Improving Management of Individual Patients with Leukemia.

Comprehensive Review Paper

Vincent N. Carrasco, MD MSIS

Carolina Health Informatics Program

&

School of Information and Library Science

University of North Carolina

Abstract:

Patients with leukemia have their best chance for a good outcome with a precise diagnosis and an individualized treatment beginning at the point-of-care. Evidence-based medicine (EBM) and genomic medicine (GenM) have done much to advance towards this goal. EBM, promised high-quality cost-effective treatment using biostatistical methods and in the 1990s replaced the mechanistic paradigm of disease (MPD). This was justified by the assumption that EBM would enable precise treatment of individual patients. However, efficacy, for a statistically derived average patient has proven insufficient for individualizing treatment for the real-world patient population. Currently, physicians have no means of making predictive point-of-care decisions for individual patients. Health analytics has the capacity to individualize patient management. Warehoused electronic health information (EHI) represents accumulated clinical experience from real-world patients. Multivariate clinical data analysis (MCA) and cohort multidimensional analysis CoMA (CoMA) has the potential to build individualized clinical patient models from EHI to support decision making and prognostication.

***Keywords*:** evidence-based medicine; leukemia; genomics; multivariate analysis; co-variate analysis; principal component analysis

CONTENTS:

I. Introduction 4
A. Evidence-based medicine 5
1. Background 5
2. Value of EBM 6
B. Genomics 7
1. Background 7
2. Value of Genomics for the treatment of Leukemia 8
C. Multivariate clinical data analysis (MCA-L) 10
II. Problem Statement, Scope and Research Questions 12
A. Problem Statement 12
B. Scope 12
1. The scope of the Research Questions 13
C. Expected Findings 13
III. METHODS. 14
1. Evidence-based Medicine for Leukemia (EBM-L) Search 14
1. Multivariate Clinical Data Analysis of Leukemia (MCA-L). 15
2. Genomics of Leukemia (GEN-L) 16
B. Data Management and Processing 17
C. Results and Analysis 20
D. Discussion 24
1. Has evidence-based medicine (EBM -L) improved a physician's ability to diagnose and treat an individual patient's leukemia? 24
2. Does the human genomics of leukemia (GEN-L) study the diagnosis and treatment of individual leukemia patients? 25
3. What are common methods of complex data analysis in the leukemia literature and how they were employed? 25
IV. Conclusion 26
V. References. 29
VI. Article list from search by category. 33
A. EBM-L 33
B. MCA-L 36
C. Gen-L 39

I. INTRODUCTION

Historically, diseases were classified by anatomical location, microscopic appearance, symptoms and progression. Medical research focused on normal anatomy, histology, biochemistry, physiology and a basic understanding of health emerged for the human body [1]. Disease was defined as an impairment of bodily function accompanied by discernable signs and symptoms. Researchers established theoretical mechanisms for disease by studying variations from normal. Basically, pathology appearance and pathophysiology referred to those things that varied in chemical processes. This was called the *Mechanistic Framework* for disease etiology. These variations were linked to objective findings (signs), subjective complaints (symptoms) and bodily changes over time (progression). This framework served as the theoretical basis for treatment.

During the 20th century the basic medical sciences[1] underpinned the "mechanistic" curriculum from which medical students were taught. Supervised apprenticeships[2] provided clinical practice experience. Unfortunately, the quality of medical education and the delivery of health care varied by institution and region.

In 1950 scientific knowledge increased at an exponential rate, Figure 2 [2]. By 1982 human

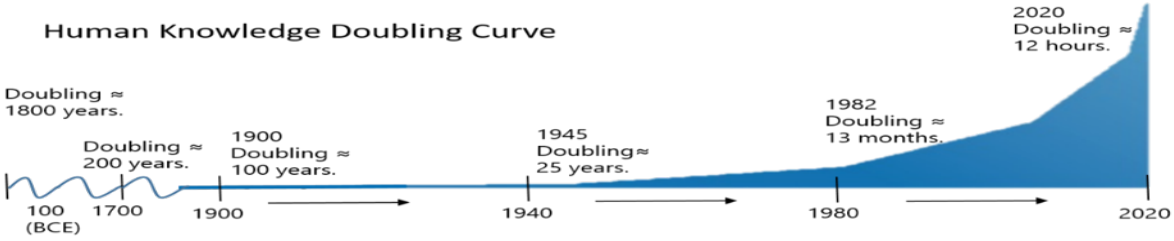


Figure 2. Knowledge went from doubling every 100 years in 1900 to doubling every 12 hours in 2020. [Fuller; IBM, VNC]

knowledge doubled every 13 months. The translation of new medical knowledge into patient care began to fall behind. A gap between research and clinical practice developed.

During the 1990's, the cost of health care skyrocketed, the translational gap became significant and the quality of care[3] declined. Confidence in the mechanistic principals of practice began to wane. As a result, a new paradigm, evidence-based medicine (EBM), dominated clinical research. EBM depended upon clinical evidence derived from biostatistical and epidemiological methodologies. It promised improved individualized high-quality care and cost reduction [3].

The 21st century brought genomics[4] and health analytics. Genomics promises to further individualize care with more accurate diagnoses, focused therapy and better outcomes. Health

[1] Gross anatomy, histology, biochemistry, physiology, etc.
[2] "see one, do one, teach one"
[3] "outcomes"
[4] The translation of genetic testing

analytics[5] also evolved during this time period. Health analytics linked EBM and genomics creating new medical knowledge. Health analytics, applied to clinical practice data has the potential to augment current clinical decision making and meet the standards of precision medicine.

Evidence-based medicine of leukemia (EBM-L), genomics of leukemia (GEN-L) and multivariate clinical data analysis of leukemia (MCA-L) are the core topics of this review.

A. Evidence-based medicine

1. Background

To Err is Human: Building a Safer Health System, in 1999 publicly revealed that the American health care system health was fragmented, unaccountable and lacked uniform quality. 98,000 people per year were dying from medical errors and the cost of care was unsustainably accelerating [4]. Americans spent twice as much per capita on health care than the European, Asian and Oceanic nations but had a shorter life span (figure 3). It was during this time that evidence-based medicine gained momentum [5].

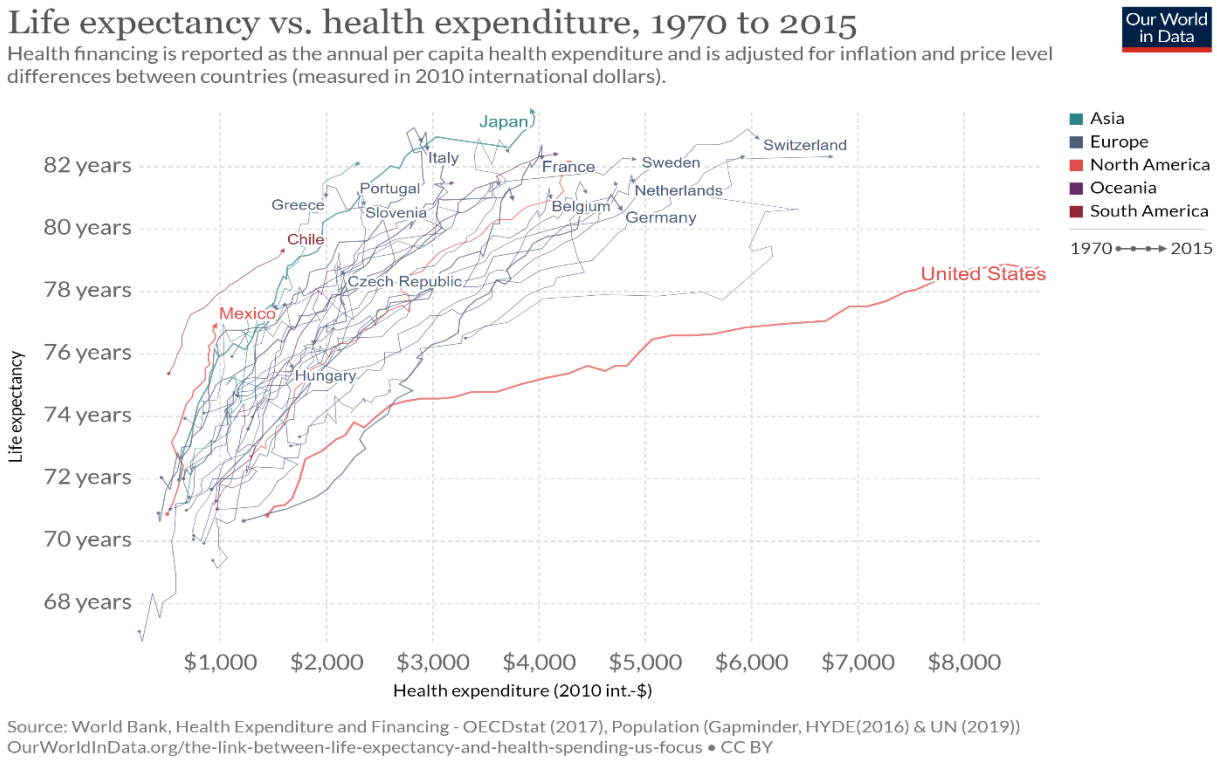


Figure 3. Health care per capita spending: US vs World [6]

[5] Mathematical and biostatistical methods for the analysis and translation of medical data into actionable knowledge

Epidemiological medical researchers decided to change physicians' thinking about the diagnosis and treatment of disease. Their decision was based upon 4 assumptions. First, embracing EBM promoted "superior patient care" (3: p.2421) by making them better informed and eliminating bias. Second, physicians were biased because of their practice experience and their mechanistic thinking [5]. Physicians were responsible for: "*expensive, ineffective or harmful decision making*" and "*relying upon pathophysiological principles may lead to adverse events or inaccurate estimates about the efficacy of interventions*" [7: p. 1122, par. 1)]. Clinical intuition and experience were biased and represented an uncritical or haphazard use of research results [5]. Third, physicians do not understand biostatistical "rules of evidence, as the basis for determining treatment efficacy. Certain types of evidence deserved privileged consideration compared to others based on their higher and the low risk of bias (Fig.4) [5]. Fourth, decisions based on mechanistic thinking were unreliable and the translational gap was a result of the slow uptake of new interventions and explained variations in practice [3].

Figure 4. Pyramid of Hierarchical Evidence [VNC]

2. Value of EBM

Essentially, academic clinical and research educators believed, mechanistic reasoning, was flawed and should be rejected [8]. As a result, the emphasis on the basic medical sciences was reduced and problem-oriented learning, based on evidence from RCTs and systematic reviews was substituted (Figure 4) [3]. In fact, RCTs showed statistically significant evidence of efficacy even in the absence of identifying or understanding causal mechanisms [9]. This became the new paradigm for medical education and clinical practice [10; 11].

EBM helped establish a strong framework for determining the safety and efficacy of treatments by defining a criterion for "quality medical evidence" and a methodology for analysis by systematic review. Relevant publications were identified, aggregated, and analyzed based on the rules of systematic review. The strength of support for treatments depended upon hierarchical grading systems that assigned levels of quality to research reports during systematic review [12]. The basic principal being that certain types of evidence were better than others and deserved privileged consideration. They were a higher quality of research that had a lower risk of bias compared to others (Fig.4) [5]. This became the gold standard for the medical treatment of people. Subsequently, a number of journals and medical organizations adopted variations of increasingly complex classification systems of the quality of medical research [13; 4].

B. Genomics

1. Background

Genomics is defined as "the branch of biotechnology concerned with the application of genetics and molecular biology to the sequencing of gene sets, specifically the genetic mapping of DNA and complete genomes" [15]. Its roots reach to the classical era of Hippocrates, Aristotle and Gregor Mendel whose experiments advanced genetics from a descriptive science to one with substantiated scientific conclusions [16].

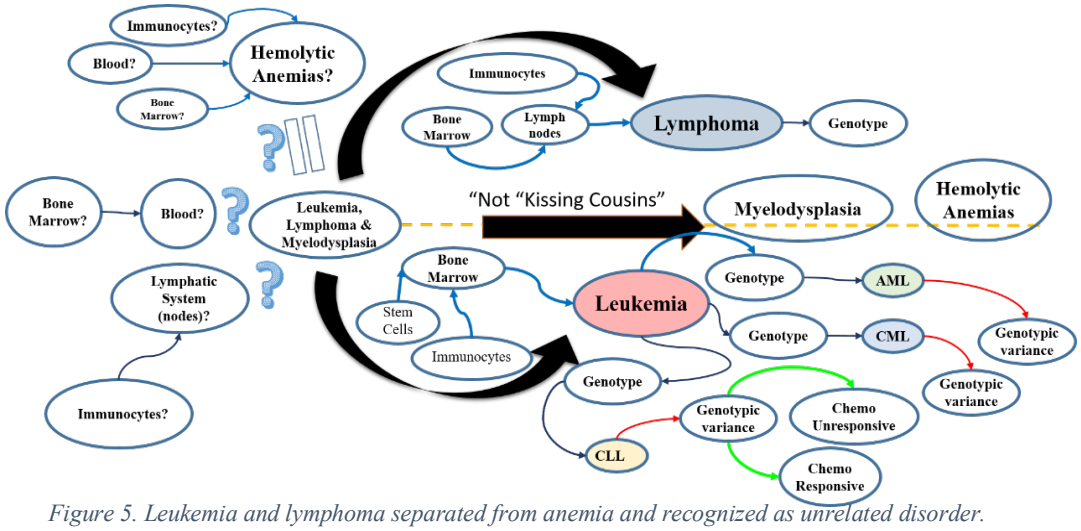


Figure 5. Leukemia and lymphoma separated from anemia and recognized as unrelated disorder.

