dCollection

Ph.D. Dissertation

Visual Analytics System Designs for Using Scientific Evidence in Medical Research

의학 연구에서의 과학적 증거의 활용을 위한
시각적 분석 시스템 디자인

August 2022

Department of Computer Science and Engineering
College of Engineering
Seoul National University

Brian J. Lee

Visual Analytics System Designs for Using Scientific Evidence in Medical Research

Advisor: Jinwook Seo

Submitting a Ph.D. Dissertation of Computer Science and Engineering

July 2022

College of Engineering
Seoul National University

Brian J. Lee

Confirming the Ph.D. Dissertation written by Brian J. Lee

June 2022

Chair	Sun Kim	(Seal)
Vice Chair	Jinwook Seo	(Seal)
Examiner	Bohyoung Kim	(Seal)
Examiner	Jongchan Lee	(Seal)
Examiner	Hyunjoo Song	(Seal)

Abstract

Brian J. Lee

Department of Computer Science and Engineering

College of Engineering | **Seoul National University**

Evidence-based medicine, *"the conscientious, explicit, and judicious use of current best evidence in healthcare and medical research"* [98], is one of the most widely accepted medical paradigms of modern times. Searching, reviewing, and synthesizing reliable and high-quality scientific evidence is the key step for the paradigm. However, despite the widespread use of the EBM paradigm, challenges remain in applying Evidence-based medicine protocols to medical research. One of the barriers to applying the best scientific evidence to medical research is the severe literature and clinical data overload that causes the evidence-based tasks to be tremendous time-consuming tasks that require vast human effort. In this dissertation, we aim to employ visual analytics approaches to address the challenges of searching and reviewing massive scientific evidence in medical research. To overcome the burden and facilitate handling scientific evidence in medical research, we conducted three design studies and implemented novel visual analytics systems for laborious evidence-based tasks.

First, we designed PLOEM, a novel visual analytics system to aid evidence synthesis, an essential step in Evidence-Based medicine, and generate an Evidence Map in a standardized method. We conducted a case study with an oncologist with years of evidence-based medicine experience. In the second study, we conducted a preliminary survey with 76 medical doctors to derive

the design requirements for a biomedical literature search. Based on the results, We designed EEEVis, an interactive visual analytic system for biomedical literature search tasks. The system enhances the PubMed search result with several bibliographic visualizations and PubTator annotations. We performed a user study to evaluate the designs with 24 medical doctors and presented the design guidelines and challenges for a biomedical literature search system design. The third study presents GeneVis, a visual analytics system to identify and analyze gene expression signatures across major cancer types. A task that cancer researchers utilize to discover biomarkers in precision medicine. We conducted four case studies with domain experts in oncology and genomics. The study results show that the system can facilitate the task and provide new insights from the data. Based on the three studies of this dissertation, we conclude that carefully designed visual analytics approaches can provide an enhanced understanding and support medical researchers for laborious evidence-based tasks in medical research.

Contents

CHAPTER 4

EEEvis: Efficacy improvement in searching MEDLINE database using a novel PubMed visual analytic system   26

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

The Evidence-based medicine (EBM) paradigm is *"the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients"* [98]. It has arisen from the call for greater objectivity in clinical decision-making and has had a positive impact and significant influence on healthcare and medicine for the last few decades [37, 101]. One of the most significant assets of the Evidence-based medicine paradigm is that it provides a robust systematic approach to translating the findings of high-quality medical research into other medical research and clinical practice [93]. The practical process of Evidence-based medicine involves the following five essential steps [93]:

1. *Searching for an answer to a clinical question.*

2. *Searching for the evidence.*

3. *Critically appraising and reviewing the existing evidence.*

4. *Combining credible evidence with clinical experience and the patients' viewpoints.*

5. *Evaluating the process.*

Each step of the Evidence-based medicine paradigm follows strict protocols and principles [1]. In addition, various literature (scientific evidence) applications and databases (*PubMed, Cochrane Library, JAMA Evidence*, Etc.) help medical researchers with these tasks. However, some challenges remain despite the widespread use of the EBM paradigm and the various efforts and applications to support EBM practice over the last thirty years [70, 86, 101, 112, 125]. In this dissertation, we aim to employ visual analytics approaches to address the challenges of searching and reviewing massive scientific evidence.

We interviewed several domain experts, such as medical oncologists and biomedical researchers, to identify the difficulties and challenges of Evidence-based medical paradigms in real life. After a series of interviews, we confirmed that there are still many obstacles that require a tremendous human effort to interpret the data insights for the medical models, and this hinders the practice of EBM.

First, searching for the proper reliable evidence is still a challenging task. The evidence in systematic reviews is derived from systematic literature searches as new research builds on previous discoveries. Without a systematically rigorous search strategy, the results or findings of the systematic review and meta-analyses should be suspect. Therefore, with the tremendous amount of new evidence in these days, rapidly finding relevant and reliable evidence has become even more critical [2]. However, formulating optimal search queries

for a given topic is a challenging task for a novice user. While searching with a query-based system such as PubMed, novice users often miss some publications or retrieve irrelevant results [24, 90]. Therefore, some guidelines even advise having a professional medical librarian or information specialist conduct the search task [47]. This task brings about a massive amount of person-hours that can be critical for the practice of Evidenced-based principles in medical research.

Second, users require broader support for various evidence synthesis approaches. Conventional evidence synthesis systems usually focus on systematic reviews and meta-analyses. Systematic reviews and meta-analyses are well-developed methods for precisely focused questions. For example, to review the accuracy of a novel diagnostic test method or the effectiveness of a new treatment. On the other hand, when it comes to much broader questions such as *"the relative effectiveness of various treatment options available for a particular condition"* [17], the systematic review methods face challenges in evidence synthesis [17]. Many other evidence synthesis models have been developed to meet various user requirements, and there are demands for systems that support these models.

However, critics have suggested that Evidence-based medicine has several limitations [15]. Evidence-based medicine collects evidence from groups of patients and summarizes the evidence. The response is for an average patient rather than an individual patient [28], so it often fails to provide an appropriate solution for outlier patients [27, 110]. Therefore, medical researchers also utilize scientific evidence from Precision medicine as complementary.

Precision medicine is another widely used medical paradigm that adapts the diagnosis and treatment to individual patients. The diagnosis and treatment are based on genetics, biomarkers, environment, and clinical information [27]. Furthermore, with the advance in patient characterizing technology with omics science (proteomics, metabolomics, genomics, etc.) [27], the amount and impact of the omics data are also rapidly growing. Therefore, the request for methods to search and appraise scientific evidence of biomedical data is also increasing. So, to continue the objective of using the current best evidence for medical research, we believe medical research frameworks should integrate genomic evidence with conventional literature and clinical evidence as complementary approaches in medical research.

We evaluate this approach by aiding tasks with genomic sequencing data. With the help of the projects such as The Cancer Genome Atlas (TCGA), Encyclopedia of DNA Element (ENCODE), and Genotype-Tissue Expression (GTEx), the research community now has an extensive dataset of transcriptomic profiles from RNA-seq experiments [109]. The community is now facing the challenge of understanding this big data. Reflecting this need with the tremendous big data, the appearance of sophisticated RNA-seq analysis tools in a visual analytics approach is increasing. However, only limited types of tasks have standardized tools that organize, visualize, and analyze the resulting big data, and there are many tasks that researchers still conduct in non-standard methods.

In this dissertation, we conducted three design studies to design and implement novel visual analytics systems for laborious scientific evidence

based tasks in medical research. The following questions are the research questions that motivated this dissertation:

RQ1. What analytic tasks in Evidence-based medical research can benefit from visual analytics systems?

RQ2. How should we design a visual analytics system to facilitate the scientific evidence search task?

RQ3. How should we design a visual analytics system to facilitate the scientific evidence review task?

RQ4. How should we design a visual analytics system to facilitate integrating genomic evidence into evidence-based medical research?

Thesis Statement Carefully designed visual analytics systems can provide a better understanding of biomedical scientific evidence and facilitate evidence-based tasks for medical research.

1.2   Dissertation Outline

This dissertation is divided into six chapters. Chapter 2 discusses previous relevant studies of this dissertation, including studies on evidence mapping, literature visualizations, and genomics visualizations. Chapter 3 introduces PLOEM (PLOting Evidence Mapping), a visual analytics system to present an evidence map in a standardized method. Chapter 4 proposes EEEVIs, a novel interactive visual analytics system for biomedical literature search tasks. Chapter 5 presents GeneVis, a visual analytics system that visualizes

a user-configured batch of gene signatures across major cancer types. Lastly, Chapter 6 concludes this dissertation and proposes our future work.

Chapter 2

Related Work

2.1 Evidence Mapping: Graphical Representation for a Scientific Evidence Landscape

During scientific research, it is a standard procedure for researchers to explore scientific publications to obtain up-to-date knowledge of a research topic [53]. Furthermore, with the exponentially growing number of scientific literature, reviews that summarize facts and insights are becoming more valuable [81]. Systematic reviews and meta-analyses are systematic methods that can extract and synthesize quantitative information from scientific publications, and it is an essential procedure for Evidence-based medicine [5, 41, 46, 57, 80, 102, 114]. However, these methods are appropriate for focused research questions and not practical for broader topics or synthesizing heterogeneous literature [81]. Therefore, the Evidence mapping method emerged to overcome these limitations.

Evidence mapping is a tool to *"systematically and comprehensively identify, organize and summarize the distribution of scientific evidence on a broad field"* [4, 51]. The difference against systematic review is that the risk-of-bias appraisal

of the included studies is not required [4, 31], so it applies to much broader research questions than systematic reviews. Evidence mapping summarizes the insights and characteristics of scientific publications in tree forms or tabular forms [104]. The summary clarifies the research gaps (or evidence gaps) [22], representing differences among the research areas where missing evidence and the research areas with sufficient evidence. Researchers can identify research fields with enough evidence to conduct a systematic review [4, 17, 31, 44, 104], or determine the research fields that need focused studies in the future [4, 17, 22, 31, 51, 104].

However, there are disparities among the terminologies and methods of the evidence mapping techniques (e.g., evidence map, systematic map, scoping map, and scoping review), and still, no methodological standards [32, 44, 51, 100, 104]. Nonetheless, by Wang et al., there are common steps that all evidence mapping variances generally follow [117]:

1. *Identifying a broad research area of interest and defining the key variables and framework for descriptive analysis*

2. *Developing a thorough, clear and reproducible literature search strategy*

3. *Establishing the a priori inclusion and exclusion criteria*

4. *Systematically extracting, coding, sorting and reporting the findings in a tabular evidence map*

In Chapter 3, based on the common key steps of evidence mapping, we designed PLOEM (PLOting Evidence Mapping), a visual analytic system to present an evidence map in a standardized method. Providing a standardized method for evidence mapping and a standard representation for

evidence maps should aid the usage and improve the interpretability of evidence mapping for clinicians and biomedical domain experts.

2.2 Scientific Literature Visualizations and Bibliography Visualizations

Searching for literature evidence is essential for evidence-based medicine and clinical decision-making. Millions of clinicians and biomedical researchers search literature repositories, such as PubMed (`https://pubmed.gov`), and explore the vast search results to find reliable literature evidence. Many users use literature management tools to manage literature collections, such as Mendeley, EndNote, Zotero, Cochrane RevMan, and Etc. However, these tools are focused on managing citation and user annotation rather than literature discovery.

Various studies support bottom-up literature discovery, a strategy to search target papers from a set of source papers [7, 58, 65, 122]. SenseMaker [7] supports iterative searching and expansion in topics of interest with a rich graphical user interface. CiteSense [122] provides features for users to explore the publications that cite a text snippet of a paper or the publications cited by a paper. Wivi [65] reflects the reading history and visualizes the connections among visited papers as a graph form. LitSense [107] focuses on research questions an build topic graphs. Organizing and visualizing bibliographic aspects is one of the most conventional methodologies in literature visualizations [13, 35, 50, 82, 96]. Many studies present the citation network as a node-link diagram, or an adjacency matrix [23, 26, 34, 48, 50, 56, 68, 75,

82, 89, 105]. Several research employ the author-paper network or co-author relationship [60, 62, 123]

In Chapter 4, we present EEEVIs, a novel interactive visual analytic system for biomedical literature search tasks. EEEvis provides enhanced literature data analysis functions, including (1) an overview of the bibliographic features, (2) an overview of the co-authorship network, (3) interactive sorting, filtering, and highlighting, and (4) a named entity recognition (NER) annotated literature visualizations that employ bio-entities as NER tags from PubTator.

2.3 Visual Anlytics Systems for Genomics Data sets and Research Tasks

With the initiatives of large genomic data sets gathered from clinical and population studies, such as TCGA, GTEx, and ENCODE, the bioinformatics community has put its effort into developing various tools to ease the task of data assembling and analyzing [39, 42, 61, 77, 83, 94, 118, 124]. Data portals, such as CBioPortal [21], GDC Data Portal [45], and ICGC Data Portal [121] are widely used effective tools for browsing centrally maintained datasets and featuring a backend API server. With advanced sequencing technology, such as whole-genome sequencing, new genomic datasets are increasing rapidly. Therefore, analyzing and exploring complex datasets is getting more challenging, and the role of visual analytics is becoming more important.

Representing dysregulated genes in a cohort with a matrix form is bioinformatics's most conventional visualization technique, and it supports iden-

tifying the primary driver genes and discovering recurrent patterns [6, 52, 54, 76, 108]. However, this representation frequently leaves out the global structure of the cohort [76, 83, 85, 92].

In Chapter 5, we designed and implemented a web-based novel visual analytics system, GeneVis. A system that users can visualize the gene expression signature of a user-defined gene set and analyze the patterns and insights across multiple major cancer types.

Chapter 3

PLOEM: An Interactive Visualization Tool for Effective Evidence Mapping with Biomedical literature

This chapter introduces a visual analytic system that supports biomedical researchers, clinical doctors, and medical scientists to produce an evidence map in a standardized method[1].

3.1 Introduction

Conducting a clinical trial that can control all variables for diseases or treatments at once on a statistically significant scale is a challenging task in the biomedical and healthcare context. Therefore clinical researchers collect and analyze secondary data on previous studies through systematic reviews (SR), an evidence synthesis method to perform and inform evidence-based medicine.

---

[1]The preliminary version of Chapter 3 was published as a journal article in KIISE Transactions on Computing Practices.

If the amount and quality of the secondary data are sufficient, researchers can conduct a meta-analysis (MA) under standardized procedures to integrate the results of independent studies to derive conclusions of the research. Meta-analysis is a widely used tool that plays a central role in evidence-based medicine. However, it is not appropriate to conduct a meta-analysis when the data sources have too much uncertainty, nor the heterogeneity among studies is too significant. Accordingly, in recent years, a methodology based on evidence mapping (EM) that applies to more diverse treatment methods and fewer samples has been introduced and spread[63].

The purpose of evidence mapping is to provide an overview of existing studies rather than quantitative result synthesis and identify gaps in each category by classifying existing studies according to one or more criteria. To this end, existing evidence mapping studies have used various visualizations to collect the number of related studies, sample size, evidence level, and effect size for each category. For example, presenting the classification of studies in a table-based[36] or tree-based[4] figure was the conventional approach. However, visualization techniques used in most existing studies could represent only the number of papers corresponding to a specific category. There was a limitation in that it was difficult to recognize changes over time or the relationship between papers. In addition, although the need and interest in the evidence mapping technique are increasing, a standardized procedure has not yet been established, so each study presents the evidence map in various ways.

In this chapter, we present PLOEM, a visual analytic system that supports biomedical researchers, clinical doctors, and medical scientists to produce an evidence map in a standardized method and the usage scenarios.

3.2 Visual Representations and Interactions of PLOEM

This study proposes PLOEM, a novel visual analytics system for evidence mapping. To derive the features that clinicians and biomedical researchers require from an evidence map, we conducted a literature survey over evidence mapping studies and a series of interviews with biomedical, clinical, and healthcare researchers. As a result, we derived the following five essential characteristics of evidence mapping that a visual analytics system should present.

- Study design

- Sample population

- Intervention of interest and comparator intervention

- Outcome of interest

- Trend over time

The system comprises four coordinated views: the overview, the timeline view, the relation view, and the study detail view.

3.2.1 Overview of the PICO Criteria

In the case of evidence mapping, it is essential to follow up the sample population of each study design that has been conducted on the intervention of
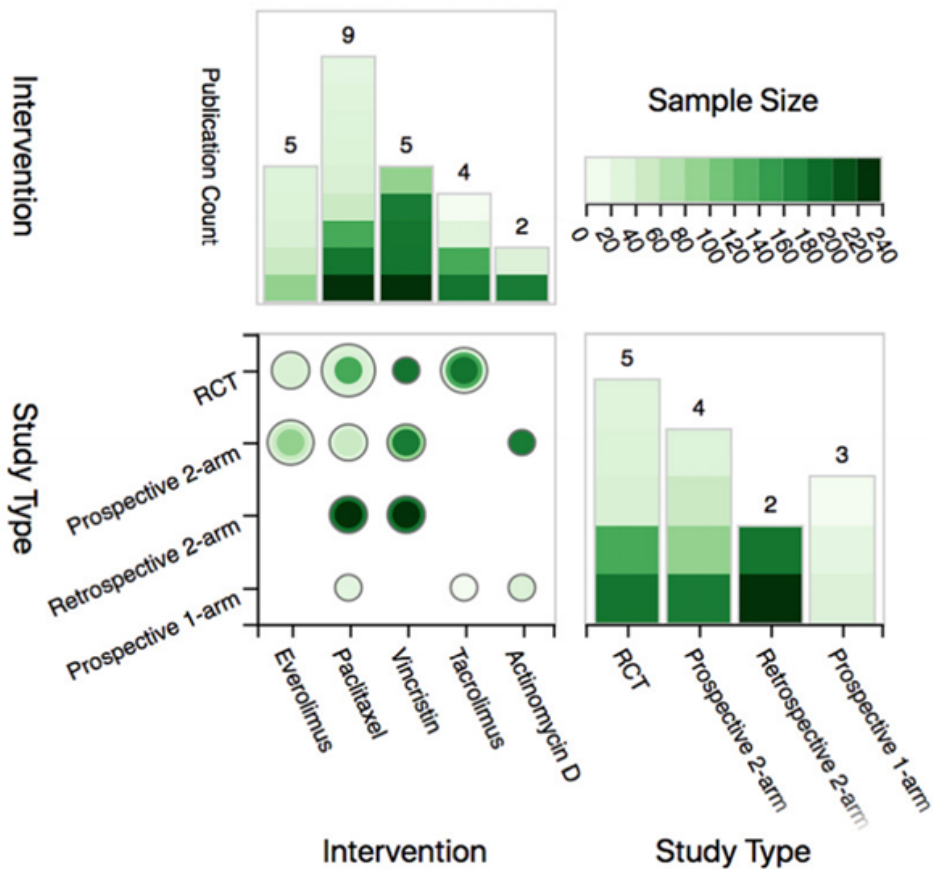
Figure 3.1: Overview. Studies are categorized with study design and intervention types. Each chart is color coded with sample size.

interest. Therefore we present the number of studies in the overview as a 2 x 2 matrix of interventions and study designs. As can be seen from Figure 3.1, the screen is divided mainly into four zones, each clockwise from the top right: 1) a color scheme legend for the sample population; 2) a stacked bar chart showing the number of each study design and the sample population of each study; 3) a heatmap table where the rows represent the study designs and the columns represent the interventions; 4) a stacked bar chart showing

the number of studies for each intervention and the sample population of each study.

