 (4.88)
		Female	13 (2.19)	2 (0.34)	15 (2.53)	0 (0)	15 (2.53)
	Unavailable	Male	12 (2.02)	1 (0.17)	12 (2.02)	1 (0.17)	13 (2.19)
		Female	5 (0.84)	0 (0)	5 (0.84)	0 (0)	5 (0.84)
5.1-15	Wood	Male	61 (10.27)	6 (1.01)	66 (11.11)	1 (0.17)	67 (11.28)
		Female	22 (3.7)	5 (0.84)	26 (4.38)	1 (0.17)	27 (4.55)
	Gas	Male	38 (6.4)	4 (0.67)	42 (7.07)	0 (0)	42 (7.07)
		Female	23 (3.87)	1 (0.17)	24 (4.04)	0 (0)	24 (4.04)
	Unavailable	Male	4 (0.67)	4 (0.67)	7 (1.18)	1 (0.17)	8 (1.35)
		Female	2 (0.34)	1 (0.17)	3 (0.51)	0 (0)	3 (0.51)
15.1-39	Wood	Male	29 (4.88)	16 (2.69)	35 (5.89)	10 (1.68)	45 (7.58)
		Female	7 (1.18)	11 (1.85)	10 (1.68)	8 (1.35)	18 (3.03)
	Gas	Male	14 (2.36)	14 (2.36)	19 (3.2)	9 (1.52)	28 (4.71)
		Female	5 (0.84)	21 (3.54)	14 (2.36)	12 (2.02)	26 (4.38)
	Unavailable	Male	16 (2.69)	8 (1.35)	19 (3.2)	5 (0.84)	24 (4.04)
		Female	2 (0.34)	5 (0.84)	6 (1.01)	1 (0.17)	7 (1.18)
39.1 & >	Wood	Male	19 (3.2)	24 (4.04)	10 (1.68)	33 (5.56)	43 (7.24)
		Female	6 (1.01)	15 (2.53)	6 (1.01)	15 (2.53)	21 (3.54)
	Gas	Male	7 (1.18)	32 (5.39)	7 (1.18)	32 (5.39)	39 (6.57)
		Female	0 (0)	23 (3.87)	3 (0.51)	20 (3.37)	23 (3.87)
	Unavailable	Male	11 (1.85)	8 (1.35)	10 (1.68)	9 (1.52)	19 (3.2)
		Female	2 (0.34)	5 (0.84)	4 (0.67)	3 (0.51)	7 (1.18)
Total			382 (64.31)	212 (35.69)	431 (72.56)	163 (27.44)	594 (100)
Family Type							
<1	Nuclear	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

	Extended	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	Unavailable	Male	4 (0.67)	1 (0.17)	4 (0.67)	1 (0.17)	5 (0.84)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
1.1-5	Nuclear	Male	15 (2.53)	1 (0.17)	16 (2.69)	0 (0)	16 (2.69)
		Female	7 (1.18)	1 (0.17)	8 (1.35)	0 (0)	8 (1.35)
	Extended	Male	37 (6.23)	2 (0.34)	39 (6.57)	0 (0)	39 (6.57)
		Female	16 (2.69)	1 (0.17)	16 (2.69)	1 (0.17)	17 (2.86)
	Unavailable	Male	23 (3.87)	2 (0.34)	24 (4.04)	1 (0.17)	25 (4.21)
		Female	10 (1.68)	1 (0.17)	11 (1.85)	0 (0)	11 (1.85)
5.1-15	Nuclear	Male	32 (5.39)	5 (0.84)	36 (6.06)	1 (0.17)	37 (6.23)
		Female	14 (2.36)	0 (0)	14 (2.36)	0 (0)	14 (2.36)
	Extended	Male	50 (8.42)	5 (0.84)	55 (9.26)	0 (0)	55 (9.26)
		Female	23 (3.87)	4 (0.67)	26 (4.38)	1 (0.17)	27 (4.55)
	Unavailable	Male	21 (3.54)	4 (0.67)	24 (4.04)	1 (0.17)	25 (4.21)
		Female	10 (1.68)	3 (0.51)	13 (2.19)	0 (0)	13 (2.19)
15.1-39	Nuclear	Male	13 (2.19)	2 (0.34)	13 (2.19)	2 (0.34)	15 (2.53)
		Female	2 (0.34)	0 (0)	2 (0.34)	0 (0)	2 (0.34)
	Extended	Male	15 (2.53)	4 (0.67)	16 (2.69)	3 (0.51)	19 (3.2)
		Female	4 (0.67)	8 (1.35)	9 (1.52)	3 (0.51)	12 (2.02)
	Unavailable	Male	31 (5.22)	32 (5.39)	44 (7.41)	19 (3.2)	63 (10.61)
		Female	8 (1.35)	29 (4.88)	19 (3.2)	18 (3.03)	37 (6.23)
39.1 & >	Nuclear	Male	8 (1.35)	6 (1.01)	4 (0.67)	10 (1.68)	14 (2.36)
		Female	0 (0)	1 (0.17)	1 (0.17)	0 (0)	1 (0.17)
	Extended	Male	12 (2.02)	8 (1.35)	3 (0.51)	17 (2.86)	20 (3.37)
		Female	5 (0.84)	8 (1.35)	4 (0.67)	9 (1.52)	13 (2.19)
	Unavailable	Male	17 (2.86)	50 (8.42)	20 (3.37)	47 (7.91)	67 (11.28)
		Female	3 (0.51)	34 (5.72)	8 (1.35)	29 (4.88)	37 (6.23)
Total			382 (64.31)	212 (35.69)	431 (72.56)	163 (27.44)	594 (100)
Carbonated Drinks							
<1	Yes	Male	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	No	Male	3 (0.51)	1 (0.17)	3 (0.51)	1 (0.17)	4 (0.67)
		Female	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
	Unavailable	Male	1 (0.17)	0 (0)	1 (0.17)	0 (0)	1 (0.17)
		Female	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
1.1-5	Yes	Male	45 (7.58)	2 (0.34)	47 (7.91)	0 (0)	47 (7.91)
		Female	20 (3.37)	1 (0.17)	20 (3.37)	1 (0.17)	21 (3.54)
	No	Male	18 (3.03)	2 (0.34)	20 (3.37)	0 (0)	20 (3.37)
		Female	8 (1.35)	2 (0.34)	10 (1.68)	0 (0)	10 (1.68)
	Unavailable	Male	12 (2.02)	1 (0.17)	12 (2.02)	1 (0.17)	13 (2.19)
		Female	5 (0.84)	0 (0)	5 (0.84)	0 (0)	5 (0.84)
5.1-15	Yes	Male	73 (12.29)	5 (0.84)	77 (12.96)	1 (0.17)	78 (13.13)
		Female	36 (6.06)	4 (0.67)	39 (6.57)	1 (0.17)	40 (6.73)
	No	Male	25 (4.21)	5 (0.84)	30 (5.05)	0 (0)	30 (5.05)
		Female	9 (1.52)	2 (0.34)	11 (1.85)	0 (0)	11 (1.85)
	Unavailable	Male	5 (0.84)	4 (0.67)	8 (1.35)	1 (0.17)	9 (1.52)
		Female	2 (0.34)	1 (0.17)	3 (0.51)	0 (0)	3 (0.51)
15.1-39	Yes	Male	30 (5.05)	19 (3.2)	38 (6.4)	11 (1.85)	49 (8.25)
		Female	8 (1.35)	20 (3.37)	16 (2.69)	12 (2.02)	28 (4.71)
	No	Male	13 (2.19)	11 (1.85)	16 (2.69)	8 (1.35)	24 (4.04)
		Female	3 (0.51)	11 (1.85)	6 (1.01)	8 (1.35)	14 (2.36)
	Unavailable	Male	16 (2.69)	8 (1.35)	19 (3.2)	5 (0.84)	24 (4.04)
		Female	3 (0.51)	6 (1.01)	8 (1.35)	1 (0.17)	9 (1.52)
39.1 & >	Yes	Male	17 (2.86)	27 (4.55)	11 (1.85)	33 (5.56)	44 (7.41)
		Female	4 (0.67)	15 (2.53)	4 (0.67)	15 (2.53)	19 (3.2)
	No	Male	9 (1.52)	28 (4.71)	5 (0.84)	32 (5.39)	37 (6.23)
		Female	2 (0.34)	23 (3.87)	5 (0.84)	20 (3.37)	25 (4.21)
	Unavailable	Male	11 (1.85)	9 (1.52)	11 (1.85)	9 (1.52)	20 (3.37)
		Female	2 (0.34)	5 (0.84)	4 (0.67)	3 (0.51)	7 (1.18)
Total			382 (64.31)	212 (35.69)	431 (72.56)	163 (27.44)	594 (100)

3.1.2.1 Age and gender-wise distribution of demographic and lifestyle among patients and controls

3.1.2.1.1 Ethnicity

According to our data, leukemia was most prevalent in the Pathan population, about 57% of the total, participants with Punjabi ethnicity were 31% and other ethnicities (Baloch, Kashmiri, Hindkuwan, Balti, Sindhi, and Afghan) were collectively only 4.2% of the total (Table 3.6). The representation of Pathan ethnicity is highly variable across each group in frequency, however, the number of Punjabis, although about half than Pathan in each age group, was nearly similar in each group. Pathans were prominent in each age group, they are 2.2 times greater in frequency in children (18.01%), which is the highest difference between Pathan and Punjabis in any other group (Table 3.7A, Figure 3.15A). In general, males with Pathans ethnicity dominated every group, while females with Pathan ethnicity were higher in children, while the difference declined as the age group is raised. The huge difference that exists between the Pathan and Punjabi races within the age groups is due to male patients, as the females in the same groups have less or even negligible difference except for females in children, where out of total females in the group, Pathan females (59.20%) were 1.8 times more prominent than Punjabi females (31.48%) in the same group (Table 3.7A).

Lymphoid leukemia (92%) is more than ten times more prevalent than myeloid leukemia (8%) in Pathan children, twice more common in AYAs males, while both are the same (51%, 49%) in Pathan ethnicity in the aged group in Pathan males. Myeloid leukemia is more common in AYAs and aged Punjabis in general. In AYAs Pathan, acute (82%) is more prevalent, while in Punjabi AYAs male, chronic (71%) is more common. Aged, chronic is Overall more common in both Pathan (73%) and Punjabis (77%) (Table 3.17B).

Among all 844 controls, 3.69% were Pathan, 55.88% were Punjabi and 6.33% were from other ethnicities collectively (Table 3.6). In age groupwise distribution, Punjabis were the most frequent participants across each group and gender (Table 3.7A, Figure 3.15B)

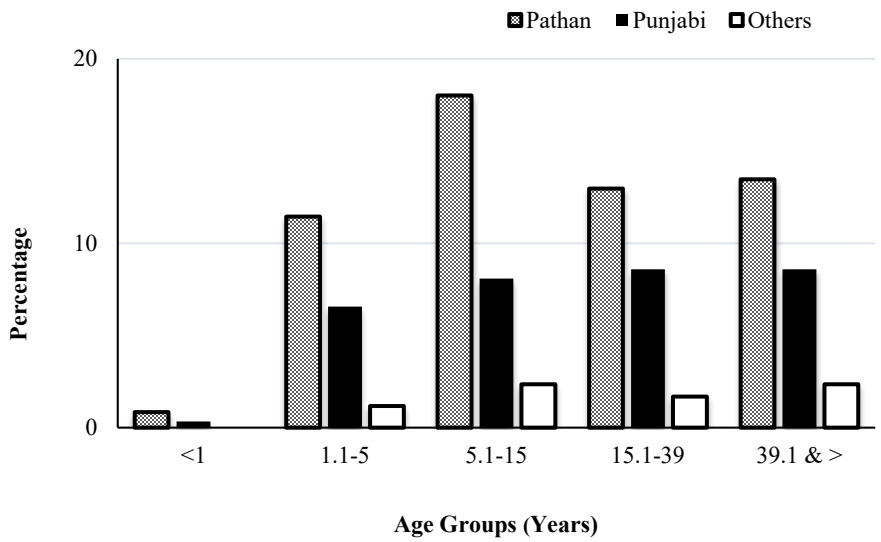


Figure 3.15A: Overall age group-wise percentage distribution of leukemia patients against ethnicity (n=594)

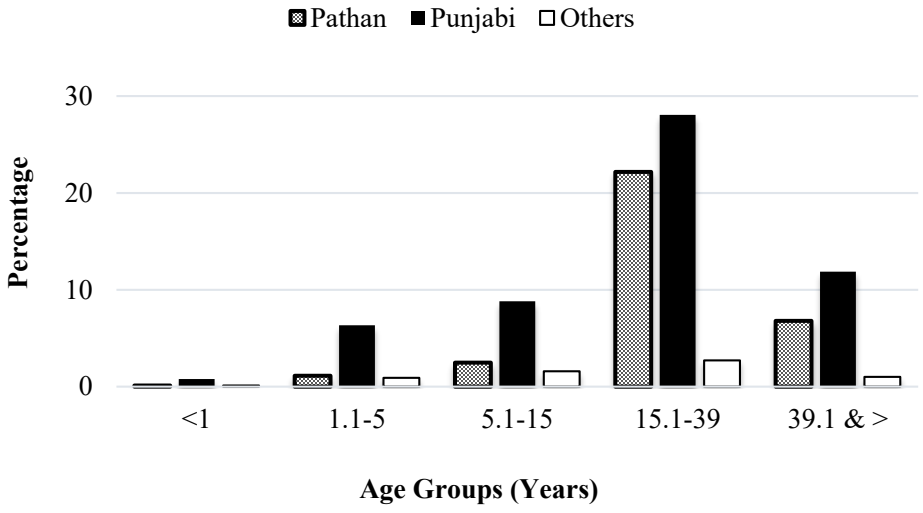


Figure 3.15B: Overall age groupwise percentage distribution of controls against ethnicity (n=884)

3.1.2.1.2 Education

In our data, 46% of all the known were educated, among which 23% were primarily passed, 19% were secondary passed and 4% were graduates (Table 3.6). About 20% of the total population, comprised of preschoolers were inadmissible by default. However, 17% were not educated. In general, the overall population was educated (Table 3.7A, Figure 3.16A).

Among all 844 controls, 78.51% were educated, 11.09% had no normal education, and 6.56% were inadmissible (Table 3.6). In age groupwise distribution, overall, a higher number of the control participants were educated. In AYAs and adults, 5.43% and 4.75% of the individuals had no formal education, comprised of both male and female controls (Table 3.7A, Figure 3.16B)

.

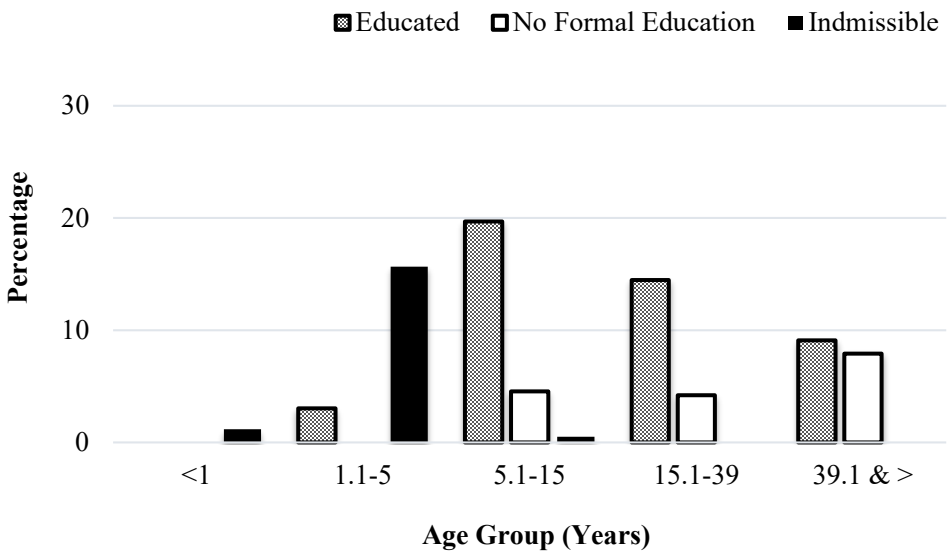


Figure 3.16A: Overall age groupwise percentage distribution of leukemia patients against education (n=594)

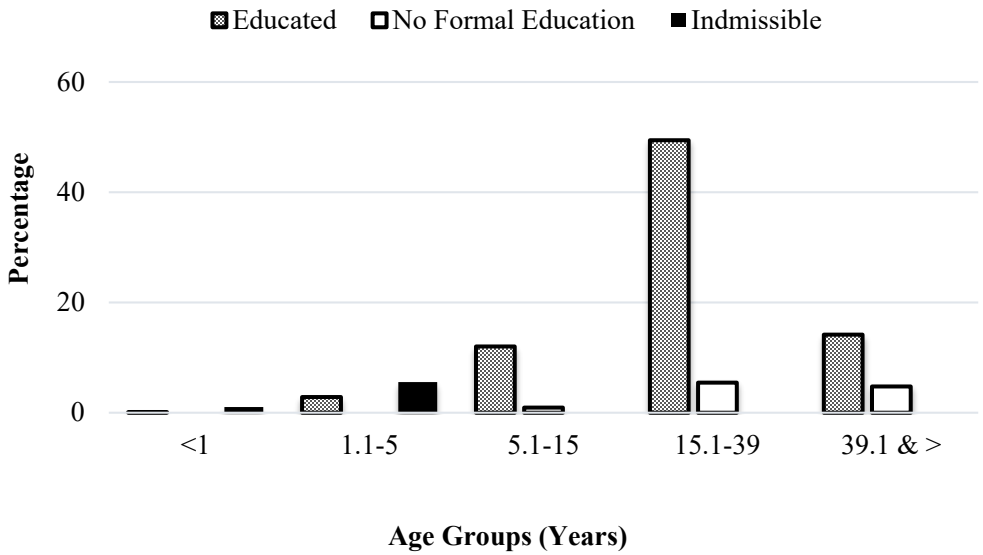


Figure 3.16B: Overall age groupwise percentage distribution of controls against education (n=884)

3.1.2.1.3 Consanguinity

Among all the known parental union types, consanguineous unions were more prevalent in children (13.47%), while non-consanguineous unions were more observed in AYAs (13.3%) and adults (14.48%). However, in preschoolers, the frequency of both union types is the same (Table 3.7A, Figure 3.17A, 3.18). In children, males are the major contributors towards the higher CU unions (CU=49%, NCU=39%), as the females in the same group also have the same percentage of CU (42%) and NCU (46%) parental unions, within each group. In general, across each group, more males with leukemia are borne to consanguineous union couples, and more females with leukemia were borne to nonconsanguineous unions (Table 3.7A).

Among all the controls, 21.83% had parental consanguinity, while 69% had non-consanguineous parental unions (Table 3.6). About 5% of the patients, comprised of preschoolers and 6.45% in the age group 5.1-15 years, and 7% in AYAs had parental consanguinity, which more frequently occurred in males (Table 3.7A, Figure 3.17B, 3.18).

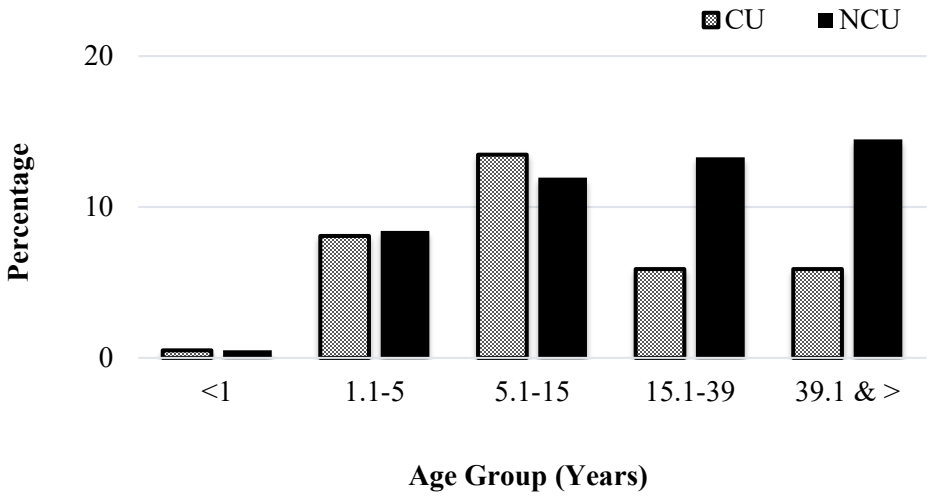


Figure 3.17A. Overall age groupwise percentage distribution of leukemia patients against parental consanguinity (n=594)

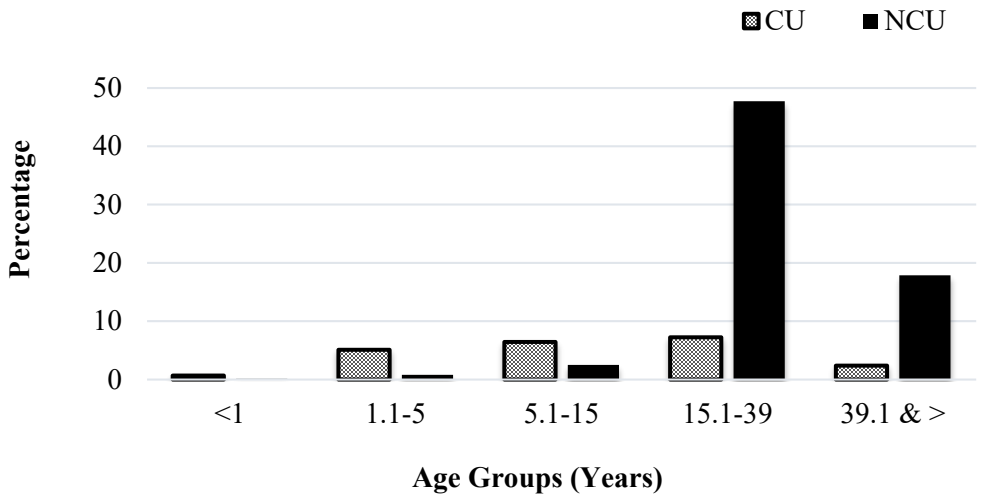


Figure 3.17B. Overall age grouwise percentage distribution of controls against parental consanguinity (n=884)

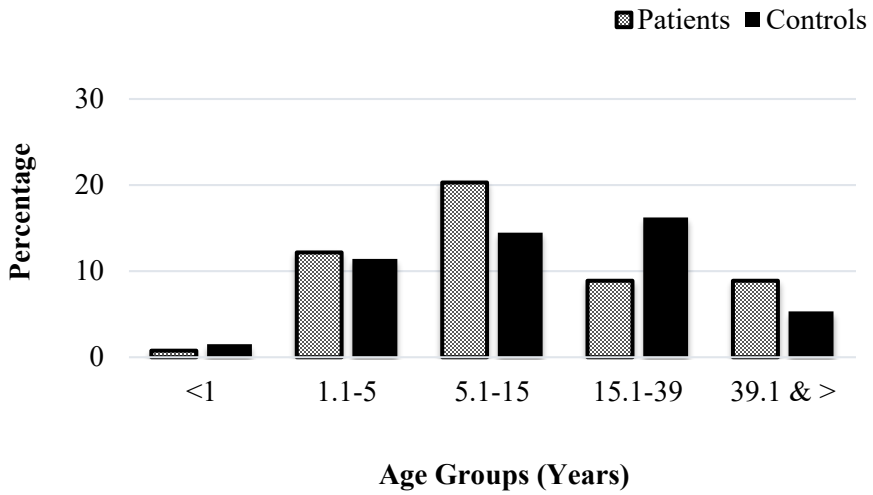


Figure 3.18: Comparison of parental consanguinity between leukemia patients and controls

3.1.2. 1.4 Family history

A family history of leukemia and other cancers was recorded for all the samples in our research population. About 5% of the patients had cancer running in their families (Table 3.6). Of which 5.39%, 2,53% were from children (Table 3.7A, Figure 3.19A, 3.20).

Among all the controls, 1.36% had a family history of cancer, while 98% has no family history of any type of cancer (Table 3.6). In age group distribution, there is a negligible number of controls having some history of cancer in the family in both genders (Table 3.7A, Figure 3.19B, 3.20).

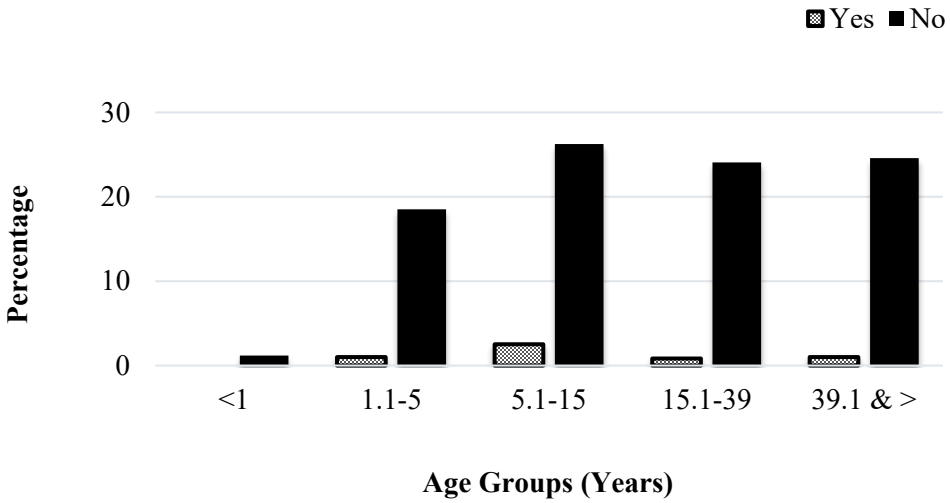


Figure 3.19A: Overall age group-wise percentage distribution of leukemia patients against family history (n=594)

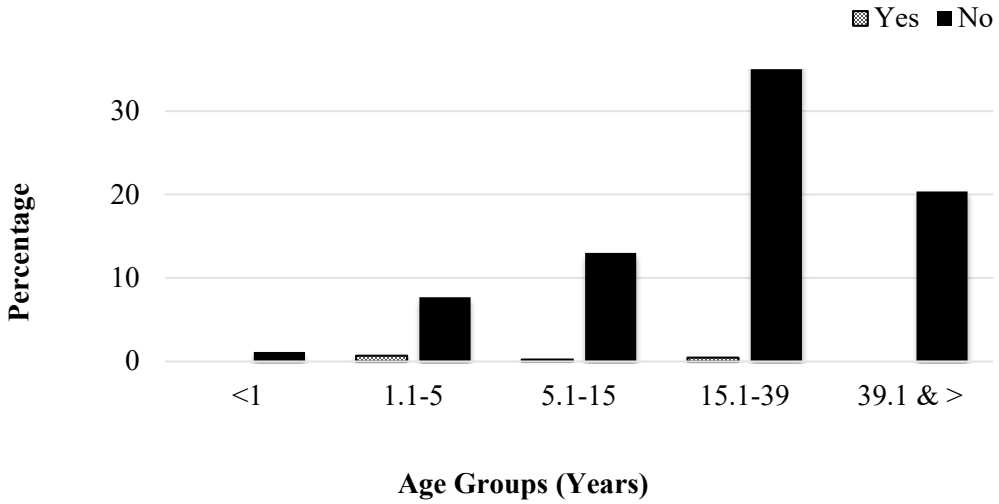


Figure 3.19B: Overall age groupwise percentage distribution of controls against family history (n=884).

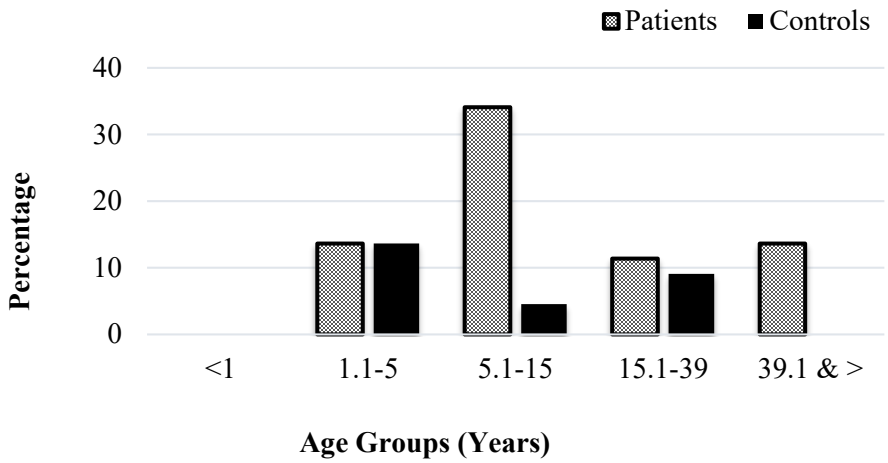


Figure 3.20: Comparison of family history of cancer between leukemia patients and controls

3.1.2. 1.5 Weight status

Body mass index (BMI) was derived from the height and weight taken during the sampling. According to international standards, individuals with a BMI <18.5kg/m2 were considered underweight, with a BMI of 18.5kg/m2-24.9kg/m2 as normal, while a BMI >24.9kg/m2 was considered overweight. According to our data, there is an increase in weight than normal as age increases. Of the individuals in the aged group, 52% of the individuals were overweight, 1.3% were underweight, and only 21% were bearing normal body weight. Here overweight is about 2.5 times higher than normal (Table 3.7A, Figure 3.21A).

About 63% of the female participants in the aged group were overweight, while among the men 47% were overweight, so here females outnumbered males in being overweight. While in AYAs, overweight comprises about 39%, which is almost twice to normal weight. About 57% of all females in AYAs were overweight, while only 30% of the males were overweight. Overall, female with leukemia, in both AYAs and Aged groups exhibits a higher tendency to overweight status (Table 3.7A).

Among controls, 23.76% had normal body weight, 52.83% were overweight, and 17.19% were underweight (Table 3.6). The weight status of control participants was also observed in age group and gender-wise distribution (Table 3.7A, Figure 3.21B).

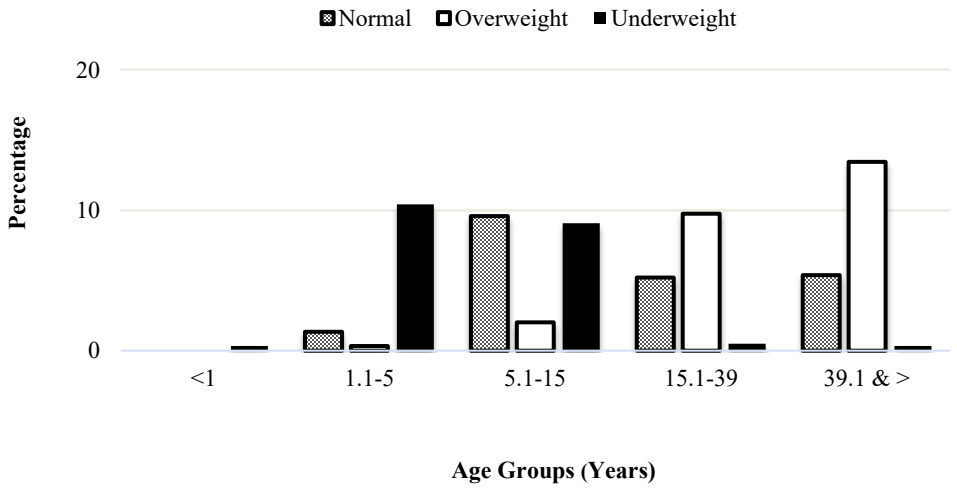


Figure 3.21A: Overall age groupwise percentage distribution of leukemia patients against body weight (n=594)

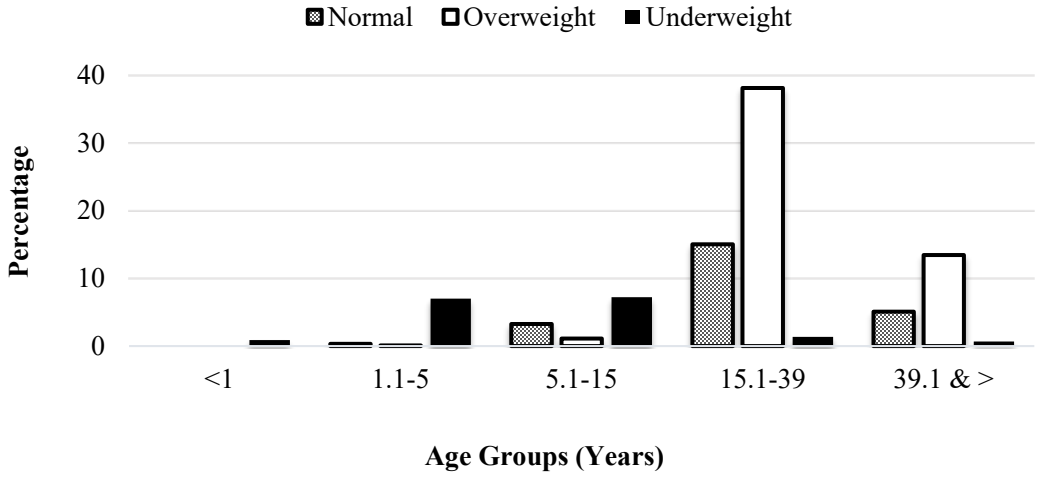


Figure 3.21B: Overall age groupwise percentage distribution of controls against body weight (n=884)

3.1.2. 1.6 Diet

Among all, 59% of the patients were using a balanced diet, however, 24% were unable to get proper nutrition (Table 3.6). Balanced nutrition is more visible across each group and gender in general (Table 3.7A, Figure 3.22A), However, in children, males had a higher peak of unbalanced diet (39%), although still lower to balanced nutrition (54%) as compared to other groups, where the difference of balanced to unbalanced is almost twice (Table 3.7A).

Among controls, 81.67% were eating a standard diet, and only 14.25% were on a poor diet (Table 3.6). The details of age group and gender-wise distribution of controls based on diet type are also provided (Table 3.7A, Figure 3.22B).

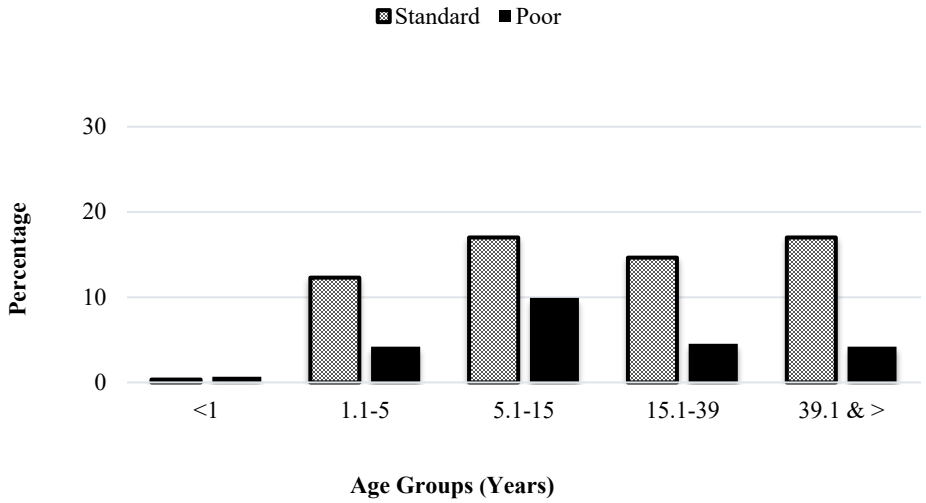


Figure 3.22A: Overall age groupwise percentage distribution of leukemia patients against diet (n=594)

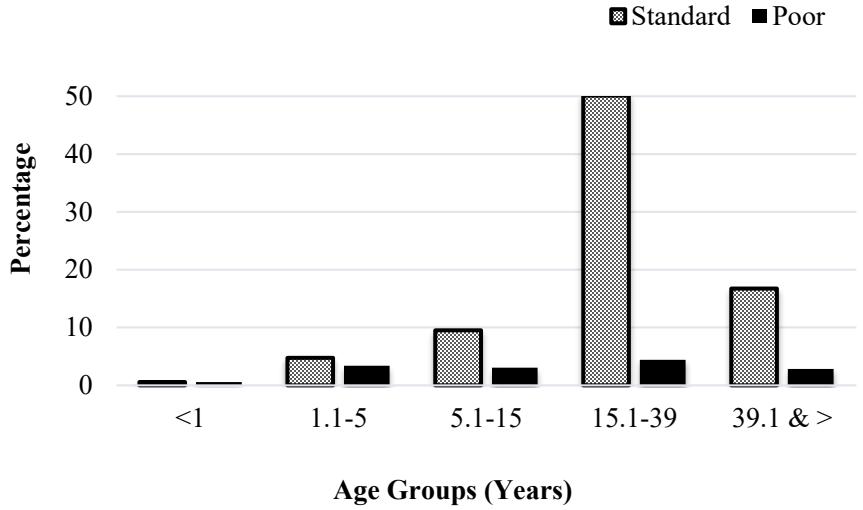


Figure 3.22B: Overall age groupwise percentage distribution of controls against diet (n=884)

3.1.2.1.7 Area

The samples were also grouped based on their residential area i.e. rural or urban. According to our data, across each group, a higher number of patients were living in rural surroundings, that is 57% of the preschoolers, 60% of children, 28% of the AYAs, and 27% of the aged. These percentages are more than thrice higher than the percentages of urban surroundings in each group (Table 3.7A, Figure 3.23A). In general, males have about a 3:1 rural-to-urban ratio when compared to females within each group (Table 3.7A). In all leukemia types, either lymphoid or myeloid and acute or chronic, the rural residential setup prevailed (Table 3.7B).

Among controls, 28.51% were living in a rural residential setup, while 19.91% were living in an urban setup (Table 3.6). The details of age group and gender-wise distribution of controls based on the residential area are provided (Table 3.7A, Figure 3.23B).

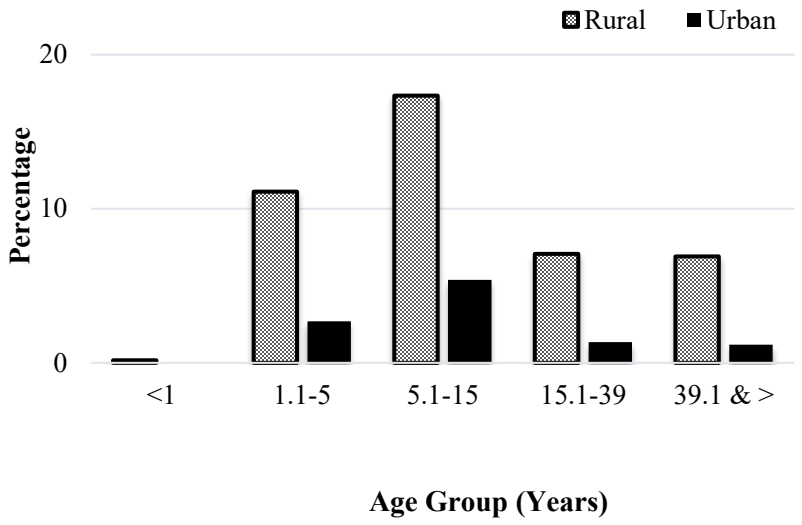


Figure 3.23A: Overall age groupwise percentage distribution of leukemia patients against area (n=594)

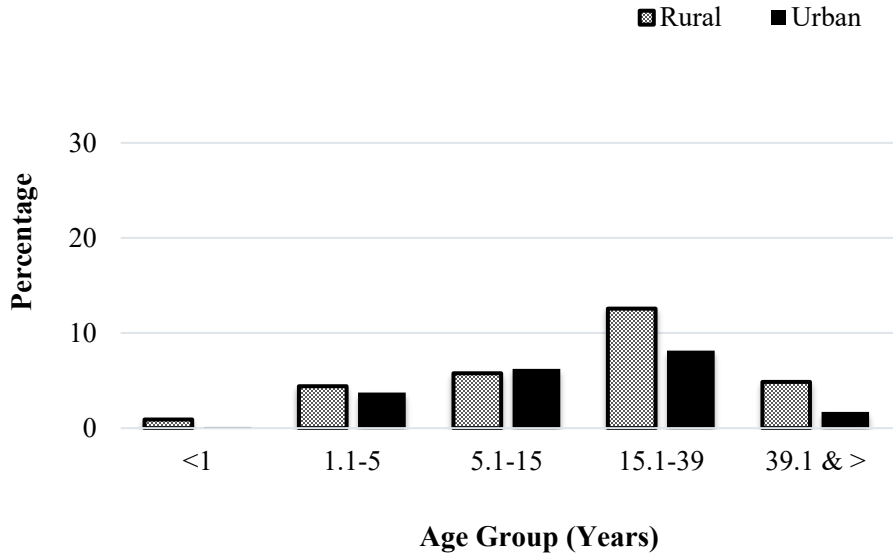


Figure 3.23B: Overall age groupwise percentage distribution of controls against area (n=884)

3.1.2.1.8 Water type

Overall, 39% of the patients were using groundwater, 11% were using surface water and 2% were using treated water (Table 3.6). In each age group, from preschooler to aged, groundwater consumption was more than twice higher than any other type in both genders (Table 3.7A, Figure 3.24A). In AYAs, the occurrence of groundwater consumption is the same (50%) for females with both lymphoid and myeloid leukemia. Overall, groundwater consumption surpasses water from other sources, in both acute and chronic leukemia in both genders across each age group (Table 3.7B).

Among controls, 27.26% were consuming groundwater, 15.72% of the individual were surface water, and 5.3% were consuming treated water (Table 3.6). The details of age group and gender-wise distribution of controls based on water type are provided (Table 3.7A, Figure 3.24B).

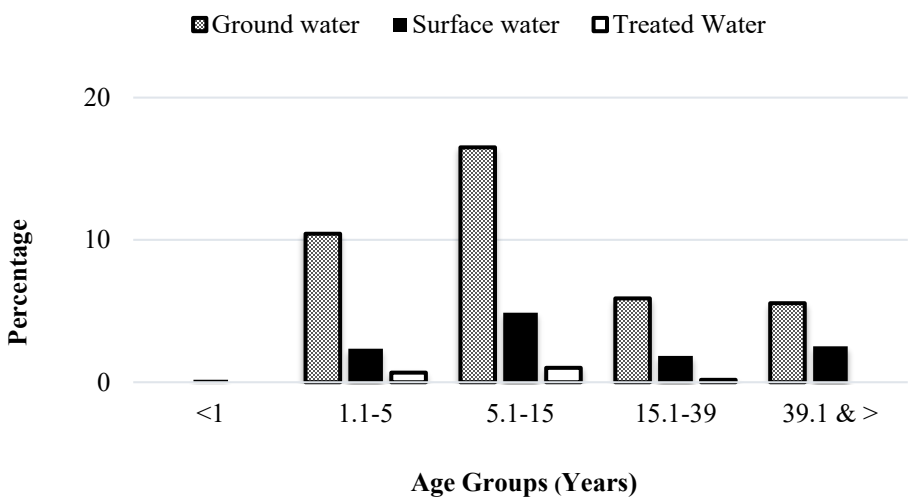


Figure 3.24A: Overall age groupwise percentage distribution of leukemia patients against water type (n=594)

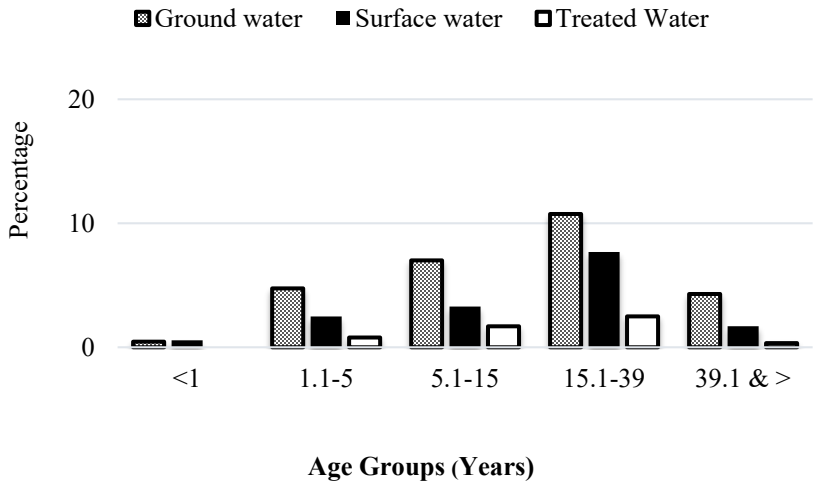


Figure 3.24B: Overall age groupwise percentage distribution of controls against water type (n=884)

3.1.2.1.9 Fuel

Among all, 46.6% were burning wood as fuel for cooking and 37.8% were burning gas as fuel (Table 3.6). The major difference in both lies in children, where 55% of the individuals were using wood and 38% were using gas as a fuel source. In the remaining groups, although wood is used more, the difference is not that considerable (Table 3.7A, Figure 3.25A). In general, the dominance of wood used is in the male gender in each group. Wood and gas are almost the same for females across each age group including children. However, in AYAs more females are using gas instead of wood (Table 3.7A).

In lymphoid leukemia, the wood use is recorded across all groups. In myeloid AYAs, more females are using gas (81%) than males (50%) within the group. In the same group, an equal percentage of gas use is observed for males with both lymphoid (50%) and myeloid origin (50%), however, for lymphoid leukemia males are higher in number than females in gas use. In myeloid aged's, gas use is higher than wood use in both males and females. In acute leukemia, more males were using wood, however, females have similar percentages for both wood and gas in children. In chronic leukemia not clearly, a visible difference is observed for both males and females in the use of either wood or gas as fuel (Table 3.7B).

Among controls, 23.3% were burning wood as a domestic fuel source, and 75.11% were using natural gas as fuel (Table 3.6). In age groupwise distribution (Table 3.7A). The details of age group and gender-wise distribution of controls based on domestic fuel type are also provided (Table 3.7A, Figure 3.25B).

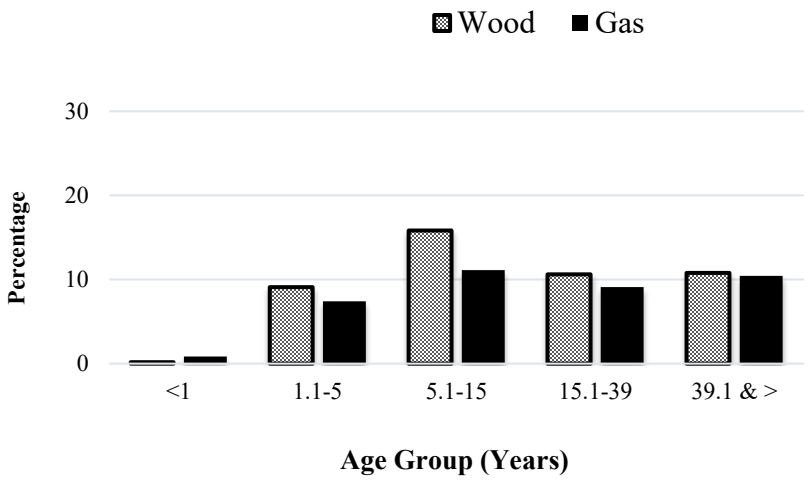


Figure 3.25A: Overall age groupwise percentage distribution of leukemia patients against fuel (n=594)

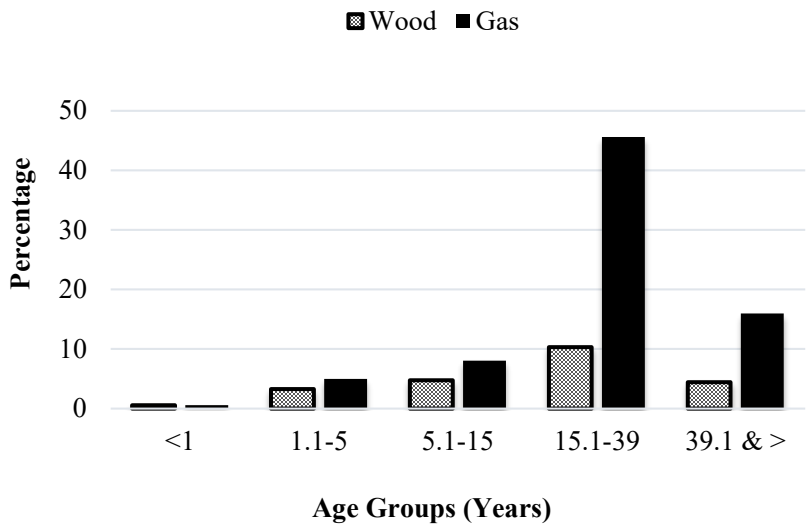


Figure 3.25B: Overall age group-wise percentage distribution of controls against fuel (n=884)

3.1.2.1.10 Family type

The family type was classified into two categories, i.e. Nuclear family was comprised of a husband, wife, and their children, while in the extended family, other relatives were also living along with them. According to our data, Overall, the highest frequency of the patients was living in an extended family across age groups in both genders. Overall, the percentage of the extended family was highest in children of age 1-15 years old, which is 47% of all within the group, while in AYAs (19%) and aged (21.7%) the percentages of extended families were almost double than nuclear family set up (11% and 9.8%, respectively) (Table 3.7A, Figure 3.26A).