Chromosomal abnormalities as a cause for cancer were hypothesized as early as 1914 [17]. Tissue staining and histochemical methods identified genetic markers that enabled some early differentiation of hematological malignancies. However, it wasn't until the 1960's that diagnostic testing became sophisticated enough to genetically differentiate leukemia and lymphoma from "anemia", see Figure 5. By the1980's oncogenes were identified as responsible for malignant transformation and clearly separated leukemia from lymphoma [17;19; 20]. Anemia, was now recognized as an end-stage state for many blood disorders.

Genomics owes much of its success to the development of precise measurement and biostatistical tools. Analysis of human tissues and body fluids detected biomarkers while biostatistical methods provided the correlation with exposure to environmental factors and specific diseases. The detection of a biomarker reflected an individual's genome or changes in their genome consistent with the tendency towards or presence of pathology.

After the discovery of chromosomal abnormalities associated with cancer, early treatment efforts focused on the pathophysiology caused by these abnormalities. For example, cancer cells have vulnerabilities compared to most normal cell systems, such as rapid replication, high metabolic demands, inappropriate angiogenesis[6] and deranged cellular repair mechanisms. Normal cell systems had razor thin advantages over cancerous cells. RCTs exploited the vulnerabilities of

[6] Capacity to form nutrient blood vessels

cancer cells with highly toxic agents. The most efficacious treatments, many with thin statistical advantages, were consolidated by expert consensus committees into evidence-based guidelines and protocols [21]. Leukemia had some of the earliest confirmations that this terminal disease was treatable [22].

Most leukemic patients, were treated with conventional chemotherapeutic agents and experienced unsustained positive responses[7]. All of the patients experienced toxic injuries. Many normal cell systems responded well to systemically toxic chemotherapeutic drugs, some did not. Cell systems (tissues or organs) with normally rapid lifecycles[8], similar to cancer cells (gastrointestinal, blood, and immune), were severely damaged causing nausea, dehydration, malnutrition, infection and even death. The "razor thin" advantage human cell systems had was their ability to recover. Treatment, with these agents were a double-edged sword; giving hope at the expense of quality of life and cost of care.

2. Value of Genomics for the treatment of Leukemia

Leukemia may be suspected from a patient's clinical presentation, diagnosed by laboratory testing but is only effectively treated after gene mapping. This is the process that identifies gene locations and their distance from related genes, essentially creating a "map or fingerprint" of the disease. Once the genes' physiological role in the cell cycle is known a theory of the disease's etiology is proposed. Work can then begin looking for less toxic targeted therapies.

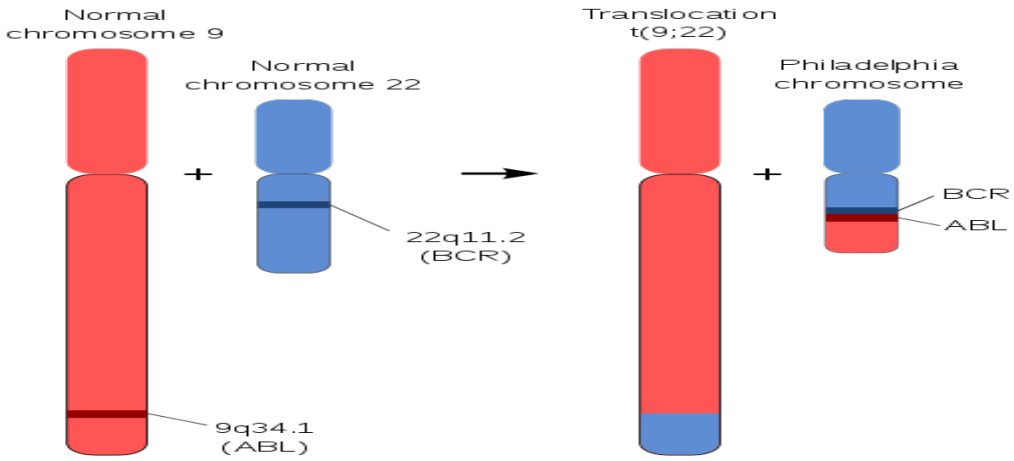


Figure 6. Reciprocal Philadelphia Translocation t(9;22) (q34;q11) (Wikimedia Commons)

Chronic myelogenous leukemia (CML) results from an asymmetrical reciprocal translocation from chromosome 9 (c9) to chromosome 22 (c22) [23]. This is called the Philadelphia translocation and is designated by the notation *t(9;22)(q34;q11)*[9] and yields a BCR-Abl fusion gene, see figure 2. The BCR-Abl fusion gene codes for the BCR-ABL protein; a constitutively

[7] Response in this sense means <u>any</u> discernible change compared to the untreated state
[8] Cell turnovers
[9] Translocation "t" from c9 to c22 with "region 3, band 4, sub-band 1" break of long arm "q" of c9 "region 1, band 1, sub-band 2" of the long arm q of c22. [26]

active tyrosine kinase. Abl (Abelson) protein, normally a self-inhibitory cell induction protein, becomes a fusion protein that de-regulates and accelerates cell division. Unchecked cell replication no longer undergoes apoptosis and rapidly supersedes Hayflick's limit for normal senescence: 50 ± 10 cell replications [24]. The bone marrow expands into all tissue spaces and is populated by immortal dysregulated leukemogenic clones. The bloodstream is flooded with poorly differentiated myeloblasts (blast cells)[10]. As normal clones senesce, and the leukemogenic expanded bone marrow exhausts its nutrient supply, bleeding, hypoxia, tissue necrosis (pain), sepsis and death ensues.

Leukemogenic oncogene translocation activation deranges intracellular signal transduction leading to leukemia and its phenotypic variants. Hematopoietic progenitor cell replication errors manifest by mutation, amplification or translocation. Some are immediately lethal; many are eliminated by senescence and others transform proto-oncogenes[11] to oncogenes. [25]. Conversely, there are *oncogenes*[12] shared by highly disparate cancers that respond to the same chemotherapeutic and targeted agents. For example, CML and gastrointestinal stromal tumor (GIST) are both linked to ABL1 and ABL2 (Abelson family of protein kinases) and respond to imatinb - the first clinically available PKI [26]. Therefore, current treatment of leukemia employs hematopoietic cell membrane tyrosine kinase inhibitors. In many cases these new targeted therapies replace systemic chemotherapeutic agents, radiation and bone marrow transplantation as first line therapy. This new class of drugs resulted from the analysis of human tissues and the identification of biomarkers; indicators that represent the presence of disease, its variants, their response to treatment and environmental factors (epigenetics). Evidence for their efficacy was determined by RCTs and oncologists moved more readily from non-specific systemically toxic treatments to more targeted therapies.

[10] Myeloblasts or blasts are primitive poorly differentiated blood cells that precede the final functioning blood cell.
[11] Proto-oncogenes are normal genes when altered by mutation become oncogenes – genes that cause cancer.
[12] I.e., oncogenes are the cancer-causing genes

A more recent example is chronic lymphocytic leukemia (CLL), a subtype of leukemia, it is typically a slow indolent disease that responds well to treatment [45]. However, it also has phenotypic variants that are extremely aggressive and resistant to chemotherapy. These are associated with genotypes possessing del(17p) and/or the TP53 gene mutations. Fortunately, these variants are sensitive to venetoclax, an orally bioavailable inhibitor of B-cell lymphoma/leukemia 2 protein (BCL-20) a kinase that blocks apoptosis [27]. Before these mutations were identified, all CLL patients underwent the conventional treatment protocols. A sub-category of patients experienced relentless disease progression and treatment failure. These patients have the mutations. Patients with CLL can be screened upfront for the mutations and provided with the additional targeted therapy as appropriate.

There were many areas impacted by genomics including bone marrow transplantation and the diagnosis of metabolic disorders. Genomics is really in its infancy and there are many exciting discoveries yet to be made that will improve the human condition.

C. Multivariate clinical data analysis (MCA-L)

Principal component analysis is a form of multivariate analysis. It is a matrix factorization method that extracts individual variables from sets of random or mixed variables. It affectively reduces the dimensions of interrelated variables while maintaining data set variation. The reduction in dimensionality occurs when the original datasets are *transformed* into new sets of ordered uncorrelated variables; *the principle components.* signals from is and is defined as an orthogonal linear transformation consisting of a data matrix "X" with *n* rows and *p* columns [28; 29; 30]. This is transformed to p – dimensional or" weighted" vectors representing the greatest variance. First greatest variance is the first component with the second greatest variance representing the second component etc. In this way the most significance features are identified and dataset dimensionality is reduced[13]. It began as an analogue to the principal axis theorem[14] [31] and later rediscovered, and named PCA, by Hotelling in 1933 [32; 33]. PCA was computationally intensive and impractical for anything but small datasets. Recent advances in computerization enable application to high-dimensional datasets. Currently, PCA is used under a variety of names; table 1 [28].

Designation		Domain
Karhunen–Loève transform		Signal processing
Eigenvalue decomposition (EVD)		Linear algebra
Empirical modal analysis		Structural engineering
Proper orthogonal decomposition (POD)		Mechanical engineering

Table 1. Variations of PCA used in other fields.

[13] Component vectors (sores) (t) are described as vectors projected onto the axes (x, y, z…n) & having a length or magnitude >0 & a direction which is an angle relative to an axis or "components" along an axis that is pos. or neg.
[14] Where a system of variables is described as a straight line – "orthogonal best fit". Pearson 1901

The medical record[15], until approximately 10 years ago, was a folder[16] organized using both structured and unstructured formats, by date and in sections[17]. Inpatient and outpatient files[18] were available in the same record. Health information, existed predominantly as unstructured physician notes, letters and reports. Health data was typically structured.

Structured health data, as part of its governance, are organized, and can be validated against expected values. This facilitates retrieval, analysis and exchange. Structured health data generally consists of lists of variables. Such as, quantitative variables; weight, blood pressure and cholesterol (continuous) or qualitative (categorical) variables; blood type, genotype (nominal), stages of a disease, social status and educational level (ordinal).

Unstructured health care data lacks the familiar formats characterized by structured data. Nonetheless, unstructured health care data exists in predictable formats determined by the nature of the data and domain. For example, diagnostic images, photographs, pathological specimens, histological slides, and scanned copies of structured data are formatted by standards of the media. For this data to become useful, currently, it requires human analysis and documentation.

Physician notes, reports and letters contain unstructured data within predictable formats ("templated" text), but most importantly there is *synthesis*. The data is transformed into a context in the form of a physician's narrative.

Hand written (dictated) unstructured physician narratives are ontological discourses that are temporally linked. Individual patients' medical problems are solved by structured and unstructured data gathering and analysis. Synthesis requires integration of internal (health record) and external resources (i.e.; EBM and genomics) resulting in a knowledgebase from which treatment decisions are made. Complex problems require input from multiple domains[19] not only for their analysis and synthesis but their experience. This knowledgebase and individual outcomes are chronicled in the EMR and are the foundation for clinical practice experience or experienced-based medicine (ExBM) [34]

Finally, there are no established methods to exploit unstructured data and structured data on leukemic patients for clinical decision making. These data exist within non-integrated databases[20]. Natural language processing (NLP) and machine learning (ML) can extract, process and aggregate this data into patient cohorts for medical modeling.

---

[15] The terms health record and medical record are used interchangeably
[16] Or a series of folders. An unhealthy patient could have a stack of folders a meter or greater in height
[17] Or folders commonly referred to as the "chart"
[18] vital signs, lab results, patient-generated lifestyle data, images and physician notes, letters and reports
[19] Peer and specialty consultation, imaging, special testing & multidisciplinary board presentation; i.e. Tumor Board.
[20] Genomic data is included as non-integrated structured data along with disease registry data, pharmacy data etc.

II. PROBLEM STATEMENT, SCOPE AND RESEARCH QUESTIONS

A. Problem Statement

Physicians cannot provide patients with the individualized diagnoses and personalized treatments that maximize their chances for the best possible outcome. Predictive point-of-care decisions for personalized care are not possible.

Genomics improved precision by enabling better disease characterization by identifying genetic abnormalities. EBM improved the overall uniformity of patient treatment and eliminated unacceptable practices. However, it also has significant weaknesses. EBM did not demonstrably reduce the cost of care or provide sufficient evidence to individualized care. The assumptions justifying the move from the mechanistic framework to EBM were flawed. [5; 35; 36].