Every visual mark(rectangles and points) of the stacked bar chart and the heatmap is color-coded by the study's sample population, and the discrete color legend is located at the top right corner. Both bar charts encode studies as a rectangle and stack them to place the study with the most sample population at the bottom. For example, in Figure 3.1 we can identify there are only two cases of Retrospective 2-arm studies, but both studies have a relatively large sample population. The heatmap table at the bottom left follows a similar visual metaphor. The table is a 2-dimensional heatmap with a study design axis(row axis) and an intervention axis(column axis). We place the studies at each coordinate as concentric circles, while the innermost circle has the largest sample population. Therefore, users can briefly overview the number of studies and the sample population. Initially, we designed the study design bar chart as a horizontal bar. However, during our studies with clinical and biomedical researchers, many of them reported difficulty at interpreting a horizontal bar chart, so, as shown in Figure 3.1, we altered the design to a vertical bar chart.

Users can confirm the details of the studies with a specialized tooltip (Figure 3.2). The tooltip contains a list of studies and presents the sample population and the evidence level of each study with a sparkline[111]. This tooltip is also used in the system's other views(timeline and relation views).
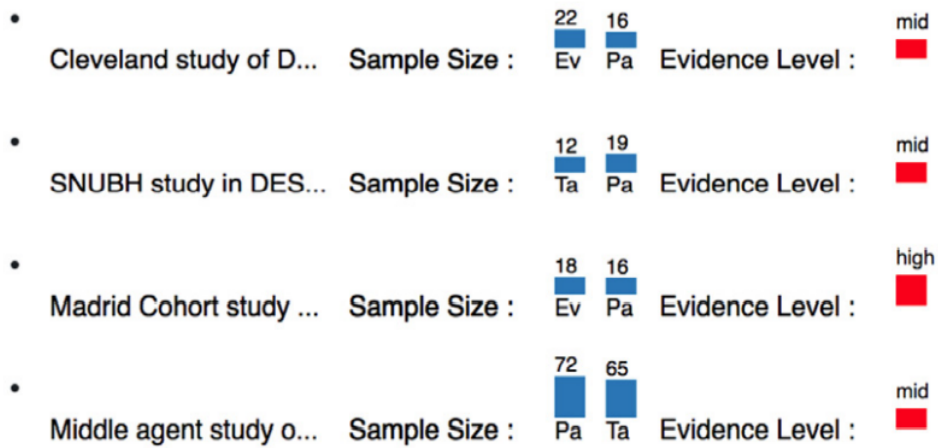
Figure 3.2: Detailed information tooltip for four papers. Titles, sample sizes, and evidence levels inferred from study design are visualized.

3.2.2 Trend Visualization with the Timeline view

The previous evidence mapping visualization techniques could not present the trend over time, a significant limitation. Therefore, one of the goals of this study was to present the temporal trends of the evidence-based medicine characteristics: sample population, medical intervention, study designs, and the number of studies.

We present a glyph-based timeline view(Figure 3.3) that places glyphs designed to represent the above features on a horizontal time axis to fulfill this goal. The view is a focus-plus-context visualization, while the topmost timeline is an overall timeline(Context), and the following timelines are the intervention timelines(Focus). The overall timeline shows the trend of all data through the entire time range. On the other hand, the intervention timelines show the study trends of each intervention for a selected time range. Users can select a range of interest in time by brushing a region of the
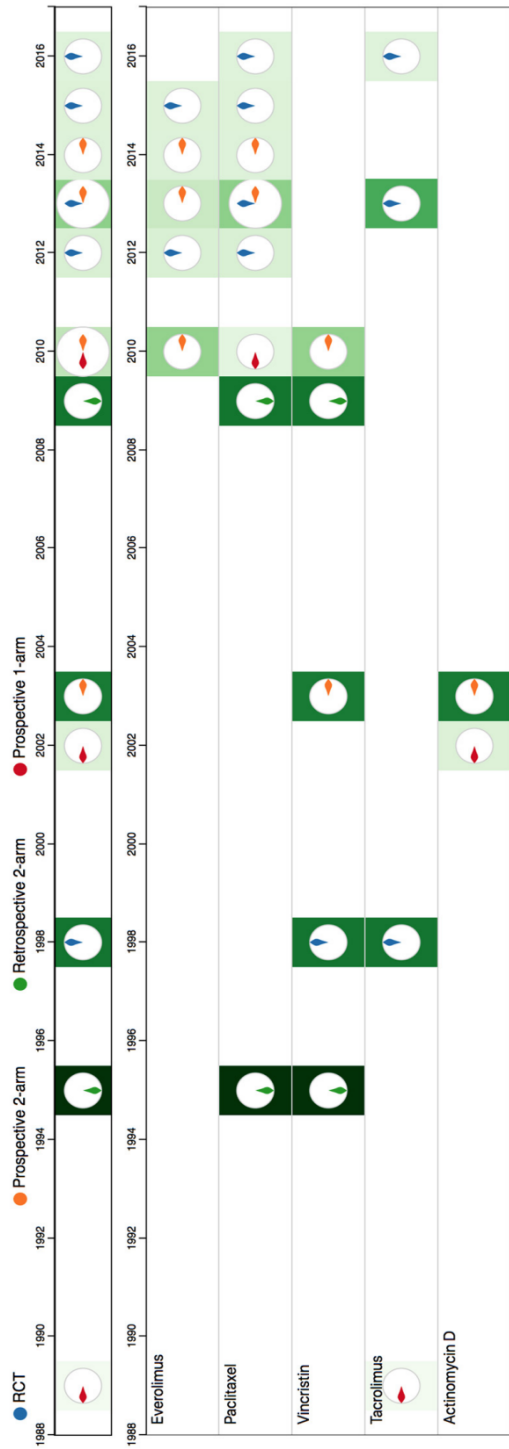
Figure 3.3: Timeline view with glyphs. Fourteen papers were classified by four study types (RCT, Prospective 2-arm, Retrospective 2-arm, Prospective 1-arm) and five interventions (Everolimus, Paclitaxel, Vincristin, Tacrolimus, Actinomycin D).

overall timeline, and the intervention timelines will reflect the selection by zooming into the selected temporal region. Other coordinated views, such as the relation view and the study detail view, also reflect this time range selection and visualize studies only within the range of interest. In addition, when users brush a region of the intervention timelines, the coordinated views visualize studies that include the intervention and within the time range.
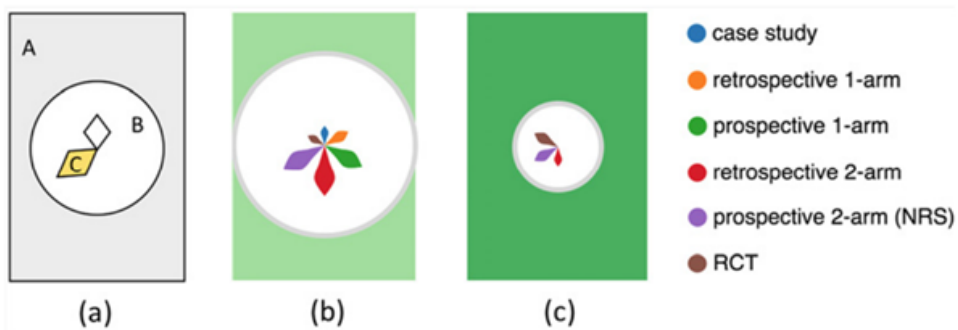
Figure 3.4: Glyph for timeline visualization. (a) Background color, radius of a circle, length of radial shape is mapped to sample size, number of whole papers, and number of papers categorized by study type, respectively, (b) Larger number of studies with all six study types, and smaller sample size, (c) Smaller number of studies with three study types, and larger sample size.

We designed the glyph to aggregate and encode the following information of a period: the sample population, the number of studies, and the number of each study design. The glyph has three main parts, and each part encodes the above three study characters. The outer rectangular figure (Figure 3.4 (a) A) represents the time period and the summation of the sample populations. The box's width encodes the length of the period the glyph aggregates, and the color saturation encodes the size of the sample population. The radius of the white center circle (Figure 3.4 (a) B) encodes the number of studies in the period, so a larger circle indicates more studies in the period.

Finally, each petal of the flower-petal glyph (Figure 3.4 (a) C) represents the number of each study design. A unique color and direction are assigned to each study design petal so users can notice the rise and fall of each study design through various timestamps.

3.2.3 Representing the PICO Co-occurrence with the Relation view

Most clinical trials are single-arm or two-arm studies, so each study confirms only a subset of the target interventions when a researcher conducts a systematic review of multiple interventions. Therefore, there is inevitably an imbalance between the number of studies for each intervention, and interpreting the relationships between the imbalanced study sets of each intervention is an important analysis task. However, with the conventional evidence-mapping visualization techniques, such as tables and trees, this is a challenging task.

We encode this set relationship among interventions in a node-link diagram and two coordinated list views (Figure 3.5). Nodes represent each intervention, and the node size indicates the sample population. Edges connect the interventions studied together, and the edge width represents the number of studies. For example, in Figure 3.5.a, we can identify there are five interventions, and Vincristine and Paclitaxel are the two treatments that have the most extensive sample population. However, the thick edge between Paclitaxel and Everolimus indicates that a study between Paclitaxel and Everolimus was the most common combination in the two-arm studies. These combinations among interventions are listed in a juxtaposed view with a bar chart of the number of studies (Figure 3.5.c). The edges and list
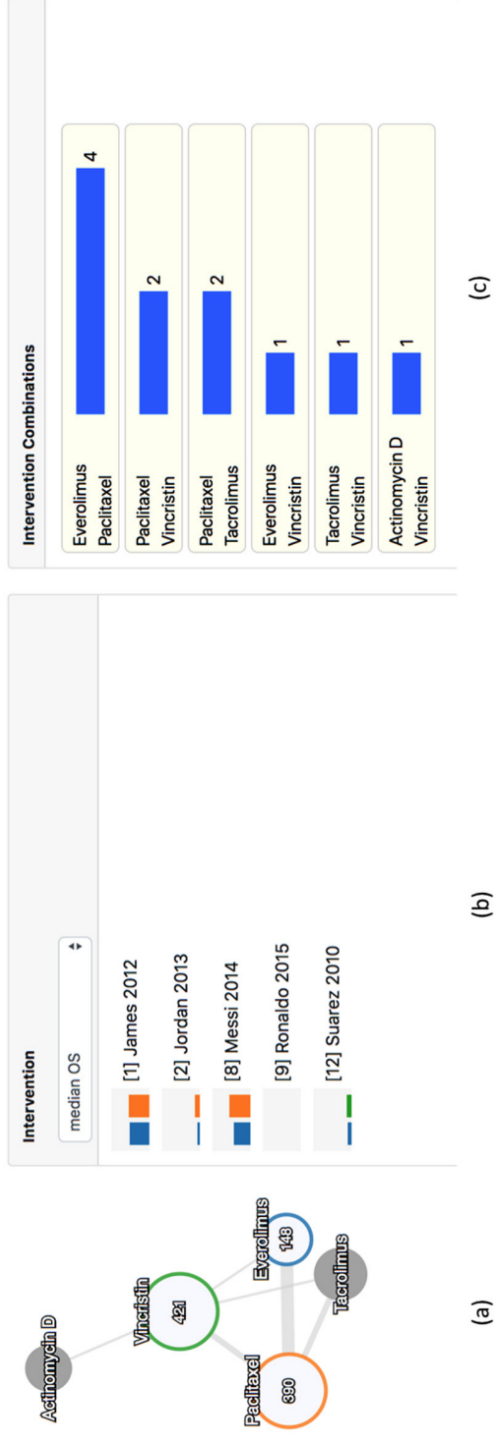
Figure 3.5: Relation view represents relationship between interventions. (a) Each intervention is represented with a circular node, with a diameter proportional to sample size. When user moves one's cursor over a node, related interventions are highlighted with unique colors, (b) Outcome from each intervention pair gets visualized with bar charts, (c) Explicitly visualizes the number of papers for each pair of interventions.

items are linked, highlighting the corresponding list item when a user hovers an edge and vice versa.

Another juxtaposed list is a study-outcome list (Figure 3.5.b) that lists the studies of a mouse selected intervention(node) or combination(edge). The corresponding studies are listed with the outcome of each study in a sparkline bar chart[111], and the outcome metric can be changed by selecting an option from the dropdown list. For example, The height of the orange bars in Figure 3.5.b indicates that Paclitaxel's overall survival rate(median OS) was usually better than Everolimus.

Study Details

#	Date	▼ Title	Au...	S...	Intervention T1	T2	Samp T1
1	2016-...	Yoga-based intervention in patients with	Contains ▼		A	TCA	5
2	2007-...	Yoga therapy as an add-on treatment in t	Filter...		RI		1
3	2012-...	Why does genetic causal information alte...	AJvvi	clinic...	Reassu...	Fam Tx	5
4	2000-...	Which clinical practice guidelines for depr...	PLCor...	case ...	Stablon		8
5	2015-...	Voluntary exercise enhances activity rhyt...	KTal-K...	case-...	SSRI	Reassu...	2
6	2009-...	VEGF serum levels in depressed patients...	MVentri	clinic...	SSRI	Fam Tx	12
7	2016-...	Variation and ethnic inequalities in treatm...	SLPr	case ...	TCA		3
8	2002-...	Vagus nerve stimulation: where are we?	SCSc...	cross...	Reassu...		14
9	2011-...	Vagus nerve stimulation: from pre-clinical ...	GMDe	case-...	TCA	SSRI	9
10	2015-...	Vagus nerve stimulation has antidepressa...	AGrim...	cross...	SSRI		5
11	2011-...	Using pre-treatment electroencephalogra...	AKho...	clinic...	TCA	Fam Tx	14
12	2016-	Using mixed methods case-series evalua	AMHarri	cross	Fam Tx		9

Figure 3.6: Detail view with entire information about all papers. Filtering and sorting can be activated on each column header.

3.2.4 Study detail view

The study detail view(Figure 3.6) is a table that users confirm the raw data of the studies when the additional information from the tooltips is insuf-

ficient. The table presents the publication data, study title, publication authors, study design, evidence level, intervention, comparator intervention, the sample population, and the outcome of each arm. The view is coordinated with other views and reflects the selection in the overview, timeline view, and relation view. In addition, users can sort and filter the rows by using the sort/filter controls in the column headers.

3.3   Usage Scenarios: Visualizing Various Study Sizes with PLOEM

This section reports the results from three usage scenarios to demonstrate how PLOEM can aid users in conducting evidence mapping and interactive visual analysis. We used three datasets about pancreatic cancer and the outcome of multiple treatments with a variant number of studies(14, 504, and 3,271), and each dataset was generated under the supervision of a domain expert in clinical research and the systematic review procedure.

The first scenario is an evidence map with 14 studies, and with the overview (Figure 3.1), researchers identified the gap of the number of studies and the imbalance of the sample population among interventions. For example, Everolimus is the runner-up treatment in the number of studies, but all five studies have a small sample size, which can conclude that these studies should have a low evidence level. On the other hand, there were only two studies of Actinomycin D, but one of the two studies showed a significant sample population.

The timeline view reveals the study trend change between the early and recent studies. For example, in Figure 3.3, we can find that the number of

studies increased over time, especially after the year 2009 and that many studies were conducted about Everolimus after that year. However, the glyph's background color difference indicates that the studies after 2009 show a smaller population than the studies before. This population change might be related to the change of study designs because we can also identify the rise of the blue petals in the glyphs after 2012, which indicates the increase of randomized control trials(RCT), and RCTs usually have a smaller population than other study designs.

The researchers who used PLOEM in all three scenarios reported that it was much easier to generate a map with PLOEM than other conventional methods, and it was effective in displaying the gaps among study designs, interventions, and sample populations. They also commented that the timeline view presented many insights that they did not catch with a conventional table view or a tree view, such as the trend changes and the breaks(the periods with no significant studies).

3.4   Conclusion

This study presents PLOEM, a novel visual analytic system for evidence mapping. Evidence mapping is a novel systematic review method spotlighted in the clinical and healthcare community, yet does not have a standardized procedure or generation tool. To this end, we interviewed domain experts in clinical research, healthcare, and systematic review. Based on the preliminary survey and interviews, we derived the main characteristics of evidence mapping and the features for a visual analytic system for it.

However, the system is still a novel approach, and several limitations should be improved. One of the limitations was the scalability of the flower glyph design in the timeline view. It was not difficult to identify the glyphs until a few hundred studies, but the visibility of the glyph decreased when the number of studies was more than thousands.

The learning curve of the system was another limitation. The participants of the usage scenarios are not experts in visualization or visual analytic systems, and the visual representations and the interactions can be an obstacle for users of a different domain[64]. The participants were not familiar with the concepts of glyphs, coordinated views, or brushing and linking interactions and needed some guidance to interpret the visualizations or utilize the interactions. We believe we can overcome this limitation by with a detailed interactive tutorial.

Finally, we plan to expand our research into more requirements of the systematic review procedure, especially the method to search and retrieve the studies for evidence mapping or other systematic reviews.

Chapter 4

EEEvis: Efficacy improvement in searching MEDLINE database using a novel PubMed visual analytic system

4.1 Introduction

4.1.1 Motivation

Literature searching is a crucial step in conducting scientific research, preparing presentation material, and selecting study topics. PubMed (`https://pubmed.ncbi.nlm.nih.gov/`), which has been developed and is administered by the National Center for Biotechnology Information (NCBI) at the United States National Library of Medicine (NLM) [38], has become the most widely used search engine and free database in the biomedical field. PubMed covers most of the published biomedical literature and more than three million of its cases receive clicks by 2.5 million users per day [38].

The advantages of PubMed can generally be described as follows:

- *An enormous amount of biomedical records including Index Medicus and Non-Index Medicus* [30]

- *Expandability to related databases such as BioProject (formerly known as the Genome Project), other genetics, proteomics, and the Medical Subject Headings (MeSH) database* [11, 88, 95]

- *An easy and intuitive user interface (UI)* [88]

- *Free accessibility to the advanced search mode* [88]

- *Individual optimization using a My NCBI account* [88]

- *Multiple filters displayed on the left side of the browser* [103]

- *A continuously evolving search algorithm* [38]

- *Various additional functions that provide user friendliness* [55]

The PubMed database provides a simple search interface. However, the search interface supports only a limited set of data analysis methods, and it is difficult to interpret and explore through a massive literature dataset. Users can narrow the overwhelming search space to the most relevant matches by successive text-based queries with a multi-bibliographic criteria syntax. Nevertheless, this task requires a highly skilled query construction ability and the domain knowledge of the search field [88]. Therefore, conducting a massive literature search task only with a text-based query language can be time-consuming and inefficient.

