Biology and Medicine

Research Article   **Open Access**

Polymorphisms of *BCL2* Gene in Acute Lymphoblastic Leukemia Patients in Pakistan and Screening of Phytochemicals to Overcome its Expression

Huma Butt[*], **Amara Khalid and Muhammad Yaqoob**

Department of Microbiology and Molecular Genetics, University of the Punjab, Lahore, Pakistan

[*]**Corresponding author:** Huma Butt, Department of Microbiology and Molecular Genetics, University of the Punjab, Lahore, Pakistan, Tel: 0324-9412877; E-mail: Humaabutt1993@gmail.com

Abstract

Leukemia is a third fatal disease in developing countries. According to Punjab cancer registry 2014 report, leukemia incidence in Pakistan is 18.8%. Many treatments are available for leukemia now days but still there is a mortality rate measured. *BCL2* controls cell apoptosis but when overexpressed due to the presence of promoter region single nucleotide polymorphism 938C>A, it causes anticancer drug resistance and represses apoptosis of malfunctioned cells resulting in immature leucocytes. The aim of the present study was genotyping of promoter region SNP in acute lymphoblastic leukemia patients of Pakistan and screening of naturally occurring phytochemical to identify *BCL2* active site targeting compounds to control its expression. Study was performed on 104 cases taking similar control size showing allelic frequency C=0.466 and A=0.534 in cases, indicating an association with the disease by having chi-square p-value 0.0032. Screening of 90 phytochemicals was performed out of which one compound was selected as a lead compound based upon drug likeliness properties, binding with protein, and natural existence that may cause the least harm to human body and greater effect on leukemia cells. The information provided by this research can be used to make new-targeted drugs to control leukemia by inhibiting the mutated proteins.

Introduction

The progress of human genome project expanded the competence of exploring the genetic basis of diseases including rare genetic diseases, syndromes and cancers leading to pathogenesis of disorders [1]. These evolvements steered new roots of treatments like targeted therapies or the personalized medicines to cure diseases like cancer. Leukemia is a hematological malignancy. It generally commences in the bone marrow and ensue high number of atypical leukocytes (white blood cells). These leukocytes are immature and undifferentiated cells called as blast cells [2]. Acute leukemia is depicted by diverse genetic and chromosomal abnormalities, with altered frequency by age. The two-step model for childhood leukemia suggests that leukemia occurs after both a first mutation, usually a chromosomal translocation taking place in uterus, and a second mutation stirring after birth [3]. In all cancer, leukemia is almost 2.7% and its frequency is 4.7 per 100,000. Numbered manifestation of leukemia in Pakistani population is 11 per 100,000 per year [4] and ALL accounts for 80% of all childhood leukemia cases [5].

BCL2 is a regulatory protein that regulates the cell apoptosis by stimulating the pro-apoptotic protein or inhibiting the anti-apoptotic protein [6]. This gene encodes for the essential outer mitochondrial membrane protein that stops the self-death program of lymphocytes [7]. Hence resulting in increased number of lymphocytes in blood [8].

BCL2 is present on chromosome 18 at location 18q21.33. It consists of 6 exons which encodes for 239 amino acid protein having weight 26266 Da. It comprises of two functional domains BH4 (amino acid 10-29) and a hydrophobic cleft formed of BH3 (amino acid 90-107),

BH2 (amino acid 184-199), and BH1 (amino acid 133-152). A helical transmembrane domain (amino acid 212-233) is also present which is required for membrane anchoring. Amino acids in-between BH1 and BH2 form a channel for the movement of Ca ions in and outside the mitochondria, and are essential for the interaction with BAX and anti-apoptotic activity. The hydrophobic cleft is also known as the receptor domain where the BH3 proteins bind and inhibit apoptosis (Figure 1) [9]. *BCL2* comprises of 2 promoter regions P1 and P2. P1 is located at position 1386 to 1423 bp upstream of the GC rich initiation translation site including TATA-less promoter comprising of multiple transcription initiation sites. P2 is located at 1300bp at the start of gene, comprises of CCAAT box, which is an octamer motif and TATA element. In normal B-cells, P1 promoter is used majorly but in B-cell lymphoma cell lines, P2 becomes active due to increase in transcription factor [10]. The imposed BCL2 expression interrupts the programmed cell death of cell lines and hence prolongs life span of cells. Along with this, it also confers resistance to cytotoxic agents and provides resistance to leukemic cell lines against anticancer drugs [11]. Single nucleotide polymorphism (SNP) rs2279115 is studied much in association with acute lymphocytic leukemia. A study on SNP 938 C/A showed that presence of C allele has a significant effect on promoter 2 activation; the increase in interaction between the promoter and transcription factors lowers the BCL2 protein level [12]. This SNP is present as 5 prime UTR variant located near promoter region [13].

Designing and marketing a drug may take about minimum 10 years. To accelerate the process of drug discovery, Insilco approaches are used, which is called computer aided drug designing (CADD). The CADD process comprises of target identification, validation, selection of ligands, screening, optimization ADME (adsorption, distribution, metabolism, and excretion) and the toxicity. It also deals with the computational proteomics, chemical properties, and physical

properties including pharmacodynamics and pharmacokinetic properties, structural and topological proteomics. CADD is further parted into two groups namely, structure based drug designing and the ligand based drug designing [14]. Molecular docking is a technical approach to prophesy the ideal binding of one molecule to the other forming stable complex. This procedure is interpreted by the scoring function based on binding affinity energies of these molecules [15]. Docking is applied for the hit identification, to screen large databases for desired ligands with specific property and lead optimization in which the location and the orientation of the ligands that binds to the protein.

Present study is based on the genotyping of BCL2 SNP in Pakistani population and screening of specific targeted inhibitors that binds at increased BH3 sites in cells due to increased expression of BCL2 protein in cell and decrease the anticancer drugs resistance in acute lymphoblastic leukemia (ALL) patients.

