



Title of Thesis

“Computational analysis and compilation of leukemia biomarkers.”

by
Anjali Walia

Under the Supervision of Prof. Gajendra P.Raghava

Indraprastha Institute of Information Technology Delhi
June, 2020

Inner Cover Page

Page No.
2



Title of Thesis

“Computational analysis and compilation of leukemia biomarkers.”

by
Anjali Walia

Submitted
in partial fulfillment of the requirements for the degree of
Master of Technology

to

Indraprastha Institute of Information Technology Delhi
June, 2020

Certificate

This is to certify that the thesis titled “**Computational analysis and compilation of leukemia biomarkers.**” being submitted by Anjali Walia to the Indraprastha Institute of Information Technology Delhi, for the award of the Master of Technology, is an original research work carried out by her under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree/diploma.

June, 2020

Prof.Gajendra P Raghava
Department of Computational Biology
Indraprastha Institute of Information Technology Delhi
New Delhi 110 020

Acknowledgement

I would like to express my sincere gratitude and respect towards Prof. Gajendra P S Raghava from Indraprastha Institute of Information and Technology, Delhi for being my supervisor and for exposing me to this wonderful topic of research and guiding me throughout. Besides my supervisor, I would like to thank the Ph.D. scholars from his esteemed lab Harpreet Kaur for her constant guidance, support, and motivation throughout the project. I would also like to thank the Centre for Computational Biology and IT administrators at IIIT Delhi for providing me with all the necessary resources. Lastly, I would also like to thank my family and friends for providing much-needed support and motivating me from time to time throughout the course of my thesis which enabled me to pursue my research in an efficient and structured manner.

Anjali Walia

M.Tech (CB)

Table of Contents

Abstract	8
Introduction	9
Systems and Methodology	13
Study Design	13
Data curation	14
LeukemiaBD: Architecture and Interface	17
Database Organization	19
Results	20
Web tools Implementation	20
A. Search tools	20
B. Browse tools	21
LeukemiaBD statistics	22
Availability	26
Working of the database	27
Discussion	33
Applications of database	34
References	35

List of Figures

Figure 1: Study design

Figure 2: Architecture of LeukemiaBD

Figure 3: Distribution of biomarkers based on biomarker type.

Figure 4: Distribution of biomarkers based on biomolecule type.

Figure 5: Distribution of biomarkers based on source type.

Figure 6: LeukemiaBD

Figure 7: Browse biomarker based on different categories i.e type,source,biomolecule etc.

Figure 8: Browse biomarker based on biomarker type.

Figure 9: Example of a keyword Search in Leukemia BD

Figure 10: Keyword search Query results

Figure 11: Example of a Complex Search in Leukemia BD

Figure 12: Complex search Query results

Abstract

LeukemiaBD, is a Database of Biomarker data for Leukemia. The aim of this database is to serve as a comprehensive repository catering to all relevant information related to Leukemia. The database currently has manually curated information about the leukemia biomarkers. It comprises an interactive interface to query each biomarker via an easy searching and browsing facility. Individual information about each biomarker comprises the associated type of leukemia, patient cohort, regulation of biomarker, type of biomarker, type of biomolecule, source, etc. We believe that LeukemiaBD will be a valuable resource for the scientific community to explore the leukemia biomarker. LeukemiaBD is built on responsive templates which are compatible with smartphones and various other gadgets (mobile, iPhone, iPad, tablets etc.).

Database URL: webs.iiitd.edu.in/raghava/leukemiabd

Objectives

1. Collection and compilation of biomarkers for leukemia from the literature.
2. Development of web-server(LeukemiaBD) to organize Biomarker information in user friendly format using Apache HTTP server on Linux Platform using the most popularly available open-source relational database management system (RDBMS), MySQL version 5.7.24.

Introduction

Leukemia is a cancer of the blood cells. It is basically a clonal proliferation of hematopoietic stem cells in the bone marrow.[2] Leukemia could be developed due to a problem with the blood cell production. It usually affects the white blood cells or leukocytes. Leukemias fall under the group of life-threatening malignant disorders of the blood and bone marrow. There is not an exact known cause of leukemia. But it seems to develop from a combination of genetic and environmental factors.[3] Basically, leukemia is thought to occur when some blood cells acquire mutations in their DNA. It develops when the DNA of developing blood cells, mainly white cells, incurs damage. This, in turn, causes the blood cells to grow and divide uncontrollably. As the healthy blood cells die, new cells replace them. These develop in the bone marrow. But the abnormal blood cells do not just die at a natural point in their life cycle. But, they build up and occupy more space. So, as the bone marrow produces more cancer cells, they start to overcrowd the blood, thus, preventing the healthy white blood cells (WBCs) from growing and functioning normally. Eventually, the cancerous cells outnumber healthy cells in the blood. In terms of how quickly it develops and gets worse, the onset of leukemia can be acute (sudden onset or fast-growing) or chronic (slow onset or slow-growing). In acute leukemia, cancer cells multiply rapidly and result in the accumulation of immature and functionless blood cells in the bone marrow. Whereas chronic leukemia progresses slowly, which results in the accumulation of mature, but still abnormal, white blood cells and the early symptoms may be very mild in this case. Further, Leukemia is also classified according to the type of cell involved. Leukemia involving myeloid cells is called myelogenous leukemia. And, leukemia involving lymphocytes

is called lymphocytic leukemia. [4] Acute lymphoblastic leukemia (ALL) occurs more often in children, whereas the others are more common in adults. The significant risk factors included here are genetic predisposition as well as environmental factors, such as exposure to ionizing radiation. The four broad subtypes of Leukemia are acute lymphoblastic, acute myeloid, chronic lymphocytic, and chronic myeloid leukemia. Acute lymphocytic leukemia (ALL) progresses rapidly by replacing the healthy cells that produce functional lymphocytes with the leukemia cells that can't mature properly. The leukemia cells are carried away in the bloodstream to other tissues and organs, including liver, lymph nodes, brain and testes, where they continue to grow and divide. The building, dividing and spreading of these leukemia cells may result in several possible symptoms.[5] Acute myeloid leukemia (AML) is known by several other names such as acute myelogenous leukemia, acute myeloblastic leukemia, acute granulocytic leukemia or acute nonlymphocytic leukemia. It is a fast-growing form of cancer of the blood and bone marrow. [6] Chronic lymphocytic leukemia (CLL) is typically a type of slow-growing cancer that begins in lymphocytes in the bone marrow and extends to the blood. It also gets spread to lymph nodes and organs such as the liver and spleen. It develops when too many abnormal lymphocytes grow, crowding out normal blood cells and making it difficult for the body to fight infection.[7] Chronic myeloid leukemia (CML) is also known as chronic myelogenous leukemia which begins in the blood-forming cells of the bone marrow and then with time spreads to the blood gradually. And eventually, the disease spreads to other areas of the body.[8]