Current dogma is based on evidence from RCTs. Efficacy for a drug is established when there is a statistically significant positive outcome in a theoretically homogenous population. In other words, RCTs measure the general affects within a cohort of average patients with a disease. Candidate selection for clinical trials are strict in order to measure the effect of a drug upon a disease.

People are evaluated and treated as individuals, not as groups. Patient populations are naturally heterogenous and therefore contain people with individual confounding factors and comorbidities. These characteristics significantly influence treatment outcomes (see table 2). Individual patients with multiple confounding variables represent skewing heterogeneous components of a statistically homogenous interpretation.

Patients with comorbidities and confounding factors populate RCT sub-cohorts. They represent small groups of phenotypic variants with effects that are statistically insignificant compared to the overall study population. The goal of a clinical trial is to determine the efficacy of a treatment upon a disease. Measuring the treatment effects on phenotypic variants is a different goal. An RCT with the capacity to yield statistically significant effects upon phenotypic variants would be orders of magnitude larger than typical studies and prohibitively expensive.

Confounding factors	Comorbidities
Race	diabetes
Age	hypertension
Gender	Heart disease
Socioeconomic Status (SES)	kidney disease
allostatic load [37]	

Table 2. List of Confounding factors and comorbidities found in phenotypic variants

B. Scope

Leukemia was selected as the center of this review because it has a rich coherent literature that has benefited from EBM, genomics and complex analytics. Many genotypic and phenotypic variants are available. In addition, the morbidity and mortality in the elderly is increasing and requires study.

Research questions centered on the management of leukemia patients tend to be broad based and with complex intertwined answers. To further narrow the scope of this review treatment of individual leukemic patient. Evidence-based medicine of leukemia (EBM-L) and the genomics of leukemia (GEN-L) are the current gold standards for decision making. Multivariate clinical data analysis (MCA-L) can bring the experience of managing patients with diverse and confounding characteristics into the clinical decision-making process. This review is restricted to a clinical perspective (i.e., PubMed) to further narrow the scope.

1. The scope of the Research Questions

This comprehensive review of the leukemia literature is divided into 3 areas of enquiry*;* *evidence-based medicine of leukemia* (EBM-L), *genomics of leukemia* (GEN-L) and *analytical techniques used in leukemia;* referred to simply as multivariate clinical data analysis (MCA-L). Two questions evaluate the EBM-L and GEN-L literature for study methodologies impacting individual patients and one question catalogues the analytical methodologies used to study leukemia. The research questions are as follows:

a) *Has* evidence-based medicine (EBM -L) improved a physician's ability to diagnose and treat an individual patient with leukemia*?*

b) Does the clinical human genomics of leukemia literature (GEN-L) study the diagnosis and treatment of individual patients with leukemia?

c) What are the common methods of multivariate clinical data analysis in the leukemia literature (MCA-L) and how were they employed?

d) From what we have learned by answering the 3 questions above, can we apply technology to improve a physician's ability to diagnose and treat an individual patient with leukemia?

C. Expected Findings

A majority of the literature will focus on EBM using hierarchical graded RCTs and systematic reviews as the way to determine best practices. There will be a positive impact on many areas of medicine including diagnosis, prognosis and therapeutics.

I expect to see documentation that genomics improves disease characterization through gene expression, phenotypic and "omic" analysis (metabolomics etc.). Its impact on the rest of the leukemia literature will be obvious. I also expect that the analytical methods will be frequently introduced in an unnecessarily complex manner.

The need for a point-of-care physicians' system to assist in the prediction of risks and outcomes for individual patients will be suggested in the literature. However, theoretical approaches will be introduced into the literature with few concrete proposals [38].

D. METHODS.

1. Literature Search Strategy

Search the literature on EBM-L, and MCA-L as separate entities, screen and evaluate the content of each search result for evidence-based management of leukemia, genomics in the management of leukemia and the use complex statistical methodologies ("MCA") relating to leukemia. Since EBM-L and MCA-L overlap significantly, with regard to "leukemia", the papers excluded from these categories will form the pool from which the GEN-L papers will be drawn. As with the other categories they will be evaluated heuristically. This will serve to maintain tight topical focus within this study.

Search strategies were designed and executed with the assistance and guidance of a UNC Health Sciences Librarian.

1. *Evidence-based Medicine for Leukemia (EBM-L) Search*

The search was formulated as evidence-based medicine + leukemia retrieved from PubMed with the MESH terms (Table 3):

[Evidence based practice, Evidence-based guidelines, Randomized control trials (RCT), Systematic reviews of RCT, Meta-analyses of RCT]

AND

[Leukemia, leukemic, myeloproliferative neoplasms, hematological malignancies, CML, chronic myeloid leukemia, chronic myelogenous leukemia, AML, acute myeloid leukemia, acute myelogenous leukemia, ALL, acute lymphocytic leukemia, acute lymphoblastic leukemia, CLL, chronic lymphocytic leukemia, chronic lymphoblastic leukemia]

The heuristic for article selection, outlined in table 2, screened for the highest levels of evidence: RCT, systematic review and guidelines present in the leukemia literature. There were 2 comprehensive reviews included in the final article set: one represented an update of an earlier systematic review and the second resulted in an important guideline [21; 39]. Both were landmark articles. The objective of this section of the paper was to review the of evidence-based literature and summarize the major study categories and methods of analysis as well as evaluate their effect on the diagnosis and treatment of individual leukemia patients.

	Search	Inclusion	Exclusion
Evidence-based medicine + leukemia (EBM-L]	[Evidence based practice, Evidence-based guidelines, Randomized control trials (RCT), Systematic reviews of RCT, Meta-analyses of RCT]	RCT, EB Systematic Review, meta-analysis, Comprehensive Review Evidence based literature (EBL), Guidelines English	Summaries, opinion, commentary, case studies, retrospective studies, non-English
	AND		
	[Leukemia, leukemic, myeloproliferative neoplasms, hematological malignancies, CML, chronic myeloid leukemia, chronic myelogenous leukemia, AML, acute myeloid leukemia, acute myelogenous leukemia, ALL, acute lymphocytic leukemia, acute lymphoblastic leukemia, CLL, chronic lymphocytic leukemia, chronic lymphoblastic leukemia]		

Table 3. EBM-L Search Result: 240 records retrieved from PubMed

1. Multivariate Clinical Data Analysis of Leukemia (MCA-L).

This search is labeled: “multivariate clinical data analysis” + leukemia and in this context includes a variety of biostatistical methods retrieved from PubMed. The MESH terms (Table 4) are as follows:

[#1 = ("principal component analysis") OR ("composite variable analysis") OR ("multivariate cohort analysis") OR ("covariate analysis") OR ("non-parametric multivariate analysis")]

[#2 = (leukemia OR leukemic OR "myeloproliferative neoplasms" OR "hematological malignancies" OR Leucocythaemia)]

[#1 AND #2]

The inclusion/exclusion heuristic is outlined in table 4. The objective of this review was to develop a clinician's view of common health data analytical methods, how they are employed and support the diagnosis and treatment of individual leukemia patients.

	Search	Inclusion	Exclusion
Multi variate Clinical data Analysis– Leukemia (MCA-L	[#1 = ("principal component analysis") OR ("composite variable analysis") OR ("multivariate cohort analysis") OR ("covariate analysis") OR ("non-parametric multivariate analysis")]	Clinical & laboratory biostatistical studies studies relevant to leukemia; analytical methods, data analysis, machine learning, systematic reviews, meta-analysis, comprehensive reviews, prospective/ retrospective/ case studies, English	Pharmacokinetics, genomics/genetic/ pathological, diagnostic tests, summaries, opinion, commentary, non-English
	[#2 = (leukemia OR leukemic OR "myeloproliferative neoplasms" OR "hematological malignancies" OR Leucocythaemia)]		
	[#1 AND #2]		

Table 4. Multivariate Clinical data analysis

2. Genomics of Leukemia (GEN-L)

As discussed, earlier EBM-L and MCA-L were broad searches and results were expected to overlap. These overlapping domains including, genomics, were excluded and saved in a separate pool. The GEN-L "search" result was a concatenated set of articles consisting of EBM-L and MCA-L exclusions. This pool was heuristically reviewed as outlined in Table 5. and represented a broad view of genomics and its related fields on our understanding and management of leukemia.

	Search	Inclusion	Exclusion
Genomics of Leukemia (GEN-L)	Concatenated from the EBM-L and PCA-L exclusions.	Targeted drug therapy, genomics/genetic/ metabolomics/proteomics miRNA pathological/ histological diagnostics, English	Diagnostic methods, spectrographic methods, summaries, opinion, commentary, non-English

Table 5. Evidence-based medicine genomics of leukemia

The objective of this section of the paper was to review the literature on the genomics of leukemia, summarize the major study categories and methods of analysis and evaluate effect on the diagnosis and treatment of individual leukemia patients.

E. Data Management and Processing

Search results were received in a spreadsheet format and converted into .CSV file. Article data was sorted, 9 fields were extracted, and exported as text:

PMID
Article Title
Abstract
Journal Name
Journal Year
Journal Volume
Journal Issue
Pagination
Authors

Text was managed in Notepad++ for processing and analysis. Article and reference data were managed with a duplicative combination of ZOTERO and Mendeley. ZOTERO provided rapid capture, organization (synch) and storage of article data with PDFs. However, ZOTERO did not consistently capture complete metadata. Mendely, is much slower synching its reference databases (crashes periodically), its capacity to automatically import ZOTERO data files served as a backup and fills in metadata by drawing from its online network. Mendely regularly frequently fails to capture PDFs and urls. A strong feature of Mendeley was its ability to highlight, capture, and condense main ideas and salient points into notes for later text analysis (ADOBE will soon be better).

EBM-L, MCA-L and GEN-L final article sets were processed individually for analysis and comparison:

- The EBM-L articles were alphabetized by title and labeled under 2 categories: "Study Category" and "Analysis Category". Studies were either Diagnostic, Prognostic or Therapeutic and the analyses were published as either a Systematic Review, Comprehensive Review, Guideline, or RCT.

EBM-L Study Categories
Therapeutic
Prognostic
Diagnostic
Total

EBM-L Analysis Category
Systematic Review
Comprehensive Review EBL
Guidelines
RCT
Total

Table 6. EBM-L Study and analysis categories

- The MCA-L articles were alphabetized by title and labeled under 3 categories: "Study

MCA-L Study Categories	MCA-L Sub-category
Data Cleaning	Gene Expression analysis
Diagnostic	Genome Analysis
Prognostic	Metabolome / Proteome
Therapeutic	Phenotype Analysis

Table 7. MCA: Study categories and sub-categories

Category", "Sub-category" and "Methods of Data Analysis". Studies were either Data Cleaning, Diagnostic, Prognostic or Therapeutic and Sub-categorized as either Gene Expression Analysis, Genome Analysis, Metabolome/Proteome Analysis or Phenotype Analysis.

MCA-L Methods of Data Analysis
ANOVA
Canonical correlation analysis
Clustering
Cronbach's alpha
Fisher exact test
Functional principal component analysis
Gaussian model
Gibbs sampling
Kernel principal component analysis
Kruskal–Wallis H test
Mann-Whitney U test
Mean residual life regression model
Multiple regression analysis
k nearest neighbor (classification)
Nonnegative matrix factorization
Partial Least Squares (PLS)
Parzen windows
Pearson's Chi Squared test
Principal Component Analysis (PCA)
Principal factor analysis
Proportional hazards model (Cox's analysis)
Quantile residual lifetime regression
Relative expression level
Sliced inverse regression
Support vector machines
Total Principal Component regression (TPCR)

Table 8. MCA-L Methods of data analysis

- The GEN-L articles were alphabetized by title and labeled under 3 categories: "Study Category", "Sub-category" and "Methods of Data Analysis" (Table 10). Categories were either *Data Cleaning, Diagnostic, Prognostic or Therapeutic* (Table 9). Sub-categories were either *Gene Expression Analysis, Genome Analysis, Metabolome/Proteome Analysis or Phenotype Analysis* (Table 9).