Figure 1: Structure of BCL2.
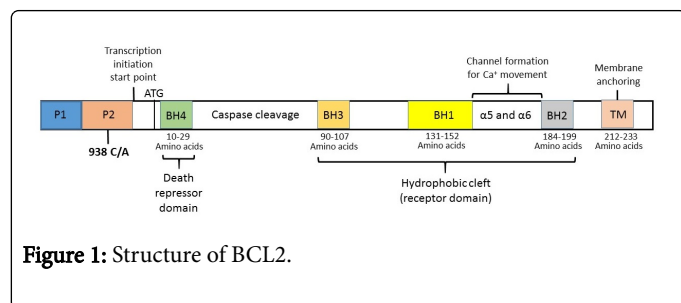
Methodology

The research work was conducted at department of microbiology and molecular genetics, university of the Punjab; Lahore, Pakistan in time duration of one-year, June 2016 to July 2017 and in this period of study, sampling was done for 6 months from November 2016 to April 2017. Practically 208 study subjects were included in this study (104 clinically diagnosed acute lymphoblastic leukemia (ALL) patients and 104 healthy controls) were conscripted from The Children's Hospital & the Institute of Child Health, Lahore, Pakistan. Patients recently diagnosed or presently being treated with ALL were included in this study. Making sure that all the patients have B cell markers (CD10, CD19, CD20, CD22, and cCD79a) positive indicating B-cell acute lymphoblastic leukemia whereas, Leukemia patients with any type other than B-ALL or any other blood related disease and ethnicity other than Pakistani were excluded. The controls were selected randomly from the general population considering the ones selected are healthy and of the same age group and sex, showing serum traits in normal range, in accordance with the cases. All the subjects were asked for a written informed consent and questionnaire. The procedures taken in use were ethically approved by the institutional ethical board and Children's Hospital & the Institute of Child Health ethical committee. We calculated a sample size 104 for the concerned disease using formula:

$$n = Z^2 * P(1-P)/d^2$$

Here Z=Z-score on the basis of confidence level 95% that is=1.96, P=it is the prevalence of the disease obtained from literature, that is 18.8% from data of 2015, d=it is the error margin also called confidence interval taken as 7.5% [16].

Biological parameters

Biochemical parameters were determined by recently performed hematology and chemical pathology test reports of patients during sampling era at the Children's hospital Lahore, Pakistan. In hematology, complete blood count was taken in notice. The biochemical parameters included total bilirubin, serum creatinine, blood urea, aminotransferase (ALT), alkaline phosphatase (ALP), uric acid, albumin, random glucose, and salts consisting, sodium, potassium, chloride, calcium, magnesium, and phosphorous.

Primer designing

Reference sequence of gene was taken from NCBI (Ref gene ID NG_009361.1) for primer designing. Using Primer1 software (http://primer1.soton.ac.uk/primer1.html), primers were designed for tetra arm PCR keeping primer length 22-25 nucleotides. Designed primers were crosschecked using UCSC Genome Browser in-silico PCR facility. Further confirmation was done regarding secondary structures using Oligo Analyzer software online.

Genotyping

Virtually 2 ml of whole blood was drawn from 104 patients and 104 healthy controls and kept in disodium EDTA vial. Sambrook 2001 organic method was used for DNA extraction and results were observed on 2% agarose gel. Tetra arm PCR was performed using following set of primers. Outer Forward primer 5'-CCGGCTCCTTCATCGTCTCC-3', outer reverse primer 5'-CCCAGGAGAGAGACAGGGGAAAT -3', inner forward primer 5'-AATAAAACCCTCCCCCACCACCT-3' and inner reverse primer 5'-CCCTTCTCGGCAATTTACACGC -3'. PCR program consisted of initial denaturation at 95°C for 5min, denaturation at 94°C for 1min, annealing at 62.5°C for 1min, extension at 72°C for 1min and final extension at 72°C for 5 minutes. Step 2 to 4 was repeated 30 times. For amplification, advanced primus 96 (PeqLab) thermal cycler was used.

In-silico analysis

Fasta sequence of BCL2 protein was retrieved from NCBI, protein ID=NP_000624.2. The sequence of amino acids was subjected to Raptorx software (http://raptorx.uchicago.edu/) for structure prediction. The 3D structures were confirmed using the SAVES v.4 (https://services.mbi.ucla.edu/SAVES/).

Retrieval of ligands

The anti-leukemic compounds were retrieved from the Duke's phytochemical and ethnobotanical databases https://phytochem.nal.usda.gov/phytochem/search) [17]. The compounds 3D structures were taken from PubChem Compound (https://pubchem.ncbi.nlm.nih.gov/) [18] and ChemSpider (http://www.chemspider.com/) [19]. Almost 90 active compounds were selected having anti-leukemic activity and IC_{50} value less than 100 μM.

Application of Drug Likeliness Rules

On selected ligands from the databases, drug likeliness filters were applied which deals with the pharmacodynamics and pharmacokinetic properties of the compounds. In pharmacokinetics, selected Lipinski rule of five, blood brain barrier, TPSA and Pgp substrate values, which were calculated using SwissADME software (http://

www.swissadme.ch/). It has an input in the form of SMILES that were taken from the PubChem compound and ChemSpider. In pharmacodynamics, minimum inhibitory concentration (IC_{50}) value and toxicity value were taken in concern. The toxicity values were calculated by PROTOX (http://tox.charite.de/tox/) [20].

Target site selection

Target site of protein was selected based on data gathered from NCBI, marking the active regions of the protein for like phosphorylation sites that helps in functioning of the protein. BH3 domain (amino acid 90-107), also known as the hydrophobic cleft was selected as the target site.

Molecular docking

AutoDock Vina 1.5.6 [21] was used for molecular docking. The grid box was drawn surrounding the active site. The grid plates per map were 8505 of grid box drawn. Grid box number of points in dimensions, x=14, y=26, and z=20 followed with centers, x=13.344, y=0.51, and z=15.606, keeping space (angstrom) 1.00. Pymol software was used to visualize the molecular docking results. Hydrogen bonding was analyzed in the present ligand format and their bond length was measured.

Statistical analysis

Genotyping analysis was done using software SPSS IBM version 22 and excel 2013. T-test, ANOVA and chi square was used to determine the association of parameters with genotype, SNPs association with ALL, and genotype effect on ALL. Odds ratio along with 95% confidence interval was calculated using logistic regression to evaluate the association potency with ALL.