Treatment for leukemia depends on many factors like age, overall health, the type of leukemia, and whether or not it has spread to other parts of the body. Common treatments used for leukemia are chemotherapy, biological therapy, targeted therapy, radiation therapy, stem cell

transplant and surgery. Chemotherapy is the primary form for the treatment of leukemia. Chemotherapy treatment uses chemicals to kill leukemia cells. Also, it is the primary treatment for Acute myeloid leukemia (AML). Sometimes, doctors may recommend a bone marrow transplant. Depending on the type of leukemia one has, one may be treated with single-drug treatment or a combination of drugs. The drugs used in this treatment may come in a pill form, or they may be injected directly into the vein. Biological therapy works by using treatments that help the immune system to recognize and attack leukemia cells. Targeted therapy uses those drugs that attack specific vulnerabilities within the cancer cells. This type of treatment uses tyrosine kinase inhibitors that target cancer cells without affecting other cells, reducing the risk of side effects. Examples include imatinib, dasatinib, and nilotinib. Radiation therapy uses X-rays or other high-energy beams to destroy leukemia cells or bone marrow tissue and stop their growth. Radiation therapy is usually used before the transplant or to prepare for a stem cell transplant. Surgery often involves removing the spleen, but this depends on the type of leukemia. In the stem cell transplant,, diseased bone marrow is replaced with healthy bone marrow. Before performing a stem cell transplant, one may receive high doses of chemotherapy or radiation therapy to destroy the diseased bone marrow. After that, one may receive an infusion of blood-forming stem cells that help to rebuild the bone marrow. A stem cell transplant is similar to a bone marrow transplant.[9]

To have a better understanding of leukemia we need to study the genes/proteins that are regulated in the disease condition. The term “biomarker”, a portmanteau of “biological marker”, refers to a broad subcategory of medical signs – that is, objective indications of medical state observed from outside the patient – which can be measured accurately and reproducibly [10].

Different types of biomarkers are prognostic, diagnostic and predictive. Diagnostic biomarkers are used to confirm that a patient has a particular health disorder [11]. Prognostic Biomarkers shows how a disease may develop in a person when a disorder is already diagnosed. The absence or presence of prognostic markers is useful for the selection of patients for treatment but it does not directly predict the response to a treatment [12]. Predictive Biomarkers helps to determine which patients are most likely to benefit from a specific treatment option. It provides information about how well a treatment is expected to work in case of a particular patient and about the likelihood of that treatment, causing an unwanted side effect [13,14].

LeukemiaBD: A Database of Biomarker data of Leukemia. This is a contribution to the scientific community who have been actively researching the area of leukemia biomarkers. LeukemiaBD is a comprehensive information system which serves as a common portal to address and understand leukemia biomarkers. It comprises all the relevant information about the biomarkers from principal data repositories like NCBI (The National Center for Biotechnology Information) PubMed. This resource has various web-based tools for searching and browsing facilities to facilitate users in extracting and analyzing the data. We hope that this portal will be a useful contribution to divert more attention to this field of research and serve as one-stop for all the information related to these biomarkers.