Within this context, we conducted a three-phased study, as follows:

1. A preliminary user survey regarding the current usability and the literature search task requirements of PubMed

2. An implementation of a novel interactive visual analytic system for biomedical literature search in the PubMed database

3. A randomized controlled prospective user survey comparing the usability of PubMed and the proposed implementation (Figure 4.1).
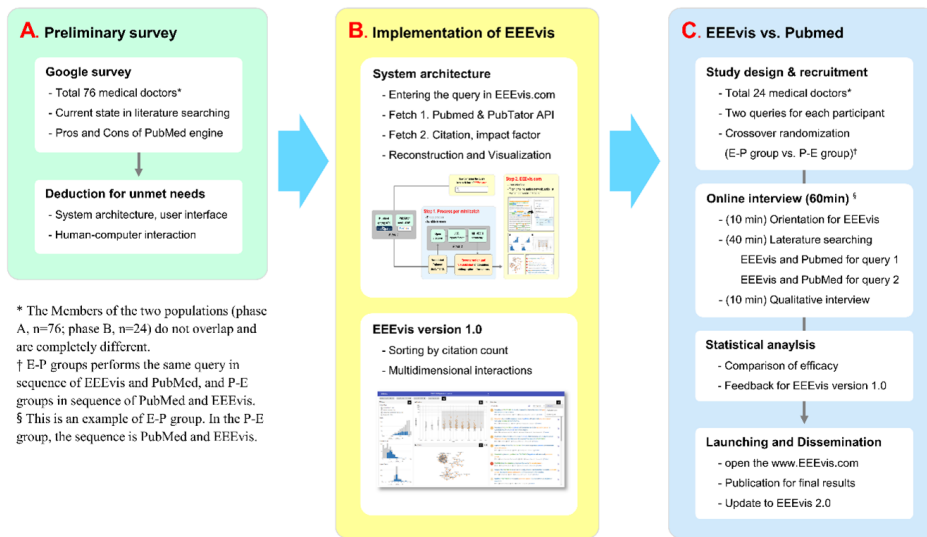
Figure 4.1: Study scheme

4.1.2 Preliminary Survey: A Questionnaire on conventional literature search methods

We conducted an online questionnaire between January and May 2019 with 76 medical doctors (MDs) who held professional qualifications in various clinical departments and had experience in medical research to derive the requirements of the biomedical literature search task. The survey questions were categorized into three sections, as follows: (1) baseline information for the research career, (2) questions for current use patterns of PubMed or other

literature search interfaces, and (3) additional requirements for a literature search task. The survey questions are summarized in Table 4.1 and the baseline demographic findings of the participants are summarized in Table 4.2.

Section 1. Research experience

Q1. What is your area of expertise? (e.g. gastroenterology, neurology, radiology, pathology, preventive medicine, medical statistics, etc.) _____

Q2. What is your age?
1) Less than 30
2) 30 ~34
3) 35 ~39
4) 40 ~44
5) 45 ~50
6) 50 ~54
7) 55 ~59
8) 60 or more

Q3. What is your year of graduation from medical school? (e.g. 2008) _____

Q4. How many research articles have you been significantly involved as the first author, the corresponding author, or co-author having invested 3 workdays or more? The papers may include unpublished research.
1) 0
2) 1
3) 2
4) 3~5
5) 6~10
6) 11~20
7) 21 or more

Section 2. Literature search engine use experience

Q1. On what purpose do you mainly use PubMed? (Multiple choices allowed)

1) Writing Article

2) Searching materials for seminar

3) Searching for knowledges (e.g. research updates, drug information)

4) Routine updates of literatures on subjects of interest (including My NCBI subscription)

5) Meta-analysis

6) Others (_____)

Q2. What search engines do you use for medical literature search? (Multiple choices allowed)

1) PubMed

2) Embase

3) Google or Google Scholar

4) Scopus

5) Web of Science

6) Dbpia (in Korean database)

6) Others (_____)

Q3. When using PubMed, how frequently do you use Boolean operators? (e.g. AND, OR, TI, TA, AU, AD, DP)

1) Almost always

2) Often

3) Sometimes

4) Rarely or never

5) I rarely use PubMed, so this question is not applicable to me.

Q4. When using PubMed, which part of the window do you usually look? (Please refer to the images for the letter indications)

1) I mostly look at B part, and rarely at A and C parts.

2) I mainly look at B part, but I often use the filter function of A part.

3) I mainly look at B part, but I often use the MeSH term or related article links of C part.

4) I mainly look at A or C parts, and rarely look at B part.

5) I mainly look at B parts, and look at other parts as needed.

6) I rarely use PubMed, so this question is not applicable to me.

Q5. When using PubMed, how do you use "Sort by" function located below the search bar?

1) I rarely use it.

2) I use it depending on the search results. (e.g. 50 abstracts per page)

3) I have the preset settings which I prefer and use frequently.

4) I rarely use PubMed, so this question is not applicable to me.

Q6. When using PubMed, which features of part B do you usually look at first? (Please refer to the images for the letter indications)

1) I skim at the article titles to look for the suitable article for the purpose of my search.

2) I look at the author names and look for the leading expertise on the subject of interest.

3) I look at the journal names and look for that with high impact factor.

4) I sort the results by "Best Match" and skim the results.

5) I let abstracts appear and search for the keywords by using control + F.

6) I rarely use PubMed, so this question is not applicable to me.

Q7. When using PubMed, do you use "Related Article" function?

1) I don't know about the function.

2) I ignore the function and instead look for the results by using current keywords.

3) I use the function after I found the appropriate article.

4) I use it simultaneously with the current keywords.

5) I use it case by case.

6) I rarely use PubMed, so this question is not applicable to me.

Q8. How do you manage literature PDF files?

1) I download the reference PDF files when necessary and save them in computer folders.

2) I download the reference PDF files when necessary and save them in EndNote.

3) I download the reference PDF files when necessary but do not save them.

4) I routinely download the reference PDF files and save them in computer folders.

5) I routinely download the reference PDF files and save them in EndNote.

6) I do not download PDF files but organize the PMID or keywords in excel or word files.

7) There is no definite management pattern for me.

8) Others (_____)

Section 3. Functions for new search engine

Q1. Do you feel any shortcomings when using PubMed? You may write "None" if there aren't any, and you may write multiple opinions.

(_____)

Q2. What functions do you think need to be included in a new search engine? (Upto 3 choices possible)

1) Visualization of interaction between articles by keywords

2) Sorting the articles by journal impact factor

3) Sorting the articles by citation counts

4) Searching by the leading expertise of the subject

5) Visualization of articles on the subject by timeline

6) Automatic word completion function (e.g. AND, OR)

Q3. Is there any other function would you want a new search engine to include that is not included in the above question? You may write "None" if there aren't any, and you may write multiple opinions.

(_____)

Table 4.1: Preliminary Google survery

Variables	Count (n=76)
Age	
30-34	28
35-39	34
40-44	11
45 or more	3
Specialty	
Gastroenterology	50
Neuropsychiatry	6
Internal medicine	3
Cardiology	2
Emergency medicine	2
Family medicine	2
General surgery	2
Pulmonology	1
Orthopedic surgery	1
Radiology	1
Pathology	1
Rheumatology	1
Plastic surgery	1
Neurosurgery	1
Thoracic surgery	1
Laboratory medicine	1
Year of obtainment of medical license	
2000 or before	2
2001-2005	8
2006-2010	43
2011-2015	23
Duration of exposure to PubMed (year)	
0–4	5
5–9	36
10–15	28

Table 4.2 continued from previous page

Variables	Count (n=76)
16 or more	7
Number of involved research articles as an author	
0	2
1~2	23
3~5	27
6~10	10
11~15	9
16 or more	5

Table 4.2: Demographic findings of preliminary survey

Among the 76 participants, 34 (45%) were between the ages of 35 and 39, and 28 (37%) were aged 30 to 34. Gastroenterologists constituted 50 (66%) of the participants. A major portion (93%) of the participants had been using PubMed for at least 5 years, and over 97% of the participants had published at least one medical article as a lead author.

As illustrated in Figure 4.2, the questionnaire consisted of the following five questions: (a) Which search tools do you use? (multiple possible responses); (b) How often do you use Boolean operators in PubMed?; (c) Which part do you see first (in the PubMed results view)?; (d) What are your urgent needs for PubMed?; and (e) Which function would you like in a new search engine? (multiple possible responses).

In question (a), the most commonly used medical literature search tools were PubMed (93%), followed by Google or Google Scholar (83%). In question (b), Boolean operators such as AND, OR, and NOT were used by 74% of participants. In the primary screening process for the result window of
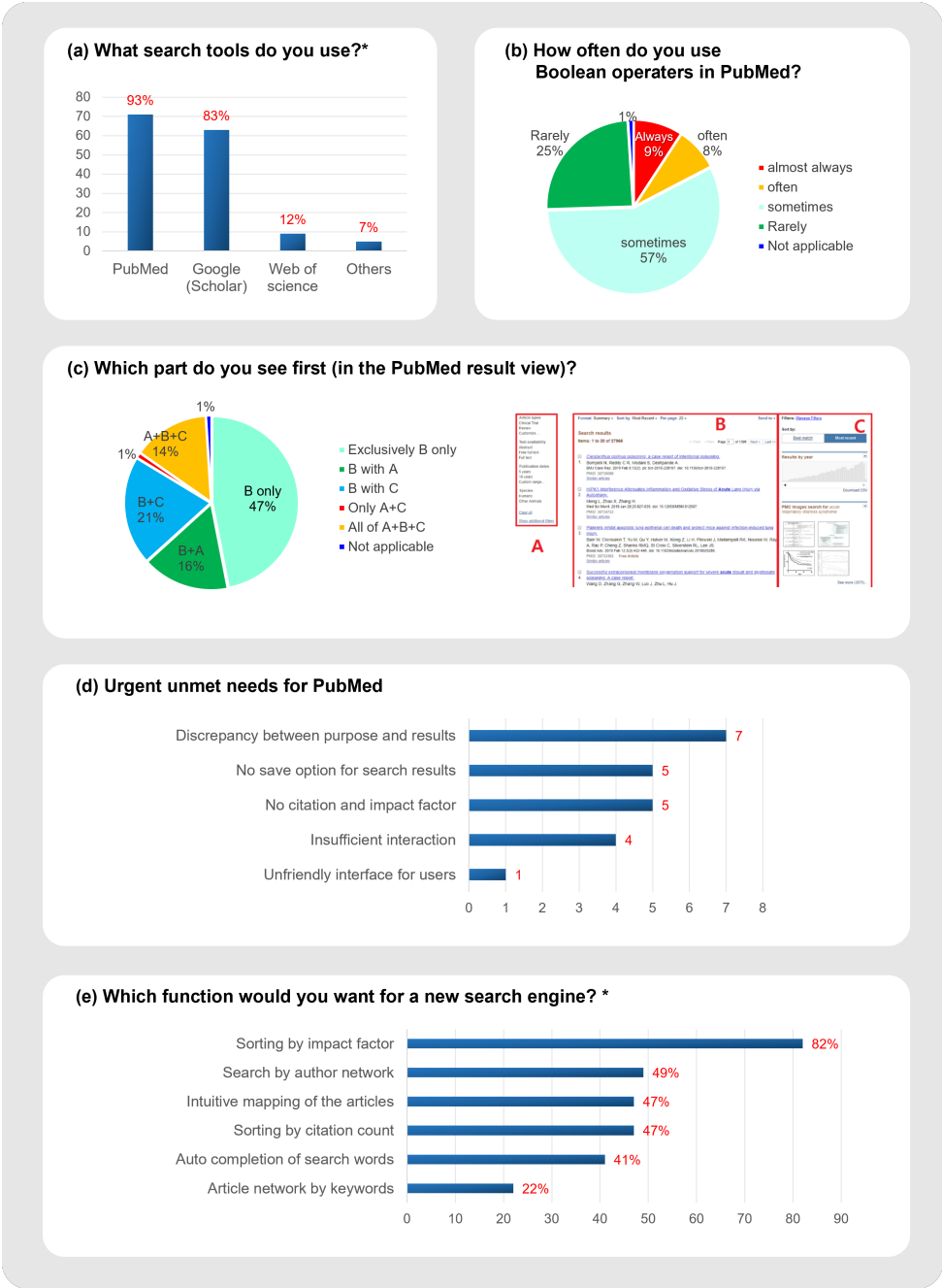
Figure 4.2: Results of preliminary survey.
∗ Multiple choice was possible up to three answers.

PubMed, in question (c), 47% of participants selected the "body" section only. Users who screened all parts of the body and both sidebars constituted 14% of all participants.

In question (d), 28% of the participants reported the demands and requests to enhance the results of PubMed. The most submitted reply was to reduce the discrepancy between the search purpose and results. Finally, in open-ended question (e), the answers were submitted in the order: sorting by impact factor (82%), searching by author network (49%), sorting by citation count (47%), intuitive mapping of articles (47%), auto-completion of search words (41%), and article network by keywords (22%).

4.1.3 Design Requirements for Biomedical Literature Search Systems

The preliminary survey aided in providing an understanding of the design requirements for a biomedical literature search system. Based on the results, we determined the following requirements that should be supported by a biomedical literature search system:

R1. Construct advanced search queries through direct manipulation.

R2. Present an overview for the bibliographic information.

R3. Present the trend and importance of the articles.

R4. Present the bibliographic relationship with the co-author network.

R5. Provide filtering and sorting functions to narrow the search.
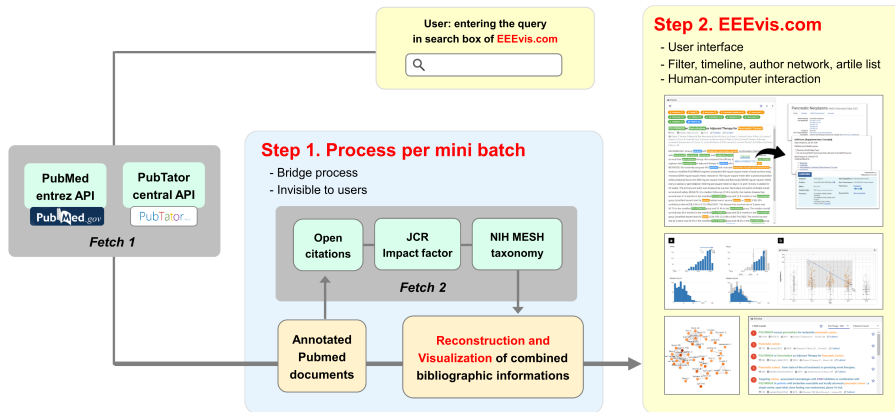
Figure 4.3: System architecture for EEEvis.

4.2 System and Interface Implementation of EEEVis

4.2.1 System Overview

We designed and implemented visual analytic system known as EEEVis to assess and enhance the biomedical literature search results from the PubMed database. Through EEEVis, users can create a search query through direct manipulation that supports the PubMed search query syntax and visualize the results.

EEEVis is a web-based system that employs a client–server architecture (Figure 4.3). With the client users input PubMed syntax-based search queries (`https://pubmed.ncbi.nlm.nih.gov/help/`) and the results are fetched from the server. The web application is implemented using Angular (`https://angular.io/`), whereas the charts are implemented using Plotly.js (`https://plotly.com/javascript/`) and Vega (`https://vega.github.io/vega/`).

The server is a RESTful web service implemented in Python that fetches the PubMed search results of the query using PubMed Entrez Programming Utilities API [99]. After retrieving the search results, the server annotates the results with PubTator Central biomedical concept annotations [119], citation counts from the OpenCitations (`https://opencitations.net/`) and NIH MeSH databases (`https://www.nlm.nih.gov/mesh/meshhome.html`), and impact factors from Journal Citation Reports (`https://jcr.clarivate.com/jcr/home`).

However, while the PubMed web search interface presents the results in a page-based list and fetches only a limited number (10 200) of articles on each page, EEEVis fetches the complete article dataset (up to 10,000 results) of the search query to hand out an overview. Larger search results lead to a longer fetch time and slower user response. EEEVis splits the fetch process into a series of tasks of fetching partial mini batches to overcome the fetch time delay and progressively updates the interface on each batch. This approach is similar to the progressive visual analytics (PVA) method that allows users to explore partial data analysis results in integrated and interactive visualizations during the execution [106].

Users can interactively explore the PubMed search results that are annotated with citation data and biomedical concept tags in an interface consisting of four coordinated views: (1) Bibliography Filters, (2) a Timeline View, (3) a Co-authorship Network View, and (4) an Article List & Detail View (Figure 4.4).

Figure 4.4: An overview of EEEVis. (a) Bibliography Filters, (b) Timeline view, (c) Co-authorship view, (d) Article list & detail view, (e) Filter status bar.
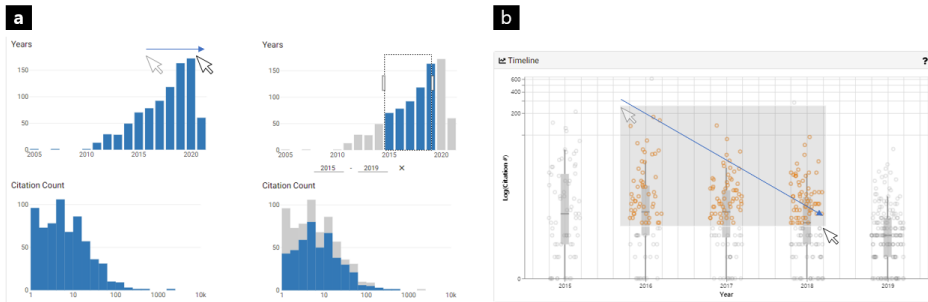
Figure 4.5: a) Filtering: the documents in the range of interest are filtered by dragging the mouse horizontally on each metadata histogram. Each histogram is coordinated and EEEVis supports a cross-filter function, so that the filter result (blue histogram) will be an intersection of each filter. b) Focusing: focusing is a bushing and linking interaction. The interaction is similar to filtering; however, it is an independent filter function that does not alter the document subset of the filtering. This mouse interaction highlights the documents between the two coordinated views, the Timeline View, and the Author Network View.

4.2.2 Bibliography Filters

This view provides an overview of the bibliographic features: the article type, year of publication, citation count, and impact factor. Each feature is mapped to a visual representation based on its data type. The visualizations are coordinated using the brushing and linking [14, 20] technique so that users can filter a user-interest subset of the search query result with specific values or ranges of the feature in a visualization and all other linked visualizations also immediately reflect the filtering result (Figure 4.5).

The article types, which are categorical values, are listed with the number of corresponding articles. The list supports multiple selections using a checkbox interface. A bar chart is employed to represent discrete values, such as the publication year, whereas histograms are used to represent continuous values, such as the citation count and impact factor. The continuous values

are binned into 20 bins, which are calculated to have similar physical widths on the x-scale. As the citation count and impact factor usually exhibit skewness towards small values in a vast range, the x-axes are implemented on a logarithmic scale. The bar charts and histograms support interactive x-axis range selection (brushing) by dragging an area on the chart or inserting specific values into the text boxes below the x-axis. The gray bars represent the number of articles in the search query result and the blue bars represent the intersection of all filters (linking), thereby constituting the user-interest subset. Every filter is displayed with chip buttons in the filter status bar (Figure 4.4). The other three coordinated views display the user-interest subset.

4.2.3 Timeline View

Figure 4.6: Timeline view. A series of box plots on an ordinal timeline x-axis. The box plots are a standard Tukey box plot (the median, the 1st quartile, and the 3rd quartile). The y-axis represents the citation counts with a logarithmic scale.

Users can already assess the annual trend of the number of articles per year using the year of publication histogram in the bibliography filters. However, as the citation count is also an important metric in literature searches, several tasks require the identification of the temporal trends of citation counts. Thus, we plot a series of boxplots according to the year of publication to reveal the annual trends (Figure 4.6).

The x-axis is an ordinal timeline of each year of publication and the y-axis is the logarithmic scale of the citation counts. Every box plot represents the statistical data of the articles that are published each year. The box plots are in the form of a standard Tukey boxplot that displays the median, first quartile, and third quartile. The whisker ranges from the smallest to largest data within the range [Q1 – 1.5 * IQR, Q3 + 1.5 * IQR], where Q1 and Q3 are the first and third quartiles, and IQR is the interquartile range (Q3 - Q1). Every article is encoded as a circular data point and jittered over the box plot to display the actual citation counts and distribution.