Results

The acute lymphoblastic leukemia cases were gathered from Children's hospital while controls were extraneous subjects but ethnologically accorded and healthy individuals from the general populace. A total of 208 subjects were included in the study from different areas of Punjab, Pakistan. In which 104 were ALL cases (66 males and 38 females), while 104 were healthy, control subjects (62 males and 42 females). Our study encompassed 34 (32.7) cases being treated with chemotherapy, 29 (27.9%) patient's fathers were found to be smokers, not a single patient or their parents had been exposed to radiations, and 43 (41.3%) patients parents have consanguinity marriages. 17 (16.3%) of the total cases had a family history of cancer. Mother and father education was also taken in consideration that if they were educated they would be more anxious about the health issue of their child and get them treated by time. Almost 39.4% of mothers and 16.3% of fathers were educated that they had almost matriculation degree and 23% mothers were graduates. When we had a birds view over the cast we found that the disease is more common in arraign cast almost 23% following Pathan 12% and jutt 16%. Compiling all the cases we analyzed that more cases of ALL were from Lahore (37%) followed by cases in Sialkot (15%) in Pakistan. Results indicates 17 patients had liver disorder and enlargement, having size 12.7 ± 3.3. Almost 14 patients has a spleen malfunctioning and enlargement, having size 11.6 ± 2.4. About 7 cases with kidney enlargement and improper functioning, having size 10.47 ± 2.9 of right kidney and 10.9 ± 2.3 of left kidney. Other symptoms like fever, abdomen pain, cough with blood, flu, epistaxis, pallor, jaundice, bruises, cervical lymphadenopathy, enlarged lymph nodes, fits, microcytic blood, hypo chromatic, night sweating, loose stools, dizziness, vomiting, dehydration, palpable spleen, swollen kidney, hepatosplenomegaly, mild abdominal ascites, pedal edema, low leukocytes, atypical blast cells and thrombocytopenia were also recorded.

The Table 1 shows that the parameters studied are linked to ALL, as the significance value is less than 0.05. All the parameters including anthropometric and sociodemographic, were related to ALL except patient age (p=0.4931), father age (p=0.051) creatine (p=0.1067), random glucose (p=0.1890), calcium (p=0.0663), and magnesium (p=0.1084).

Parameters	ALL (n=104)	normal (n=104)	p-value
Patient age	5.9 ± 4.3	6.3 ± 4.1	0.4931
Weight (kg)	18.01 ± 0.9	20 ± 8.5	0.0185*
Mother age	32.7 ±7.3	30.8 ± 4.0	0.0209*
Father age	38.1 ± 7.8	36.1 ± 6.9	0.0515*
Parental Smoking	0.29 ± 4.5	0.72 ± 4.52	0.4925
Radiation exposure	0	0	-
Bilirubin (mg/dl)	0.80 ± 0.1	0.4 ± 0.23	0.0001*
Creatine	0.584 ± 0.037	0.54 ± 0.25	0.1067
Urea (mg/dl)	21.96 ± 1.85	15 ± 5.36	0.0001*
ALK. Phosphate (U/L)	203.4 ± 12.2	156 ± 34.6	0.0001*
A.L.T (U/L)	50.85 ± 6.7	43 ± 6.0	0.0001*
Uric acid (mg/dl)	4.57 ± 0.7	4.98 ± 1.74	0.0269*
Albumin (g/dl)	6.97 ± 3.7	4.3 ± 0.9	0.0001*
Random glucose (mg/dl)	111.73 ± 10.4	106 ± 43.1	0.189
RBC (× 10^6/µl)	3.4 ± 0.1	5.25 ± 0.56	0.0001*
HGB (g/dL)	9.15 ± 0.26	13.9 ± 1.5	0.0001*
WBC (× 10^3/µl)	17.68 ± 4.87	8.7 ± 1.4	0.0001*
Platelets (× 10^3/µl)	134.09 ± 14.7	310 ± 57.9	0.0001*
CRP (mg/l)	22.2 ± 2.6	12 ± 1.2	0.0001*
Sodium (mEq/L)	135.3 ± 0.66	136 ± 3.1	0.0254*
Potassium (mEq/L)	3.7 ± 0.09	3.5 ± 0.1	0.0001*
Chloride (mEq/L)	102.38 ± 1.6	100.6 ± 2.1	0.0001*
Calcium (mg/dl)	10.86 ± 2.14	10.2 ± 3.0	0.0663
Magnesium (mg/dl)	2.1 ± 0.4	1.9 ± 1.2	0.1084
Phosphorous (mg/dl)	3.8 ± 0.13	2.7 ± 0.45	0.0001*
*Indicates significant association			

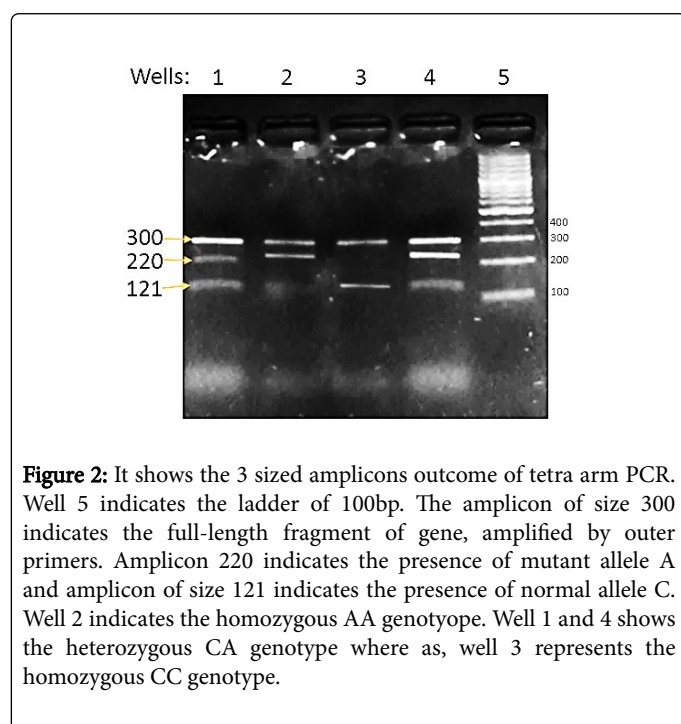**Table 1:** Anthropometric, sociodemographic and serum traits association with acute lymphocytic leukemia.

Genotyping, done by tetra arm PCR (Figure 2) were analyzed using statistically significant differences (p=0.001) in the genotype prevalence for the cases and controls, for single nucleotide polymorphism along with the odds ratio and confidence interval with 0.05 significance. Allelic frequency in cases was C=0.466 and A=0.534 showing an association with the disease by having chi-square p=0.0032 and Genotype frequency CC=0.26, CA=0.413, and AA=0.327, showing odds ratio 0.55(0.30 to 1.03) with a p=0.06 (Table 2).

Gene/SNP	Alleles/ Genotype	Controls %age	Cases %age	Odds ratio	95% CI	Chi square	P value
BCL2 rs2279115	C	60.05	46.6	------------	------------	8.7	0.0032*
	A	39.5	53.4				
	CC	42.3	26	0.55	0.30 to 1.03	3.52	0.06
	CA	37.5	41.3				
	AA	20.2	32.7				
*Indicates significant association							

Table 2: Allelic and genotypic frequencies in study subjects.