Systems and Methodology

Study Design

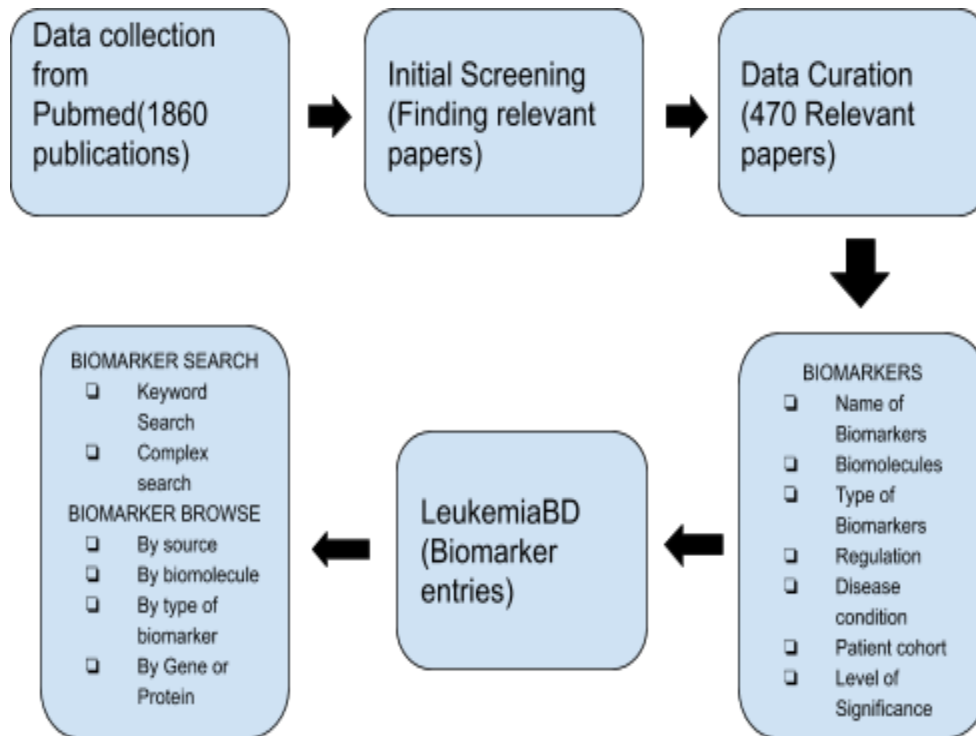


Figure 1: Study design

Data curation

The data were manually annotated, covering all aspects of the leukemia biomarkers from broad perspectives, namely Diagnostic Biomarkers, Prognostic Biomarkers & Predictive Biomarkers. For the annotation of data, we used NCBI (The National Center for Biotechnology Information) PubMed. To filter out the articles outside our targeted area of research literature, PubMed was queried using the keywords “(leukemia[Title/Abstract]) AND biomarker[Title/Abstract],(leukemia[Title/Abstract]) AND signature[Title/Abstract],(blood cancer[Title/Abstract]) AND signature[Title/Abstract],(blood cancer[Title/Abstract]) AND biomarker[Title/Abstract]”. After sorting the queries further based on relevant keywords and identifying recent publications from the last 15 years we were able to narrow down information from 1860 publications published until 2019. During the initial screening, review articles and the articles lacking relevant information with respect to leukemia biomarkers were excluded. Based on the screening that we did we narrowed down to 470 relevant papers out of the total 1860 research papers. Finally the data was curated from 470 relevant research papers. We manually retrieved information from the research papers. The information to be retrieved from the relevant research papers are PMID, year, biomarker, biomolecule, source, subjects, regulation in cancerous condition, odds ratio, effect on pathway, experiment, type of biomarker, cohort, sensitivity, specificity, accuracy, p-value and validity. PMID is the Pubmed ID of the paper. It is the unique identifier number used in PubMed that can be used to look up abstracts or articles in PubMed. Year refers to the year in which research paper was Published. Biomolecule is basically whether it's a gene, protein, Enzyme, Hormone etc all are biomolecules, so it is a term

for molecules and ions present in organisms that are essential to one or more typically biological processes, such as cell division, development etc. Source is the sample from which it is taken from like Bone marrow, Serum, Blood etc. Subjects refers to whether the patients under study were Humans, Mice, Rat. Regulation in cancerous condition means whether the biomarker is up or down Regulated (including fold change). Gene regulation comprises a wide range of mechanisms that are used by cells to either increase or to decrease the quantity of specific gene products such as protein or RNA. Down-regulation is the process by which a cell decreases the production of specific gene products in response to an external stimulus. And upregulation involves increases of such components.[15]Odds Ratio/ Hazard Ratio describes the relative risk of the complication based on the comparison of event rates. It is also used to describe the outcome of therapeutic trials where the main focus is to extent treatment can shorten the duration of the illness.[16]Effect on pathways is when the expression of the particular biomarker regulates a particular pathway in cancer.Experiment refers to the conditions in which a biomarker is differentially expressed. A gene is said to be a differentially expressed gene if an observed difference or change in read counts or expression levels between two experimental conditions is statistically significant.[17] The types of biomarker are diagnostic, prognostic or predictive. Cohort is the description of the dataset. The sensitivity Sensitivity of the biomarker is also known as the true positive rate, it measures how often a test correctly generates positive result for those people who have the condition that they were tested for.Specificity of the biomarker is also known as the true negative rate, it measures a test's ability to correctly generate a negative result for people who don't have the condition that's being tested for.[17] AUC isArea Under Curve of the Biomarker. It is a useful tool to evaluate the diagnostic accuracy. Many biomarker

combination methods rely on the maximization of the area under the ROC curve (AUC).[18] Accuracy of the Biomarker refers to the closeness of a measured value to a standard or known value.p-value describes how statistically significant was the differentiation.[19] The p-value or probability value is the probability of obtaining results at least as extreme as the results observed of a statistical hypothesis test, assuming that the null hypothesis is correct.Validity refers to the fact that how good a biomarker it is whether it was validated on some internal or external cohort.[20]

LeukemiaBD: Architecture and Interface

LeukemiaBD was built using the Apache HTTP server on Linux Platform. The front end was maintained using CSS3, HTML5, XAMPP, PHP version 4.0.10 and Javascript (1.7).

At the back-end, Relational Database Management system (RDBMS): MySQL version 5.7.24, was used to organize all the data into the table. It is the most popularly available open-source Database Management system. The tabular organization of data allows better and efficient data retrieval and thus, clear visualization on the interface. The main table with 25 different fields was maintained associated with each of the unique entries. The overall architecture of the LeukemiaBD is represented in Figure 2.