As the citation count is also correlated with the age of the article, the use of a single global range filter may lead to the omission of recent important articles or noise from many uninteresting old research papers. Thus, we implement a secondary filter function known as "Focus" in this view to dynamically explore a particular subspace of the time by citation count space. Users can focus on any timespan or citation count range by dragging a rectangular area on the chart. (Figure 4.5.b) The data points that correspond to the focus subset are color-coded in orange, but the complementary ones are in gray. These secondary filters are also displayed in the filter status bar

with the prefix "Focus." As a secondary brushing and linking technique, the Author Network View and Article List View also reflect this focused subset.

4.2.4 Co-authorship Network View
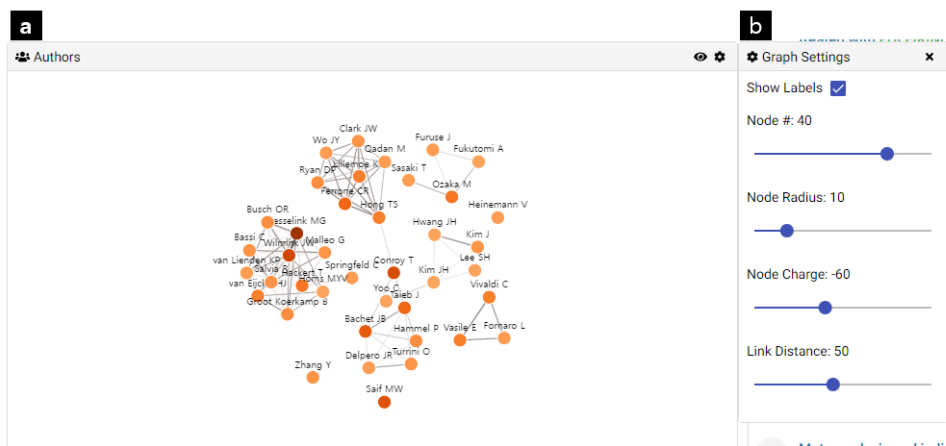
Figure 4.7: Co-authorship Network View. a) the view represents the co-authorship network among the most publishing authors of the targeted subset with a force-directed node-link graph. The number of authorships and co-authorships are visually encoded into the node color and the link strength. b) users can adjust the graph properties with the controls in the optional configuration side panel.

The analysis of bibliometric networks [12, 40, 78, 91], such as co-authorship or co-citation networks, has always been of major interest in the literature analysis domain. We implement an interactive force-directed node-link graph to reveal the co-authorship network, and to identify the hubs and authorities of the user-filtered subset (Figure 4.7).

The node-link graph is constructed using the top-k publishing authors. The default value of the k-value, which is the number of nodes, is 40. The number of articles by each author is color-coded into the color of each node using a sequential orange color scheme. The links between the author nodes

indicate co-authorship between the authors, whereas the width and saturation of each link represent the strength of the link. In this view, users can identify the leading authorities of the user-interest subset, highly connected subgroups, hub nodes of the subgroups, and bridge authors among the subgroups.

However, the graph properties, namely the number of nodes, node radius, node charge strength, and link distance, may need to be adjusted to reveal the network topology efficiently. Users can configure these properties in real time using the Graph Settings panel to monitor the network topology. Users can select each node and link to focus on the articles by an author and the articles to which authors contributed together. As the Co-authorship Network View, Timeline View, and Article List View are coordinated views, they all reflect this focused subset.

4.2.5   Article List and Detail View

The articles of the search result are presented (Figure 4.4.d). The list responds to the filter and focus interactions, and displays the user-interested subset of the articles. Users can sort the list according to several options: PubMed relevance [69], publication date, and citation count. Each list item includes the title, citation count, publication journal, impact factor, publication year, and authors. In an item, a circular glyph is placed ahead of the title, which is color-coded using an orange color scheme. The glyph color encodes the citation count of the article, which serves as an indicator of popular articles. Users can bookmark specific articles of interest by toggling the star-shaped button that follows each item. The list can be switched between dis-

playing all articles or bookmarked articles by toggling the star-shaped button in the list toolbar.

The title is annotated with biomedical concepts using the PubTator Central application programming interface (API) [119]. PubTator Central parses natural language and recognizes biomedical named entities [120]. A total of six biomedical concept categories are annotated using different colors: genes, diseases, chemicals, mutations, species, and cell lines.
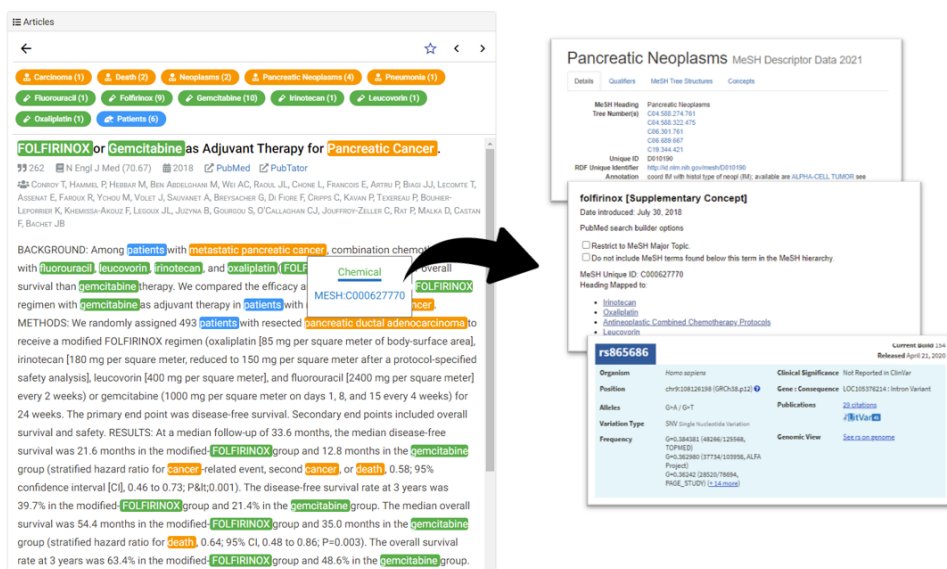
Figure 4.8: Article List & Detail View. The Detail View displays the title, authors, citation counts, publishing source and data, external links to PubMed and PubTator, the abstract, and the PubTator annotations. Users can click on each annotation, which provides external links to the information of the biomedical entity.

When a user selects an article from the list or clicks on an article point from the Timeline View, the List View is transformed into the Detail View (Figure 4.8), which displays the annotated title, abstract, and more detailed information of the biomedical concept tags. A list of the concepts that appear in the title and abstract text is placed above the title so that users can com-

prehend the main keywords without reading the entire text. Users can also click on a tag and identify the ID of each biomedical term, such as MeSH term or NCBI Taxonomy IDs. The detailed information of each biomedical concept is provided, with a link to an external database.

4.3   User Study

4.3.1   Participants

We conducted a randomized controlled user survey of 24 MDs between November to December 2020. This population was completely different from that of the preliminary user survey. All participants were first randomly assigned to one of two groups and a crossover randomization method (1:1) using a random number table was adopted for random allocation.

As indicated in Table 4.3, all of the participants were in the first or second year of their fellowship and were specialists who were actively conducting medical research. The mean age was 36 years, and 33% of the participants were female. The participants had obtained a doctor's license and had written an average of three papers as a first author. The participants had a median of four years of experience using PubMed; however, they had no exposure to EEEvis prior to the randomized controlled study. The population of this randomized controlled study (n = 24) was completely different from that of the preliminary survey (n = 76).

Variables	Values (n=24)
Age (year)	36 (33–38)*
Gender	

Table 4.3 continued from previous page

Variables	Values (n=24)
Female	8
Male	1
Research career	
Years after medical college graduate	7 (5–9)*
Published articles as a main author	3 (1–6)*
Exposure to search browser	
Previous exposure to PubMed	100 (100.0)
Duration of exposure to PubMed (years)	4 (2–6)*
Subspecialty of fellowship	
Gastroenterology (non-HBP)	3
Gastroenterology (pancreatobiliary)	3
Gastroenterology (hepatology)	4
Oncology	2
Hematology	2
Cardiology	1
Pulmonology	1
Nephrology	1
Surgery, HBP	1
Surgery, gastric	1
Surgery, orthostatic	1
Surgery, thoracic	1
Pediatrics	1
Obstetrics	1
Anesthesiology	1

Table 4.3: Demographics findings of randomized controlled user test
∗ Data are demonstrated as median (range)

4.3.2 Procedures

The two groups were as follows: (1) the E-P group, which performed the same query in the sequence of EEEvis and PubMed, and (2) the P-E group, which performed the same query in the sequence of PubMed and EEEvis. Once the participants had signed the consent form, the actual study was conducted as a 1:1 online Zoom meeting between the interviewer and interviewee. The interview was conducted for a total of 60 minutes, and it consisted of 10 minutes of orientation and free use, 40 minutes of actual literature searching, and 10 minutes of user feedback (Figure 4.9).
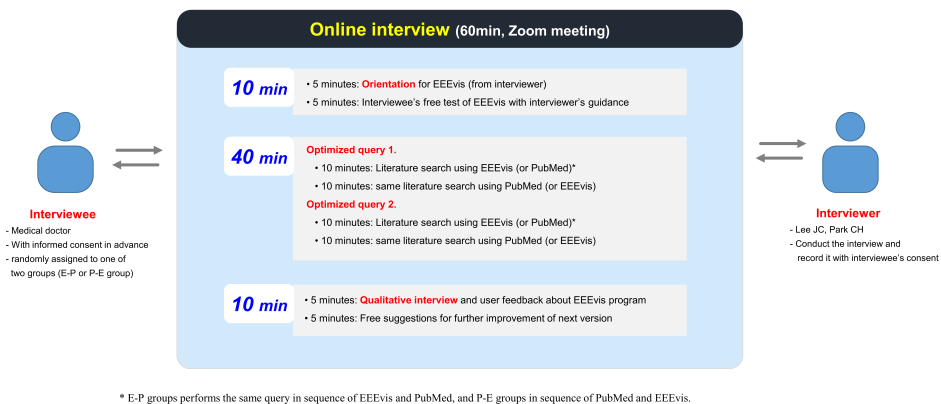
Figure 4.9: Schematic flow for randomized controlled user study.
∗ The E-P group performed the same query in the sequence of EEEvis and PubMed, and the P-E group performed the query in the sequence of PubMed and EEEvis.

In the first 10 minutes, as the participants had not been exposed to EEEvis before, the interviewer taught them how to use EEEvis for 5 minutes, following which they could freely use the system and test it for the remain-

ing 5 minutes. If any question arose during use, the interviewer immediately provided an answer.

For the 40 minutes of literature searching, two optimized queries were provided according to the subspecialty of the interviewee. The first query was presented on the assumption that when participants prepare for a presentation at a conference or seminar, they find main reference articles on the subject. The interviewee could select up to 5 suitable papers for their requirements and performed a paper search using EEEvis or PubMed for a maximum time of 10 minutes. For the following 10 minutes, the same query was executed using another search program that was not previously used. If the same process was repeated twice, the previous 10 minutes could be considered as learning about the corresponding query, which could lead to a bias (i.e., practice effect) whereby the results appeared to be better in the later 10 minutes. We randomly assigned the entire group to either the E-P or P-E group to correct this bias.

The second query was also optimized according to the subspecialty of the interviewee. In this case, when an academic thesis is written and a specific sentence is used, the researcher is searching for a reference article that can serve as a citation for that sentence. EEEvis and PubMed were assigned the same time of 10 minutes each and the order of the search interfaces was equally assigned according to the E-P or P-E group. The two queries that were presented for each field were provided blindly by assistant or associate professors in the corresponding subspecialty.

In the final 10 minutes, user feedback was provided based on open questions, and the results were qualitatively analyzed and summarized.

4.3.3 Results and Observations

Methodologies

Once all 24 participants had been interviewed, two researchers analyzed the video according to the standardized format, whereas two other researchers cross-checked the video. The overall analysis results were divided into two parts. The first part was related to the phase of entering search keywords, during which the time to achieve the optimal search strategy and loading time were measured as specific variables. The second part was the stage after the search results were presented, whereby the size of the search results, the number of selected articles, time to reach the first article, and time to complete the article selection were measured and analyzed. The existing assistant and associate professors who submitted the query evaluated how effectively the finally determined papers met the requirements and fit the topic. The papers were graded into three categories: (1) appropriate, (2) sub-optimal, and (3) inappropriate; subsequently, the results were aggregated and analyzed. Finally, the results of the user feedback were classified as advantages in comparison to PubMed and suggestions for further improvement in the subsequent version according to their frequency.

Task Results

Following the randomization in the sequence of PubMed and EEEvis, two queries were provided to each participant (Table 4.6), the search efficiency of which is compared in Table 4.4. None of the participants had experienced EEEvis prior to this trial, which means that this was the first exposure to EEEvis for every participant. However, the participants had an average of 4

years of experience in using PubMed. The median numbers of articles that were listed as query results were 1,423 (EEEvis) and 1,446 (PubMed). The median times to reach the first targeted article were 93 (36 to 125) and 90 (35 to 119) seconds, respectively. The median times to complete the search were 298 (140 to 415) and 306 (102 to 397) seconds, respectively (P = 0.771). A comparison of the results of the subjects to those of experts demonstrated that the relevance was 87.9% vs. 90.5% (P = 0.637), respectively.

Interview Observations

During the last 10 minutes of online interview, the participants were asked open questions regarding the advantages and disadvantages of EEEvis and for suggestions for further improvement in the future version (Table 4.54). A total of 21 (87.5%) participants mentioned the sorting and filtering by the citation count and impact factor as an advantage. Increasing the accuracy and granularity of the filtering interaction (58.3%) and reducing the search loading time (45.8%) were the most frequent suggestions for further improvement in the future version. Finally, 22 of 24 participants (91.7%) responded that they are willing to use EEEVis as their first choice for a biomedical literature search task.

	EEEvis	PubMed	P-value
Phase of entering search keywords			
Time to reach optimal search strategy (sec)	82 (10–119)	114 (11–184)	0.039
Loading time (sec)*	12 (8–18)	2 (1–2)	–
Phase after showing the results			
Number of listed articles per each query	1,423 (684 –2,205)	1,446 (712–2,428)	0.822
Number of selected articles per each query	4 (2–5)	4 (2–5)	0.537
Time to reach the first article (sec)	93 (36–125)	90 (35–119)	0.535
Time to complete article selection (sec)	298 (140–415)	306 (102 –397)	0.771
Quality of searched literature			
Sum of selected articles of overall 48 queries †	N = 207	N = 199	–
Number of optimal selections of articles §	182 (87.9)	180 (90.5)	0.637
Number of optimal plus suboptimal selections §	197 (95.1)	192 (96.5)	0.715

Table 4.4: Results of literature search task. Data are demonstrated as median (interquartile range) or number (%).
* The loading time of EEEvis is the duration of complete demonstration of all the results (1000 or more literature), whereas that of PubMed is the duration of dynamic demonstration of the first page results (initial 20 literature). Therefore, P-value was not presented in this row.
† Two queries were given per person out of a total of 24 participants.
§ Individual faculties in each subspecialty refereed the appropriateness of the article selection.

Advantages compared to PubMed	No (%)
Sorting and filtering by citation count and impact factor	21 (87.5)
Clean and intuitive interface	13 (54.2)
Article bookmark function	9 (37.5)
Co-authorship Network View	4 (16.6)
Accurate search output of important and target articles	3 (12.5)
PubTator Central word annotation	1 (4.2)
Quick access to abstract	1 (4.2)
Suggestion for further improvement in next version	**No (%)**
Mouse dragging sensitivity in filter section	14 (58.3)
Loading time of whole demonstration of time	11 (45.8)
Save and export bookmarked articles	8 (33.3)
Not suitable interface for mobile screen	3 (12.5)
Messy PubTator Central word annotation	3 (12.5)
No advanced search function, thereby, impossible to apply to meta-analysis	2 (8.3)
Misspellings even a little results in wrong results	2 (8.3)
No exclude option	1 (4.2)
Difficult to modulate Co-authorship Network View	1 (4.2)
Abstracts are not visible in the first place	1 (4.2)
No related or cited by articles	1 (4.2)
No direct full text link	1 (4.2)
Unable to save search options	1 (4.2)
Willingness to use EEEvis as the first choice for literature search	**No (%)**
Yes	22 (91.7)
No	2 (8.3)

Table 4.5: Feedback for EEEvis

4.4 Discussion

We have presented the literature searching tool called EEEVis, a novel interactive visual analytic system for the biomedical literature search task. We compared the search performance of EEEVis and PubMed with medical researchers and evaluated their usability. The search performance measured as the time to obtain the appropriate search results was comparable between the two systems, despite EEEVis being new to the subjects. Also, in the qualitative questionnaire after user study, the feedback of participants showed considerably favorable trends toward EEEVis.

With the developments in biomedical science, computer science, and information technology, the number of scientific papers published annually has increased by 8% to 9% each year, which is equivalent to over one million papers being entered into the PubMed database per year in the biomedical field alone [59]. Such number of papers cannot be handled by libraries, which has led to the development of websites for managing literature databases. These sites include PubMed, Web of Science (`https://www.webofscience.com/`), Scopus (`https://www.scopus.com/`), and Google Scholar (`https://scholar.google.com/`). Medical researchers may use one or more of these websites according to their purposes and preferences. Among these sites, PubMed is a free resource that is supported by the NIH, United States, and it covers most medical articles that are published. Furthermore, PubMed provides Boolean operators and a multi-criteria query syntax for detailed searching of the appropriate literature. These characteristics of broadness and specificities have resulted in the fact that almost all meta-analysis research includes PubMed as a search strategy.

However, the recent flooding of article publication has made it difficult for medical researchers to conduct accurate searches for key articles or to learn about a subject effectively. There are at least two reasons for this: First, the credibility of articles and journals varies. As an article with well-designed methodology and that with low reliable data is counted as one, it is becoming more time consuming and difficult for the medical researchers to identify "good" articles. Moreover, so-called predator journals do not employ sufficiently credible review processes, such as the careful selection of editors and thorough peer-review. The list of such journals may not be familiar to many researchers, which may cause other difficulties in medical research. Second, as the number of articles continues to increase exponentially, it is difficult to explore and narrow the massive set of biomedical literates with the current search interface. To narrow the search result and retrieve the most relevant target articles, the multi-criteria bibliographic query syntax is non-intuitive and comes with a steep learning curve.

We reflected the design requirements from the preliminary survey into EEEVis. The responses to the post-user-study questionnaire show the users found the best strength of EEEVis is to be the sorting and filtering functions (R5) of the impact factor and citation counts (R2). The runner-up response was the intuitive interface (R1), so we believe users found the system easy to use. Despite there being no significant difference between the task completion time of PubMed and EEEVis, users answered a preference for EEEVis over PubMed as a biomedical literature search tool (91.7%). We believe this shows that EEEVis fulfills the design requirements as a literature search tool and improves the user experience.

EEEVis follows a progressive visual analytic approach to reduce the time delay during a large-scale query search and provide interactivity to the system. However, despite the users being informed that the search results of EEEVis are updated progressively and users can explore the results during the fetch process, most users waited until the fetch process was completed. There was no explicit feedback on why users hesitated to explore the results during the fetch process. We suppose that the search query results during the user study tasks were not big enough(684 2205 articles) to invoke a severe fetching delay, and users did not have the necessity of exploring the approximate partial batch results.

4.4.1 Design Implications

According to our findings, we present two implications for improving the design of visual analytics systems for biomedical literature search.