Figure 2: It shows the 3 sized amplicons outcome of tetra arm PCR. Well 5 indicates the ladder of 100bp. The amplicon of size 300 indicates the full-length fragment of gene, amplified by outer primers. Amplicon 220 indicates the presence of mutant allele A and amplicon of size 121 indicates the presence of normal allele C. Well 2 indicates the homozygous AA genotyope. Well 1 and 4 shows the heterozygous CA genotype where as, well 3 represents the homozygous CC genotype.

The serum traits were evaluated and associated with BCL2 rs2279115 polymorphism. The SNP is significantly associated with CRP level of ALL patients (p=0.037) and magnesium (p=0.002). However, it shows no significant relation with other parameters including biochemical and blood cell counts. The Table 3 shows the relation of different traits with polymorphism which is evaluated by independent T test and cross examined by one way ANOVA.

Parameters	Genotype	Mean ± Std. Deviation	p-value
Bilirubin (mg/dl)	CC	1.1 ± 1.4	0.266
	CA	0.7 ± 0.3	
	AA	0.8 ± 1.1	
Creatine (mg/dl)	CC	0.6 ± 0.4	0.585
	CA	0.6 ± 0.4	
	AA	0.5 ± 0.2	
Urea (mg/dl)	CC	22.3 ± 16.6	
	CA	20.9 ± 7.7	
	AA	23.1 ± 19.8	
ALK. Phs (U/L)	CC	188.8 ± 55.8	0.288
	CA	224.5 ± 142.3	
	AA	184.0 ± 65.1	
ALT (U/L)	CC	40.4 ± 36.6	0.528
	CA	58.8 ± 64.3	
	AA	46.1 ± 60.9	
Uric acid (mg/dl)	CC	3.5 ± 1.3	0.299
	CA	5.8 ± 6.2	
	AA	3.6 ± 1.4	
Albumin (g/dl)	CC	3.0 ± 0.6	0.283
	CA	3.4 ± 0.8	
	AA	16.9 ± 23.4	
Glucose (mg/dl)	CC	124.7 ± 28.0	0.627
	CA	100.3 ± 64.2	
	AA	120.1 ± 35.0	
RBC (× 10^6/µl)	CC	3.1 ± 1.2	0.331
	CA	3.4 ± 1.2	
	AA	3.7 ± 1.6	
HGB (g/dl)	CC	8.5 ± 2.6	0.347

	CA	9.2 ± 2.6	
	AA	9.6 ± 2.2	
WBC (× 10³/µl)	CC	35.8 ± 87.4	0.095
	CA	17.1 ± 34.7	
	AA	5.7 ± 5.8	
Platelets (× 10³/µl)	CC	179.7 ± 177.2	0.269
	CA	126.1 ± 142.0	
	AA	114.9 ± 116.5	
CRP (mg/l)	CC	15.5 ± 13.4	0.037*
	CA	26.7 ± 18.4	
	AA	31.1 ± 19.8	
Neutrophil (%)	CC	41.9 ± 30.5	0.371
	CA	30.9 ± 26.0	
	AA	30.9 ± 20.9	
Lymphocyte (%)	CC	42.7 ± 27.7	0.512
	CA	35.5 ± 22.2	
	AA	43.1 ± 24.9	
Monocyte (%)	CC	3.8 ± 4.1	0.488
	CA	5.7 ± 5.7	
	AA	5.4 ± 4.4	
Eosinophil (%)	CC	2.8 ± 1.9	0.494
	CA	4.3 ± 3.8	
	AA	4.5 ± 6.2	
Blast cells (%)	CC	79.0 ± 19.8	0.356
	CA	46.3 ± 33.5	
	AA	40.0 ± 30.8	
Sodium (mEq/L)	CC	134.1 ± 6.6	0.415
	CA	136.3 ± 5.0	

	AA	134.9 ± 5.5	
Potassium (mEq/L)	CC	4.0 ± 1.0	0.259
	CA	3.7 ± 0.8	
	AA	3.6 ± 0.5	
Chloride (mEq/L)	CC	102.4 ± 9.8	0.36
	CA	105.0 ± 5.2	
	AA	99.7 ± 17.3	
Calcium (mg/dl)	CC	8.6 ± 0.9	0.715
	CA	12.6 ± 20.2	
	AA	9.3 ± 0.7	
Magnesium (mg/dl)	CC	1.7 ± 0.1	0.002*
	CA	1.7 ± 0.2	
	AA	4.1	
Phosphorous (mg/dl)	CC	4.0 ± 0.6	0.863
	CA	3.8 ± 0.6	
	AA	3.8 ± 0.8	
*Indicates significant association			

Table 3: Association of BCL2 rs2279115 polymorphism with traits.

Drug likeliness

Phytochemicals were selected based on their anti-leukemic property. Ninety Phytochemicals present naturally in fruits, vegetables, herbs, spices, weeds and mushrooms were selected (supplementary Table). These phytochemicals were passed through the drug likeliness filters, which ended giving 20 phytochemicals that passed all the filters. They were selected on the property having IC_{50} value, Lipinski rule of five, LD_{50} value and blood brain barrier. Pharmacophore properties were also determined using ligand scout (Table 4). These 20 phytochemicals were docked with BCL2 protein at BH3 domain. Binding energies of each phytochemical were determined out of which the ones with more binding energy were selected for bonding analysis (Tables 5 and 6).

Sr#	Active Chemical Compound	Formula	Lipinski rule of five.						Toxicity value			Blood brain barrier		
			Mol. weight	H-bond acceptor	H-bond donor	Molar refractivity	LOGP	LD_{50}	class 4	class 5	BBB permeant	TPSA	Pgp substrate
1	AJOENE	C9H14OS3	234.4	1	0	67.41	2.1	1600 mg/kg	✓		No	86.88	No
2	ALOE-EMODIN	C15H10O5	270.24	5	3	69.92	0.1	5000 mg/kg		✓	No	94.83	No
3	APIGENIN	C15H10O6	286.24	6	4	76.01	-0.03	3919 mg/kg		✓	No	111.13	No
4	ASTRAGALIN	C21H20O11	448.38	11	7	108.13	-2.1	5000 mg/kg		✓	No	190.28	No
5	BAICALEIN	C15H10O5	270.24	5	3	73.99	0.52	3919 mg/kg		✓	No	90.9	No

