Computational and Structural Biotechnology Journal 24 (2024) 639–660

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research Article

BioKGrapher: Initial evaluation of automated knowledge graph construction from biomedical literature

Henning Schäfer [a,b], Ahmad Idrissi-Yaghir [b,c], Kamyar Arzideh [d], Hendrik Damm [b,c], Tabea M.G. Pakull [a,b], Cynthia S. Schmidt [a,d], Mikel Bahn [d], Georg Lodde [f], Elisabeth Livingstone [f], Dirk Schadendorf [f], Felix Nensa [d,e], Peter A. Horn [a,1], Christoph M. Friedrich [b,c,*,1]

[a] Institute for Transfusion Medicine, University Hospital Essen, Hufelandstraße 55, Essen, 45147, Germany
[b] Department of Computer Science, University of Applied Sciences and Arts Dortmund (FHDO), Emil-Figge Str. 42, Dortmund, 44227, Germany
[c] Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital Essen, Hufelandstraße 55, Essen, 45147, Germany
[d] Institute for AI in Medicine (IKIM), University Hospital Essen, Girardetstraße 2, Essen, 45131, Germany
[e] Institute of Interventional and Diagnostic Radiology and Neuroradiology, University Hospital Essen, Hufelandstraße 55, Essen, 45147, Germany
[f] Department of Dermatology, University Hospital Essen, Hufelandstraße 55, Essen, 45147, Germany

ARTICLE INFO

Keywords:
Knowledge graph
Named entity recognition
Entity linking
Clinical guidelines
Software

ABSTRACT

Background The growth of biomedical literature presents challenges in extracting and structuring knowledge. Knowledge Graphs (KGs) offer a solution by representing relationships between biomedical entities. However, manual construction of KGs is labor-intensive and time-consuming, highlighting the need for automated methods. This work introduces BioKGrapher, a tool for automatic KG construction using large-scale publication data, with a focus on biomedical concepts related to specific medical conditions. BioKGrapher allows researchers to construct KGs from PubMed IDs.

Methods The BioKGrapher pipeline begins with Named Entity Recognition and Linking (NER+NEL) to extract and normalize biomedical concepts from PubMed, mapping them to the Unified Medical Language System (UMLS). Extracted concepts are weighted and re-ranked using Kullback-Leibler divergence and local frequency balancing. These concepts are then integrated into hierarchical KGs, with relationships formed using terminologies like SNOMED CT and NCIt. Downstream applications include multi-label document classification using Adapter-infused Transformer models.

Results BioKGrapher effectively aligns generated concepts with clinical practice guidelines from the German Guideline Program in Oncology (GGPO), achieving F_1-Scores of up to 0.6. In multi-label classification, Adapter-infused models using a BioKGrapher cancer-specific KG improved micro F_1-Scores by up to 0.89 percentage points over a non-specific KG and 2.16 points over base models across three BERT variants. The drug-disease extraction case study identified indications for Nivolumab and Rituximab.

Conclusion BioKGrapher is a tool for automatic KG construction, aligning with the GGPO and enhancing downstream task performance. It offers a scalable solution for managing biomedical knowledge, with potential applications in literature recommendation, decision support, and drug repurposing.

1. Introduction

In the rapidly evolving field of healthcare and biomedical research, the exponential growth of data presents both opportunities and challenges. The advance of biomedical research methods, increase in clinical trials and the continuous generation of electronic health records have led to an exponential growth in biomedical information [1]. This flood of data contains a broad spectrum of knowledge, from clinical trial results and clinical observations to detailed patient records and results published in scientific journals. Among these resources, the PubMed

* Corresponding author at: Department of Computer Science, University of Applied Sciences and Arts Dortmund (FHDO), Emil-Figge Str. 42, Dortmund, 44227, Germany.
E-mail address: christoph.friedrich@fh-dortmund.de (C.M. Friedrich).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2024.10.017
Received 28 August 2024; Received in revised form 11 October 2024; Accepted 11 October 2024
Available online 17 October 2024

database[2] stands out as a treasure of biomedical knowledge, providing insights into a variety of conditions, treatments, and outcomes. However, the volume and complexity of this information, much of which is only available in unstructured forms, often obscures valuable connections and insights, making it increasingly difficult for researchers and practitioners to navigate this vast and complicated landscape. One possible approach to deal with complexity is the representation of knowledge at the conceptual level through ontologies and knowledge graphs (KGs). At their core, knowledge graphs represent a collection of interlinked descriptions of entities, objects, events, or concepts allowing for a structured way to manage knowledge [2–4].