In lymphoid leukemia, patients living in an extended family setup have the highest percentages in all groups. The highest peak was observed in both male and female children (43%). Both acute and chronic leukemia, although both males and females living in extended families were higher in number, however in females, within the group the tendency towards extended family is much higher (Table 3.7B).

Among controls, 18.33% were living in a nuclear family setup, and 30.66 were living in extended families (Table 3.6). The details of age group and gender-wise distribution of controls based on family type are provided (Table 3.7A, Figure 3.26B).

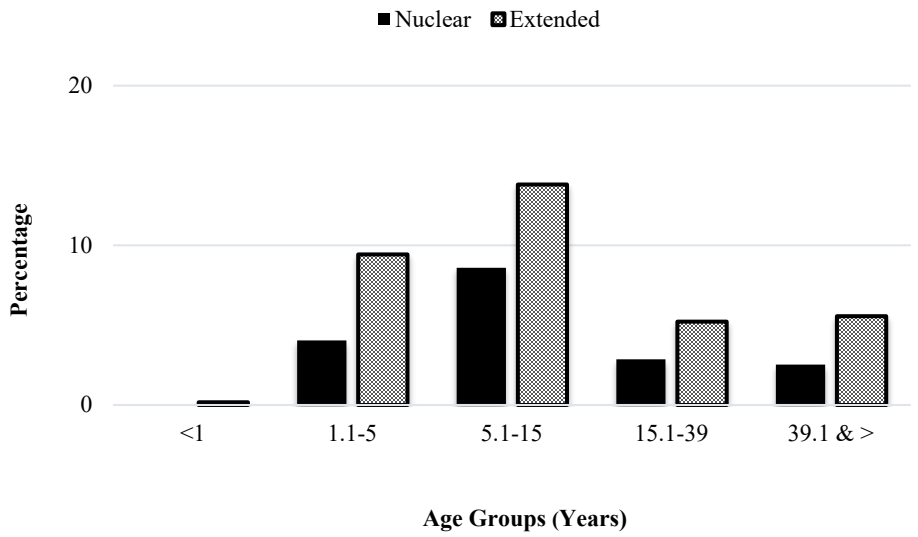


Figure 3.26A: Overall age groupwise percentage distribution of leukemia patients against family type (n=594)

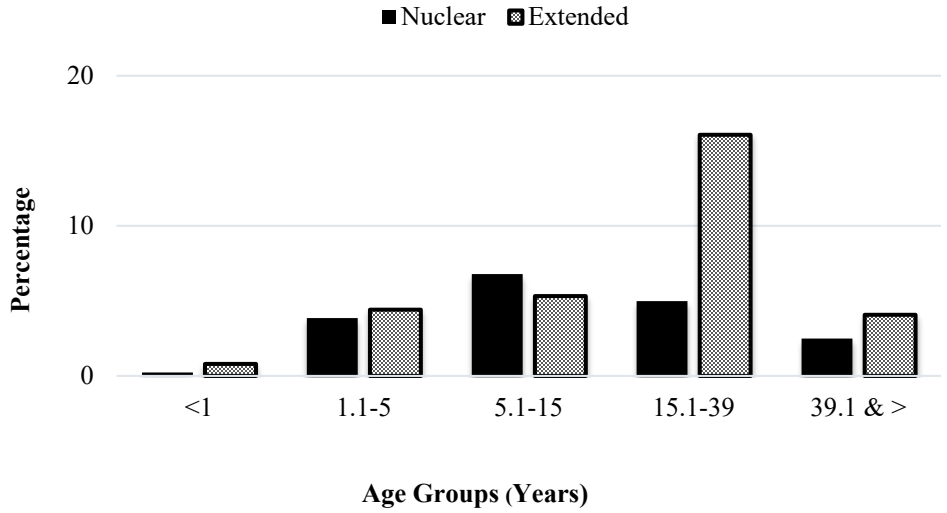


Figure 3.26B: Overall age groupwise percentage distribution of controls against family type (n=884)

3.1.2.1.11 Carbonated drinks

Among all, 55% of the cases were consuming carbonated drinks and 30% were not consuming any such drink (Table 3.6). There is an increasing trend in drink consumption in both genders, from preschoolers (57%) to children (69%), after which there is a decrease in consumption in AYAs (52%) and aged (41%) (Table 3.7A). There is no difference in carbonated drinks consumption and no consumption for the aged. Males are twice more consuming carbonated drinks as compared to females in all the leukemia subtypes, within the groups (Table 3.7A). Both in disease nature and origin, the percentage of carbonated drink consumption is greater than no consumption (Table 3.7B).

Among controls, 58.71% were drinking carbonated drinks, and 39.59% were not consuming any kind of carbonated drinks (Table 3.6). The details of age group and gender-wise distribution of controls based on consumption of carbonated drinks are provided (Table 3.7A).

3.1.2.1.12 Tobacco use

Overall, 30% of the patients were exposed to different forms of tobacco, either directly or indirectly (Table 3.6). In this 30%, preschoolers (5.7%) and children (10.4%) were most highly exposed. About 8.6% of aged individuals were exposed to tobacco. However, in AYAs the exposure is not heavy. Within each group, the exposure is much lesser than non-exposure for both genders. Although in aged males both exposure and non-exposure are equal (40.5%) (Table 3.7A, Figure 3.27A).

Among controls, 26.36% were using tobacco in different forms, while 72.29% were not exposed to tobacco use (Table 3.6). The details of age group and gender-wise distribution of controls based on tobacco use are provided (Table 3.7A, Figure 3.27B).

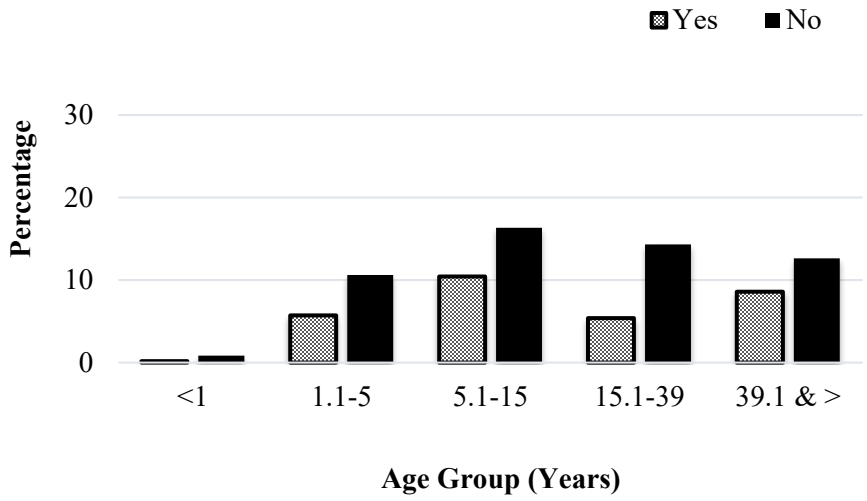


Figure 3.27A: Overall age groupwise percentage distribution of leukemia patients against tobacco use (n=594)

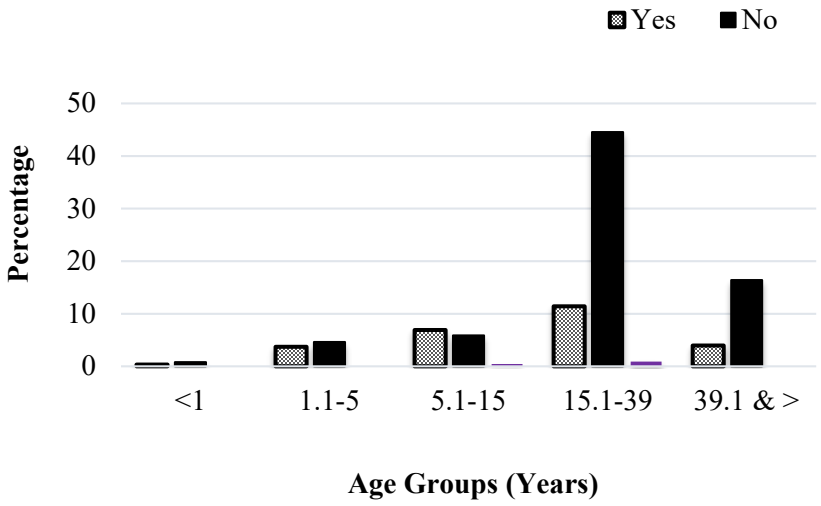


Figure 3.27B: Overall age groupwise percentage distribution of controls against tobacco use (n=884)

3.1.2.1.1.13 Microwave use

Among all, 8.6% of the patients were using a microwave oven, and 65% were not using a microwave at all (Table 3.6). The highest number of patients using microwaves was in the age group 5.1-15, which is only 2.86%. In all age groups, most of the patients were not using a microwave oven (Table 3.7A, Figure 3.28A).

Among controls, 33.26% were using microwaves and 64% were not using microwaves (Table 3.6). In age groupwise distribution, in all of the control groups, a higher number of controls were using microwaves (although lower than no use in controls) as compared to patients, particularly, in the age group comprised of 15.1-39 years, 21.04% of the controls were using microwave (Table 3.7A, Figure 3.28B).

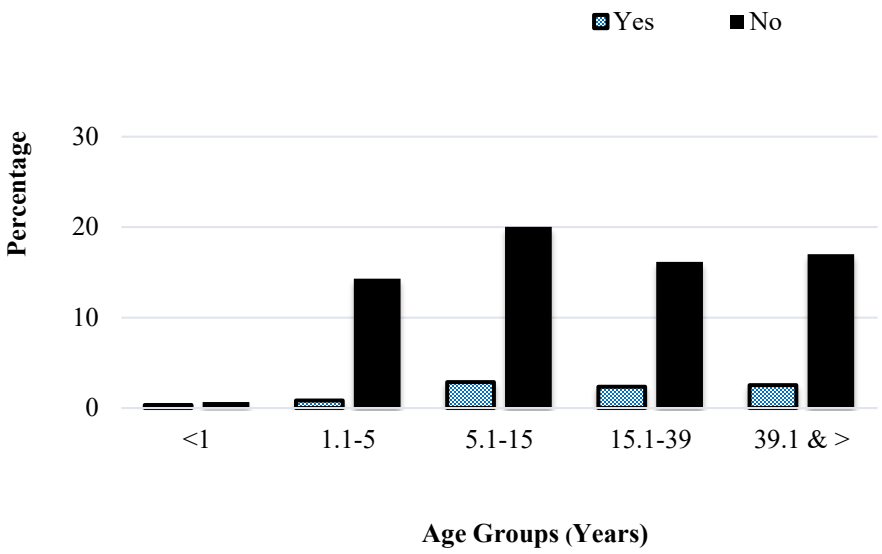


Figure 3.28A: Overall age groupwise percentage distribution of leukemia patients against microwave use (n=594)

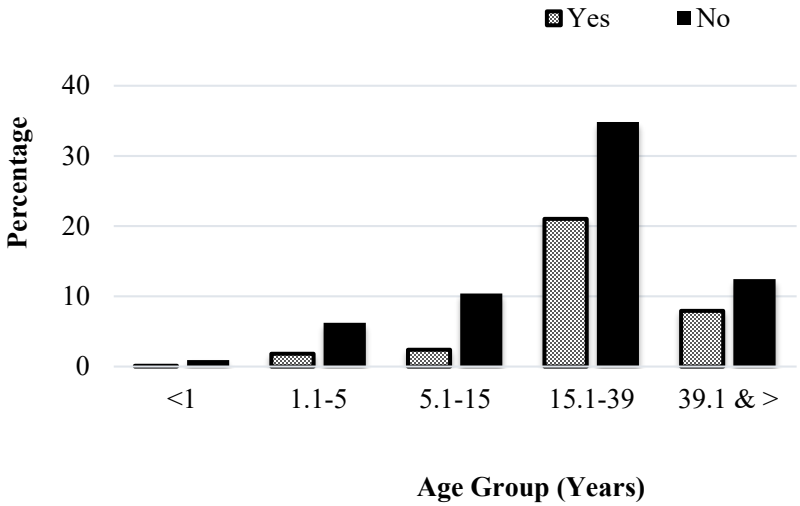


Figure 3.28B: Overall age groupwise percentage distribution of controls against microwave use (n=884)

3.1.2.1.14 Perfume use

Among all, 42.86% of the patients were wearing perfumes in different forms, and 40.42% were not wearing any perfume (Table 3.6). In age group-wise distribution, the highest number of perfume use, 15.32% has been observed in the age group 5.1-15, followed by 10.44% use in the age group 1.1-5 years, however, a decrease in the number has been observed to use perfume in comparison to no use (Table 3.7A, Figure 3.29A).

Among all, 70.14% of the controls were wearing perfumes in different forms, and 28.05 % were not using perfume (Table 3.6). In age group-wise distribution, a higher number of controls were observed to use perfume as compared to controls with no perfume use (Table 3.7A, Figure 3.29B).

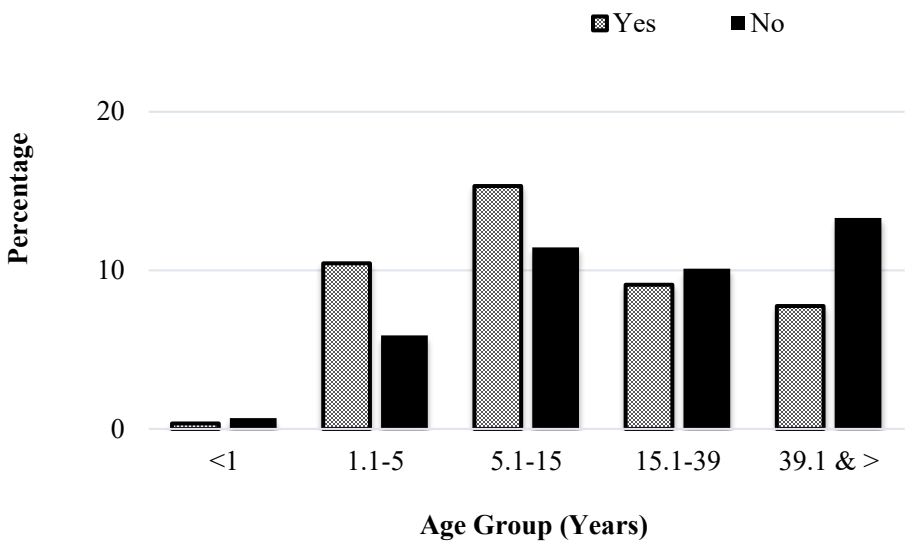


Figure 3.29A: Overall age groupwise percentage distribution of leukemia patients against perfume use (n=594)

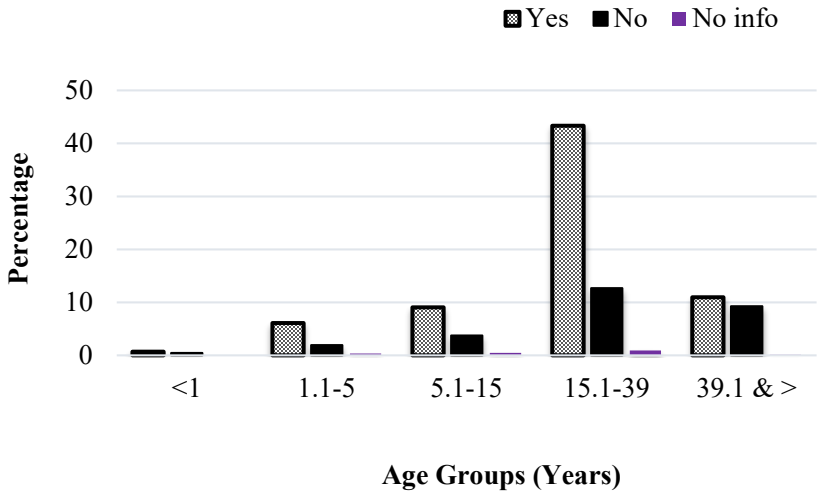


Figure 3.29B: Overall age groupwise percentage distribution of controls against perfume use (n=884)

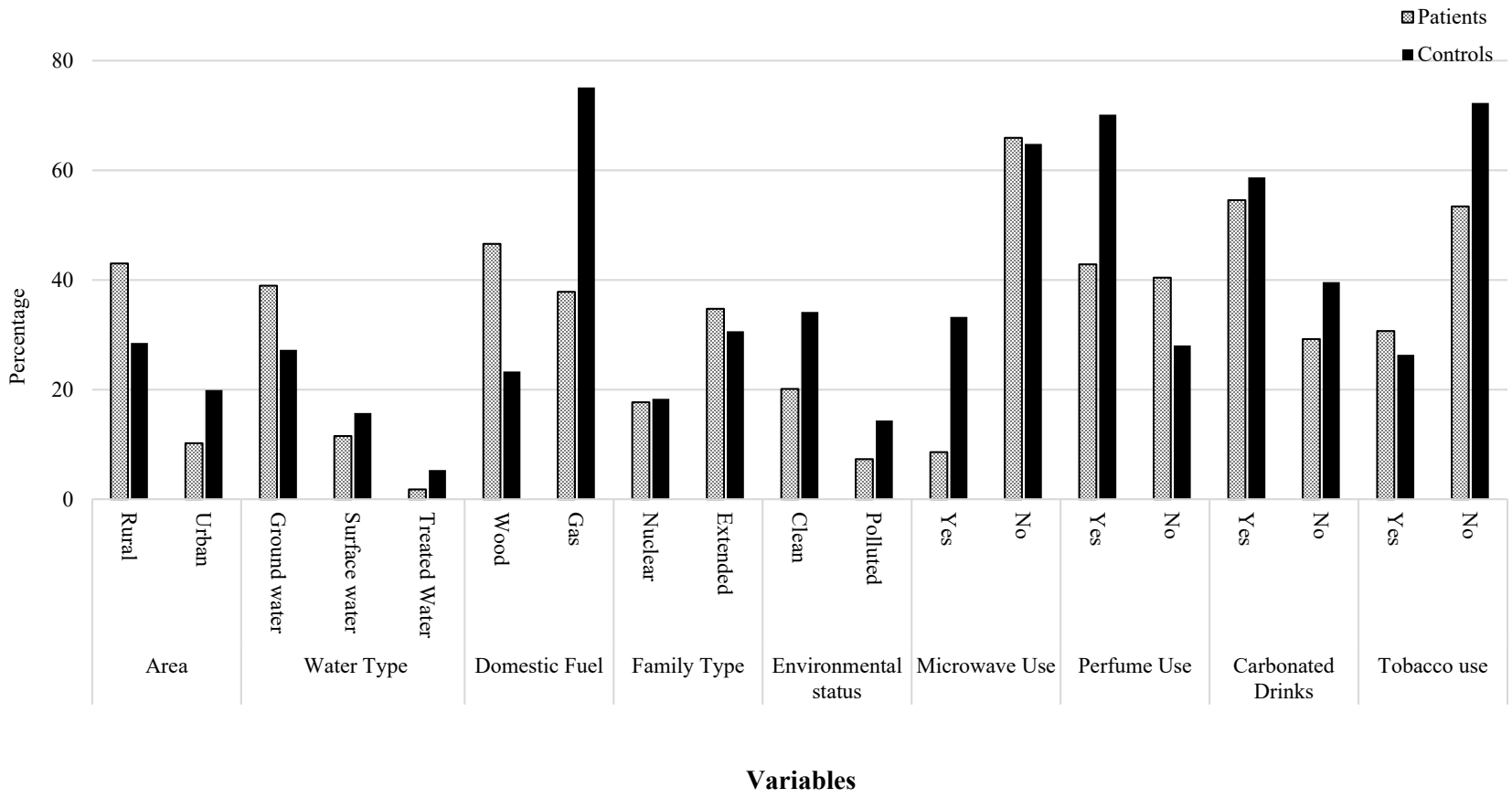


Figure 3.30: Comparison of environmental and lifestyle style-related risk factors between patients and controls

3.1.2.2 Association of demographical and lifestyle factors with leukemia risk

The association of demographic and lifestyle variables to the risk of leukemia was investigated in comparison to controls for each variable. To identify the risk factors, odds ratios for the patients were calculated in comparison to controls, while considering the group with the highest number of participants as a reference group.

The high-risk age groups based on demographic and lifestyle variables, from within the patients were also identified by comparing the frequencies of the variables between the patient age groups and calculating odds ratios.

Table 3.8: Odds ratios for demographic and lifestyle related variables for patients

Categories	Variable	Patients	Control	Odds Ratio	95 % CI	P-Value
		N (%)	N (%)			
Age Group	Children	294 (48)	201 (23)	3.10	2.48 - 3.87	P < 0.0001
	Adult	322 (52)	683 (77)	1.00		
Gender	Male	411 (66.72)	623 (70.47)	0.84	0.67 - 1.05	P = 0.1223
	Female	205 (33.27)	261 (29.52)	1.00		
Blood Group	B	83 (13.47)	63 (7.12)	1.11	0.74-1.68	P = 0.6011
	All others	143 (23.21)	121 (13.69)	1.00		
Rh-Factor	+ve	206 (33.44)	166 (18.77)	1.12	0.57-2.18	P = 0.7460
	-ve	20 (3.24)	18 (2.03)	1.00		
Ethnicity	Pathan	354 (57.46)	289 (32.69)	2.85	2.3-3.55	P < 0.0001
	All Others	236 (38.31)	550 (62.22)	1.00		
Education	No Education	208 (33.77)	156 (17.65)	3.36	2.62-4.32	P < 0.0001
	Educated	275 (44.64)	694 (78.5)	1.00		
Diet	Poor Diet	150 (24.35)	126 (14.25)	2.34	1.79-3.06	P < 0.0001
	Standard Diet	367 (59.57)	722 (81.67)	1.00		
Weight Status	Underweight	123 (19.96)	152 (17.19)	1.96	1.49-2.58	P < 0.0001
	Normal, Overweight	280 (45.45)	677 (76.58)	1.00		
Parental Consanguinity	Yes	202 (32.79)	193 (21.83)	2.14	1.68-2.72	P < 0.0001
	No	299 (48.53)	610 (69)	1.00		
Family History	Yes	34 (5.51)	12 (1.35)	4.25	2.18-8.27	P < 0.0001
	No	582 (94.48)	872 (98.64)	1.00		
Area	Rural	265 (43.01)	252 (28.5)	2.94	2.1-4.11	P < 0.0001
	Urban	63 (10.22)	176 (19.9)	1.00		
Environment	Polluted	124 (20.12)	302 (34.16)	0.86	0.58-1.29	P = 0.4692
	Clean	45 (7.3)	127 (14.36)	1.00		
Water Type	Groundwater	240 (38.96)	241 (27.26)	2.26	1.65-3.1	P < 0.0001
	All Others	82 (13.31)	186 (21.04)	1.00		

Fuel	Wood	287 (46.59)	206 (23.3)	3.97	3.15-5.01	P < 0.0001
	Gas	233 (37.82)	664 (75.11)	1.00		
Family Set up	Extended	214 (34.74)	271 (30.65)	1.17	0.87-1.59	P = 0.2984
	Nuclear	109 (17.69)	162 (18.32)	1.00		
Tobacco use	Yes	189 (30.68)	233 (26.35)	1.58	1.25-1.99	P < 0.0001
	No	329 (53.4)	639 (72.28)	1.00		
Carbonated drinks	Yes	336 (54.54)	519 (58.71)	1.26	1-1.58	P = 0.0461
	No	180 (29.22)	350 (39.59)	1.00		
Microwave Use	Yes	53 (8.6)	294 (33.25)	0.25	0.18-0.35	P < 0.0001
	No	406 (65.9)	573 (64.81)	1.00		
Perfume Use	Yes	264 (42.85)	620 (70.13)	0.42	0.34-0.53	P < 0.0001
	No	249 (40.42)	248 (28.05)	1.00		

3.1.2.2.1 Age, Gender, ethnicity, education, and leukemia risk

A thorough comparison of patients, based on gender, ethnicity, and education was done with controls (Figure 3.31). According to the present study, leukemia is more recurrent in children than adults (OR=3.10; 95% CI=2.48 - 3.87). It was also found that the male gender has an about 2-fold higher risk for lymphoid leukemia as compared to myeloid leukemia (OR=1.97; 95%CI=1.38-2.80). AYAs have a 2.31-fold higher risk for myeloid leukemia compared to all other age groups of patients (OR=2.31; 95%CI 1.58-3.38). The age group >39.1 has a 21.46-fold high risk for chronic leukemia as compared to all other age groups of patients (OR= 21.46; 95% CI=13.49-34.14; Table 3.10, Figure. 3.36). The risk for Pathan ethnicity has a 2.8-fold higher risk as compared to controls (OR=2.85; 95% CI=2.29-3.54). The risk of no education in leukemia patients has higher as compared to controls (OR=3.36; 95% CI=2.62-4.32; Table 3.8, Figure. 3.34).

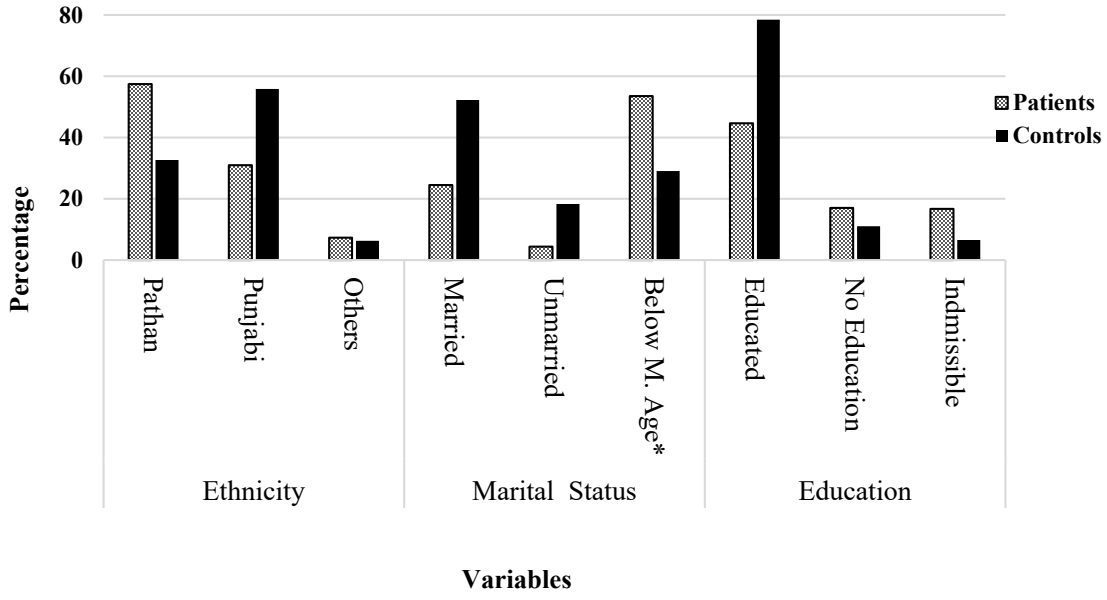


Figure 3.31: Comparison of ethnicity, marital status, and education between patients and controls

3.1.2.2.2 Consanguinity and family history as a risk factor for leukemia

The parental consanguinity of the patients was compared with controls (Figure 3.32). In the present study, patients with the parental consanguineous union had a 2-fold higher, significant risk for leukemia as compared to controls (OR=2.13; 95% CI=1.67-2.71; Table 3.8, Figure 3.34). Particularly, patients in the age group 5.1-15 with parental consanguinity have been observed to have an increased risk for leukemia in comparison to controls (OR= 1.6, 95% CI=1.03-2.39: Table 3.9, Figure 3.35). In this study, the patients with a family history of leukemia/solid cancers, have a 4-fold increased risk for leukemia as compared to controls (OR= 4.24, 95% CI=2.18-8.26: Table 3.8, Figure 3.34).

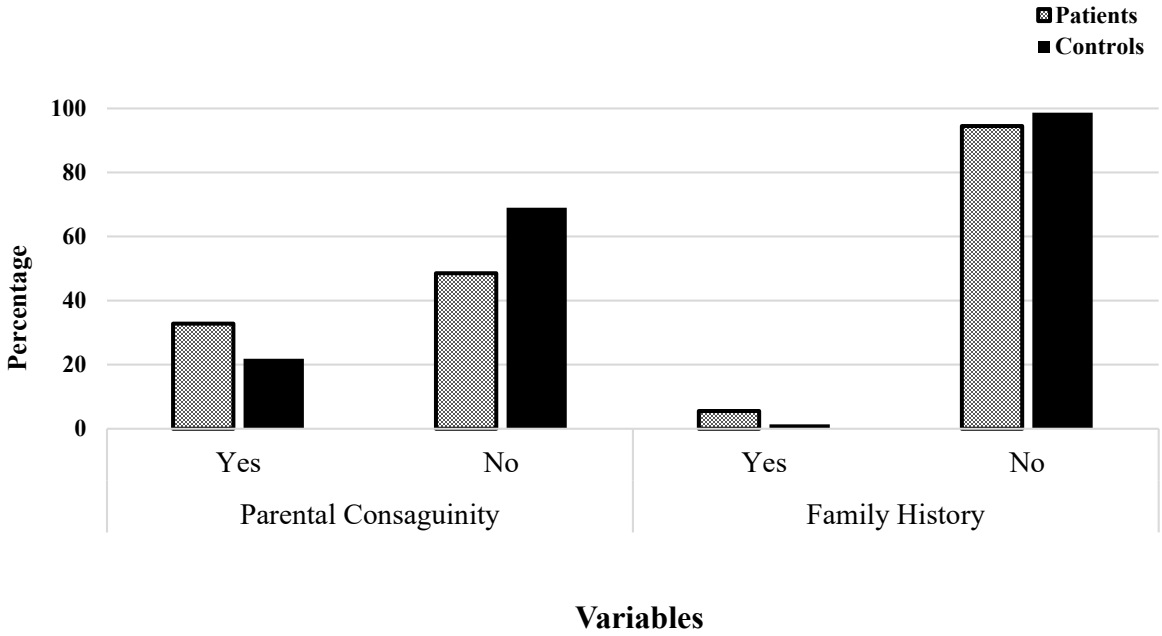


Figure.3.32: Comparison of parental consanguinity and family history of cancer between patients and controls

3.1.2.2.3 Diet, weight, and leukemia risk

The diet and weight of patients were compared with controls (Figure 3.33). Overall, patients with a poor diet have a 2.34-fold higher risk for leukemia as compared to controls (OR=2.34; 95% CI=1.79-3.06; Table 3.8, Figure 3.34), while patients in the age group 5.1-15 with a poor diet face a higher risk for leukemia in comparison to matched controls (OR=2.67; 95% CI=1.55-4.59; Table 3.9, Figure 3.35).

Lower BMI significantly increased the risk for leukemia in this study as compared to controls (OR=1.95; 95% CI=1.50-2.60; Table 3.8, Figure 3.34). Pre-schoolers (1.1-5 years) with lower BMI have a significant, 7.9-fold higher risk for leukemia, as compared to patients in the age group 5.1-15 years (OR=7.92; 95% CI=3.71-16.88) and 28.37-fold higher risk in comparison to the additive risk of all other patients age groups (OR=28.37; 95% CI=3.74-58.57) of the patients (Table 3.10, Figure 3.36).

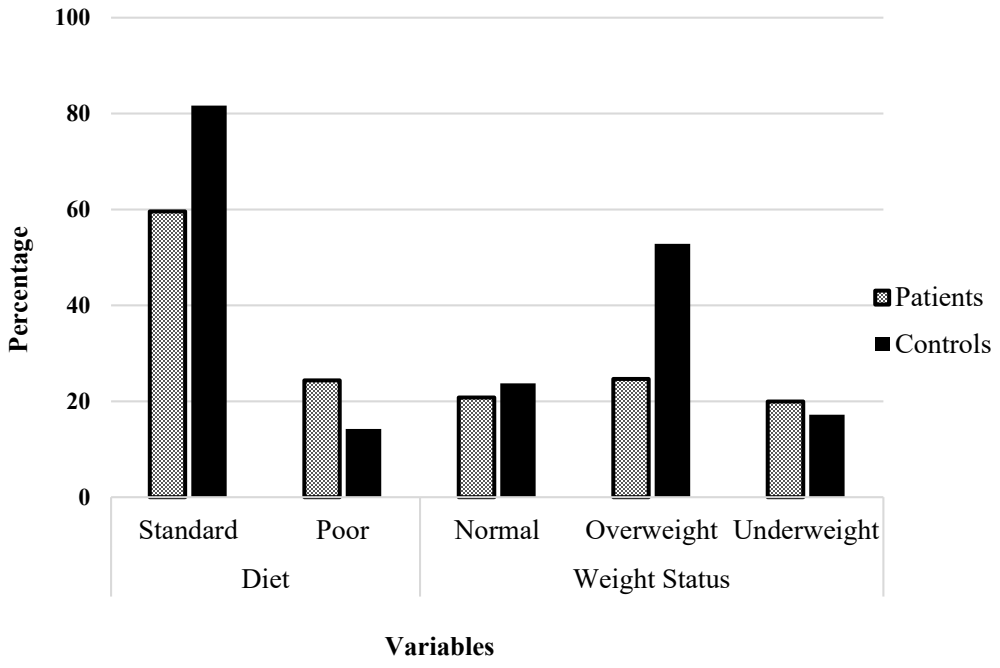


Figure 3.33: Comparison of diet and body weight of cancer between leukemia patients and controls

3.1.2.2.4 Lifestyle and residential risk factors associated with leukemia

In this study, the exposure to lifestyle-related risk factors was thoroughly compared between patients and controls (Figure 3.30).

It has been observed that patients living in the rural residential setup have a 2.9-fold increase, a significant risk for leukemia as compared to controls (OR=2.93, 95% CI=2.10-4.10; Table 3.8, Figure 3.34). The leukemia patients in the age group 5.1-15, residing in a rural area, have 1.7-fold higher risk (OR=2.70; 95% CI=1.82-4.02), as compared to age-matched controls. A similar relationship was shown by the preschoolers (1.1-5 years) (OR=1.92; 95% CI=1.23-2.99). However, no such association was identified in adult patients (Table 3.9, Figure 3.35).

Consuming groundwater as a primary drinking water source has been associated with an increased risk for leukemia in this study (OR=2.25; 95% CI=1.6479-3.0964; Table 3.8, Figure 3.34). Particularly, the risk is significant for children (OR= 2.17; 95% CI=1.47-3.21) and preschoolers (OR= 1.76, 95% CI 1.13-2.75), respectively, in comparison to controls. No such association was seen in adults (OR=0.90; 95% CI=0.54-1.49; Table 3.9, Figure 3.35).

Wood use as a domestic source of fuel has a significant association with higher leukemia risk as compared to controls (OR= 3.97; 95% CI=3.14-5.01; Table 3.8, Figure 3.34). Particularly, children in families doing wood burning, have an elevated risk for leukemia in comparison to controls (OR= 2.01; 95% CI=1.32-3.07). No such significant association was seen in any other age group (Table 3.9, Figure 3.35).

The extended family setup has no overall significant risk for leukemia as compared to controls in this study; however, the preschoolers and children living in an extended family setup, have 2 (OR=2.26; 95% CI=1.43-3.58), and 3-fold increased risk (OR=3.12; 95% CI=2.05-4.77), to get the disease, respectively, when compared to the same age groups in controls. However, in adult patients, no significant increased risk was observed in comparison to adult controls (Table 3.9, Figure 3.35).

The consumption of carbonated drinks has been linked to an increase in leukemia risk by 25% in the present study (OR=1.25; 95% CI=1.00-1.57; Table 3.8, Figure 3.34).

Tobacco use (active/passive) is observed to significantly increase leukemia risk by 57% (OR=1.57; 95% CI=1.24-1.98; Table 3.8, Figure 3.34). In age group comparison, the risk was significant for the patients belonging to age group >39.1 those who were using tobacco in comparison to age-matched controls (OR=2.23; 95% CI=1.37-3.62; Table 3.9, Figure 3.35).

3.1.2.2.5 Factors reducing leukemia risk

In the present study, perfume use (OR= 0.42; 95% CI=0.33-0.53) and the use of microwave oven (OR= 0.25; 95% CI=0.18-0.35) were associated with a lower risk of leukemia and identified as a protective factor for the studied population (Table 3.8, Figure 3.34).

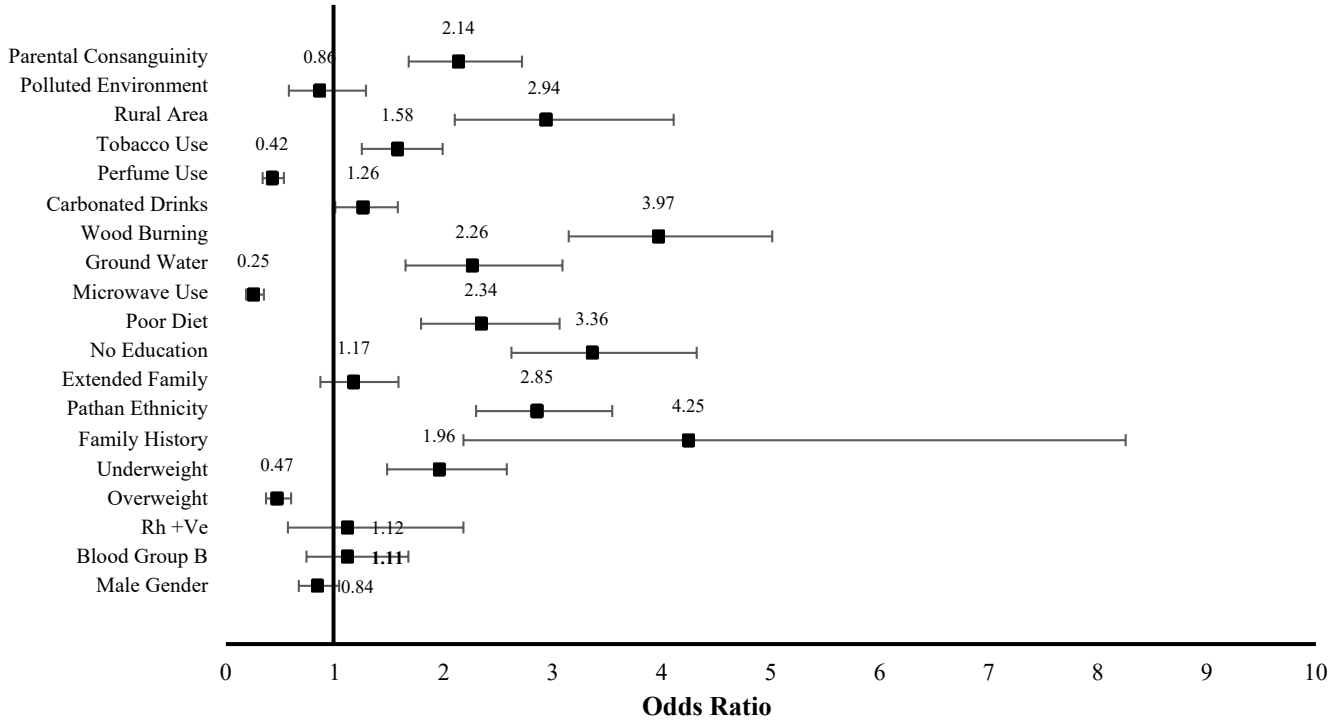


Figure 3.34: Odds ratios depicting risk factors

Odds ratios were calculated by comparing the exposure of risk factors to patients and controls (CI 95%, p*= 0.05). It is indicated that lower BMI (ref. normal and higher BMI), poor diet (ref. standard BMI), all forms of tobacco use (ref. no use), consanguineous parental union (non-consanguineous parental union), family history of leukemia and solid cancers (ref. no family history), lack of basic education (ref. educated) and hailing to Pathan ethnicity (ref. other ethnicities additively) were high risks variables significantly associated to leukemia. In lifestyle-related factors, wood burning, consumption of unprocessed groundwater for drinking purposes, and living in a residential rural setup were found to increase leukemia risk significantly, and use of carbonated drinks was only observed to cause a slight but significant risk for leukemia; However, wearing perfume and microwave use has been observed to have a protective effect for leukemia risk. The reference used for the above factors were gas burning, water from all other sources, urban residential setup, no consumption of carbonated drinks, no use of perfume, and no use of microwave, respectively. No significant risk-based association was noticed for extended family type, blood group B, Rh +ve, and gender in the studied population for leukemia patients in comparison to controls taking nuclear family, all other blood groups than B, Rh -ve, and female gender as reference.

Table 3.9: Odds ratios for high-risk patient age groups vs control groups (additive model)

Variables	Risk	Risk Group (Age Category)	Patients	Controls	Odd Ratios	95 % CI	P -Value
			N (%)	N (%)			
Diet	Poor	1.1-5	25 (4.2)	30 (3.39)	0.70	0.38-1.26	P = 0.2326
		All Other Groups	115 (19.36)	96 (10.86)	1.00		
	Poor	5.1-15	59 (9.93)	27 (3.05)	2.67	1.55-4.59	P = 0.0004
		All Other Groups	81 (13.64)	99 (11.2)	1.00		
Parental Consanguinity	Consanguineous union	1.1-5	48 (12.18)	45 (11.42)	1.03	0.65-1.64	P = 0.8951
		All Other Groups	153 (38.83)	148 (37.56)	1.00		
	Consanguineous union	5.1-15	80 (20.3)	57 (14.46)	1.58	1.04-2.4	P = 0.0329
		All Other Groups	121 (30.71)	136 (34.52)	1.00		
	Consanguineous union	39.1 & >	35 (8.88)	21 (5.32)	1.73	0.97-3.09	P = 0.0656
		All Other Groups	166 (42.13)	172 (43.65)	1.00		
Family History	Family History, +	5.1-15	15 (34.09)	2 (4.54)	4.41	0.83-23.42	P = 0.0814
		All Other Groups	17 (38.64)	10 (22.73)	1.00		
	Family History, +	5.1-15	15 (34.09)	2 (4.54)	0.48	0.02-11.37	P = 0.6473
		39.1 & >	6 (13.63)	0 (0)	1.00		
	Family History, +	39.1 & >	6 (13.63)	0 (0)	6.13	0.32-117.64	P = 0.2289
		All Other Groups	26 (59.09)	12 (27.27)	1.00		
Family type	Extended	5.1-15	82 (13.8)	47 (5.31)	3.13	2.05-4.77	P < 0.0001
		All Other Groups	121 (20.37)	224 (25.34)	1.00		
	Extended	1.1-5	56 (9.42)	39 (4.41)	2.27	1.43-3.58	P = 0.0005
		All Other Groups	147 (24.75)	232 (26.24)	1.00		
	Extended	15.1-39	31 (5.21)	142 (16.06)	0.16	0.1-0.26	P < 0.0001

		All Other Groups	172 (28.96)	129 (14.59)	1.00		
	Extended	39.1 & >	33 (5.55)	36 (4.07)	1.27	0.76-2.11	P = 0.3647
		All Other Groups	170 (28.62)	235 (26.58)	1.00		
Area	Rural	1.1-5	66 (11.11)	39 (4.41)	1.93	1.24-3	P = 0.0036
		All Other Groups	187 (31.48)	213 (24.1)	1.00		
	Rural	5.1-15	103 (17.34)	51 (5.76)	2.71	1.82-4.02	P < 0.0001
		All Other Groups	150 (25.25)	201 (22.74)	1.00		
	Rural	15.1-39	42 (7.07)	111 (12.55)	0.25	0.17-0.38	P < 0.0001
		All Other Groups	211 (35.52)	141 (15.95)	1.00		
	Rural	39.1 & >	41 (6.9)	43 (4.86)	0.94	0.59-1.5	P = 0.7957
		All Other Groups	212 (35.69)	209 (23.64)	1.00		
Fuel	Wood	1.1-5	54 (9.09)	29 (3.28)	1.48	0.91-2.43	P = 0.1159
		All Other Groups	222 (37.37)	177 (20.02)	1.00		
	Wood	5.1-15	94 (15.82)	42 (4.75)	2.02	1.32-3.07	P = 0.0011
		All Other Groups	182 (30.64)	164 (18.55)	1.00		
	Wood	5.1-15	94 (15.82)	42 (4.75)	1.20	0.67-2.15	P = 0.5339
		1.1-5	54 (9.09)	29 (3.28)	1.00		
Water Type	Ground	1.1-5	62 (10.43)	42 (4.75)	1.77	1.14-2.75	P = 0.0115
		All Other Groups	166 (27.95)	199 (22.51)	1.00		
	Ground	5.1-15	98 (16.49)	62 (7.01)	2.18	1.47-3.21	P = 0.0001
		All Other Groups	130 (21.89)	179 (20.25)	1.00		
	Ground	15.1-39	35 (5.89)	95 (10.74)	0.28	0.18-0.43	P < 0.0001
		All Other Groups	193 (32.49)	146 (16.52)	1.00		
	Ground	39.1 & >	33 (5.55)	38 (4.29)	0.90	0.55-1.5	P = 0.6961
		All Other Groups	195 (32.83)	203 (22.96)	1.00		
Tobacco use	Tobacco use	5.1-15	62 (10.43)	61 (6.9)	1.48	0.97-2.26	P = 0.0693

	All Other Groups	118 (19.87)	172 (19.46)	1.00		
Tobacco use	1.1-5	34 (5.72)	33 (3.73)	1.41	0.84-2.38	P = 0.1977
	All Other Groups	146 (24.58)	200 (22.62)	1.00		
Tobacco use	39.1 & >	51 (8.58)	35 (3.95)	2.24	1.38-3.63	P = 0.0011
	All Other Groups	129 (21.72)	198 (22.4)	1.00		
Tobacco use	15.1-39	32 (5.38)	101 (11.42)	0.28	0.18-0.45	P < 0.0001
	All Other Groups	148 (24.92)	132 (14.93)	1.00		
Tobacco use	5.1-15	62 (10.43)	61 (6.9)	0.70	0.4-1.22	P = 0.2048
	39.1 & >	51 (8.58)	35 (3.95)	1.00		
Tobacco use	39.1 & >	51 (8.58)	35 (3.95)	1.43	0.82-2.5	P = 0.2048
	5.1-15	62 (10.43)	61 (6.9)	1.00		

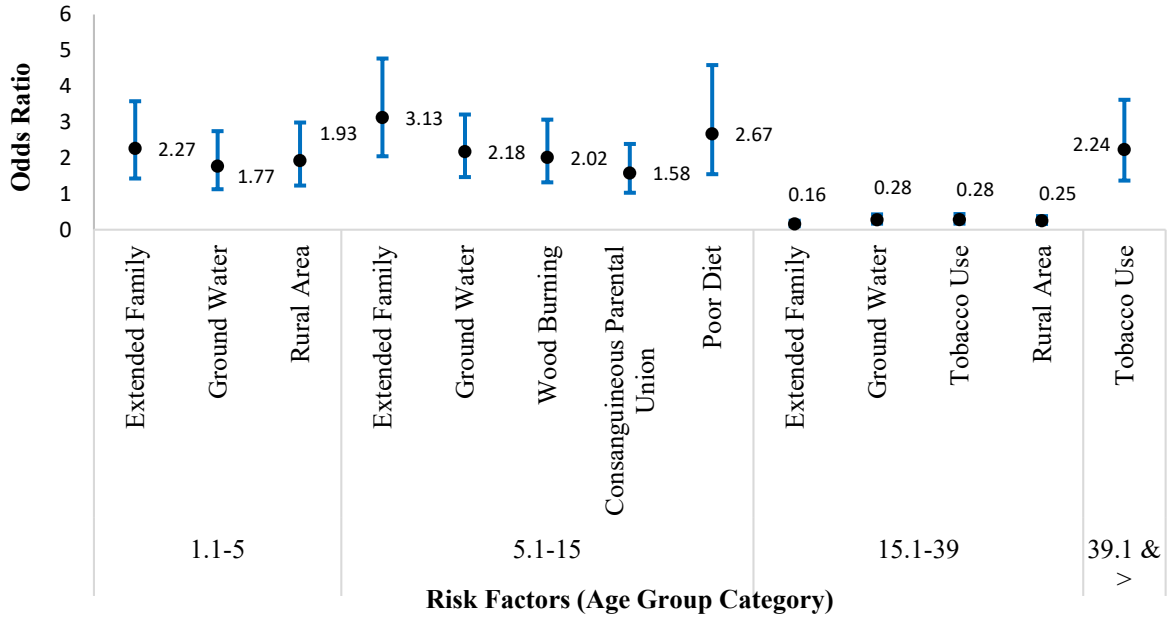


Figure 3.35: Odds ratios were calculated to identify age group-related risk factors for patients compared to controls

Odds ratios calculated to identify age group-related risk factors for patients compared to controls (CI 95%, p* = 0.05), indicated that although, the extended family type was overall nonsignificant, is a significant risk factor for children below the age of 5 years with leukemia (ref. extended family type in remaining age groups), along with groundwater consumption (ref. groundwater consumption in all remaining age groups) and rural residential set up (ref. rural residential set up in remaining age groups). For patients of age 5.1-15 years old, extended family, groundwater, wood burning (ref. wood burning in remaining age groups) as domestic fuel, consanguineous parental unions (ref. consanguineous parental unions in all remaining age groups), and poor diet (ref. poor diet in remaining age groups) were associated with a high risk of leukemia. For patients older than 39 years of age, tobacco use (ref. tobacco use in remaining age groups) was significantly associated compared to controls; however, for the Adolescents and young adults (AYAs; 15.1-39 years), designated as a unique group in oncology, although being a huge group in the present study, the pattern was changed. Risk factors for other age groups, e.g., extended family, groundwater, tobacco use, and rural residency are exhibiting protective effects when compared to age-matched controls. The occurrence of the respective risk factor in all of the remaining age groups was treated as a reference in each case.