#	Name	Formula	MW			PSA	logP	LD50					
6	CURCUMIN	C21H20O6	368.38	6	2	102.8	1.47	2000 mg/kg	✓		No	93.06	No
7	ELLAGIC ACID	C14H6O8	302.19	8	4	75.31	0.14	2991 mg/kg	✓		No	141.34	No
8	GALLIC ACID	C7H6O5	170.12	5	4	39.47	-0.16	2000 mg/kg	✓		No	97.99	No
9	GENISTEIN	C15H10O5	270.24	5	3	73.99	0.52	2500 mg/kg		✓	No	90.9	No
10	KAEMPFEROL	C15H10O6	286.24	6	4	76.01	-0.03	3919 mg/kg		✓	No	111.13	No
11	LUTEOLIN	C15H10O6	286.24	6	4	76.01	-0.03	3919 mg/kg		✓	No	111.13	No
12	MATAIRESINOL	C20H34O6	370.48	6	2	97.69	1.37	4400 mg/kg		✓	No	85.22	No
13	METHYL GALLATE	C8H8O5	184.15	5	3	43.79	0.18	1700 mg/kg	✓		No	86.99	No
14	RHEIN	C15H8O6	284.22	6	3	70.75	0.29	5000 mg/kg		✓	No	111.9	No
15	ROSMARINIC ACID	C18H16O8	360.31	8	5	91.4	0.9	5000 mg/kg		✓	No	144.52	No
16	S-ALLYMERCAPTOCYSTEINE	C6H11NO2S2	193.29	3	2	50.14	-1.93	660 mg/kg	✓		No	113.92	No
17	SCUTELLAREIN	C15H10O6	286.24	6	4	76.01	-0.03	3919 mg/kg		✓	No	111.13	No
18	SULFORAPHANE	C6H11NOS2	177.29	2	0	48.4	1.84	1000 mg/kg	✓		No	80.73	No
19	TRICIN	C17H14O7	330.29	7	3	86.97	-0.07	4000 mg/kg		✓	No	109.36	No
20	WOGONIN	C16H12O5	284.26	5	2	78.46	0.77	3919 mg/kg		✓	No	79.9	No

Table 4: Drug likeliness rules application upon phytochemicals.

Active Chemical Compound	Binding energies of *BCL2*								
	1	2	3	4	5	6	7	8	9
AJOENE	-3.6	-3.5	-3.5	-3.1	-3.1	-2.8	-2.8	-2.8	-2.8
ALOE-EMODIN	-5.4	-5.3	-5.1	-5.1	-3.9	-3.5	-3.5	-3	--------
APIGENIN	-5	-5	-4.9	-4.9	-4.8	-4.8	-4.8	-4.7	-4.2
ASTRAGALIN	-1.7	0.2	1.1	--------	--------	--------	--------	--------	--------
BAICALEIN	-5	-5	-4.8	-4.7	-4.6	-4.5	-4.4	-4.1	-3.7
CURCUMIN	-4.3	-3.6	-3.1	--------	--------	--------	--------	--------	--------
ELLAGIC ACID	-5.2	-5.2	-5.1	-4.9	-4.7	-4.6	-4.5	-4.4	-4.4
GALLIC ACID	-5.1	-4.9	-4.8	-4.8	-4.7	-4.6	-4.6	-4.5	-4.4
GENISTEIN	-5.4	-5.1	-4.9	-4.6	-4.3	-4	-3.2	-3	--------
KAEMPFEROL	-4.8	-4.6	-4.6	-4.5	-4.4	-4.2	-4.1	-3.8	-3.6
LUTEOLIN	-5.2	-5.1	-5	-4.9	-4.9	-4.9	-4.9	-4.5	-4.1
MATAIRESINOL	-4.1	-3.4	-2.9	-2.6	-2.4	-2.4	-2.1	-1.7	-1.7
METHYL GALLATE	-4.8	-4.8	-4.5	-4.5	-4.4	-4.3	-4.1	-3.6	-3.4
RHEIN	-5.3	-5.1	-5.1	-4.7	-4.5	-4.1	-3.7	-3.6	--------
ROSMARINIC ACID	-4	-3.7	-2.5	-1.3	-1	--------	--------	--------	--------

S-ALLYMERCAPTOCYSTEINE	-3.9	-3.8	-3.7	-3.6	-3.5	-3.5	-3.5	-3.3	-3.3
SCUTELLAREIN	-5.1	-5	-4.8	-4.7	-4.7	-4.6	-4.5	-4.5	-4.1
SULFORAPHANE	-3.7	-3.4	-3.2	-3.2	-3.2	-2.9	-2.9	-2.6	-2.6
TRICIN	-5.2	-5	-4.7	-4.6	-3.5	--------	--------	-------	--------
WOGONIN	-5.1	-4.9	-4.9	-4.9	-4.3	-3.9	-3.8	-3.7	-3.7

Table 5: Binding energies of phytochemicals docked with BCL2 protein.

Active Chemical Compound	Confirmations								
	1	2	3	4	5	6	7	8	9
ALOE-EMODIN	--------	SER 51 CO-OH 3.5	SER 51 CO-OH 3.2	--------	--------	--------	--------	LEU 119 CO-OH 2.4 THR 125 CO-OH 3.1 CO-OH 2.4 ARG 129 CN-OH 3.1 THR 132 CO-OH 2.4	--------
APIGENIN	ARG 146 CN-OH 3.1 ARG 139 CN-OH 3.4	--------	--------	ARG 139 CN-OH 3.3 CN-OH 3.1 CN-OH 2.8	--------	ASP 140 CO-HO 2.3 ARG 139 CN-OH 3.1	ARG 146 CN-OH 3.3	--------	ARG 139 CN-OH 2.8
BAICALEIN	ARG 139 CN-OH 3.0 CN-OH 3.4 CN-OH 2.9 ASP 140 CO-HO 2.6 ARG 146 CN-OH 3.2	--------	ASP 140 CO-HO 2.2 CO-HO 2.4 ARG 139 CN-OH 3.2 SER 51 CO-OH 2.9	--------	--------	SER 51 CO-OH 3.2	--------	--------	--------
ELLAGIC ACID	--------	--------	--------	ASP 140 CO-OH 3.4 CO-HO 2.7 ARG 139 CN-OH 2.8 CN-OH 2.9 CN-OH 3.0 ARG 146 CN-OH 3.4	GLU 136 CO-OH 3. ARG 139 CN-OH 3.0 CN-OH 3.1 CN-OH 3.4	ARG 139 CN-OH 3.1 CN-OH 3.0 ASP 140 CO-HO 2.5 CO-OH 3.4	--------	ASP 140 CO-HO 2.3 ARG 139 CN-OH 2.8 CN-OH 3.0 GLU 136 CO-HO 2.5 SER 51 CO-OH 3.3	ASP 140 CO-HO 2.7 ARG 139 CN-OC 3.3
GALLIC ACID	TRP 176 CO-OH 2.7 ARG 127 CN-OH 3.3	GLU 136 CO-OC 2.8 ARG 127 CN-OH 3.3	ARG 146 CN-OH 2.8 ASP 140 CO-HO 2.0 CO-HO 2.0 LEU 137	ARG 139 CN-OH 3.1 CN-OH 3.1 CN-OH 3.1 ASP 140 CO-HO 2.1	ARG 127 CN-OH 3.1 CN-OH 3.3 CN-OH 2.7	ARG 146 CN-OC 3.1 ASP 140 CO-HO 2.3 ARG 139 CN-OH 3.2	ARG 139 CN-OH 3.2 CN-OH 3.1 CN-OH 3.0 GLU 136 CO-HO 2.2	PHE 104 CO-HO 2.6 SER 51 CN-OH 2.9	SER 51 CN-HO 3.3