GEN-L Study Categories
Data Cleaning
Diagnostic
Prognostic
Therapeutic

GEN-L Sub-categories
Gene Expression analysis
Genome Analysis
Metabolome
Phenotype Analysis

Table 9. GEN-L Study categories and sub-categories

Methods of Data Analysis
Statistical analysis of clusters
ANOVA
Covariate analysis
Cox's analysis
Dynamic Cross-Correlation Map (DCCM)
Feature clustering using the Bayes information criterion
Fisher's ratio and fold change
Genomic nucleotide sequence
Levene test
Linear Discriminant Analysis (LDA)
Linear-mixed model
Logistic Discrimination
Mann-Whitney test
multivariate statistical analysis
Partial Least Squares (PLS)
Per-residue Root Mean Square Deviation (per-residue RMSD)
Principal Component Analysis (PCA)
Principal component discriminant analysis (PCDA)
Proteomic analysis
Quadratic Discriminant Analysis
Rissanen's MDL
Root-mean-square deviation (RMSD)
Root mean square fluctuation (RMSF)
Tukey HSD method of one-way ANOVA
Unsupervised hierarchical clustering analysis (HCA)

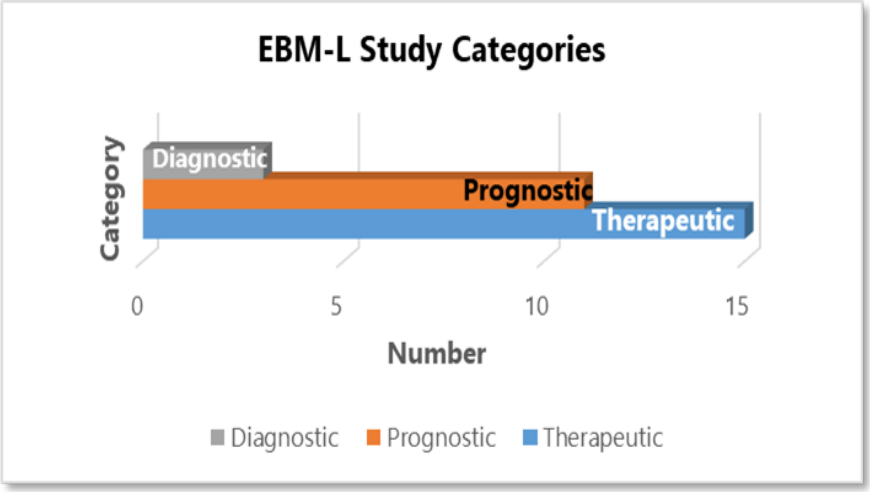
Table 10. GEN-L Methods of data analysis

Methods of data analysis were collated and listed in table 10.

F. Results and Analysis

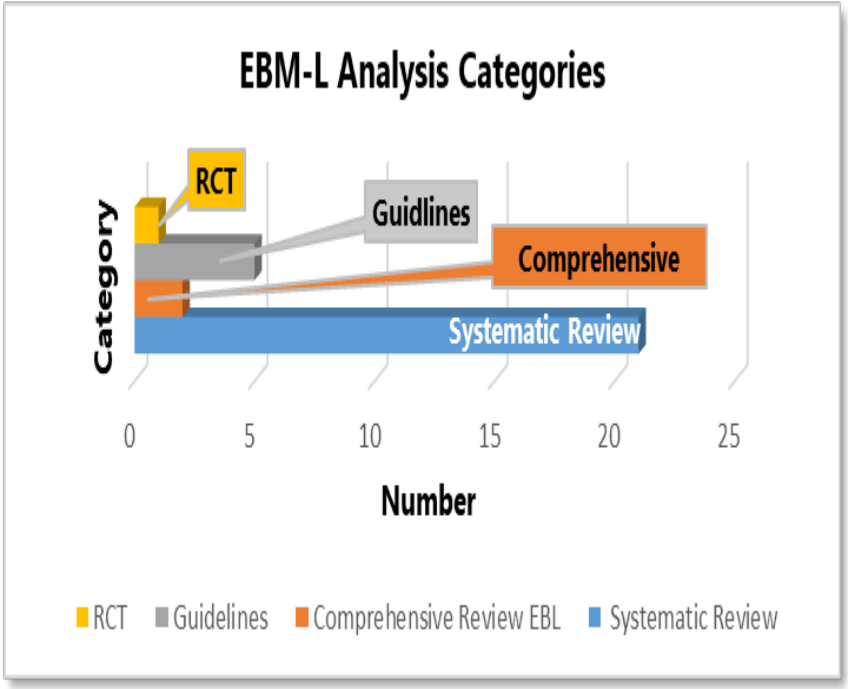
The evidence-based medicine- leukemia (EBM-L) search yielded 240 records. The multivariate clinical data analysis– leukemia (MCA-L) search yielded 215 records. Following the inclusion/exclusion process 29 EBM-L and 33 MCA-L relevant papers resulted. The GEN-L "search" result was a concatenated set of 392 articles consisting of 211 EBM-L and 181 MCA-L exclusions. After screening there were 33 relevant GEN-L papers.

2. The 29 EBM-L articles were analyzed for study type (category), illustrated in "graph and table 1 "and analysis category, illustrated in "graph and table 2":



EBM-L Study Categories	Number
Therapeutic	15
Prognostic	11
Diagnostic	3
Total	29

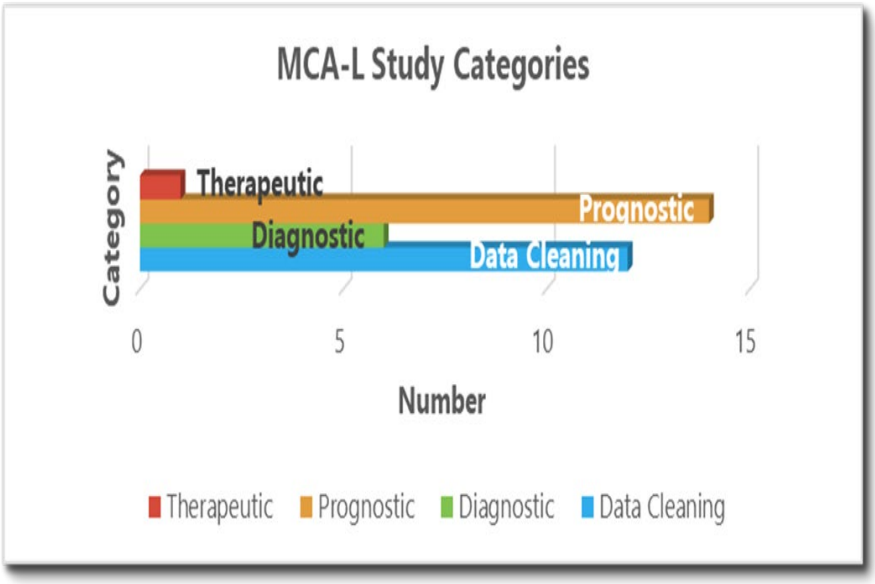
Graph and table 1. GEN-L study types or categories



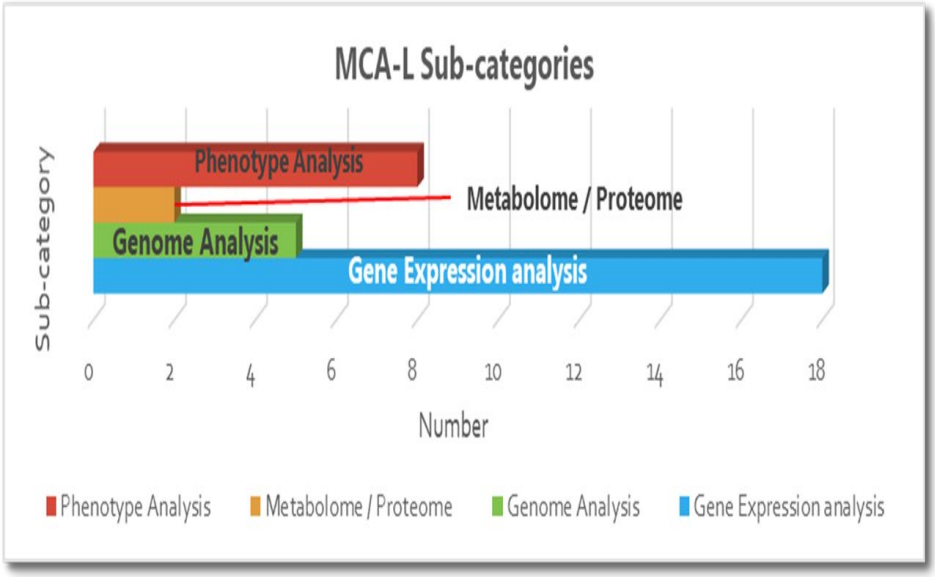
EBM-L Analysis Category	Number
Systematic Review	21
Comprehensive Review EBL	2
Guidelines	5
RCT	1
Total	29

Graph and table 2. GEN-L Analysis categories

- The 33 MCA-L articles were analyzed for study category, sub-category and methods of data analysis:

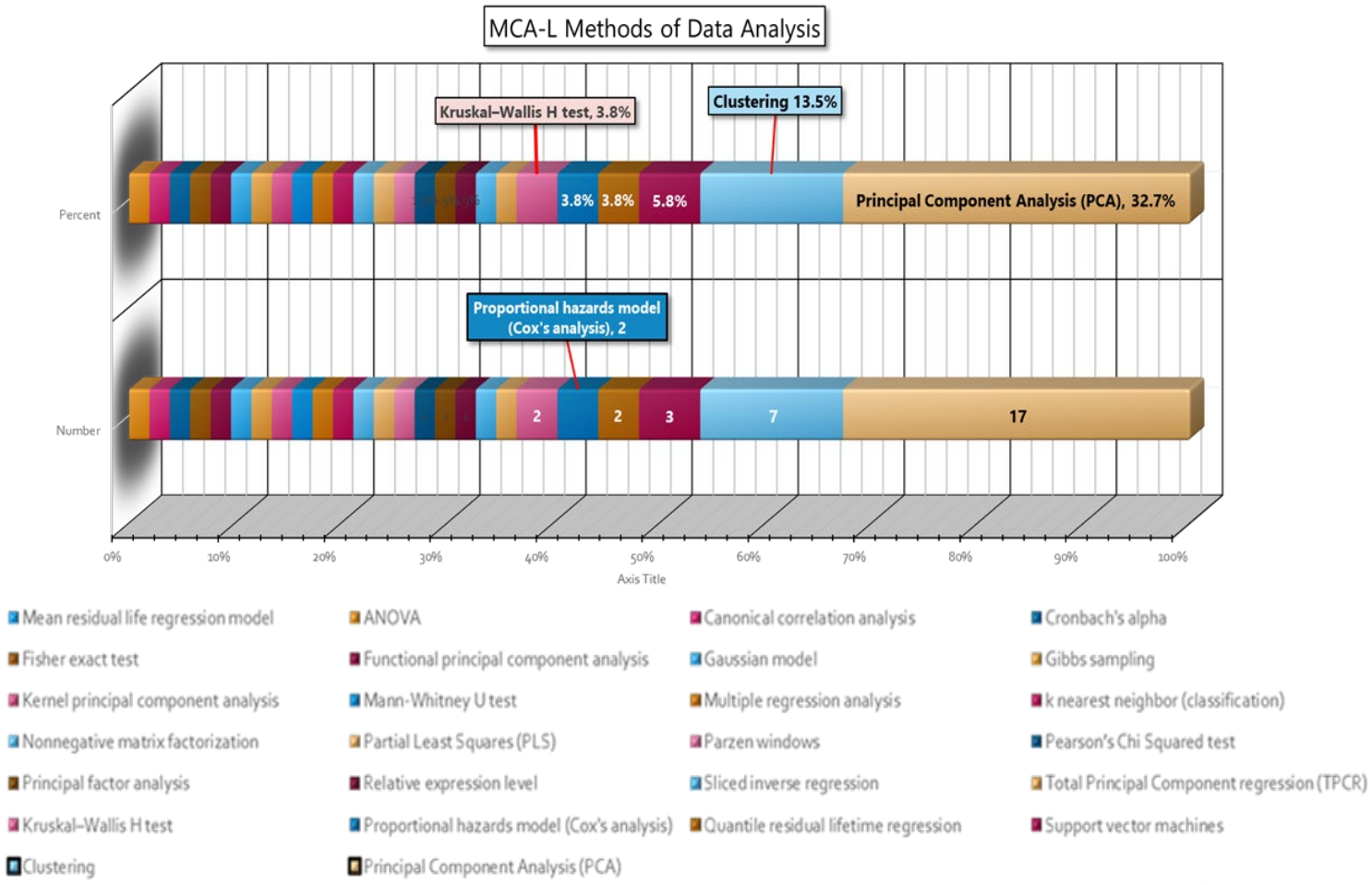


MCA-L Study Categories	Number
Data Cleaning	12
Diagnostic	6
Prognostic	14
Therapeutic	1
Total	33



MCA-L Sub-category	Number
Gene Expression analysis	18
Genome Analysis	5
Metabolome / Proteome	2
Phenotype Analysis	8
Total	33