The manual construction of knowledge graphs is a labor-intensive and time-consuming process, often requiring extensive domain expertise and continuous updates to reflect new discoveries or changes in recent publications. Given the rapid pace of data generation, there is a critical need for more efficient methods of knowledge graph construction [5], considering that the publication rate in the biomedical field is too high for curators to keep up with, making it difficult to find information using manual approaches. [6]. This necessity has led to the exploration and development of automatic knowledge graph construction methodologies [7].

Automatic and semi-automatic knowledge graph construction aims to minimize human intervention by leveraging algorithms and computational techniques to extract, organize, and link data automatically [8,9,5]. This approach involves the processes of data acquisition, entity recognition, relationship extraction, and integration into a graph structure [10]. Advances in machine learning, and particularly in natural language processing (NLP), have opened new ways for extracting meaningful information from unstructured data sources. Once constructed, KGs can support a variety of applications, such as improved information retrieval, semantic search functions and expert systems [11]. In addition, the relationships between entities enable a range of applications, including predictive modeling [12], drug discovery [13], drug repurposing [13], decision support [14], patient stratification [15,16], and precision medicine, and the extraction of genotype-phenotype correlations [17]. A particular example would be their use in molecular tumor boards [18]. Structured knowledge representations in the form of ontologies and KGs also open up important areas in the field of machine learning, where KGs can form a semantic framework for tasks such as relation extraction [19], question answering [20] and improving large language models (LLMs) [21–24]. In the area of recommendation systems using machine learning methods, ontologies and KGs are increasingly being applied [25].

In this study, the BioKGrapher application is introduced and evaluated for constructing knowledge graphs from large-scale publication data, focusing on biomedical concepts that are prevalent, meaning they are widespread and commonly occurring, in relation to specific medical conditions. Six conditions, including Breast Cancer, Colorectal Cancer, Actinic Keratosis and Cutaneous Squamous Cell Carcinoma of the Skin, Malignant Melanoma, and Adult Soft Tissue Sarcomas, have been selected for evaluation purposes based on corresponding evidence-based clinical practice guidelines. Publications are identified using Medical Subject Headings (MeSH) terms, and a Named Entity Recognition and Linking (NER+NEL) pipeline extracts and ranks biomedical concepts from abstracts, weighted against occurrences in the entire PubMed database. The concept rankings are assessed by aligning them with the reference clinical practice guideline concepts. The study also examines the distribution and relationships of concepts across different semantic types within the KGs. The effectiveness of BioKGrapher is demonstrated through two downstream tasks: document multi-label classification and drug repurposing. A cancer-specific KG, derived from 3.6 million neoplasm-related publications, is used to pre-train BioBERT [26], SciBERT [27], and PubMedBERT [28] models by infusing KG triples into the encoder-decoder architectures through a Mixture-of-Partitions [29] approach with Adapter [30] modules, followed by fine-tuning on a multi-label classification task. In the drug repurposing case study, BioKGrapher analyzes publications on two example drugs, re-ranking associated neoplastic processes. Top results are validated against existing literature, highlighting BioKGrapher's potential in a practical application setting. By linking to the Unified Medical Language System (UMLS) [31], several target mappings to knowledge bases such as Systematized Nomenclature of Medicine and Clinical Terms (SNOMED CT) [32], National Cancer Institute Thesaurus (NCIt) or MeSH are available for BioKGrapher, whereby individual hierarchies and relations are generated in each case. The BioKGrapher application code is available online and requires a UMLS license.[3]

The remainder of this work is structured as follows. Section 2 reviews existing methodologies for automatic knowledge graph construction in the biomedical domain, highlighting recent advancements and challenges. Section 3 details the datasets used, including the PubMed corpus and clinical practice guidelines. Section 4 describes the BioKGrapher application pipeline, covering NER+NEL, concept weighting, re-ranking, and the construction of hierarchical knowledge graphs with relation triples. Section 5 presents the outcomes of the approach, including semantic type distribution analysis, concept re-ranking, clinical practice guideline alignment, and performance in multi-label document classification and drug repurposing downstream tasks. Finally, Sections 7 and 8 discuss the findings, state limitations, and suggest future directions.