Table 3.10: Odds ratios comparing high-risk groups in patients cohort (additive model)

Categories	Variables	Risk group (Age Category)	Patients (Group 1)[a]	Patients (Group 2)[b]	Odds Ratios	95 % CI	P -Value
			N (%)	N (%)			
Gender	Male	Lymphoid	278 (46.8)	122 (20.53)	1.97	1.38-2.81	P = 0.0002
		Myeloid	104 (17.5)	90 (15.15)	1.00		
	Male	Acute	298 (50.16)	102 (17.17)	1.34	0.92-1.95	P = 0.1285
		Chronic	133 (22.39)	61 (10.26)	1.00		
Diet	Poor diet, Family type, Extended vs Nuclear	15.1-39	19 (3.19)	7 (1.17)	1.73	0.62-4.81	P = 0.2959
		5.1-15	33 (5.55)	21 (3.53)	1		
Weight Status	Underweight	1.1-5	62 (10.43)	2 (0.33)	6.89	1.48-32.16	P = 0.0141
		5.1-15	54 (9.09)	12 (2.02)	1.00		
	Underweight (Normal + Overweight)	1.1-5	62 (10.43)	10 (1.68)	7.92	3.72-16.89	P < 0.0001
		5.1-15	54 (9.09)	69 (11.61)	1.00		
	Underweight (Normal + Overweight)	1.1-5	62 (10.43)	10 (1.68)	27.44	13.31-56.57	P < 0.0001
		All Other Groups	61 (10.27)	270 (45.45)	1.00		
	Overweight CML vs CLL	39.1 & >	53 (8.92)	13 (2.18)	0.18	0.02-1.44	P = 0.1050
		15.1-39	23 (3.87)	1 (0.16)	1.00		
	Overweight CLL vs CML	39.1 & >	13 (2.18)	53 (8.92)	5.64	0.7-45.7	P = 0.1050
		15.1-39	1 (0.16)	23 (3.87)	1.00		
Consanguinity and Ethnicity	Consanguinity, Pathan vs Punjabi	5.1-15	48 (8.08)	26 (4.37)	1.45	0.78-2.66	P = 0.2368
		All Other Groups	60 (10.1)	47 (7.91)	1		
Fuel, Water, and Area	Wood, groundwater, rural area vs Gas	5.1-15	59 (9.93)	21 (3.53)	1.04	0.54-1.99	P = 0.9047
		All Other Groups	81 (13.64)	30 (5.05)	1		
	Wood, groundwater, rural area vs Gas	15.1-39	26 (4.37)	4 (0.67)	2.68	0.89-8.1	P = 0.0807

		All Other Groups	114 (19.19)	47 (7.91)	1		
	Wood, groundwater, rural area vs Gas	39.1 & >	24 (4.04)	4 (0.67)	2.43	0.8-7.39	P = 0.1172
		All Other Groups	116 (19.53)	47 (7.91)	1		
	Wood, groundwater, rural area vs Gas	5.1-15	59 (9.93)	21 (3.53)	0.47	0.15-1.51	P = 0.2036
		39.1 & >	24 (4.04)	4 (0.67)	1		
	Rural, Ground Water vs Urban	5.1-15	103 (17.34)	18 (3.03)	1.02	0.35-2.99	P = 0.9686
		39.1 & >	28 (4.71)	5 (0.84)	1		
	Area Rural, Ground Water vs Urban	1.1-5	53 (8.92)	9 (1.51)	1.02	0.45-2.31	P = 0.9541
		All Other Groups	161 (27.1)	28 (4.71)	1		
	Area Rural, Ground Water vs Urban	5.1-15	103 (17.34)	18 (3.03)	0.98	0.49-1.97	P = 0.9536
		All Other Groups	111 (18.69)	19 (3.2)	1		
Area	Area, Rural vs Urban	1.1-5	66 (11.11)	16 (2.69)	1.28	0.65-2.52	P = 0.4713
		5.1-15	103 (17.34)	32 (5.38)	1		
Tobacco use	Tobacco use	39.1 & >	51 (8.58)	75 (12.62)	1.06	0.66 - 1.72	P = 0.7994
		5.1-15	62 (10.43)	97 (16.32)	1		
Disease Origin	Lymphoid	15.1-39	73 (12.28)	75 (12.62)	0.43	0.3-0.63	P < 0.0001
		All Other Groups	309 (52.02)	137 (23.06)	1.00		
	Myeloid	15.1-39	75 (12.62)	73 (12.28)	2.32	1.58-3.39	P < 0.0001
		All Other Groups	137 (23.06)	309 (52.02)	1.00		
Disease Nature	Acute	15.1-39	103 (17.34)	45 (7.58)	0.82	0.55-1.24	P = 0.3514
		All Other Groups	328 (55.22)	118 (19.87)	1.00		
	Chronic	15.1-39	45 (7.58)	103 (17.34)	1.21	0.81-1.83	P = 0.3514
		All Other Groups	118 (19.87)	328 (55.22)	1.00		
	Chronic	39.1 & >	112 (18.85)	40 (6.73)	21.47	13.49-34.15	P < 0.0001
		All Other Groups	51 (8.59)	391 (65.82)	1.00		
Blast Cells %	Blast Cells >50%, Lymphoid	5.1-15	76 (12.79)	21 (3.53)	1.25	0.68-2.32	P = 0.4720

	vs ≤50% Lymphoid						
		All Other Groups	104 (17.52)	36 (6.06)	1.00		
	Blast Cells >50%, Lymphoid vs ≤50% Lymphoid	5.1-15	76 (12.79)	21 (3.53)	1.11	0.44-2.82	P = 0.8203
		15.1-39	26 (4.37)	8 (1.34)	1.00		
	Blast Cells >50%, Lymphoid vs ≤50% Lymphoid	5.1-15	76 (12.79)	21 (3.53)	2.41	0.87-6.67	P = 0.0895
		39.1 & >	12 (2.02)	8 (1.34)	1.00		
	Blast cell ≤50% Myeloid vs ≤50%Lymphoid	39.1 & >	14 (2.35)	8 (1.34)	5.72	2.01-16.23	P = 0.0011
		All Other Groups	15 (2.53)	49 (8.25)	1.00		
	Blast cells ≤50%, Chronic vs Acute	39.1 & >	17 (2.86)	5 (0.84)	27.69	7.78-98.49	P < 0.0001
		All Other Groups	7 (1.17)	57 (9.59)	1.00		
	Blast Cells >50%, Acute VS ≤50% Acute	5.1-15	87 (14.64)	26 (4.37)	1.04	0.58-1.85	P = 0.8978
		All Other Groups	116 (19.53)	36 (6.06)	1.00		
	Blast Cells ≤50% Chronic vs >50%, Chronic	39.1 & >	17 (2.86)	6 (1.01)	0.18	0.01-3.61	P = 0.2619
		All Other Groups	7 (1.18)	0 (0)	1.00		
	Blast Cells >50%, Chronic VS ≤50% Chronic	39.1 & >	6 (1.01)	17 (2.86)	5.57	0.28-112.02	P = 0.2619
		All Other Groups	0 (0)	7 (1.18)			
Hemoglobin Status	Hb Decrement, Blast Cells >50%, Lymphoid vs ≤50% Lymphoid	5.1-15	61 (10.26)	16 (2.69)	1.39	0.52-3.69	P = 0.5130
		15.1-39	22 (3.7)	8 (1.34)	1		
	Hb Decrement, Blast Cells >50%, Lymphoid vs ≤50% Lymphoid	5.1-15	61 (10.26)	16 (2.69)	1.91	0.57-6.37	P = 0.2946
		39.1 & >	10 (1.68)	5 (0.84)	1		
	Hb Decrement, Blast cell ≤50% Myeloid vs >50%myeloid	39.1 & >	10 (1.68)	7 (1.17)	1.63	0.4-6.63	P = 0.4927
		15.1-39	7 (1.17)	8 (1.34)	1		
Platelets Status	PLT decrement, Blast cell ≤50% Lymphoid vs >50%Lymphoid	5.1-15	15 (2.52)	58 (9.76)	0.77	0.37-1.58	P = 0.4685

	All Other Groups	26 (4.38)	77 (12.96)	1		
PLT decrement, Blast cell >50%Myeloid vs >50% Lymphoid	15.1-39	8 (1.34)	22 (3.7)	2.42	0.93-6.29	P = 0.0705
	All Other Groups	17 (2.86)	113 (19.02)	1		
PLT Decrement, Blast cell ≤50%, Myeloid vs ≤50% Lymphoid	15.1-39	2 (0.33)	7 (1.17)	1.62	0.27-9.75	P = 0.5988
	All Other Groups	6 (1.06)	34 (5.72)	1		

[a] Group 1: Patients exposed to the risk variable; [b] Patients not exposed to the variable

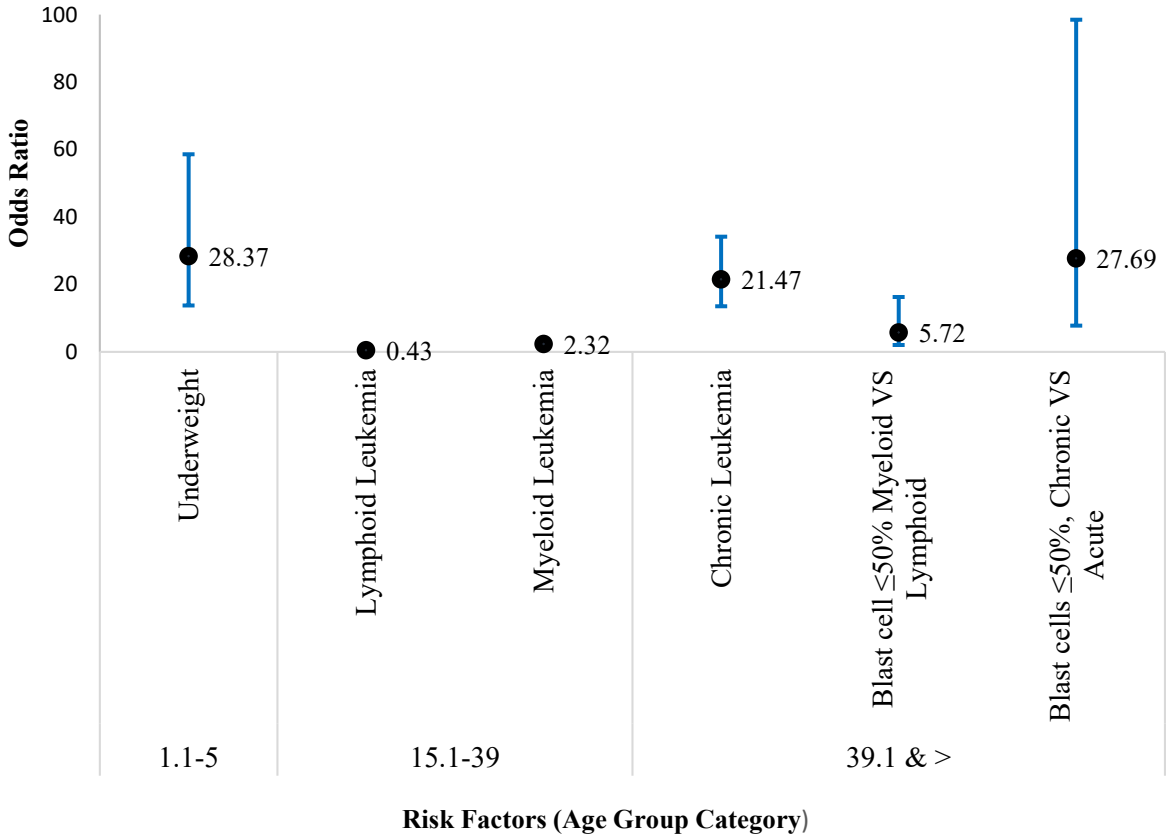


Figure 3.36: Odds ratios of high-risk patients' age group within the patient cohort

Odds ratios for identification of high-risk patients' age group within the patient cohort (CI 95%, p* = 0.05), showed that among patients' age groups, children (1.1-5 years) with lower BMI (ref. lower BMI in remaining age groups additively) were at higher risk for the disease. Patients in age group 15.1-39 have a higher tendency to get myeloid leukemia (ref. myeloid leukemia in all other age groups additively) and patients older than 39 years of age have a higher risk for chronic leukemia (ref. chronic leukemia in remaining age groups additively). In patients older than 39 years old, the blast cells % for myeloid leukemia (ref. Blast cells ≤50% in myeloid leukemia in remaining age groups additively) and chronic leukemia (ref. Blast cells ≤50% chronic leukemia in remaining age groups additively) were significantly lower than lymphoid and acute in the same age groups, respectively.

3.2 Genetic analysis

This section includes the results of molecular genetic analysis of selected leukemia patients' DNA to check the mutational status of crucial genes. The genetic analysis was performed through four different techniques to explore as much as possible

(i). Mutational screening through ARMS

(ii). Mutation screening through targeted NGS (Tar-Seq)

(iii). Mutational screening through microarray

3.2.1 Mutation Screening through ARMS

3.2.1.1 Results

3.2.1.1.1 Samples characteristics

Patient samples (n 276) containing good-quality DNA were used for mutation screening, of .which197 (71.38%) were male and 79 (28.62) female, 56% with <14 years of age, 21% 15-39 years age group and 21% >40 years old. A high frequency (78%) of the patients were affected by leukemia of lymphoid lineage, and 21% of the myeloid lineage. Early-stage leukemia (acute leukemia) was present in 79% of the patients, while chronic leukemia in 20%. Pathan race comprised 69% of the patients, 25% were Punjabis, and 5% of other ethnicities. Family history was observed in 26 (9%), and parental consanguinity in 98 (35%) of the patients included in the study (Table 3.11; Figure, 3.37).

Table 3.11: Demographic details of leukemia patients included in the study

Variables	Male	Female	Total	OR	95% CI
	N (%)	N (%)	N (%)		
Age categories (Years)					
Up to 15	110 (71)	46 (29)	100 (57)	1.15	0.590 to 2.243
>15-39	43 (72)	17 (28)	60 (22)	1.09	0.488 to 2.424
>39	44 (73)	16 (27)	60 (22)	Ref.	
Total	197 (71)	79 (29)	276 (100)		
Disease Origin*					
Lymphoid	162 (75)	55 (25)	217 (79)	Ref.	
Myeloid	35 (59)	24 (41)	59 (21)	2.02	1.105 to 3.691
Disease Nature					
Acute	158 (72)	62 (28)	220 (80)	Ref.	
Chronic	39 (70)	17 (30)	56 (20)	1.11	0.585 to 2.109
Ethnicity					
Pathan	137 (72)	53 (28)	190 (69)	1.16	0.358 to 3.760
Punjabi	48 (69)	22 (31)	70 (25)	1.38	0.398 to 4.748
Others	12 (75)	4 (25)	16 (6)	Ref.	
Family History					
Yes	18 (69)	8 (31)	26 (9)	1.12	0.466 to 2.694
No	179 (72)	71 (28)	250 (91)	Ref.	
Parental Consanguinity					
Yes	70 (71)	28 (29)	98 (45)	0.81	0.454 to 1.446
No	81 (67)	40 (33)	121 (55)	Ref.	

*Distribution was statistically significant in Chi-test statistics

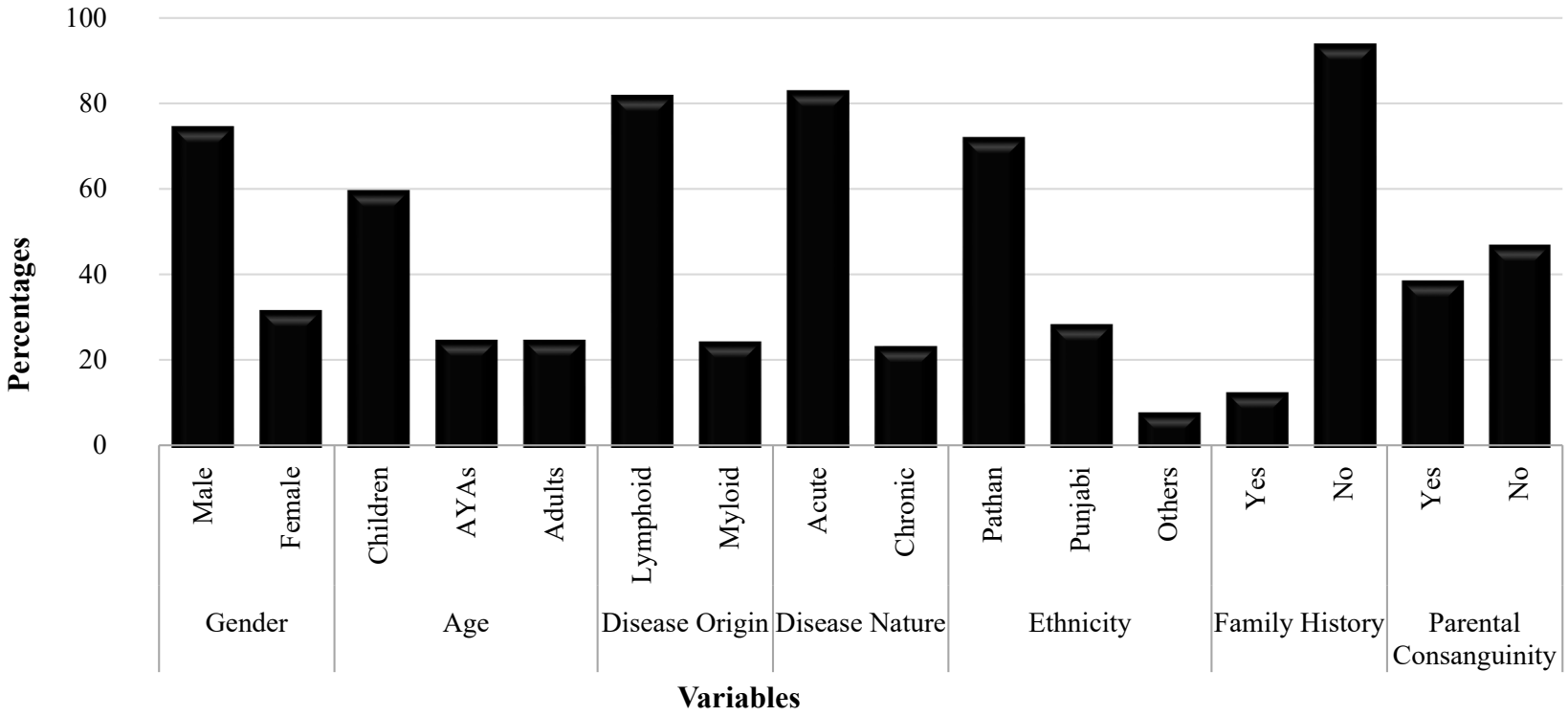


Figure 3.37: Percentage distribution of demographic risk factors of the leukemia patients screened for *JAK1*^{V623A}, *JAK2*^{S473}, and *STAT5B*^{N642H}

Percentages were calculated to find the frequencies of demographic risk factors. The male gender (71.38%) was the dominant group in disease. Children (0-14 years; 56.52%) were the most recurrent age group affected by leukemia. Lymphoid (78.62%) and acute (79.71%) forms of leukemia were more prevalent in the studied population as compared to lymphoid and chronic leukemia, respectively. Leukemia was highly recurrent in the Pathan race (68.84) than in other local ethnicities in the area. However, 9.42% and 35.51% of the leukemia patients had a family history of leukemia and solid cancers, and consanguineous parental union, correspondingly.

3.2.1.1.2 Identification of cases with STAT5B mutations

Among the patients, all 276 were homozygous (normal) for the *JAK* mutations *JAK1*^{V623A} and *JAK2* ^{S473}. However, in the case of STAT5B in leukemia, 6 (2.17%) patients were heterozygous for STAT5B^{N642H} and the remaining 270 (97.8%) were homozygous (wild type) for the same genetic location (Table 3.12; Figure 3.38). STAT5B^{N642H} is located on exon 16 of the *STAT5B* gene, which transcribes and translates into the SH2 domain of the STAT5B protein (Figure 3.39).

Acute lymphoblastic leukemia (ALL) comprised almost 70% of the total sample. All 6 of the patients having the mutations were from the same subtype. Acute myeloid leukemia (AML) and chronic myeloid leukemia (CML) were present in similar frequencies (10%), while the least occurring leukemia subtype observed was chronic lymphocytic leukemia (CLL; 9%). Unlike ALL, all these subtypes were homozygous (wt) for all screened mutations (Table 3.13, Figure 3.40).

Table.3.12: Mutational status of JAK1^{V623A}, JAK2 ^{S473} and STAT5B^{N642H} in leukemia

Mutations	Patients		Total
	Homozygous (Wild type) N (%)	Heterozygous (Mutant) N (%)	N
JAK1$^{V623A\ (T/C)}$	276 (100)	0	276
JAK2 $^{S473\ (A/T)}$	276 (100)	0	276
STAT5B$^{N642H(A/C)}$	270 (97.8)	6 (2.17)	276

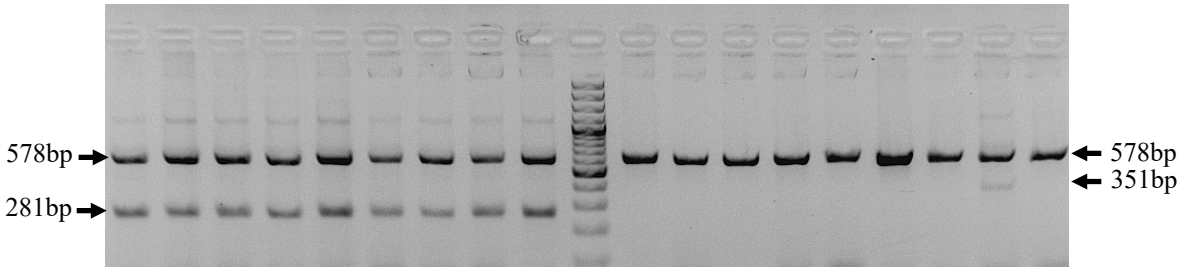


Figure 3.38: Gel picture showing the bands for allele-specific PCR for STATFB^{N642}

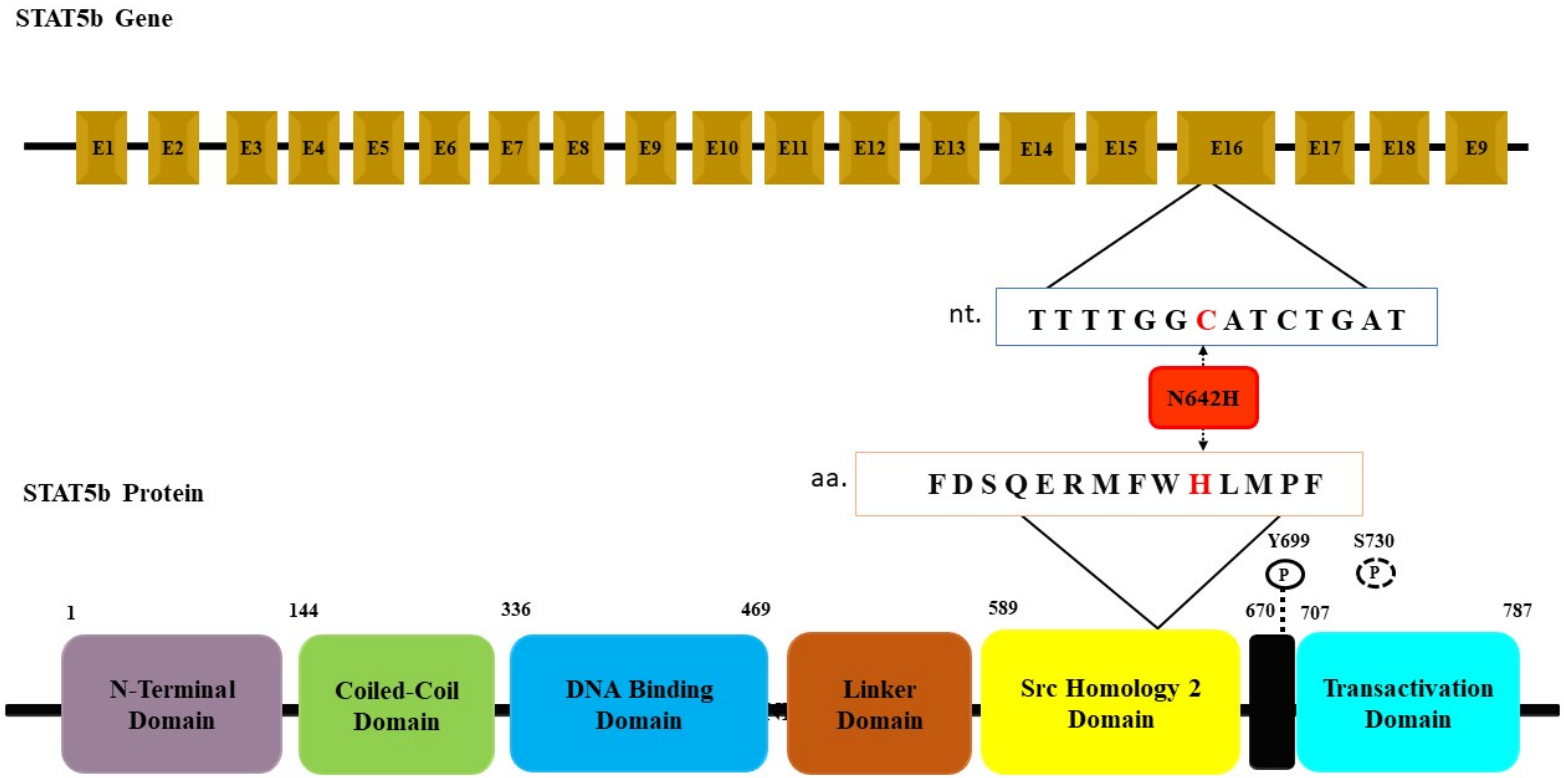


Figure 3.39: Percentage distribution of demographic risk factors of the leukemia patients screened for *JAK1*^{V623A}, *JAK2*^{S473}, and *STAT5B*^{N642H}

Table 3.13: Mutational status of *JAK1*^{V623A}, *JAK2* ^{S473} and *STAT5B*^{N642H} in leukemia subtypes

Leukemia Subtype	*JAK1*$^{V623A\ (T/C)}$		*JAK2* $^{S473\ (A/T)}$		*STAT5B*$^{N642H(A>C)}$		Total Patients
	Homozygous	**Heterozygous**	**Homozygous**	**Heterozygous**	**Homozygous**	**Heterozygous**	
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
ALL	190 (68.84)	0	190 (68.84)	0	184 (66.67)	6 (2.17)	190 (68.84)
AML	30 (10.87)	0	30 (10.87)	0	30 (10.87)	0	30 (10.87)
CLL	27 (9.78)	0	27 (9.78)	0	27 (9.78)	0	27 (9.78)
CML	29 (10.51)	0	29 (10.51)	0	29 (10.51)	0	29 (10.51)
Total	276 (100)	0	276 (100)	0	270 (97.83)	6 (2.17)	276 (100)

*Homozygous; Normal, Heterozygous; Mutant

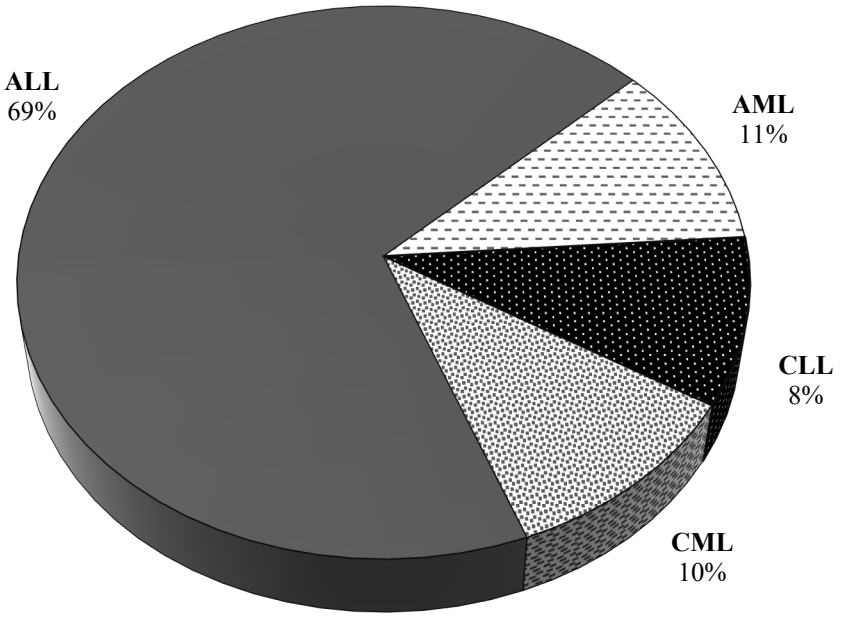


Figure 3.40: Percentage distribution of leukemia subtypes in the studied population screened for *JAK1*^{V623A}, *JAK2* ^{S473,} and *STAT5B*^{N642H} in leukemia.

Acute lymphoblastic leukemia (ALL) comprised almost 70% of the total sample. Acute myeloid leukemia; AML and Chronic myeloid leukemia; CML, were present in similar frequency, while the least occurring leukemia subtype observed was chronic lymphocytic leukemia (CLL).

3. 2.1.1.3 Clinical features associated with STAT5B^{N642H}

All 6 of the patients carrying the missense STAT5B^{N642H} mutation were children, aging 10 to 11 years, and affected with acute lymphoblastic leukemia. There was only one female patient, the remaining five patients harboring the STAT5B^{N642H} were male. Five patients, male, with known blood groups, were Rh+), and three of them were O^+ group. Three of the patients were underweight, two were within normal weight range and 1 was overweight. Parental consanguinity was present in only 1 patient among mutants. The paternal age at childbirth (patient) was >25 for all the patients, in which 4 of the fathers were 30 to 40 years old. In the case of maternal age at the patient's birth, 5 were >25 categories, among which, 2 were in the range of 30 - 40 years old (Table 3.14).

In clinical information, the lymphocyte was higher than normal (30-40%) for 3 of the patients, hemoglobin range was lower than normal (11.5-15.5 g/dL) for 4 of the patients, and platelets also were lower than normal (170-450 $\times$ 10^3/ul) for 4 of the patients. One of the patients carrying STAT5B^{N642H} was a Down syndrome, another had weak bones, and the remaining four has no significant complaints other than leukemia (Table 3.14).

In lifestyle-related risk factors, 4 out of 6 patients were having a poor diet, similarly, 4 of 6 patients were drinking water directly from underground sources, and 4 of the 6 patients were living in rural areas. Parental smoking was observed in 3 of the patients. Of all, 50% of the mutated patients were using wood/biomass as domestic fuel and the remaining 50% were using natural gas for the same purpose. All the patients were exposed to 2 or more 2 lifestyle-related risk factors individually (including the single female patient, exposed to all the mentioned risk factors), except for LP21, which is not exposed to any of the mentioned risk factors, and also, overweight and had Down syndrome (Table 3.14).

Table 3.14: Clinical findings and lifestyle-related risk factors of the acute lymphoid leukemia patients harboring STAT5B^{N642H}

Patient's IDs		LP27	LP72	LP74	LP163	LP279	LP400
Mutational Status	N642H(A/C) *	Heterozygous	Heterozygous	Heterozygous	Heterozygous	Heterozygous	Heterozygous
Demographical Variables	Gender	Male	Male	Male	Female	Male	Male
	Age (Years)	10	11	10	10	11	10
	Blood Group	A+	O+	B+	O+	-	O+
	Weight Status	Normal	Overweight	Underweight	Under\weight	Normal	Underweight
	Parental Consanguinity	No	No	No	No	No	Yes
	Parity	1	4	3	2	3	7
	Father age **	40	39	28	33	28	44
	Mother age	28	38	27	28	24	36
Clinical Information	ALL	ALL	ALL	ALL	ALL	ALL	ALL
	Lymphocytes Status (%)	59	80	14	78		
	Hemoglobin Status (g/dL)	10.3	10.1	11.7	7.3	7.7	-
	Platelets Status (10^3/ul)	74	68	172	38	73	-
	Patients Other Diseases	Weak bones	Down syndrome	Weakness	No	N0	No
Lifestyle Related Risk Factors	Diet	Poor	Standard	Poor	Poor	Standard	Poor
	Domestic Fuel Source	Gas	Gas	Wood	Wood	Gas	Wood
	Water Type	Groundwater	Municipal supply	Groundwater	Groundwater	Groundwater	Municipal supply
	Parental Smoking	Yes	No	Yes	Yes	No	No
	Area	Rural	Urban	Urban	Rural	Rural	Rural

*Homozygous; Normal, Heterozygous; Mutant, ** Age of the parents at childbirth.

3. 2.1.1.4 Association of parental ages with leukemia

There were 62 (41.89%) mothers, who were ≤25 years old at the time of the patient's birth, among which 30 (20.27%) were those, whose partners were of the same age at the time, and for 32 (21.62%) women, the partner was >25 years old. Only 1 of the 33 women, in the age range 25 to 30 years old had a partner ≤25 years old, for the remaining 32, the male partners were of the age >25 years, 17 of whom were >30 to 40 years old. Of 39 women, in the age group >30 to 40 years, 38 had a partner age of >25, of which 24 were in the same range as the women. For all the women > 40 years, the male partner at the time of childbirth was also in the same age group (Table 3.15A).

Unlike maternal age, where the largest group was ≤25 for patients 62 (41.89) and controls 88 (46.32), for the patient age at birth, 30 to 40 was the most recurrent group for both patients 53 (35.81) and controls 77 (40.53). The statistical analysis showed that for maternal age, the group > 40 years old is significantly at lower risk than the reference group (OR= 0.10; 95% CI= -3.81- -0.78), however, the size of the group is very small. There was no other significant parental age group. The detailed results of the regression analysis are provided in Table 3.15B.

Table 3.15A: Relationship of maternal to paternal age, at birth for leukemia patients in comparison to matched controls (n=338)

Maternal Age	Paternal Age									
	Patients					Controls				
	Paternal Age ≤25	All Age Groups		Paternal Age >25	Total	Paternal Age ≤25	All Age Groups		Paternal Age >25	Total
Age Range (Years)	N (%)	Range	N (%)	N (%)	N (%)	N (%)	Range	N (%)	N (%)	N (%)
≤25	30 (20.27)	25 to30	20 (13.51)	32 (21.62)	62 (41.89)	33 (17.37)	25 to30	34 (17.89)	55 (28.94)	88 (46.32)
		30 to 40	12 (8.11)				30 to 40	18 (9.47)		
		>40	0				>40	3 (1.58)		
25 to 30	1 (0.68)	25 to30	15 (10.14)	32 (21.62)	33 (22.3)	3 (1.58)	25 to30	14 (7.37)	44 (23.15)	47 (24.74)
		30 to 40	17 (11.49)				30 to 40	28 (14.74)		
		>40	0				>40	2 (1.05)		
>30 to 40	1 (0.68)	25 to30	3 (2.03)	38 (25.68)	39 (26.35)	0	25 to30	0	53 (27.89)	53 (27.89)
		30 to 40	24 (16.22)				30 to 40	31 (16.32)		
		>40.1	11 (7.43)				>40	22 (11.58)		
>40	0	25 to30	0	14 (9.46)	14 (9.46)	0	25 to30	0	2 (1.052)	2 (1.05)
		30 to 40	0				30 to 40	0 (0)		
		>40	14 (9.46)				>40	2 (1.05)		
Total	32 (21.62)		116 (78.38)	116 (78.38)	148 (100)	36 (18.95)		154 (81.05)	154 (18.05)	190 (100)

Patients, whose parental age information was not available are excluded.

Table 3.15B: Comparison of parental age at birth for leukemia patients to matched controls (n=338)

Parental Age	Case	Control	Total	Coefficient	OR	Std. Err.	z	P>|z|	[95% CI]
	N (%)	N (%)							
Maternal age									
≤25	62 (41.89)	88 (46.32)	150	Ref.					
_Cons*									
25 to 30	33 (22.3)	47 (24.74)	81	0.0342	1.03	0.2829	0.12	0.904	-0.52 - 0.59
_Cons				-0.6614	0.52	0.2177	-3.04	0.002	-1.09 - -0.23
30 to 40	39 (26.35)	53 (27.89)	92	-0.0175	0.98	0.2696	-0.06	0.948	-0.55 - 0.51
_Cons				-0.4895	0.61	0.206	-2.38	0.017	-0.89 - -0.09
>40	14 (9.46)	2 (1.05)	16	-2.2961	0.10	0.7739	-2.97	0.003*	-3.81 - -0.78
_Cons				-1.4881	0.23	0.2959	-5.03	0.000	-2.07 - -0.91
Total	148 (100)	190 (100)	339	-		-	-	-	-
Paternal Age									
≤25	32 (21.62)	36 (18.95)	68	-0.1912	0.83	0.3049	-0.63	0.531	-0.79 - 0.41
_Cons				-0.5691	0.57	0.2285	-2.49	0.013	-1.02 - -0.12
25 to 30	38 (25.68)	48 (25.26)	86	-0.1399	0.87	0.2811	-0.5	0.619	-0.69 - 0.41
_ Cons				-0.3327	0.72	0.2126	-1.57	0.118	-0.75 - 0.08
30 to 40	53 (35.81)	77 (40.53)	130	Ref.	`				
_Cons									
>40	25 (16.89)	29 (15.26)	54	-0.2251	0.80	0.3261	-0.69	0.490	-0.86 - 0.41
_ Cons				-0.7514	0.47	0.2426	-3.1	0.002	-1.23 - 0.28
Total	148 (100)	190 (100)	338	-		-	-	-	-

Cons*: The logit value of the multinomial logit model if all the other variables are equal to zero.

3.2.2 Mutation screening through targeted NGS (TarSeq)

3.2.2.1 Results

The DNAs of three male toddlers (age; 2-2.5 years) were selected for screening through a targeted next-generation technique by using Ion Ampliseq Cancer Hot Spot Panel 2. A total of 1653 positions, spanning 50 genes, were sequenced through a targeted NGS panel and evaluated by comparing genotype combinations to find any genetic variations present.

3.2.2.1.1 Patient 1

The patient I was a 2-year-old male, Punjabi, underweight, diagnosed with acute lymphoblastic leukemia, with 78% blast cell at diagnosis. The patient's blood group was B+, and the blood profile showed the presence of leukocytopenia, anemia, neutropenia, and lymphocytosis; however, platelets level was in the normal range according to the standard for the age and gender-specific reference accepted for the population. The immunophenotyping showed the presence of LCA, CD20, Ki67, Pax5, CD30, MPO, ALK markers.

The patient had no other disease or family history of any cancer and was born to a consanguineous union and was the youngest offspring. The patient was living in a nuclear family in a rural neighborhood, nutrition was poor, and natural gas was used by the family as a primary fuel source, along with groundwater consumption for drinking. The patient was homozygous for all three *JAK/STAT* mutations screened through ARMS analysis previously. The patient presented genetic changes at 22 positions, among which 20 were single nucleotide variations, and 2 were multiple nucleotide variations. Among all, 14 were exonic variations, of which 3 (including *TP53*; rs1042522) were missense mutations and 11 were synonymous. Of the remaining 8 mutations, 6 were intronic and two were localized in the downstream/UTR[3] region. Genotyping revealed that 7 variations were homozygous, comprising of 4 exonic, 1 intronic, and 2 downstream/3-UTR positioned variations, Though, the remaining 15 mutations were heterozygous, three of them were exonic missense mutations. The details of all the variations for patient 1 are given in Table 3.16.

Table 3.16: The mutations found in Patient I; 2 years old male child with acute lymphoblastic leukemia, through targeted NGS.

Locus	*Gene*	REF [a]	Geno type	MAF [c]	Zygosity (mutant)	Structural Location	Exon	Coding [b]	Codon	Nature	Clinvar	P-value	Transcript (mRNA)	OMIM	DB-SNP	Reported in Leukemia*
chr17: 7579472	***TP53***	G	**G/C**	0.457	Heterozygous	Exonic	4	*c.215C>G*	CGC	Missense	Drug response	7.45E-283	NM_000546.5	191170	rs1042522	Yes
chr2: 29443666	***ALK***	C	**C/T**	-	Heterozygous	Exonic	23	*c.3551G>A*	GAG	Missense		9.60E-07	NM_004304.4	105590	rs1670600278	Not directly/ Lymphoma
chr18: 48584594	***SMAD4***	A	**A/T**	-	Heterozygous	Exonic	6	c.767A>T	CTG	Missense		1.75E-05	NM_005359.5	600993		Not direct
chr4: 1807894	***FGFR3***	G	**A/A**	0.044	Homozygous	Exonic	14	c.1953G>A	ACA	Synonymous		0	NM_000142.4	134934	rs7688609	Not directly
chr4: 55141050	***PDGFRA***	**A**	**G/G**	0.042	Homozygous	Exonic	12	c.1701A>G	CCG	Synonymous		0	NM_006206.5	173490	rs606231209	Not directly
chr4: 55152040	***PDGFRA***	C	**C/T**	0.24	Heterozygous	Exonic	18	c.2472C>T	GTT	Synonymous	Benign	0	NM_006206.5	173490	rs2228230	Not directly
chr10: 43613843	***RET***	G	**T/T**	0.287	Homozygous	Exonic	13	c.2307G>T	CTT	Synonymous	Benign	0	NM_020975.4	164761	rs1800861	Not directly
chr5: 112175769	***APC***	G	**A/A**	0.335	Homozygous	Exonic	16	c.4479G>A	ACA	Synonymous		0	NM_000038.5	611731	rs41115	No
chr3: 178952020	***PIK3CA***	C	**C/T**	0.041	Heterozygous	Exonic	21	c.3075C>T	ACT	Synonymous	Benign	0	NM_006218.3	171834	rs17849079	Not directly
chr7: 55249063	***EGFR***	G	**G/A**	0.433	Heterozygous	Exonic	20	c.2361G>A	CAA	Synonymous	Benign	5.06E-187	NM_005228.4	131550	rs1050171	No
chr7: 55249063	***EGFR-AS1***	G	**G/A**	0.433	Heterozygous	Exonic _nc	2	c.2361G>A	CAA	Synonymous	Likely benign	5.06E-187	NR_047551.1	131550	rs1050171	No
chr7: 116339672	***MET***	C	**C/T**	0.088	Heterozygous	Exonic	2	c.534C>T	AGT	Synonymous	Benign	0	NM_001127500.2	164860: 603514	rs35775721	Yes/Myeloid
chr9: 133750318	***ABL1***	C	**C/T**	-	Heterozygous	Exonic	7	c.1149C>T	GGT	Synonymous		9.31E-09	NM_005157.5	189980		Yes/ABL+
chr11: 534242	***HRAS***	A	**A/G**	0.297	Heterozygous	Exonic	2	*c.81T>C*	CAC	Synonymous	Benign	0	NM_001130442.2	190020	rs12628	No
chr5: 149433596	***CSF1R***	TG	**GA/GA**	0.203	Homozygous	UTR_3	22	c.*1841TG>GA		Unknown		2.62E-302	NM_005211.3	164770	rs386693509	Yes/Acute myeloid
chr4: 55980239	***KDR***	C	**T/T**	0.456	Homozygous	Intronic		*c.798+54G>A*		Unknown		0	NM_002253.2	191306	rs7692791	No
chr4: 55946354	***KDR***	G	**G/T**	0.072	Heterozygous	Intronic		*c.3849-24C>A*		Unknown		0	NM_002253.2	191306	rs10006115	No
chr3: 178917005	***PIK3CA***	A	**A/G**	0.273	Heterozygous	Intronic		c.352+40A>G		Unknown		7.29E-269	NM_006218.3	171834	rs3729674	No
chr13: 28610183	***FLT3***	A	**A/G**	0.437	Heterozygous	Intronic		*c.1310-3T>C*		Unknown		0	NM_004119.2	136351	rs2491231	Yes
chr19: 1220321	***STK11***	T	**T/C**	0.36	Heterozygous	Intronic		c.465-51T>C		Unknown		1.81E-189	NM_000455.4	602216	rs2075606	No
chr2: 212812097	***ERBB4***	T	**T/C**	0.355	Heterozygous	Intronic		*c.421+58A>G*		Unknown		0	NM_005235.2	600543	rs839541	No

[a] Wildtype genomic reference, [b] Genes with reverse orientation (minus strand) are presented in Italics, [C] Minor allele frequency.

3.2.2.1.2 Patient II

Patient II was a 2-year-old male, Pathan, diagnosed with acute lymphoblastic leukemia, with 95% blast cell at diagnosis. The patient's blood group was A+, and the blood profile showed the presence of anemia, neutropenia, and lymphocytopenia; however, platelet levels were in the normal range according to the standard for the age and gender-specific reference accepted for the population.