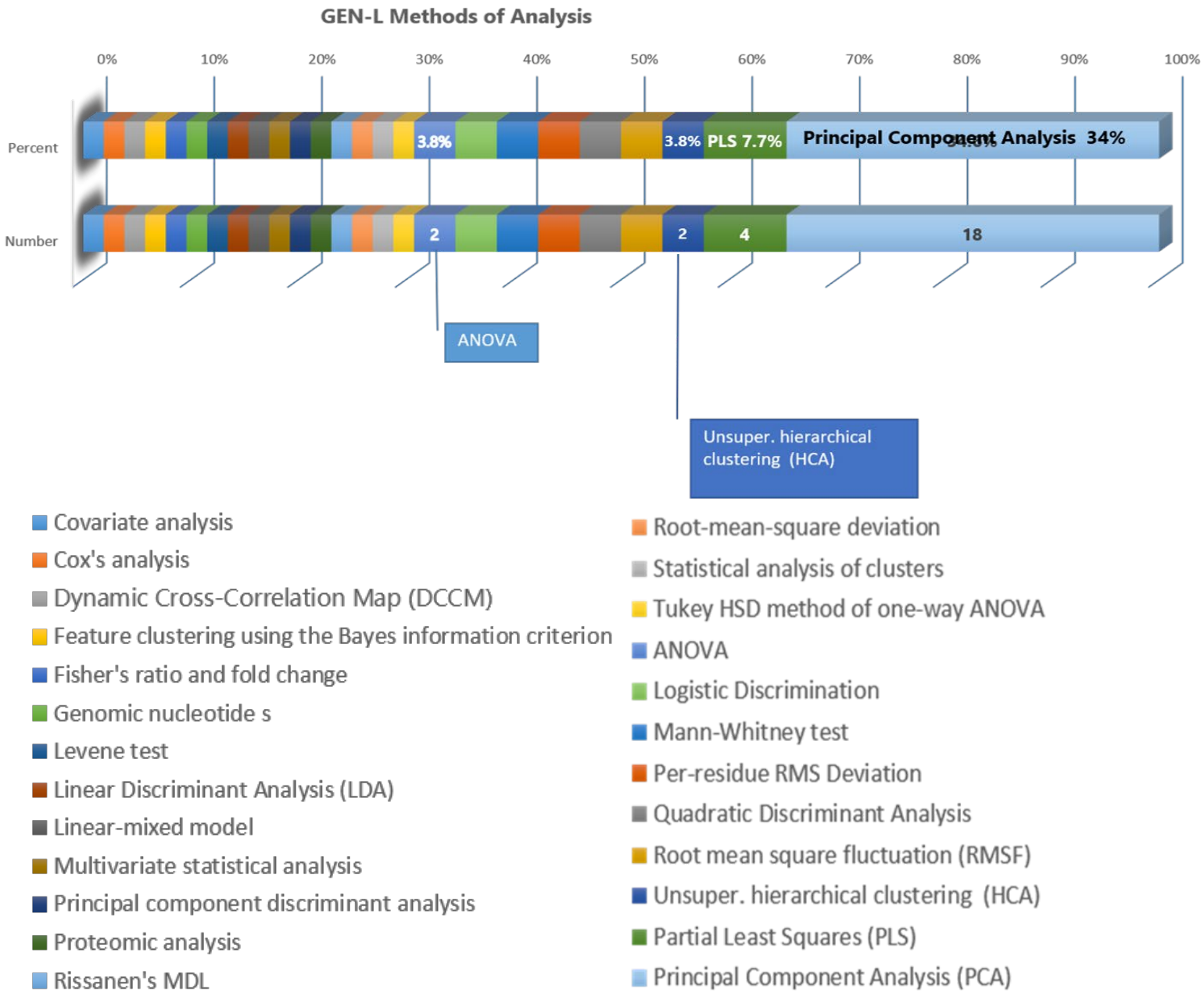
Graph and table 3. MCA-L study types or categories and sub-categories



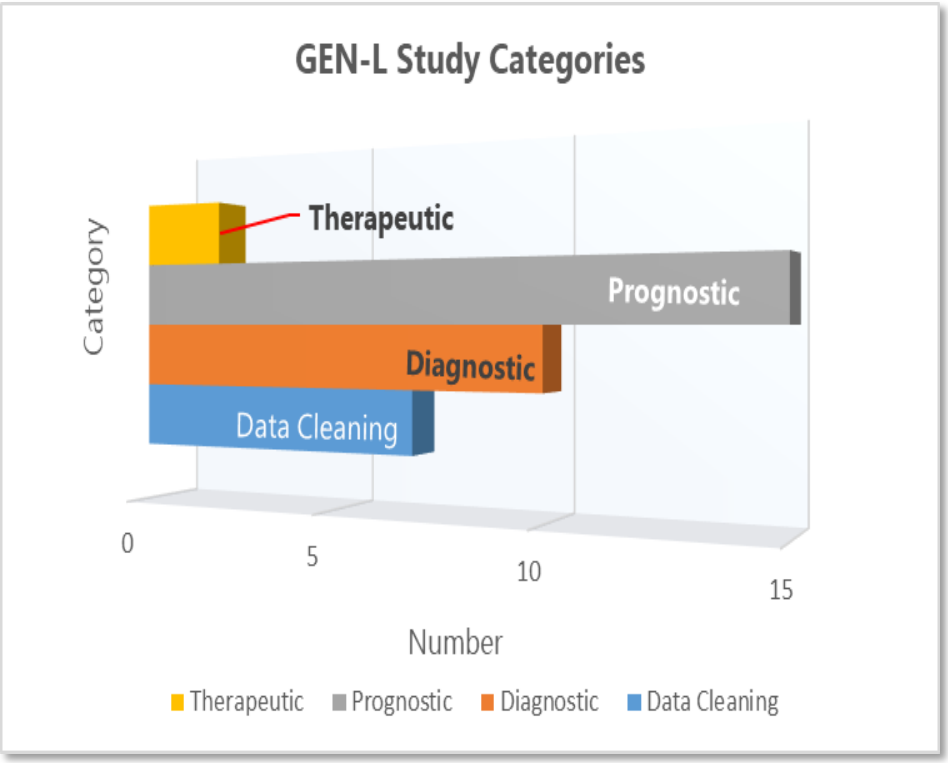
Graph and table 4. MCA-L Methods of data analysis

There was a gratifying finding within this group of papers. Rigolin et al in their 1994 paper used cluster analysis to group patient presenting features (symptoms/finding) to determine prognosis [38]. Principal component analysis was used as a method for feature reduction and Cox's proportional hazard model was used to evaluate prognosticating features. These were subsequently clustered in to groups to simplify interpretation.

- The 33 GEN-L articles were analyzed for methods of analysis (graph and table 5) and study category, sub-category (graph and table 6)



Graph and table 5. GEN-L methods of analysis



GEN-L Study Categories	Number
Data Cleaning	7
Diagnostic	10
Prognostic	15
Therapeutic	2
Total	34

Graph and table 6. GEN-L study types or categories

There was no literature that addressed the concept of cohort multidimensional analysis (CoMA) of warehoused EHR data for the precision treatment of individual patients with leukemia

G. Discussion

1. Has evidence-based medicine (EBM -L) improved a physician's ability to diagnose and treat an individual patient with leukemia?

This comprehensive review documents a robust evidence-based literature for determining the efficacy of a wide variety of therapeutic, prognostic and diagnostic approaches to the management of leukemia derived from population based biostatistical methodologies from systematic reviews. It demonstrates that efficacy is calculated as a population statistic for the "average patient", a theoretical homogenous representor for what is really a heterogenous population. The flaws and limitation of evidence-based medicine for leukemia are not really addressed. Some studies discuss limitations associated with quality or abundance of evidence but rarely, is the lack of evidence for the risks and outcomes associated with the generalization of these treatment to more genialized populations.

Efficacy is determined for the treatment of a disease, not an individual plus the disease. The main effects are measured. The effects associated with individual patients with confounding factors such as comorbidities, age, race, gender, socioeconomic and environmental exposure. These are the phenotypic variants that put patients into sub-categories within RCTS. There are a large number of variants populated by small numbers of patients within the context of RCTs. The

effects from these groups become statistically insignificant and therefore have little impact on final study outcomes. They are washed out in EBM RCTs. The ability to treat these patients safely and with predictable risks and outcomes is what defines precision medicine.

Frequently these phenotypic variants have risks and outcomes inconsistent with published results. When treatment populations expand from 1,000s to 100s of millions, those statistically insignificant effects during RCTs become unexpected consequences and result in recalls. Those sub-categories with confounding factors rapidly populate.

Although intuitively attractive, skeptics noted very little supporting data accompanied the promises of waste reduction, cost savings and better patient outcomes. The ethical justification, physician decision making would be better informed and less biased, to change medical education and clinical practice was unsubstantiated [5].

After nearly 3 decades of EBM, a growing voice in the medical literature was bringing attention to the notion that evidence based medicine has failed to meet its original expectations. *To provide practicing clinicians sufficient evidence to predict the risks and outcomes faced by individual patients in response to treatment. [40]* RCTs supply population-based risk and outcomes data on diseases (phenotypes) and limited risk and outcomes data on subcategories of disease (phenotypic variants) i.e.; individual patients. They have a wide range of individual risks and responsiveness based on confounding factors and the doctor-patient relationship was not encouraged. As a result, EBM by itself, lacked sufficient power to accurately predict an individual patient's outcome from any diagnostic or treatment option.

2. Does the clinical human genomics of leukemia (GEN-L) study the diagnosis and treatment of individual patients with leukemia?

As genomics expanded, pharmaceutical companies increased drug screening. This made thousands of active compounds available. RCTs vetted promising drugs and the number of systematic reviews evaluating potentially new drugs exponentially increased.

To qualify for RCTs subjects must meet narrowly defined criteria attempting to create as homogenous a population as possible. Human populations and their diseases are naturally heterogeneous. Populations invariably contain subcategories (phenotypic variants (5). As a result, the main affects (efficacy, low toxicity etc) of a drug achieve statistical significance for a representative average patient that may or may not exist but rarely achieve statistical significance for phenotypic variants (subcategories) [35]. However, it should be noted that the greater the homogeneity of the study pool the higher the probability targeted affects are statistically significant and predict risks and outcomes. This raises questions of patient selection bias because pharmaceutical companies fund a substantial number of these studies. Some studies are not reproducible [5].

3. What are common methods of multivariate clinical data analysis in the leukemia literature (MCA-L) and how were they employed?

Principal component analysis dominated the statistical landscape in more than 30% of MCA-L and GEN-L publications. PCA determined the impact of variables on outcomes without consideration of correlation. It was used to reduce dimensionality of datasets.

Within the GEN-L literature the next most frequent method of analysis, at nearly 8%, was partial least squares or PLS (regression). PLS regression is related to principal component regression (PCR) and therefore PCA. In short PLS /PCR are applied upon correlations where as PCA depends upon the rank of projected maximum variance between independent variables and response. In the MCA-L literature clustering was important for cohort determination and was used in 13.5% of publications.

Clearly the literature demonstrates that the methodologies exist for analysis of complex datasets. Multivariate analytics/PCA made gene expression analysis results meaningful and genomics comprehensible from a clinical standpoint.

Finally, no methodologies for precision medicine will approximate cohort multidimensional analysis (CoMA).

4. From what we have learned by answering the 3 questions above, can we apply technology to improve a physician's ability to diagnose and treat an individual patient with leukemia?

Rigolin et al.'s 1994 paper, approximated the methodology proposed in CoMA. Symptoms and findings of AML patients were ranked by principal component analysis. Prognostic value was derived by Cox's proportional hazard model [38]. In this example age was found to be a prognosticating factor and clustering patient features modeled patient presentations. This study was, enlightening because Cox's univariate models was also compared to other multi-variate models including PCA. It was encouraging, because on a very small scale, patient matched cohort modeling was used to individualized care.

Clearly methodologies exist for statistical analysis of complex datasets and coupled with natural language processing for processing of clinical unstructured electronic health data the analytical technology exists to provide individualized patient care using the proper baseline model.

III. CONCLUSION

Evidence-based medicine for leukemia (EBM-L) is currently the *sine qua non* for disease management, relying predominantly upon systematically reviewed hierarchically graded randomized controlled trials (RCTs). and yield population-based risks and outcomes for disease treatments. EBM promised cost effective individualized "superior patient care" (3, p. 2421) by vetting diagnostic and treatment methodologies using biostatistically derived evidence from RCTs. This brought uniformity to practice but was ineffective for precision management of the individual leukemic patients. This characteristic of population-based results is a strength and weakness of EBM.

For determining treatment efficacy, it is a strength. The RCT "average patient", does not resemble the "every day" patients seen by physicians [35]. It is a weak methodology for predicting the clinical risks and outcomes for individual patients. The patients with the highest

need are within the sub-cohorts[21]. In the RCT driven EBM model they are statistically insignificant. Abandoning the mechanistic understanding of disease and clinical experience, for epidemiological and biostatistical medical research evidence alone has significant limitations. The phenotypic variants, the patients with comorbidities, are the real-world patients that build wisdom and experience [41]. It has failed to advance medicine towards its original goal of improving the physician's ability to manage the individual patient [42]. EBM lacks sufficient power to accurately predict an individual patient's outcome from a particular diagnostic or treatment option.

Cohort multidimensional analysis (CoMA) based on principal component analysis leverages real world clinical experience by mining data from EHRs, health data warehouses, disease data marts, genomic and other specialty databases (pharmacy, oncology, diabetes, etc.). Data is extracted aggregated and analyzed based on an individual patient's characteristics, providing evidence of treatment efficacy on patients that vary from the idealized RCT patient[22].