Provide Overview for Bibliographic Features

While searching for optimal articles using EEEVis, every participant used the citation count and impact factor as key indicators for their search strategy. Novices may find it difficult to establish a search strategy when the search space is too wide and noisy. Conventional importance metrics, such as the citation counts and impact factor, may provide a cornerstone to explore the literature search space. Most of the participants identified the appropriate range of the values in which they were not interested with the bibliographic feature overviews and reduced the information noise of the search space by filtering out articles in which they were not interested. A total of 21 out of 24 participants pointed out that "Sorting and filtering by citation count and

impact factor" was the main advantageous feature of EEEVis compared to PubMed.

Represent Features in Multiple Coordinated Views

As novices often experience interpretation problems [43], presenting the search dataset with a single bibliographic feature at a time is not sufficient to provide a clear understanding. Presenting multiple bibliographic features in multiple coordinated views helped the users to understand the dataset more precisely. Users utilized the brushing and linking techniques to find specific regions of interest, and to reveal the connections among different bibliographic representations. Our findings suggest that biomedical literature search tools should provide different viewpoints of the data with multiple coordinated views.

4.4.2 Limitations and Future Work

EEEVis is still a work in progress, which means that several features require improvement. These include improving the range selection accuracy and granularity of the range filters and enabling to save the user bookmarked(stared) articles. Furthermore, the information on the impact factors and citation counts should be updated from an external database periodically, as these are not provided by the PubMed API.

Artificial intelligence-based models can enhance the quality of the user experience in literature search systems. Conventional search systems like PubMed or EEEVis provide search results with ranking algorithms that rank the relevance of the documents from the search keyword [38]. However, content-based ranking results are usually not self-explanatory, and users will

still need to seek through the text and bibliographic information to determine whether the documents are an appropriate match for them. With the attention mechanism-based language models and progress in the explainable AI field [67, 87, 113, 116], it is possible to visualize which part of the text is responsible for the relevance rankings. We believe that making search results more explainable will provide intuitive visual cues, and users will be able to develop more efficient search strategies.

The pilot study exhibits several limitations. First, the pre-development survey was conducted using a relatively small number of medical researchers with a limited background. Various additional opinions on PubMed or other search tools could be obtained with greater numbers and more diverse backgrounds, such as countries, mother languages, and ethnicities of medical researchers.

Second, the results of the user study should be interpreted with caution. The number of subjects was small, the participants had various fields of expertise and areas of interest, and the queries that were presented to the subjects were individualized. This resulted in heterogeneity, which may have masked the true difference in performance between the two systems. Furthermore, the methods that were used to measure the performance of the systems represented only a part of the overall performance, as the performance of search systems can be measured by many other means. The optimal methods for measuring the search performance remains a controversial subject.

Third, the task size might have been too small to observe different user behaviors between EEEVis and PubMed. In the case of PubMed, the median

number of query results was 1,446, and the median task completion time was 306 seconds. As we supposed that the task size might be why users did not explore the dataset between progressive updates, different task sizes might affect user behaviors. So, conducting case studies with real-world scenarios that usually have more extensive search results and longer task time might reveal significant insights between systems.

Finally, comparing behaviors among populations with different biomedical literature search experiences might be an interesting approach. In the user study, we recruited participants with a similar amount of experience in biomedical literature searching to control the effect of the expert level. However, several participants commented that EEEVis would be more effective for novice users than expert users who have already established an optimal search strategy.

4.5 Conclusions

In this study, we propose EEEVis, which is a novel interactive visual analytic system for a biomedical literature search task and a design guideline for the task. We demonstrated that the proposed system could improve the user experience in searching the appropriate literature by conducting a controlled user study. With the application of more future work, we expect EEEVis to become a system that can efficiently aid medical researchers in searching important articles of interest in the era of information overload.

Table 4.6. Two optimized queries for each participant

No	Subspecialty	Group*	Query 1†	Query 2§
1	Gastroenterology	P-E	(A) Are there any benefits using ursodeoxycholic acid after cholecystectomy?	(B) Helicobacter pylori infection is a definite risk factor for gastric cancer. VS. (C) It is controversial whether Helicobacter pylori infection is a risk factor for gastric cancer.
2	Gastroenterology	P-E	(A) Which regimen would be the best second-line chemotherapy for patients who progressed on lenvatinib?	(B) Helicobacter pylori infection is a definite risk factor for gastric cancer. VS. (C) It is controversial whether Helicobacter pylori infection is a risk factor for gastric cancer.
3	Gastroenterology	P-E	(A) What are the systemic treatment regimens available for patients with pancreatic neuroendocrine tumor?	(B) Neoadjuvant chemotherapy is beneficial for patients with pancreatic cancer. VS (C) It is controversial whether neoadjuvant chemotherapy is beneficial for patients with pancreatic cancer.
4	Hematology-Oncology	P-E	(A) What are the systemic treatment regimens available for patients with pancreatic neuroendocrine tumor?	(B) Neoadjuvant chemotherapy is beneficial for patients with pancreatic cancer. VS (C) It is controversial whether neoadjuvant chemotherapy is beneficial for patients with pancreatic cancer.

No	Subspecialty	Group*	Query 1†	Query 2§
5	Gastroenterology	E-P	(A) What are the systemic treatment regimens available for patients with pancreatic neuroendocrine tumor?	(B) Neoadjuvant chemotherapy is beneficial for patients with pancreatic cancer. VS (C) It is controversial whether neoadjuvant chemotherapy is beneficial for patients with pancreatic cancer.
6	Gastroenterology	P-E	(A) Which regimen would be the best second-line chemotherapy for patients who progressed on lenvatinib?	(B) Helicobacter pylori infection is a definite risk factor for gastric cancer. VS. (C) It is controversial whether Helicobacter pylori infection is a risk factor for gastric cancer.
7	Gastroenterology	E-P	(A) What are the systemic treatment regimens available for patients with pancreatic neuroendocrine tumor?	(B) Neoadjuvant chemotherapy is beneficial for patients with pancreatic cancer. VS (C) It is controversial whether neoadjuvant chemotherapy is beneficial for patients with pancreatic cancer.
8	Gastroenterology	P-E	(A) What are the possible systemic treatment options for the patients with biliary tract cancer who progressed on the combination of gemcitabine and cisplatin?	(B) Neoadjuvant chemotherapy is beneficial for patients with pancreatic cancer. VS (C) It is controversial whether neoadjuvant chemotherapy is beneficial for patients with pancreatic cancer.

61

No	Subspecialty	Group*	Query 1†	Query 2§
9	Gastroenterology	P-E	(A) What is the relationship between ulcerative colitis and colorectal cancer?	(B) Helicobacter pylori eradication for infected patients prevents stomach cancer development. VS. (C) It is controversial whether Helicobacter pylori eradication for infected patients prevents stomach cancer development.
10	Cardiology	E-P	(A) What is the appropriate duration of dual antiplatelet therapy for patients who underwent percutaneous coronary intervention?	(B) It is recommended for clinicians to start angiotensin receptor-neprilysin inhibitor for patients with acute decompensated reduced ejection fraction heart failure at the time of admission. VS. (C) It is controversial whether angiotensin receptor-neprilysin inhibitor should be started for patients with acute decompensated reduced ejection fraction heart failure at the time of admission.
11	Gastroenterology	E-P	(A) How many days are required for patients to stop antiplatelet or anticoagulation agents before undergoing endoscopic submucosal dissection for early gastric cancer?	(B) Intestinal metaplasia is a risk factor for gastric cancer. VS (C) It is controversial whether intestinal metaplasia is a risk factor for gastric cancer.

No	Subspecialty	Group*	Query 1†	Query 2§
12	Gastroenterology	E-P	(A) Which regimen would be the best second-line chemotherapy for patients who progressed on lenvatinib?	(B) Intestinal metaplasia is a risk factor for gastric cancer. VS. (C) It is controversial whether intestinal metaplasia is a risk factor for gastric cancer.
13	Nephrology	P-E	(A) Which one of parenteral or oral iron replacement is better for patients with chronic kidney disease?	(B) There are survival benefits of early formation of arteriovenous fistula in patients with end-stage renal disease who initiated hemodialysis. VS. (C) It is controversial whether there are survival benefits of early formation of arteriovenous fistula in patients with end-stage renal disease who initiated hemodialysis.
14	Obstetrics-Gynecology	P-E	(A) What is the optimal antibiotic regimen for intrauterine infection?	You are working on an article and planning to write a sentence on the first line optimal chemotherapeutic agent for mucinous type ovarian cancer.
15	Hematology-Oncology	E-P	(A) What is the optimal first-line treatment for chronic myeloid leukemia?	(B) Diagnosis of smoldering multiple myeloma is an indication for treatment initiation. VS (C) It is controversial whether the diagnosis of smoldering multiple myeloma is an indication for treatment initiation.

63

No	Subspecialty	Group*	Query 1†	Query 2§
16	Hematology-Oncology	E-P	(A) What is CAR-T cell therapy and what are the indications?	(B) Tumor mutation burden is a biomarker for immune checkpoint inhibitor response. VS. (C) It is controversial whether tumor mutation burden is a biomarker for immune checkpoint inhibitor response.
17	Pulmonology	P-E	(A) What is the indication of remdesivir for COVID-19?	(B) Steroid administration improves survival in patients with acute exacerbation of idiopathic pulmonary fibrosis. VS. (C) It is controversial whether steroid administration improves survival in patients with acute exacerbation of idiopathic pulmonary fibrosis.
18	Hematology-Oncology	E-P	(A) What are the treatment options for thrombotic thrombocytopenic purpura?	(B) In patients with hormone receptor positive breast cancer, the recommended duration of adjuvant hormone therapy is 5 years. VS (C) It is controversial on the duration of adjuvant hormone therapy for patients with hormone receptor positive breast cancer.

No	Subspecialty	Group*	Query 1†	Query 2§
19	Thoracic surgery	E-P	(A) What are the long-term complications in patients with left ventricular assist devices?	(B) Artery graft shows superior outcome compared with venous graft in coronary artery bypass graft surgery. VS. (C) It is controversial whether artery graft shows superior outcome compared with venous graft in coronary artery bypass graft surgery.
20	Pediatrics	E-P	(A) What is the biomarker used for monitoring acute kidney injury in premature infants?	(B) Younger children are more susceptible to urinary tract infection by ESBL-producing bacteria. VS. (C) It is controversial whether younger children are more susceptible to urinary tract infection by ESBL-producing bacteria.
21	Orthopedic surgery	P-E	(A) What are the novel treatment options for patients with osteosarcoma?	(B) Pelvic angioplasty showed superior outcome compared with peritoneal pelvic packing in pelvic fracture patients with hemodynamic instability. VS. (C) It is controversial whether pelvic angioplasty is superior outcome compared with peritoneal pelvic packing in pelvic fracture patients with hemodynamic instability.

65

No	Subspecialty	Group*	Query 1†	Query 2§
22	Anesthesiology	E-P	(A) Which one of inhalation or intravenous anesthesia is better for one-lung ventilation?	(B) Propofol has more protective effect on postoperative acute kidney injury compared with inhalant agents. VS. (C) It is controversial whether propofol has more protective effect on postoperative acute kidney injury compared with inhalant agents.
23	Hepatobiliary surgery	P-E	(A) What are the effects of preoperative biliary tract drainage on prognosis according to the location of extrahepatic biliary cancer?	(B) Pancreatic enzyme medication reduces the risk of fatty liver in patients who had pancreatectomy. VS. (C) It is controversial whether pancreatic enzyme medication reduces the risk of fatty liver in patients who had pancreatectomy.
24	Gastric surgery	P-E	(A) What are the surgical options for patients with stomach cancer and single liver metastasis?	(B) Long term fat-soluble vitamin supplements are necessary for patients who underwent bariatric surgery. VS. (C) It is controversial whether long term fat-soluble vitamin supplements are necessary for patients who underwent bariatric surgery.

Table 4.6: Two optimized queries for each participant

Table 4.6 Two optimized queries for each participant

∗ All the participants were randomly allocated into two groups: E-P means the sequence of EEEvis to Pubmed, and P-E means the sequence of PubMed to EEEvis.

† Query 1 assumes a situation where the participant makes a presentation on the 'topic (A)' at a specific conference. If you are unfamiliar with the topic (A), what key article would you find in that field to prepare for a presentation? For the same query, perform a literature search in the order of E-P or P-E. The time limit for each search engine is 10 minutes.

§ Query 2 assumes a situation where the participant writes an introduction or discussion part of a certain paper. If you were to select and use either 'sentence (B)' or 'sentence (C)' in a particular paragraph, which article would you cite (except participants 14)? For the same query, perform a literature search in the order of E-P or P-E. The time limit for each search engine is 10 minutes.

Chapter 5

GeneVis: A Visual Analytics System for Gene Signature Analysis in Cancers

This chapter introduces GeneVis, a novel web-based visual analytics system for gene signature analysis from RNA-Seq data in cancers genomics.

5.1   Motivation

A gene signature is *"a single or a group of genes in a cell having a unique gene expression pattern that is the consequence of either changed biological process or altered pathogenic medical terms"* [73]. Identifying gene signatures from RNA-Seq data is one of the most commonly used biomedicine techniques and has often been used to group samples in clinical research or treatment [73]. Many statistical analysis methods are already developed and utilized to determine differentially expressed transcripts among groups of samples [3, 8, 9, 16, 49, 66, 71, 72, 74, 79, 97, 115]. However, detecting the most informative signatures from this complex data is still challenging.

Identifying gene signatures and signature scores across multiple major cancer types is a commonly used technique in cancer genomics. However, this task is usually done by plotting and interpreting the gene expressions values across cancer types for a single gene and repeating it over multiple target genes. Because of the massive size of the gene candidates, this process requires tremendous human effort and burden to complete. An appealing way to make sense of this big data is to project it into a two-dimensional domain, where a single point represents each gene signature. Such representations can provide a perspective to interpret the patterns and similarities among the biomarker candidates. To aid in analyzing and interpreting gene signatures from RNA-Seq data, we introduce GeneVis, a web-based visual analytics system that visualizes a user-configured batch of gene signatures across major cancer types.

5.2   System and Interface Implementation

5.2.1   System Overview

We designed and implemented *GeneVis*, a novel web-based RNA-Seq gene expression visual analytic system. The system's objective is to visualize RNA-Seq gene expression data and aid cancer genomics studies. The system uses *The Cancer Genome Atlas(TCGA)* RNA-Seq data, a tabular data of 10,327 rows(10,327 tissue samples) and 20,527 columns(20,502 genes + 25 sample clinical variables). Each sample is one of the three tissue types(Tumor, Normal adjacent to the tumor, and blood) and one of the 32 cancer types. We design the user interface as a primary gene expression box-plot view surrounded by

Figure 5.1: An overview of GeneVis

two coordinated views that present the contextual information of a gene set overview (Figure 5.1).

When users input the genes of interest, a subset of the 20,502 genes, the system will fetch the expression data of the input genes from the server and visualize the results. The tabular expression data of the input genes can be exported as a CSV file. Users can search a gene name and add one by one (5.2.A) or input a list of gene names at once(5.2.B). The input gene list appears at the left sidebar(Figure 5.1.A) with an additional list item, 'Overall Average', the arithmetic mean of all input genes. Users can confirm the RNA-Seq gene expression data of each gene(or the overall average) with the expression detail view(Figure 5.1.F,G).
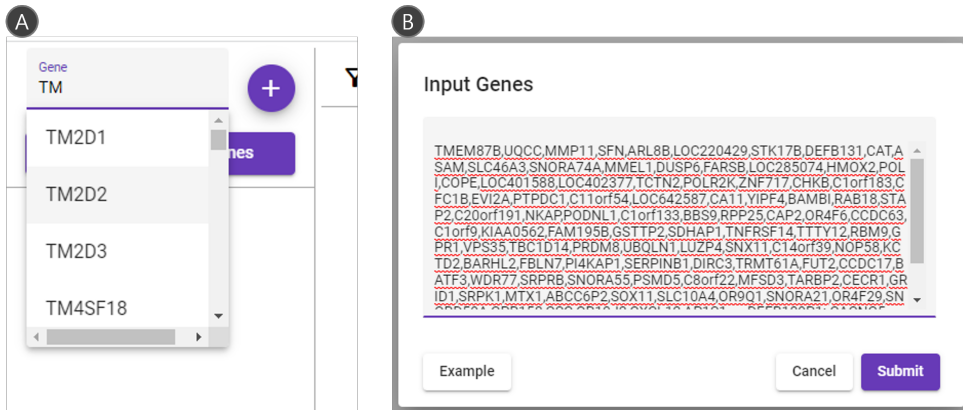

Figure 5.2: Input dialog for genes of interest

5.2.2   Gene Expression Detail View

The expression detail view presents the dispersion of the gene expression data(or the average). It contains two group of box plots, the per tissue type box plots(Figure 5.1.F) and the per cancer type box plots(Figure 5.1.F). The

71

box encodes the lower quartile(Q1), median, upper quartile(Q3), and the whiskers represent the extent values within the range [Q1 − 1.5 * IQR, Q3 + 1.5 * IQR] where IQR is the interquartile range (Q3 - Q1). Users can sort the plots alphabetically or numerically, zoom the plot to full screen, export the plot image, and export the plot data. In addition, users can plot the expression values with a jitter option(Figure 5.3. A) or group the per cancer type box plot by tissue types to see the subgroups of each cancer type(Figure 5.3.B).

5.2.3 Gene Vector Projection View

The expression detail view can only present the data of a single gene. Therefore, the two coordinated views, The projection view and the gene x cancer type heatmap, visualize the overview of the input gene list. The projection view shows a 2D UMAP scatter plot(Figure 5.1.D) of the gene expression vectors. The clusters reveal the expression pattern similarity among genes and indicate possible predictive biomarkers or relationships among genes. The scatter plot supports pan & zoom interaction and brushing & linking interactions with the mouse behavior.