The patient had no other disease or family history of any cancer, was born to a consanguineous union and was the youngest offspring. The patient was living with an extended family in a rural neighborhood, nutrition was standard, and natural gas was used by the family as a primary domestic fuel source, along with groundwater consumption for drinking. The patient was passively exposed to smoking and other forms of tobacco use. Like Patient I, Patient II was also homozygous (wildtype) for all three *JAK/STAT* mutations screened through Allele-specific analysis previously.

Genetic variations were presented at 18 different positions, among which 16 were single nucleotide variations, and 2 were multiple nucleotide variations (MNV), including deletion and inversion in one of the MNV loci. Among all, 8 were exonic mutations, of which 1 (*ALK* was a missense mutation, and 7 variations were synonymous. Of the remaining 10 variations, 9 were intronic and 1 was localized in the UTR^3 region. Genotyping revealed that 6 variations were homozygous, comprising 2 exonic, 3 intronic, and 1 UTR^3 variations, Though, the remaining 12 mutations were heterozygous, 6 of them were exonic synonymous mutations. The details of all the variations for patient II are given in Table 3.17.

Table 3.17: The mutations found in patient II; 2 years old male child with acute lymphoblastic leukemia, through Targeted NGS

Locus	Gene	REF [a]	Genotype	MAF [b]	Zygosity	Structural Location	Exon	Coding	Codon	Nature	ClinVar	P-Value	Transcript	*OMIM*	dbSNP	Reported in Leukemia
chr2: 29443666	*ALK*	C	C/T		Heterozygous	Exonic	23	**c.3551G>A**	GAG	Missense		0.00001	NM_004304.5	*105590*	**rs1670600278**	Lymphoma
chr2: 29432625	*ALK*	C	C/A	0.103	Heterozygous	Intronic		c.3836+27G>T	-	Unknown		0.00001	NM_004304.5	*105590*	rs3738868	Lymphoma
chr2: 212578379	*ERBB4*	T14AG	T13AG/T11AG3A	0.0, 0.443	Heterozygous	Intronic		c.884-7delT, c.884-21_884-18delinsTTTC	-	Unknown	Benign	0.00001	NM_005235.3	600543	rs1412472143 rs67894136	No
chr2: 212812097	*ERBB4*	T	C/C	0.355	Homozygous	Intronic		**c.421+58A>G**	-	Unknown		0.00001	NM_005235.3	600543	**rs839541**	No
chr4: 1807894	*FGFR3*	G	A/A	0.044	Homozygous	Exonic	14	**c.1953G>A**	ACA	Synonymous	Likely benign	0.00001	NM_000142.5	134934	**rs7688609**	Not directly
chr4: 55141050	*PDGFRA*	AGA	GGA/GGA	0.042	Homozygous	Exonic	12	**c.1701A>G**	CCG	Synonymous		0.00001	NM_006206.6	173490	**rs606231209**	Hematologic Diseases
chr5: 112175769	*APC*	CGG	CGG/CAG	0.335	Heterozygous	Exonic	16	**c.4479G>A**	ACA	Synonymous	Benign	0.00001	NM_000038.6	611731	**rs41115**	No
chr7: 55249063	*EGFR*	G	G/A	0.433	Heterozygous	Exonic	20	**c.2361G>A**	CAA	Synonymous	Benign	0.00001	NM_005228.5	131550	**rs1050171**	AML
chr7: 116339672	*MET*	C	C/T	0.088	Heterozygous	Exonic	2	**c.534C>T**	AGT	Synonymous	Benign	0.00001	NM_001127500.3	164860	**rs35775721**	Myeloid Leukemia
chr9: 133750318	*ABL1*	C	C/T		Heterozygous	Exonic	7	**c.1149C>T**	GGT	Synonymous		0.00032	NM_005157.6	189980		Leukemia
chr11: 534242	*HRAS*	A	A/G	0.297	Heterozygous	Exonic	2	**c.81T>C**	CAC	Synonymous	Benign	0.00001	NM_001130442.2	190020	**rs12628**	AML
chr4: 55946354	*KDR*	G	G/T	0.072	Heterozygous	Intronic		**c.3849-24C>A**	-	Unknown		0.00001	NM_002253.3	191306	**rs10006115**	No
chr4: 55980239	*KDR*	C	T/T	0.456	Homozygous	Intronic		**c.798+54G>A**		unknown		0.00001	NM_002253.3	191306	**rs7692791**	No
chr3: 178917005	*PIK3CA*	A	A/G	0.273	Heterozygous	Intronic		**c.352+40A>G**		unknown	Uncertain	0.00001	NM_006218.4	171834	**rs3729674**	Not directly
chr13: 28610183	*FLT3*	A	G/G	0.437	Homozygous	Intronic		**c.1310-3T>C**		unknown		0.00001	NM_004119.3	136351	**rs2491231**	Myeloid Leukemia
chr19: 1220321	*STK11*	T	T/C	0.36	Heterozygous	Intronic		**c.465-51T>C**		unknown	Uncertain	0.00001	NM_000455.5	602216	**rs2075606**	No
chr22: 24176287	*DERL3*	G	G/A	0.152	Heterozygous	Intronic		c.1119-41G>A		unknown	Benign	0.00001	NM_003073.5	610305	rs5030613	AML
chr5: 149433596	*CSF1R*	TG	GA/GA	0.203	Homozygous	UTR_3		c.*1841TG>GA		unknown	Uncertain	0.00001	NM_005211.3	164770	rs386693509	Myeloid Leukemia

[a] Wildtype genomic reference, [b] Minor allele frequency

3.2.3 Mutational screening through microarray assay

The cytoscan microarray assay was performed on the DNA from a male ALL patient. The microarray screening identified the presence of 97 loss of heterozygosity (LOH) at 97 different loci, and three copy number variations in comparison to the reference genome. Loss of heterozygosity is a common type of allelic imbalance in which one of the two alleles in a heterozygous somatic cell is lost, resulting in homozygosity. Copy number variation (CNV) is a term that refers to sections of the genome that are repeated and the number of repetitions varies from individual to individual. CNVs are divided into two major types: Gain of copy number and loss of copy number variations. Among the three CNVs detected, 2 were loss of copy number variations, and 1 was gain of copy number variation.

Both of the loss of copy number variations were located on chromosome 8 of the studied ALL patient genome, located at positions p11.22 and q24.23 (Figure 3.41). The copy number loss at q24.23 is flanked by segments with LOH regions, p11.1 and p12. This particular fragment harbors 17 OMIM genes including *KAT6A* (601408), *PLAT* (173370), *DKK4* (605417), *IKBKB* (603258), *POLB* (174760), *VDAC3* (610029), *CHRNB3* (118508), *CHRNA6* (606888), *SLC20A2* (158378), *THAP1* (609520), *RNF170* (614649), *HOOK3* (607825), *FNTA* (134635), *POMK* (615247), *HGSNAT* (610453), and *POTEA* (608915). The stretch lost at position q24.23 resides 2 OMIM genes named *UNC5D* (616466) and *KCNU1* (615215).

A gain of copy number variation identified in the present study was located on the q arm of chromosome 14, q32.33 (Figure 3.41). This particular location codes for gene *FAM30A* (616623). The same locus is carrying a LOH region, harboring 8 OMIM genes, including JAG2 (602570), *BRF1* (604902), *NUDT14* (609219), *PACS2* (610423), *CRIP2* (601183), *MTA1* (603526), *CRIP1* (123875), *FAM30A* (616623). The details of the copy number variations and LOH segments detected on chr. 8 and chr.14 are listed in Table 3.18.

Table 3.18: Copy number aberrations spanning LOH positions

Locus	Band Start[a]	Band End	Type	CNS[b]	Size (kbp)	Marker Count	Gene Count	OMIM Genes (ID) in the segment
Chr.8	p11.21	p11.1	LOH [d]		1884.17	273	21	*KAT6A (601408), AP3M2 (610469), PLAT (173370), IKBKB (603258), POLB (174760), DKK4 (605417), VDAC3 (610029), SLC20A2 (158378), CHRNB3 (118508), CHRNA6 (606888), THAP1 (609520), RNF170 (614649), HOOK3 (607825), FNTA (134635), POMK (615247), HGSNAT (610453), POTEA (608915)*
	p11.22	p11.22	Loss [e]	1	138.601	73	2	*ADAM5, ADAM3A*
	p12	p11.23	LOH		1465.6	297	3	*UNC5D (616466), KCNU1 (615215)*
	q12.3	q13.1	LOH		1236.05	250	7	*BHLHE22 (613483), CYP7B1 (603711)*
	q22.2	q22.2	LOH		1191.09	192	7	*STK3 (605030), OSR2 (611297), VPS13B (607817), COX6C (124090), RGS22 (615650)*
	q24.23	q24.23	Loss	1	179.926	119	1	*LINC02055*
Chr.14	q23.1	q23.1	LOH		1819.08	429	21	*DAAM1 (606626), GPR135 (607970), L3HYPDH (614811), JKAMP (611176), RTN1 (600865), DHRS7 (612833), PPM1A (606108), C14orf39 (617307), SIX6 (606326), SIX1 (601205), SIX4 (606342), MNAT1 (602659), TRMT5 (611023), SLC38A6 (616518)*
	q23.2	q23.3	LOH		1102.62	281	9	*PPP2R5E (601647), SGPP1 (612826), SYNE2 (608442), ESR2 (601663), MTHFD1 (172460)*
	q23.3	q24.1	LOH		1269.37	278	8	*GPHN (603930), PALS1 (606958), ATP6V1D (609398), EIF2S1 (603907), PLEK2 (608007)*
	q32.33	q32.33	LOH		1452.58	81	25	*JAG2 (602570), NUDT14 (609219), BRF1 (604902), PACS2 (610423), MTA1 (603526), CRIP2 (601183), CRIP1 (123875), FAM30A (616623)*
	q32.33	q32.33	Gain [f]	3	861.82	268	10	*FAM30A(616623), MIR8071-1, MIR8071-2, ELK2AP, MIR4539, MIR4507, MIR4538, MIR4537, ADAM6, LINC00226*

[a]; Cytoband, [b]; Copy number state, [c]; All genes names are provided for copy number variation only, [d]; Loss of heterozygosity, [e]; Copy number loss, [f]; Copy number gain

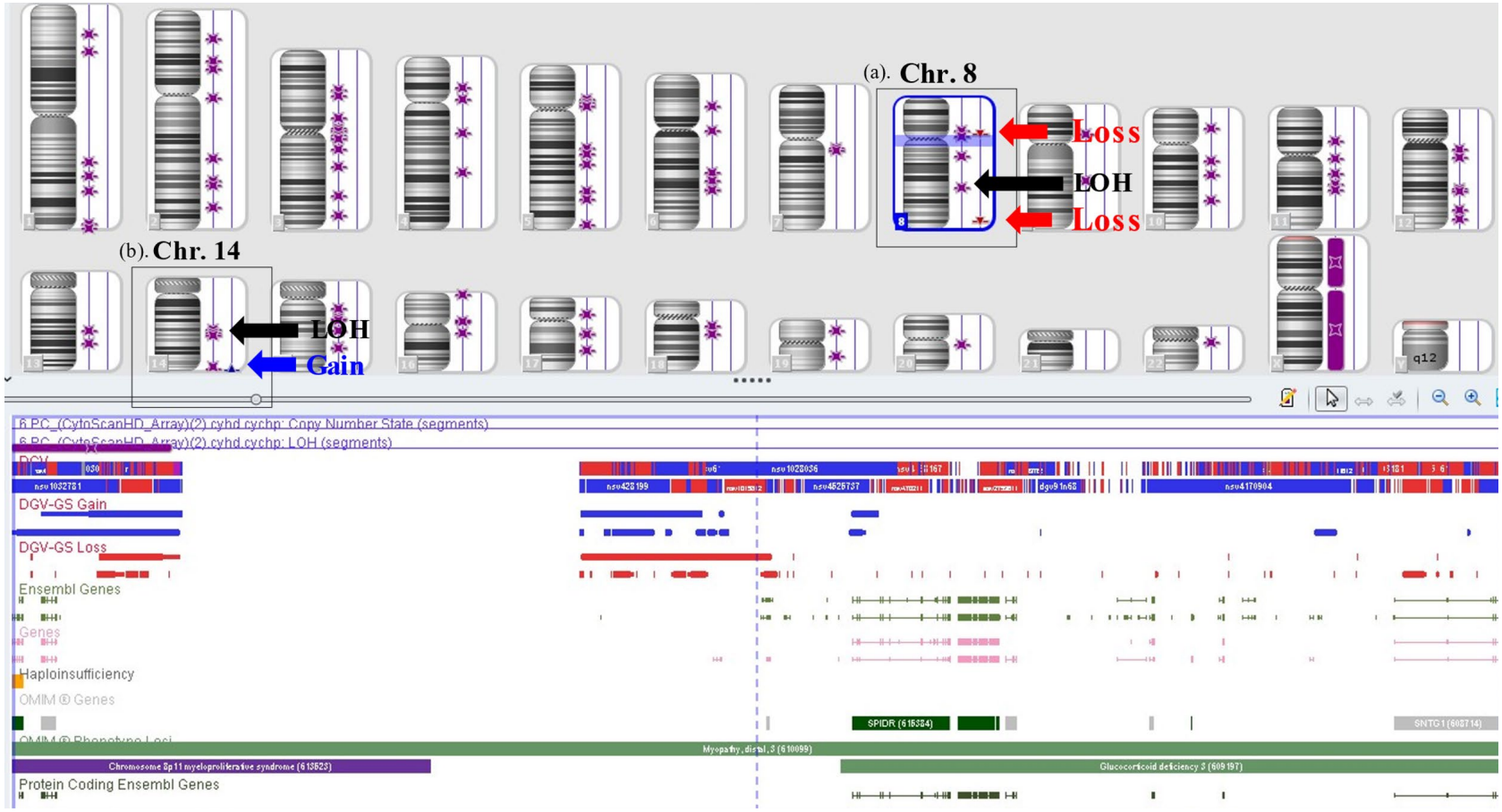


Figure 3.41: The graphical interface of the ChAS, showing the positions of copy number variation.

The loss of copy number on chromosome 8 is shown in red, and the gain of copy number variation on chromosome 14 is shown in blue. The loss of heterozygosity is highlighted with a black pointer.

3.3 *In silico* protein drug interaction analysis

STAT5B is interacting in vivo with a good number of downstream biomolecules, either via direct physical binding or by contribution to a shared role, thus contributing to the mediation of crucial signaling pathways, e.g. cell growth, proliferation, apoptosis, through both canonical and non-canonical mechanisms as evident by the protein interaction based STRING analysis (Szklarczyk et al., 2021). Hence, the effect of mutation on the expression or activation of STAT5B, not only poses a direct influence on JAK/STAT pathway but also causes disruptions in all these cascade reactions (Figure 3.42).

3.3.1 Molecular docking analysis

In the current study, in silico virtual screening through molecular docking, was performed for all eight ligands with the target wild type STAT5B and STAT5B^{N642H+} protein individually. The ligands were probed for their binding affinity and stability in the STAT5B and STAT5B ^{N642H} active sites. Molecular docking studies revealed that among 8 selected compounds, 4 compounds showed adequate, but differential binding affinities towards the normal and mutated STAT5B. The ligands with the highest affinity, lower RMSD value, and higher number of strong interactions observed in the present study were selected as the potential inhibitors.

3.3.2 Molecular interaction of the ligands with wildtype STAT5B

The study revealed that the ligands, 17F (-8.197 Kcal/mol; 1 H bond-Lys600, 1 electrostatic interaction, 9 hydrophobic interactions), AC_4_130 (-8.295 Kcal/mol; 6 H bonds, C-H bonds, 5 hydrophobic interactions) and Pimozide (-8.324 Kcal/mol; 3 H bonds-Ser541, 1 C-H bonds, 6 hydrophobic interactions) has better docking, but not necessarily to SH2 domain, as compared to other ligands, like AC_3_019 (-7.594 Kcal/mol; 1 C-H bonds, 12 hydrophobic interactions, 5 halogen interactions), Indirubin (-7.028 Kcal/mol; 3 H bonds-GLN601, HIS604, PHE640, 7 hydrophobic interactions), and Stafib_2 (-7.643 Kcal/mol; 4 H bonds-ASN642,

ASP621, ILE629, 1 electrostatic interaction, 7 hydrophobic interactions) (Table 3.19, Figure 3.43.& 3.44).

3.3.3 Molecular interaction of the ligands with STAT5B ^{N642H}

In the case of STAT5B^{N642H} domain and ligands interactions, it is found that 17F (-8.047 Kcal/mol; 1 C-H bond, 5 hydrophobic interactions) has a comparable affinity for the mutant protein. AC_4_130 (-7.483 Kcal/mol; 1 C-H bond, 4 hydrophobic interactions, 5 halogen interactions), has a lower affinity for the mutated protein than for the normal. Stafib_2 (-7.527 Kcal/mol; 6 H bonds-GLN340, GLU269, GLY263, GLU335, 2 C-H bonds, 4 electrostatic interaction, 2 hydrophobic interactions) has similar docking affinity for both the wildtype and mutant protein; however, Indirubin (-7.097 Kcal/mol; 2 H bonds-GLN601, HIS604, PHE640, 2 electrostatic interaction, 5 hydrophobic interactions), has a slightly higher affinity for the mutant version of the protein. Apart from these, Piceatannol and Sulforaphane, both have lower affinities for SH2-domains of both the wildtype and mutated STAT5B proteins (Table 3.19, Figure 3.43.& 3.44).

The present docking study indicated that AC_4_130 (-8.295 Kcal/mol; 6 H bonds, C-H bonds, 5 hydrophobic interactions) has the most stable interaction, and higher affinity for wildtype STAT5B protein, Stafib_2 has stable, strong docking interaction with mutant protein as compared to normal due to presence of extra 2 interacting H-bonds and 3 electrostatic interactions as compared to the wildtype, but the interaction is outside of the SH2 domain. Indirubin and Pimozide have good but differential docking affinities for both the normal and mutant protein; however, the interaction of pimozide with normal protein is stronger due to the presence of 3 hydrogen bonds (Table 3.19, Figure 3.43.& 3.44).

The present study concludes that AC_4_130, pimozide, indirubin, and stafib 2 have higher, but differential docking affinities for the SH2-domain of both normal and mutated STAT5B. The AC_4_130 has the potential to inhibit the STAT5B signaling via SH2-AC_4_130 interactions. The strong interactions, maintained by multiple hydrogen bonds and electrostatic interactions, of stafib-2 and indirubin with the SH2 domain of STAT5B $^{N642H+}$

proteins, indicate that these compounds with slight modifications can achieve even better interactions. These compounds in combination can play a suggestive role in the targeted inhibition of STAT5B $^{N642H+}$ proteins and hence can prove helpful in aggressive acute lymphoblastic leukemia harboring the mutated STAT5B $^{N642H+}$ protein (Table 3.19, Figure 3.43 & 3.44).

Table 3.19: List of STAT5B inhibitors, associated with STAT5B, in resistant leukemia

S#	Ligand	Protein STAT5B	Mode	affinity (kcal/mol)	Distance from best mode		Protein-Ligand Interaction				Amino acids (H bond)
					RMSD l. b	RMSD u. b	*C./CH H-Bonding	Electrostatic	Halogen	Hydrophobic	
1	**17F**	Normal	2	-8.197	10.02	12.87	1/1	1		9	LYS600
		N642H+	4	-8.047	2.501	10.7	0/1			5	-
2	**AC_3_019**	Normal	2	-7.594	1.152	2.218	0/1		5	12	-
		N642H+	5	-6.998	4.279	7.963	3/1	0	5	4	LYS514*, GLU518
3	**AC_4_130**	Normal	2	-8.295	1.237	2.195	6/3			5	GLN337, GLN340, GLN340, GLN340, ASP575, GLU579
		N642H+	5	-7.483	13.18	15.29	0/1	0	5	4	
4	**Indirubin**	Normal	2	-7.028	1.857	6.123	3/0	0		7	GLN601, HIS604, PHE640
		N642H+	5	-7.097	1.87	6.059	2/0	2		5	ASN567, LYS536
5	**Piceatannol**	Normal	4	-5.8	2.058	2.829	5/0	0		2	ASP621, PHE640*, VAL598, THR628
		N642H+	2	-6.262	13.84	15.96	3/0	1		1	HIS604, ARG618, TRP631,
6	**Pimozide**	Normal	2	-8.324	2.661	3.977	3/1	0	2	6	SER541*
		N642H+	3	-8.14	1.85	4.318	0/2	0	2	5	-
7	**Stafib_2**	Normal	2	-7.643	7.672	9.622	4/0	1		7	ASN642, ASP621, ILE629
		N642H+	3	-7.527	5.678	7.728	6/2	4		2	GLN340, GLU269, GLY263, GLU335
8	**Sulforaphane**	Normal	2	-3.217	11.8	12.92	1/0	0		0	PHE640
		N642H+	3	-3.257	3.45	5.048	2/1	0		0	HIS584, TYR682

* Conventional Hydrogen bonding/ Carbon Hydrogen bonding

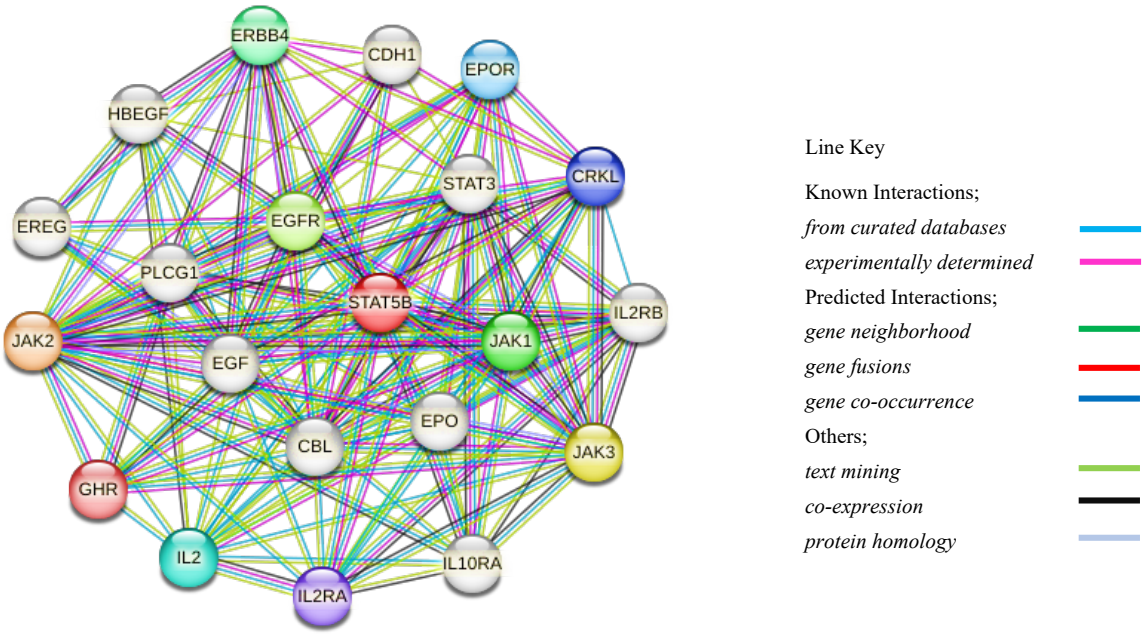


Figure 3.42: STAT5B interactome with other proteins regulating several transduction pathways.

Signal transducer and activator of transcription 5B; Carries out a dual function: signal transduction and activation of transcription. Mediates cellular responses to the cytokine KITLG/SCF and other growth factors. Binds to the GAS element and activates PRL-induced transcription. Positively regulates hematopoietic/erythroid differentiation (787 aa). Each node is a protein, and each edge is an interaction between those proteins replicated either in multiple publications or multiple bioassays as reported by the integrated interactions database (Szklarczyk et al., 2021). The color of the lines is representative of different types of protein interactions. The length of the edges is arbitrary. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding to each other.

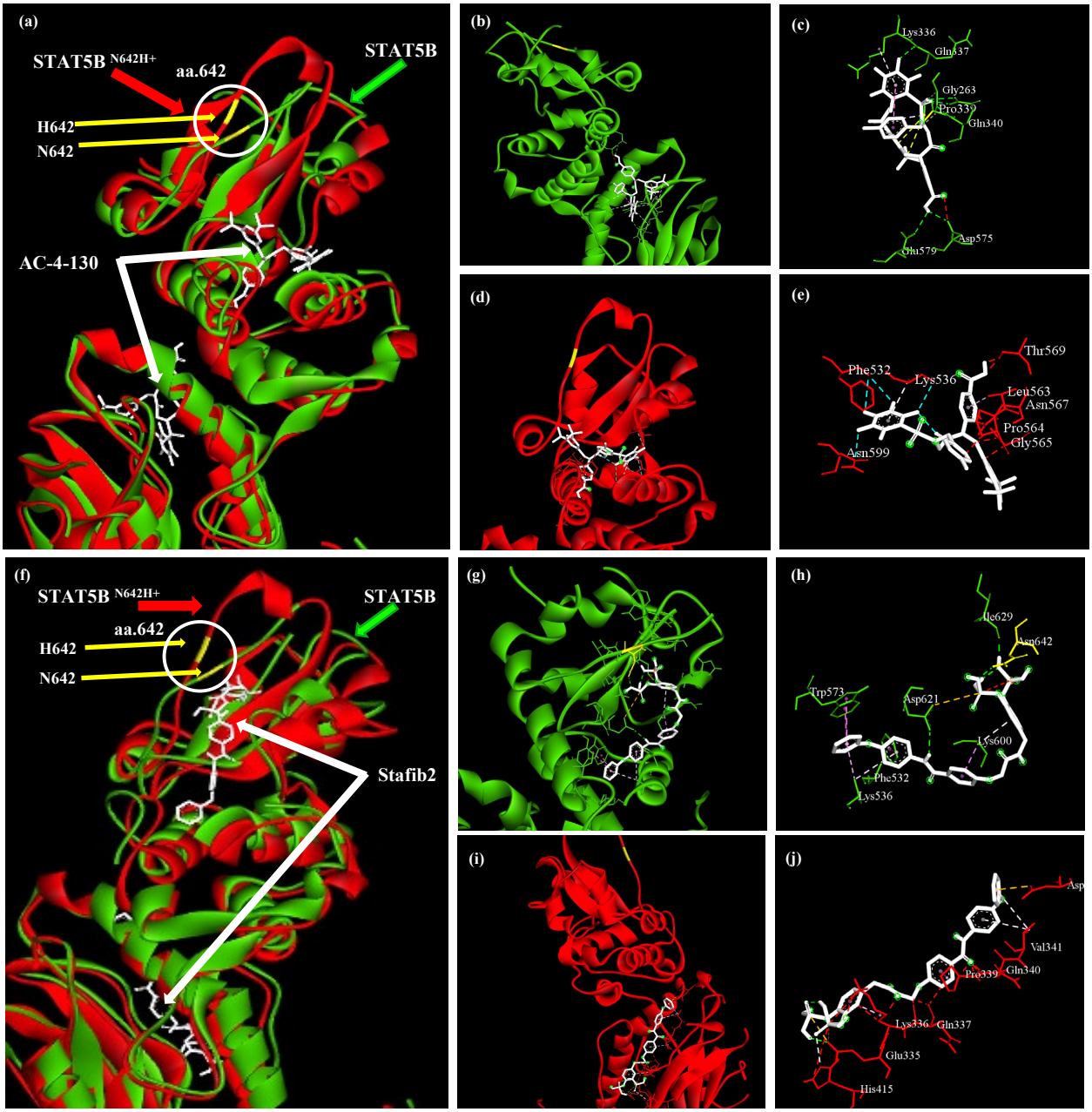


Figure 3.43: Interaction of AC-4-130 and Stafib-2 with wildtype STAT5B and STAT5B^{N642H+} proteins.

Visual 3D structural overlay of SH2 domains of wildtype STAT5B (green) and STAT5B^{N642H+} (red) proteins, highlighting the position of the respective mutated amino acid in both proteins (yellow), and the position of ligand interaction (white). Images **a-e** show the interaction of **AC-4-130,** and **f-j** show the interaction of **Stafib-2** to the wildtype and STAT5B^{N642H+}, respectively.

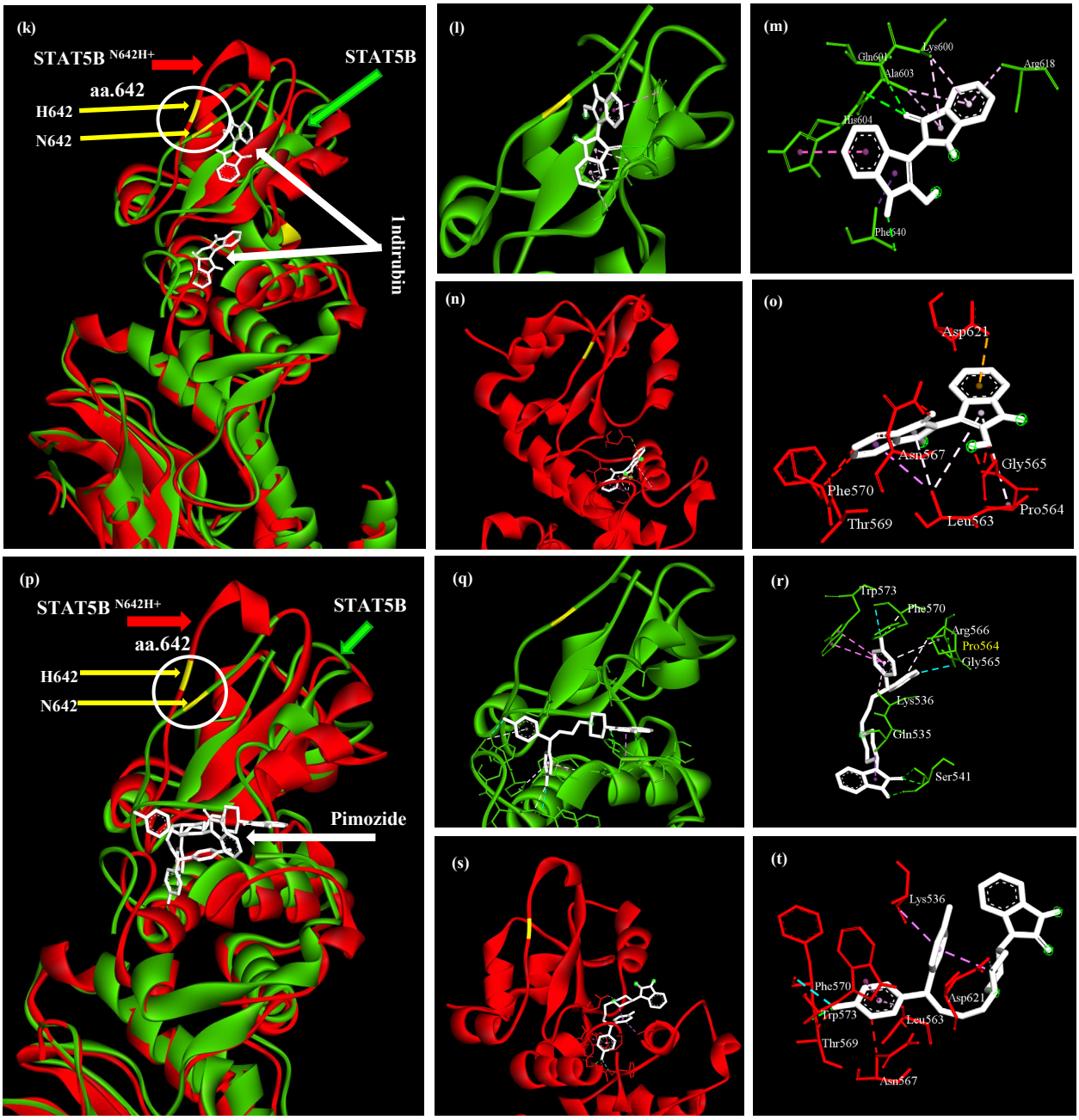


Figure 3.44: Interaction of Indirubin, and Pimozide with wildtype STAT5B and STAT5B^{N642H+} proteins.

Visual 3D structural overlay of SH2 domains of wildtype STAT5B (green) and STAT5B^{N642H+} (red) proteins, highlighting the position of the respective mutated amino acid in both proteins (yellow), and the position of ligand interaction (white). Images **k-o** are showing the interaction of **Indirubin,** and **p-t** are showing the interaction of **Pimozide** to the wildtype and STAT5B^{N642H+}, respectively.

CHAPTER 4

DISCUSSION

4. Discussion

4.1 Epidemiology

We investigated demography, clinical aspects, and environmental and lifestyle risk factors for leukemia patients in comparison to matched controls. This study provides a measured magnitude of clinical-epidemiological risk factors associated with leukemia in the Pakistani population.

ALL was the most frequent leukemia subtype in this study population. The age groups used in present study were sub-classified based on developmental status into three major groups: Children (<1-15), Adolescents and young adults (AYAs) (15.1-39), and aged (>39). AYAs serve as a bridge between children and the aged and have been widely accepted as a unique group when it comes to cancer pathology, genetics, and treatment outcomes. Like our population, neighboring countries also report ALL being the highest occurring leukemia subtype, particularly in white children (Jha and Kumar, 2021). Geographical variation play role in leukemia subtype distribution across the globe. The male-to-female ratio was 2.1:1. This is consistent with the findings of other studies conducted in Pakistan (Sultan *et al.*, 2016a; Pervez *et al.*, 2020). There could be several possible explanations for the increased preponderance of male patients. Differential genetic susceptibility, occupational exposure to carcinogens, smoking, and gender bias in seeking health care may be the contributing factors. It was further observed that ALL was common in children and adolescents while CML was frequent in adults. This finding is consistent with the published data reported from other countries (Höglund *et al.*, 2015; Katz *et al.*, 2015).

In this study, the difference in the occurrence of leukemia was also evident among the ethnic groups. Pathan ethnicity (white with debatable ancestry) was at higher risk for leukemia. Previous literature showed that white ethnicities are more prone to develop leukemia than black and colored (Oksuzyan *et al.*, 2015; Bispo *et al.*, 2020), but have a good overall prognosis and disease-free survival (Acharya *et al.*, 2016). Differences in leukemia incidence on the base of

race/ethnicity could be explained by variances in genetic susceptibility to environmental risk factors.

In the present study, no education is a significant risk factor for leukemia. Only a few studies in the past have investigated the relationship between subject literacy and leukemia. The observations of the present study are supported by previous literature (Barrera *et al.*, 2005). Recent studies in the country, also reported no education or lower education level in leukemia patients in different localities of Pakistan, particularly in poor socioeconomic classes (Nasir *et al.*, 2015) (Saeed *et al.*, 2019). Higher education level has been inversely associated with leukemia risk in the Columbian population (Kabat *et al.*, 2013). The research focused on Ugandans concluded that lower literacy is directly associated with the advanced cancer stage at diagnosis (Mwaka *et al.*, 2016). These studies elaborate on the protective role of education in leukemia and the higher risk of leukemia with no education. In Pakistan, lack of education is the direct outcome of lower socioeconomic status. Previous studies conducted in the country reported that the highest number of patients belonged to the lower middle class (Nasir *et al.*, 2015). Lower socioeconomic status along with no or less education also contributes to financial problems leading to late presentations of disease or early cessation of treatment. No education can lead to higher incidence and lower survival of leukemia patients directly due to lack of awareness of care and precautionary measures, unhealthy lifestyle adopted by such people who have no access to even basic education, and indirectly through financial inabilities.

Poor diet has been identified as a significant risk factor for leukemia in the present study. Studies reported the prevalence of malnutrition in children and adolescents (0-19 years of age) (Ferlay *et al.*, 2015). High frequency of undernutrition in children from developing countries, with ALL, has been reported and associated with worse treatment outcomes, and poor survival due to lower tolerance to chemotoxicity and anti-neoplastic agents when presented at the time of diagnosis (Loeffen *et al.*, 2015; Yazbeck *et al.*, 2016). The mechanisms by which nutritional status might influence cancer outcomes are hypothesized to be the differential metabolic effects based on body composition (Joffe *et al.*, 2019). Undernutrition is linked to poorer survival rates among young people with leukemia in low- and middle-income countries (Barr *et al.*, 2016). Since nutritional status is a potentially modifiable prognostic

factor, nutritional interventions, including medical nutrition therapy, are considered important supportive measures. Contrary to our observation, some previous cohort studies have recommended that diet is not associated with leukemogenesis (Saberi Hosnijeh *et al.*, 2014) and leukemia subtypes (Kabat *et al.*, 2013). These inconsistencies might be due to differences in study design and the population under study.

Poor diet is directly related to lower BMI. The present study reports lower BMI as a significant risk factor for leukemia, particularly in children; however, contrary to our observation, recent meta-analyses reported the presence of obesity at diagnosis in leukemia patients, but they associated it with poor prognosis (Castillo *et al.*, 2012; Amankwah *et al.*, 2016). The weight status of patients and its impact on risk and treatment outcomes is highly debatable.

Consanguineous marriages are practiced worldwide with a variation in prevalence. The effect of consanguinity on the occurrence of different neoplasms is not well understood. In the present study, the parental consanguineous union is seen to significantly increase leukemia risk. This observation is supported by the findings of other researchers in the country (Yasmeen and Ashraf, 2009; Mahmood *et al.*, 2020); however, studies exist that reported no direct association of consanguinity with leukemia in diverse populations (Bener *et al.*, 2009; Kakaje *et al.*, 2020). Consanguinity increases the probability of having homozygosity. Many recessive genes that predispose to leukemia and other hematological malignancies, can be found more frequently in patients with consanguineous parents (Stieglitz and Loh, 2013). A recent study stated no direct association of consanguineous parental union to leukemia but associated positive family history of leukemia patients with consanguinity which may indicate a family history of hereditary recessive cancer genes (Kakaje *et al.*, 2020). Consanguinity has complicated interactions that might either increase or decrease the susceptibility to certain cancers. All of these studies are suggesting the genetic base of leukemia. Extensive studies with large datasets are required to fully appreciate this phenomenon.

In the present study, a family history of leukemia and other cancers of the patients is reported as a significant risk factor. Similar findings have been reported in the past (Brown,

2008). The aim of obtaining the family history of oncological patients is to identify patients with a genetic predisposition or patients at risk of second malignancies. It has been reported that, though generally considered a sporadic disease, the presence of hematologic and solid cancers, in the family history of leukemia patients is relatively high (Sandner *et al.*, 2019). This indicates the potential role of genetic factors in the incidence of leukemia; It also can indicate the association between environmental and infectious agents with leukemia, because infectious and environmental factors are common between children and relatives, especially first-degree relatives.

In the present study, rural residential setups have been identified as a significant risk factor for leukemia. Leukemia has been positively associated with rural residency in Pakistan previously (Mahmood *et al.*, 2020). Similar associations have been reported by populations, geographically residing on other continents (Stubbins *et al.*, 2018; Kassahun *et al.*, 2020). A recent study identified rural residential setup as a risk factor for financial toxicity caused due to hematological malignancies (Ouchveridze *et al.*, 2022). Pakistan is an agricultural country. In rural setups where agriculture is a major profession of the inhabitants, exposure to pesticides and other agrochemicals is common. Residential proximity and prolonged exposure to agriculture-based pesticides and insecticides have been associated with increased leukemia risk (Rull *et al.*, 2009; Metayer *et al.*, 2013; Park *et al.*, 2020). In the United States, leukemia patients residing in rural area has a higher rate of early mortality (Jamy *et al.*, 2022). Less access to supportive care in rural areas is also likely a contributing factor. On the contrary, some studies denied any significant association between rurality and hematologic cancer (Mao *et al.*, 2018).

In the current study, groundwater is identified as a risk factor for leukemia. Leukemia has been associated with contaminants in groundwater in different parts of the world (Wagner *et al.*, 2011) and unfiltered water in Pakistan (Mahmood *et al.*, 2020). A study reported a positive association between organic drinking water contaminants and leukemia incidence (Cohn *et al.*, 1994). A recent study conducted in Iran concluded that exposure to high levels of heavy metals, particularly in drinking water was significantly associated with the mortality rate due to leukemia (Rahmani *et al.*, 2022). Heavy metal derivatives can react with side groups of

proteins and enzymes, leading to genetic mutations (Chen et al., 2021). Pakistan is one of the countries in Asia, which is exposed to heavy metals contaminated groundwater (Rahmani *et al.*, 2022). The continued use of contaminated groundwater with heavy metals, insecticides, herbicides, and their naturally derived compounds for drinking and irrigation can trigger several health manifestations in the human body. Contaminated groundwater with agricultural chemicals and heavy metals is the potential cause of it being a risk factor for leukemia in the population.

Biomass combustion bi-product exposure turned out to be the most ignored and consistent risk factor for cancer. The present study reports a significantly higher risk of leukemia for people who are burning wood as a household fuel source. Due to the lack of energy resources like natural gas and electricity, common men in Pakistan have no choice, but to use wood/biomass as an alternative fuel and energy source, particularly in rural settings. Inhalation of woodsmoke promotes systemic oxidative stress that increases the risk for diseases like chronic inflammation and cancer (Rabha *et al.*, 2018). Biomass use as a fuel has been considered a risk factor for cancer in recent studies (Saeed *et al.*, 2019). The generation of woodsmoke produces complex mixtures of various toxicants e.g. gaseous pollutants such as nitric oxide (NO), carbon monoxide (CO), and carbon dioxide (CO_2) are released into the atmosphere, as well as aldehydes, polycyclic aromatic hydrocarbons (PAHs), and particulate matter (PM) (Peters *et al.*, 2018; Scott and Reilly, 2019), linked to adverse health outcomes, including cancer (Avenbuan and Zelikoff, 2020), among which PAHs and carbon monoxide are well known to play a role in increased leukemia risk (Bailey *et al.*, 2011). A study directly associated leukemia-related mortality with the concentration of CO and NO in the air (Dehghani *et al.*, 2017). There is not much literature available addressing the association of fuel type with leukemia; however, on the contrary, several studies associated a higher risk of leukemia with unconventional oil and natural gas development (Elliott *et al.*, 2017; Mckenzie *et al.*, 2017).

Little is known about the relationship between family characteristics and leukemia. The size of the family and the order of birth mostly impact the risk of some cancers (Feller *et al.*, 2010). In the current study, there was no overall association between extended family to

increased leukemia risk; however, the risk for children (1-15 years) living in extended families was significantly increased. In concordance with our study, an elevated risk of AML, particularly for older siblings, living in larger families has been reported (Altieri *et al.*, 2006). There could be multiple logics for this observation. Children living in extended families must compete for common resources. The providers might not be able to provide proper nutrition and care and maintain a hygienic living environment for the children. Also, a major number of study participants were from rural setups where people are usually financially vulnerable and live tighter in large families to cope with demands in minimal resources. Extended family set up is also supported as a cultural value in Pakistan, especially in villages.

Carbonated drinks have been observed to significantly increase leukemia risk in the present study. Like our study, a previous study reported a slightly increased risk for childhood ALL due to the consumption of cola-based drinks (Thomopoulos *et al.*, 2015). Although there is not much literature available directly associating carbonated drinks with leukemia studies show an association of artificial sweeteners like aspartame and methyl-imidazole, which are used in carbonated drinks, to increased risk of leukemia in laboratory animals (Jacobson and Michael, 2011), and are potential carcinogens to human (Grosse *et al.*, 2011); However, there are also studies, who denies any positive association between components of carbonated drinks and hematopoietic cancer (Bernardo *et al.*, 2016).

In the current study, tobacco use in any form and state is associated with an increased risk of leukemia. Cigarettes contain numerous harmful constituents and tobacco use is well known to cause a variety of cancers in adults, including leukemia, via both direct and secondary means of exposure (Colamesta *et al.*, 2016). Smoking has been reported to be a strong risk factor for leukemia in Pakistani patients (Nasir *et al.*, 2015). Benzene from smoking contributes to increasing leukemia risk (Fiebelkorn and Meredith, 2018). Apart from benzene, radioactive metals like polonium-210, present in tobacco smoke are also known to promote carcinogenesis through benzene interactions (Hecht, 1999). A recent study associated smokeless tobacco use with cancer (Saeed *et al.*, 2019). In the current study, there was a high occurrence of tobacco use in subjects with CLL and AML. As the subjects with CLL were predominantly adult males, this observation explains the higher exposure. Previous studies

have reported a positive association between smoking with CLL compared to other leukemia subtypes (Richardson *et al.*, 2008). Parental smoking is associated with increased leukemia risk in children (Metayer *et al.*, 2016; Chunxia *et al.*, 2019). In the Puerto Rican population, parental smoking has been associated with an increased risk of AML children (Frederiksen *et al.*, 2020).

In the present study, perfume use has been observed to have a protective effect on leukemia. There is no previous study available, evaluating perfume use as a protective factor for leukemia in the general population. There could be several potential justifications for these protective effects. The bases/mediums used in perfumery in Asian countries mostly consist of plant-based oils and fragrances e.g. davanone, cedar oil, and linalool are some of the plant extract-based fragrances used in the subcontinent perfume production industry. These compounds have recently been reported to have anticancer/antileukemic properties and the potential to induce apoptosis in vitro *(*Hung *et al.,* 2020*) (*Gong *et al.,* 2020; Xiao *et al.,* 2020). Several other plant-based perfume bases also have been reported to have similar properties (N Adham *et al.*, 2020; Mileva *et al.*, 2021). Along with this, perfume use by the population is linked to general cleanliness and health consciousness, which is direct prevention. The study found no significant correlation between perfume use to other variables like smoking, microwave use, education, or other risk factors. Also, there is no study available in previous literature that associates these factors together to leukemia or one another in cancer epidemiology.

According to our study using a microwave oven for food has a protective effect on leukemia. There could be several possible explanations for this i.e., a larger portion of our research population was from a rural setup and lower socioeconomic status, where there is no affordability for a microwave oven, and in our population, the trend is to use microwave oven occasionally, for heating food items, instead of cooking it. The occasional use and the fact that microwave uses weak electromagnetic waves to heat the food can potentially explain why microwave oven do not pose any risk of leukemia in the studied population; However, we haven't any information regarding the number and frequency of daily/weekly use of microwave oven by the study participants. Similar explanations, regarding the association of

microwave to leukemia, have been claimed by previous studies (Kaufman *et al.*, 2009; Tabrizi and Hosseini, 2015). The latest study claimed that the use of a microwave oven for food processing is not a risk for carcinogenesis (Guzik *et al.*, 2022). However, exposure to strong electromagnetic fields (EMF) has been considered to increase the risk of leukemia (Ghahremani *et al.*, 2020).

In our study, a significant association was observed between leukemia and the higher blast percentage. More than 70% of the cases were carrying >50% blast cells. A study showed previously that bone marrow blasts ≥67% were associated with shorter overall survival (Ren *et al.*, 2017). In the current study, the blast percentage was significantly lower than 50% for myeloid and chronic leukemias. It is evident that these diseases affect cells at a more advanced stage of differentiation and maturation. These subtypes are comparatively more aggressive and difficult to treat than acute leukemia. However, studies have associated a lower blast percentage with poor prognosis.

Higher WBCs counts have been associated with leukemia previously. A high WBC count is known to have a high risk for poor prognosis (Jha and Kumar, 2021). The current study reported leukocytosis, lymphocytosis, and neutropenia as significantly associated with leukemia. Like our study, a higher WBCs count has also been reported by the studies conducted in Pakistan (Fadoo *et al.*, 2015). Other studies conducted in Pakistan also reported leukocytosis in all four major subtypes of leukemia in the recent past (Chang *et al.*, 2016a; Munir and Khan, 2019; Mahmood *et al.*, 2020). Leukocytosis is associated with leukemia in other populations, along with anemia, and thrombocytopenia for all subtypes in other populations e.g. Brazil, Sudan (Basabaeen *et al.*, 2019a; Silva *et al.*, 2019; Ebrahim *et al.*, 2022).