Diagnoses, demographics, confounding factors and comorbidities are patient variables or attributes. Observable attributes are signs and symptoms for example. When grouped, these presenting traits or observable expressions represent disease and define a disease's phenotype. If the attributes are general or non-specific, they represent a group of similar diseases or phenotypes. If the attributes are unique or detailed, they may define a specific phenotype or even a phenotypic variant. Individual patients present with patterns of attributes. When these attributes are transformed into a *composite variable,* a template or fingerprint is created representing that patient and their phenotypic variant. When aggregated and clustered they define cohorts. An individual patient is the *index composite variable (ICV).* ICVs are used to extract and aggregate data from information sources into attributable cohorts[23]. Attributable cohorts contain specific patient matched data: disease designation, demographics, treatments, genotype, phenotype, confounding factors and comorbidities[24]. This data when recorded over time represents the course of a patient's disease. Time series data is important because it can be used to forecast a patient's disease trajectory given a set of circumstances or hypothesizing a set of circumstances. In other words, an outcome may be forecasted with the current treatment and stage of disease or a hypothetical outcome may be forecasted given a proposed change in treatment with the same stage of disease.

---

[21] Sub-cohorts contain phenotypic variations and represent matched patients with confounding factors and comorbidities.
[22] These are the phenotypic variants or sub-cohorts within RCTs
[23] another general description is "phenotypically matched cohorts"
[24] These can be individual entities or sets of related entities (disease, phenotype, treatment etc)

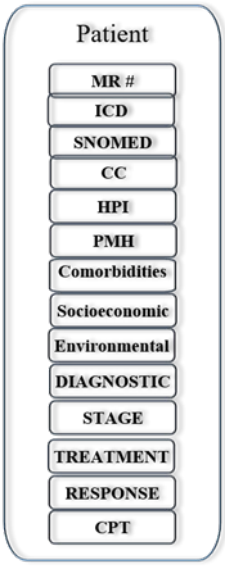


Figure 7. Index Composite Variable from patient attributes

Modified principal component analysis, using the index composite variable has the potential to distinguish cohorts with attributes that match (patient matched, disease matched, phenotype matched, treatment match, and genotype matched cohorts) or approximate the index composite variable. Clinical decision models and predictive models can be built for an individual patient.

A methodology is therefore needed that can build attributable cohorts using ICVs. PCA is a likely candidate to enable this process. The methodology will be applied to a single known health information dataset either sourced from the Carolina Data Warehouse for Health (CDW) or synthetically derived for testing. Then to multiple information sources. The long-term goal is to develop a point-of-care system that presents models based on patients' ICVs for medical decision making and predictive analytics.

There are significant challenges associated with this task. For example, using categorical variables with continuous variables needs to be addressed. The uniformity and quality of data within information sources can present impediments. Patient data is not stored in longitudinal sets and will require a degree of access that permits assemblage of individual patient timelines and patient matched timelines.

A brief example of how to overcome the categorical/continuous variable problem would be to use the categorical variable for cohort development and the continuous variable for PCA

The evidence for personalizing care lies within the EHR and specialized clinical databases [43]. This structured and unstructured data represents real-world treatment experience of complex diversified patients. Experience-based evidence derived from clinical health data using cohort multi-dimensional analysis (CoMA-HD) is proposed as a methodology to use this data. Medicine has reached an asymptote in quality and personalization and will not advance without a paradigm change – the inclusion of clinical experience-based evidence in decision making.

IV. REFERENCES.

1. Klausner R. Leukemia, Lymphoma, and Multiple Myeloma: Toward a New Understanding. Washington, DC: Senate Appropriations Subcommittee on Labor, Health and Human Services, Education, and Related Agencies; 2001. https://www.cancer.gov/about-nci/legislative/hearings/2001-leukemia-lymphoma-multiple-myeloma-new-understanding.pdf .

2. Fuller RB, Kuromiya K. *Critical Path (PDF) | Senses | Human*. 1st ed. St. Martin Press; 1981. Accessed August 4, 2019. https://www.scribd.com/doc/310920820/R-Buckminster-Fuller-Critical-Path-PDF

3. Gordon Guyatt, Cairns J, Churchill D, et al. Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine (EB Working Group-1992). JAMA. 1992;268(17):2420-2425. doi:10.1001/jama.1992.03490170092032

4. IOM C on I of HC. *1st Report To Err Is Human: Building A Safer Health System*.; 1999:312. Accessed January 6, 2013. https://tinyurl.com/yak9ynkd

5. Rogers, W., & Hutchison, K. (2017). Evidence-Based Medicine in Theory and Practice: Epistemological and Normative Issues. In T. Schramme & S. Edwards (Eds.), Handbook of the Philosophy of Medicine (pp. 851–872). https://doi.org/10.1007/978-94-017-8688-1_40

6. Link between health spending and life expectancy: The US is an outlier. Our World in Data. Accessed June 27, 2020. https://ourworldindata.org/the-link-between-life-expectancy-and-health-spending-us-focus

7. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *BMJ*. 1995;310(6987):1122-1126. doi:10.1136/bmj.310.6987.1122

8. Djulbegovic B, Guyatt GH, Ashcroft RE. Epistemologic Inquiries in Evidence Based Medicine. Cancer Control. 2009;16(2):158-168. doi:10.1177/107327480901600208

9. Goodman KW. Ethics and Evidence-Based Medicine: Fallibility and Responsibility in Clinical Science. New York: Cambridge University Press; 2003.

10. Bennett KJ, Sackett D, Haynes B, Neufeld V, Tugwell P, Roberts R. A controlled trial of teaching critical appraisal of the clinical literature to medical students. JAMA: The Journal of the American Medical Association. 1987;257(18):2451-2454. doi:10.1001/jama.257.18.2451

11. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. BMJ. 1995;310(6987):1122-1126. doi:10.1136/bmj.310.6987.1122

12. Cochrane Reviews | Cochrane Library. Cochrane Library. https://www.cochranelibrary.com/. Published August 3, 2019. Accessed August 7, 2019

13. George JN, Woolf SH, Raskob GE. The Evidence-Based Analysis of Treatment for Chronic Myeloid Leukemia: An Introduction to Its Methods and Clinical Implications. Blood. 1999;94(5):1515-1516. http://www.bloodjournal.org/content/94/5/1515. Accessed September 25, 2019.

14. Burns PB, Rohrich RJ, Chung KC. The Levels of Evidence and their role in Evidence-Based Medicine. *Plast Reconstr Surg*. 2011;128(1):305-310. doi:10.1097/PRS.0b013e318219c171

15. Merriam-Webster's collegiate dictionary. GENOMICS. In: Merriam-Webster's Collegiate Dictionary. Merriam-Webster, Incorporated; 2019. https://www.merriam-webster.com/dictionary/genomics .

16. Mendel, Gregor. 1866. Versuche über Plflanzenhybriden. Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865, Abhandlungen, 3–47

17. Boveri T. Concerning the Origin of Malignant Tumours by Theodor Boveri. Translated and annotated by Henry Harris. Journal of Cell Science. 2008;121(Supplement 1):1-84. doi:10.1242/jcs.025742

18. Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*. 1975;98(3):503-517. doi:10.1016/S0022-2836(75)80083-0

19. Desforges JF, Ross JD, Moloney WC. Mechanisms of anemia in leukemia and malignant lymphoma. The American Journal of Medicine. 1960;28(1):69-76. doi:10.1016/0002-9343(60)90224-2

20. Levinson AD, Oppermann H, Levintow L, Varmus HE, Bishop JM. Evidence that the transforming gene of avian sarcoma virus encodes a protein kinase associated with a phosphoprotein. *Cell*. 1978;15(2):561-572. doi:10.1016/0092-8674(78)90024-7

21. Skoetz N, Kluge S, Monsef I, et al. Twelfth Biannual Report of the Cochrane Haematological Malignancies Group—Focus on Tyrosine Kinase Inhibitors. J Natl Cancer Inst. 2011;103(4):E1-E1. doi:10.1093/jnci/djq531

22. Silver RT, Woolf SH, Hehlmann R, et al. An Evidence-Based Analysis of the Effect of Busulfan, Hydroxyurea, Interferon, and Allogeneic Bone Marrow Transplantation in Treating the Chronic Phase of Chronic Myeloid Leukemia: presented at ASH, December 5, 1998, Miami Beach, FL. *Blood*. 1999;94(5):1517-1536. http://www.bloodjournal.org/content/94/5/1517

23. Slovak ML, Theisen A, Shaffer LG. Human Chromosome Nomenclature: An Overview and Definition of Terms. In: Gersen SL, Keagle MB, eds. The Principles of Clinical Cytogenetics. New York, NY: Springer; 2013:23-49. doi:10.1007/978-1-4419-1688-4_3

24. Hayflick L. The limited in vitro lifetime of human diploid cell strains. Experimental Cell Research. 1965;37(3):614-636. doi:10.1016/0014-4827(65)90211-9

25. Pierotti MA, Sozzi G, Croce CM. Mechanisms of oncogene activation. In: Kufe D, Pollock R, Weichselbaum R, eds. Holland-Frei Cancer Medicine. 6th ed. BC Decker; 2003. https://www.ncbi.nlm.nih.gov/books/NBK12538/.

26. Hochhaus A. Imatinib mesylate (Glivec, Gleevec) in the treatment of chronic myelogenous leukemia (CML) and gastrointestinal stromal tumors (GIST). Ann Hematol. 2004;83 Suppl 1:S65-66. doi:10.1007/s00277-004-0850-2

27. Gentile M, Petrungaro A, Uccello G, et al. Venetoclax for the treatment of chronic lymphocytic leukemia. Expert Opinion on Investigational Drugs. 2017;26(11):1307-1316. doi:10.1080/13543784.2017.1386173

28. Shlens J. A Tutorial on Principal Component Analysis. arXiv:14041100 [cs, stat]. Published online 2014. http://arxiv.org/abs/1404.1100

29. Principal component analysis. In: *Wikipedia*. ; 2020. Accessed June 29, 2020. https://en.wikipedia.org/w/index.php?title=Principal_component_analysis&oldid=964850741

30. Naik GR, ed. *Advances in Principal Component Analysis*. Springer Singapore; 2018. Accessed April 9, 2020. http://link.springer.com/10.1007/978-981-10-6704-4

31. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 1901;2(11):559-572. doi:10.1080/14786440109462720

32. Hotelling H. Analysis of a complex of statistical variables into principal components Part I. Journal of Educational Psychology. 1933;24(6):417-441. doi:10.1037/h0071325

33. Hotelling H. Analysis of a complex of statistical variables into principal components Part II. Journal of Educational Psychology. 1933;24(7):498-520. doi: 10.1037/h0070888

34. Pepe PE, Aufderheide TP. EBM vs. EBM: combining evidence-based and experienced-based medicine in resuscitation research. Current Opinion in Critical Care. 2017;23(3):199-203. doi:10.1097/MCC.0000000000000413

35. Horwitz RI, Singer BH. Why evidence-based medicine failed in patient care and medicine-based evidence will succeed. J Clin Epidemiol. 2017;84:14-17. doi:10.1016/j.jclinepi.2017.02.003

36. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. BMJ Evidence-Based Medicine. 2016;21(4):125-127. doi:10.1136/ebmed-2016-110401

37. McEwen BS, Stellar E. Stress and the Individual: Mechanisms Leading to Disease. Arch Intern Med. 1993;153(18):2093-210. doi:10.1001/archinte.1993.00410180039004

38. Rigolin GM, Fagioli F, Spanedda R, et al. Study of prognosis in acute myeloid leukemias (AML) by cluster analysis. Haematologica. 1994;79(3):233-240. http://www.haematologica.org/content/79/3/233.

39. Oliansky, D. M., Larson, R. A., Weisdorf, D., Dillon, H., Ratko, T. A., Wall, D., McCarthy, P. L., & Hahn, T. (2012). The Role of Cytotoxic Therapy with Hematopoietic Stem Cell Transplantation in the Treatment of Adult Acute Lymphoblastic Leukemia: Update of the 2006 Evidence-Based Review. *Biology of Blood and Marrow Transplantation*, *18*(1), 18-36.e6. https://doi.org/10.1016/j.bbmt.2011.07.019

40. Sackett DL. Evidence-based medicine. Seminars in Perinatology. 1997;21(1):3-5. doi:10.1016/S0146-0005(97)80013-4

41. Current Guidelines Have Limited Applicability to Patients with Comorbid Conditions: A Systematic Analysis of Evidence-Based Guidelines - ProQuest. https://search-proquest-com.libproxy.lib.unc.edu/docview/1310231377?pq-origsite=summon .

42. Simons MR, Zurynski Y, Cullis J, Morgan MK, Davidson AS. Does evidence-based medicine training improve doctors' knowledge, practice and patient outcomes? A systematic review of the evidence. Medical Teacher. 2019;41(5):532-538. doi:10.1080/0142159X.2018.1503646

43. Ross MK, Wei W, Ohno-Machado L. "Big Data" and the Electronic Health Record. Year b Med Inform. 2014;9(1):97-104. doi:10.15265/IY-2014-0003

44. Barillot E., Laurence Calzone L., Philippe Hupé P., Jean-Philippe Vert J., Andrei Zinovyev A, Computational Systems Biology of Cancer Chapman & Hall/CRC Mathematical & Computational Biology , 2012

45. Janssens A, Andre M, Berman Z, et al. Effectiveness and safety of ibrutinib for chronic lymphocytic leukemia (CLL) in routine clinical practice: interim analysis (IA) of the Belgian ibrutinib real-world data (BIRD) study. In: EHA; 2019. https://library.ehaweb.org/eha/2019/24th/266184/ann.janssens.effectiveness.and.safety.of.ibrutinib.for.chronic.lymphocytic.html?f=listing%3D0%2Abrowseby%3D8%2Asortby%3D1%2Asearch%3Dpf384.