The color of a point represents the central value(median, mean) and its variation(IQR, standard deviation) using the Value-Suppressing Uncertainty Palettes(VSUPs) [29]. The VSUP encodes the central value of the gene expressions using the hue of the Viridis colormap and the variation of the value using the saturation of the color [29]. In Figure 5.1.B, users can confirm the color legend and select the central value and variation metric to encode. In addition, users can always deselect the variation value and encode
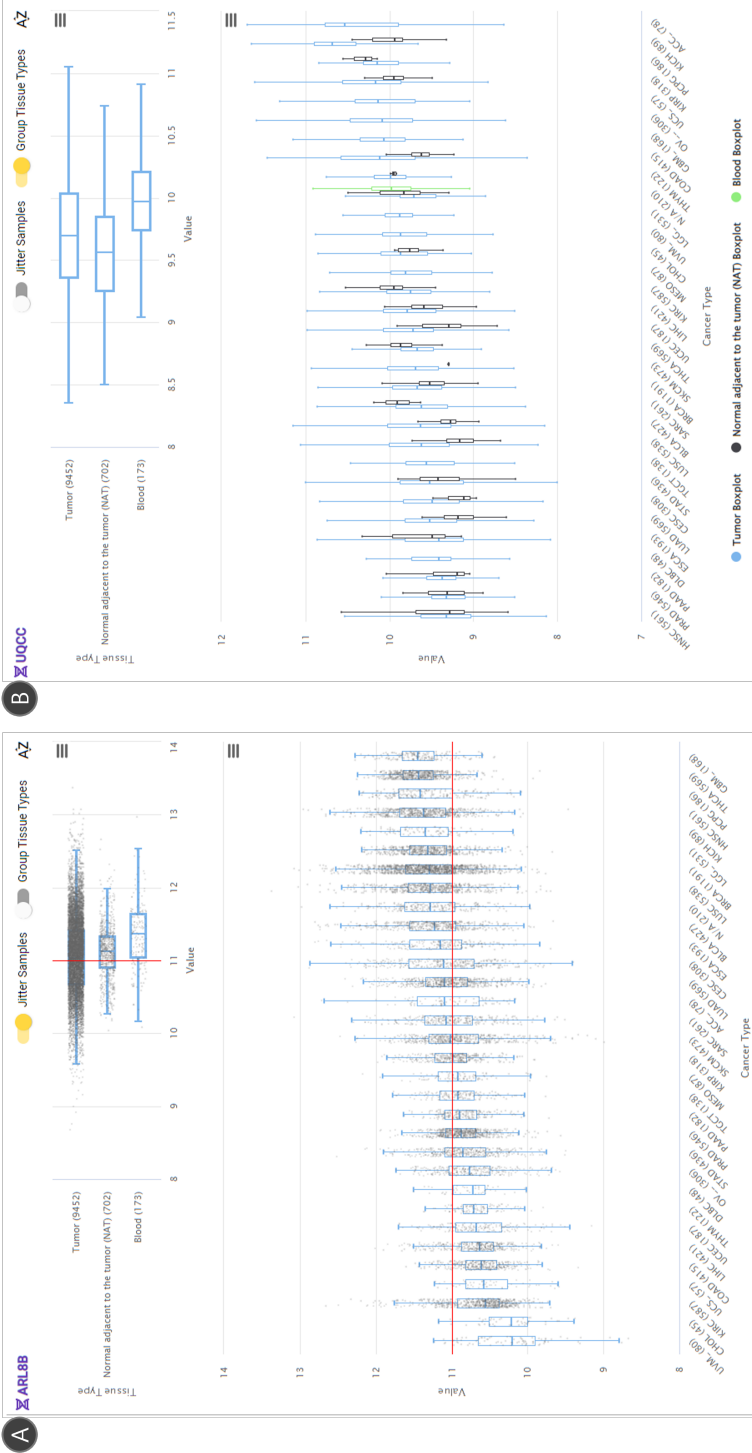
Figure 5.3: Gene Expression detail view. The panel displays an the distribution of the expression values (boxplot) for (1) each tissue type and (2) each cancer type each sample by each. (a) An interface that displays each sample on the boxplot. (b) A cancer type boxplot grouped by tissue types.

the points with a simple 1D discrete color scheme. With the 2D projection coordinates and VSUP color scheme, Users can explore the gene expression space and discover patterns such as gene clusters, genes with an anomaly or dispersed expression values.

5.2.4 Gene x Cancer Type Heatmap view

The gene-cancer heatmap(Figure 5.1. E) is a heatmap table where the rows represent each gene and the columns represent the cancer types. The color of each cell represents the expression value of the samples that match the corresponding gene and cancer type, a color from the global VSUP color legend. The cells can be shrunken to fit the table into the window and present the overview of the whole gene set or enlarged to increase the visibility of each row(Figure 5.4). With the heatmap, users can skim through cells and quickly capture the genes or cancer types with insight expression values. For example, users can identify a gene that has a very high expression value with a particular cancer type at a glance.

We applied hierarchical clustering on the gene vectors and placed a dendrogram on the right side of the table. The height of the dendrogram indicates the distance among the gene expression vectors, and the distance is calculated with the same distance function we used in the UMAP projection. We provide three distance options(Pearson correlation, cosine distance, and euclidean distance) from a dropdown list in the toolbar(Figure 5.5.C). Once a user changes the distance metric, the UMAP projection and dendrogram hierarchical clusters are recalculated, and the views are updated.
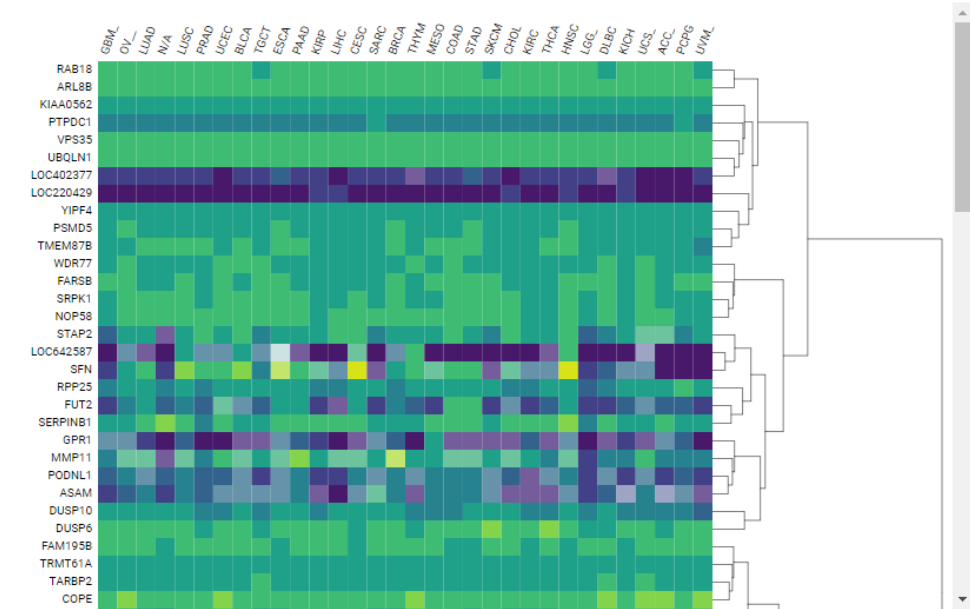
Figure 5.4: Heatmap + Dendrogram. A cancer type (columns) x gene (rows) heatmap of the expression values.

Figure 5.5: Toolbar

5.2.5 User Interaction in Multiple Coordinated Views

The filter controls in the toolbar(Figure 5.1.C) filter the samples and affect the other coordinated views. Once a user applies a filter on a tissue type or cancer type(Figure 5.5.A,B), the system recalculates the VSUP color legend, UMAP projection, and the hierarchical clustering. After the recalculation, the views and boxplots are redrawn.

The projection view supports brushing & linking, and once a user drags a region of interest on the scatter plot, the points within the region and the heatmap cells of those genes will be highlighted, and the others will be fade out(Figure 5.6). The expression cutoff filter in the toolbar fades out the points of the projection views and the cells of the heatmap that have the central value under the cutoff value.

5.3 Case Studies

5.3.1 Participants

We recruited experts with years of experience in cancer genomics and were familiar with the TCGA data set. We asked four experts to use GeneVis and share their experience with the system. The four experts were a bioinformatics system designer(P1), one clinical oncologist(P2), one cancer biomedical researcher with a clinical oncology background(P3), and a cancer bioinformatics researcher(P4).

5.3.2 Task and Procedures

The experts used GeneVis in their own work for a range of 1-3 days, and we conducted semi-structured interviews with each in a teleconference. Each
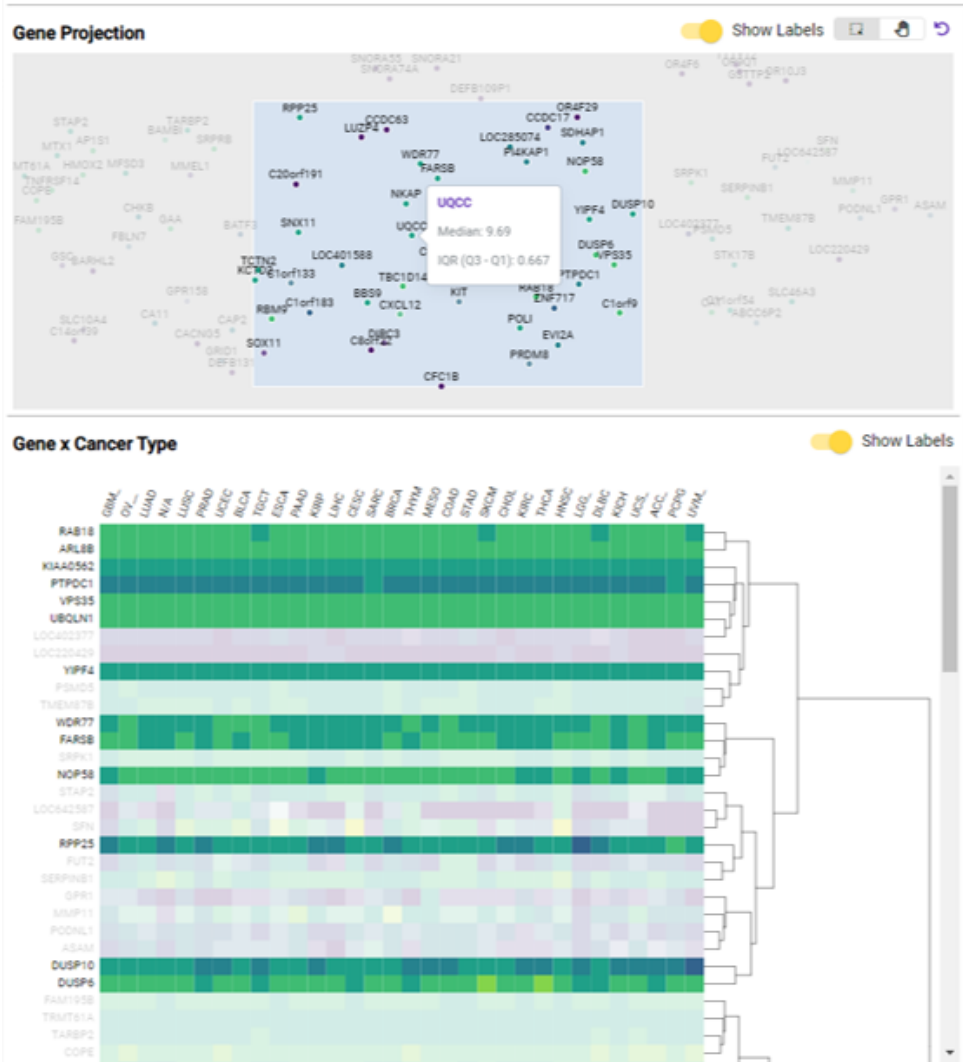
Figure 5.6: Brushing and Linking

session lasted for about 40-60 minutes, and each participant freely described what kind of data they used GeneVis for and demonstrated how they used GeneVis. Participants demonstrated one or more gene signature analyses. The participants were instructed to follow a think-aloud protocol and report anything they found interesting or felt like a barrier during the demonstration. After the demonstration, we conducted an open question semi-structured interview, and while the interview, we asked the following common questions.

Q1. For which tasks have you used GeneVis?

Q2. How many genes did you explore? What were those genes?

Q3. What did you goal to accomplish with GeneVis?

Q4. Before using GeneVis, how did you perform the tasks?

Q5. What is the advantages and disadvantages of the conventional methods?

Q6. Which feature or interface did you use most? How did it support your work?

Q7. What obstacles did you encounter? How did you address them?

Q8. How has your usage changed over time?

Q9. What features do you wish to add to GeneVis?

While we took notes during the sessions, all think-aloud demos and interview conversations were recorded for further analysis. Participants re-

sponded to a SUS (System Usability Scale) questionnaire After completing the interview.
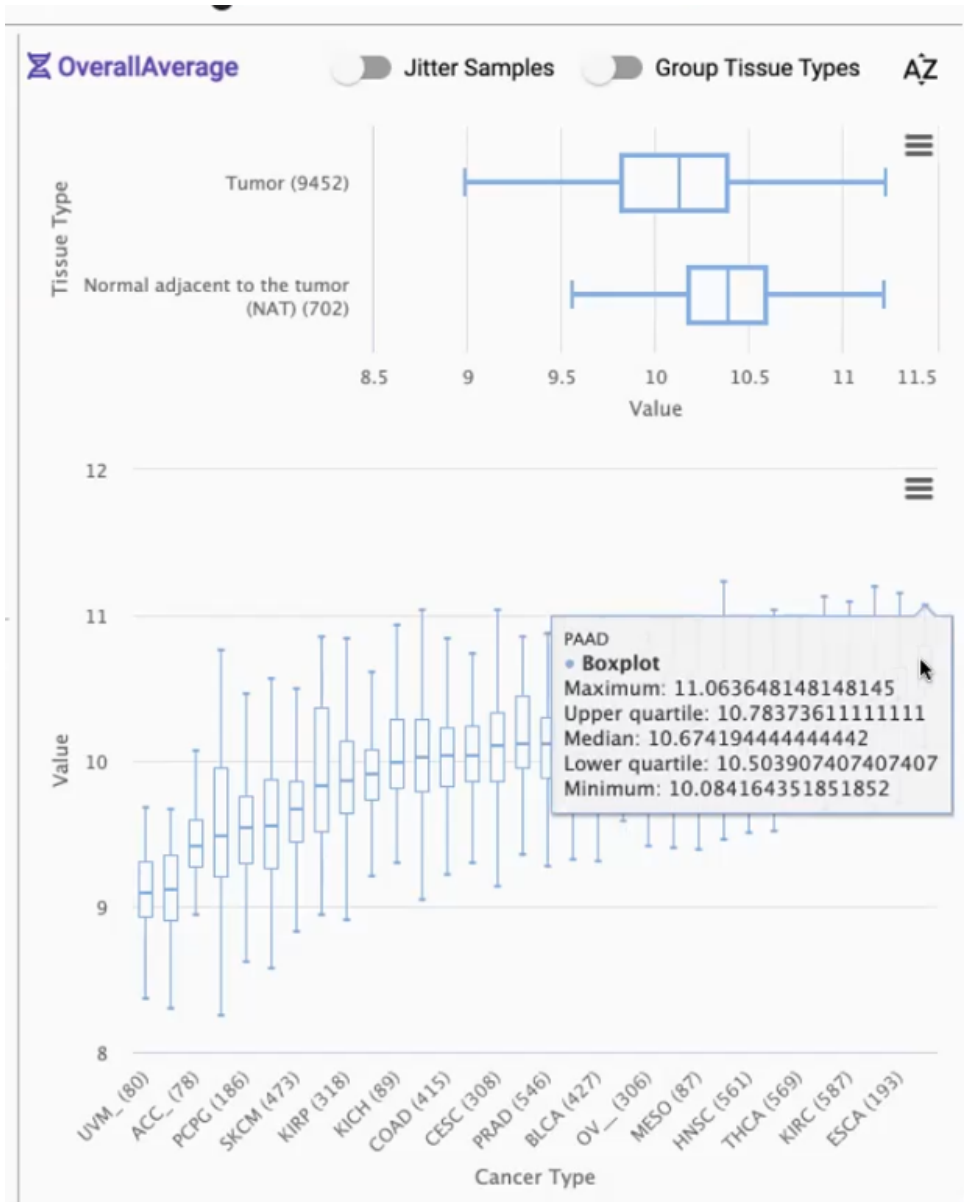
Figure 5.7: Case study P1: Data Validation based on prior knowledge

5.3.3   Case1: Identifying Similar Gene Signatures with TGFB1 in Hallmark Gene Sets

P3 is a Bio-informatics researcher working on various cancer precision medicine projects. P3 is also a medical doctor who has practiced as a clinical oncologist. P3's goal in this study was to identify 54 gene signatures across multiple cancer types of the hallmark gene sets up-regulated in response to the TGF-β cytokine.

After the gene set was loaded, P3 first checked the overall average option's cancer type box plots in the 'Gene Expression Detail View' to validate the data. With the expression data per tissue type plots, P3 confirmed the overall expression data showed a higher value in the 'Normal Adjacent to Tumor (NAT)' tissue types than the tumor tissue types (Figure 5.7). Moreover, with the per cancer type plots, P1 identified that expression values were higher in lung cancers (LUSC, LUAD) and pancreatic cancers (PAAD). These insights match the already known facts of the TGF-β cytokine, so P1 was assured that GeneVis could present the expression patterns in the RNA-Seq data.

From there on, P3 performed analysis using the 'Gene x Cancer type Heatmap' and found the essential genes (SMAD6, SMAD7, TGFB1) that he already knew as important regulators of TGF-β show a high expression value across most of the cancer types. By investigating the genes clustered with the essential genes in the dendrogram and those located near the Projection view, P3 found a gene of interest 'SERPINE1' (Figure 5.8). The SERPINE1 gene shows a similar gene signature as the TGFB1, and P3 now hypothesizes that SERPINE1 might also be an important regulator of TGF-β. After
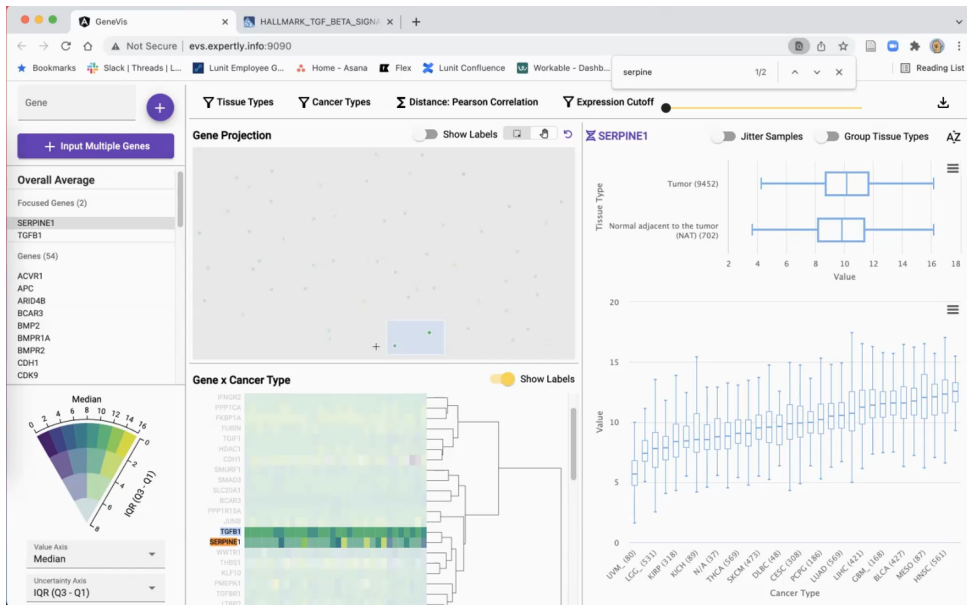
Figure 5.8: Case study P1: gene signature of SERPINE1

all, P3 downloaded the raw data and performed further analyses on SER-
PINE1 with R, such as gene expression correlation analysis or overall sur-
vival comparison analysis.

5.3.4 Case2: Identifying Cluster Patterns in the HRD data set

P2 is a clinical oncologist working at a hospital. P2's goal in this study was to
explore the gene expression patterns within the 566 gene sets of Homologous
recombination deficiency (HRD). HRD is a frequently observed common
characteristic of many tumor samples, especially in breast and ovarian cancer
[84].

At first, P2 also followed a path similar to P1 and checked the expres-
sion patterns of the tissue types and cancer types in the overall average item.
However, P2 performed his analysis mainly using the Gene Projection view
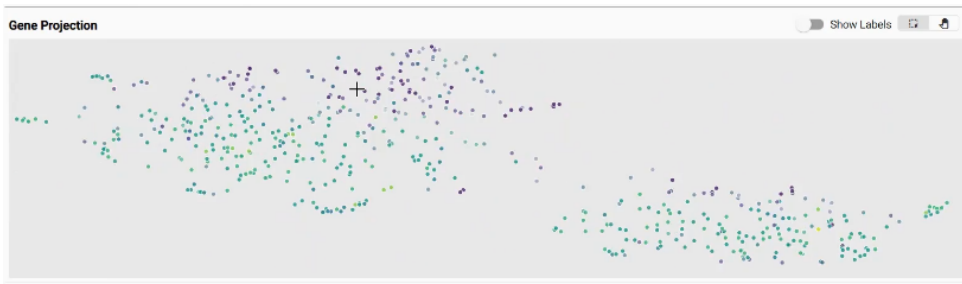
Figure 5.9: Case study P2: Positive correlation gene set and negative correlation gene set

after that step. P2 focused on identifying clusters and subgroups in the data set. However, there were no explicit, clear clusters at the beginning, P2 identified two significant clusters that divide the projection space when he filtered out other cancer types and focused on the visualization of ovarian cancer and breast cancer. P2 was interested in this result because each group's size and genes seemed similar to the gene sets in this group that have a positive/negative correlation with ovarian cancer and breast cancer incidence. P2 was also curious that the average expression value of each gene in the left group showed significantly different expression values, even though they were grouped as similar patterns.