Increased WBCs count (>10x109/l) is also associated with the risk of thrombosis (Breen *et al.*, 2012). In Acute leukemia, neutropenia (59.1%) and thrombocytopenia (75.7%) were found in Yemeni patients (Al-Maktari *et al.*, 2021). This finding is supporting the biological process that in acute leukemia, due to the infiltration of immature malignant cells in the bone marrow and arrest of maturation, there will be neutropenia, anemia, and thrombocytopenia in the peripheral circulation (Döhner *et al.*, 2015; De Kouchkovsky and

Abdul-Hay, 2016). Neutrophils are critical in providing host defense against infection. The risk of infection increases with the depth and duration of neutropenia. Neutropenia has also been associated with acute leukemia in a recent study based on an Ethiopian cohort (Kassahun *et al.*, 2020). People with neutropenia are more susceptible to bacterial infections, and without immediate medical care, the condition may become life-threatening (neutropenic sepsis). It is also known that mortality increases during leukemia treatment if neutropenia is also present (Fung *et al.*, 2015). High death rates of leukemia patients in complete remission are mainly due to infection (Lins *et al.*, 2019), neutropenia could be the possible reason.

In the current study, anemia and lower HB levels are significant clinical risk factors for leukemia. In accordance with our study, other studies also reported a high frequency of anemia (Dodhy *et al.*, 2011; Ghosh *et al.*, 2013; Zeeshan *et al.*, 2015). In previous studies, a hemoglobin level of less than 7 gm/dl was seen in 54% (Sultan *et al.*, 2016a). A Nigerian study indicated up to 74% of leukemias rearing anemia (Salawu *et al.*, 2010). Poor prognosis increased morbidity, and advanced stages are some of the manifestations exhibited, particularly by lymphoid leukemia because of anemia.

Thrombocytopenia has also been emphasized in other studies (Noetzli *et al.*, 2015). In this case-control study, an association of thrombocytopenia with leukemia has been identified. Platelet count abnormality can be considered a major factor in predicting pediatric ALL (Mahmood *et al.*, 2020). Recent studies based on Pakistani and other populations have also observed thrombocytopenia (Kassahun *et al.*, 2020; Mahmood *et al.*, 2020). This was in line with the findings reported in Kenya in which patients with CML had low platelet count and higher bleeding tendency (Othieno-Abinya *et al.*, 2016).

Regarding the etiology of leukemia, as this study indicated, there is the interaction of genetic (family history, consanguinity), nutritional (poor diet, carbonated drinks), and factors like poor lifestyle (lack of education, consumption of underground water, indoor burning of wood and biomass) in the development of leukemia in our population. There are changing trends in leukemia incidence due to changing demographic and lifestyle factors. Improvements in these factors hold the potential to improve the cancer burden.

With the advent of proper epidemiological and trans-disciplinary research, using evidence-based local data, we can expect the identification of associated risk factors and improvement in leukemia incidence and mortality. The identification of risk factors for our population will help the health department to address the high incidence of leukemia and design preventive strategies based on exposure to these risk factors encountered by the local population. Policies for primary prevention and health promotion, particularly for the unprivileged class should be a top priority to lower the incidence.

Government should make executable strategies and regulations, to efficiently deal with preventable risk factors timely, e.g. cessation program for tobacco (large-scale efforts should be made to eliminate the availability of such cancer-causing substances), provision of clean affordable fuel, decontaminated drinking water supply, Post-use processing of agriculture-based products to neutralize extra residues and toxic byproducts to avoid their mixing with groundwater channels and food chain, provide alternative animal-friendly herbicides and insecticides, proper waste management, improved education policies to include and encourage lower socioeconomic class. Disease and associated risk factor awareness should be communicated to the common man to enhance instinctive precaution. Genetic counseling should be added to basic health infrastructure to educate the general population regarding the risks and practices in our society that are potentially causing the vertical spread of diseases having genetics involved.

Initiating educational (epidemiology and cancer prevention courses) and planning programs, which involve collaboration between established cancer treatment centers and public hospitals can help in convenient management, reduce economic burden, and improve the survival of cancer patients in the country. Such courses will not only enhance knowledge and produce field experts but also will spread public awareness about the precautionary measures and manageable risk factors to which local communities are exposed, which in turn will presumably slow down leukemia incidence through improvement in quality of life and management of locally exposed risk factors. Health card advanced cancer coverage like government initiatives could be better utilized for unavoidable diseases.

4.2. Genetic analysis

This section includes the discussion for mutational analysis through all techniques listed in *sec.* methodology

4.2.1 Mutational screening through amplification refractory system

STAT5, encoded by two distinct genes STAT5A and STAT5B is one of the fundamental components of cytokine-induced signal transduction cascades, and a key downstream mediator of oncogenic Tyrosine Kinases leading to transformation. The present study reports the presence of Gain-of-function, driver mutation $STAT5B^{N642H}$ in children with ALL, transiting to adolescence. The presence of this mutation in hematological malignancies has previously been reported in the literature.

Gain-of-function mutation, $STAT5B^{N642H}$, is located in the phospho-tyrosine-binding domain, positioned in the center of the STAT5B SH2 domain, which is essential for the formation of parallel STAT5 dimers and efficient nuclear translocation, initiates enhanced and sustained tyrosine phosphorylation (Pham et al., 2018). $STAT5B^{N642H}$ mutation has a role in stabilizing the dimer interface via sampling of two different conformations that either promote or regulate dimerization, as well as prolonging activation lifetime in blood cancer cells by allowing a conformation that resists dephosphorylation, hence continually binding DNA and promoting transcription. Hyper-regulating STAT5B activity by cancer cells promotes cellular growth, metabolism, survival, and drug resistance. STAT5 signaling not only enhances protein transcription but also can repress the expression of antioxidant enzymes including glutaredoxin-1 (Glrx1) and catalase, which has been observed to promote oscillation of reactive oxygen stress in Bcr-Abl-positive CML patients (Bourgeais *et al.*, 2017).

Gain-of-function alterations in SH2, including $STAT5B^{N642H}$, and transactivation domain in STAT5B are previously acknowledged as a driver mutation in both pediatric and adult acute leukemia (Bandapalli et al., 2014; Ma et al., 2015). $STAT5B^{N642H}$ is common in pediatric leukemia and causes a higher risk of relapse (Bandapalli et al., 2014). STAT5B is

actively involved in growth and developmental pathways, and the absence of STAT5B is associated with growth retardation and dwarfism. The potential reason for the presence of the mutation STAT5B^{N642H} only in children of a certain age, observed in the present study, might be due to the reason that STAT5B is actively recruited by the growth hormone machinery as children are actively growing, and complex signaling cascades is happening to cope with growth at this crucial stage of human transit and development.

There are several mechanisms involved in leukemia aggression. One of the major contributing factors to aggressive leukemia is a mutation in genes involved in cell signaling pathways. The hyperactivation or gene silencing resulting in the constitutive expression of a defective protein or missing an important participant can lead to aggression. However, the level of aggression is dependent on the nature of the mutated gene and the type of mutation incorporated. Datasets obtained through high throughput technologies provide an overall picture of the genetic landscape of individual patients, highlighting the mutations and the genes that have the potential to induce leukemia aggression. The expression of STAT5B^{N642H} can transform cells to cytokine-independent proliferation (Bandapalli et al., 2014; Kiel et al., 2014). STAT5B ^{N642H} is involved in the rapid progression of aggressive mature T-cell neoplasia (Hennighausen and Robinson, 2008). The findings of the present study, indicate the presence aggressive form of pediatric leukemia in the studied population, as patients with the STAT5B^{N642H} mutations are being reported to suffer relatively from more aggressive and fatal disease (Rajala et al., 2013; Rajala et al., 2014).

Though in the present study, the STAT5B^{N642H} mutation is observed in the lymphoid leukemia subtype STAT5B^{N642H} has also been reported to be present in myeloid leukemia, e.g., chronic neutrophilic and eosinophilic leukemia (Luo et al., 2018; Cross et al., 2019) and clonal hematopoiesis following aplastic anemia (Babushok et al., 2015).

In the present study, all of the patients were homozygous (wt) for the screened JAKs mutations, including *JAK1*^{V623A} and *JAK2* ^{S473}. Though these mutations have been associated with leukemia, their absence in the presently studied population indicates the possibility that, as the mutation is a rare phenomenon, the mutations might have been masked due to a smaller

samples window, and a limited number of leukemia subtype, e.g., AML, CML, CLL representation in the screened samples. Leukemia involves complex multigenic and signaling interplay, the absence of these mutations signifies the chance that the present population might have mutations in other positions rather than the mentioned, which needs to be screened.

One of the expanding fields in the pathogenesis of pediatric cancer is parental reproductive factors (Hargreave et al., 2013). Various studies examining the association between leukemia risk and parental age (paternal or maternal) have drawn controversial conclusions. Previously studies have analyzed maternal age, without addressing paternal age or analyzing both maternal and paternal age separately. The present study analyzed the relative relationship between maternal age to paternal age influencing leukemia risk. No significant association was seen for any specific age group to increase leukemia risk. Some previous studies addressing the association between maternal factors and risk for childhood ALL claimed an insignificant relationship between maternal age and ALL risk (Yan et al., 2015), which is similar to the findings of the recent study, as about 70% of our research population was ALL. On the contrary, the risk of ALL has been reported to increase with increasing parental, specifically maternal age (Sergentanis *et al.*, 2015; Contreras *et al.*, 2017). An extensive meta-analysis, analyzing both cohort and case-control studies, observed a direct relation between older parental age and increased risk of acute lymphoblastic leukemia (ALL) in offspring. Furthermore, children of younger fathers also had an elevated risk of ALL, and an increased risk of AML with younger mothers (Sergentanis et al., 2015).

4.2.2. Mutational screening through targeted NGS approach

In this section of the study, we did a comprehensive screening of the genome of toddlers with acute lymphoblastic leukemia by targeting a 50 genes panel, designed specifically for cancers, comprising genes having an established relationship with oncogenesis, through targeted NGS. The mutations and genetic variations in larger genomic regions. Which requires a lot of time, expertise, and expense to be screened through traditional techniques, and can be readily done with NGS. In comparison to traditional techniques, NGS is not only economical and quick, but it also does not require high sample quality to provide reliable output (Ivanova et al., 2016). In the study, it was established that there were several hits in the genes on the panel, including, a single, single nucleotide variation in *TP53, ALK, SMAD4, FGFR3, PDGFRA, RET, APC, PIK3CA, PDGFRA, EGFR, EGFR-AS1, MET, ABL1, HRAS, PIK3CA, FLT3, STK11, ERBB4*, each, multiple single nucleotides variations in *PDGFRA, KDR* (all Exonic and intronic in location), and multiple nucleotide variations in the genes *HMGXB3* and *CSF1R*, who were located downstream of exon 22, and in the 3-UTR region following, respectively.

Genetic susceptibility to cancer is conferred both by inherited (germline) and tumor-specific (somatic) variants and as such, it is evident in most individuals, not just in those individuals with a personal or family history of cancer. Although the deleterious alleles of cancer-risk genes are generally not highly penetrant, the presence of genetic susceptibility variants at multiple loci is generally assumed to increase an individual's overall risk of cancer.

The present study reports the presence of single-nucleotide polymorphism (SNP; rs1042522) located at exon 4 on codon 72 of the *TP53* gene. The chromosomal location of *TP53* is 17p13, which plays a vital role in genome stability and maintenance through encoding the TP53 protein. By mediating the DNA damage towards their cellular response, the TP53 protein regulates cell cycle arrest, apoptosis, repair of DNA, and genetic transcription. The observed polymorphism is well established in the progression of cancer and is also investigated in leukemia. rs1042522 encodes a transverse mutation of G to C change (Arg to Pro) and its protein is known to express differentially in malignant tumors, including

leukemia (Pim and Banks, 2004). The polymorphism Pro72Arg shows functional variation, and its role is crucial in the pathophysiology of leukemia.

Data on rs1042522 polymorphism in the *TP53* gene and cancer risk are controversial. A recent metanalysis, based on 10 case controls studies (8 Asian), concluded a significant association between lower risks of chronic leukemia and Pro72Arg polymorphism in Asian and Caucasian populations, in a recessive comparison model (Akhtar and Alharbi, 2022). Earlier reported that the variant Arg72 more effectively induces apoptosis in comparison to variant Pro72 (Dumont *et al.*, 2003). A current study claimed a significant protective effect of rs1042522 in childhood acute leukemia in Moroccan children (Skhoun *et al.*, 2022). This protective effect is further strengthened by a study based on the Sudanese population, that reported the Pro/Pro genotype contribution to increased susceptibility to B-Chronic Lymphocytic Leukemia risk in the population tenfold higher than those had Arg/Arg genotype (Basabaeen *et al.*, 2019b).

The *TP53* Pro72 allele potentially increases the prognostic significance of *TP53* mutations in chronic lymphocytic leukemia (Dong *et al.*, 2014). However, in AML patients, the variant genotypes of TP53 rs1042522 were associated with adverse molecular and cytogenetic risk and with Nucleophosmin (NPM1) mutations (Tripon *et al.*, 2020). A current meta-analysis disclosed evidence of a positive association between rs1042522 and leukemia susceptibility in both recessive and overdominant genetic models, among Chinese, Japanese, American, Indian, and African populations (Drokow *et al.*, 2020). A recent study, conducted by a well-known institute in the country, reported that missense SNV rs1042522, in heterozygous genotype (GC), is more crucial, as it decreases the response to cisplatin, paclitaxel, capecitabine, and oxaliplatin anti-cancer drugs as compared the homozygous (CC) genotype (Shahid *et al.*, 2020). In contrast to these studies, no significant associations with acute leukemia are also being reported (Ruan *et al.*, 2015).

Apart from leukemia, significant associations were reported for rs1042522 with oral squamous cell carcinoma, breast cancer in Pakistani and Chinese populations (Saleem *et al.*, 2013; Gonçalves *et al.*, 2014; Peng *et al.*, 2015), and gastric and colorectal cancer in the Korean

population (Song *et al.*, 2011). Some authors suggested the existence of interactions between rs1042522 and environmental and other genetic factors (Wang *et al.*, 2015).

Another mutation reported in the present study was an *ALK* (2q23), harboring somatic, missense mutation, c.3551G>A (p. Gly1184Glu), located on exon 23, of unknown clinical impact. The protein ALK (2q23), was originally implicated in the carcinogenesis of anaplastic large-cell lymphoma as a fusion partner of nucleophosmin after a chromosomal rearrangement (Morris *et al.*, 1994). This is the first study reporting the presence of missense mutation c.3551G>A (p. Gly1184Glu), in acute lymphoblastic leukemia. There is very little literature available regarding this mutation. The presence of ALK p. Gly1184Glu was initially reported by a case study in metastatic thymic squamous cell carcinoma (Hu *et al.*, 2013). Another case study also reported the presence of the mutation of metastatic colorectal cancer along with *TP53* (p.Arg175His), *KRAS* mutation (p.Gly12Val) synonymous *ERBB2* (p.Lys860Lys) alteration (De Figueiredo Barros *et al.*, 2018); however, the mutation was not detected in circulating tumor DNA (ctDNA) and cell-free DNA (cfDNA; that has been released into circulation) of the same patient. None of the only two studies reporting the c.3551G>A (p. Gly1184Glu), evaluated the potential clinical impact of the mutation. The presence of *ALK* fusions is associated with EGFR tyrosine kinase inhibitor (TKI) resistance. However, whether this missense mutation acts like the ALK gene fusion is not known.

The next missense exonic mutation identified in this study was c.767A>T (p. Gln256Leu), located on the *SMAD4* gene. The *SMAD4* gene on chromosome 18q21.2 encodes a member of the *Smad4* (Sma (*Caenorhabditis elegans*) Mothers Against Decapentaplegia homolog 4) or *DPC4* (Deleted in Pancreatic Cancer, locus 4) (SMAD4), a signal transduction protein, that is phosphorylated and activated by transmembrane serine-threonine receptor kinases in response to TGF-β (Wrana, 2009). *SMAD4* is a tumor suppressor gene that functions as a pivotal intracellular transcriptional mediator of transforming growth factor-β(TGF-β). *SMAD 4* (DPC4) forms a complex with *SMAD 3* which can bind to DNA and modify the expression of several genes related to cellular activities such as proliferation or differentiation (Lin *et al.*, 2003).

Point mutations of *SMAD4* have only rarely been observed in leukemia. There is no previous study reporting *SMAD4* c.767A>T (p. Gln256Leu) in leukemia; though c.767A>T (p. Gln256Leu) has been affirmed in breast cancer (Konstantinova *et al.*, 2017; Orditura *et al.*, 2018), and Advanced Non-Small Cell Lung Cancers in Octogenarians (Fumagalli *et al.*, 2019). Other *SMAD4* mutations, like 4P102L, have been studied in AML (Imai *et al.*, 2001), IVS9 (1139 +2 T>G) in AYAs, ph+ ALL (Sartor *et al.*, 2013). A current study reported the p.K88X (NM_005359: exon3: c. A262T) mutations in myelodysplastic syndrome, as it progressed to AML (Zhang *et al.*, 2022). The clinical impact of *SMAD4* c.767A>T (p. Gln256Leu) mutation has yet to be studied.

Once called "silent mutations" and assumed to not affect protein structure and function, synonymous variants are now recognized to be drivers for some cancers. There have been significant advances in the understanding of the numerous mechanisms by which synonymous single nucleotide variants (sSNVs) can affect protein structure and function by affecting micro-RNA (miRNA) binding and regulation of gene expression (Gartner *et al.*, 2013), pre-mRNA splicing (Pagani *et al.*, 2005), mRNA structure and stability (Hamasaki-Katagiri *et al.*, 2017), translation rate, co-translational folding, protein conformation (Buhr *et al.*, 2016), phosphorylation (Karakostis *et al.*, 2019), protein function and localization (Kimchi-Sarfaty, 2011), and cell fitness (Parmley and Hurst, 2007).

The study identified a synonymous genetic alteration, rs7688609 (c.1953G>A; Thr651Thr) at fibroblast growth factor receptor 3 (*FGFR3*) exon 14. FGFR3 is a receptor tyrosine kinase (RTK) of the FGF receptor family activated by ligand binding to their extracellular domain, that commonly induces dimerization and activation of the intracellular kinase domain. RTK activation can induce cell responses e.g., cell migration, proliferation, altered survival, or differentiation. The clinical impact of the current genetic variation is likely benign according to the Clinvar database. No previous study is available identifying rs7688609 (c.1953G>A; Thr651Thr) in leukemia. However, some recent studies reported the presence of rs7688609 in anaplastic ependymoma (Butt *et al.*, 2019). Craniopharyngioma (Jastania *et al.*, 2020), and lung cancer patients in the Uzbek population (Mirakbarova and Turdikulova, 2021).

The study identified a synonymous SNP rs2228230 (c.2472C>T; p. Val824 Val), located on exon 18 of the Platelet-derived growth factor receptorα (*PDGFRA)* gene, a class III TK receptor. Platelet-derived growth factor (PDGF) is an angiogenesis-inducing cytokine that exerts its effects by interacting with receptors PDGFR. This cytokine and receptor system is crucial for the migration, proliferation, and recruitment of fibroblasts along with other cell types (Östman, 2017), also, implicated in wound healing and tissue regeneration and involved in survival, tumor growth, and metastasis when deregulated (Andrae *et al.*, 2008; Jing *et al.*, 2016). The platelet-derived growth factor is involved both in physiological processes, such as fibrosis, and in the pathogenesis of hematological and solid tumors. The polymorphism identified in the present study has been previously identified in AML patients with eosinophilia (Monma *et al.*, 2006; Trojani *et al.*, 2008). The polymorphism has been associated with Altered signaling via MAPK and PI3K/AKT pathways, decreased mRNA stability, predicted to affect splicing by the creation of ESS, predicted to affect mRNA secondary structure (Ben Jemii *et al.*, 2020; Dai *et al.*, 2020). The mutation has also been reported in several other cancers (Ben Jemii *et al.*, 2020; Dai *et al.*, 2020; Ye *et al.*, 2020; Mazurenko *et al.*, 2022). It has been associated with poor clinical outcomes in Head and Neck Squamous Cell Carcinoma (Butkiewicz *et al.*, 2022). On the contrary, a study linked rs2228230:T to reduced expression of PDGFRA and downstream signaling activity, which in turn is associated with good prognosis in Acral Melanoma patients (Dai *et al.*, 2020).

The current study also identified rs1800861 (c.2307G>T; Leu769Leu) located on exon 13 of the *RET* gene. The human gene Rearranged during transfection (RET) gene is localized on chromosome 10 (10q11.2) and contains 21 exons (Ceccherini *et al.*, 1993), RET is the receptor for members of the glial cell-derived neurotrophic factor (GDNF) family of ligands (GFLs): namely GDNF, Neurturin, Persephin, and Artemin (Baloh *et al.*, 2000). The RET proto-oncogene encodes a tyrosine kinase receptor (Arighi *et al.*, 2005; Ibáñez, 2013), and RET activation promotes downstream signaling, leading to cell proliferation, differentiation, and survival (Arighi *et al.*, 2005). The role of RET in cancers was first described when somatic rearrangements of RET (named RET/PTC) showing constitutive tyrosine kinase activity were found in papillary thyroid carcinomas (PTCs) (Grieco *et al.*, 1990). The *RET* gene plays a critical role in cell signaling, and mutations in the gene lead to the disruption of cell

differentiation of tissues (neural crest cells) and cell proliferation. The polymorphism is not reported in leukemia; however, it has been reported in case studies of Medullary Thyroid Carcinoma and is associated with reduced risk of the disease in the Italian population (Das *et al.*, 2017; Gemignani *et al.*, 2020) and craniopharyngioma (Jastania *et al.*, 2020).

The current study reported an SNP, rs41115 (c.4479G>A; p. Thr1493Thr) in the adenomatous polyposis coli protein isoform (*APC*) gene, located on chromosome 5(q22.2). Adenomatous polyposis coli (APC) is a tumor suppressor gene and one of the key players in the Wnt signaling pathway. The Wnt pathway has repeatedly been implicated in brain tumor genesis. Mutations in the APC gene are often insertions, frameshifts, or deletions that introduce premature stop codons and lead to the production of a truncated APC protein (Hankey *et al.*, 2018). The SNP has not reported leukemia before. The prognostic impact of APC tumor suppressor gene mutation was not well-established yet. The polymorphism is reported in Craniopharyngioma (Jastania *et al.*, 2020), Breast cancer (Chang *et al.*, 2016b), Colorectal cancer (Ashktorab *et al.*, 2017; Gerami *et al.*, 2020), Lynch syndrome (Wheeler *et al.*, 2016), Familial adenomatous polyposis (Ghadamyari *et al.*, 2021), and it is also predicted to alter splicing (Liu *et al.*, 2018; Ghadamyari *et al.*, 2021). The clinical status of the SNP is benign.

The current study identified the presence of 2 genetic variations in Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform *(PIK3CA),* including an intronic; c.352+40A>G rs3729674 and a synonymous, exonic variation c.3075C>T/p.T1025T/rs17849079, located on exon 21 of the *PIK3CA* gene. The c.352+40A>G rs3729674 is a known variant, reported in anaplastic ependymoma, and thyroid nodules, however, rs17849079 is reported in colorectal cancer, along with thyroid nodules (Borowczyk *et al.*, 2019; Butt *et al.*, 2019; Shetty *et al.*, 2022). *PIK3CA* is one of the most studied and targeted oncogenes, having a role in a broad range of cellular processes, including cell growth, proliferation, and apoptosis. Though the variation found in the present study has not been directly linked to leukemia before, the role of *PIK3CA* has been implicated in various cancers (Fruman and Rommel, 2014; Arafeh and Samuels, 2019).

The current study identified the presence of a single nucleotide variation, c.2361G>A/ rs1050171 located on exon 20 of epidermal growth factor receptor (*EGFR*) and its long

noncoding RNA *EGFR-AS1*. The *EGFR* gene is positioned in chromosome 7p11.2 and consists of thirty-two exons that encode the epidermal growth factor receptor of 1186 amino acid residues. EGFR is a critical transmembrane glycoprotein that acts as a receptor for the members of the EGF family (Herbst, 2004). It is one of the highly critical polymorphic loci that has been demonstrated to be correlated with various dysfunctions throughout the body (Sigismund et al., 2018). The *EGFR* gene is involved as a causative factor in various types of cancers (Zhang *et al.*, 2013; Wu *et al.*, 2019b). Recently EGFR-AS1 has been claimed to play a crucial role in cancer progression (Dong *et al.*, 2019; Wang *et al.*, 2020). This particular mutation is associated with a better response to anti-EGFR therapy in patients with metastatic cancers (Bonin *et al.*, 2016), glioma susceptibility (Baek *et al.*, 2022), and lung cancer (Lawi *et al.*, 2022). However, there is no study available associating this variation with leukemia.

The current also identified the presence of synonymous exonic polymorphism, c.534C>T/ p. Ser178Ser/ rs35775721, located on exon 2 of the mesenchymal-epithelial transition factor (MET) gene. The MET gene, located on chromosome 7, consists of 21 exons separated by 20 introns, coding for a receptor tyrosine kinase, with hepatocyte growth factor (HGF) as its natural ligand. Several studies suggest that activation of the MET signaling pathway mediates a diversity of biological functions, for instance, cell proliferation, survival, invasion, and tumor metastasis (Trusolino et al., 2010). The *MET* activation is involved in oncogenesis in many human cancers. It plays a critical role in the development and progression of primary tumors and metastases. There is very minute literature available on rs35775721, however, the polymorphism is also detected in Craniopharyngioma (Jastania et al., 2020).

All of the variations identified in the present study, either novel, reported, or reported in leukemia or any other disease, are structurally located on genes, that are either tumor suppressors, proto-oncogenes, oncogenes, or genes performing in crucial pathways, controlling growth, cell division, apoptosis, cell differentiation, and other important biological processes. The variation affects cells in both direct and indirect manner.

4.2.3 Cytogenetic analysis through microarray assay

Cytogenetic aberrations could play a critical role in the process of tumorigenesis, through activation/gain of oncogenes or inactivation/deletion of tumor suppressor genes coming within its jurisdiction. One of the important mechanisms involved in leukemogenesis is gene amplification achieved through copy number variations. A better understanding of genomic variations in leukemia, through modern technology like microarray, is important for more sensitive and precise classification of its subtypes (Usvasalo et al., 2010).

The present study identified the copy number alteration in the genome of a male leukemia patient, which included two copy number loss alterations at chromosome 8 (p11.22, q24.23), flanked with LOH segments. One of the common loci associated with the aggressive phenotype of cancer is the LOH segments positioned at the short arm of chromosome 8 (8p) (Cai *et al.*, 2016; Moelans *et al.*, 2018). The 8p11-12 segment is a frequent breakpoint region, distinguished for higher affinity for allelic loss and lower for amplifications/copy number gains towards 8pter. The gathering of genomic anomalies in this region can play a peculiar role in oncogenesis, even if the LOH and gene amplifications are absent.

The impact of copy number change in 8q24.3 could be implemented in the form of altered transcriptional programming or through triggering certain genomic shifts like inversions or translocations that might hold the potential to favor tumor growth or induce resistance. Though there is no literature available on the association of ch8;q24.23 to leukemia; however, CNVs detection in genes positioned at 8q24.3 can affect patient treatment outcomes and can prove a helpful predictor for cancer prognosis (Brusselaers et al., 2019).

The current study identifies a copy number gain (CNV; 3) alteration at chromosome 14 (q32.33), in share with LOH in the exactly same locus in male leukemia patients, with the presence of an OMIM gene, *FAM30A* (616623). Rearrangements at 14q32.33, involving specific genes, acquire unique biological attributes and have correlated to have allegiance with clinical, immunophenotypic, and morphological characteristics. The presence of a cancer-associated oncogene at 14q32.33 is deemed critical in the process of leukemogenesis (Chapiro

et al., 2010). Alterations in this position are rare in leukemia, however occurring more in adults than children (Chapiro et al., 2010; Woo et al., 2014).

Though the segment gained in the present study do not carry IGH, 14q32.33 rearrangements appear to be connected to good clinical outcome in leukemia when involving IGH (Othman et al., 2016). There is no evidence of this segment gaining a direct association with leukemia, however, loss/cryptic deletion of 14q32 has been linked to leukocytosis and shorter overall survival in leukemia patients (Forero-Castro et al., 2016). The upregulation of miRNA clusters, located on the 14q32.33, may influence the expressional pattern of the genes mediating critical signaling pathways such as cellular differentiation, apoptosis, and cell division due to the gain of copy number. The loss of heterozygosity on the 14q32/miRNA cluster may be an alternative mechanism employed for the transformation of lymphoid cells, and therefore it can be potentially used as a marker for diagnostic and chemotherapeutic targets in leukemia (Agueli et al., 2010).

Cancer cells do context-dependent selection for copy number loss in some settings and amplifications in others. The presence of cytogenetic aberrations accountable for leukemogenesis may be one of the causes that lead to treatment failure (Kuchinskaya et al., 2008). The genes being part of this role shift could be oncogenes or tumor suppressors in nature, thus complicating the development of new strategies for treatment. Therefore, one of the current challenges is to identify hidden genomic lesions that may affect the patient outcome. Studies that could explain whether copy number loss/gain is directly related, or associated with targettable genetic/epigenetic changes are required to design to more precise treatment.

4.3 *In silico* drug-protein interaction

The present study was conducted to test and suggest the interaction of chemical entities/drugs having the potential inhibitors STAT5B signaling, either in clinical trials or currently on the market, with the normal and mutated (N642H+) STAT5B protein. Direct pharmaceutical inhibitors that target STAT5B directly are currently in different stages of drug development (Nam *et al.*, 2012; Nelson *et al.*, 2012). The well-established role of STAT5B in leukemia, particularly regarding aggression and drug resistance emphasizes the need for specific compounds targeting STAT5B. Structure-based drug design and drug screening and focusing on targeting highly active domains, like the Src homology domain 2 (SH2) domain, can lead to the discovery/identification of worthwhile effective STAT5B inhibitors. Several STAT5 B-specific inhibitors are being identified by such approaches, in cell lines or preclinical studies. However, there are very limited details on the target specificity and selectivity of these ligands.

The present study observed that AC-4-130 was strongly docking with STAT5B. this salicylic acid-based compound could inhibit STAT5 dimerization in AML cells, resulting in decreased STAT5 phosphorylation as well as nuclear translocation, leading to inhibit clonogenic growth and proliferation in both primary AML patient cells and AML cell lines (Wingelhofer *et al.*, 2018). Pro-B murine cell line, Ba/F3 cells have been reported to be sensitive, while Ba/F3 STAT5B^{N642H+} cells were resistant towards AC-4–130 (Wingelhofer et al., 2018). The inhibitory effects of AC-4–130, were synergized and introduced in combination with the Ruxolitinib (JAK1/2 inhibitor), hence suggested as an effective approach to target AML cell growth and survival (Wingelhofer et al., 2018).

The present study also reports the docking efficiency of pimozide, FDA approved neuroleptic drug, for STAT5B and STAT5B^{N642H}. Pimozide has been previously screened as a potent inhibitor of STAT5 phosphorylation, resulting in the expressional downregulation of its downstream target genes in hematopoietic cell lines (Nelson *et al.*, 2011; Walker *et al.*, 2014). Pimozide induced growth arrest and decreased the tyrosine phosphorylation of STAT5 apoptosis in CML cells (Mistry *et al.*, 2013).

Pimozide is also reported to be effective in imatinib-resistant Bcr-Abl^{+} CML cells, with increased efficacy when used in combination with tyrosine kinase inhibitors (Nelson *et al.*, 2011). Pimozide is reported to manage resistance that Flt3 confers to AML cells (Nogami *et al.*, 2015). Pimozide is claimed to be effective against aggressive heterogenous non-Hodgkin's lymphoma by inhibiting STAT5 (Simpson *et al.*, 2018). Hence, targeting leukemia dependent on aberrant STAT5 signaling, with pimozide could prove to be an effective strategy. Despite the promising data on hematological malignancies, the mechanism of action of pimozide is still not clearly understood.

Stafib-2 is another small molecule inhibitor, that posed good efficiency in docking to STAT5B in the present study. Stafib-2 is a bisphosphate-containing small molecule, derived from Stafib-1. Stafib-2, similar to its predecessor, has shown high specificity for STAT5B (Elumalai *et al.*, 2017). These small molecules were, though not effective enough for clinical translation.

The present study observed a stable interaction, through H bonding, between the indirubin and STAT5B/STAT5B^{N642H}, with a slightly higher affinity for the mutant protein. Indirubin derivatives are previously been reported to impede STAT5 phosphorylation in CML (cell line), but the inhibitory mechanism was suggested to be the suppression of upstream tyrosine kinases (Nam *et al.*, 2012).

The rapidly evolving understanding of cancer has led to the discovery of new approaches to cancer therapy that directly target cancer cells, which is frequently referred to as precision medicine. Ideally, the use of precision medicines is contingent on the identification of specific changes in the cancer cell genome to ensure that the individual to whom the drug is prescribed is likely to benefit. The key to these approaches is finding genomic or proteomic features that are unique to the cancer cell that can be used to target the therapy to the cancer cell. These features are often proteins, playing a vital role in key cellular pathways, being expressed at higher levels than in normal cells, making them easy to recognize.

Currently, active research is happening on the use of tyrosine kinase inhibitors for treating aberrated STAT5 B-involved cancers, even though tyrosine kinase inhibitors are significantly superior to classical chemotherapy, however; their application is known to induce resistance and long-lasting side effects. Therefore, the development of more specific and effective inhibitors that, narrow down the window and shut down only the disrupted gene/protein is highly desirable to overcome the limitations of current strategies.

Finding the potential therapeutic window for leukemic cells, without killing the normal cells, for direct targeting of STAT5 signaling is one of the greatest challenges in the treatment of leukemia, where STAT5B mutations are involved. Optimization of STAT5 inhibitors by chemical modifications and highly specific, selective, and sophisticated drug delivery systems both required targeted delivery to tumor cells. Developments and improvements in targeted drug delivery, e.g., antibody-conjugated nanoparticles, might be advantageous in delivering the agent of choice to the target of choice.

We envisage that with added in vitro examination, other biophysical & structural studies on the inhibitory potential of these ligands on selective $STAT5B^{N642H+}$ inhibition could lead the way to the use of these compounds, individually or in combination for individualized compassionate use in the future for more specific inhibition.

Strengths and limitations of the study

To the best of our knowledge, this is the first descriptive case-control study, presenting a comprehensive account of demographic, clinical, lifestyle-related, and environmental risk factors associated with leukemia in the Pakistani population and measuring the magnitude of the risk. Previously, epidemiologic studies on the disease, have been executed to elaborate on common characteristics and manifestations of leukemia while the present study was designed and conducted to explore detailed spectra of the risk factors to contribute to the already known etiology of leukemia. The study was designed very carefully to include the maximum aspects of leukemia, e.g., epidemiology, genetics, cytogenetics, and drug prediction. Aberrated signaling pathways play the most important role in leukemia initiation and progression. The sequence of 50 crucial genes has been scanned through NGS to find any mutation/polymorphisms, and microarray technology has been employed to identify any variations in the patients' genomes. The *in silico* analysis docked and compared all of the main inhibitor groups, currently the focus of targeted pathway inhibition of the pioneer research groups in the field.

However, some confines and concerns in our study should be considered, e.g., short time frame, small sample size from selected regions, and the possibility of misclassification of race/ethnicity due to unawareness of the responders. Data regarding the quality/quantity of perfumes, carbonated drinks used per day, and tobacco exposure per day were not obtained. Some of the leukemia subtypes e.g., CLL were underrepresented in genetics analysis. The risk factor association study would have been stronger if an equal number of leukemia subtypes had been included because each leukemia subtype is unique in its pathogenesis and genetics.

Only a limited number of patient samples were analyzed through the high-throughput, modern genetic techniques employed in the study due to the scarcity of resources. The study would have produced highly meaningful and clearer results if more samples were processed through modern genetic techniques. Including translational research with basic research would have been more beneficial for clinicians and patients.

Conclusion and future recommendations

The present study signifies a preliminary investigation of a wide range of leukemia-associated risk factors in Pakistan. The study observed that leukemia is most common in male children, with acute lymphoblast being the most frequent subtype. Poor diet, wood use, drinking unprocessed ground water, rural residential setups, and lower education are significant factors increasing leukemia risk. Leukocytosis, anemia, thrombocytopenia, and neutropenia were significant, recurrent, paraclinical observations in leukemia patients. Children, AYAs, and adults exhibit different patterns of exposure to the identified risk factors.

A comprehensive evaluation of the risk factors that could be contributing to leukemia was the basic reason and essence of the study. Incorporating new tools and complex study designs with high statistical powers into traditional epidemiological approaches will be more helpful in the elucidation of impacting factors. An extensive study with an extended time frame and targeting country-wide populations will be more helpful in studying leukemia etiology, comprehensive characterization, and contribution of the involved risk factors in the Pakistani population.

At present, it is much more important for a country like Pakistan, which is already going through an economic crunch, to identify crucial risk factors and implement preventive strategies to address them, which will ease the financial burden, the unstable economy of the country must bear to deal with alarmingly increasing cancer incidence. Impacting influences on leukemia epidemiology should be pursued with large prospective studies.

The study identified the presence of *STAT5B*^{N64H} mutation in pediatric leukemia. The gain of function, STAT5B^{N642H} mutation in leukemia is an indicator of more aggressive and resistant leukemia and, is responsible for drug resistance and poor prognosis. The presence of this mutation in pediatric leukemia in the studied population indicates the aggressive nature of leukemia in the Pakistani population.

Our data extend and strengthen the spectrum of detrimental mutations/polymorphisms, either previously identified or novel employing a targeted sequencing approach for the diagnosis of leukemia. The study identified missense mutations/alterations in genes e.g., tp53, playing a role in key signal transduction pathways in two acute pediatric leukemia patients through NGS. The present study also highlights the usefulness of panel sequencing in cases where the prognosis becomes challenging. Targeted NGS is both a time and cost-efficient platform, which can be used to perform sensitive, reproducible, reliable, and simultaneous detection of multiple mutations hotspot genes in a go. Thus, targeted NGS at diagnosis may provide useful information. NGS is valuable for identifying individual genomic profiles and clustering the patients for targeted therapies. NGS techniques are valuable for identifying individual genomic profiles and clustering the patients for targeted therapies.

According to the Molecular Analysis for Therapy Choice (MATCH) Program conducted by the U.S. National Cancer Institute, the choice of a therapeutic agent is based on the specific molecular findings obtained using targeted NGS analysis rather than on the type of cancer. NGS should not be restricted to specific patients because many of the non-validated molecular alterations susceptible to targeted therapies have low prevalence. Personalized medicine in cancer has major challenges. Clinical care data should be collected and implemented in specific databases to report all situations with or without obvious clinical impact and to make molecular data meaningful for direct patients' benefit. Despite all its usefulness, However, the implementation of NGS in clinical laboratories still poses specific challenges. In parallel, molecular tumor boards are needed to supervise treatment decisions and to support off-labeled treatments because not all patients have access to clinical trials.

Microarray platforms provide high-resolution detection of genome-wide LOHs and CNAs, which are crucial for understanding the molecular mechanisms behind carcinogenesis and can lead to the identification of genetic markers that predict clinical outcomes. The present study reported copy number alterations, e.g., loss at chromosome 8 and gain at chromosome 14, along with 97 loss of heterozygosity regions, among which 8 were present on the same chromosome, and 3 were in close vicinity of the copy number variations. The comprehensive genomic analysis during diagnosis and treatment can provide new and relevant information

about how these aberrations could manipulate the course of treatment response and relapse risk. Microarray platforms provide high-resolution detection of genome-wide LOHs and CNAs, which are crucial for understanding the molecular mechanisms behind carcinogenesis and can lead to the identification of genetic markers that predict clinical outcomes. The comprehensive genomic analysis during diagnosis and treatment can provide new and relevant information about how these aberrations could manipulate the course of treatment response and relapse risk.

The design and development of STAT5 inhibitors are evolving. Approaches led to the identification of selective STAT5 inhibitors with potent anticancer agents are in dire need. The present study was carried out by utilizing the in silico molecular docking tools to identify the affinity of previously recorded ligands to inhibit STAT5B signaling in leukemia, with normal and mutated STAT5B. The study predicted that Stafib 2 and AC-4-130 have the potential to bind to STAT5B product, and hence can be predicted for potential inhibition. However, the above study's findings are based only on computer simulations, validation of which with an adequately designed experimental protocol is necessary. Finally, the optimization of combination therapies using STAT5 inhibitors with molecules targeting tyrosine kinases or other key players in cancer will be required to find the right combination that safely unlocks drug resistance in hematologic cancers. *In silico* prediction of the impact of potential inhibitors on an aberrated gene product, before the start time and money investment could be a promising savior

REFERENCES

Abelson, S., Collord, G., Ng, S. W., Weissbrod, O., Cohen, N. M., Niemeyer, E., Barda, N., Zuzarte, P. C., Heisler, L., Sundaravadanam, Y. (2018). Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature*. **559**: 400-404.

Acharya, S., Hsieh, S., Shinohara, E. T., Dewees, T., Frangoul, H., Perkins, S. M. (2016). Effects of race/ethnicity and socioeconomic status on outcome in childhood acute lymphoblastic leukemia. *Journal of pediatric hematology/oncology*. **38**: 350-354.

Afzal, M. S. (2020). Childhood Cancer in Pakistan. *Iranian journal of public health*. Aug; **49(8)**: 1579.

Agueli, C., Cammarata, G., Salemi, D., Dagnino, L., Nicoletti, R., La Rosa, M., Messana, F., Marfia, A., Bica, M. G., Coniglio, M. L. (2010). 14q32/miRNA clusters loss of heterozygosity in acute lymphoblastic leukemia is associated with up-regulation of BCL11a. *American journal of hematology*. **85**: 575-578.

Ahmad, S., Shah, K. A., Hussain, H., Haq, A. U., Ullah, A., Khan, A., Rahman, N. U. (2019). Prevalence of Acute and Chronic Forms of Leukemia in Various Regions of Khyber Pakhtunkhwa, Pakistan: Needs Much More to be done! *Bangladesh journal of medical science*. **18**: 222-227.

Ahmed, A. T., Yassin, A. K., Mohammed, N. S., Hasan, K. M. (2019). Acute promyelocytic leukemia: Epidemiology, clinical presentation, and outcome over a 10-year period of follow-up at Nanakali Hospital of Erbil city "Single-center study". *Iraqi journal of hematology*. **8**: 7.

Akada, H., Akada, S., Hutchison, R. E., Sakamoto, K., Wagner, K. U., Mohi, G. (2014). Critical role of Jak2 in the maintenance and function of adult hematopoietic stem cells. *Stem cells*. **32**: 1878-1889.

Akhtar, M. S., Alharbi, R. A. (2022). Genetic association of TP53 Pro72Arg polymorphism (rs1042522) in Leukemia: An updated meta-analysis of 10 case-control studies. *Human gene*. **34**: 201130.

Akram, A. M., Iqbal, Z., Akhtar, T., Khalid, A. M., Sabar, M. F., Qazi, M. H., Aziz, Z., Sajid, N., Aleem, A., Rasool, M. (2017). Presence of novel compound BCR-ABL mutations in late chronic and advanced phase imatinib sensitive CML patients indicates their possible role in CML progression. *Cancer biology & therapy*. **18**: 214-221.

Al-Abady, I. (2021). Clinico-Hematological Profile in Patients with Chronic Myeloid Leukemia. *Annals of the college of medicine, Mosul*. **43**: 1-9.

Al-Maktari, L. A., Al-Nuzaili, M. A., Al-Shamahy, H. A., Al-Hadi, A. A., Ishak, A. A., Bamashmoos, S. A. (2021). Distribution of hematological parameters counts for children with leukemia in children's cancer units at Al-Kuwait Hospital, Sana'a City: A Cross-Sectional Study. *Advance cancer reseatch clinical image*. **3**: 2021.

Al-Shehab, M. A., Alhadi, A. M., Elnemr, M. a. M., Al-Qadasi, F. A., Alabsi, N. A., Aqlan, M. A. (2020). Epidemiological, Clinical and Paraclinical Characteristics of Childhood Acute Lymphoblastic Leukaemia in Sana'a, Yemen. *Journal of clinical & diagnostic research*. **14**.

Ali, A., Gale, R. E., Shakoori, A. R. (2017). Detection of FLT3/TKD and IDH1 mutations in Pakistani acute myeloid leukemia patients by denaturing HPLC. *Journal of cellular biochemistry*. **118**: 1174-1181.

Alnajjar, R., Mostafa, A., Kandeil, A., Al-Karmalawy, A. A. (2020). Molecular docking, molecular dynamics, and in vitro studies reveal the potential of angiotensin II receptor blockers to inhibit the COVID-19 main protease. *Heliyon*. **6**: e05641.

Altieri, A., Castro, F., Bermejo, J. L., Hemminki, K. (2006). Number of siblings and the risk of lymphoma, leukemia, and myeloma by histopathology. *Cancer epidemiology and prevention biomarkers*. **15**: 1281-1286.

Amankwah, E. K., Saenz, A. M., Hale, G. A., Brown, P. A. (2016). Association between body mass index at diagnosis and pediatric leukemia mortality and relapse: a systematic review and meta-analysis. *Leukemia & lymphoma*. **57**: 1140-1148.

American Cancer Society, (2016). Cancer Facts & Figures 2016. U.S: American Cancer Society.

American Cancer Society, (2019). Cancer Facts & Figures 2019. U.S: American Cancer Society.

Amitay, E. L., Keinan-Boker, L. (2015). Breastfeeding and childhood leukemia incidence: a meta-analysis and systematic review. *JAMA pediatrics*. **169**: e151025-e151025.

Andrae, J., Gallini, R., Betsholtz, C. (2008). Role of platelet-derived growth factors in physiology and medicine. *Genes & development*. **22**: 1276-1312.

Aoki, Y., Feldman, G. M., Tosato, G. (2003). Inhibition of STAT3 signaling induces apoptosis and decreases survivin expression in primary effusion lymphoma. *Blood, The journal of the american society of hematology*. **101**: 1535-1542.

Arafeh, R., Samuels, Y., (2019). PIK3CA in cancer: the past 30 years. *Seminars in cancer biology: Elsevier*. pp. 36-49.

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

Argetsinger, L. S., Campbell, G. S., Yang, X., Witthuhn, B. A., Silvennoinen, O., Ihle, J. N., Carter-Su, C. (1993). Identification of JAK2 as a growth hormone receptor-associated tyrosine kinase. *Cell*. **74**: 237-244.

Arighi, E., Borrello, M. G., Sariola, H. (2005). RET tyrosine kinase signaling in development and cancer. *Cytokine & growth factor reviews*. **16**: 441-467.

Ashktorab, H., Mokarram, P., Azimi, H., Olumi, H., Varma, S., Nickerson, M. L., Brim, H. (2017). Targeted exome sequencing reveals distinct pathogenic variants in Iranians with colorectal cancer. *Oncotarget*. **8**: 7852.

Aslam, R., Raza, S. M., Naeemi, H., Mubarak, B., Afzal, N., Khaliq, S. (2016). SOCS3 mRNA expression and polymorphisms as pretreatment predictor of response to HCV genotype 3a IFN-based treatment. *Springerplus*. **5**: 1-9.

Avenbuan, O. N., Zelikoff, J. T. (2020). Woodsmoke and emerging issues. *Current opinion in toxicology*. **22**: 12-18.