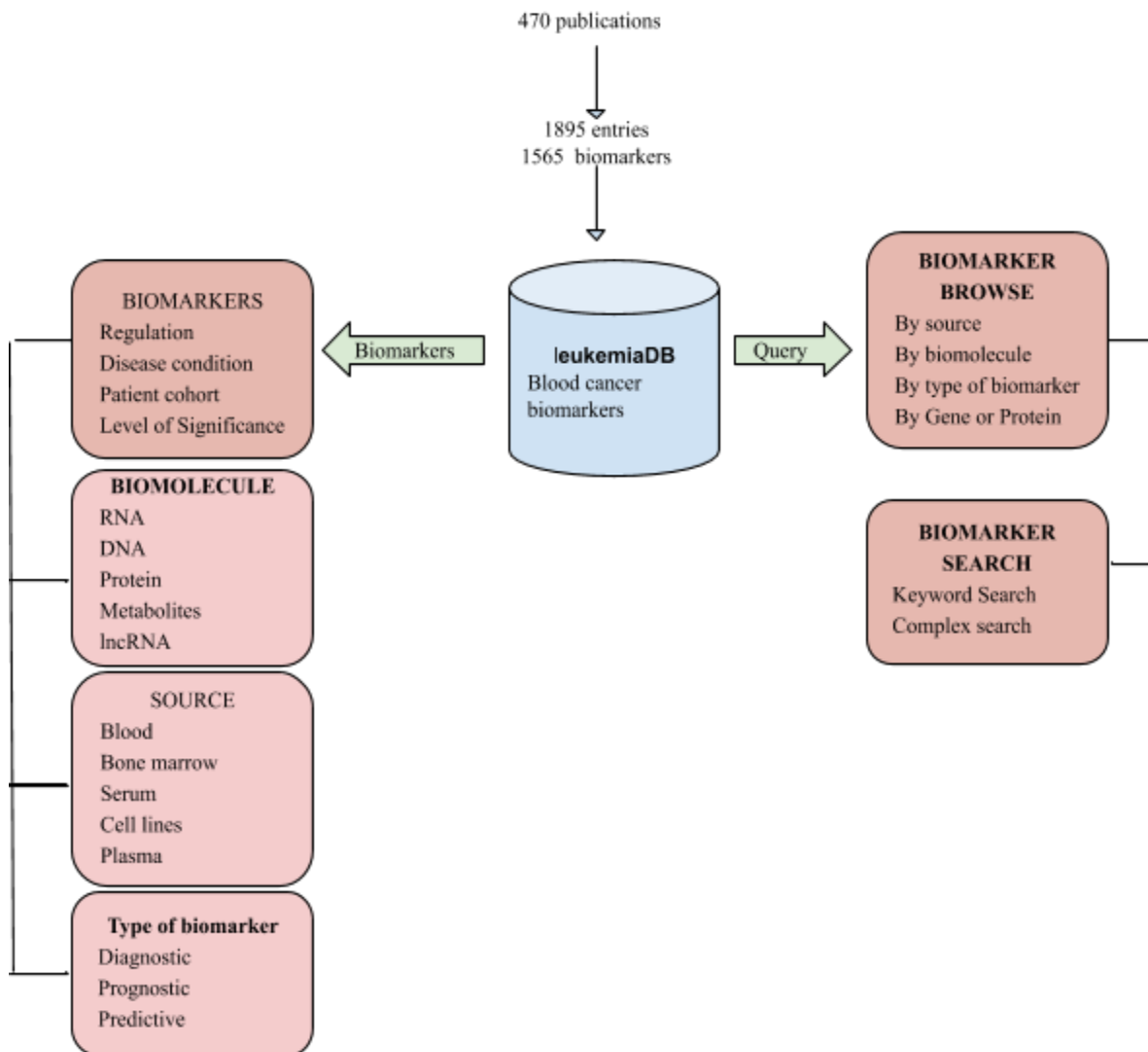


Figure 2: Architecture of LeukemiaDB

Database Organization

In leukemiaBD, the data is organized into four broad categories. The categorization of data into four browse domains which further categorized into different subtypes gene type, source, biomarker type and biomolecule type. Gene name has all the names of the genes that are regulated in leukemia are under this category. Leukemia biomarkers as a whole can be viewed by a particular user on the interface based on the gene/protein name of the biomarker that is regulated in leukemia. The visualization, of biomarker data based on the sources from which the sample is taken from are categorised into 5 main categories "Blood", "Bone marrow", "Cell lines", "Plasma" and "Serum". The visualization of biomarker data based on their types are categorised into 3 main categories "Diagnostic", "Prognostic", and "Predictive". Also we can visualise the biomarker data based on the type of biomolecule. It has five different categories "RNA", "Protein", "miRNA", "Metabolites" and others like "Long non-coding RNA".

The research articles from which data was extracted to incorporate in the database which has a unique PUBMED ID is also returned along with the results. Also each PUBMED ID has a hyperlink which redirects it to their relevant site domains. Each biomarker is individually linked to their PUBMED IDs wherever available in order to provide users an efficient, hassle free and oriented search via the platform.

Results

Web tools Implementation

For providing flexible and hassle-free data searching experience to the users, various tools such as retrieval, browsing, and analysis are integrated with LeukemiaBD.

A. Search tools

A.1 Keyword search

This tool represents keywords for data retrieval modules from LeukemiaBD. This can be executed by search queries such as Cancer Type, Biomarker, Biomolecule, Type of Biomarker, PMID, Regulation and condition and etc. Moreover, this module also allows the users to select various fields such as Pubmed ID, Biomarker, Biomolecule, Subjects, Cancer Type, Pathway, Level of significance, Patients Cohort, Sensitivity, Specificity, Accuracy, AUC, Degree of Validity and etc., that they want to get displayed for the result.

A.2 Complex search

For performing complex searchers using two or more query keywords related to the biomarker, the user can perform multiple structured query systems using the complex search option of

leukemiaBD. It performs four query searches simultaneously, by default, but it also allows the user to select a desired field/keyword against which search can be established. Apart from this facility in the complex search module, it allows the user to apply basic standard logical operators (e.g. LIKE, >, <, and =). Also, the queries to be implemented can also be added or removed to perform precise searches.

B. Browse tools

In leukemiaBD, we have implemented a browsing facility, which helps the user for convenient and flexible data navigation within the database in an orderly manner. In this module, a user can retrieve information on leukemia biomarkers by browsing following 5 main categories:

B.1 Browse by GENE NAME:

This page provides the option to browse the entries of LeukemiaBD based on the gene/miRNA and protein entity.

B.2 Browse by SOURCE:

This page provides the option to browse the entries of LeukemiaBD based on the Source of Samples from which Biomarkers were extracted. Further after clicking on the number of entities, the page shows the result of browsing biomarkers from different fields. It will display different fields such as Biomarker ID, Biomarker, Biomolecule, Subject, Regulation, Biomarker type, Experiment, Level of significance, Source, Year and PMID. On clicking PMID, it will redirect the user to the respective article.

B.3 Browse by BIOMARKER TYPE:

It provides the option to browse the entries of LeukemiaBD based on the Types of Biomarkers

(Diagnostic/Prognostic/Predictive). On further clicking on the level of entities, the page will display fields like Biomarker ID, Biomarker, Biomolecule, Subject, Regulation, Biomarker type, Experiment, Level of significance, Source, Year, PMID and etc.

B.4 Browse by BIOMOLECULE TYPE:

It provides the option to browse the biomarkers of LeukemiaBD based on the type of biomolecule (Protein/RNA/miRNA,DNA). Here also, after going further the page will display fields like Biomarker ID, Biomarker, Biomolecule, Subject, Regulation, Biomarker type, Experiment, Level of significance, Source and PMID.

There are respective hyperlinks to each of their PubMed ID's which redirects the research articles to their relevant site domains. Also, the biomarker and the associated disease type are individually linked to their PubMed IDs wherever they are available to enable a flexible and oriented search via the interface.

LeukemiaBD statistics

LeukemiaBD comprises a total of 1895 entries of the potential biomarker candidates of leukemia. During the initial screening, review articles and the articles lacking relevant information with respect to leukemia were excluded. Based on the screening that we did we narrowed down to 470 relevant papers out of the total 1860 research papers. Finally the data was curated from 470 relevant research papers. This resource has approximately 1568 unique biomarkers or biomarker candidates.

There are many studies available in literature that further elucidated the potential of a specific biomarker under different experimental conditions and cohorts. This resource includes information from all such studies. For example, there total 18 different entries of RUNX1 gene from 9 different studies in our database. The genes that were more prevalent in leukemia are FLT3, RUNX1, TP53, ABL1, NOTCH1, WT1, CEBPA, JAK2, ETV6, ZAP70, IDH1, STAT3, DNMT3A, IDH2, MCL1, CD38 etc. Some of the genes regulated in Acute Lymphocytic leukemia are CD24, BCL2L11, CDK6, IKZF1, CEBPA etc. Some of the genes regulated in acute myeloid leukemia are DNMT3, PARN, CNOT7, CNOT6, CNOT6L, Survivin etc. Some of the genes regulated in Chronic Lymphocytic Leukemia are miR-103, miR-181a, miR-181b, SFTPB etc and in Chronic Myeloid Leukemia are CECR1, FCN1, ADM, ANGPT1, S100A10, VWF, CLEC7A, MUC4, EGFL6 etc.