		CO-HO 2.7 GLU 136 CO-HO 1.9 ARG 139 CN-OH 3.1 CN-OH 3.1 CN-OH 3.1	ARG 146 CN-OH 2.8			CN-OH 3.1 CN-OH 3.0	ARG 146 CN-OC 3.1		
GENISTEIN	--------	SER 51 CO-OH 3.2	--------	--------	ARG 139 CN-OH 3.0 ARG 146 CN-OH 3.1	--------	--------	ARG 139 CN-OH 3.1	--------
LUTEOLIN	THR 125 CO-OH 2.7 CO-OH 2.6 ARG 129 CN-OH 3.1	ARG 139 CN-OH 3.2 CN-OH 3.0	THR 132 CO-OH 3.0 GLN 118 CO-OH 3.1	SER 51 CO-OH 2.8 GLU 136 CO-HO 2.2 ARG 139 CN-OH 3.0	ARG 139 CN-OH 3.0 SER 51 CO-OH 2.9	ARG 146 CN-OH 3.3	SER 51 CO-OH 3.4 ARG 139 CN-OH 3.0 CN-OH 2.9	ARG 146 CN-OH 2.9	ASP 140 CO-HO 2.4 SER 51 CO-OH 3.3
RHEIN	---------	---------	---------	---------	---------	---------	ASP 140 CO-HO 2.5 ARG 146 CN-HO 3.3	ARG 146 CO-OH 2.9 ARG 139 CN-HO 2.9	---------
SCUTELLAREIN	ARG 139 CN-OH 3.0 SER 51 CO-OH 2.9	ARG 139 CN-HO 3.1	---------	---------	ARG 139 CN-OH 2.9	---------	ARG 139 CN-OH 3.6	---------	---------
TRICIN	ARG 146 CN-OC 3.2 ARG 139 CN-OH 2.9	ARG 146 CN-OC 3.2 ARG 139 CN-OC 3.3 ASP 140 CO-HO 2.6 SER 51 CO-OH 3.4	ARG 146 CN-OC 3.3	---------	---------	---------	---------	---------	---------
WOGONIN	SER 51 CO-OC 3.5 CO-OH 3.0 ARG 139 CN-OH 3.1 ASP 140 CO-HO 2.8	---------	ARG 146 CN-OH 3.3 SER 51 CO-OC 3.0 CO-OH 3.4	SER 51 CO-OH 3.5 ARG 139 CN-OH 3.2 CN-OH 3.1 CN-OH 2.6 ARG 146 CN-OH 3.4 CN-OC 3.2	---------	---------	---------	---------	---------

Table 6: Hydrogen bonding between ligands and amino acids of protein target site.

Almost 11 phytochemicals were found with about 5-6 binding energy and confirmations range from 2-9. Comparison was made based on energy and IC_{50} value. We selected Ellagic acid as the main phytochemical based upon its rest of drug likeliness properties and hydrogen bonding with the amino acids of *BCL2* target site BH3.

Discussion

The study conducted provides the allele and a genotype frequency of *BCL2* promoter rs2279115 polymorphism that is previously studied in other Asian populace but no relevant data is available of this single

nucleotide polymorphism for Pakistani population. We intended to locate its association, if any, with biochemical parameters to draft a probable mechanism of action. The pervious genotyping results in various populations about this variant are conflicting that whether it is associated with onset of acute lymphoblastic leukemia or not. Older studies revealed that parental smoking and the time duration since smoking elevates the risk of developing ALL in children, on the other hand another study shows a contradiction related to maternal or parental cigarette smoking being a risk [22]. Some studies found an association between parental exposure to radiation before conception, that may be due to their work [23] or the X-rays and ultrasound done [24]. In our analysis, smoking of parents was not associated with increased risk of ALL and none of the patient or their parents were ever exposed to radiation. In Pakistan, the concept of consanguinity marriages is still present especially in rural areas from where most of the cases were reported, because of consanguinity marriages, the minor alleles pool up and becomes a cause of genetic diseases. Mother and father education was also taken in consideration that if they were educated they would be more anxious about the health issue of their child and get them treated by time. Many cases were reported at initial stages of cancer. Patients of ALL suffers from organs enlargements namely liver, kidney, and spleen. These are the blood cells forming and degrading organs where blood cells are filtered based upon their functional ability. Accumulation of malfunctioned and immature white blood cells resulted in these organs enlargement. To determine the capability of organs functioning, biochemical parameters were taken in concern. All the biological parameters including hematology and body salts studied were found to be associated with ALL as also indicated in previous study [25,26]. The association between the allelic frequency and the study groups, differentiated into cases and controls was evaluated by crosstabs. Logistic regression and chi square was also used to analyze the association SNPs polymorphism with acute lymphocytic leukemia showing significant p-value less than 0.05. Presence of A allele enhances the risk of one to have ALL. However, the odds ratio shows that the allele is protective but when we consider the p-value, it is not significant so we can say that if we study a larger sample size it would be possible that we find rare allele as the risk allele. The risk increasing allele A, when A is present in the genotype the value of CRP increases CC=15.5, CA=26.7 and AA=31.1 (mg/l). Similarly magnesium level also rises CC=1.7, CA=1.7 and AA=4.1 (mg/dL).

Raptorx software used to build 3D structure is a gateway for protein structure and function prediction. It predicts the tertiary structure of a protein sequence using templates. The 3D structures were confirmed using the SAVES v.4., a web server consisting of PROCHECK that determines the stereo chemical quality of a protein and VERYFY_3D, which determines the compatibility of atomic models in amino acids of protein by assigning structural class and comparison to good structures. In SNP rs2279115, cytosine (pyrimidine) nucleotide is replaced with adenine (purine) nucleotide at position 4777. This change in nucleotide does not alter the structure of the protein but effects the gene expression, as it has reported in previous studies [10]. It is assumed that the presence of adenine enhances the promoter to regulate transcription due to which cells undergo premature proliferation passing through the cell checkpoints for mutation or faulty cell formation.