Awan, T., Iqbal, Z., Aleem, A., Sabir, N., Absar, M., Rasool, M., Tahir, A. H., Basit, S., Khalid, A. M., Sabar, M. F. (2012). Five most common prognostically important fusion oncogenes are detected in the majority of Pakistani pediatric acute lymphoblastic leukemia patients and are strongly associated with disease biology and treatment outcome. *Asian pacific journal of cancer prevention*. **13**.

Ayub, A., Ahmad, Q. M., Javed, T., Hayat, M. Z., Farooq, M. A., Anwar, H. M. Z., Khan, M. A. (2020). Evaluation of diet as a risk factor in the development of childhood leukaemia: a case control study. *JPMA*. **2019**.

Babushok, D. V., Perdigones, N., Perin, J. C., Olson, T. S., Ye, W., Roth, J. J., Lind, C., Cattier, C., Li, Y., Hartung, H. (2015). Emergence of clonal hematopoiesis in the majority of patients with acquired aplastic anemia. *Cancer genetics*. **208**: 115-128.

Badar, F., Mahmood, S., Yusuf, M. A., Sultan, F. (2016). Epidemiology of cancers in Lahore, Pakistan, 2010–2012: a cross-sectional study. *BMJ Open*. **6**: e011828.

Baek, I. K., Cheong, H. S., Namgoong, S., Kim, J.-H., Kang, S.-G., Yoon, S.-J., Kim, S. H., Chang, J. H., Kim, L. H., Shin, H. D. (2022). Two independent variants of epidermal growth factor receptor associated with risk of glioma in a Korean population. *Scientific reports*. **12**: 19014.

Bailey, H. D., De Klerk, N. H., Fritschi, L., Attia, J., Daubenton, J. D., Armstrong, B. K., Milne, E., Consortium, A. A. (2011). Refuelling of vehicles, the use of wood burners and the risk of acute lymphoblastic leukaemia in childhood. *Paediatric and perinatal epidemiology*. **25**: 528-539.

Baloh, R. H., Enomoto, H., Johnson Jr, E. M., Milbrandt, J. (2000). The GDNF family ligands and receptors—implications for neural development. *Current opinion in neurobiology*. **10**: 103-110.

Bandapalli, O. R., Schuessele, S., Kunz, J. B., Rausch, T., Stütz, A. M., Tal, N., Geron, I., Gershman, N., Izraeli, S., Eilers, J. (2014). The activating STAT5B N642H mutation is a common abnormality in pediatric T-cell acute lymphoblastic leukemia and confers a higher risk of relapse. *Haematologica*. **99**: e188.

Bannon, S. A., Foglesong, J., Dinardo, C. D. (2017). Germline mutations associated with leukemia in childhood: new discoveries and emerging phenotypes. *Current genetic medicine reports*. **5**: 59-65.

Barr, R. D., Gomez-Almaguer, D., Jaime-Perez, J. C., Ruiz-Argüelles, G. J. (2016). Importance of nutrition in the treatment of leukemia in children and adolescents. *Archives of medical research*. **47**: 585-592.

Barrera, M., Shaw, A. K., Speechley, K. N., Maunsell, E., Pogany, L. (2005). Educational and social late effects of childhood cancer and related clinical, personal, and familial characteristics. *Cancer*. **104**: 1751-1760.

Barrows, N., Campos, R., Powell, S., Prasanth, K., Schott-Lerner, G., Soto-Acosta, R. (2016). A screen of FDA-approved drugs for inhibitors of Zika virus infection. *Cell host microbe*. 259-270.

Basabaeen, A. A., Abdelgader, E. A., Babekir, E. A., Eltayeb, N. H., Altayeb, O. A., Fadul, E. A., Bahashwan, O. S., Ibrahim, I. K. (2019a). Clinical presentation and hematological profile among young and old chronic lymphocytic leukemia patients in Sudan. *BMC research notes*. **12**: 1-6.

Basabaeen, A. a. M., Abdelgader, E. A., Babekir, E. A., Abdelrahim, S. O., Eltayeb, N. H., Altayeb, O. A., Fadul, E. A., Sabo, A., Ibrahim, I. K. (2019b). TP53 gene 72 arg/pro (rs1042522) single nucleotide polymorphism contribute to increase the risk of b-

chronic lymphocytic leukemia in the Sudanese population. *Asian pacific journal of cancer prevention: APJCP*. **20**: 1579.

Baxter, E. J., Scott, L. M., Campbell, P. J., East, C., Fourouclas, N., Swanton, S., Vassiliou, G. S., Bench, A. J., Boyd, E. M., Curtin, N. (2005). Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *The lancet*. **365**: 1054-1061.

Bejjani, B. A., Shaffer, L. G. (2008). Clinical utility of contemporary molecular cytogenetics. *Annual review of genomics and human genetics.* **9**: 71-86.

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

Belurkar, S., Mantravadi, H., Manohar, C., Kurien, A. (2013). Correlation of morphologic and cytochemical diagnosis with flowcytometric analysis in acute leukemia. *Journal of cancer research and therapeutics*. **9**: 71.

Ben Jemii, N., Tounsi-Kettiti, H., Yaiche, H., Mezghanni, N., Jaballah Gabteni, A., Fehri, E., Ben Fayala, C., Abdelhak, S., Boubaker, S. (2020). Dysregulated PDGFR alpha expression and novel somatic mutations in colorectal cancer: association to RAS wild type status and tumor size. *Journal of translational medicine*. **18**: 1-20.

Bener, A., El Ayoubi, H. R., Chouchane, L., Ali, A. I., Al-Kubaisi, A., Al-Sulaiti, H., Teebi, A. S. (2009). Impact of consanguinity on cancer in a highly endogamous population. *Asian pacific journal of cancer prevention*. **10**: 35-40.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R., Sultan, C. (1976). Proposals for the classification of the acute leukaemias French-American-British (FAB) co-operative group. *British journal of haematology*. **33**: 451-458.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H. R., Sultan, C. (1985). Proposed revised criteria for the classification of acute myeloid

leukemia: a report of the French-American-British Cooperative Group. *Annals of internal medicine*. **103**: 620-625.

Bergmann, A. K., Schneppenheim, S., Seifert, M., Betts, M. J., Haake, A., Lopez, C., Maria Murga Penas, E., Vater, I., Jayne, S., Dyer, M. J. (2014). Recurrent mutation of JAK3 in T-cell prolymphocytic leukemia. *Genes, chromosomes and cancer*. **53**: 309-316.

Bernardo, W., Simões, R. D. S., Buzzini, R., Nunes, V., Glina, F. (2016). Adverse effects of the consumption of artificial sweeteners-systematic review. *Revista da associação médica brasileira*. **62**: 120-122.

Bewicke-Copley, F., Kumar, E. A., Palladino, G., Korfi, K., Wang, J. (2019). Applications and analysis of targeted genomic sequencing in cancer studies. *Computational and structural biotechnology journal*. **17**: 1348-1359.

Bhatia, S., Neglia, J. P. (1995). Epidemiology of childhood acute myelogenous leukemia. *Journal of pediatric hematology/oncology*. **17**: 94-100.

Bhattacharya, K., Bordoloi, R., Chanu, N. R., Kalita, R., Sahariah, B. J., Bhattacharjee, A. (2022). In silico discovery of 3 novel quercetin derivatives against papain-like protease, spike protein, and 3C-like protease of SARS-CoV-2. *Journal of genetic engineering and biotechnology*. **20**: 1-20.

Bhurgri, Y., Bhurgri, A., Hassan, S. H., Zaidi, S., Rahim, A., Sankaranarayanan, R., Parkin, D. M. (2000). Cancer incidence in Karachi, Pakistan: first results from Karachi cancer registry. *International journal of cancer*. **85**: 325-329.

Biovia, D. S. (2021). BIOVIA Discovery Studio Academic Research Suite. *San diego: dassault systèmes*.

Bispo, J. a. B., Pinheiro, P. S., Kobetz, E. K. (2020). Epidemiology and etiology of leukemia and lymphoma. *Cold spring harbor perspectives in medicine*. **10**: a034819.

Black, J. S., Salto-Tellez, M., Mills, K. I., Catherwood, M. A. (2015). The impact of next generation sequencing technologies on haematological research–a review. *Pathogenesis*. **2**: 9-16.

Bonin, S., Donada, M., Bussolati, G., Nardon, E., Annaratone, L., Pichler, M., Chiaravalli, A. M., Capella, C., Hoefler, G., Stanta, G. (2016). A synonymous EGFR polymorphism predicting responsiveness to anti-EGFR therapy in metastatic colorectal cancer patients. *Tumor biology*. **37**: 7295-7303.

Bonner, M. R., Williams, B. A., Rusiecki, J. A., Blair, A., Freeman, L. E. B., Hoppin, J. A., Dosemeci, M., Lubin, J., Sandler, D. P., Alavanja, M. C. (2010). Occupational exposure to terbufos and the incidence of cancer in the Agricultural Health Study. *Cancer causes & control*. **21**: 871-877.

Borowczyk, M., Szczepanek-Parulska, E., Dębicki, S., Budny, B., Verburg, F. A., Filipowicz, D., Wrotkowska, E., Janicka-Jedyńska, M., Więckowska, B., Gil, L. (2019). Genetic heterogeneity of indeterminate thyroid nodules assessed preoperatively with next-generation sequencing reflects the diversity of the final histopathologic diagnosis. *Polish archives of internal medicine.* **129**: 761-769.

Bourgeais, J., Ishac, N., Medrzycki, M., Brachet-Botineau, M., Desbourdes, L., Gouilleux-Gruart, V., Pecnard, E., Rouleux-Bonnin, F., Gyan, E., Domenech, J. (2017). Oncogenic STAT5 signaling promotes oxidative stress in chronic myeloid leukemia cells by repressing antioxidant defenses. *Oncotarget*. **8**: 41876.

Bousoik, E., Montazeri Aliabadi, H. (2018). "Do we know jack" about JAK? A closer look at JAK/STAT signaling pathway. *Frontiers in oncology*. **8**: 287.

Brachet-Botineau, M., Deynoux, M., Vallet, N., Polomski, M., Juen, L., Hérault, O., Mazurier, F., Viaud-Massuard, M.-C., Prié, G., Gouilleux, F. (2019). A novel inhibitor of STAT5 signaling overcomes chemotherapy resistance in myeloid leukemia cells. *Cancers*. **11**: 2043.

Braggio, E., Egan, J. B., Fonseca, R., Stewart, A. K. (2013). Lessons from next-generation sequencing analysis in hematological malignancies. *Blood cancer journal*. **3**: e127-e127.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A cancer Journal For Clinicians*. **68**: 394-424.

Breen, K. A., Grimwade, D., Hunt, B. J. (2012). The pathogenesis and management of the coagulopathy of acute promyelocytic leukaemia. *British journal of haematology*. **156**: 24-36.

Brown, J. R. (2008). Inherited predisposition to chronic lymphocytic leukemia. *Expert review of hematology*. **1**: 51-61.

Brusselaers, N., Ekwall, K., Durand-Dubief, M. (2019). Copy number of 8q24. 3 drives HSF1 expression and patient outcome in cancer: an individual patient data meta-analysis. *Human genomics*. **13**: 1-12.

Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., Rodnina, M. V., Komar, A. A. (2016). Synonymous codons direct cotranslational folding toward different protein conformations. *Molecular cell*. **61**: 341-351.

Burns, A., Alsolami, R., Becq, J., Stamatopoulos, B., Timbs, A., Bruce, D., Robbe, P., Vavoulis, D., Clifford, R., Cabes, M. (2019). Correction: Whole-genome sequencing of chronic lymphocytic leukaemia reveals distinct differences in the mutational landscape between IgHV mut and IgHV unmut subgroups. *Leukemia*. **33**: 2342-2342.

Butkiewicz, D., Gdowicz-Kłosok, A., Krześniak, M., Rutkowski, T., Łasut-Szyszka, B., Składowski, K. (2022). Germline Variants in Angiogenesis-Related Genes Contribute to Clinical Outcome in Head and Neck Squamous Cell Carcinoma. *Cancers*. **14**: 1844.

Butt, E., Alyami, S., Nageeti, T., Saeed, M., Alquthami, K., Bouazzaoui, A., Athar, M., Abduljaleel, Z., Al-Allaf, F., Taher, M. (2019). Mutation profiling of anaplastic ependymoma grade III by Ion Proton next generation DNA sequencing. *F1000 research*. **8**.

Cai, Y., Crowther, J., Pastor, T., Asbagh, L. A., Baietti, M. F., De Troyer, M., Vazquez, I., Talebi, A., Renzi, F., Dehairs, J. (2016). Loss of chromosome 8p governs tumor progression and drug response by altering lipid metabolism. *Cancer cell*. **29**: 751-766.

Calabretta, B., Perrotti, D. (2004). The biology of CML blast crisis. *Blood*. **103**: 4010-4022.

Carpenter, R. L., Lo, H.-W. (2014). STAT3 target genes relevant to human cancers. *Cancers*. **6**: 897-925.

Castillo, J. J., Reagan, J. L., Ingham, R. R., Furman, M., Dalia, S., Merhi, B., Nemr, S., Zarrabi, A., Mitri, J. (2012). Obesity but not overweight increases the incidence and mortality of leukemia in adults: a meta-analysis of prospective cohort studies. *Leukemia research*. **36**: 868-875.

Ceccherini, I., Bocciardi, R., Luo, Y., Pasini, B., Hofstra, R., Takahashi, M., Romeo, G. (1993). Exon structure and flanking intronic sequences of the human RET proto-oncogene. *Biochemical and biophysical research communications*. **196**: 1288-1295.

Chandran, R. K., Geetha, N., Sakthivel, K. M., Aswathy, C. G., Gopinath, P., Raj, T. V. A., Priya, G., Nair, J. K. K. M., Sreedharan, H. (2019). Genomic amplification of BCR-ABL1 fusion gene and its impact on the disease progression mechanism in patients with chronic myelogenous leukemia. *Gene*. **686**: 85-91.

Chang, F., Shamsi, T., Waryah, A. (2016a). Clinical and Hematological Profile of Acute Myeloid Leukemia (AML) Patients of Sindh. *Hematology and thromboembolic diseases.* 4: 239.

Chang, Y.-S., Lin, C.-Y., Yang, S.-F., Ho, C.-M., Chang, J.-G. (2016b). Analysing the mutational status of adenomatous polyposis coli (APC) gene in breast cancer. *Cancer cell international*. **16**: 1-6.

Chapiro, E., Russell, L., Struski, S., Cave, H., Radford-Weiss, I., Valle, V., Lachenaud, J., Brousset, P., Bernard, O., Harrison, C. (2010). A new recurrent translocation t (11; 14)(q24; q32) involving IGH@ and miR-125b-1 in B-cell progenitor acute lymphoblastic leukemia. *Leukemia*. **24**: 1362-1364.

Charalambous, A., Vasileiou, P. (2012). Risk factors for childhood leukemia: a comprehensive literature review. *Health science journal*. **6**: 432.

Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S. M., Savidge, T. C. (2016). Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Science advances*. **2**: e1501240.

Chihara, D., Ito, H., Matsuda, T., Shibata, A., Katsumi, A., Nakamura, S., Tomotaka, S., Morton, L. M., Weisenburger, D. D., Matsuo, K. (2014). Differences in incidence and trends of haematological malignancies in J apan and the U nited S tates. *British journal of haematology*. **164**: 536-545.

Chung, Y.-J., Park, B.-B., Kang, Y.-J., Kim, T.-M., Eaves, C. J., Oh, I.-H. (2006). Unique effects of Stat3 on the early phase of hematopoietic stem cell regeneration. *Blood*. **108**: 1208-1215.

Chunxia, D., Meifang, W., Jianhua, Z., Ruijuan, Z., Xiue, L., Zhuanzhen, Z., Linhua, Y. (2019). Tobacco smoke exposure and the risk of childhood acute lymphoblastic leukemia and acute myeloid leukemia: A meta-analysis. *Medicine*. **98**.

Coccaro, N., Anelli, L., Zagaria, A., Specchia, G., Albano, F. (2019). Next-generation sequencing in acute lymphoblastic leukemia. *International journal of molecular sciences*. **20**: 2929.

Cocco, P., T'mannetje, A., Fadda, D., Melis, M., Becker, N., De Sanjosé, S., Foretová, L., Mareckova, J., Staines, A., Kleefeld, S. (2010). Occupational exposure to solvents and risk of lymphoma subtypes: results from the Epilymph case–control study. *Occupational and environmental medicine*. **67**: 341-347.

Cohn, P., Klotz, J., Bove, F., Berkowitz, M., Fagliano, J. (1994). Drinking Water Contamination and the Incidence of Leukemia and Non-Hodgkin's Lymphoma. *Environmental health perspectives*. **102**: 556-561.

Colamesta, V., D'aguanno, S., Breccia, M., Bruffa, S., Cartoni, C., La Torre, G. (2016). Do the smoking intensity and duration, the years since quitting, the methodological quality and the year of publication of the studies affect the results of the meta-analysis on cigarette smoking and Acute Myeloid Leukemia (AML) in adults? *Critical reviews in oncology/hematology*. **99**: 376-388.

Contreras, Z. A., Hansen, J., Ritz, B., Olsen, J., Yu, F., Heck, J. E. (2017). Parental age and childhood cancer risk: A Danish population-based registry study. *Cancer epidemiology*. **49**: 202-215.

Corsello, S. M., Bittker, J. A., Liu, Z., Gould, J., Mccarren, P., Hirschman, J. E., Johnston, S. E., Vrcic, A., Wong, B., Khan, M. (2017). The Drug Repurposing Hub: a next-generation drug library and information resource. *Nature medicine*. **23**: 405-408.

Cortes, J. E., Jones, D., O'brien, S., Jabbour, E., Konopleva, M., Ferrajoli, A., Kadia, T., Borthakur, G., Stigliano, D., Shan, J. (2010). Nilotinib as front-line treatment for patients with chronic myeloid leukemia in early chronic phase. *Journal of clinical oncology*. **28**: 392.

Cross, N. C., Hoade, Y., Tapper, W. J., Carreno-Tarragona, G., Fanelli, T., Jawhar, M., Naumann, N., Pieniak, I., Lübke, J., Ali, S. (2019). Recurrent activating STAT5B N642H mutation in myeloid neoplasms with eosinophilia. *Leukemia*. **33**: 415-425.

Cumaraswamy, A. A., Lewis, A. M., Geletu, M., Todic, A., Diaz, D. B., Cheng, X. R., Brown, C. E., Laister, R. C., Muench, D., Kerman, K. (2014). Nanomolar-potency small molecule inhibitor of STAT5 protein. *ACS medicinal chemistry letters*. **5**: 1202-1206.

Dai, J., Yang, L., Xu, T., Si, L., Cui, C., Sheng, X., Chi, Z., Mao, L., Lian, B., Tang, B. (2020). A functional synonymous variant in PDGFRA is associated with better survival in acral melanoma. *Journal of cancer*. **11**: 2945.

Das, S., Kelly, D., Moran, B., Han, K., Mulligan, N., Barrett, C., Buckley, P. G., Mcmahon, P., Mccaffrey, J., Van Essen, H. F. (2017). Postmortem examination of an aggressive case of medullary thyroid carcinoma characterized by catastrophic genomic abnormalities. *JCO precision oncology*. **1**.

De Figueiredo Barros, B. D., Kupper, B. E., Aguiar Junior, S., De Mello, C. A., Begnami, M. D., Chojniak, R., De Souza, S. J., Torrezan, G. T., Carraro, D. M. (2018). Mutation detection in tumor-derived cell free DNA anticipates progression in a patient with metastatic colorectal cancer. *Frontiers in oncology*. **8**: 306.

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

De Paula Junior, M. R., Nonino, A., Nascimento, J. M., Bonadio, R. S., Pic-Taylor, A., De Oliveira, S. F., Pereira, R. W., Do Couto Mascarenhas, C., Mazzeu, J. F. (2018). High frequency of copy-neutral loss of heterozygosity in patients with myelofibrosis. *Cytogenetic and genome research*. **154**: 62-70.

Dehghani, M., Keshtgar, L., Javaheri, M. R., Derakhshan, Z., Oliveri Conti, G., Zuccarello, P., Ferrante, M. (2017). The effects of air pollutants on the mortality rate of lung cancer and leukemia. *Molecular medicine reports*. **15**: 3390-3397.

Deschler, B., Lübbert, M. (2006). Acute myeloid leukemia: epidemiology and etiology. *Cancer: interdisciplinary international journal of the american cancer society*. **107**: 2099-2107.

Dingli, D., Traulsen, A., Lenaerts, T., Pacheco, J. M. (2010). Evolutionary dynamics of chronic myeloid leukemia. *Genes & cancer*. **1**: 309-315.

Dodhy, M., Zafar, H., Aslam, W. (2011). Chronic lymphocytic leukemia: an experience of a decade at a tertiary care hospital. *Annals of pakistan institute of medical sciences*. **7**: 196-199.

Döhner, H., Weisdorf, D. J., Bloomfield, C. D. (2015). Acute myeloid leukemia. *New england journal of medicine*. **373**: 1136-1152.

Dong, H.-J., Fang, C., Wang, L., Fan, L., Xu, J., Wu, J.-Z., Lu, T.-X., Li, J.-Y., Xu, W. (2014). TP53 Pro72 allele potentially increases the poor prognostic significance of TP53 mutation in chronic lymphocytic leukemia. *Medical oncology*. **31**: 1-10.

Dong, Z.-Q., Guo, Z.-Y., Xie, J. (2019). The lncRNA EGFR-AS1 is linked to migration, invasion and apoptosis in glioma cells by targeting miR-133b/RACK1. *Biomedicine & pharmacotherapy*. **118**: 109292.

Dores, G. M., Anderson, W. F., Curtis, R. E., Landgren, O., Ostroumova, E., Bluhm, E. C., Rabkin, C. S., Devesa, S. S., Linet, M. S. (2007). Chronic lymphocytic leukaemia and small lymphocytic lymphoma: overview of the descriptive epidemiology. *British journal of haematology*. **139**: 809-819.

Drokow, E. K., Chen, Y., Waqas Ahmed, H. A., Oppong, T. B., Akpabla, G. S., Pei, Y., Kumah, M. A., Neku, E. A., Sun, K. (2020). The relationship between leukemia and TP53 gene codon Arg72Pro polymorphism: analysis in a multi-ethnic population. *Future oncology*. **16**: 923-937.

Dugoff, L., Norton, M. E., Kuller, J. A., Medicine, S. F. M.-F. (2016). The use of chromosomal microarray for prenatal diagnosis. *American journal of obstetrics and gynecology*. **215**: B2-B9.

Dumont, P., Leu, J., Della Pietra, A. C., George, D. L., Murphy, M. (2003). The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. *Nature genetics*. **33**: 357-365.

Dunbar, A., Nazir, A., Levine, R. (2017). Overview of transgenic mouse models of myeloproliferative neoplasms (MPNs). *Current protocols in pharmacology*. **77**: 14.40. 11-14.40. 19.

Duncavage, E., Tandon, B. (2015). The utility of next-generation sequencing in diagnosis and monitoring of acute myeloid leukemia and myelodysplastic syndromes. *International journal of laboratory hematology*. **37**: 115-121.

Dzierzak, E., Philipsen, S. (2013). Erythropoiesis: development and differentiation. *Cold Spring Harbor perspectives in medicine*. **3**: a011601.

Eberhardt, J., Santos-Martins, D., Tillack, A. F., Forli, S. (2021). AutoDock Vina 1.2. 0: New docking methods, expanded force field, and python bindings. *Journal of chemical information and modeling*. **61**: 3891-3898.

Ebrahim, H., Fisha, T., Debash, H., Bisetegn, H. (2022). Patterns of Bone Marrow Confirmed Malignant and Non-Malignant Hematological Disorders in Patients with Abnormal Hematological Parameters in Northeast Ethiopia. *Journal of blood medicine*. **13**: 51.

Ehret, G. B., Reichenbach, P., Schindler, U., Horvath, C. M., Fritz, S., Nabholz, M., Bucher, P. (2001). DNA Binding Specificity of Different STAT Proteins: COMPARISON OF IN VITRO SPECIFICITY WITH NATURAL TARGET SITES* 210. *Journal of biological chemistry*. **276**: 6675-6688.

Eliaa, S. G., Al-Karmalawy, A. A., Saleh, R. M., Elshal, M. F. (2020). Empagliflozin and doxorubicin synergistically inhibit the survival of triple-negative breast cancer cells via interfering with the mTOR pathway and inhibition of calmodulin: in vitro and molecular docking studies. *ACS pharmacology & translational science*. **3**: 1330-1338.

Elliott, E. G., Trinh, P., Ma, X., Leaderer, B. P., Ward, M. H., Deziel, N. C. (2017). Unconventional oil and gas development and risk of childhood leukemia: Assessing the evidence. *Science of the total environment*. **576**: 138-147.

Elumalai, N., Berg, A., Rubner, S., Blechschmidt, L., Song, C., Natarajan, K., Matysik, J., Berg, T. (2017). Rational development of Stafib-2: a selective, nanomolar inhibitor of the transcription factor STAT5b. *Scientific reports*. **7**: 1-9.

Fabbri, G., Rasi, S., Rossi, D., Trifonov, V., Khiabanian, H., Ma, J., Grunn, A., Fangazio, M., Capello, D., Monti, S. (2011). Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *Journal of experimental medicine*. **208**: 1389-1401.

Fadoo, Z., Nisar, I., Yousuf, F., Lakhani, L. S., Ashraf, S., Imam, U., Zaheer, J., Naqvi, A., Belgaumi, A. (2015). Clinical features and induction outcome of childhood acute lymphoblastic leukemia in a lower/middle income population: A multi-institutional report from Pakistan. *Pediatric blood & cancer*. **62**: 1700-1708.

Faiz, M., Qureshi, A., Qazi, J. I., (2011). Molecular characterization of different fusion oncogenes associated with childhood Acute Lymphoblastic leukaemia from Pakistan. *IJAVMS.* **5**: 497-507

Fasouli, E. S., Katsantoni, E. (2021). JAK-STAT in early hematopoiesis and leukemia. *Frontiers in cell and developmental biology*. **9**.

Feller, M., Adam, M., Zwahlen, M., Brazzola, P., Niggli, F., Kuehni, C., Group, S. P. O. (2010). Family characteristics as risk factors for childhood acute lymphoblastic leukemia: a population-based case-control study. *Plos one*. **5**: e13156.

Feng, J., Witthuhn, B. A., Matsuda, T., Kohlhuber, F., Kerr, I. M., Ihle, J. N. (1997). Activation of Jak2 catalytic activity requires phosphorylation of Y1007 in the kinase activation loop. *Molecular and cellular biology*. **17**: 2497-2501.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*. **136**: E359-E386.

Fiebelkorn, S., Meredith, C. (2018). Estimation of the leukemia risk in human populations exposed to benzene from tobacco smoke using epidemiological data. *Risk analysis*. **38**: 1490-1501.

Fitzmaurice, C., Allen, C., Barber, R. M., Barregard, L., Bhutta, Z. A., Brenner, H., Dicker, D. J., Chimed-Orchir, O., Dandona, R., Dandona, L. (2017). Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. *JAMA oncology*. **3**: 524-548.

Forero-Castro, M., Robledo, C., Benito, R., Abaigar, M., Africa Martin, A., Arefi, M., Fuster, J. L., De Las Heras, N., Rodriguez, J. N., Quintero, J. (2016). Genome-wide DNA copy number analysis of acute lymphoblastic leukemia identifies new genetic markers associated with clinical outcome. *Plos one*. **11**: e0148972.

Frederiksen, L. E., Erdmann, F., Wesseling, C., Winther, J. F., Mora, A. M. (2020). Parental tobacco smoking and risk of childhood leukemia in Costa Rica: a population-based case-control study. *Environmental research*. **180**: 108827.

Friedbichler, K., Kerenyi, M. A., Kovacic, B., Li, G., Hoelbl, A., Yahiaoui, S., Sexl, V., Müllner, E. W., Fajmann, S., Cerny-Reiterer, S. (2010). Stat5a serine 725 and 779 phosphorylation is a prerequisite for hematopoietic transformation. *Blood, The journal of the american society of hematology*. **116**: 1548-1558.

Fruman, D. A., Rommel, C. (2014). PI3K and cancer: lessons, challenges and opportunities. *Nature reviews drug discovery*. **13**: 140-156.

Fumagalli, C., Catania, C., Ranghiero, A., Bosi, C., Viale, G., De Marinis, F., Barberis, M., Guerini-Rocco, E. (2019). Molecular profile of advanced non-small cell lung cancers in octogenarians: the door to precision medicine in elderly patients. *Journal of clinical medicine*. **8**: 112.

Fung, M., Kim, J., Marty, F. M., Schwarzinger, M., Koo, S. (2015). Meta-analysis and cost comparison of empirical versus pre-emptive antifungal strategies in hematologic malignancy patients with high-risk febrile neutropenia. *Plos one*. **10**: e0140930.

Gale, R. P., Cozen, W., Goodman, M. T., Wang, F. F., Bernstein, L. (2000). Decreased chronic lymphocytic leukemia incidence in Asians in Los Angeles County. *Leukemia research*. **24**: 665-669.

Gamero, A. M., Potla, R., Wegrzyn, J., Szelag, M., Edling, A. E., Shimoda, K., Link, D. C., Dulak, J., Baker, D. P., Tanabe, Y. (2006). Activation of Tyk2 and Stat3 is required for the apoptotic actions of interferon-β in primary pro-B cells. *Journal of biological chemistry*. **281**: 16238-16244.

Gartner, J. J., Parker, S. C., Prickett, T. D., Dutton-Regester, K., Stitzel, M. L., Lin, J. C., Davis, S., Simhadri, V. L., Jha, S., Katagiri, N. (2013). Whole-genome sequencing identifies a recurrent functional synonymous mutation in melanoma. *Proceedings of the national academy of sciences*. **110**: 13481-13486.

Gemignani, F., Romei, C., Ciampi, R., Corrado, A., Melaiu, O., Figlioli, G., Bonotti, A., Foddis, R., Cristaudo, A., Pellegrini, G. (2020). Polymorphisms within the RET Proto-Oncogene and Risk of Sporadic Medullary Thyroid Carcinoma. *Thyroid*. **30**: 1579-1588.

Gerami, S. M. N., Somi, M. H., Vahedi, L., Farassati, F., Dolatkhah, R. (2020). The APC gene rs41115 polymorphism is associated with survival in Iranian colorectal cancer patients. *Biomedical research and therapy*. **7**: 3962-3970.

Ghadamyari, F., Heidari, M. M., Zeinali, S., Khatami, M., Merat, S., Bagherian, H., Rejali, L., Ghasemi, F. (2021). Mutational screening through comprehensive bioinformatics analysis to detect novel germline mutations in the APC gene in patients with familial adenomatous polyposis (FAP). *Journal of clinical laboratory analysis*. **35**: e23768.

Ghahremani, S., Shiroudbakhshi, K., Kordasiabi, A. H. S., Firoozbakht, M., Hosseinzadegan, M., Ashrafinia, F., Rahafard, S. (2020). Exposure to magnetic fields and childhood leukemia: An overview of meta-analysis. *International journal of pediatrics-mashhad*. **8**: 11361-11365.

Ghanem, A., Emara, H. A., Muawia, S., Abd El Maksoud, A. I., Al-Karmalawy, A. A., Elshal, M. F. (2020). Tanshinone IIA synergistically enhances the antitumor activity of doxorubicin by interfering with the PI3K/AKT/mTOR pathway and inhibition of topoisomerase II: in vitro and molecular docking studies. *New journal of chemistry*. **44**: 17374-17381.

Ghanem, S., Friedbichler, K., Boudot, C., Bourgeais, J., Gouilleux-Gruart, V., Régnier, A., Herault, O., Moriggl, R., Gouilleux, F. (2017). STAT5A/5B-specific expansion and transformation of hematopoietic stem cells. *Blood cancer journal*. **7**: e514-e514.

Ghoreschi, K., Laurence, A., O'shea, J. J. (2009). Janus kinases in immune cell signaling. *Immunological reviews*. **228**: 273-287.

Ghosh, J., Singh, R., Saxena, R., Gupta, R., Vivekanandan, S., Sreenivas, V., Raina, V., Sharma, A., Kumar, L. (2013). Prevalence and aetiology of anaemia in lymphoid malignancies. *National medical journal of India*. **26**: 79-81.

Globocan, (2021). Cancer Incidence in Five Continents. Global Cancer Observatory.

Gonçalves, M. L., Borja, S. M., Cordeiro, J. a. B. L., Saddi, V. A., Ayres, F. M., Vilanova-Costa, C. a. S. T., Silva, A. M. T. C. (2014). Association of the TP53 codon 72 polymorphism and breast cancer risk: a meta-analysis. *Springerplus*. **3**: 1-8.

Gong, X., Wang, B., Yan, L., Lu, X., Zhao, X. (2020). Linalool inhibits the growth of human T cell acute lymphoblastic leukemia cells with involvement of the MAPK signaling pathway. *Oncology letters*. **20**: 1-1.

Gouilleux-Gruart, V., Gouilleux, F., Desaint, C., Claisse, J.-F., Capiod, J.-C., Delobel, J., Weber-Nordt, R., Dusanter-Fourt, I., Dreyfus, F., Groner, B. (1996). STAT-related transcription factors are constitutively activated in peripheral blood cells from acute leukemia patients. Blood. 87 **(5)**: 1692–1697.

Granatowicz, A., Piatek, C. I., Moschiano, E., El-Hemaidi, I., Armitage, J. D., Akhtari, M. (2015). An overview and update of chronic myeloid leukemia for primary care physicians. *Korean journal of family medicine*. **36**: 197.

Gresham, D., Dunham, M. J., Botstein, D. (2008). Comparing whole genomes using DNA microarrays. *Nature reviews genetics*. **9**: 291-302.

Grieco, M., Santoro, M., Berlingieri, M. T., Melillo, R. M., Donghi, R., Bongarzone, I., Pierotti, M. A., Della Ports, G., Fusco, A., Vecchiot, G. (1990). PTC is a novel rearranged form of the ret proto-oncogene and is frequently detected in vivo in human thyroid papillary carcinomas. *Cell*. **60**: 557-563.

Gross, S. A., Zhu, X., Bao, L., Ryder, J., Le, A., Chen, Y., Wang, X. Q., Irons, R. D. (2008). A prospective study of 728 cases of non-Hodgkin lymphoma from a single laboratory in Shanghai, China. *International journal of hematology*. **88**: 165-173.

Grosse, Y., Baan, R., Secretan-Lauby, B., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Islami, F., Galichet, L., Straif, K. (2011). Carcinogenicity of chemicals in industrial and consumer products, food contaminants and flavourings, and water chlorination byproducts. *The lancet oncology*. **12**: 328-329.

Gullotta, F., Biancolella, M., Costa, E., Colapietro, I., Nardone, A. M., Molinaro, P., Pietropolli, A., Narcisi, M., Di Rosa, C., Novelli, G. (2007). Prenatal diagnosis of

genomic disorders and chromosome abnormalities using array-based comparative genomic hybridization. *Journal of prenatal medicine*. **1**: 16.

Güngör, D., Nadaud, P., Dreibelbis, C., Lapergola, C. C., Wong, Y. P., Terry, N., Abrams, S. A., Beker, L., Jacobovits, T., Järvinen, K. M. (2019). Infant milk-feeding practices and childhood leukemia: a systematic review. *The American journal of clinical nutrition*. **109**: 757S-771S.

Gutierrez-Camino, A., Martin-Guerrero, I., García-Orad, A. (2017). Genetic susceptibility in childhood acute lymphoblastic leukemia. *Medical oncology*. **34**: 1-9.

Guzik, P., Szymkowiak, A., Kulawik, P., Zając, M., Migdał, W. (2022). The confrontation of consumer beliefs about the impact of microwave-processing on food and human health with existing research. *Trends in food science & technology*. **119**: 110-121.

Haan, C., Kroy, D. C., Wüller, S., Sommer, U., Nöcker, T., Rolvering, C., Behrmann, I., Heinrich, P. C., Haan, S. (2009a). An unusual insertion in Jak2 is crucial for kinase activity and differentially affects cytokine responses. *The journal of immunology*. **182**: 2969-2977.

Haan, S., Wüller, S., Kaczor, J., Rolvering, C., Nöcker, T., Behrmann, I., Haan, C. (2009b). SOCS-mediated downregulation of mutant Jak2 (V617F, T875N and K539L) counteracts cytokine-independent signaling. *Oncogene*. **28**: 3069-3080.

Haavisto, A., Henriksson, M., Heikkinen, R., Puukko-Viertomies, L. R., Jahnukainen, K. (2016). Sexual function in male long-term survivors of childhood acute lymphoblastic leukemia. *Cancer*. **122**: 2268-2276.

Habbel, J., Arnold, L., Chen, Y., Möllmann, M., Bruderek, K., Brandau, S., Dührsen, U., Hanoun, M. (2020). Inflammation-driven activation of JAK/STAT signaling reversibly accelerates acute myeloid leukemia in vitro. *Blood advances*. **4**: 3000-3010.

Hall Jr, D. C., Ji, H.-F. (2020). A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease. *Travel medicine and infectious disease*. **35**: 101646.

Hamasaki-Katagiri, N., Lin, B. C., Simon, J., Hunt, R. C., Schiller, T., Russek-Cohen, E., Komar, A. A., Bar, H., Kimchi-Sarfaty, C. (2017). The importance of mRNA structure in determining the pathogenicity of synonymous and non-synonymous mutations in haemophilia. *Haemophilia*. **23**: e8-e17.

Han, L., Wierenga, A. T., Rozenveld-Geugien, M., Van De Lande, K., Vellenga, E., Schuringa, J. J. (2009). Single-cell STAT5 signal transduction profiling in normal and leukemic stem and progenitor cell populations reveals highly distinct cytokine responses. *Plos one*. **4**: e7989.

Hankey, W., Frankel, W. L., Groden, J. (2018). Functions of the APC tumor suppressor protein dependent and independent of canonical WNT signaling: implications for therapeutic targeting. *Cancer and metastasis reviews*. **37**: 159-172.

Hargreave, M., Jensen, A., Toender, A., Andersen, K. K., Kjaer, S. K. (2013). Fertility treatment and childhood cancer risk: a systematic meta-analysis. *Fertility and sterility*. **100**: 150-161.

Hecht, S. S. (1999). Tobacco smoke carcinogens and lung cancer. *JNCI: Journal of the national cancer institute*. **91**: 1194-1210.

Hennighausen, L., Robinson, G. W. (2008). Interpretation of cytokine signaling through the transcription factors STAT5A and STAT5B. *Genes & development*. **22**: 711-721.

Herbst, R. S. (2004). Review of epidermal growth factor receptor biology. *International journal of radiation oncology* biology* physics*. **59**: S21-S26.

Hoa, N. T. K. (2020). Clinical, laboratory features and minimal residual disease levels in childhood acute lymphoblastic leukemia at Hue Central Hospital, Viet Nam.

Hoelbl, A., Kovacic, B., Kerenyi, M. A., Simma, O., Warsch, W., Cui, Y., Beug, H., Hennighausen, L., Moriggl, R., Sexl, V. (2006). Clarifying the role of Stat5 in lymphoid development and Abelson-induced transformation. *Blood*. **107**: 4898-4906.

Höglund, M., Sandin, F., Simonsson, B. (2015). Epidemiology of chronic myeloid leukaemia: an update. *Annals of hematology*. **94**: 241-247.

Hu, Z., Wang, J., Yao, T., Hong, R.-L., Zhang, K., Gao, H., Wu, X., Li, J., Bai, C., Yen, Y. (2013). Identification of novel mutations of TP53, ALK and RET gene in metastatic thymic squamous cell carcinoma and its therapeutic implication. *Lung cancer*. **81**: 27-31.

Huang, Z., Richmond, T. D., Muntean, A. G., Barber, D. L., Weiss, M. J., Crispino, J. D. (2007). STAT1 promotes megakaryopoiesis downstream of GATA-1 in mice. *The journal of clinical investigation*. **117**: 3890-3899.

Hughes, J. P., Rees, S., Kalindjian, S. B., Philpott, K. L. (2011). Principles of early drug discovery. *British journal of pharmacology*. **162**: 1239-1249.

Hung, P.-H., Hsieh, M.-C., Lee, S.-C., Huang, X.-F., Chang, K.-F., Chen, S.-Y., Lee, M.-S., Tsai, N.-M. (2020). Effects of Cedrus atlantica extract on acute myeloid leukemia cell cycle distribution and apoptosis. *Molecular biology reports*. **47**: 8935-8947.

Ibáñez, C. F. (2013). Structure and physiology of the RET receptor tyrosine kinase. *Cold Spring Harbor perspectives in biology*. **5**: a009134.

Idris, M., Farid, J., Sarwar, J., Ahmed, S., Wiqar, M. A., Badsha, S. (2010). Response rate of Pakistani children with acute lymphoblastic leukaemia to medical research council acute lymphoblastic leukaemia 97 chemotherapy protocol. *Journal of ayub medical college abbottabad*. **22**: 8-11.

Imai, Y., Kurokawa, M., Izutsu, K., Hangaishi, A., Maki, K., Ogawa, S., Chiba, S., Mitani, K., Hirai, H. (2001). Mutations of the Smad4 gene in acute myelogeneous leukemia and their functional implications in leukemogenesis. *Oncogene*. **20**: 88-96.

Iqbal, Z. (2014). Molecular genetic studies on 167 pediatric ALL patients from different areas of Pakistan confirm a low frequency of the favorable prognosis fusion oncogene TEL-AML1 (t 12; 21) in underdeveloped countries of the region. *Asian pacific journal of cancer prevention*. **15**: 3541-3546.

Ishfaq, M., Malik, A., Faiz, M., Sheikh, I. A., Asif, M., Khan, M. N., Qureshi, M. S., Zahid, S., Manan, A., Arooj, M. (2012). Molecular characterization of FLT3 mutations in acute leukemia patients. *Asian pacific journal of cancer prevention*. **13**: 4581-4585.

Ivanova, M., Shivarov, V., Pavlov, I., Lilakos, K., Naumova, E. (2016). Clinical evaluation of a novel nine-gene panel for Ion Torrent PGM sequencing of myeloid malignancies. *Molecular diagnosis & therapy*. **20**: 27-32.

Ivey, A., Hills, R. K., Simpson, M. A., Jovanovic, J. V., Gilkes, A., Grech, A., Patel, Y., Bhudia, N., Farah, H., Mason, J. (2016). Assessment of minimal residual disease in standard-risk AML. *New england journal of medicine*. **374**: 422-433.

Jacobson, M. F., Michael, F. (2011). Petition to bar the use of caramel colorings produced with ammonia and containing the carcinogens 2-methylimidazole and 4-methylimidazole. Center for Science in the Public Interest. Available at: *http://cspinet. org/new/pdf/caramel_coloring_petition. pdf*. [Accessed 11 January 2011].

Jaffe, E. S. (2001). World Health Organization classification of tumours. Pathology and genetics of tumours of haematopoietic and lymphoid tissues.

Jaime-Pérez, J. C., García-Arellano, G., Herrera-Garza, J. L., Marfil-Rivera, L. J., Gómez-Almaguer, D. (2019). Revisiting the complete blood count and clinical findings at diagnosis of childhood acute lymphoblastic leukemia: 10-year experience at a single center. *Hematology, transfusion and cell therapy*. **41**: 57-61.

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

Jamy, O. H., Dhir, A., Costa, L. J., Xavier, A. C. (2022). Impact of sociodemographic factors on early mortality in acute promyelocytic leukemia in the United States: A time-trend analysis. *Cancer*. **128**: 292-298.

Jang, M.-A., Yoo, E.-H., Kim, K., Kim, W. S., Jung, C. W., Kim, S.-H., Kim, H.-J. (2013). Chronic lymphocytic leukemia in Korean patients: frequent atypical immunophenotype and relatively aggressive clinical behavior. *International journal of hematology*. **97**: 403-408.

Jastania, R. A., Saeed, M., Al-Khalidi, H., Alquthami, K., Nageeti, T. H., Al-Allaf, F. A., Valerie, K., Taher, M. M. (2020). Adamantinomatous Craniopharyngioma in an Adult: A Case Report with NGS Analysis. *International medical case reports journal*. **13**: 123.

Jatiani, S. S., Baker, S. J., Silverman, L. R., Reddy, E. P. (2010). Jak/STAT pathways in cytokine signaling and myeloproliferative disorders: approaches for targeted therapies. *Genes & cancer*. **1**: 979-993.

Jawaid, A., Arif, K., Amjad, N. (2017). Clinical Presentations of Acute Leukemia in Pediatric Emergency Department of Pakistan. *Bone*. **29**: 27.27-23.23.

Jemal, A., Siegel, R., Xu, J., Ward, E. (2010). Cancer statistics, 2010. *CA: A cancer journal for clinicians*. **60**: 277-300.

Jensen, C. D., Block, G., Buffler, P., Ma, X., Selvin, S., Month, S. (2004). Maternal dietary risk factors in childhood acute lymphoblastic leukemia (United States). *Cancer causes & control*. **15**: 559-570.

Jha, S., Kumar, D. (2021). Acute lymphoblastic leukemia in Indian children at a tertiary care center: A multiparametric study with prognostic implications. *National journal of clinical anatomy*. **10**: 214.

Jing, Q., Wang, Y., Liu, H., Deng, X., Jiang, L., Liu, R., Song, H., Li, J. (2016). FGF s: crucial factors that regulate tumour initiation and progression. *Cell proliferation*. **49**: 438-447.

Joffe, L., Schadler, K. L., Shen, W., Ladas, E. J. (2019). Body composition in pediatric solid tumors: state of the science and future directions. *JNCI monographs*. **2019**: 144-148.

Jones, A. V., Kreil, S., Zoi, K., Waghorn, K., Curtis, C., Zhang, L., Score, J., Seear, R., Chase, A. J., Grand, F. H. (2005). Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders. *Blood*. **106**: 2162-2168.

Juliusson, G., Lazarevic, V., Hörstedt, A.-S., Hagberg, O., Höglund, M. (2012). Acute myeloid leukemia in the real world: why population-based registries are needed. *Blood, The journal of the american society of hematology*. **119**: 3890-3899.

Kaatsch, P. (2010). Epidemiology of childhood cancer. *Cancer treatment reviews*. **36**: 277-285.

Kabat, G. C., Wu, J. W., Moore, S. C., Morton, L. M., Park, Y., Hollenbeck, A. R., Rohan, T. E. (2013). Lifestyle and dietary factors in relation to risk of chronic myeloid leukemia in the NIH-AARP Diet and Health Study. *Cancer epidemiology and prevention biomarkers*. **22**: 848-854.

Kakaje, A., Alhalabi, M. M., Ghareeb, A., Karam, B., Mansour, B., Zahra, B., Hamdan, O. (2020). Interactions of Consanguinity and Number of Siblings with Childhood Acute Lymphoblastic Leukemia. *Biomed research international*. **2020**.

Kallen, M. E., Dulau-Florea, A., Wang, W., Calvo, K. R., (2019). Acquired and germline predisposition to bone marrow failure: diagnostic features and clinical implications. *Seminars in hematology: Elsevier.* pp. 69-82.

Kamran, S., Raca, G., Nazir, K. (2015). RCSD1-ABL1 translocation associated with IKZF1 gene deletion in B-cell acute lymphoblastic leukemia. *Case reports in hematology*. **2015**.

Kang, J.-U. (2018). Overview of cytogenetic technologies. *Korean journal of clinical laboratory science*. **50**: 375-381.

Kang, J. U., Koo, S. H. (2012). Evolving applications of microarray technology in postnatal diagnosis. *International journal of molecular medicine*. **30**: 223-228.

Karakostis, K., Vadivel Gnanasundram, S., López, I., Thermou, A., Wang, L., Nylander, K., Olivares-Illana, V., Fåhraeus, R. (2019). A single synonymous mutation determines the phosphorylation and stability of the nascent protein. *Journal of molecular cell biology*. **11**: 187-199.