Another important aspect of these biomarkers is their differential regulation in various conditions like cancer versus normal. Wherever available, we have collected the information whether the biomarkers that are involved in leukemia were upregulated or downregulated under different experimental conditions. Also along with this we have specified the sensitivity, accuracy, specificity of the biomarkers as mentioned in the respective research paper wherever possible to give an idea of the accuracy of the respective biomarker. The distribution of biomarkers can be done based on the gene/protein name, type of biomarker, type of biomolecule, and based on the source from which it is extracted. Biomarkers are of three types namely prognostic, diagnostic, predictive. In leukemiaBD database there are 855 biomarkers which are of type prognostic, 1519 of type diagnostic and 115 are of type predictive. The different types of Biomolecule are included in leukemiaBD, such as DNA, RNA, Proteins, miRNAs, Metabolites,

lncRNA. Total 1027 RNA, 439 protein, 369 miRNA, 50 DNA, 12 lncRNA and 3 metabolites. Biomarkers were extracted from various sources such as blood, bone marrow, cell lines, etc. 904 biomarkers were extracted from blood, 76 from Serum, 777 from bone marrow, and 234 from Cell Lines and 15 from plasma.

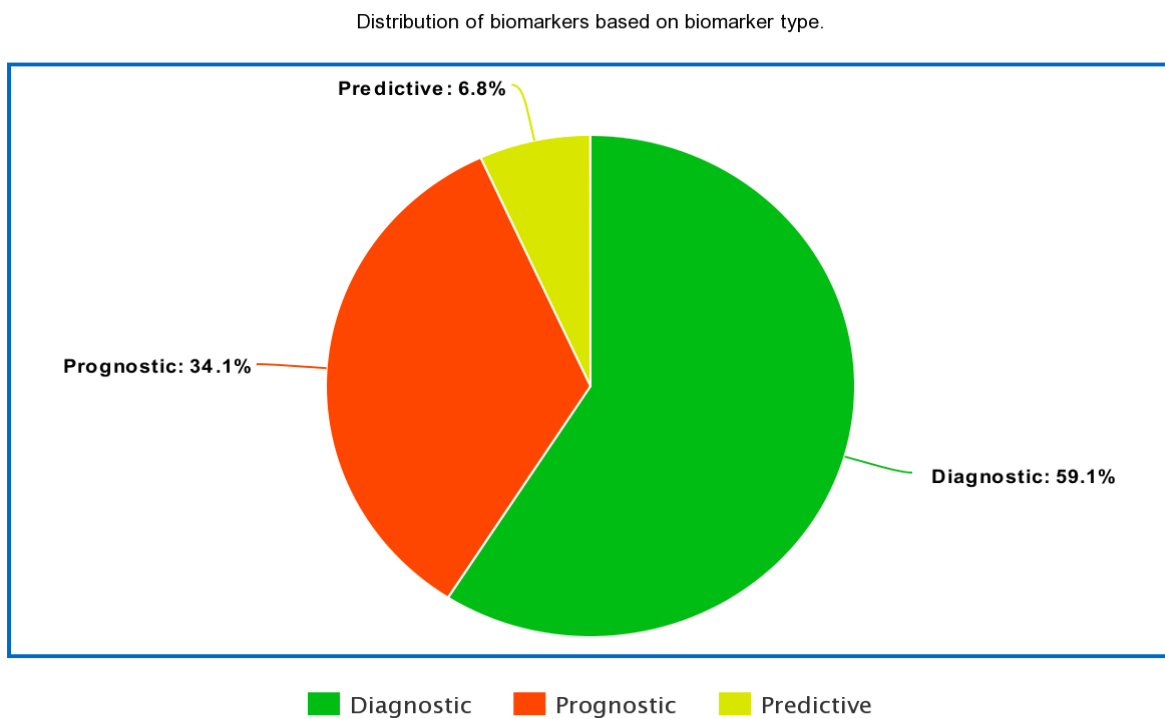


Figure 3: Distribution of biomarkers based on biomarker type.

Distribution of biomarkers based on biomolecule.

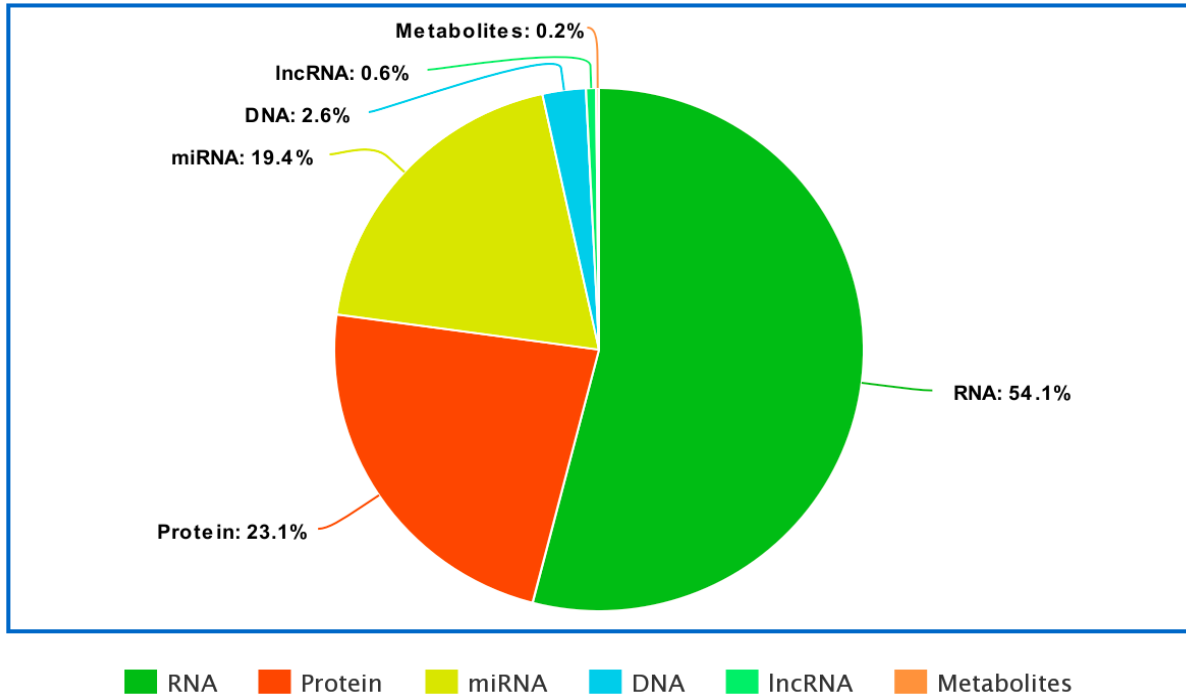


Figure 4: Distribution of biomarkers based on biomolecule.