5.3.5 Results

Usability

We used the System Usability Score (SUS) [18], a highly valuable and robust tool in helping assess the quality of system interfaces [10]. At the end of the interview, we asked the participants to answer a standard SUS questionnaire. The overall SUS score of GeneVis is 76.25 5.10), which is placed

#	Survey Question	A	B	C	D
1	I think that I would like to use this system frequently.	4	4	4	4
2	I found the system unnecessarily complex.	2	2	2	2
3	I thought the system was easy to use.	4	4	4	4
4	I think that I would need the support of a technical person to be able to use this system.	2	4	4	4
5	I found the various functions in this system were well integrated.	4	5	4	4
6	I thought there was too much inconsistency in this system.	1	2	1	2
7	I would imagine that most people would learn to use this system very quickly.	4	4	4	5
8	I found the system very cumbersome to use.	1	2	2	1
9	I felt very confident using the system.	5	4	4	5
10	I needed to learn a lot of things before I could get going with this system.	2	3	2	1
	SUS Score :	82.5	70	72.5	80

- A: Bioinformatics System Designer
- B: Oncology M.D.
- C: Oncology M.D.
- D: Bioinformatics Researcher

Average SUS Score: 76.25
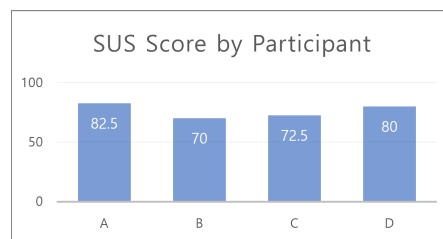
Figure 5.10: The SUS survey results of the four case studies.

in the acceptable range and the good–excellent range based on the study of Bangor et al. [10].

All four participants replied that GeneVis's user interface was intuitive and easy to learn. They found that the visualization features, except the VSUP color scheme, were familiar, easy to interpret, and tightly integrated. In addition, every participant responded that GeneVis reduces the vast human effort and time for the gene signature analysis in the cancer domain and would like to continue using the system with other real-life data sets.

Expert Feedback

During the semi-structured interview, we received various feedback from the participants. Most comments were about the VSUP color scheme in the

gene-cancer heatmap and new feature suggestions for an improved version of GeneVis. We address the following common comments.

- All four participants answered the common question Q4 (Before using GeneVis, how did you perform the tasks?) that they usually write a series of their own Python or R codes for an analysis task to process and visualize gene signature data. In addition, P3 and P4 answered that they usually repeat to conduct similar analysis tasks one by one for each gene of interest or each target cancer type by writing similar Python or R codes for each analysis. P2 answered similarly, but he also mentioned that he occasionally uses UCSC Xena [42].

- The participants commented that GeneVis was easy to learn and use to explore the gene signature patterns among multiple cancer types. They also commonly mentioned that compared to the conventional methods from (Q4), GeneVis reduced the burden of the vast human effort and shortened the task completion time (Q5).

- One common answer of P2 and P3 (whom both have a background as clinical oncologists) for Q9 (What features do you wish to add to GeneVis?) was a method to explore and seek the clinical outcomes (sex, age, overall survival rate Etc.) of the samples. P1 and P2 both wanted to plot and confirm the clinical outcome of the samples, and they also wanted to select a group of samples on-demand to compare the clinical outcomes of the groups.

- The most common suggestion for a future version of GeneVis (Q9) was adding a subgroup analysis feature. All four participants said they

wanted to continue the gene signature analysis with subgroups divided by the genes, cancer types, or clinical outcomes. While the case study, the participants commonly asked if they could select a group of samples by selecting the elements or selecting a region of interest in the projection view or the gene-caner heatmap view.

- Participant P2 suggested a statistical evaluation method (ANOVA, MANOVA, Etc.).

5.4   Summary

This chapter implemented GeneVis, a novel web-based interactive gene signature visual analytics system that visualizes the gene expression patterns from RNA-Seq data across major cancer types. We conducted four case studies with experienced clinical oncologists and biomedical researchers, and we asked them about their use of GeneVis. Based on the interview responses and the SUS survey results, the case studies showed that GeneVis facilitates the multi-cancer type gene signature analysis and provides a better understanding of the gene set RNA-Seq data.

Chapter 6

Conclusion and future work

6.1 Conclusion

This dissertation presents the results of three design studies on visual analytics systems for laborious evidence-based tasks in medical research. In chapter 3, we introduce PLOEM, a visual analytics system to synthesize broad evidence and generate an evidence map in a standardized procedure and visual representation. The use cases show how the system can facilitate the task and how the users can better understand the evidence gaps and the trends. In chapter 4, we conducted a preliminary survey with 76 medical doctors to derive the requirements for a biomedical literature search. Based on the results, We designed and implemented EEEVis, an interactive visual analytic system for biomedical literature search tasks. A system enhances the PubMed search result with several bibliographic visualizations and PubTator annotations. We performed a user study with 24 medical doctors to evaluate the system and presented the design guidelines and challenges for a biomedical literature search system design. In chapter 5, we implemented GeneVis, a novel web-based interactive gene signature visual analytics sys-

tem that visualizes the gene expression patterns from RNA-Seq data across major cancer types. Case studies with oncology medical doctors and biomedical researchers showed that GeneVis facilitates the multi-cancer type gene signature analysis and provides a better understanding of the gene set RNA-Seq data.

6.2   Future Work

We use machine learning techniques with GeneVis in chapter 5. We employ the UMAP algorithm for a gene vector 2D projection view and the hierarchical clustering algorithms for the dendrogram. However, there are more evidence-based tasks that state-of-the-art machine learning models can facilitate in medical research.

For example, visualizing the similarity among literature abstract sections with the state-of-the-art language models, such as BERT [33] or GPT-3 [19], can provide the researchers with a better visual landscape of the literature evidence space. In addition, researchers could utilize this feature in clustering evidence or facilitating the literature search task by drilling down the evidence space into the target literature.

Another interesting approach would be utilizing the biological entities we displayed in EEEVis (chapter 4). Chen et al. [25] trained and developed a concept embedding model, BioConceptVec, that captures the semantics of biological concepts by training the model with 30 million PubMed abstracts. We hypothesize with projecting the biological concept embeddings that belong to the literature of an evidence search task, a researcher can have a brief overview of the semantics of the evidence literature. Therefore we are im-

Figure 6.1: A new EEEVis version with SOTA machine learning based document similarity visualizations and BioC entity projections

plementing and integrating these visualizations into EEEVis and planning to study the impacts on medical researchers in the future (Figure 6.1).

Bibliography

[1] A. K. Akobeng. Principles of evidence based medicine, 2005.

[2] Alexis Allot, Kyubum Lee, Qingyu Chen, Ling Luo, and Zhiyong Lu. Lit-suggest: a web-based system for literature recommendation and curation using machine learning. *Nucleic Acids Research*, 49(W1):W352–W358, 2021.

[3] U. Alon, N. Barka, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences of the United States of America*, 96(12):6745–6750, 1999.

[4] Michelle D. Althuis and Douglas L. Weed. Evidence mapping: method-ologic foundations and application to intervention and observational research on sugar-sweetened beverages and health outcomes. *American Journal of Clinical Nutrition*, 98(3):755–768, 2013.

[5] Göran Arnqvist and David Wooster. Meta-analysis: synthesizing research findings in ecology and evolution, 1995.

[6] Matthew H. Bailey et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*, 173(2):371–385, April 2018.

[7] Michelle Q.Wang Baldonado and Terry Winograd. Sensemaker: an information-exploration interface supporting the contextual evolution of a user's interests. en. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 11–18. ScholarOpenURL Seoul National University, 1997.

[8] Sanghamitra Bandyopadhyay and Saurav Mallik. Integrating multiple data sources for combinatorial marker discovery: a study in tumorigenesis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(2):673–687, 2018.

[9] Sanghamitra Bandyopadhyay, Saurav Mallik, and Anirban Mukhopadhyay. A survey and comparative study of statistical tests for identifying differen-

tial expressionfrom microarray data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 11(1):95–115, 2014.

[10] Aaron Bangor, Philip T. Kortum, and James T. Miller. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6):574–594, 2008.

[11] Tanya Barrett, Karen Clark, Robert Gevorgyan, Vyacheslav Gorelenkov, Eugene Gribov, Ilene Karsch-Mizrachi, Michael Kimelman, Kim D. Pruitt, Sergei Resenchuk, Tatiana Tatusova, Eugene Yaschenko, and James Ostell. Bioproject and biosample databases at ncbi: facilitating capture and organization of metadata. *Nucleic Acids Research*, 40(D1), 2012.

[12] Donald De B. Beaver. Reflections on scientific collaboration (and its study): past, present, and future. *Scientometrics*, 52(3):365–377, 2001.

[13] Fabian Beck, Sebastian Koch, and Daniel Weiskopf. Visual analysis and dissemination of scientific literature collections with survis. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):180–189, January 2016.

[14] Richard A. Becker and William S. Cleveland. Brushing scatterplots. *Technometrics*, 29(2):127–142, 1987.

[15] Jacques S. Beckmann and Daniel Lew. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. *Genome Medicine*, 8(1), 2016.

[16] Soumya Bhattacharjee, Kaushik Renganaath, Rajesh Mehrotra, and Sandhya Mehrotra. Combinatorial control of gene expression, 2013.

[17] Peter Bragge, Ornella Clavisi, Tari Turner, Emma Tavender, Alex Collie, and Russell L. Gruen. The global evidence mapping initiative: scoping research in broad topic areas. *BMC Medical Research Methodology*, 11, 2011.

[18] John Brooke. Sus—a quick and dirty usability scale. 1996. en. *Usability evaluation in industry*, 189(194):4–7, 1996.

[19] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 2020-Decem, 2020. arXiv: 2005.14165.

[20] Andreas Buja, John Alan McDonald, John Michalak, and Werner Stuetzle. Interactive data visualization using focusing and linking. In *Proceedings of the 2nd Conference on Visualization 1991, VIS 1991*, 1991.

[21] Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, Benjamin E. Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, Caitlin J. Byrne, Michael L. Heuer, Erik Larsson, Yevgeniy Antipin, Boris Reva, Arthur P. Goldberg, Chris Sander, and Nikolaus Schultz. The cbio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, 2(5):401–404, May 2012.

[22] Evelina Chapman, Ludovic Reveiz, Amy Chambliss, Stephanie Sangalang, and Xavier Bonfill. Cochrane systematic reviews are useful to map research gaps for decreasing maternal mortality. *Journal of Clinical Epidemiology*, 66(1):105–112, 2013.

[23] Duen Horng Chau, Aniket Kittur, Jason I. Hong, and Christos Faloutsos. Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 167–176, 2011.

[24] Qingyu Chen, Alexis Allot, and Zhiyong Lu. Litcovid: an open database of covid-19 literature. *Nucleic Acids Research*, 49(D1):D1534–D1540, 2021.

[25] Qingyu Chen, Kyubum Lee, Shankai Yan, Sun Kim, Chih Hsuan Wei, and Zhiyong Lu. Bioconceptvec: creating and evaluating literature-based biomedical concept embeddings on a large scale. *PLoS Computational Biology*, 16(4), 2020. arXiv: 1912.10846.

[26] J. K. Chou and C. K. Yang. Papervis: literature review made easy. *Computer Graphics Forum*, 30(3):721–730, 2011.

[27] Ngai Chow, Lucas Gallo, and Jason W Busse. Evidence-based medicine and precision medicine: complementary approaches to clinical decision-making. *Precision Clinical Medicine*, 1(2):60–64, 2018.

[28] Francis S. Collins. Exceptional opportunities in medical science: a view from the national institutes of health, 2015.

[29] Michael Correll, Dominik Moritz, and Jeffrey Heer. Value-suppressing uncertainty palettes, 2018.

[30] Jo Cross. Medline, pubmed, pubmed central, and the nlm. *Editors' Bulletin*, 2(1):1–5, 2006.

[31] Issa J. Dahabreh, Nira Hadar, and Mei Chung. Emerging magnetic resonance imaging technologies for musculoskeletal imaging under loading stress: scope of the literature, November 2011.

[32] Helena M.L. Daudt, Catherine Van Mossel, and Samantha J. Scott. Enhancing the scoping study methodology: a large, inter-professional team's experience with arksey and o'malley's framework, 2013.

[33] Jacob Devlin, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: pre-training of deep bidirectional transformers for language understanding. In *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, volume 1, pages 4171–4186, 2019. arXiv: 1810. 04805.

[34] Angelo Di Iorio, Raffaele Giannella, Francesco Poggi, Silvio Peroni, and Fabio Vitali. Exploring scholarly papers through citations. In *DocEng 2015 - Proceedings of the 2015 ACM Symposium on Document Engineering*, pages 107–116. Association for Computing Machinery, Inc, September 2015.

[35] Ao Dong, Wei Zeng, Xi Chen, and Zhanglin Cheng. Vistory: interactive storyboard for exploring visual information in scientific publications. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, September 2019.

[36] Wei Duan-Porter, Karen Goldstein, Jennifer McDuffie, Megan Clowse, Jaime M Hughes, Ruth Klap, Varsha Masilamani, Nancy M Allen LaPointe, and John W C N - NBK362890 Williams. Mapping the evidence: sex effects in high-impact conditions for women veterans – depression, diabetes, and chronic pain. *VA Evidence-based Synthesis Program Reports*, 2015.

[37] Evidence-Based Medicine Working Group. Evidence-based medicine. a new approach to teaching the practice of medicine. evidence-based medicine working group. *JAMA: The Journal of the American Medical Association*, 268(17):2420–2425, 1992.

[38] Nicolas Fiorini, Kathi Canese, Grisha Starchenko, Evgeny Kireev, Won Kim, Vadim Miller, Maxim Osipov, Michael Kholodov, Rafis Ismagilov, Sunil Mohan, James Ostell, and Zhiyong Lu. Best match: new relevance search for pubmed. *PLoS Biology*, 16(8), 2018.

[39] Jianjiong Gao, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S. Onur Sumer, Yichao Sun, Anders Jacobsen, Rileen Sinha, Erik Larsson, Ethan Cerami, Chris Sander, and Nikolaus Schultz. Integrative analysis of complex cancer genomics and clinical profiles using the cbioportal. *Science Signaling*, 6(269), April 2013.

[40] Wolfgang Glänzel and András Schubert. Analysing scientific networks through co-authorship. In *Handbook of Quantitative Science and Technology Research*, pages 257–276. 2006.

[41] Gene V. Glass. Meta-analysis at middle age: a personal history. *Research Synthesis Methods*, 6(3):221–231, September 2015.

[42] Mary J. Goldman, Brian Craft, Mim Hastie, Kristupas Repečka, Fran Mc-Dade, Akhil Kamath, Ayan Banerjee, Yunhai Luo, Dave Rogers, Angela N. Brooks, Jingchun Zhu, and David Haussler. Visualizing and interpreting cancer genomics data via the xena platform, May 2020.

[43] Lars Grammel, Melanie Tory, and Margaret Anne Storey. How information visualization novices construct visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):943–952, 2010.

[44] Maria J. Grant and Andrew Booth. A typology of reviews: an analysis of 14 review types and associated methodologies, June 2009.

[45] Robert L. Grossman, Allison P. Heath, Vincent Ferretti, Harold E. Varmus, Douglas R. Lowy, Warren A. Kibbe, and Louis M. Staudt. Toward a shared vision for cancer genomic data. *New England Journal of Medicine*, 375(12):1109–1112, September 2016.

[46] Jessica Gurevitch, Julia Koricheva, Shinichi Nakagawa, and Gavin Stewart. Meta-analysis and the science of research synthesis, March 2018.

[47] Ted Haines, Leslie McKnight, Eric Duku, Lenora Perry, and Achilleas Thoma. The role of systematic reviews in clinical research and practice, 2008.

[48] Jiangen He, Qing Ping, Wen Lou, and Chaomei Chen. Paperpoles: facilitating adaptive visual exploration of scientific publications by citation links. *Journal of the Association for Information Science and Technology*, 70(8):843–857, August 2019.

[49] Zengyou He and Weichuan Yu. Stable feature selection for biomarker discovery, 2010. arXiv: 1001.0887.

[50] Florian Heimerl, Qi Han, Steffen Koch, and Thomas Ertl. Citerivers: visual analytics of citation patterns. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):190–199, January 2016.

[51] Sarah E. Hetrick, Alexandra G. Parker, Patrick Callahan, and Rosemary Purcell. Evidence mapping: illustrating an emerging methodology to improve evidence-based practice in youth mental health. *Journal of Evaluation in Clinical Practice*, 16(6), 2010.

[52] Katherine A. Hoadley et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, 158(4):929–944, August 2014.

[53] Michael D. Jennions, Christopher J. Lortie, and Julia Koricheva. Role of meta-analysis in interpreting the scientific literature. In *Handbook of Meta-analysis in Ecology and Evolution*, pages 364–380. 2013.

[54] Cyriac Kandoth, Michael D. McLellan, Fabio Vandin, Kai Ye, Beifang Niu, Charles Lu, Mingchao Xie, Qunyuan Zhang, Joshua F. McMichael, Matthew A. Wyczalkowski, Mark D.M. Leiserson, Christopher A. Miller, John S. Welch, Matthew J. Walter, Michael C. Wendl, Timothy J. Ley, Richard K. Wilson, Benjamin J. Raphael, and Li Ding. Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471):333–339, 2013.

[55] Pauline Kang, Michael Kalloniatis, and Gordon S. Doig. Using updated pubmed: new features and functions to enhance literature searches, 2021.

[56] Taraneh Khazaei and Orland Hoeber. Supporting academic search tasks through citation visualization and exploration. *International Journal on Digital Libraries*, 18(1):59–72, March 2017.

[57] Julia Koricheva, Jessica Gurevitch, and Kerrie Mengersen. *Handbook of meta-analysis in ecology and evolution*. 2013.

[58] Christin Katharina Kreutz, Peter Boesten, Alex Witry, and Ralf Schenkel. Facetsearch: a faceted information search and exploration prototype. en. In *CEUR Workshop Proceedings*, volume 2191, pages 215–226, 2018.

[59] Esther Landhuis. Scientific literature: information overload, 2016.

[60] Shahid Latif and Fabian Beck. Vis author profiles: interactive descriptions of publication records combining text and visualization. en. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):152–161, 2019.

[61] Jessica W. Lau, Erik Lehnert, Anurag Sethi, Raunaq Malhotra, Gaurav Kaushik, Zeynep Onder, Nick Groves-Kirkby, Aleksandar Mihajlovic, Jack DiGiovanna, Mladen Srdic, Dragan Bajcic, Jelena Radenkovic, Vladimir Mladenovic, Damir Krstanovic, Vladan Arsenijevic, Djordje Klisic, Milan Mitrovic, Igor Bogicevic, Deniz Kural, and Brandi Davis-Dusenbery. The cancer genomics cloud: collaborative, reproducible, and democratized - a new paradigm in large-scale computational research. *Cancer Research*, 77(21):e3–e6, November 2017.