V. ARTICLE LIST FROM SEARCH BY CATEGORY.

A. EBM-L

1. Breems DA, Löwenberg B. Acute Myeloid Leukemia and the Position of Autologous Stem Cell Transplantation. Seminars in Hematology. 2007;44(4):259-266. doi: 10.1053/j.seminhematol.2007.08.002

2. Berry SM, Broglio KR, Berry DA. Addressing the Incremental Benefit of Histamine Dihydrochloride When Added to Interleukin-2 in Treating Acute Myeloid Leukemia: A Bayesian Meta-Analysis. Cancer Investigation. 2011;29(4):293-299. doi:10.3109/07357907.2011.568563

3. Silver RT, Woolf SH, Hehlmann R, et al. An Evidence-Based Analysis of the Effect of Busulfan, Hydroxyurea, Interferon, and Allogeneic Bone Marrow Transplantation in Treating the Chronic Phase of Chronic Myeloid Leukemia: Developed for the American Society of Hematology: Presented in part at the Education Session of the American Society of Hematology, December 5, 1998, Miami Beach, FL. Blood. 1999;94(5):1517-1536.

4. Sousa DC, Rodrigues FB, Duarte G, Campos F, Pinto F, Vaz-Carneiro A. Anterior ischemic optic neuropathy and hematologic malignancy: a systematic review of case reports and case series. Canadian Journal of Ophthalmology. 2016;51(6):459-466. doi: 10.1016/j.jcjo.2016.05.018

5. Thakar MS, Forman SJ. ASH evidence-based guidelines: is there a role for second allogeneic transplant after relapse? Hematology. 2009;2009(1):414-418. doi:10.1182/asheducation-2009.1.414

6. Xu X, Bryke C, Sukhanova M, et al. Assessing copy number abnormalities and copy-neutral loss-of-heterozygosity across the genome as best practice in diagnostic evaluation of acute myeloid leukemia: An evidence-based review from the cancer genomics consortium (CGC) myeloid neoplasms working group. Cancer Genetics. 2018;228-229:218-235. doi:10.1016/j.cancergen.2018.07.005

7. Owen C, Gerrie AS, Banerji V, et al. Canadian evidence-based guideline for the first-line treatment of chronic lymphocytic leukemia. Curr Oncol. 2018;25(5):e461-e474. doi:10.3747/co.25.4092

8. Adam M, Rebholz CE, Egger M, Zwahlen M, Kuehni CE. Childhood leukaemia and socioeconomic status: what is the evidence? Radiat Prot Dosimetry. 2008;132(2):246-254. doi:10.1093/rpd/ncn261

9. Kheifets L, Shimkhada R. Childhood leukemia and EMF: review of the epidemiologic evidence. Bioelectromagnetics. 2005;Suppl 7:S51-59. doi:10.1002/bem.20139

10. Baumann FT, Bloch W, Beulertz J. Clinical exercise interventions in pediatric oncology: a systematic review. Pediatr Res. 2013;74(4):366-374. doi:10.1038/pr.2013.123

11. Hehlmann R. CML-Where do we stand in 2015? Ann Hematol. 2015;94(2):103-105. doi:10.1007/s00277-015-2331-1

12. Sasse EC, Sasse AD, Brandalise SR, Clark OAC, Richards S. Colony-stimulating factors for prevention of myelosuppressive therapy-induced febrile neutropenia in children with acute lymphoblastic leukaemia. Cochrane Database of Systematic Reviews. 2005;(3). doi:10.1002/14651858.CD004139.pub2

13. Visani G, Olivieri A, Malagola M, et al. Consolidation therapy for adult acute myeloid leukemia: A systematic analysis according to evidence-based medicine. Leukemia & Lymphoma. 2006;47(6):1091-1102. doi:10.1080/10428190500513595

14. Gale RP, Edward Park R, Dubois RW, et al. Delphi-panel analysis of appropriateness of high-dose therapy and bone marrow transplants in adults with acute myelogenous leukemia in 1st remission. Leukemia Research. 1999;23(8):709-718. doi:10.1016/S0145-2126(99)00044-2

15. Turner M, Gagnon D, Lagace M, Gagnon I. Effect of treatment for paediatric cancers on balance: what do we know? A review of the evidence. European Journal of Cancer Care. 2013;22(1):3-11. doi:10.1111/ecc.12019

16. Abiri B, Kelishadi R, Sadeghi H, Azizi-Soleiman F. Effects of Maternal Diet During Pregnancy on the Risk of Childhood Acute Lymphoblastic Leukemia: A Systematic Review. Nutr Cancer. 2016;68(7):1065-1072. doi:10.1080/01635581.2016.1206581

17. Zhang X, Li Y, Zhang Y, et al. Etoposide in combination with low-dose CAG (cytarabine, aclarubicin, G-CSF) for the treatment of relapsed or refractory acute myeloid leukemia: A multicenter, randomized control trial in southwest China. Leukemia Research. 2013;37(6):657-664. doi:10.1016/j.leukres.2013.03.005

18. Cornely OA, Ullmann AJ, Karthaus M. Evidence-based assessment of primary antifungal prophylaxis in patients with hematologic malignancies. Blood. 2003;101(9):3365-3372. doi:10.1182/blood-2002-05-1356

19. Shelton BK. Evidence-based care for the neutropenic patient with leukemia. Seminars in Oncology Nursing. 2003;19(2):133-141. doi:10.1016/S0749-2081(03)00026-3

20. Kharfan-Dabaja MA, Hamadani M, Reljic T, et al. Gemtuzumab ozogamicin for treatment of newly diagnosed acute myeloid leukaemia: a systematic review and meta-analysis. British Journal of Haematology. 2013;163(3):315-325. doi:10.1111/bjh.12528

21. Oscier D, Fegan C, Hillmen P, et al. Guidelines on the diagnosis and management of chronic lymphocytic leukaemia. British Journal of Haematology. 2004;125(3):294-317. doi:10.1111/j.1365-2141.2004.04898.x

22. Vlaanderen J, Vermeulen R, Heederik D, Kromhout H. Guidelines to Evaluate Human Observational Studies for Quantitative Risk Assessment. Environ Health Perspect. 2008;116(12):1700-1705. doi:10.1289/ehp.11530

23. Dreger P, Corradini P, Kimby E, et al. Indications for allogeneic stem cell transplantation in chronic lymphocytic leukemia: the EBMT transplant consensus. Leukemia. 2007;21(1):12-17. doi:10.1038/sj.leu.2404441

24. Hahn T, Wall D, Camitta B, et al. The Role of Cytotoxic Therapy with Hematopoietic Stem Cell Transplantation in the Therapy of Acute Lymphoblastic Leukemia in Adults: An Evidence-based Review. Biology of Blood and Marrow Transplantation. 2006;12(1):1-30. doi:10.1016/j.bbmt.2005.10.018

25. Hahn T, Wall D, Camitta B, et al. The Role of Cytotoxic Therapy with Hematopoietic Stem Cell Transplantation in the Therapy of Acute Lymphoblastic Leukemia in Children: An Evidence-Based Review. Biology of Blood and Marrow Transplantation. 2005;11(11):823-861. doi:10.1016/j.bbmt.2005.08.035

26. Oliansky DM, Appelbaum F, Cassileth PA, et al. The Role of Cytotoxic Therapy with Hematopoietic Stem Cell Transplantation in the Therapy of Acute Myelogenous Leukemia in Adults: An Evidence-Based Review. Biology of Blood and Marrow Transplantation. 2008;14(2):137-180. doi:10.1016/j.bbmt.2007.11.002

27. Oliansky DM, Rizzo JD, Aplan PD, et al. The Role of Cytotoxic Therapy with Hematopoietic Stem Cell Transplantation in the Therapy of Acute Myeloid Leukemia in Children: An Evidence-Based Review. Biology of Blood and Marrow Transplantation. 2007;13(1):1-25. doi:10.1016/j.bbmt.2006.10.024

28. Oliansky DM, Larson RA, Weisdorf D, et al. The Role of Cytotoxic Therapy with Hematopoietic Stem Cell Transplantation in the Treatment of Adult Acute Lymphoblastic Leukemia: Update of the 2006 Evidence-Based Review. Biology of Blood and Marrow Transplantation. 2012;18(1):18-36.e6. doi:10.1016/j.bbmt.2011.07.019

29. Skoetz N, Kluge S, Monsef I, et al. Twelfth Biannual Report of the Cochrane Haematological Malignancies Group—Focus on Tyrosine Kinase Inhibitors. J Natl Cancer Inst. 2011;103(4):E1-E1. doi:10.1093/jnci/djq531

B. MCA-L

1. Chen W-L, Wang J-H, Zhao A-H, et al. A distinct glucose metabolism signature of acute myeloid leukemia with prognostic value. Blood. 2014;124(10):1645-1654. doi:10.1182/blood-2014-02-554204

2. Zuo Z, Jones D, Yao H, et al. A pathway-based gene signature correlates with therapeutic response in adult patients with Philadelphia chromosome-positive acute lymphoblastic leukemia. Mod Pathol. 2010;23(11):1524-1534. doi:10.1038/modpathol.2010.137

3. Yang JJ, Cheng C, Devidas M, et al. Ancestry and Pharmacogenomics of Relapse in Acute Lymphoblastic Leukemia. Nat Genet. 2011;43(3):237-241. doi:10.1038/ng.763

4. Costa ES, Pedreira CE, Barrena S, et al. Automated pattern-guided principal component analysis vs expert-based immunophenotypic classification of B-cell chronic lymphoproliferative disorders: a step forward in the standardization of clinical immunophenotyping. Leukemia. 2010;24(11):1927-1933. doi:10.1038/leu.2010.160

5. Liu Z, Chen D, Bensmail H, Xu Y. Clustering gene expression data with kernel principal components. J Bioinform Comput Biol. 2005;03(02):303-316. doi:10.1142/S0219720005001168

6. Henriques A, Rodríguez-Caballero A, Nieto WG, et al. Combined Patterns of IGHV Repertoire and Cytogenetic/Molecular Alterations in Monoclonal B Lymphocytosis versus Chronic Lymphocytic Leukemia. PLoS One. 2013;8(7). doi:10.1371/journal.pone.0067751

7. Wei T, Liao B, Ackermann BL, et al. Data-driven analysis approach for biomarker discovery using molecular-profiling technologies. Biomarkers. 2005;10(2-3):153-172. doi:10.1080/13547500500107430

8. Meng Q, Catchpoole D, Skillicorn D, Kennedy PJ. DBNorm: normalizing high-density oligonucleotide microarray data based on distributions. BMC Bioinformatics. 2017;18. doi:10.1186/s12859-017-1912-5

9. Wang M, Kornblau SM, Coombes KR. Decomposing the Apoptosis Pathway Into Biologically Interpretable Principal Components. Cancer Inform. 2018;17. doi:10.1177/1176935118771082

10. Xu Y, Brereton RG. Diagnostic Pattern Recognition on Gene-Expression Profile Data by Using One-Class Classification. J Chem Inf Model. 2005;45(5):1392-1401. doi:10.1021/ci049726v

11. Dai JJ, Lieu L, Rocke D. Dimension Reduction for Classification with Gene Expression Microarray Data. Statistical Applications in Genetics and Molecular Biology. 2006;5(1). doi:10.2202/1544-6115.1147

12. Jaumot J, Tauler R, Gargallo R. Exploratory data analysis of DNA microarrays by multivariate curve resolution. Analytical Biochemistry. 2006;358(1):76-89. doi:10.1016/j.ab.2006.07.028

13. Nagai A, Zou N, Kubota M, et al. Fatigue in survivors of childhood acute lymphoblastic and myeloid leukemia in Japan. Pediatrics International. 2012;54(2):272-276. doi:10.1111/j.1442-200X.2011.03530.x

14. Wu C-J, Fu Y, Murali TM, Kasif S. Gene expression module discovery using gibbs sampling. Genome Inform. 2004;15(1):239-248.