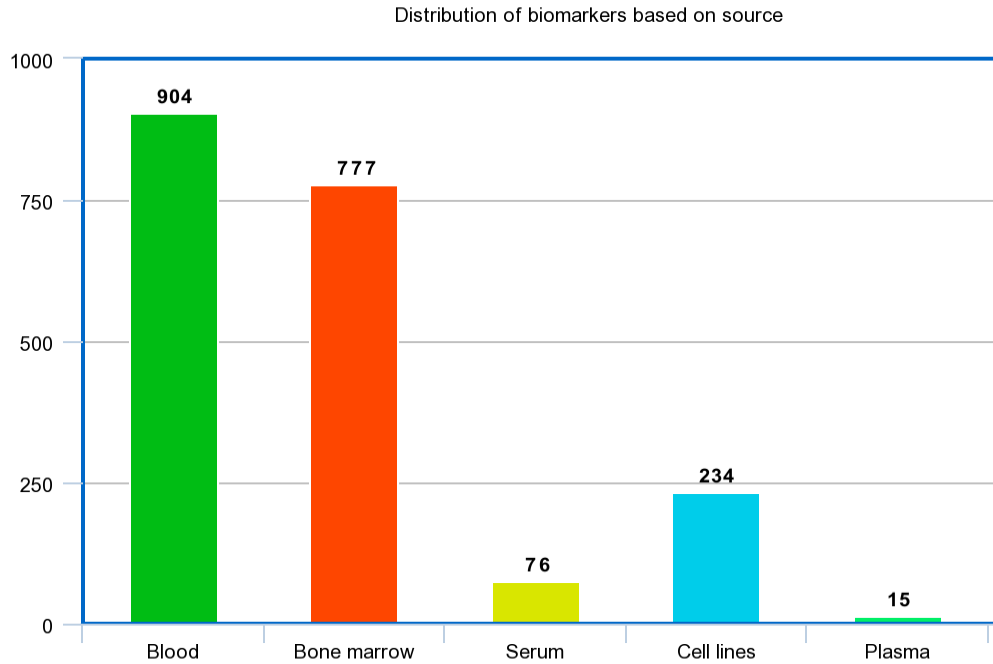


Figure 5: Distribution of biomarkers based on source

Availability

LeukemiaBD is compatible and flexible with all the latest gadgets and appliances. All prospective users such as students, researchers, clinicians, counsellors etc. can access the data provided in leukemiaBD and interface without any prior login credentials required.

LeukemiaBD can thus, be accessed freely at:

<https://webs.iiitd.edu.in/raghava/leukemiaBD/>

Working of the database

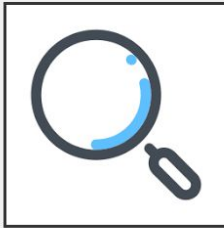
The database was prepared in such a way that all the information related to the leukemia biomarkers could be available at a single web location. The search and browse option displays all information about the queried biomarker in a tabular representation. Thus it saves time and makes information readily available to the users in a hassle-free way.

The screenshot displays the leukemiaBD website interface. At the top, there is a navigation bar with the logo 'leukemiaBD' and links for 'Home', 'Search', 'Browse', 'General', and 'Contact Us'. Below the navigation bar, the main heading reads 'leukemiaBD: A Database of Biomarker data of blood Cancer'. A descriptive paragraph states: 'leukemiaBD is a database of blood cancer that maintains gene expression datasets and biomarkers curated from public repositories and literature respectively. It contains following two modules for extracting data.' Two main modules are presented side-by-side. The left module, 'Browse biomarker', features a magnifying glass icon and a green button labeled 'Browse biomarker'. Below it, a text box explains: 'This module allows the users to browse for biomarker based on gene/protein name, source, biomolecule and type of biomarker.' The right module, 'Biomarker search', features a blue box with a diagram showing 'Diagnostic', 'Prognostic', and 'Predictive' biomarkers, and a green button labeled 'Search biomarker'. Below it, a text box explains: 'This module allows the users to retrieve biomarkers in leukemiaBD. It has both searching facilities for users.' At the bottom of the page, a footer contains a list of links: 'Home', 'About Us', 'Contact Us', 'Privacy Policy', 'Terms of Use', 'Disclaimer', 'Sitemap', 'Feedback', 'Help', 'FAQ', 'Glossary', 'Index', 'Search', 'Browse', 'General', 'Contact Us'.

Figure 6: LeukemiaBD

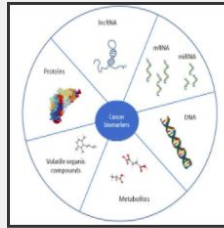
leukemiaBD: A Database of Gene expression and Biomarker data of blood Cancer

leukemiaBD is a database of leukemia that maintains gene expression datasets and biomarkers curated from public repositories and literature respectively. It contains following three modules for extracting and analyzing data.



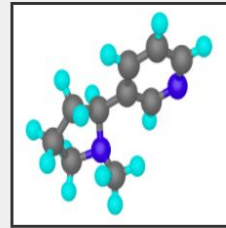
KeywordSearch

This tool represents keywords for data retrieval modules from LeukemiaBD. This can be executed by search queries such as Cancer Type, Biomarker, Biomolecule, Type of Biomarker, PMID, Regulation and condition and etc. Moreover, this module also allows the users to select various fields such as Pubmed ID, Biomarker, Biomolecule, Subjects, Cancer Type, Pathway, Level of significance, Patients



Type of Biomarker

It provides the option to browse the entries of LeukemiaBD based on the Types of Biomarkers (Diagnostic/Prognostic/Predictive). On further clicking on the level of entities, the page will display fields like Biomarker ID, Biomarker, Biomolecule, Subject, Regulation, Biomarker's type, Experiment, Level of significance, Source, Year, PMID and etc.