[62] Bongshin Lee, Mary Czerwinski, George Robertson, and Benjamin B. Bederson. Understanding research trends in conferences using paperlens. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 1969–1972, 2005.

[63] Jong Chan Lee, Soyeon Ahn, In Kuk Cho, Jongchan Lee, Jaihwan Kim, and Jin Hyeok Hwang. Management of recurrent pancreatic cancer after surgi-

cal resection: a protocol for systematic review, evidence mapping and meta-analysis. *BMJ Open*, 8(4), April 2018.

[64] Sukwon Lee, Sung Hee Kim, and Bum Chul Kwon. Vlat: development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):551–560, 2017.

[65] Simon Lehmann, Ulrich Schwanecke, and Ralf Dörner. Interactive visualization for opportunistic exploration of large document collections. *Information Systems*, 35(2):260–269, April 2010.

[66] Jie Li, Xianglong Tang, Wei Zhao, and Jianhua Huang. A new framework for identifying differentially expressed genes. *Pattern Recognition*, 40(11):3249–3262, 2007.

[67] Zhouhan Lin, Minwei Feng, Cicero Nogueira Dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. A structured self-attentive sentence embedding. In *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017. arXiv: 1703.03130.

[68] Chun Hung Lu, Chih Chien Wang, Min Yuh Day, Chorng Shyong Ong, and Wen Lian Hsu. Using "cited by" information to find the context of research papers. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5075 LNCS, pages 343–354, 2008.

[69] Zhiyong Lu, Won Kim, and W. John Wilbur. Evaluating relevance ranking strategies for medline retrieval. *Journal of the American Medical Informatics Association*, 16(1):32–36, 2009.

[70] Lauren A. Maggio, Olle Ten Cate, H. Carrie Chen, David M. Irby, and Bridget C. O'Brien. Challenges to learning evidence-based medicine and educational approaches to meet these challenges: a qualitative study of selected ebm curricula in u.s. and canadian medical schools, 2016.

[71] Saurav Mallik, Sagnik Sen, and Ujjwal Maulik. Idpt: insights into potential intrinsically disordered proteins through transcriptomic analysis of genes for prostate carcinoma epigenetic data. *Gene*, 586(1):87–96, 2016.

[72] Saurav Mallik and Zhongming Zhao. Congems: condensed gene co-expression module discovery through rule-based clustering and its application to carcinogenesis, 2018.

[73] Saurav Mallik and Zhongming Zhao. Identification of gene signatures from rna-seq data using pareto-optimal cluster algorithm. *BMC Systems Biology*, 12, 2018.

[74] Saurav Mallik and Zhongming Zhao. Traprm: transcriptomic and proteomic rule mining using weighted shortest distance based multiple minimum supports for multi-omics dataset. In *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017*, volume 2017-Janua, pages 2187–2194, 2017.

[75] Justin Matejka, Tovi Grossman, and George Fitzmaurice. Citeology: visualizing paper genealogy. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 181–189, 2012.

[76] Lidia Mateo, Oriol Guitart-Pla, Miquel Duran-Frigola, and Patrick Aloy. Exploring the oncogenomic landscape of cancer. *Genome Medicine*, 10(1):1–8, August 2018.

[77] Lisa G. McFerrin, Michael Zager, Jianan Zhang, Gretchen Krenn, Robert McDermott, Desert Horse-Grant, Emily Silgard, Kara Colevas, Paul Shannon, Hamid Bolouri, and Eric C. Holland. Analysis and visualization of linked molecular and clinical cancer data by using oncoscape, September 2018.

[78] G. Melin and O. Persson. Studying research collaboration using co-authorships. *Scientometrics*, 36(3):363–377, 1996.

[79] Anirban Mukhopadhyay and Monalisa Mandal. Identifying non-redundant gene markers from microarray data: a multiobjective variable length pso-based approach. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 11(6):1170–1183, 2014.

[80] Shinichi Nakagawa, Daniel W.A. Noble, Alistair M. Senior, and Malgorzata Lagisz. Meta-evaluation of meta-analysis: ten appraisal questions for biologists, March 2017.

[81] Shinichi Nakagawa, Gihan Samarasinghe, Neal R. Haddaway, Martin J. Westgate, Rose E. O'Dea, Daniel W.A. Noble, and Malgorzata Lagisz. Research weaving: visualizing the future of research synthesis, March 2019.

[82] Rina Nakazawa, Takayuki Itoh, and Takafumi Saito. A visualization of research papers based on the topics and citation network. In *Proceedings of the International Conference on Information Visualisation*, volume 2015-Septe, pages 283–289. Institute of Electrical and Electronics Engineers Inc., September 2015.

[83] Yulia Newton, Adam M. Novak, Teresa Swatloski, Duncan C. McColl, Sahil Chopra, Kiley Graim, Alana S. Weinstein, Robert Baertsch, Sofie R. Salama, Kyle Ellrott, Manu Chopra, Theodore C. Goldstein, David Haussler, Olena Morozova, and Joshua M. Stuart. Tumormap: exploring the molecular similarities of cancer samples in an interactive portal. *Cancer Research*, 77(21):e111–e114, November 2017.

[84] Luan Nguyen, John W. M. Martens, Arne Van Hoeck, and Edwin Cuppen. Pan-cancer landscape of homologous recombination deficiency. *Nature Communications*, 11(1), 2020.

[85] Monica Nicolau, Arnold J. Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences of the United States of America*, 108(17):7265–7270, April 2011.

[86] Katrien Oude Rengerink, Shakila Thangaratinam, Gemma Barnfield, Katja Suter, Andrea R. Horvath, Jacek Walczak, Anna Wełmińska, Susanne Weinbrenner, Berit Meyerrose, Theodoros N. Arvanitis, Rita Ónody, Gianni Zanrei, Regina Kunz, Chantal Arditi, Bernard Burnand, Harry Gee, Khalid S. Khan, and Ben W.J. Mol. How can we teach ebm in clinical practice? an analysis of barriers to implementation of on-the-job ebm teaching and learning. *Medical Teacher*, 33(3), 2011.

[87] Cheonbok Park, Jaegul Choo, Inyoup Na, Yongjang Jo, Sungbok Shin, Jaehyo Yoo, Bum Chul Kwon, Jian Zhao, Hyungjong Noh, and Yeonsoo Lee. Sanvis: visual analytics for understanding self-attention networks. In *2019 IEEE Visualization Conference, VIS 2019*, pages 146–150, 2019. arXiv: 1909 . 09595.

[88] Maksim V. Plikus, Zina Zhang, and Cheng Ming Chuong. Pubfocus: semantic medline/pubmed citations analytics through integration of controlled biomedical dictionaries and ranking algorithm. *BMC Bioinformatics*, 7, 2006.

[89] Antoine Ponsard, Francisco Escalona, and Tamara Munzner. Paperquest: a visualization tool to support literature review. In *Conference on Human Factors in Computing Systems - Proceedings*, volume 07-12-May-, pages 2264–2271. Association for Computing Machinery, May 2016.

[90] Sylvain Poux, Cecilia N Arighi, Michele Magrane, Alex Bateman, Chih Hsuan Wei, Zhiyong Lu, Emmanuel Boutet, Hema Bye-A-Jee, Maria Livia Famiglietti, Bernd Roechert, and The UniProt Consortium. On expert curation and scalability: uniprotkb/swiss-prot as a case study. *Bioinformatics (Oxford, England)*, 33(21):3454–3460, 2017.

[91] D. J. Price and D. D. Beaver. Collaboration in an invisible college. *The American psychologist*, 21(11):1011–1018, 1966.

[92] Dmitry Prokopenko, Julian Hecker, Edwin K. Silverman, Marcello Pagano, Markus M. Nöthen, Christian Dina, Christoph Lange, and Heide Loehlein Fier. Utilizing the jaccard index to reveal population stratification in sequencing data: a simulation study and an application to the 1000 genomes project. *Bioinformatics*, 32(9):1366–1372, May 2016.

[93] Fakher Rahim, Akbar Soltani, and Vahid Haghpanah. Developing organized level of biomedical evidence: evidence-based biomedicine. *Frontiers in Physiology*, 3 NOV, 2012.

[94] Sheila M. Reynolds, Michael Miller, Phyliss Lee, Kalle Leinonen, Suzanne M. Paquette, Zack Rodebaugh, Abigail Hahn, David L. Gibbs, Joseph Slagel, William J. Longabaugh, Varsha Dhankani, Madelyn Reyes, Todd Pihl, Mark Backus, Matthew Bookman, Nicole Deflaux, Jonathan Bingham, David Pot, and Ilya Shmulevich. The isb cancer genomics cloud: a flexible cloud-based platform for cancer genomics research. *Cancer Research*, 77(21):e7–e10, November 2017.

[95] Randy R. Richter and Tricia M. Austin. Using mesh (medical subject headings) to enhance pubmed search strategies for evidence-based practice in physical therapy. *Physical Therapy*, 92(1):124–132, 2012.

[96] Alexander Rind, Andrea Haberson, Kerstin Blumenstein, Christina Niederer, Markus Wagner, and Wolfgang Aigner. Pubviz: lightweight visual presentation of publication data. en. In *EuroVis* (*Short Papers*), volume 8, 4, pages 333–352. ScholarOpenURL Seoul National University, 2017.

[97] Matthew E. Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47, 2015.

[98] David L. Sackett, William M.C. Rosenberg, J. A.Muir Gray, R. Brian Haynes, and W. Scott Richardson. Evidence based medicine: what it is and what it isn't. 1996. *Clinical orthopaedics and related research*, 455:3–5, 2007.

[99] Eric Sayers. E-utilities quick start entrez programming utilities help entrez programming utilities help. *The Journal of Systems and Software*, 85:1930–1952, 2012.

[100] C. Schmucker, E. Motschall, G. Antes, and J. J. Meerpohl. Methoden des evidence mappings: eine systematische übersichtsarbeit. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 56(10):1390–1397, October 2013.

[101] Shashi S. Seshia and G. Bryan Young. The evidence-based medicine paradigm: where are we 20 years later? part 1, 2013.

[102] William R. Shadish and Jesse D. Lecy. The meta-analytic big bang. *Research Synthesis Methods*, 6(3):246–264, September 2015.

[103] Salimah Z. Shariff, Jessica M. Sontrop, R. Brian Haynes, Arthur V. Iansavichus, K. Ann McKibbon, Nancy L. Wilczynski, Matthew A. Weir,

Mark R. Speechley, Amardeep Thind, and Amit X. Garg. Impact of pubmed search filters on the retrieval of evidence by physicians. *CMAJ*, 184(3), 2012.

[104] Birte Snilstveit, Martina Vojtkova, Ami Bhavsar, Jennifer Stevenson, and Marie Gaarder. Evidence & gap maps: a tool for promoting evidence informed policy and strategic research agendas. *Journal of Clinical Epidemiology*, 79:120–129, 2016.

[105] John Stasko, Jaegul Choo, Yi Han, and Mengdie Hu. Citevis: exploring conference paper citation data visually. en. *Posters of IEEE InfoVis.*, 2:2–3, 2013.

[106] Charles D. Stolper, Adam Perer, and David Gotz. Progressive visual analytics: user-driven visual exploration of in-progress analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1653–1662, 2014.

[107] Nicole Sultanum, Christine Murad, and Daniel Wigdor. Understanding and supporting academic literature review workflows with litsense. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, September 2020.

[108] S. M. Sweeney, E. Cerami, A. Baras, T. J. Pugh, N. Schultz, T. Stricker, J. Lindsay, C. Del Vecchio Fitz, P. Kumari, C. Micheel, K. Shaw, J. Gao, N. Moore, T. Stricker, C. Kandoth, B. Reardon, E. Lepisto, S. Gardos, K. Dang, J. Guinney, L. Omberg, T. Yu, B. Gross, Z. Heins, D. Hyman, B. Rollins, C. Sawyers, D. Solit, D. Schrag, V. Velculescu, F. Andre, P. Bedard, M. Levy, G. Meijer, B. Rollins, and K. Shaw. Aacr project genie: powering precision medicine through an international consortium. *Cancer Discovery*, 7(8):818–831, 2017.

[109] Diem Trang Tran, Tian Zhang, Ryan Stutsman, Matthew Might, Umesh R. Desai, and Balagurunathan Kuberan. Anexvis: visual analytics framework for analysis of rna expression. *Bioinformatics*, 34(14):2510–2512, 2018.

[110] Harsha Trivedi, Heather M. Kling, Tina Treece, William Audeh, and Gordan Srkalovic. Changing landscape of clinical-genomic oncology practice, 2019.

[111] Edward R Tufte. *Beautiful evidence*, volume 1. Graphics Press Cheshire, CT, 2006.

[112] Nynke Van Dijk, Lotty Hooft, and Margreet Wieringa-De Waard. What are the barriers to residents' practicing evidence-based medicine? a systematic review, 2010.

[113] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 2017-Decem, pages 5999–6009, 2017. arXiv: 1706.03762.

[114] Daniela Vetter, Gerta Rucker, and Ilse Storch. Meta-analysis: a need for well-defined usage in ecology and conservation biology. *Ecosphere*, 4(6), June 2013.

[115] Andrew J. Vickers. Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. *BMC Medical Research Methodology*, 5, 2005.

[116] Jesse Vig. A multiscale visualization of attention in the transformer model. In *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of System Demonstrations*, pages 37–42, 2019. arXiv: 1906 . 05714.

[117] Ding Ding Wang, Marissa Shams-White, Oliver John M. Bright, J. Scott Parrott, and Mei Chung. Creating a literature database of low-calorie sweeteners and health studies: evidence mapping. *BMC Medical Research Methodology*, 16(1):1–11, January 2016.

[118] Yaoyu E. Wang, Lev Kutnetsov, Antony Partensky, Jalil Farid, and John Quackenbush. Webmev: a cloud platform for analyzing and visualizing cancer genomic data. *Cancer Research*, 77(21):e11–e14, November 2017.

[119] Chih Hsuan Wei, Alexis Allot, Robert Leaman, and Zhiyong Lu. Pubtator central: automated concept annotation for biomedical full text articles. *Nucleic Acids Research*, 47(W1):W587–W593, 2019.

[120] Chih Hsuan Wei, Hung Yu Kao, and Zhiyong Lu. Pubtator: a web-based text mining tool for assisting biocuration. *Nucleic acids research*, 41(Web Server issue), 2013.

[121] Junjun Zhang, Rosita Bajari, Dusan Andric, Francois Gerthoffert, Alexandru Lepsa, Hardeep Nahal-Bose, Lincoln D. Stein, and Vincent Ferretti. The international cancer genome consortium data portal, April 2019.

[122] Xiaolong Zhang, Yan Qu, C. Lee Giles, and Piyou Song. Citesense: supporting sensemaking of research literature. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 677–680, 2008.

[123] Jian Zhao, Michael Glueck, Fanny Chevalier, Yanhong Wu, and Azam Khan. Egocentric analysis of dynamic networks with egolines. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 5003–5014. Association for Computing Machinery, May 2016.

[124] Xin Zhou, Michael N. Edmonson, Mark R. Wilkinson, Aman Patel, Gang Wu, Yu Liu, Yongjin Li, Zhaojie Zhang, Michael C. Rusch, Matthew Parker, Jared Becksfort, James R. Downing, and Jinghui Zhang. Exploring genomic alteration in pediatric cancer using proteinpaint, December 2015.

[125] Sandra Zwolsman, Ellen Te Pas, Lotty Hooft, Margreet Wieringa-De Waard, and Nynke Van Dijk. Barriers to gps' use of evidence-based medicine: a systematic review. *British Journal of General Practice*, 62(600), 2012.

국문 초록

근거중심의학(Evidence-Based Medicine)이란 "임상 치료 및 의학 연구에서 현재 존재하는 최고의 증거를 양심적이고, 명백하며, 분별 있게 이용하는 방법론"이며 [98], 현대 의학에서 가장 널리 받아들여지는 의학 패러다임이다. 신뢰할 수 있는 고수준의 과학적 근거를 검색, 검토, 합성하는 것이야 말로 근거중심의학의 핵심이다. 하지만, 근거중심의학이 이미 광범위하게 사용되고 있음에도 불구하고, 의학 연구에 근거중심의학의 프로토콜을 실천하는 데에는 여전히 많은 어려움이 따른다. 의료 문헌 정보, 임상 정보 및 유전체학 정보까지 연구자가 검토해야 할 근거의 양은 방대하며 광범위하다. 또한 의학과 기술의 발전으로 인해 점차 더 빠른 속도로 늘어나고 있기에, 이를 모두 엄밀히 검토하기 위해서는 막대한 양의 시간과 인력이 있어야 한다.

본 논문은 시각적 분석 방법론을 접목하여 의학 연구에서 방대한 과학적 증거를 검색하고 검토할 시 발생하는 막대한 인적 자원의 과부하 문제를 완화하고자 한다. 이를 위하여 근거중심의학의 절차 중 특히 인력 소모가 막심한 절차들을 선정하고, 이러한 난관을 극복하고 보다 효율적이고 효과적으로 데이터에서 유의미한 정보를 도출할 수 있게끔 보조하는 세 가지 시각적 분석 시스템들을 구현하였으며, 각각의 시스템에 관한 디자인 연구를 수행하였다.

우선 첫 디자인 연구에서는 근거중심의학 연구에 있어 필수적 단계인 근거 합성 방법론의 하나인 근거 매핑(Evidence Mapping) 과정을 지원하기 위한 시각적 분석 시스템 PLOEM을 설계했다. 그리고 이를 검증하기 위해 다년간의 근거 기반 의료 경험이 있는 종양학자와 함께 사례 연구를 수행했다. 두 번째 디자인 연구에서는 의학 문헌 검색 시스템의 요구사항 분석을 위해 총 76명의 의사를 상

대로 설문조사를 진행하였고, 이러한 분석을 바탕으로 대화형 시각적 분석 시스템인 EEEVis를 설계했다. 이 시스템은 여러 종의 서지 정보 시각화 인터페이스와 PubTator의 주석 정보를 활용하여 PubMed 검색 엔진의 검색 결과를 증강하는 시스템이며, 이를 평가하기 위해 총 24명의 의사와 함께 사용자 연구를 수행하였다. 이 연구 결과를 바탕으로 의학 문헌 검색 시스템에 대한 설계 지침과 과제를 제시한다. 마지막으로 세 번째 디자인 연구에서는 임의의 유전자군의 유전자 발현 패턴을 주요 암 유형에 따라 시각화하고 분석할 수 있는 시스템인 GeneVis를 설계하였다. 암 유형에 따른 유전자 발현 패턴의 분석과 비교는 암 연구자들이 정밀 의학에서 생체 지표(Biomarker)를 발견하기 위해 빈번히 수행하는 작업이다. 우리는 종양학 전문가 및 유전체학 전문가 총 4인을 대상으로 사례 연구를 진행하였고, 그 결과 GeneVis가 해당 작업을 더 수월하게 수행하는 것과 기존의 데이터에서 새로운 정보를 도출하는 것에 도움이 되었음을 확인하였다.

위의 세 디자인 연구의 결과를 바탕으로, 본 논문은 사용자 분석과 작업 분석을 동반한 시각적 분석 방법론이 의학 연구의 근거 관련 작업의 어려움을 해소하고, 분석 데이터에 대한 보다 나은 이해를 제공하는 것이 가능하다고 결론 내린다.