2. Background

This section reviews the methodologies and advances in automatic KG construction, highlighting the role of NLP and ontology integration.

Initial efforts in biomedical KG construction were largely manual [5]. These early systems were either labor-intensive or limited in scope, but they established foundational ontologies and controlled vocabularies that continue to be integral to modern approaches. Building on this, semi-automated methods with expert curation and clinician-in-the-loop approaches have been increasingly used in recent years. Most modern methods for automatically constructing knowledge graphs rely on NLP techniques [33,34]. When deriving KGs from natural language, NER is typically employed to identify entities, which are then classified (entity typing) and uniquely resolved (entity linking) [35]. After basic knowledge discovery, approaches vary, often using some form of relationship extraction [36] and methods to refine the resulting KG [37].

For this work, when selecting NER+NEL tools and methods, it is important that entities are linked to large concept databases such as UMLS to capture a wide semantic range of concepts related to various conditions across the large amounts of publication abstracts. Tools for NER+NEL, such as MetaMap [38], SemEHR [39], cTAKES [40], CLAMP [41], and ScispaCy [42], have contributed to advancements in this area. MetaMap was an early tool developed to map biomedical text to the UMLS Metathesaurus, but it struggles with spelling variations and ambiguous concepts. SemEHR further advanced the approach by applying manual rules, although the rule-based approach can be time-consuming. cTAKES, CLAMP and ScispaCy offer open-source approaches, with CLAMP being developed with focus to narrative patient reports and ScispaCy providing modern supervised NER approaches but only limited linking capabilities. Additionally, BioPortal [43,44] offers extensive support for a wide range of ontologies, making it a versatile tool for use cases beyond UMLS. The Medical Concept Annotation Toolkit (MedCAT) [45] has emerged with a self-supervised approach to address the limitations of earlier tools, showing recent improved performance in extracting UMLS concepts from various datasets. A critical component of the NER+NEL tools is contextualization, which enhances the understanding and extraction of entities within text. Approaches

[2] https://pubmed.ncbi.nlm.nih.gov/about/ (last accessed: 2024-07-28).

[3] https://github.com/rtg-wispermed/BioKGrapher.

Table 1
Descriptive statistics of the corpora used for constructing and evaluating knowledge graphs. The PubMed annual baseline dataset serves to build global concept frequency weights. These weights are then applied to concepts found in six publication sets related to the conditions: Breast Cancer, Colorectal Cancer, Actinic Keratosis and Cutaneous Squamous Cell Carcinoma of the Skin, Malignant Melanoma and Adult Soft Tissue Sarcomas, collected through MeSH descriptors. Clinical practice guidelines of the same conditions serve as a baseline for the evaluation of the identified prevalent concepts. All six guidelines meet the highest possible level of the AWMF (S3). † denotes that the document was translated with the fairseq Transformer ported model wmt19-de-en, * denotes statistics are based on citations with abstract.