Ninety Phytochemicals present naturally in fruits, vegetables, herbs, weeds and mushrooms were selected on their pharmacokinetic and pharmacodynamics properties. Almost 11 phytochemicals were found with about 5-6 binding energy and confirmations range from 2-9. Based on energy we selected Ellagic acid as the main phytochemical

based upon its rest of properties like IC_{50} value=1 μM as shown in the graph (Figure 3). IC_{50} stands for the inhibitory concentration or the half-maximal inhibitory concentration that measures the effectiveness of the drug in hindering the target pathway or the biological function for which it is meant.
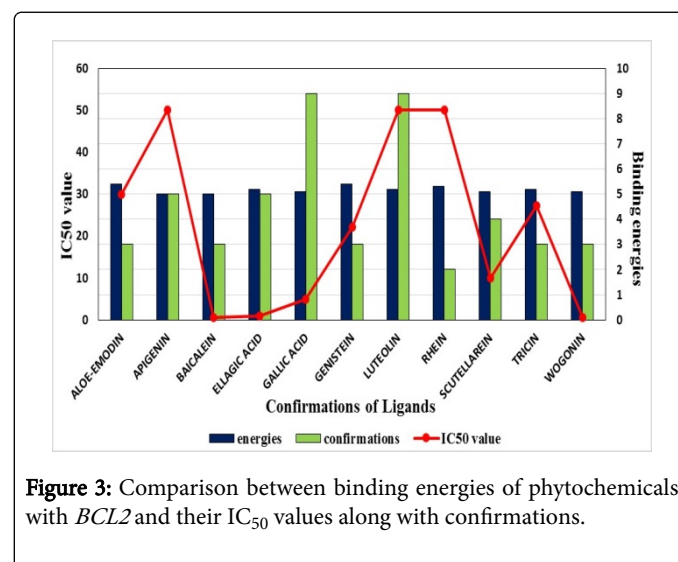

Figure 3: Comparison between binding energies of phytochemicals with *BCL2* and their IC_{50} values along with confirmations.

It is commonly isolated from strawberries, walnuts [27], and pomegranate [28] commonly found in Pakistan. Ellagic acid obeys Lipinski rule of five. Molecular mass should be less than 500 Dalton, High lipophilicity (expressed as Log P less than 5), hydrogen bond donors less than 5, hydrogen bond acceptors less than 10, and Molar refractivity in-between 40-130 [29]. Here molecular weight determines the movement of the molecule from the pores of different membranes while the log P is an important physiochemical property that measures the partition coefficient between water and n-octanol required for well adsorption and rapid metabolism. Solubility of drug in water can be assumed by the number of hydrogen bond donors compared to alkyl sidechains present in the molecule. Low water solubility represents slow absorption of drug and low action. Low molecular weight drugs are more selected because the molecular weight of compound directly effects the diffusion. Usually marketed drugs have molecular weight between 200 to 600 Daltons, however Lipinski rule indicates that molecular weight should be less than 500 Dalton. Ellagic acid have molecular weight 302.19 g/mol, 8 hydrogen bond acceptors, 4 hydrogen bond donors, 75.31 molar refractivity, and 0.14 logp. It comprises of 22 heavy atoms, 4 aromatic rings, 6 rotatable bonds, 16 aromatic ring atoms, and do not consist of any anionic or cationic ion and hydrophobic region. The rotatable bonds in pharmacophore model deals with the bioavailability of the compound. Less than 7 rotatable bonds are considered to have a good bioavailability. Aromatic rings are important for the structure of a drug for its stability, polarity, and hydrophobicity, not changing the solubility of the drug. Along with it, biological molecules present in the body are ring structures usually so when dealing with target drugs its useful to select phenyl group caring compounds.

It shows low class 4 toxicity as compared to other phytochemicals. Toxicity of a compound is measured as the lethal dose (LD_{50}). LD_{50} values are in mg/kg body weight. Ellagic acid do not crosses the blood brain barrier it remains in blood stream with entering to other tissues and brain with topological surface area 141.34 and it is not a pgp substrate (multi-drug resistant protein). TPSA stands for the

topological polar surface area, which controls the movement of the drug in different organs of the body. Compounds having TPSA<140 Å2 have a good intestinal absorption. While the compounds having TPSA<70 Å2 have good brain penetration. Blood brain barrier is a highly selective brain barrier that prevents the circulating blood from entering the brain [30]. For normal drugs it is important that it do not reaches the brain instead being designed for central nervous system (CNS). The permeability glycoprotein is abbreviated as Pgp. It is an important protein, located on cell membrane, which pumps the foreign substances out of the cell. Pgp is important factor in pharmacokinetics. It reduces the efficacy of drugs for ADME properties. Moreover, cancer cells express more Pgp as it helps in multi-drug resistance [31]. A protein 3D structure is stabilized to some extent by hydrogen bonding. Binding of ligand to protein involves the breakdown of these bonds and formation of new bonds between protein and ligand. This basic phenomenon facilitates the ligand binding to target protein and inhibit its biological activity. Formation of hydrogen bond between protein and ligand depicts stronger interaction. Hydrogen bond is an electrostatic force between two polar groups that arises due to covalent bonding of hydrogen atom with highly electronegative atom like nitrogen, oxygen, or fluorine. Figure 4 shows its best confirmation hydrogen bonding with the amino acids of *BCL2* target site BH3.
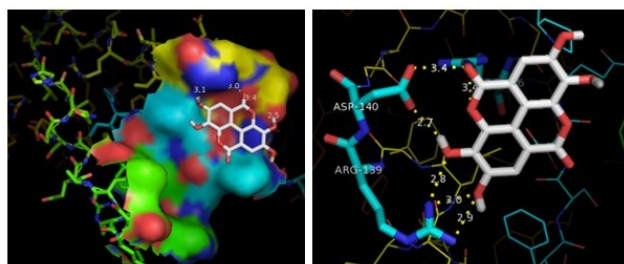

Figure 4: BCL2 ligand binding pocket (solid surface) on left and right one is showing hydrogen bonds between active cite amino acids and Ellagic acid phytochemical.