Biomolecule

It provides the option to browse the biomarkers of LeukemiaBD based on the type of biomolecule (Protein/RNA/miRNA,DNA). Here also, after going further the page will display fields like Biomarker ID, Biomarker, Biomolecule, Subject, Regulation, Biomarker's type, Experiment, Level of significance, Source and PMID.

Figure 7: Browse biomarker based on different categories i.e type,source,biomolecule etc.

The browsing option has been developed keeping in mind the different needs of a clinician, researcher, counselor, or a patient, to view biomarkers and the associated leukemia subtypes. This allows the user to experience hassle free and complete view about these biomarkers from a single database (Figure 7).

Browse Datasets in CancerLeukemia based on types of Biomarker

This page provides the option to browse the entries of CancerLeukemia based on the Types of Biomarkers (Diagnostic/Prognostic/Predictive).
For more information see HELP page.

Type of Biomarker	No of Entries
Diagonostic	1519
Prognostic	855
Predictive	115

Figure 8: Browse biomarker based on biomarker type.

Figure 9 shows a graphical representation, to exhibit how a basic search for the PMID 30840284 can be carried out. figure 10 gives query results i.e all the information about the PMID.

leukemiaBD Home Search Browse General Contact Us

Keyword search against Biomarkers

This page is designed for performing keyword search against Biomarker in leukemia. It allows the user to perform search against biomarkers on any field or against multiple fields. It also allows the user to select fields for DISPLAY in biomarkers. If you need any help, please visit HELP page.

Query Submission Form

Paste/insert/type your query to be searched:

Select Fields to Search

Disease type[AML,ALL,CLL,CML]
 PMID

Biomarker
 Regulation in Cancerous Condition[Upregulation/Downregulation]

Biomolecule[DNA/RNA/Protein/Metabolite/lncRNA]
 Experiment

Type of Biomarker[Prognostic/Diagnostic/Predictive]

Select Fields to Display

PMID
 Disease Condition
 Cohort
 Accuracy

Biomarker
 Regulation in Cancerous Condition
 Clinical
 Degree of Validity

Biomolecule
 Effect on Pathway
 Sensitivity
 P-Value

Subjects
 Type of Biomarker
 Specificity

Figure 9: Example of a keyword Search in Leukemia BD

leukemiaBD Home Search Browse General Contact Us

Welcome!

This page shows the result of browsing biomarker from different fields.

Please click on PMID to view detailed information.

The total number of entries retrieved from this search are: 1

Row	PMID	Biomarker	biomolecule	Subjects	Disease condition	Regulation in cancerous condition	Type of biomarker	Cohort
36	30840284	MiR-592	miRNA	humans	acute myeloid leukemia	downregulation	Diagnostic	94 patients,healthy control

Raghava's group CancerCSP CancerTSP CancerPDF CancerLSP CancerSPP

Figure 10: Keyword search Query results

Figure 11 shows a graphical representation, to exhibit how a complex search for the biomarkers that are diagnostic and is a miRNA can be carried out. figure 12 gives query results i.e all the information about the PMID.

The screenshot shows the LeukemiaBD website's search interface. At the top, there is a navigation bar with links for Home, Search, Browse, General, and Contact Us. The main heading is "Complex Search Module of Biomarker". Below this, a paragraph explains that the page is designed for advanced searching of biomarkers, allowing users to perform complex queries with multiple fields and conditions. The core of the interface is a table titled "Complex Search Module of Biomarker" with the following structure:

No.	Field	Condition	Query	Operator	Delete Row
1	Biomolecule	LIKE	miRNA	NO OPERATOR	Remove
2	Type of Biomarker	LIKE	Diagnostic	AND	Remove

Below the table are two buttons: "Search" and "Add new Row". At the bottom of the page, there is a footer with links for Raghava's group, CancerCSP, CancerTSP, CancerPDF, CancerSP, and CancerSPP, along with a small logo.

Figure 11: Example of a complex Search in Leukemia BD

Welcome!

This page shows the result of browsing biomarker from different fields.

Please click on PMID to view detailed information.

RN	Biomarker	Biomolecule	Subjects	Regulation	Type of Biomarker	Experiment	Level of Significance	Disease Condition	Source	PMID
10	miR-34a, miR-155, and miR-342-3p	miRNA	humans	upregulation	Diagnostic	CLL vs normal	NA	chronic Lymphocytic Leukemia	blood	21408091
11	miR-105, miR-181a and miR-181b	miRNA	humans	downregulation	Diagnostic	CLL vs normal	NA	chronic Lymphocytic Leukemia	blood	21408091
36	MiR-592	miRNA	humans	downregulation	Diagnostic	AML vs normal	(p < 0.01)	acute myeloid leukemia	bone marrow,serum	30840284
38	miR-335-3p	miRNA	humans	downregulation	Diagnostic	c-ALL vs control	P=0.018	Acute lymphoblastic leukemia	bone marrow	30639603

Figure 12: Complex search Query results

Discussion

LeukemiaBD is a comprehensive resource of biomarkers that are regulated in leukemia. The identification of new leukemia biomarkers contributes to a better understanding of the molecular basis of the disease. Also, it is significantly useful in screening, diagnosis, prognosis and monitoring of leukemia, as well as the possibility of predicting each individual's response to the treatment. LeukemiaBD consists and maintains potential biomarkers for leukemia that includes diagnostic, prognostic and predictive biomarkers.[14] These biomarkers are based on the different type of samples (like blood, bone marrow, cell lines, serum and plasma) and different kinds of biomolecules (like RNA, Protein, miRNA, metabolites, lncRNA).[21]

When we went through the literature we found an existing database of leukemia biomarkers. One such database is LGL database.[22] In this study they have a database of the leukemia genes and have provided information about the different biomarkers that are regulated in different subtypes. But it nowhere provides information about whether the biomarker is upregulated or down regulated, the different experimental condition, the different cohorts, sensitivity, specificity and accuracy of the particular biomarker. On the other hand our database of Leukemia biomarkers (leukemiaBD) enables the users to access all the information about the biomarkers that are involved in Leukemia. The information about all these biomarkers are stored in the database and can be accessed freely through our web interface. All the data in the leukemiaBD is extracted from the published biomedical research literature.[23] All the data extracted from these publications are categorically organised and stored in our database. All the extracted data is

categorised based on the type of biomarker, type of biomolecule ,the gene/protein name the sources from which the samples of biomarker are extracted from and various other important biological and clinical information.Each biomarker entry is linked to the original PubMed citation. The leukemiaDB platform provides a collection of relevant information that spans all possible aspects of the biomarkers of leukemia. It stores information ranging from the subtype of leukemia, type of biomarker to the experimental pathway which shall enable, hunt for searching biomarkers matches, which could further helpful in the development of drugs for the treatment of leukemia So, this database is developed to help the biological and medical sciences community and other users to easily access all information about biomarkers that are involved in Leukemia.In the absence of leukemiaBD,in order to make inferences about the biomarkers and their regulation in leukemia a researcher has to go through a huge amount of literature. Also since all the information is scattered on different possible sites it would make the task even more tough. The information about the leukemia biomarkers in leukemiaBD is displayed in a tabular format which makes it more feasible for any researcher to glance at all the relevant information in a more convenient and systematic manner.