Data Set Descriptor	Citations/ Documents	Citation has Abstract	Token	Token per Document (std)*	Concepts [distinct]	Concepts per Document (std)*	Distinct Concepts per Document (std)*
MEDLINE	37,522,738	26,171,810	6,399,477,076	239.432 (97.874)	1,557,162,536 [150,628]	58.085 (24.948)	36.782 (13.393)
PubMed Baseline 24n							
Breast Cancer MeSH:D001943	344,172	273,194	72,423,131	261.885 (94.614)	18,284,652 [33,283]	65.959 (24.047)	40.379 (12.217)
Colorectal Cancer MeSH:D015179	117,332	102,972	28,623,408	276.152 (89.047)	6,979,430 [21,791]	67.231 (22.183)	41.063 (11.290)
Endometrial Cancer MeSH:D016889	25,326	22,420	6,153,435	272.816 (95.686)	1,523,742 [12,669]	67.471 (23.564)	41.122 (12.032)
AK and SCC of the Skin MeSH:D055623;D002294	146,815	115,902	29,109,540	247.698 (95.575)	7,308,888 [29,284]	62.032 (23.392)	39.338 (12.549)
Malignant Melanoma MeSH:D008545	104,531	78,341	19,080,198	239.594 (96.327)	4,949,795 [23,888]	61.949 (24.397)	39.106 (12.958)
Adult Soft Tissue Sarcomas MeSH:D012509	35,681	20,980	5,205,761	239.567 (105.747)	1,366,514 [16,719]	62.618 (26.704)	40.753 (14.398)
AWMF: 032/045OL v4.4 [58] Breast Cancer	-	-	120,617	-	24,566 [2,302]	-	-
AWMF: 021/007OL v2.1 [59] Colorectal Cancer	-	-	84,182	-	16,343 [1,726]	-	-
AWMF: 032/034OL v2.0 [60] Endometrial Cancer	-	-	135,257	-	20,925 [1,897]	-	-
AWMF: 032/022OL v2.0 [61] AK and SCC of the Skin	-	-	211,265	-	39,566 [2,164]	-	-
AWMF: 032/024OL v3.3 [62] Malignant Melanoma†	-	-	69,847	-	14,045 [1,858]	-	-
AWMF: 032/044OL v1.1 [63] Adult Soft Tissue Sarcomas	-	-	65,849	-	13,663 [1,828]	-	-

like Word2Vec [46], fastText [47], GloVe [48], and more recently, Bidirectional Encoder Representations from Transformers (BERT) [49,50], provide contextual embeddings that capture semantic relationships between words in context, rather than treating them as isolated tokens. These embeddings can be used for architectures such as Bidirectional Long Short-Term Memory networks (Bi-LSTM) [51,52] and Transformers to improve the accuracy and robustness of NER models. While BERT-based models generally excel in NLP tasks, the authors of MedCAT [45] found that the Word2Vec embeddings outperformed Bio_ClinicalBERT [53] on average for the task of linking ambiguous mentions to a large database. This was attributed to the use of pre-trained static BERT embeddings that may not effectively capture context-specific similarities, especially with tokenization challenges in the medical domain, where terminology often includes complex and rare structures. Tokenization is the process of breaking down text into smaller units, such as words or sub-words, that can be processed by models like BERT.

These advancements in NER+NEL tools and methods are also reflected in the construction of KGs within the biomedical domain. Rossanez et al. [54] developed a semi-automatic method using NLP and linking to biomedical ontologies for generating KGs from biomedical texts, particularly focusing on Alzheimer's disease research. It was evaluated against manual extractions by physicians. Their results demonstrated a high efficiency in extracting relevant triples and linking them to ontologies. Health Knowledge Graph Builder (HKGB) [55] is a platform for constructing disease-specific health knowledge graphs from various sources. This platform incorporates a clinician-in-the-loop methodology to leverage medical expertise during the graph construction process to improve the accuracy. The authors applied HKGB for cardiovascular diseases and knee osteoarthritis. Xu et al. [56] focused on enhancing knowledge discovery in the medical field by developing a PubMed Knowledge Graph. While it integrates bio-entities extracted from 29 million PubMed abstracts using BioBERT [26], along with supplementary article metadata such as authors, funding sources, and affiliations, it takes a broader approach to biomedical knowledge integration. In contrast, this work focuses on constructing KGs specific to medical conditions for more targeted applications. Integrating existing medical ontologies into knowledge graphs enhances their applicability in specialized domains. For example, Maghawry et al. [57] developed a framework that integrates the Human Disease Ontology and Symptom Ontology into a knowledge graph. This framework allows for the extraction of specific subgraphs, such as those focused on organ cancer, which can be particularly useful for specialized healthcare applications.

3. Data

This work utilizes two types of data sources: PubMed, for automatic KG construction, and clinical practice guideline that are used for reference alignment and evaluation purposes. The descriptive statistics of the corpora used are summarized in Table 1. The PubMed annual baseline dataset, comprising over 37 million citations, serves to build global concept frequency weights. These weights are then applied to concepts identified in six publication sets related to specific conditions, including Breast Cancer, Cervical Cancer, Colorectal Cancer, Actinic Keratosis and Cutaneous Squamous Cell Carcinoma of the Skin (AK and SCC of the Skin), Malignant Melanoma, and Adult Soft Tissue Sarcomas. These publication sets were collected through MeSH descriptors.