15. Mehta RS, Holtan SG, Wang T, et al. GRFS and CRFS in alternative donor hematopoietic cell transplantation for pediatric patients with acute leukemia. Blood Adv. 2019;3(9):1441-1449. doi:10.1182/bloodadvances.2018030171

16. Lugthart S, Cheok MH, den Boer ML, et al. Identification of genes associated with chemotherapy crossresistance and treatment response in childhood acute lymphoblastic leukemia. Cancer Cell. 2005;7(4):375-386. doi:10.1016/j.ccr.2005.03.002

17. Sint Nicolaas SM, Schepers SA, van den Bergh EMM, et al. Illness cognitions and family adjustment: psychometric properties of the Illness Cognition Questionnaire for parents of a child with cancer. Support Care Cancer. 2016;24:529-537. doi:10.1007/s00520-015-2795-5

18. Soneson C, Lilljebjörn H, Fioretos T, Fontes M. Integrative analysis of gene expression and copy number alterations using canonical correlation analysis. BMC Bioinformatics. 2010; 11:191. doi:10.1186/1471-2105-11-191

19. Varma S, Simon R. Iterative class discovery and feature selection using Minimal Spanning Trees. BMC Bioinformatics. 2004; 5:126. doi:10.1186/1471-2105-5-126

20. Lin X, Lu T, Yan F, Li R, Huang X. Mean residual life regression with functional principal component analysis on longitudinal data for dynamic prediction. Biometrics. 2018;74(4):1482-1491. doi:10.1111/biom.12876

21. Tan Y, Shi L, Tong W, Wang C. Multi-class cancer classification by total principal component regression (TPCR) using microarray gene expression data. Nucleic Acids Res. 2005;33(1):56-65. doi:10.1093/nar/gki144

22. Komura D, Nakamura H, Tsutsumi S, Aburatani H, Ihara S. Multidimensional support vector machines for visualization of gene expression data. Bioinformatics. 2005;21(4):439-444. doi:10.1093/bioinformatics/bti188

23. Ollenschläger G, Thomas W, Konkol K, Diehl V, Roth E. Nutritional behaviour and quality of life during oncological polychemotherapy: Results of a prospective study on the efficacy of oral nutrition therapy in patients with acute leukaemia. European Journal of Clinical Investigation. 1992;22(8):546-553. doi:10.1111/j.1365-2362.1992.tb01504.x

24. Thiele J, Kvasnicka HM, Zankovich R, Fischer R, Diehl V. Parameters of Predictive Value in Chronic Myeloid Leukemia—The Prognostic Impact of Histopathological Variables in a Multivariate Regression Analysis. Leukemia & Lymphoma. 1991;4(1):63-74. doi:10.3109/10428199109107059

25. Hopcroft LEM, McBride MW, Harris KJ, et al. Predictive response-relevant clustering of expression data provides insights into disease processes. Nucleic Acids Res. 2010;38(20):6831-6840. doi:10.1093/nar/gkq550

26. Degos L, Feingold N, Bastin C, Rain JD. Principal clinical features and prognosis in chronic lymphocytic leukemia. Nouv Rev Fr Hematol. 1978;20(3):359-376.

27. Ishiyama K -i, Kitawaki T, Sugimoto N, et al. Principal component analysis uncovers cytomegalovirus-associated NK cell activation in Ph+ leukemia patients treated with dasatinib. Leukemia; London. 2017;31(1):203-212. doi:http://dx.doi.org.libproxy.lib.unc.edu/10.1038/leu.2016.174

28. Mejstrikova E, Volejnikova J, Fronkova E, et al. Prognosis of children with mixed phenotype acute leukemia treated on the basis of consistent immunophenotypic criteria. Haematologica. 2010;95(6):928-935. doi:10.3324/haematol.2009.014506

29. Lin X, Li R, Yan F, Lu T, Huang X. Quantile residual lifetime regression with functional principal component analysis of longitudinal data for dynamic prediction. Stat Methods Med Res. 2019;28(4):1216-1229. doi:10.1177/0962280217753466

30. Liu W, Yuan K, Ye D. Reducing microarray data via nonnegative matrix factorization for visualization and clustering analysis. Journal of Biomedical Informatics. 2008;41(4):602-606. doi:10.1016/j.jbi.2007.12.003

31. Rigolin GM, Fagioli F, Spanedda R, et al. Study of prognosis in acute myeloid leukemias Haematologica. 1994;79(3):233-240. (AML) by cluster analysis.

32. Bannur Z, Teh LK, Hennesy T, et al. The differential metabolite profiles of acute lymphoblastic leukaemic patients treated with 6-mercaptopurine using untargeted metabolomics approach. Clinical Biochemistry. 2014;47(6):427-431. doi:10.1016/j.clinbiochem.2014.02.013

33. Fontes M, Soneson C. The projection score - an evaluation criterion for variable subset selection in PCA visualization. BMC Bioinformatics. 2011;12:307. doi:10.1186/1471-2105-12-307

34. Wang M, Abrams ZB, Kornblau SM, Coombes KR. Thresher: determining the number of clusters while removing outliers. BMC Bioinformatics. 2018;19. doi:10.1186/s12859-017-1998-9

C. Gen-L

1. Bonardi F, Fusetti F, Deelen P, Gosliga D van, Vellenga E, Schuringa JJ. A Proteomics and Transcriptomics Approach to Identify Leukemic Stem Cell (LSC) Markers. Molecular & Cellular Proteomics. 2013;12(3):626-637. doi:10.1074/mcp.M112.021931

2. Chanock SJ. A twist on admixture mapping. Nat Genet. 2011;43(3):178-179. doi:10.1038/ng0311-178

3. Metzeler KH, Hummel M, Bloomfield CD, et al. An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. Blood. 2008;112(10):4193-4201. doi:10.1182/blood-2008-02-134411

4. Lhermitte L, Mejstrikova E, Sluijs-Gelling AJ van der, et al. Automated database-guided expert-supervised orientation for immunophenotypic diagnosis and classification of acute leukemia. Leukemia. 2018;32(4):874-881. doi:10.1038/leu.2017.313

5. Carlson SM, Najmi A, Cohen HJ. Biomarker clustering to address correlations in proteomic data. PROTEOMICS. 2007;7(7):1037-1046. doi:10.1002/pmic.200600514

6. Bai Y, Zhang H, Sun X, Sun C, Ren L. Biomarker identification and pathway analysis by serum metabolomics of childhood acute lymphoblastic leukemia. Clinica Chimica Acta. 2014;436:207-216. doi:10.1016/j.cca.2014.05.022

7. Pardo LM, Voigt AP, Alonzo TA, et al. Deciphering the Significance of CD56 Expression in Pediatric Acute Myeloid Leukemia: A Report from the Children's Oncology Group. Cytometry. July 2019:cyto.b.21829. doi:10.1002/cyto.b.21829

8. Fuka G, Farias-Vieira TM, Hummel L, et al. Evaluation of multiplex ligation dependent probe amplification (MLPA) for identification of acute lymphoblastic leukemia with an intrachromosomal amplification of chromosome 21 (iAMP21) in a Brazilian population. Mol Cytogenet. 2015;8. doi:10.1186/s13039-015-0147-2

9. Li W-Y, Wei H-Y, Sun Y-Z, Zhou H, Ma Y, Wang R-L. Exploring the effect of E76K mutation on SHP2 cause gain-of-function activity by a molecular dynamics study. Journal of Cellular Biochemistry. 2018;119(12):9941-9956. doi:10.1002/jcb.27316

10. Wang R-R, Ma Y, Du S, et al. Exploring the reason for increased activity of SHP2 caused by D61Y mutation through molecular dynamics. Computational Biology and Chemistry. 2019;78:133-143. doi:10.1016/j.compbiolchem.2018.10.013

11. Model F, Adorján P, Olek A, Piepenbrock C. Feature selection for DNA methylation based cancer classification. Bioinformatics. 2001;17(suppl_1):S157-S164. doi:10.1093/bioinformatics/17.suppl_1.S157

12. Maity A, Majumdar S, Ghosh Dastidar S. Flexibility enables to discriminate between ligands: Lessons from structural ensembles of Bcl-xl and Mcl-1. Computational Biology and Chemistry. 2018;77:17-27. doi:10.1016/j.compbiolchem.2018.08.007

13. Kornblau SM, Tibes R, Qiu YH, et al. Functional proteomic profiling of AML predicts response and survival. Blood. 2009;113(1):154-164. doi:10.1182/blood-2007-10-119438

14. Vasconcelos Y, Vos JD, Vallat L, et al. Gene expression profiling of chronic lymphocytic leukemia can discriminate cases with stable disease and mutated Ig genes from those with progressive disease and unmutated Ig genes. Leukemia. 2005;19(11):2002-2005. doi:10.1038/sj.leu.2403865

15. Yang M, Li X, Li Z, et al. Gene Features Selection for Three-Class Disease Classification via Multiple Orthogonal Partial Least Square Discriminant Analysis and S-Plot Using Microarray Data. PLoS One. 2013;8(12). doi:10.1371/journal.pone.0084253

16. Bentaboulet M, Mihaesco E, Gendron M-C, Brouet J-C, Tsapis A. Genomic alterations in a case of α heavy chain disease leading to the generation of composite exons from the jh region. European Journal of Immunology. 1989;19(11):2093-2098. doi:10.1002/eji.1830191119

17. Prasad P, Rönnerblad M, Arner E, et al. High-throughput transcription profiling identifies putative epigenetic regulators of hematopoiesis. Blood. 2014;123(17):e46-e57. doi:10.1182/blood-2013-02-483537

18. Lilljebjörn H, Henningsson R, Hyrenius-Wittsten A, et al. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. Nat Commun. 2016;7. doi:10.1038/ncomms11790

19. Andersson A, Olofsson T, Lindgren D, et al. Molecular signatures in childhood acute leukemia and their correlations to expression patterns in normal hematopoietic subpopulations. Proc Natl Acad Sci U S A. 2005;102(52):19069-19074. doi:10.1073/pnas.0506637102

20. McClure BJ, Heatley SL, Kok CH, et al. Pre-B acute lymphoblastic leukaemia recurrent fusion, EP300-ZNF384, is associated with a distinct gene expression. Br J Cancer. 2018;118(7):1000-1004. doi:10.1038/s41416-018-0022-0

21. Chadeau-Hyam M, Vermeulen RCH, Hebels DGAJ, et al. Prediagnostic transcriptomic markers of Chronic lymphocytic leukemia reveal perturbations 10 years before diagnosis. Ann Oncol. 2014;25(5):1065-1072. doi:10.1093/annonc/mdu056

22. Kaplan D, Sun Z, Tallman MS, et al. Prognostic information and biological insights in chronic lymphocytic leukemia by high-resolution immunophenotypic analysis of ZAP70. Cytometry Part A. 2014;85(9):798-808. doi:10.1002/cyto.a.22485

23. Delft FW van, Bellotti T, Luo Z, et al. Prospective gene expression analysis accurately subtypes acute leukaemia in children and establishes a commonality between hyperdiploidy and t(12;21) in acute lymphoblastic leukaemia. British Journal of Haematology. 2005;130(1):26-35. doi:10.1111/j.1365-2141.2005.05545.x

24. Cochran DAE, Evans CA, Blinco D, et al. Proteomic Analysis of Chronic Lymphocytic Leukemia Subtypes with Mutated or Unmutated Ig VH Genes. Molecular & Cellular Proteomics. 2003;2(12):1331-1341. doi:10.1074/mcp.M300055-MCP200

25. Mayer RL, Schwarzmeier JD, Gerner MC, et al. Proteomics and metabolomics identify molecular mechanisms of aging potentially predisposing for chronic lymphocytic leukemia. Mol Cell Proteomics. 2018;17(2):290-303. doi:10.1074/mcp.RA117.000425

26. Rombouts EJC, Pavic B, Löwenberg B, Ploemacher RE. Relation between CXCR-4 expression, Flt3 mutations, and unfavorable prognosis of adult acute myeloid leukemia. Blood. 2004;104(2):550-557. doi:10.1182/blood-2004-02-0566

27. Xue Y, Ge Y, Kang M, et al. Selection of three miRNA signatures with prognostic value in non-M3 acute myeloid leukemia. BMC Cancer. 2019;19. doi:10.1186/s12885-019-5315-z

28. Jörnsten R, Yu B. Simultaneous gene clustering and subset selection for sample classification via MDL. Bioinformatics. 2003;19(9):1100-1109. doi:10.1093/bioinformatics/btg039

29. Choi J, Pacheco CM, Mosbergen R, et al. Stemformatics: visualize and download curated stem cell data. Nucleic Acids Res. 2019;47(Database issue):D841-D846. doi:10.1093/nar/gky1064

30. Xu C-J, Hoefsloot HC, Smilde AK. To aggregate or not to aggregate high-dimensional classifiers. BMC Bioinformatics. 2011;12:153. doi:10.1186/1471-2105-12-153

31. Eagle GL, Zhuang J, Jenkins RE, et al. Total Proteome Analysis Identifies Migration Defects as a Major Pathogenetic Factor in Immunoglobulin Heavy Chain Variable Region (IGHV)-unmutated Chronic Lymphocytic Leukemia. Mol Cell Proteomics. 2015;14(4):933-945. doi:10.1074/mcp.M114.044479

32. Hsu C-H, Nguyen C, Yan C, et al. Transcriptome Profiling of Pediatric Core Binding Factor AML. PLoS One. 2015;10(9). doi:10.1371/journal.pone.0138782

33. Nguyen DV, Rocke DM. Tumor classification by partial least squares using microarray gene expression data. Bioinformatics. 2002;18(1):39-50. doi:10.1093/bioinformatics/18.1.39