Kassahun, W., Tesfaye, G., Bimerew, L. G., Fufa, D., Adissu, W., Yemane, T. (2020). Prevalence of leukemia and associated factors among patients with abnormal hematological parameters in Jimma Medical Center, Southwest Ethiopia: A cross-sectional study. *Advances in hematology*. **2020**.

Kato, Y., Iwama, A., Tadokoro, Y., Shimoda, K., Minoguchi, M., Akira, S., Tanaka, M., Miyajima, A., Kitamura, T., Nakauchi, H. (2005). Selective activation of STAT5 unveils its role in stem cell self-renewal in normal and leukemic hematopoiesis. *The Journal of experimental medicine*. **202**: 169-179.

Katz, A. J., Chia, V. M., Schoonen, W. M., Kelsh, M. A. (2015). Acute lymphoblastic leukemia: an assessment of international incidence, survival, and disease burden. *Cancer causes & control*. **26**: 1627-1642.

Kaufman, D. W., Anderson, T. E., Issaragrisil, S. (2009). Risk factors for leukemia in Thailand. *Annals of hematology*. **88**: 1079-1088.

Kawamura, M., Taki, T., Kaku, H., Ohki, K., Hayashi, Y. (2015). Identification of SPAG9 as a novel JAK2 fusion partner gene in pediatric acute lymphoblastic leukemia with t (9; 17)(p24; q21). *Genes, chromosomes and cancer*. **54**: 401-408.

Kawamura, M., Mcvicar, D. W., Johnston, J. A., Blake, T. B., Chen, Y.-Q., Lal, B. K., Lloyd, A. R., Kelvin, D. J., Staples, J. E., Ortaldo, J. R. (1994). Molecular cloning of L-JAK, a Janus family protein-tyrosine kinase expressed in natural killer cells and activated leukocytes. *Proceedings of the national academy of sciences*. **91**: 6374-6378.

Kearney, L., Gonzalez De Castro, D., Yeung, J., Procter, J., Horsley, S. W., Eguchi-Ishimae, M., Bateman, C. M., Anderson, K., Chaplin, T., Young, B. D. (2009). Specific JAK2 mutation (JAK2 R683) and multiple gene deletions in Down syndrome acute lymphoblastic leukemia. *Blood, the journal of the american society of hematology*. **113**: 646-648.

Khaliq, J., Samad, S., Siddique, S., Imam, M., Shamsi, T. S. (2019). Frequency of Zap-70 and CD38 in Newly Diagnosed Cases of B-Cell Chronic Lymphocytic Leukemia. *National journal of health sciences*. **4**: 52-55.

Kiel, M. J., Velusamy, T., Rolland, D., Sahasrabuddhe, A. A., Chung, F., Bailey, N. G., Schrader, A., Li, B., Li, J. Z., Ozel, A. B. (2014). Integrated genomic sequencing reveals mutational landscape of T-cell prolymphocytic leukemia. *Blood, The journal of the american society of hematology*. **124**: 1460-1472.

Kiem Hao, T., Nhu Hiep, P., Kim Hoa, N. T., Van Ha, C. (2020). Causes of death in childhood acute lymphoblastic leukemia at Hue Central Hospital for 10 years (2008-2018). *Global pediatric health*. **7**: 2333794X20901930.

Kimchi-Sarfaty, C. (2011). " A'silent'polymorphism in the MDR1 gene changes substrate specificity"(January, pg 525, 2007). *Science*. **333**: 39-39.

Kisseleva, T., Bhattacharya, S., Braunstein, J., Schindler, C. (2002). Signaling through the JAK/STAT pathway, recent advances and future challenges. *Gene*. **285**: 1-24.

Kiu, H., Nicholson, S. E. (2012). Biology and significance of the JAK/STAT signalling pathways. *Growth factors*. **30**: 88-106.

Kiyoi, H., Ohno, R., Ueda, R., Saito, H., Naoe, T. (2002). Mechanism of constitutive activation of FLT3 with internal tandem duplication in the juxtamembrane domain. *Oncogene*. **21**: 2555-2563.

Kleppe, M., Spitzer, M. H., Li, S., Hill, C. E., Dong, L., Papalexi, E., De Groote, S., Bowman, R. L., Keller, M., Koppikar, P. (2017). Jak1 integrates cytokine sensing to regulate hematopoietic stem cell function and stress hematopoiesis. *Cell stem cell*. **21**: 489-501. e487.

Kollmann, S., Grundschober, E., Maurer, B., Warsch, W., Grausenburger, R., Edlinger, L., Huuhtanen, J., Lagger, S., Hennighausen, L., Valent, P. (2019). Twins with different personalities: STAT5B—but not STAT5A—has a key role in BCR/ABL-induced leukemia. *Leukemia*. **33**: 1583-1597.

Konstantinova, A. M., Vanecek, T., Martinek, P., Kyrpychova, L., Spagnolo, D. V., Stewart, C. J., Portelli, F., Michal, M., Kazakov, D. V. (2017). Molecular alterations in lesions of anogenital mammary-like glands and their mammary counterparts including hidradenoma papilliferum, intraductal papilloma, fibroadenoma and phyllodes tumor. *Annals of diagnostic pathology*. **28**: 12-18.

Koskela, H. L., Eldfors, S., Ellonen, P., Van Adrichem, A. J., Kuusanmäki, H., Andersson, E. I., Lagström, S., Clemente, M. J., Olson, T., Jalkanen, S. E. (2012). Somatic STAT3 mutations in large granular lymphocytic leukemia. *New england journal of medicine*. **366**: 1905-1913.

Kovacic, B., Stoiber, D., Moriggl, R., Weisz, E., Ott, R. G., Kreibich, R., Levy, D. E., Beug, H., Freissmuth, M., Sexl, V. (2006). STAT1 acts as a tumor promoter for leukemia development. *Cancer cell*. **10**: 77-87.

Kreil, S., Waghorn, K., Ernst, T., Chase, A., White, H., Hehlmann, R., Reiter, A., Hochhaus, A., Cross, N. C. (2010). A polymorphism associated with STAT3 expression and response of chronic myeloid leukemia to interferon α. *Haematologica*. **95**: 148.

Kriegsmann, J., Kriegsmann, M., Kriegsmann, K., Longuespée, R., Deininger, S. O., Casadonte, R. (2019). MALDI imaging for proteomic painting of heterogeneous tissue structures. *Proteomics–clinical applications*. **13**: 1800045.

Kuchinskaya, E., Heyman, M., Nordgren, A., Schoumans, J., Staaf, J., Borg, Å., Söderhäll, S., Grander, D., Nordenskjöld, M., Blennow, E. (2008). Array-CGH reveals hidden gene dose changes in children with acute lymphoblastic leukaemia and a normal or failed karyotype by G-banding. *British journal of haematology*. **140**: 572-577.

Kumar, A., Abbas, A. K., Jon, C. (2015). Aster: Robbins and Cotran pathologic basis of disease. *Elsevier Health Sciences.* Ed (**9**).

Kumar, S., Mahto, N., Bharti, A., Meena, L. P. (2020a). Hematological malignancies in relation with ABO blood group at a teaching hospital, Varanasi, India. *Journal of family medicine and primary care*. **9**: 2309.

Kumar, Y., Singh, H., Patel, C. N. (2020b). In silico prediction of potential inhibitors for the main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing. *Journal of infection and public health*. **13**: 1210-1223.

Kwan, M. L., Block, G., Selvin, S., Month, S., Buffler, P. A. (2004). Food consumption by children and the risk of childhood acute leukemia. *American journal of epidemiology*. **160**: 1098-1107.

Lai, S. Y., Johnson, F. M. (2010). Defining the role of the JAK-STAT pathway in head and neck and thoracic malignancies: implications for future therapeutic approaches. *Drug resistance updates*. **13**: 67-78.

Landau, D. A., Tausch, E., Taylor-Weiner, A. N., Stewart, C., Reiter, J. G., Bahlo, J., Kluth, S., Bozic, I., Lawrence, M., Böttcher, S. (2015). Mutations driving CLL and their evolution in progression and relapse. *Nature*. **526**: 525-530.

Lawi, Z. K., Al-Shuhaib, M. B. S., Amara, I. B., Alkhammas, A. H. (2022). Two missense variants of the epidermal growth factor receptor gene are associated with non small cell lung carcinoma in the subjects from Iraq. *Molecular biology reports*. **49**: 11653-11661.

Leonard, W. J., O'shea, J. J. (1998). Jaks and STATs: biological implications. *Annual review of immunology*. **16**: 293-322.

Levine, R. L., Pardanani, A., Tefferi, A., Gilliland, D. G. (2007). Role of JAK2 in the pathogenesis and therapy of myeloproliferative disorders. *Nature reviews cancer*. **7**: 673-683.

Levine, R. L., Wadleigh, M., Cools, J., Ebert, B. L., Wernig, G., Huntly, B. J., Boggon, T. J., Wlodarska, I., Clark, J. J., Moore, S. (2005). Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer cell*. **7**: 387-397.

Levy, D. E., Darnell, J. (2002). Stats: transcriptional control and biological impact. *Nature reviews Molecular cell biology*. **3**: 651-662.

Li, M., Pinkel, D. (2006). Clinical cytogenetics and molecular cytogenetics. *Journal of zhejiang university science B*. **7**: 162-163.

Li, M. M., Andersson, H. C. (2009). Clinical application of microarray-based molecular cytogenetics: an emerging new era of genomic medicine. *The Journal of pediatrics*. **155**: 311-317.

Li, Q., Li, B., Hu, L., Ning, H., Jiang, M., Wang, D., Liu, T., Zhang, B., Chen, H. (2017). Identification of a novel functional JAK1 S646P mutation in acute lymphoblastic leukemia. *Oncotarget*. **8**: 34687.

Lim, J. Y. S., Bhatia, S., Robison, L. L., Yang, J. J. (2014). Genomics of racial and ethnic disparities in childhood acute lymphoblastic leukemia. *Cancer*. **120**: 955-962.

Lin, P., Luthra, R., Nussenzveig, R. H., Medeiros, L. J. (2010). JAK2 V617F mutation is uncommon in patients with the 3q21q26 syndrome. *Human pathology*. **41**: 758-762.

Lin, S.-Y., Lee, C.-N., Peng, A.-Y., Yuan, T.-J., Lee, D.-J., Lin, W.-H., Ma, G.-C., Chen, M. (2018). Application of molecular cytogenetic techniques to characterize the aberrant Y chromosome arising de novo in a male fetus with mosaic 45, X and solve the discrepancy between karyotyping, chromosome microarray, and multiplex ligation dependent probe amplification. *Journal of the formosan medical association*. **117**: 1027-1031.

Lin, X., Liang, M., Liang, Y.-Y., Brunicardi, F. C., Melchior, F., Feng, X.-H. (2003). Activation of transforming growth factor-β signaling by SUMO-1 modification of tumor suppressor Smad4/DPC4. *Journal of biological chemistry*. **278**: 18714-18719.

Linet, M. (2006). The leukemias. *Cancer epidemiology and prevention*.

Lins, M. M., Mello, M. J. G., Ribeiro, R. C., De Camargo, B., Thuler, L. C. S. (2019). Survival and risk factors for mortality in pediatric patients with acute myeloid leukemia in a single reference center in low–middle-income country. *Annals of hematology*. **98**: 1403-1411.

Liu, W. Q., Dong, J., Peng, Y. X., Li, W. L., Yang, J. (2018). Synonymous mutation adenomatous polyposis coliΔ486s affects exon splicing and may predispose patients to adenomatous polyposis coli/mutY DNA glycosylase mutation-negative familial adenomatous polyposis. *Molecular medicine reports*. **18**: 4931-4939.

Ljungström, V., Cortese, D., Young, E., Pandzic, T., Mansouri, L., Plevova, K., Ntoufa, S., Baliakas, P., Clifford, R., Sutton, L.-A. (2016). Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations. *Blood, The journal of the american society of hematology*. **127**: 1007-1016.

Loeffen, E. A., Brinksma, A., Miedema, K. G., De Bock, G., Tissing, W. J. (2015). Clinical implications of malnutrition in childhood cancer patients—infections and mortality. *Supportive care in cancer*. **23**: 143-150.

Loghavi, S., Kanagal-Shamanna, R., Khoury, J. D., Medeiros, L. J., Naresh, K. N., Nejati, R., Patnaik, M. M. (2023). Fifth edition of the world health classification of tumors of the hematopoietic and lymphoid tissues. *Modern pathology*. 100397.

López, C., Bergmann, A. K., Paul, U., Murga Penas, E. M., Nagel, I., Betts, M. J., Johansson, P., Ritgen, M., Baumann, T., Aymerich, M. (2016). Genes encoding members of the JAK-STAT pathway or epigenetic regulators are recurrently mutated in T-cell prolymphocytic leukaemia. *British journal of haematology*. **173**: 265-273.

Louvigné, M., Rakotonjanahary, J., Goumy, L., Tavenard, A., Brasme, J.-F., Rialland, F., Baruchel, A., Auclerc, M.-F., Despert, V., Desgranges, M. (2020). Persistent osteoarticular pain in children: early clinical and laboratory findings suggestive of acute lymphoblastic leukemia (a multicenter case-control study of 147 patients). *Pediatric rheumatology*. **18**: 1-8.

Lugo, T. G., Pendergast, A.-M., Muller, A. J., Witte, O. N. (1990). Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. *Science*. **247**: 1079-1082.

Luo, Q., Shen, J., Yang, Y., Tang, H., Shi, M., Liu, J., Liu, Z., Shi, X., Yi, Y. (2018). CSF3R T618I, ASXL1 G942 fs and STAT5B N642H trimutation co-contribute to a rare chronic neutrophilic leukaemia manifested by rapidly progressive leucocytosis, severe infections, persistent fever and deep venous thrombosis. *British journal of haematology*. **180**: 892-894.

Ma, X., Wen, L., Wu, L., Wang, Q., Yao, H., Wang, Q., Ma, L., Chen, S. (2015). Rare occurrence of a STAT5B N642H mutation in adult T-cell acute lymphoblastic leukemia. *Cancer genetics*. **208**: 52-53.

Mahmood, N., Shahid, S., Bakhshi, T., Riaz, S., Ghufran, H., Yaqoob, M. (2020). Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach. *Medical & biological engineering & computing*. **58**: 2631-2640.

Mao, L., Yang, J., Deng, G. (2018). Mapping rural–urban disparities in late-stage cancer with high-resolution rurality index and GWR. *Spatial and spatio-temporal epidemiology*. **26**: 15-23.

Margolin, J. F. (2011). Molecular diagnosis and risk-adjusted therapy in pediatric hematologic malignancies: a primer for pediatricians. *European journal of pediatrics*. **170**: 419-425.

Maurer, B., Nivarthi, H., Wingelhofer, B., Pham, H. T. T., Schlederer, M., Suske, T., Grausenburger, R., Schiefer, A.-I., Prchal-Murphy, M., Chen, D. (2020). High activation of STAT5A drives peripheral T-cell lymphoma and leukemia. *Haematologica*. **105**: 435.

Mazurenko, N. N., Tsyganova, I. V., Mikhailova, I. N., Anurova, O. A., Demidov, L. V., Lushnikova, A. A. (2022). Molecular Profile of Cutaneous Melanoma. *Eurasian journal of medicine and oncology*. (1):43–49

Mckenzie, L. M., Allshouse, W. B., Byers, T. E., Bedrick, E. J., Serdar, B., Adgate, J. L. (2017). Childhood hematologic cancer and residential proximity to oil and gas development. *Plos one*. **12**: e0170423.

Mercher, T., Wernig, G., Moore, S. A., Levine, R. L., Gu, T.-L., FröHling, S., Cullen, D., Polakiewicz, R. D., Bernard, O. A., Boggon, T. J. (2006). JAK2 T875N is a novel activating mutation that results in myeloproliferative disease with features of megakaryoblastic leukemia in a murine bone marrow transplantation model. *Blood*. **108**: 2770-2779.

Merker, J. D., Valouev, A., Gotlib, J. (2012). Next-generation sequencing in hematologic malignancies: what will be the dividends? *Therapeutic advances in hematology*. **3**: 333-339.

Messinger, Y., Gaynon, P., Sposto, R., Van Der Giessen, J., Eckroth, E., Malvar, J., Bostrom, B. (2012). Therapeutic Advances in Childhood Leukemia & Lymphoma (TACL) Consortium. Bortezomib with chemotherapy is highly active in advanced B-precursor acute lymphoblastic leukemia: Therapeutic Advances in Childhood Leukemia & Lymphoma (TACL) Study. *Blood*. **120**: 285-290.

Metayer, C., Colt, J. S., Buffler, P. A., Reed, H. D., Selvin, S., Crouse, V., Ward, M. H. (2013). Exposure to herbicides in house dust and risk of childhood acute lymphoblastic leukemia. *Journal of exposure science & environmental epidemiology*. **23**: 363-370.

Metayer, C., Petridou, E., Aranguré, J. M. M., Roman, E., Schüz, J., Magnani, C., Mora, A. M., Mueller, B. A., De Oliveira, M. S. P., Dockerty, J. D. (2016). Parental tobacco smoking and acute myeloid leukemia: the childhood leukemia international consortium. *American journal of epidemiology*. **184**: 261-273.

Miklossy, G., Hilliard, T. S., Turkson, J. (2013). Therapeutic modulators of STAT signalling for human diseases. *Nature reviews drug discovery*. **12**: 611-629.

Mileva, M., Ilieva, Y., Jovtchev, G., Gateva, S., Zaharieva, M. M., Georgieva, A., Dimitrova, L., Dobreva, A., Angelova, T., Vilhelmova-Ilieva, N. (2021). Rose flowers—A delicate perfume or a natural healer?. *Biomolecules*. **11**: 127.

Miller, S., Dykes, D., Polesky, H. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids research*. **16**: 1215.

Mirakbarova, Z., Turdikulova, S. (2021). Targeted Hot Spot Sequencing of Uzbek Lung Cancer Patients. *Biomedical and pharmacology journal*. **14**: 385-390.

Mistry, H., Hsieh, G., Buhrlage, S. J., Huang, M., Park, E., Cuny, G. D., Galinsky, I., Stone, R. M., Gray, N. S., D'andrea, A. D. (2013). Small-Molecule Inhibitors of USP1 Target ID1 Degradation in Leukemic CellsTargeting ID1 in Leukemia by USP1 Inhibitor. *Molecular cancer therapeutics*. **12**: 2651-2662.

Mizuki, M., SchwäBle, J., Steur, C., Choudhary, C., Agrawal, S., Sargin, B. L., Steffen, B. R., Matsumura, I., Kanakura, Y., BöHmer, F. D. (2003). Suppression of myeloid transcription factors and induction of STAT response genes by AML-specific Flt3 mutations. *Blood, the journal of the american society of hematology*. **101**: 3164-3173.

Moelans, C. B., Van Maldegem, C. M., Van Der Wall, E., Van Diest, P. J. (2018). Copy number changes at 8p11-12 predict adverse clinical outcome and chemo-and radiotherapy response in breast cancer. *Oncotarget*. **9**: 17078.

Monge, P., Wesseling, C., Rodríguez, A. C., Cantor, K. P., Weiderpass, E., Reutfors, J., Ahlbom, A., Partanen, T. (2002). Childhood leukaemia in Costa Rica, 1981–96. *Paediatric and perinatal epidemiology*. **16**: 210-218.

Monma, F., Nishii, K., Lorenzo, F., Usui, E., Ueda, Y., Watanabe, Y., Kawakami, K., Oka, K., Mitani, H., Sekine, T. (2006). Molecular analysis of PDGFRα/β genes in core binding factor leukemia with eosinophilia. *European journal of haematology*. **76**: 18-22.

Moriggl, R., Sexl, V., Kenner, L., Duntsch, C., Stangl, K., Gingras, S., Hoffmeyer, A., Bauer, A., Piekorz, R., Wang, D. (2005). Stat5 tetramer formation is associated with leukemogenesis. *Cancer cell*. **7**: 87-99.

Morra, M. E., Kien, N. D., Elmaraezy, A., Abdelaziz, O. a. M., Elsayed, A. L., Halhouli, O., Montasr, A. M., Vu, T. L.-H., Ho, C., Foly, A. S. (2017). Early vaccination protects against childhood leukemia: a systematic review and meta-analysis. *Scientific reports*. **7**: 1-9.

Morris, R., Kershaw, N. J., Babon, J. J. (2018). The molecular details of cytokine signaling via the JAK/STAT pathway. *Protein science*. **27**: 1984-2009.

Morris, S. W., Kirstein, M. N., Valentine, M. B., Dittmer, K. G., Shapiro, D. N., Saltman, D. L., Look, A. T. (1994). Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science*. **263**: 1281-1284.

Mrozek, K., Heerema, N. A., Bloomfield, C. D. (2004). Cytogenetics in acute leukemia. *Blood reviews*. **18**: 115-136.

Mullighan, C. G., Zhang, J., Harvey, R. C., Collins-Underwood, J. R., Schulman, B. A., Phillips, L. A., Tasian, S. K., Loh, M. L., Su, X., Liu, W. (2009). JAK mutations in high-risk childhood acute lymphoblastic leukemia. *Proceedings of the national academy of sciences*. **106**: 9414-9418.

Munir, A. H., Khan, M. I. (2019). Pattern of basic hematological parameters in acute and chronic leukemias. *Journal of medical sciences*. **27**: 125-129.

Musharraf, S. G., Siddiqui, A. J., Shamsi, T., Naz, A. (2017). SERUM metabolomics of acute lymphoblastic leukaemia and acute myeloid leukaemia for probing biomarker molecules. *Hematological oncology*. **35**: 769-777.

Mwaka, A. D., Garimoi, C. O., Were, E. M., Roland, M., Wabinga, H., Lyratzopoulos, G. (2016). Social, demographic and healthcare factors associated with stage at diagnosis of cervical cancer: cross-sectional study in a tertiary hospital in Northern Uganda. *BMJ open*. **6**: e007690.

N Adham, A., F Hegazy, M. E., Naqishbandi, A. M., Efferth, T. (2020). Induction of apoptosis, autophagy and ferroptosis by Thymus vulgaris and Arctium lappa extract in leukemia and multiple myeloma cell lines. *Molecules*. **25**: 5016.

Nadeu, F., Clot, G., Delgado, J., Martin-Garcia, D., Baumann, T., Salaverria, I., Beà, S., Pinyol, M., Jares, P., Navarro, A. (2018). Clinical impact of the subclonal architecture and mutational complexity in chronic lymphocytic leukemia. *Leukemia*. **32**: 645-653.

Naeem, R., Naeem, S., Sharif, A., Rafique, H., Naveed, A. (2017). Acute Myeloid leukemia. *The professional medical journal*. **24**: 1302-1305.

Nahajevszky, S., Andrikovics, H., Batai, A., Adam, E., Bors, A., Csomor, J., Gopcsa, L., Koszarska, M., Kozma, A., Lovas, N. (2011). The prognostic impact of germline 46/1 haplotype of Janus kinase 2 in cytogenetically normal acute myeloid leukemia. *Haematologica*. **96**: 1613.

Nakamura-Ishizu, A., Takizawa, H., Suda, T. (2014). The analysis, roles and regulation of quiescence in hematopoietic stem cells. *Development*. **141**: 4656-4666.

Nam, S., Scuto, A., Yang, F., Chen, W., Park, S., Yoo, H.-S., Konig, H., Bhatia, R., Cheng, X., Merz, K.-H. (2012). Indirubin derivatives induce apoptosis of chronic myelogenous leukemia cells involving inhibition of Stat5 signaling. *Molecular oncology*. **6**: 276-283.

Nasir, M., Jabeen, F., Hussain, S. M., Shaheen, T., Samiullah, K., Chaudhry, A. S. (2015). Impact of Consanguinity, Environment, Socio-Economic and Other Risk Factors on Epidemiology of Leukemia. *Pakistan journal of zoology*. **47**.

Nelson, E. A., Walker, S. R., Weisberg, E., Bar-Natan, M., Barrett, R., Gashin, L. B., Terrell, S., Klitgaard, J. L., Santo, L., Addorio, M. R. (2011). The STAT5 inhibitor pimozide decreases survival of chronic myelogenous leukemia cells resistant to kinase inhibitors. *Blood, The journal of the american society of hematology*. **117**: 3421-3429.

Nelson, E. A., Walker, S. R., Xiang, M., Weisberg, E., Bar-Natan, M., Barrett, R., Liu, S., Kharbanda, S., Christie, A. L., Nicolais, M. (2012). The STAT5 inhibitor pimozide displays efficacy in models of acute myelogenous leukemia driven by FLT3 mutations. *Genes & cancer*. **3**: 503-511.

Network, C. G. a. R. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *New england journal of medicine*. **368**: 2059-2074.

Nips, I. (2019). Pakistan demographic and health survey 2017-18. Islamabad, Pakistan, and Rockville, Maryland, USA: NIPS, ICF. **2**.

Noetzli, L., Lo, R. W., Lee-Sherick, A. B., Callaghan, M., Noris, P., Savoia, A., Rajpurkar, M., Jones, K., Gowan, K., Balduini, C. L. (2015). Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nature genetics*. **47**: 535-538.

Nogami, A., Oshikawa, G., Okada, K., Fukutake, S., Umezawa, Y., Nagao, T., Kurosu, T., Miura, O. (2015). FLT3-ITD confers resistance to the PI3K/Akt pathway inhibitors by protecting the mTOR/4EBP1/Mcl-1 pathway through STAT5 activation in acute myeloid leukemia. *Oncotarget*. **6**: 9189.

O'shea, J. J., Schwartz, D. M., Villarino, A. V., Gadina, M., Mcinnes, I. B., Laurence, A. (2015). The JAK-STAT pathway: impact on human disease and therapeutic intervention. *Annual review of medicine*. **66**: 311-328.

Oh, B. Y., Shin, H.-T., Yun, J. W., Kim, K.-T., Kim, J., Bae, J. S., Cho, Y. B., Lee, W. Y., Yun, S. H., Park, Y. A. (2019). Intratumor heterogeneity inferred from targeted deep sequencing as a prognostic indicator. *Scientific reports*. **9**: 1-8.

Oksuzyan, S., Crespi, C. M., Cockburn, M., Mezei, G., Vergara, X., Kheifets, L. (2015). Race/ethnicity and the risk of childhood leukaemia: a case–control study in California. *Journal of epidemiology and community health*. **69**: 795-802.

Orditura, M., Della Corte, C., Diana, A., Ciaramella, V., Franzese, E., Famiglietti, V., Panarese, I., Franco, R., Grimaldi, A., Lombardi, A. (2018). Three dimensional primary cultures for selecting human breast cancers that are sensitive to the anti-tumor activity of ipatasertib or taselisib in combination with anti-microtubule cytotoxic drugs. *The breast*. **41**: 165-171.

Orkin, S. H., Zon, L. I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*. **132**: 631-644.

Östman, A. (2017). PDGF receptors in tumor stroma: biological effects and associations with prognosis and response to treatment. *Advanced drug delivery reviews*. **121**: 117-123.

Othieno-Abinya, N. A., Mwanda, W. O., Maina, M., Odhiambo, A. O., Mwanzi, S. A., Oyiro, P. O., (2016). Haematological parameters in chronic myeloid leukaemia as determinants of clinical manifestations in this disease. *Journal of clinical oncology.* **34:** *15*

Othman, M. A., Grygalewicz, B., Pienkowska-Grela, B., Rygier, J., Ejduk, A., Rincic, M., Melo, J. B., Carreira, I. M., Meyer, B., Liehr, T. (2016). A novel IGH@ gene rearrangement associated with CDKN2A/B deletion in young adult B-cell acute lymphoblastic leukemia. *Oncology letters*. **11**: 2117-2122.

Ouchveridze, E., Banerjee, R., Desai, A., Aziz, M., Lee-Smith, W., Mian, H., Berger, K., Mcclune, B., Sborov, D., Qazilbash, M. (2022). Financial toxicity in hematological malignancies: a systematic review. *Blood cancer journal*. **12**: 1-9.

Pagani, F., Raponi, M., Baralle, F. E. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proceedings of the national academy of sciences*. **102**: 6368-6372.

Paracha, R. Z., Ahmad, J., Ali, A., Hussain, R., Niazi, U., Tareen, S. H. K., Aslam, B. (2014). Formal modelling of toll like receptor 4 and jak/stat signalling pathways: insight into the roles of socs-1, interferon-β and proinflammatory cytokines in sepsis. *Plos one*. **9**: e108466.

Pardanani, A., Lasho, T., Finke, C., Hanson, C., Tefferi, A. (2007). Prevalence and clinicopathologic correlates of JAK2 exon 12 mutations in JAK2 V617F-negative polycythemia vera. *Leukemia*. **21**: 1960-1963.

Park, A. S., Ritz, B., Yu, F., Cockburn, M., Heck, J. E. (2020). Prenatal pesticide exposure and childhood leukemia–A California statewide case-control study. *International journal of hygiene and environmental health*. **226**: 113486.

Parks, J., Smith, J. (2020). Clinical Implications of Basic Research How to Discover. Antiviral *Drugs quickly*. 1-4.

Parmley, J. L., Hurst, L. D. (2007). How do synonymous mutations affect fitness?. *Bioessays*. **29**: 515-519.

Peng, J., Xue, L., Liu, D., Lin, Y. (2015). Association of the p53 Arg72Pro polymorphism with esophageal cancer in Chinese populations: a meta-analysis. *Genetics and molecular research*. **14**: 9024-9033.

Pervez, S., Jabbar, A. A., Haider, G., Ashraf, S., Qureshi, M. A., Lateef, F., Bashir, I., Zaidi, M., Khurshid, M., Quraishy, M. S. (2020). Karachi cancer registry (KCR): Age-Standardized incidence rate by age-group and gender in a Mega city of Pakistan. *Asian pacific journal of cancer prevention: APJCP*. **21**: 3251.

Peters, B., Ballmann, C., Quindry, T., Zehner, E. G., Mccroskey, J., Ferguson, M., Ward, T., Dumke, C., Quindry, J. (2018). Experimental woodsmoke exposure during exercise and blood oxidative stress. *Journal of occupational and environmental medicine*. **60**: 1073.

Pham, H. T. T., Maurer, B., Prchal-Murphy, M., Grausenburger, R., Grundschober, E., Javaheri, T., Nivarthi, H., Boersma, A., Kolbe, T., Elabd, M. (2018). STAT5B N642H is a driver mutation for T cell neoplasia. *The Journal of clinical investigation*. **128**: 387-401.

Phillips, S. M., Padgett, L. S., Leisenring, W. M., Stratton, K. K., Bishop, K., Krull, K. R., Alfano, C. M., Gibson, T. M., De Moor, J. S., Hartigan, D. B. (2015). Survivors of childhood cancer in the United States: prevalence and burden of morbidity. *Cancer epidemiology and prevention biomarkers*. **24**: 653-663.

Pim, D., Banks, L. (2004). p53 polymorphic variants at codon 72 exert different effects on cell cycle progression. *International journal of cancer*. **108**: 196-199.

Pinz, S., Unser, S., Rascle, A. (2014). The natural chemopreventive agent sulforaphane inhibits STAT5 activity. *Plos one*. **9**: e99391.

Poorcheraghi, H., Hekmatpou, D., Mehrabi, F. (2019). The Quality of Life of Patients With Different Leukemia Types. *Journal of client-centered nursing care*. **5**: 97-104.

Preethi, C. (2014). Clinico-hematological study of acutemyeloid leukemias. *Journal of clinical and diagnostic research*. **8**: FC14.

Puente, X. S., Pinyol, M., Quesada, V., Conde, L., Ordóñez, G. R., Villamor, N., Escaramis, G., Jares, P., Beà, S., González-Díaz, M. (2011). Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. **475**: 101-105.

Pui, C.-H., Robison, L. L., Look, A. T. (2008). Acute lymphoblastic leukaemia. *The lancet*. **371**: 1030-1043.

Quesada, V., Conde, L., Villamor, N., Ordóñez, G. R., Jares, P., Bassaganyas, L., Ramsay, A. J., Beà, S., Pinyol, M., Martínez-Trillos, A. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature genetics*. **44**: 47-52.

Quintana, J., Advis, P., Becker, A., Beresi, V., Campbell, M., Vinés, E., García, H., Salgado, C., Vargas, L., Rojas, J. (2005). Acute myelogenous leukemia in Chile PINDA protocols 87 and 92 results. *Leukemia*. **19**: 2143-2146.

Rabha, R., Ghosh, S., Padhy, P. K. (2018). Indoor air pollution in rural north-east India: Elemental compositions, changes in haematological indices, oxidative stress and health risks. *Ecotoxicology and environmental safety*. **165**: 393-403.

Rahman, F., Tabrez, S., Ali, R., Alqahtani, A. S., Ahmed, M. Z., Rub, A. (2021). Molecular docking analysis of rutin reveals possible inhibition of SARS-CoV-2 vital proteins. *Journal of traditional and complementary medicine*. **11**: 173-179.

Rahmani, A., Doosti-Irani, A., Shokoohizadeh, M. J., Razdari, R. A., Niksiar, S. (2022). Association between arsenic concentration of groundwater and mortality from leukemia and urological cancers in the northwest of Iran. *International journal of environmental health engineering.***12:** 18

Raj, S., Sasidharan, S., Dubey, V. K., Saudagar, P. (2019). Identification of lead molecules against potential drug target protein MAPK4 from L. donovani: An in-silico approach using docking, molecular dynamics and binding free energy calculation. *Plos one*. **14**: e0221331.

Rajala, H. L., Porkka, K., Maciejewski, J. P., Loughran Jr, T. P., Mustjoki, S. (2014). Uncovering the pathogenesis of large granular lymphocytic leukemia—novel STAT3 and STAT5b mutations. *Annals of medicine*. **46**: 114-122.

Rajala, H. L., Eldfors, S., Kuusanmäki, H., Van Adrichem, A. J., Olson, T., Lagström, S., Andersson, E. I., Jerez, A., Clemente, M. J., Yan, Y. (2013). Discovery of somatic STAT5b mutations in large granular lymphocytic leukemia. *Blood, The journal of the american society of hematology*. **121**: 4541-4550.

Rana, S., Shahid, A., Ullah, H., Mahmood, S. (2014). Lack of association of the NPAS2 gene Ala394Thr polymorphism (rs2305160: G> A) with risk of chronic lymphocytic leukemia. *Asian pacific journal of cancer prevention*. **15**: 7169-7174.

Rane, S. G., Reddy, E. P. (2002). JAKs, STATs and Src kinases in hematopoiesis. *Oncogene*. **21**: 3334-3358.

Rathee, R., Vashist, M., Kumar, A., Singh, S. (2014). Incidence of acute and chronic forms of leukemia in Haryana. *International journal of pharmacy and pharmaceutical sciences.* **6**: 323-325.

Rehman, H., Ilyas, M., Shah, M., Ali, W., Khan, A. A. (2020). Prevalence of Cancer in District Banu khyber pakhtunkhwa, Pakistan. *Preprints*.

Ren, X., Zhao, T., Wang, J., Zhu, H., Jiang, H., Jia, J., Yang, S., Jiang, B., Wang, D., Huang, X. (2017). Prognostic significance of blood count at the time of achieving morphologic leukemia-free state in adults with acute myeloid leukemia. *Zhonghua xue ye xue za zhi= Zhonghua xueyexue zazhi*. **38**: 185-191.

Richardson, D. B., Terschüren, C., Pohlabeln, H., Jöckel, K.-H., Hoffmann, W. (2008). Temporal patterns of association between cigarette smoking and leukemia risk. *Cancer causes & control*. **19**: 43-50.

Riera, L., Lasorsa, E., Bonello, L., Sismondi, F., Tondat, F., Di Bello, C., Di Celle, P. F., Chiarle, R., Godio, L., Pich, A. (2011). Description of a novel Janus kinase 3 P132A mutation in acute megakaryoblastic leukemia and demonstration of previously reported Janus kinase 3 mutations in normal subjects. *Leukemia & lymphoma*. **52**: 1742-1750.

Robinette, M. L., Cella, M., Telliez, J. B., Ulland, T. K., Barrow, A. D., Capuder, K., Gilfillan, S., Lin, L.-L., Notarangelo, L. D., Colonna, M. (2018). Jak3 deficiency blocks innate lymphoid cell development. *Mucosal immunology*. **11**: 50-60.

Rowe, J. M. (1999). Bone marrow transplantation for acute lymphoblastic leukemia (ALL) in first complete remission. *Autologous blood and marrow transplantation*. 57.

Ruan, X.-L., Li, S., Meng, X.-Y., Geng, P., Gao, Q.-P., Ao, X.-B. (2015). The role of TP53 gene codon 72 polymorphism in leukemia: a PRISMA-compliant systematic review and meta-analysis. *Medicine*. **94**.

Rull, R. P., Gunier, R., Von Behren, J., Hertz, A., Crouse, V., Buffler, P. A., Reynolds, P. (2009). Residential proximity to agricultural pesticide applications and childhood acute lymphoblastic leukemia. *Environmental research*. **109**: 891-899.

Russo, A., Degrassi, F. (2018). Molecular cytogenetics of the micronucleus: Still surprising. *Mutation research/genetic toxicology and environmental mutagenesis*. **836**: 36-40.

Saberi Hosnijeh, F., Peeters, P., Romieu, I., Kelly, R., Riboli, E., Olsen, A., Tjønneland, A., Fagherazzi, G., Clavel-Chapelon, F., Dossus, L. (2014). Dietary intakes and risk of lymphoid and myeloid leukemia in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Nutrition and cancer*. **66**: 14-28.

Sabir, N., Iqbal, Z., Aleem, A., Awan, T., Naeem, T., Asad, S., Tahir, A. H., Absar, M., Hasanato, R. M., Basit, S. (2012). Prognostically significant fusion oncogenes in Pakistani patients with adult acute lymphoblastic leukemia and their association with disease biology and outcome. *Asian pacific journal of cancer prevention*. **13**.

Sadras, T., Heatley, S. L., Kok, C. H., Mcclure, B. J., Yeung, D., Hughes, T. P., Sutton, R., Ziegler, D. S., White, D. L. (2017). A novel somatic JAK2 kinase-domain mutation in pediatric acute lymphoblastic leukemia with rapid on-treatment development of LOH. *Cancer genetics*. **216**: 86-90.

Saeed, S., Khan, J. A., Iqbal, N., Irfan, S., Shafique, A., Awan, S. (2019). Cancer and how the patients see it; prevalence and perception of risk factors: a cross-sectional survey from a tertiary care centre of Karachi, Pakistan. *BMC public health*. **19**: 1-7.

Sahar, M., Khan, F. S. (2019). Significance of Neuropilin-1 (CD 304) Expression in Paediatric B-Lineage Acute Lymphoblastic Leukemia (ALL). *Journal of Rawalpindi Medical College*. **23**: 2-7.

Salawu, L., Bolarinwa, R., Durosinmi, M. (2010). Chronic lymphocytic leukaemia: a-twenty-years experience and problems in Ile-Ife, South-Western Nigeria. *African health sciences*. **10**.

Saleem, S., Azhar, A., Hameed, A., Khan, M. A., Abbasi, Z. A., Qureshi, N. R., Ajmal, M. (2013). P53 (Pro72Arg) polymorphism associated with the risk of oral squamous cell carcinoma in gutka, niswar and manpuri addicted patients of Pakistan. *Oral oncology*. **49**: 818-823.

Samad, S., Khaliq, J., Taj, M., Shamsi, T. S. (2019). Frequency of Tumor Lysis Syndrome in Acute Leukemia. *National journal of health sciences*. **4**: 67-70.

Sanda, T., Tyner, J. W., Gutierrez, A., Ngo, V. N., Glover, J., Chang, B. H., Yost, A., Ma, W., Fleischman, A. G., Zhou, W. (2013). TYK2–STAT1–BCL2 pathway dependence in T-cell acute lymphoblastic leukemia. *Cancer discovery*. **3**: 564-577.

Sandner, A.-S., Weggel, R., Mehraein, Y., Schneider, S., Hiddemann, W., Spiekermann, K. (2019). Frequency of hematologic and solid malignancies in the family history of 50 patients with acute myeloid leukemia–a single center analysis. *Plos one*. **14**: e0215453.

Sartor, C., Papayannidis, C., Abbenante, M. C., Iacobucci, I., Broccoli, A., Venturi, C., Testoni, N., Ferrari, A., Martinelli, G. (2013). Recurrent gastrointestinal hemorrhage in treatment with dasatinib in a patient showing SMAD4 mutation with acute lymphoblastic leukemia Philadelphia positive and juvenile polyposis hereditary hemorrhagic telangiectasia syndrome. *Hematology reports*. **5**: e7.

Sarwar, L. C. S. R., Lona, H., Halder, D., Selim, S. Hematological Malignancies Pattern among Bangladeshi Adults. *leukemia (ALL)*. **5**: 6.

Schindler, C., Darnell Jr, J. (1995). Transcriptional responses to polypeptide ligands: the JAK-STAT pathway. *Annual review of biochemistry*. **64**: 621-652.

Schindler, C., Plumlee, C., (2008). Inteferons pen the JAK–STAT pathway. *Seminars in cell & developmental biology: Elsevier*. pp. 311-318.

Schindler, C., Levy, D. E., Decker, T. (2007). JAK-STAT signaling: from interferons to cytokines. *Journal of biological chemistry*. **282**: 20059-20063.

Schindler, C., Shuai, K., Prezioso, V. R., Darnell, J. E. (1992). Interferon-dependent tyrosine phosphorylation of a latent cytoplasmic transcription factor. *Science*. **257**: 809-813.

Scott, A. F., Reilly, C. A., (2019). Wood and biomass smoke: addressing human health risks and exposures. *ACS publications*. pp. 219-221.

Seavey, M. M., Dobrzanski, P. (2012). The many faces of Janus kinase. *Biochemical pharmacology*. **83**: 1136-1145.

Seif, F., Khoshmirsafa, M., Aazami, H., Mohsenzadegan, M., Sedighi, G., Bahar, M. (2017). The role of JAK-STAT signaling pathway and its regulators in the fate of T helper cells. *Cell communication and signaling*. **15**: 1-13.

Seita, J., Weissman, I. L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. *Wiley interdisciplinary reviews: systems biology and medicine*. **2**: 640-653.

Sergentanis, T. N., Thomopoulos, T. P., Gialamas, S. P., Karalexi, M. A., Biniaris-Georgallis, S.-I., Kontogeorgi, E., Papathoma, P., Tsilimidos, G., Skalkidou, A., Iliadou, A. N. (2015). Risk for childhood leukemia associated with maternal and paternal age. *European journal of epidemiology*. **30**: 1229-1261.

Shaffer, L., Bejjani, B. (2006). Medical applications of array CGH and the transformation of clinical cytogenetics. *Cytogenetic and genome research*. **115**: 303-309.

Shahab, F., Raziq, F. (2014). Clinical presentations of acute leukemia. *Journal of college of physicians and surgeons pakistan*. **24**: 472-476.

Shahid, M., Raqib, F., Aneel, Y., Ainul, Q., Hina, A., Adna, A., Natasha, P., Roma, T., Azam, H., Ayesha, Z. *et al.*, (2021). Annual Cancer Registry Report-2021: Shaukat Khanum Memorial Cancer Hospital and Research Center, Pakistan. Annual Cancer Registry Report. Pakistan: Shaukat Khanum Memorial Cancer Hospital and Research Center.

Shahid, S., Shakeel, M., Siddiqui, S., Ahmed, S., Sohail, M., Khan, I. A., Abid, A., Shamsi, T. (2020). Novel genetic variations in acute myeloid leukemia in Pakistani population. *Frontiers in genetics*. **11**: 560.

Shaikh, M. R., Haider, G., Memon, P., Rahool, R., Nouman, M., Beg, S., Meher, K., Zahoor, S., Sami, A., Pavan, B. (2020). Cytogenetic Abnormalities In Acute Myeloid Leukaemia Patients. *Journal of Ayub Medical College Abbottabad*. 33-37.

Shaikh, M. S., Adil, S. N., Shaikh, M. U., Khurshid, M. (2014a). Frequency of chromosomal abnormalities in Pakistani adults with acute lymphoblastic leukemia. *Asian pacific journal of cancer prevention*. **15**: 9495-9498.

Shaikh, M. S., Ali, S. S., Khurshid, M., Fadoo, Z. (2014b). Chromosomal abnormalities in Pakistani children with acute lymphoblastic leukemia. *Asian pacific journal of cancer prevention*. **15**: 3907-3909.

Shawahna, R. (2020). Clinical Characteristics and Outcome After Induction of Chemotherapy in Pediatric Acute Lymphoblastic Leukemia: A Descriptive and Correlational Study From the West Bank of Palestine. *Research square.*

Shetty, O., Vengurlekar, V., Kapoor, A., Kamble, V., Gurav, M., Bhargava, P., Srinivas, S., Ramaswamy, A., Ramadwar, M., Saklani, A. P. (2022). The Prevalence of BRAF, PIK3CA, and RAS Mutations in Indian Patients with Colorectal Cancer. *South asian journal of cancer*. **11**: 190-194.

Sic, H., Speletas, M., Cornacchione, V., Seidl, M., Beibel, M., Linghu, B., Yang, F., Sevdali, E., Germenis, A. E., Oakeley, E. J. (2017). An activating janus kinase-3 mutation is associated with cytotoxic T lymphocyte antigen-4-dependent immune dysregulation syndrome. *Frontiers in immunology*. **8**: 1824.

Siegel, R., Miller, K., Jemal, A. (2017a). Cancer Statistics, 2017 CA Cancer J Clin 2017; 67: 7-30. *External resources pubmed/medline (NLM).)*

Siegel, R. L., Miller, K. D., Jemal, A. (2019). Cancer statistics, 2019. *CA: A cancer journal for clinicians*. **69**: 7-34.

Siegel, R. L., Miller, K. D., Fuchs, H. E., Jemal, A. (2021). Cancer statistics, 2021. *CA: A cancer journal for clinicians*. **71**: 7-33.

Siegel, R. L., Miller, K. D., Fedewa, S. A., Ahnen, D. J., Meester, R. G., Barzi, A., Jemal, A. (2017b). Colorectal cancer statistics, 2017. *CA: A cancer journal for clinicians*. **67**: 177-193.

Sigismund, S., Avanzato, D., Lanzetti, L. (2018). Emerging functions of the EGFR in cancer. *Molecular oncology*. **12**: 3-20.

Silva, A. S. J., Júnior Jr, L. S. D. S., De Medeiros Oliveira, G. H., Silva, A. E. M. O. E., Lima, J. P. A., Lima Soares, V., Freitas, R. V., Bahia, F., Oliveira, T. M., Silva, D. G. (2019). Evaluation of Clinical, Hematological and Flow Cytometry Dates in Elderly Patients with Acute Myeloid Leukemia: Impact on Clinical Outcome. *Blood*. **134**: 5115.

Simons, A., Sikkema-Raddatz, B., De Leeuw, N., Konrad, N. C., Hastings, R. J., Schoumans, J. (2012). Genome-wide arrays in routine diagnostics of hematological malignancies. *Human mutation*. **33**: 941-948.

Simpson, H. M., Furusawa, A., Sadashivaiah, K., Civin, C. I., Banerjee, A. (2018). STAT5 inhibition induces TRAIL/DR4 dependent apoptosis in peripheral T-cell lymphoma. *Oncotarget*. **9**: 16792.

Skhoun, H., Khattab, M., Belkhayat, A., Takki Chebihi, Z., Bakri, Y., Dakka, N., El Baghdadi, J. (2022). Association of TP53 gene polymorphisms with the risk of acute lymphoblastic leukemia in Moroccan children. *Molecular biology reports*. **49**: 8291-8300.