3.1. PubMed

For the biomedical literature foundation of this work, the annual NLM PubMed baseline version of December 14, 2023, is used. Daily update files were included to incorporate citations up to May 2024 (Files 24n0001-24n1424). This corresponds to 37,522,738 citations, of which 26,171,810 citations contain an abstract. Descriptive statistics for the collected PubMed data, including extracted and linked concepts, are presented in Table 1. The PubMed abstracts feature approximately 6.4 billion tokens, with an average of 239 tokens per abstract. It identifies over 1.5 billion concepts, 150,628 of which are distinct. Each condition-specific publication set varies in size, with Breast Cancer having the

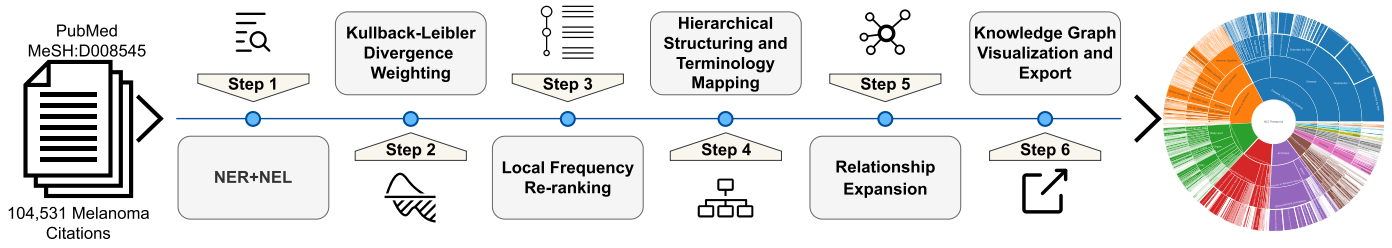


Fig. 1. Steps in the BioKGrapher application pipeline for constructing KGs from PubMed citations. The process begins with NER+NEL, followed by concept weighting using Kullback-Leibler divergence. It continues with re-ranking through local frequency weighting, hierarchical structuring, terminology mapping, and the incorporation of relationships. The final step involves visualization and export options of the constructed KG. In the sunburst chart, different colors represent top-level categories within the hierarchy.

largest number of documents (344,172) and Adult Soft Tissue Sarcomas the fewest (35,681). The token count per abstract and the average number of concepts per abstract are consistent across conditions, with minor variations. The top semantic types across the entire PubMed can be found in Appendix A.3.

3.2. Clinical practice guidelines

Evidence-based clinical practice guidelines from the German Guideline Program in Oncology (GGPO),[4] are used to align and evaluate the automatically constructed concept sets. GGPO is jointly funded by the Association of the Scientific Medical Societies in Germany (AWMF), the German Cancer Society and the German Cancer Aid. They provide clinical practice guidelines in machine-readable JSON formats, which are also used in the German Guideline Program in Oncology NLP Corpus (GGPONC) project [64,65]. From 30 German oncology guidelines, there are six English versions available in machine-readable format. Table 1 shows the available publication sets of conditions along with descriptive statistics on the extracted concepts. The guidelines, all meeting the highest quality level of AWMF (S3), provide a benchmark for evaluating the identified prevalent concepts. These guidelines range from around 65,849 to 211,265 tokens and encompass a substantial number of distinct concepts, averaging around 2,000 distinct concepts per guideline, that describe each condition. The guideline for Malignant Melanoma was translated using the fairseq Transformers port of wmt19de_en [66], a model that has previously demonstrated reasonable performance in translating biomedical and clinical data between English and German [67,68]. The evaluation using these guidelines ensures that the identified concepts are based on the best available research evidence, forming a valid representation of each condition.