Conclusion

Study of rs227195 polymorphism is not done in Pakistani populace in past. Our results are somehow similar to previous studies done on different populations but at some points are results differ, it may be due to difference in gene pool and different expression. Considering the position of SNP at the promoter, we could assume it as increasing the expression of protein resulting in immature leukocytes causing leukemia. The increased expression increases the anti-apoptotic sites where anti apoptotic proteins binds hence preventing the faulty cells from going under apoptosis and allows them to accumulate in blood stream. To deal with a problem we screened some natural phytochemicals having anti leukemic property that would bind specifically to the BH3 domain of BCL2 where anti apoptotic proteins binds. The compound identified for SNP can be further checked in laboratory ALL cell lines to analyze their effect on the cells with genome variants and the effect of variants on the drug metabolism including study of other factors like mutagenicity or teratogenic.

Author Contributions

Huma Butt performed experiments, analyzed results, and wrote the manuscript. Ammara Khalid supervised the study and proofread the manuscript. Both authors read and approved the final manuscript.

Conflict of Interest

The authors declare that they have no competing interests.

References

1. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, et al. (2013) Clinical whole-exome sequencing for the diagnosis of mendelian disorders. New England Journal of Medicine 369: 1502-1511.

2. Mohamed RMAS (2014) Using Probability Binning and Bayesian Inference to measure Euclidean Distance of Flow Cytometric data.

3. Chen M, Chang CH, Tao L, Lu C (2015) Residential exposure to pesticide during childhood and childhood cancers: A meta-analysis. Pediatrics 136: 719-729.

4. Nasir M, Jabeen F, Hussain SM, Shaheen T, Samiullah K, et al. (2015) Impact of Consanguinity, Environment, Socio-Economic and Other Risk Factors on Epidemiology of Leukemia. Pakistan J Zool 47: 1117-1124.

5. Campana D, Pui CH (2014) Childhood leukemia. Abeloff's Clinical Oncology (5thedn) Elsevier Saunders, Philadelphia, PA.

6. Kang MH , Reynolds CP (2009) Bcl-2 inhibitors: targeting mitochondrial apoptotic pathways in cancer therapy. Clin Cancer Res 15: 1126-1132.

7. Campos L, Rouault JP, Sabido O, Oriol P, Roubi N, et al. (1993) High expression of bcl-2 protein in acute myeloid leukemia cells is associated with poor response to chemotherapy. Blood 81: 3091-3096.

8. Li A, Ojogho O, Escher A (2006) Saving death: apoptosis for intervention in transplantation and autoimmunity. Clin Dev Immunol 13: 273-282.

9. Oltvai ZN, Milliman CL, Korsmeyer SJ (1993) Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programed cell death. Cell 74: 609-619.

10. Majid A, Tsoulakis O, Walewska R, Gesk S, Siebert R, et al. (2008) BCL2 expression in chronic lymphocytic leukemia: lack of association with the BCL2– 938A> C promoter single nucleotide polymorphism. Blood 111: 874-877.

11. Coustan-Smith E, Kitanaka A, Pui CH, McNinch L, Evans WE, et al. (1996) Clinical relevance of BCL-2 overexpression in childhood acute lymphoblastic leukemia. Blood 87: 1140-1146.

12. Moazami-Goudarzi M, Farshdousti-Hagh M, Hoseinpour-Feizi A, Talebi M, Movassaghpour-Akbari AA, et al. (2016) The acute lymphoblastic leukemia prognostic scoring whether it is possible by BCL-2, BAX gene promoter genotyping. Caspian J Intern Med 7: 105.

13. Stelmach P, Wedemeyeret C, Fuest L, Kurscheid G, Gehrke T, et al. (2016) The BCL2-938C> A Promoter Polymorphism Is Associated with Risk for and Time to Aseptic Loosening of Total Hip Arthroplasty. PloS one 11: e0149528.

14. Singh A, Singh S, Wahi M (2013) In Silico Structure Based Drug Designing of A Potent Inhibitor for Purine Nucleoside Phosphorylase A Therapeutic Target for Schistosomiasis. International Journal of Bioinformatics and Biological Sciences 1: 263.

15. Lengauer T, Rarey M (1996) Computational methods for biomolecular docking. Curr Opin Struct Biol 6: 402-406.

16. Yasmeen N, Ashraf S (2009) Childhood acute lymphoblastic leukaemia; epidemiology and clinicopathological features. JPMA 59: 150.

17. Duke JA (1994) Duke's phytochemical and ethnobotanical databases.

18. Bolton EE (2008) PubChem: integrated platform of small molecules and biological activities. Annual reports in computational chemistry 4: 217-241.

19. Pence HE, Williams A (2010) ChemSpider: an online chemical information resource. ACS Publications.

20. Drwal MN, Banerjee P, Dunkel M, Wettig MR, Preissner (2014) ProTox: a web server for the in silico prediction of rodent oral toxicity. Nucleic acids research 42: W53-W58.

21. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of comput Chem 31: 455-461.

22. Belson M, Kingsley B, Holmes A (2007) Risk factors for acute leukemia in children: a review. Environ Health Perspect 115: 138.

23. Gardner MJ (1991) Father's occupational exposure to radiation and the raised level of childhood leukemia near the Sellafield nuclear plant. Environ Health Perspect 94: 5-7.

24. Shu XO, Potter JD, Linet MS, Severson RK, Han D, et al. (2002) Diagnostic X-rays and ultrasound exposure and risk of childhood acute lymphoblastic leukemia by immunophenotype. Cancer Epidemiol Biomarkers Prev 11: 177-185.

25. Rasool M, Farooq S, Malik A, Shaukat A, Manan A, et al. (2015) Assessment of circulating biochemical markers and antioxidative status in acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) patients. Saudi J Biol Sci 22: 106-111.

26. Rafiq N, Iqbal T, Shahid M, Muhammad F (2014) Hematological and biochemical parameters in Pakistani chronic lymphoblastic leukemia patients. Pakistan Journal of Life & Social Sciences 12: 16-19.

27. Daniel EM, Krupnick AS, Heur YH, Blinzler JA (1989) Extraction, stability, and quantitation of ellagic acid in various fruits and nuts. Journal of Food Composition and Analysis 2: 338-349.

28. Yoshimura M, Watanabe Y, Kasai K, Yamakoshi J, Koga T (2005) Inhibitory effect of an ellagic acid-rich pomegranate extract on tyrosinase activity and ultraviolet-induced pigmentation. Bioscience, biotechnology, and biochemistry 69: 2368-2373.

29. Ghose AK, Viswanadhan VN, Wendoloski JJ (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. J Comb Chem 1: 55-68.

30. Bradbury MW (1979) The concept of a blood-brain barrier.

31. Szakács G, Paterson JK, Ludwig JA, Booth-Genthe C, Gottesman MM (2006) Targeting multidrug resistance in cancer. Nat Rev Drug Discov 5: 219-34.