Snow, J. W., Abraham, N., Ma, M. C., Abbey, N. W., Herndier, B., Goldsmith, M. A. (2002). STAT5 promotes multilineage hematolymphoid development in vivo through effects on early hematopoietic progenitor cells. *Blood, The journal of the american society of hematology*. **99**: 95-101.

Song, H.-R., Kweon, S.-S., Kim, H. N., Piao, J.-M., Yun, W.-J., Choi, J.-S., Hwang, J.-E., Yoon, J.-Y., Kim, H.-R., Park, Y.-K. (2011). p53 codon 72 polymorphism in patients with gastric and colorectal cancer in a Korean population. *Gastric cancer*. **14**: 242-247.

Stark, G. R., Darnell Jr, J. E. (2012). The JAK-STAT pathway at twenty. *Immunity*. **36**: 503-514.

Steele, M., Narendran, A. (2012). Mechanisms of defective erythropoiesis and anemia in pediatric acute lymphoblastic leukemia (ALL). *Annals of hematology*. **91**: 1513-1518.

Steliarova-Foucher, E., Colombet, M., Ries, L. A., Moreno, F., Dolya, A., Bray, F., Hesseling, P., Shin, H. Y., Stiller, C. A., Bouzbid, S. (2017). International incidence of childhood cancer, 2001–10: a population-based registry study. *The lancet oncology*. **18**: 719-731.

Stieglitz, E., Loh, M. L. (2013). Genetic predispositions to childhood leukemia. *Therapeutic advances in hematology*. **4**: 270-290.

Stoiber, D., Kovacic, B., Schuster, C., Schellack, C., Karaghiosoff, M., Kreibich, R., Weisz, E., Artwohl, M., Kleine, O. C., Muller, M. (2004). TYK2 is a key regulator of the surveillance of B lymphoid tumors. *The Journal of clinical investigation*. **114**: 1650-1658.

Stubbins, R. J., Lee, L., Abou Mourad, Y., Barnett, M. J., Broady, R., Forrest, D. L., Gerrie, A. S., Hogge, D. E., Nantel, S. H., Narayanan, S. (2018). Older adults with acute myeloid leukemia in rural areas are less likely to receive azacitidine with worsened overall survival. *Blood*. **132**: 3989.

Su, L., David, M. (2000). Distinct mechanisms of STAT phosphorylation via the interferon-α/β receptor: selective inhibition of STAT3 and STAT5 by piceatannol. *Journal of biological chemistry*. **275**: 12661-12666.

Sueur, G., Boutet, A., Gotanègre, M., Mansat-De Mas, V., Besson, A., Manenti, S., Bertoli, S. (2020). STAT5-dependent regulation of CDC25A by miR-16 controls proliferation and differentiation in FLT3-ITD acute myeloid leukemia. *Scientific reports*. **10**: 1-12.

Sultan, S., Irfan, S. M., Parveen, S., Mustafa, S. (2016a). Acute lymphoblastic leukemia in adults-an analysis of 51 cases from a tertiary care center in pakistan. *Asian pacific journal of cancer prevention*. **17**: 2307-2309.

Sultan, S., Zaheer, H. A., Irfan, S. M., Ashar, S. (2016b). Demographic and clinical characteristics of adult acute Myeloid Leukemia-tertiary care experience. *Asian pacific journal of cancer prevention*. **17**: 357-360.

Sun, Y., Xiao, J., Li, Z., Fan, S., Shen, Y. (2015). Treatment of childhood leukemia with haploidentical hematopoietic stem cell transplantation using parent as donor: a single-center study of 111 case. *European review for medical and pharmacological sciences*. **19**: 4379-4384.

Surveillance Epidemiology and End Results (Seer), (2019). Cancer Stat Facts: Leukemia: Surveillance, Epidemiology, and End Results. Surveillance Research Program, . U.S: National Cancer Institute.

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research*. **49**: D605-D612.

Tabassum, N., Saboor, M., Ghani, R., Moinuddin, M. (2014). Frequency of JAK2 V617F mutation in patients with Philadelphia positive chronic myeloid leukemia in Pakistan. *Pakistan journal of medical sciences*. **30**: 185.

Tabrizi, M. M., Hosseini, S. A. (2015). Role of electromagnetic field exposure in childhood acute lymphoblastic leukemia and no impact of urinary alpha amylase-a case control study in Tehran, Iran. *Asian pacific journal of cancer prevention*. **16**: 7613-7618.

Tahira, B., Asif, M., Khan, S., Hussain, A., Shahwani, M. N., Malik, A., Inayatullah, S., Iqbal, Z., Rasool, M. (2015). Detection of BCR/ABL fusion gene by hematological and

cytogenetical analysis in chronic myeloid leukemia patients in Quetta, Pakistan. *Asian pacific journal of cancer prevention*. **16**: 3793-3797.

Taketani, T., Taki, T., Sugita, K., Furuichi, Y., Ishii, E., Hanada, R., Tsuchida, M., Sugita, K., Ida, K., Hayashi, Y. (2004). FLT3 mutations in the activation loop of tyrosine kinase domain are frequently found in infant ALL with MLL rearrangements and pediatric ALL with hyperdiploidy. *Blood*. **103**: 1085-1088.

Tapper, W., Jones, A. V., Kralovics, R., Harutyunyan, A. S., Zoi, K., Leung, W., Godfrey, A. L., Guglielmelli, P., Callaway, A., Ward, D. (2015). Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. *Nature communications*. **6**: 1-11.

Tausch, E., Mertens, D., Stilgenbauer, S. (2014). Advances in treating chronic lymphocytic leukemia. *F1000 prime reports*. **6**.

Tefferi, A., Lasho, T., Patnaik, M., Finke, C., Hussein, K., Hogan, W., Elliott, M., Litzow, M., Hanson, C., Pardanani, A. (2010). JAK2 germline genetic variation affects disease susceptibility in primary myelofibrosis regardless of V617F mutational status: nullizygosity for the JAK2 46/1 haplotype is associated with inferior survival. *Leukemia*. **24**: 105-109.

Terwilliger, T., Abdul-Hay, M. (2017). Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood cancer journal*. **7**: e577-e577.

Thomopoulos, T. P., Ntouvelis, E., Diamantaras, A.-A., Tzanoudaki, M., Baka, M., Hatzipantelis, E., Kourti, M., Polychronopoulou, S., Sidi, V., Stiakaki, E. (2015). Maternal and childhood consumption of coffee, tea and cola beverages in association with childhood leukemia: a meta-analysis. *Cancer epidemiology*. **39**: 1047-1059.

Tomasson, M. H., Xiang, Z., Walgren, R., Zhao, Y., Kasai, Y., Miner, T., Ries, R. E., Lubman, O., Fremont, D. H., Mclellan, M. D. (2008). Somatic mutations and germline sequence

variants in the expressed tyrosine kinase genes of patients with de novo acute myeloid leukemia. *Blood, the journal of the american society of hematology*. **111**: 4797-4808.

Tripon, F., Iancu, M., Trifa, A., Crauciuc, G. A., Boglis, A., Balla, B., Cosma, A., Dima, D., Candea, M., Lazar, E. (2020). Association analysis of TP53 rs1042522, MDM2 rs2279744, rs3730485, MDM4 rs4245739 variants and acute myeloid leukemia susceptibility, risk stratification scores, and clinical features: An exploratory study. *Journal of clinical medicine*. **9**: 1672.

Trojani, A., Ripamonti, C. B., Penco, S., Beghini, A., Nadali, G., Di Bona, E., Viola, A., Castagnola, C., Colapietro, P., Grillo, G. (2008). Molecular Analysis of PDGFRA and PDGFRB Genes by Rapid Single-strand Conformation Polymorphism (SSCP) in Patients with Core-binding Factor Leukaemias without KIT or FLT3 Mutation. *Anticancer research*. **28**: 2745-2751.

Trusolino, L., Bertotti, A., Comoglio, P. M. (2010). MET signalling: principles and functions in development, organ regeneration and cancer. *Nature reviews molecular cell biology*. **11**: 834-848.

Ullah, K., Ahmed, P., Raza, S., Satti, T. M., Chaudhry, Q.-U.-N., Akhtar, F., Kamal, M. K., Akhtar, F. M., Khan, B. (2007). Management of acute myeloid leukaemia--5 years experience at Armed Forces Bone Marrow Transplant Centre, Rawalpindi. *The journal of the pakistan medical association*. **57**: 434.

Usman, T., Wang, Y., Liu, C., Wang, X., Zhang, Y., Yu, Y. (2015). Association study of single nucleotide polymorphisms in JAK 2 and STAT 5B genes and their differential mRNA expression with mastitis susceptibility in Chinese Holstein cattle. *Animal genetics*. **46**: 371-380.

Usvasalo, A., Elonen, E., Saarinen-Pihkala, U. M., Räty, R., Harila-Saari, A., Koistinen, P., Savolainen, E.-R., Knuutila, S., Hollmén, J. (2010). Prognostic classification of patients with acute lymphoblastic leukemia by using gene copy number profiles

identified from array-based comparative genomic hybridization data. *Leukemia research*. **34**: 1476-1482.

Vaclavicek, A., Bermejo, J. L., Schmutzler, R. K., Sutter, C., Wappenschmidt, B., Meindl, A., Kiechle, M., Arnold, N., Weber, B. H., Niederacher, D. (2007). Polymorphisms in the Janus kinase 2 (JAK)/signal transducer and activator of transcription (STAT) genes: putative association of the STAT gene region with familial breast cancer. *Endocrine-related cancer*. **14**: 267-277.

Vainchenker, W., Dusa, A., Constantinescu, S. N., (2008). JAKs in pathology: role of Janus kinases in hematopoietic malignancies and immunodeficiencies. *Seminars in cell & developmental biology: Elsevier.* pp. 385-393.

Van Maele-Fabry, G., Duhayon, S., Lison, D. (2007). A systematic review of myeloid leukemias and occupational pesticide exposure. *Cancer causes & control*. **18**: 457-478.

Vardiman, J. W., Thiele, J., Arber, D. A., Brunning, R. D., Borowitz, M. J., Porwit, A., Harris, N. L., Le Beau, M. M., Hellström-Lindberg, E., Tefferi, A. (2009). The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood, The journal of the american society of hematology*. **114**: 937-951.

Vicente, C., Schwab, C., Broux, M., Geerdens, E., Degryse, S., Demeyer, S., Lahortiga, I., Elliott, A., Chilton, L., La Starza, R. (2015). Targeted sequencing identifies associations between IL7R-JAK mutations and epigenetic modulators in T-cell acute lymphoblastic leukemia. *Haematologica*. **100**: 1301.

Vlaanderen, J., Lan, Q., Kromhout, H., Rothman, N., Vermeulen, R. (2011). Occupational benzene exposure and the risk of lymphoma subtypes: a meta-analysis of cohort studies incorporating three study quality dimensions. *Environmental health perspectives*. **119**: 159-167.

Waanders, E., Scheijen, B., Jongmans, M., Venselaar, H., Van Reijmersdal, S., Van Dijk, A., Pastorczak, A., Weren, R., Van Der Schoot, C., Van De Vorst, M. (2017). Germline activating TYK2 mutations in pediatric patients with two primary acute lymphoblastic leukemia occurrences. *Leukemia*. **31**: 821-828.

Wagner, S. E., Burch, J. B., Bottai, M., Puett, R., Porter, D., Bolick-Aldrich, S., Temples, T., Wilkerson, R. C., Vena, J. E., Hébert, J. R. (2011). Groundwater uranium and cancer incidence in South Carolina. *Cancer causes & control*. **22**: 41-50.

Walker, S. R., Xiang, M., Frank, D. A. (2014). Distinct roles of STAT3 and STAT5 in the pathogenesis and targeted therapy of breast cancer. *Molecular and cellular endocrinology*. **382**: 616-621.

Walters, D. K., Mercher, T., Gu, T.-L., O'hare, T., Tyner, J. W., Loriaux, M., Goss, V. L., Lee, K. A., Eide, C. A., Wong, M. J. (2006). Activating alleles of JAK3 in acute megakaryoblastic leukemia. *Cancer cell*. **10**: 65-75.

Wang, A., Jiang, A., Gan, X., Wang, Z., Huang, J., Dong, K., Liu, B., Wang, L., Chen, M. (2020). EGFR-AS1 promotes bladder cancer progression by upregulating EGFR. *Biomed research international*. **2020**.

Wang, C., Li, L., Li, M., Shen, X., Liu, Y., Wang, S. (2021). Inactivated STAT5 pathway underlies a novel inhibitory role of EBF1 in chronic lymphocytic leukemia. *Experimental cell research*. **398**: 112371.

Wang, L., Lawrence, M. S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., Werner, L., Sivachenko, A., Deluca, D. S., Zhang, L. (2011). SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *New england journal of medicine*. **365**: 2497-2506.

Wang, X., Hao, L.-R., Yue, K. (2015). The TP53 codon 72 Pro/Pro genotype may be associated with an increased lung cancer risk in North China: an updated meta-analysis. *International journal of clinical and experimental medicine*. **8**: 3120.

Wang, Z., Bunting, K. D. (2013). STAT5 in hematopoietic stem cell biology and transplantation. *Jak-stat*. **2**: e27159.

Weinkauff, R., Estey, E. H., Starostik, P., Hayes, K., Huh, Y. O., Hirsch-Ginsber, C., Andreeff, M., Keating, M., Kantarjjan, H. M., Freireich, E. J. (1999). Use of peripheral blood blasts vs bone marrow blasts for diagnosis of acute leukemia. *American journal of clinical pathology*. **111**: 733-740.

Wheeler, S. R., Shi, C., Holt, J. A., Vnencak-Jones, C. L. (2016). Mutation profiles of synchronous colorectal cancers from a patient with Lynch syndrome suggest distinct oncogenic pathways. *Journal of gastrointestinal oncology*. **7**: E64.

Whitehead, T. P., Metayer, C., Wiemels, J. L., Singer, A. W., Miller, M. D. (2016). Childhood leukemia and primary prevention. *Current problems in pediatric and adolescent health care*. **46**: 317-352.

Wierenga, A. T., Vellenga, E., Schuringa, J. J. (2008). Maximal STAT5-induced proliferation and self-renewal at intermediate STAT5 activity levels. *Molecular and cellular biology*. **28**: 6668-6680.

Wilch, E. S., Morton, C. C. (2018). Historical and clinical perspectives on chromosomal translocations. *Chromosome translocation*. 1-14.

Wingelhofer, B., Maurer, B., Heyes, E. C., Cumaraswamy, A. A., Berger-Becvar, A., De Araujo, E. D., Orlova, A., Freund, P., Ruge, F., Park, J. (2018). Pharmacologic inhibition of STAT5 in acute myeloid leukemia. *Leukemia*. **32**: 1135-1146.

Wippold, F., Perry, A. (2007). Neuropathology for the neuroradiologist: fluorescence in situ hybridization. *American journal of neuroradiology*. **28**: 406-410.

Wiraatmadja, A., Hidayat, N. S., Sugianli, A. K. (2019). Clinical Manifestations and Hematological Profiles of Pediatric Acute Myeloblastic Leukemia Patients: 3 Years

Observational Study in A West Java Tertiary Hospital, Indonesia. *International journal of integrated health sciences*. **7**: 61-66.

Wiyono, M. R., Bintoro, S. U. Y., Hernaningsi, Y. (2020). Characteristic of chronic myelogenous leukemia patients at the polyclinic of oncology, dr. Soetomo general hospital, Surabaya in 2017. *Majalah biomorfologi*. **30**: 31-38.

Woo, J. S., Alberti, M. O., Tirado, C. A. (2014). Childhood B-acute lymphoblastic leukemia: a genetic update. *Experimental hematology & oncology*. **3**: 1-14.

World Health Organization (2019). Healthy diet (No. WHO-EM/NUT/282/E). World Health Organization. Regional Office for the Eastern Mediterranean.

Wouters, B. J., Löwenberg, B., Erpelinck-Verschueren, C. A., Van Putten, W. L., Valk, P. J., Delwel, R. (2009). Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood, The journal of the american society of hematology*. **113**: 3088-3091.

Wrana, J. L. (2009). The secret life of Smad4. *Cell*. **136**: 13-14.

Wu, S.-J., Huang, S.-Y., Lin, C.-T., Lin, Y.-J., Chang, C.-J., Tien, H.-F. (2010). The incidence of chronic lymphocytic leukemia in Taiwan, 1986-2005: a distinct increasing trend with birth-cohort effect. *Blood, The journal of the american society of hematology*. **116**: 4430-4435.

Wu, S., Ou, T., Xing, N., Lu, J., Wan, S., Wang, C., Zhang, X., Yang, F., Huang, Y., Cai, Z. (2019a). Whole-genome sequencing identifies ADGRG6 enhancer mutations and FRS2 duplications as angiogenesis-related drivers in bladder cancer. *Nature communications*. **10**: 1-12.

Wu, Y.-L., Chien, M.-H., Chou, Y.-E., Chang, J.-H., Liu, T.-C., Tsao, T. C.-Y., Chou, M.-C., Yang, S.-F. (2019b). Association of EGFR mutations and HMGB1 genetic polymorphisms in lung adenocarcinoma patients. *Journal of cancer*. **10**: 2907.

Xiao, Y., Deng, T., Wang, D. (2020). Davanone terpenoid inhibits cisplatin-resistant acute myeloid leukemia cancer cell growth by inducing caspase-dependent apoptosis, loss of mitochondrial membrane potential, inhibition of cell migration and invasion and targeting PI3K/AKT/MAPK signalling pathway. *Journal of Buon*. **25**: 1607-1613.

Xu, Y., Li, Y., Xu, Q., Chen, Y., Lv, N., Jing, Y., Dou, L., Bo, J., Hou, G., Guo, J. (2017). Implications of mutational spectrum in myelodysplastic syndromes based on targeted next-generation sequencing. *Oncotarget*. **8**: 82475.

Yan, K., Xu, X., Liu, X., Wang, X., Hua, S., Wang, C., Liu, X. (2015). The associations between maternal factors during pregnancy and the risk of childhood acute lymphoblastic leukemia: A meta-analysis. *Pediatric blood & cancer*. **62**: 1162-1170.

Yang, F., Anekpuritanang, T., Press, R. D. (2020). Clinical utility of next-generation sequencing in acute myeloid leukemia. *Molecular diagnosis & therapy*. **24**: 1-13.

Yang, S.-M., Li, J.-Y., Gale, R. P., Huang, X.-J. (2015). The mystery of chronic lymphocytic leukemia (CLL): Why is it absent in Asians and what does this tell us about etiology, pathogenesis and biology? *Blood reviews*. **29**: 205-213.

Yasmeen, N., Ashraf, S. (2009). Childhood acute lymphoblastic leukaemia; epidemiology and clinicopathological features. *Journal of pakistan medical association*. **59**.

Yazbeck, N., Samia, L., Saab, R., Abboud, M. R., Solh, H., Muwakkit, S. (2016). Effect of malnutrition at diagnosis on clinical outcomes of children with acute lymphoblastic leukemia. *Journal of pediatric hematology/oncology*. **38**: 107-110.

Ye, J., Lin, M., Zhang, C., Zhu, X., Li, S., Liu, H., Yin, J., Yu, H., Zhu, K. (2020). Tissue gene mutation profiles in patients with colorectal cancer and their clinical implications. *Biomedical reports*. **13**: 43-48.

Zaen-Al-Abideen Pahore, T. S., Shamsi, M. T., Tasneem Farzana, S. H., Nadeem, M., Ahmad, M., Naz, A. (2011). JAK2V617F mutation in chronic myeloid leukemia predicts early disease progression. *Journal of the college of physicians and surgeons Pakistan*. **21**: 472-475.

Zeeshan, R., Sultan, S., Irfan, S. M., Kakar, J., Hameed, M. A. (2015). Clinico-hematological profile of patients with B-chronic lymphoid leukemia in Pakistan. *Asian pacific journal of cancer prevention*. **16**: 793-796.

Zeidner, J. F., Roy, D., Perl, A., Gojo, I. (2018). Myeloid Malignancies. *The American Cancer Society's Oncology in Practice: Clinical management*. **399**.

Zenz, T., Mertens, D., Küppers, R., Döhner, H., Stilgenbauer, S. (2010). From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nature reviews cancer*. **10**: 37-50.

Zhang, F., Li, C., Halfter, H., Liu, J. (2003). Delineating an oncostatin M-activated STAT3 signaling pathway that coordinates the expression of genes involved in cell cycle regulation and extracellular matrix deposition of MCF-7 cells. *Oncogene*. **22**: 894-905.

Zhang, H., Zhan, Q., Wang, X., Gao, F., Yu, J., Wang, J., Fu, W., Wang, P., Wei, X., Zhang, L. (2022). TLS/FUS-ERG fusion gene in acute leukemia and myelodysplastic syndrome evolved to acute leukemia: report of six cases and a literature review. *Annals of hematology*. 1-18.

Zhang, J., Mullighan, C. G., Harvey, R. C., Wu, G., Chen, X., Edmonson, M., Buetow, K. H., Carroll, W. L., Chen, I.-M., Devidas, M. (2011). Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Blood, the journal of the american society of hematology*. **118**: 3080-3087.

Zhang, J., Zhan, Z., Wu, J., Zhang, C., Yang, Y., Tong, S., Sun, Z., Qin, L., Yang, X., Dong, W. (2013). Association among polymorphisms in EGFR gene exons, lifestyle and risk of gastric cancer with gender differences in Chinese Han subjects. *Plos one*. **8**: e59254.

Zhong, Y., Wu, J., Chen, B., Ma, R., Cao, H., Wang, Z., Cheng, L., Ding, J., Feng, J. (2012a). Investigation and analysis of single nucleotide polymorphisms in Janus kinase/signal transducer and activator of transcription genes with leukemia. *Leukemia & lymphoma*. **53**: 1216-1221.

Zhong, Y., Wu, J., Ma, R., Cao, H., Wang, Z., Ding, J., Cheng, L., Feng, J., Chen, B. (2012b). Association of Janus kinase 2 (JAK2) polymorphisms with acute leukemia susceptibility. *International journal of laboratory hematology*. **34**: 248-253.

Zhong, Y., Chen, B., Feng, J., Cheng, L., Li, Y., Qian, J., Ding, J., Gao, F., Xia, G., Chen, N. (2010). The associations of Janus kinase-2 (JAK2) A830G polymorphism and the treatment outcomes in patients with acute myeloid leukemia. *Leukemia & lymphoma*. **51**: 1115-1120.

Annexure 1: Consent form

(A). Quaid-I-Azam University, Islamabad.

Informed Consent for Participation in a Research Study

Study Title: RehanaYasmin, Ph.D. Scholar, Human Genetics, Department of Animal Sciences, Quaid-i-Azam University, Islamabad

Contact Number:

Supervisor: Dr.Sajid Malik, Associate Professor, Animal Sciences (Biological Sciences), Quaid-I-Azam University, Islamabad.

Co-supervisor: Dr.Nafees Ahmad, Senior Scientist, Institute of Biomedical and Genetic Engineering (IBGE), KRL, Islamabad.

Introduction: You have been asked to participate in a research study. Before agreeing to participate in this study you must read and understand the following explanation regarding the purpose, procedure, benefits, and risks of the study. Participation in this study is completely voluntary.

Purpose of Research Study:

Genetic and epigenetic abnormalities are reported in leukemia. We are conducting the present study to determine the genetic polymorphisms and epigenetic regulation of genes in leukemia.

Why have you been asked to take part in this research study?

As you have been diagnosed with a case of Leukemia, therefore, you have been requested to take part in this research study.

Duration of Study: You will not be recruited in this research study, only the remaining samples taken for your diagnosis will be used by the investigator. About 500 people will take part in this study.

Risks and Discomforts of Study: As this study is non-intervening you have no risk due to participating in this research study. Also, information about you and your test reports will be kept secret.

Benefits of Study: You will have no direct benefit from this study. However, you may have indirect benefits from this study. So before visiting a physician, you will show your genetic test that will help him in your treatment.

Maintenance of Confidentiality: Information will be stored in the investigator's research file.

All specimens and data collected as part of this study will be given a unique code number and then each patient will also be given a unique code number. All the data and results of the study will be kept in a locked area accessible only to the investigator.

For statistical analysis, information will be reported as grouped data and individuals will not be identified. The information from the research study may be published; however, you will not be identified in such publication. Any information about you will not be disclosed to anybody without your permission.

Costs of the study: No cost will be paid by you for this research. All the expenses will be paid by the investigator and Quaid-i-Azam University/IBGE/KRL. There is no plan of any financial assistance for participation in this research study.

If you agree to participate, the following information and tests will be required:

1 Personal information: such as name, address, ethnicity, and contact number.

2 Clinical record: we will require your diagnostic reports.

3 Treatment Information: We will also require complete information about the different medicines that you are receiving for Leukemia.

4 Blood and Bone-marrow (drawn for diagnostic purposes)

Signatures:

I have read the information given above. The investigator or his designee has personally discussed with me the research study and has answered my questions. I hereby give my consent to take part in this study as a research study subject.

_____	_____	_____
Name of Subject	Date	Signature
_____	_____	_____
Name of Investigator	Date	Signature

Annexure 2. Data collection proforma/questionnaire

Collecting Officer: _____ Date: _____

Facility: _____ Lab ID: _____

Leukemia Type: AML, ALL, CML, CLL, Other Blood group_____

Patient name_____ Age_____

Contact_____ Gender_____

Weight _____ Height_____

1. **Leukemia Status (data from physicians reports)**

S#	Parameter	Status	Disease Status
1	Family history		
2	LDL		
3	HDL		
4	Chromosomal abnormalities		
5	Blood chemistry		
6	Liver and spleen		
7	Cytogenetic Analysis		
8	Flow cytometery		

Pedigree (draw on a separate sheet if the case is familial)

1.1. Current Blood Count readings

1	RBCs		3	Platelet Count	
2	WBCs		4	Hg	
Differential leukocyte count					
1	Neutrophils		3	mixed	
2	lymphocytes		4	T.L.C	

1.2. Urine examination

1	Color		5	Leukocytes	
2	pH		6	Pus cells	
3	Protein		7	Microbes	
4	Blood		8	Glucose	

1.3. Treatment history/record

S#	Treatments/Medications	Yes/No	Drugs	Dosage	Duration (from-to)	Effects
1	Surgery					
2	Bone Marrow Transplant					
3	Biological Therapy					
4	Chemotherapy					
5	Radiation					

2. Other diseases/complications in the patient

S#	Disease	Present status	Prior	After	Duration
1	Cardiac disorders				
2	Bone anomalies				
3	Depression				
4	Diabetes				
5	Asthma:				
6	Arthritis or other rheumatic disease				
7	Down syndrome				
8	Emphysema or COPD, Other lung disease				
9	Cancer				
10	Fertility treatment				
11	Hormone replacement therapy				
12	Birth control hormones				
13	Other diseases				

3. Demographic information

S#	Parameter	Patient	Father	Mother
1	Birthplace			
2	Mother tongue			
3	Ethnic group (caste)			
4	Family type			
5	Marital status			
6	Parental consanguinity			
7	Education			
8	Religion			

4. Life style

S#	Parameter	Status	Frequency
1	Occupational group		
4	Nutrition		
7	Microwave use		Frequent, Sometimes, Rare
8	Fuel type used in home		
9	Pure Water		
10	Carbonated drinks		Frequent, Sometimes, Rare
11	Perfumes, deodorants		Frequent, Sometimes, Rare
12	Detergents/ washing soap		
13	Smoking (Cigarette, *paan, beri, naswar*)		Frequent, Sometimes, Rare
14	Alcohol consumption		Frequent, Sometimes, Rare
15	Environmental conditions		

Notes [For office use only]
Hazard exposure level (1=lowest, 5=highest): 1 2 3 4 5
Type of exposure (if known)

Name & Signature: ----------

Date: ----------

Annexure 3: List of genes on the cancer panel

S#	Chr.	Gene	Symbol	Gene ID	OMIM ID	Exons	Strand	Biological Pathway
1	9	ABL Proto-Oncogene 1, Non-Receptor Tyrosine Kinase	*ABL1*	25	189980	12	+	Mitotic cell cycle, apoptotic process, DNA damage induced protein phosphorylation
2	14	AKT Serine/Threonine Kinase 1	*AKT1*	207	16473	13	-	Cell proliferation, survival, metabolism, angiogenesis in both normal and malignant cells
3	2	ALK Receptor Tyrosine Kinase	*ALK*	238	105590	29	-	Peptidyl-tyrosine phosphorylation, tyrosine kinase signaling pathway
4	5	APC Regulator of WNT Signaling Pathway	*APC*	324	611731	20	+	Mitotic cytokinesis, cellular response to DNA damage stimulus, cell fate specification
5	11	ATM Serine/Threonine Kinase	*ATM*	472	6	67	+	DNA damage checkpoint signaling, telomere maintenance
6	7	B-Raf Proto-Oncogene Serine/Threonine Kinase	*BRAF*	673	164757	24	-	Regulation of the MAP kinase/ERK signaling pathway, positive regulation of protein phosphorylation, epidermal growth factor receptor signaling pathway

7	16	Cadherin 1	*CDH1*	999	192090	16	+	Negative regulation of cell-cell adhesion,regulation of gene expression,neuron projection development
8	9	Cyclin Dependent Kinase Inhibitor 2A	*CDKN2A*	1029	600160	10	-	Regulation of cyclin-dependent protein serine/threonine kinase activity, apoptotic process,rRNA processing
9	5	Colony Stimulating Factor 1 Receptor	*CSFIR*	1436	164770	24	-	Positive regulation of protein phosphorylation,hematopoietic progenitor cell differentiation, inflammatory response
10	20	Catenin Beta Like 1	*CTNNB1*	56259	611537	19	+	Protein polyubiquitination,cell morphogenesis involved in differentiation,embryonic axis specification, osteoblast differentiation
11	7	Epidermal Growth Factor Receptor	*EGFR*	1956	131550	32	+	Cell morphogenesis, ossification, hair follicle development,receptor-mediated endocytosis
12	17	Erb-B2 Receptor Tyrosine Kinase 4	*ERBB2*	2064	16487	35	+	Positive regulation of protein phosphorylation, cell surface receptor signaling pathway, transmembrane receptor protein tyrosine kinase signaling pathway

13	2	Erb-B2 Receptor Tyrosine Kinase 4	*ERBB4*	2066	600543	31	-	Neural crest cell migration,positive regulation of protein phosphorylation, apoptotic process,signal transduction
14	7	Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit	*EZH2*	2146	6015	26	-	Negative regulation of transcription by RNA polymerase II, chromatin remodeling, regulation of protein phosphorylation,DNA methylation
15	4	F-box and WD Repeat Domain Containing 7	*FBXW7*	55294	606278	20	-	Vasculogenesis, Notch signaling pathway,sister chromatid cohesion,DNA repair
16	8	Fibroblast Growth Factor Receptor 1	*FGFR1*	2260	136350	24	-	Angiogenesis,hematopoietic progenitor cell differentiation,protein phosphorylation,positive regulation of cell population proliferation
17	10	Fibroblast Growth Factor Receptor 2	*FGFR2*	2263	176943	26	-	Ureteric bud development,angiogenesis,transmembrane receptor protein tyrosine kinase signaling pathway,cell-cell signaling
18	4	Fibroblast Growth Factor Receptor 3	*FGFR3*	2261	13493	19	+	Endochondral ossification, chondrocyte differentiation,

								protein phosphorylation, skeletal system development
19	13	Fms Related Receptor Tyrosine Kinase 3	*FLT3*	2322	136	27	-	Apoptosis, proliferation, differentiation of hematopoietic cells in the bone marrow.
20	19	G Protein Subunit Alpha 11	*GNA11*	2767	139313	7	+	Skeletal system development, heart development, cellular response to pH, endothelin receptor signaling pathway
21	9	G Protein Subunit Alpha Q	*GNAQ*	2776	600998	9	-	Action potential, negative regulation of protein kinase activity, blood coagulation, G protein-coupled receptor signaling pathway
22	20	GNAS Complex Locus	*GNAS*	2778	139320	22	+	G protein-coupled receptor signaling pathway, activation of adenylate cyclase activity, regulation of cyclase activity, intracellular transport
23	12	HNF1 Homeobox A	*HNF1A*	6927	142410	9	+	Blastocyst development, chromatin remodeling, regulation of transcription, liver development
24	11	HRas Proto-Oncogene GTPase	*HRAS*	3265	190020	7	-	Endocytosis, regulation of transcription by RNA polymerase II, negative regulation of gene expression, positive

								regulation of protein phosphorylation
25	2	Isocitrate Dehydrogenase (NADP(+)) 1	*IDH1*	3417	147700	12	-	Regulation of phospholipid biosynthetic process, response to steroid hormone, female gonad development, NADP metabolic process
26	15	Isocitrate Dehydrogenase (NADP(+)) 2	*IDH2*	3418	147650	12	-	Carbohydrate metabolic process, NADP metabolic process, isocitrate metabolic process,2-oxoglutarate metabolic process
27	9	Janus Kinase 2	*JAK2*	3717	147796	28	+	Adaptive immune response, response to oxidative stress, G protein-coupled receptor signaling pathway, apoptotic process
28	19	Janus Kinase 3	*JAK3*	3718	600173	24	-	Adaptive immune response, protein phosphorylation,cytokine-mediated signaling pathway,erythrocyte differentiation
29	4	Kinase Insert Domain Receptor	*JAK4*	3791	191306	30	-	Angiogenesis, ovarian follicle development, lymph vessel development, hematopoietic progenitor cell differentiation

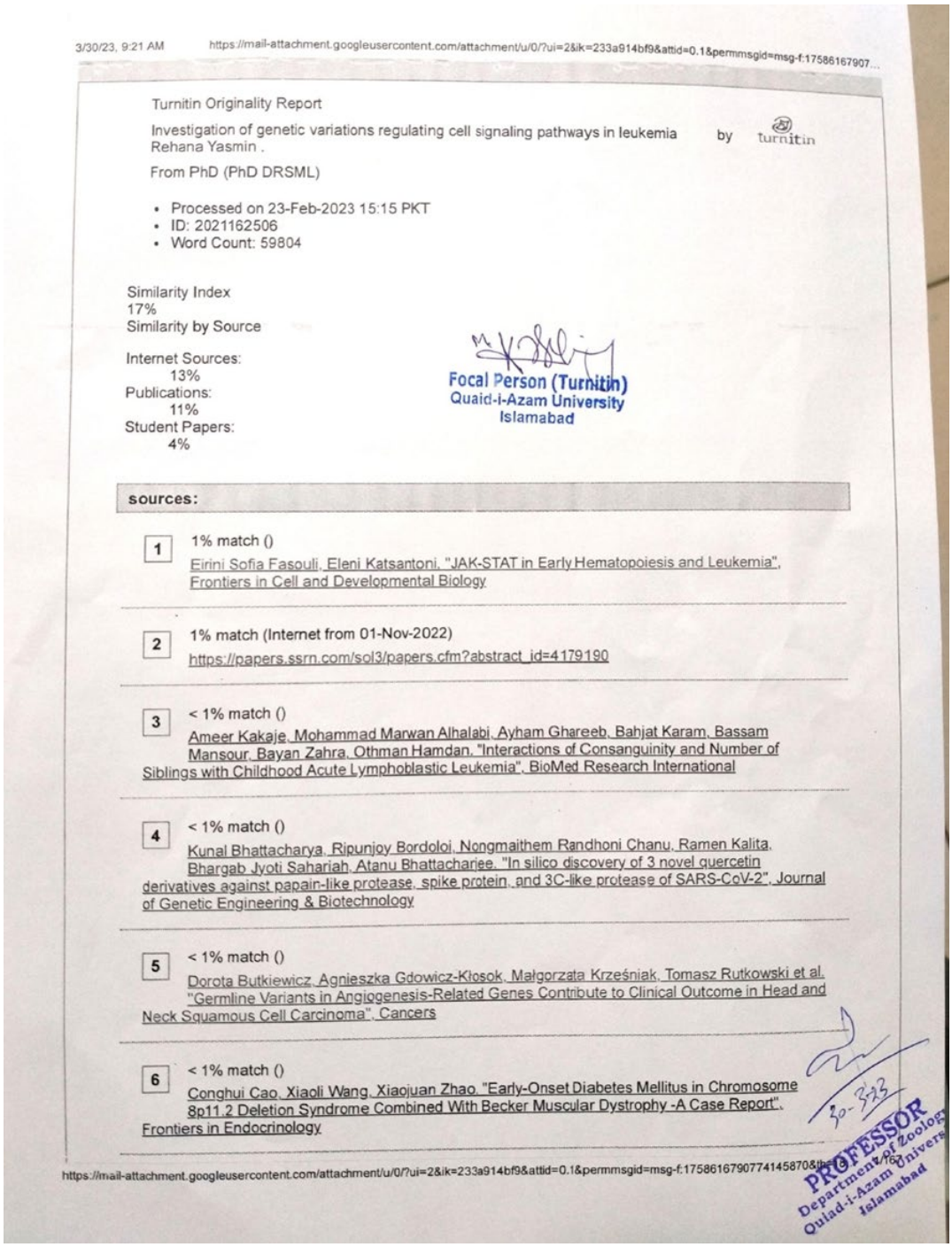
30	4	KIT Proto-Oncogene Receptor Tyrosine Kinase	*JAK5*	3815	164920	21	+	Ovarian follicle development, hematopoietic progenitor cell differentiation, positive regulation of dendritic cell cytokine production, inflammatory response
31	12	KRAS Proto-Oncogene GTPase	*JAK6*	3845	190070	7	-	Positive regulation of cell population proliferation, visual learning, skeletal muscle cell differentiation, forebrain astrocyte development
32	7	MET Proto-Oncogene Receptor Tyrosine Kinase	*JAK7*	4233	164860	22	+	Endothelial cell morphogenesis, protein phosphorylation, nervous system development, cell migration
33	3	MutL Homolog 1	*JAK8*	4292	120436	24	+	Resolution of meiotic recombination intermediates, DNA repair, resolution of meiotic recombination intermediates,double-strand break repair via nonhomologous end joining
34	1	MPL Proto-Oncogene Thrombopoietin Receptor	*JAK9*	4352	159530	12	+	Neutrophil homeostasis,cytokine-mediated signaling pathway, hemopoiesis, platelet aggregation

35	9	Notch Receptor 1	*JAK10*	4851	190198	34	-	Angiogenesis, liver development, heart looping, epithelial to mesenchymal transition
36	5	Nucleophosmin 1	*JAK11*	4869	164040	13	+	Regulation of cell growth, DNA repair, nucleosome assembly, chromatin remodeling
37	1	NRAS Proto-Oncogene GTPase	*JAK12*	4893	164790	7	-	Autoimmune lymphoproliferative syndrome, Noonan syndrome, and juvenile myelomonocytic leukemia
38	4	Platelet-Derived Growth Factor Receptor Alpha	*JAK13*	5156	173490	30	+	Organ development, wound healing, and tumor progression
39	3	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	*PIK3CA*	5290	171834	22	+	
40	10	Phosphatase and Tensin Homolog	*PTEN*	5728	601728	10	+	Tumor suppression, dephosphorylation, and regulation of energy metabolism in the mitochondria
41	12	Protein Tyrosine Phosphatase Non-Receptor Type 11	*PTPN11*	5781	176876	16	+	Cell growth, differentiation, mitotic cycle, and oncogenic transformation
42	13	RB Transcriptional Corepressor 1	*RB1*	5925	614041	27	+	Tumor suppression, stabilization of constitutive

								heterochromatin, and mutation in this gene causes retinoblastoma (RB
43	10	Ret Proto-Oncogene	*RET*	5979	164761	20	+	Cell differentiation, growth and development of organs and tissues derived from the neural crest
44	18	SMAD Family Member 4	*SMAD4*	4089	600993	12	+	Tumor suppression, inhibition of epithelial cell proliferation, reduces angiogenesis
45	22	SWI/SNF Related Matrix Associated Actin Dependent Regulator Of Chromatin Subfamily B, Member 1	*SMARCB 1*	6598	601607	9	+	Relieves the chromatin structures
46	7	Smoothened Frizzled Class Receptor	*SMO*	6608	601500	13	+	Involved in signal transduction pathway, broad expression in ovary and endometrium
47	20	SRC Proto-oncogene Non-Receptor Tyrosine Kinase	*SRC*	6714	190090	21	+	Proto-oncogene, development, and cell growth
48	19	Serine/Threonine Kinase 11	*STK11*	6794	602216	10	+	Regulates cell polarity and energy metabolism and functions as a tumor suppressor

49	17	Tumor Protein p53	*TP53*	7157	191170	11	-	Cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism
50	3	Von Hippel-Lindau Tumor Suppressor	*VHL*	7428	608537	4	+	Cytokine signaling, regulation of senescence

Annexure 4. Plagerism Report



3/30/23, 9:21 AM https://mail-attachment.googleusercontent.com/attachment/u/0/?ui=2&ik=233a914bf9&attid=0.1&permmsgid=msg-f:17586167907...

Turnitin Originality Report

Investigation of genetic variations regulating cell signaling pathways in leukemia by turnitin
Rehana Yasmin .

From PhD (PhD DRSML)

- Processed on 23-Feb-2023 15:15 PKT
- ID: 2021162506
- Word Count: 59804

Similarity Index
17%
Similarity by Source

Internet Sources:
13%
Publications:
11%
Student Papers:
4%

Focal Person (Turnitin)
Quaid-i-Azam University
Islamabad

sources:

1 1% match ()
Eirini Sofia Fasouli, Eleni Katsantoni, "JAK-STAT in Early Hematopoiesis and Leukemia", Frontiers in Cell and Developmental Biology

2 1% match (Internet from 01-Nov-2022)
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4179190

3 < 1% match ()
Ameer Kakaje, Mohammad Marwan Alhalabi, Ayham Ghareeb, Bahjat Karam, Bassam Mansour, Bayan Zahra, Othman Hamdan, "Interactions of Consanguinity and Number of Siblings with Childhood Acute Lymphoblastic Leukemia", BioMed Research International

4 < 1% match ()
Kunal Bhattacharya, Ripunjoy Bordoloi, Nongmaithem Randhoni Chanu, Ramen Kalita, Bhargab Jyoti Sahariah, Atanu Bhattacharjee, "In silico discovery of 3 novel quercetin derivatives against papain-like protease, spike protein, and 3C-like protease of SARS-CoV-2", Journal of Genetic Engineering & Biotechnology

5 < 1% match ()
Dorota Butkiewicz, Agnieszka Gdowicz-Kłosok, Małgorzata Krześniak, Tomasz Rutkowski et al. "Germline Variants in Angiogenesis-Related Genes Contribute to Clinical Outcome in Head and Neck Squamous Cell Carcinoma", Cancers

6 < 1% match ()
Conghui Cao, Xiaoli Wang, Xiaojuan Zhao, "Early-Onset Diabetes Mellitus in Chromosome 8p11.2 Deletion Syndrome Combined With Becker Muscular Dystrophy -A Case Report", Frontiers in Endocrinology

https://mail-attachment.googleusercontent.com/attachment/u/0/?ui=2&ik=233a914bf9&attid=0.1&permmsgid=msg-f:1758616790774145870&th... 1/16

30-3-23

PROFESSOR
Department of Zoology
Quaid-i-Azam University
Islamabad

Annexure 5. Article Abstract

Pakistan J. Zool., pp 1-11, 2023. DOI: https://dx.doi.org/10.17582/journal.pjz/20221215141236

Epidemiological and Clinical Correlates of Leukemia Ascertained in a Multiethnic Cohort of Pakistan

Rehana Yasmin[1,2], Rashda Abbasi[2]*, Tariq Saeed[2], Madiha Sadiq[2], Nuzhat Yasmeen[3], Muhammad Iqbal[4], A. Khuzaim Alzahrani[5], Nadeem Kizilbash[5], Ugur Bilge[6], Nafees Ahmad[2] and Sajid Malik[1]*

[1]Human Genetics Program, Department of Zoology, Faculty of Biological Sciences, Quaid-I-Azam University, 45320, Islamabad, Pakistan
[2] Institute of Biomedical and Genetic Engineering, 44000, Islamabad, Pakistan
[3] Department of Oncology, Pakistan Institute of Medical Sciences, 44080, Islamabad. Pakistan
[4] Institute of Radiotherapy and Nuclear Medicine, 25121, Peshawar, Pakistan
[5] Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, Northern Border University, Arar-91431, Saudi Arabia
[6] Department of Biostatistics and Medical Informatics, Faculty of Medicine, Akdeniz University, Antalya, Turkey

Article Information
Received 02 February 2023
Revised 20 February 2023
Accepted 29 March 2023
Available online 13 May 2023
(early access)

Authors' Contribution
RY conceptualization, data collection, data curation, formal analysis, methodology, writing original draft. RA conceptualization, methodology, investigation. TS and MS data collection, data curation. AKA, NK and UB data analysis. MI and NY data curation, visualization. NA writing review and editing, validation. SM conceptualization, formal analysis, software, writing review and editing.

Key words
Leukemia, Epidemiology, Cross-sectional study, Retrospective, Risk factors

ABSTRACT

Leukemia is one of the most prevalent cancers among pediatric malignancies and causes huge economic burden. In this case-control study epidemiological, environmental, life-style related risk factors and phenotypic characteristics of leukemia subtypes were investigated in Pakistani population. A total of 1500 subjects, including 616 patients and 884 controls were recruited through a retrospective cross-sectional sampling design. Descriptive summaries were generated, and risk factors were analyzed through logistic regression. We identified Pathan ethnicity (OR=2.85; 95%CI=2.29-3.54), no formal education (OR=3.36; 95%CI=2.62-4.32), poor diet (OR=2.34; 95%CI=1.79-3.06), lower BMI (OR=1.95; 95%CI=1.50-2.60), parental consanguinity (OR=2.13; 95%CI=1.67-2.71), positive family history (OR=4.24; 95%CI=2.18-8.26), rural residential setup (OR=2.93; 95%CI=2.10-4.10), drinking of groundwater (OR=2.25; 95%CI=1.6479-3.0964), wooden fuel (OR= 3.97; 95%CI=3.14-5.01), carbonated drinks (OR=1.25, 95%CI=1.00-1.57) and tobacco usage (OR=1.57, 95%CI=1.24-1.98) as significant risk factors for leukemia. However, odds ratios were significantly lower for patients using microwave oven (OR=0.25; 95%CI=0.18-0.35), and perfumes (OR=0.42; 95%CI=0.33-0.53). Males exhibit an increased risk for lymphoid leukemia as compared to myeloid leukemia (OR=1.97; 95%CI=1.38-2.80). Paraclinical parameters indicated that 71% of the cases had >50% of blast cells. Leukocytosis (OR= 9.06; 95% CI=6.46-12.71), anemia (OR= 15.84; 95% CI=11.84-21.21), low hemoglobin (OR=8.11; 95% CI=6.35-10.37), thrombocytopenia (OR=32.40; 95% CI=21.57-48.68), lymphocytosis (OR= 3.41; 95% CI=2.55-4.57), and neutropenia (OR=7.32; 95% CI=5.59-9.60) had significantly higher odd ratio for leukemia patients. Leukemia risk factors are mainly relevant to exposure due to rural residence, poor lifestyle, and family history of the disease. The disease incidence can be minimized by designing and implementing risk mitigation strategies.

INTRODUCTION

Leukemia is a heterogeneous group of clonal hematologic malignancies characterized by the abnormal proliferation of hematopoietic cells interrupting the normal function of blood and bone marrow. General manifestations of leukemia include fever or chills, dyspnea, persistent fatigue, weakness, recurrent infections, weight loss, swollen lymph nodes, enlarged liver or spleen, and bruising (Castro *et al.*, 2015; Louvigne *et al.*, 2020). It is subdivided into acute lymphoblastic

* Corresponding author: r.abbasi@daad-alumni.de, malik@qau.edu.pk
0030-9923/2023/0001-0001 $ 9.00/0