Applications of database

Applications of leukemiaBD are the following :

- ❑ LeukemiaBD is a data portal with a collection of information about all the biomarkers of Leukemia at a single website, would make it easier and save a lot of time to extract information related to them and draw further inferences.

- ❑ The tabular visualization of all biomarkers of leukemia makes it easier for the researcher or clinician to access all significant aspects like the type of biomarker, its regulation in leukemia, the experimental pathway, the subtypes of leukemia with which it is associated.
- ❑ To the best of my knowledge, currently, no such broad database exists, which would solely cater to the needs of niche research groups working actively in the domain of leukemia biomarkers.
- ❑ This portal is a small contribution which would divert the attention of clinical practitioners and existing genetic experts to dwell more in-depth inferences in this domain using our platform.
- ❑ LeukemiaBD provides an opportunity to help the sciences community and other users to look for involving biomarkers across all the main subtypes of Leukemia

References

1. Davis, Amanda S., Anthony J. Viera, and Monica D. Mead. "Leukemia: An overview for primary care." *American family physician* 89, no. 9 (2014): 731-738.
2. *Types Of Leukemia: Common, Rare And More Varieties*. [online] Cancer Treatment Centers of America. Available at: <<https://www.cancercenter.com/cancer-types/leukemia/types>> [Accessed 3 June 2019].
3. Zupanec, Sue, and Deborah Tomlinson. "Leukemia." In *Pediatric Oncology Nursing*, pp. 1-32. Springer, Berlin, Heidelberg, 2010.

4. Leung, K. N., N. K. Mak, and M. C. Fung. "Cytokines in the differentiation therapy of leukemia: from laboratory investigations to clinical applications." *Critical reviews in clinical laboratory sciences* 42, no. 5-6 (2005): 473-514.
5. Enyakoit, George Ojula. *Cytogenetic analysis of DNA copy number aberrations in high malignancy grade astrocytomas*. University of London, University College London (United Kingdom), 2008.
6. Tom, A. M. L. "Acute myeloid leukemia."
7. Deshantri, Anil Kumar. "Liposomal Drug Delivery in Cancer: Hematological Malignancies and Beyond." PhD diss., Utrecht University, 2019.
8. Mayoclinic.org. 2020. *Leukemia - Diagnosis And Treatment - Mayo Clinic*. [online] Available at: <<https://www.mayoclinic.org/diseases-conditions/leukemia/diagnosis-treatment/drc-20374378>> [Accessed 15 May 2020].
9. Devine, Steven M., and Richard A. Larson. "Acute leukemia in adults: recent developments in diagnosis and treatment." *CA: a cancer journal for clinicians* 44, no. 6 (1994): 326-352.
10. Strimbu, Kyle, and Jorge A. Tavel. "What are biomarkers?." *Current Opinion in HIV and AIDS* 5, no. 6 (2010): 463.
11. Winter, Jordan M., Charles J. Yeo, and Jonathan R. Brody. "Diagnostic, prognostic, and predictive biomarkers in pancreatic cancer." *Journal of surgical oncology* 107, no. 1 (2013): 15-22.
12. Weigel, Marion T., and Mitch Dowsett. "Current and emerging biomarkers in breast cancer: prognosis and prediction." *Endocrine-related cancer* 17, no. 4 (2010): R245-R262.
13. Buyse, Marc, Daniel J. Sargent, Axel Grothey, Alastair Matheson, and Aimery De Gramont. "Biomarkers and surrogate end points—the challenge of statistical validation." *Nature reviews Clinical oncology* 7, no. 6 (2010): 309.
14. Aronson, Jeffrey K., and Robin E. Ferner. "Biomarkers—a general review." *Current protocols in pharmacology* 76, no. 1 (2017): 9-23.

15. En.wikipedia.org. 2020. *Regulation Of Gene Expression*. [online] Available at: <https://en.wikipedia.org/wiki/Regulation_of_gene_expression> [Accessed 3 June 2020].
16. Spruance, Spotswood L., Julia E. Reid, Michael Grace, and Matthew Samore. "Hazard ratio in clinical trials." *Antimicrobial agents and chemotherapy* 48, no. 8 (2004): 2787-2792.
17. Anjum, A., Jaggi, S., Varghese, E., Lall, S., Bhowmik, A. and Rai, A., 2016. Identification of Differentially Expressed Genes in RNA-seq Data of Arabidopsis thaliana: A Compound Distribution Approach. *Journal of Computational Biology*, 23(4), pp.239-247.
18. HealthNewsReview.org. 2020. *Understanding Medical Tests: Sensitivity, Specificity, And Positive Predictive Value*. [online] Available at: <<https://www.healthnewsreview.org/toolkit/tips-for-understanding-studies/understanding-medical-tests-sensitivity-specificity-and-positive-predictive-value/>> [Accessed 3 May 2020].
19. Liu, Danping, and Xiao-Hua Zhou. "ROC analysis in biomarker combination with covariate adjustment." *Academic radiology* 20, no. 7 (2013): 874-882.
20. Goodman, Steven. "A dirty dozen: twelve p-value misconceptions." In *Seminars in hematology*, vol. 45, no. 3, pp. 135-140. WB Saunders, 2008.
21. Tan, Siun Chee, and Beow Chin Yiap. "DNA, RNA, and protein extraction: the past and the present." *BioMed Research International* 2009 (2009).
22. Liu, Yining, Mingyu Luo, Zhaochen Jin, Min Zhao, and Hong Qu. "dbLGL: an online leukemia gene and literature database for the retrospective comparison of adult and childhood leukemia genetics with literature evidence." *Database* 2018 (2018).
23. Falagas, Matthew E., Eleni I. Pitsouni, George A. Malietzis, and Georgios Pappas. "Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses." *The FASEB journal* 22, no. 2 (2008): 338-342.