4. Methods

The methods are organized into components that collectively describe how user-provided PubMed IDs (PMIDs) are transformed into structured knowledge representations. The “BioKGrapher Pipeline” shows an overview of the workflow, detailing how the tool processes biomedical literature to create KGs that can be explored or exported for further analysis. Next, “Named Entity Recognition and Entity Linking” explains how the application uses NLP to identify and normalize biomedical concepts within the provided document set. Following this, “Prevalent Concept Weights” describes the statistical methods used to weight the identified concepts based on their relevance to the specific document set, ensuring that the important concepts are highlighted. “Frequency Re-ranking” further refines these weights by incorporating local frequency data, addressing the challenge of balancing global and local relevance of concepts. “Hierarchical Structuring and Semantic Triples” covers the process of building the KGs hierarchical structure and defining the relationships between concepts, which are represented as semantic triples. Finally, “Guideline Alignment for Concept Evaluation” outlines the methods used to evaluate the relevance of the generated concepts by comparing them to those found in clinical practice guidelines.

4.1. BioKGrapher pipeline

The proposed pipeline as shown in Fig. 1 for processing uploaded PMIDs, begins with the use of pre-cached concept annotations that are normalized and linked to the UMLS for each citation. Different mentions that refer to the same concept are consistently identified. After normalization, these concepts undergo a weighting process to determine their importance within the specific context of the document set, for example 104k Melanoma PMIDs. This is achieved by calculating the prevalence of each concept within the provided document set relative to its frequency across the entire PubMed database. The weighting is done using Kullback-Leibler divergence (KLD) [69], a statistical measure that quantifies how much one probability distribution diverges from a second, expected probability distribution. In this case, KLD highlights terms that are disproportionately prevalent in the specific document set compared to the general PubMed database, thereby identifying concepts that are particularly relevant to the provided PMID set. To address the challenge of filtering out relevant concepts that are also common across other conditions due to their high global frequency, the pipeline includes a re-ranking process that incorporates local frequency weighting. This step adjusts the importance of concepts by considering their frequency within the local document set, ensuring that globally common terms, which may still be highly relevant in the specific context, are not downweighted or excluded.

Once the top concepts are identified and ranked, the BioKGrapher application proceeds to map these concepts to a standardized biomedical terminology, such as SNOMED CT or NCIt. This mapping is done using a path to root approach, which traces each concept to its hierarchical ancestors within the selected terminology. This approach constructs a hierarchical structure and forms a view on the KG, where each node represents a concept and its place within the broader condition specific biomedical ontology.

To further enrich the KG, BioKGrapher incorporates semantic relationships between concepts as defined in the chosen set of concepts. These relationships are represented as edges connecting the nodes in the graph in the form of semantic triples. By default, BioKGrapher restricts triples to concepts within the chosen set, ensuring that relationships remain consistent with the established ontology. However, this approach also limits the introduction of new concepts into the graph when building the relations.

The final Knowledge Graph can be explored using a variety of interactive visualization tools, such as treemaps[5] or sunburst charts[6] (as

[4] https://www.krebsgesellschaft.de/gcs/german-cancer-society/guidelines.html (last accessed: 2024-07-28).

[5] https://plotly.com/python/treemaps/ (last accessed: 2024-07-28).

[6] https://plotly.com/python/sunburst-charts/ (last accessed: 2024-07-28).

shown in Figs. 1 and 4), which provide a hierarchical view of the concepts related to a condition, with branches colored to represent different top-level categories within the hierarchy.

The software provides sorted tables for each semantic type, allowing users to intuitively navigate the hierarchical relationships and the underlying data. For those who require further analysis or integration with other systems, the KG can also be exported in the Web Ontology Language (OWL) [70] format.

4.2. Named entity recognition and entity linking

MedCAT [45] is used for NER+NEL to identify and normalize biomedical concepts to the UMLS. The process is built upon two core components: the Vocabulary (VCB) and the Concept Database (CDB). The VCB is a comprehensive list of potential words that can appear in the documents, derived from Wikipedia and UMLS terms, and is used for spell-checking. The CDB is a structured repository of biomedical concepts, where each concept is represented by a UMLS Concept Unique Identifier (CUI) and can be referenced by multiple synonymous terms. NER begins with text preprocessing, by using a lightweight spell checker that relies on word frequency and edit distance to correct errors, ensuring that variations in terminology or misspellings still allow accurate concept identification. Spell-checking is performed against the VCB, but corrections are applied only if the corrected word matches a concept in the CDB. This process preserves critical abbreviations.

In the preprocessing step, MedCAT uses SciSpaCy for tokenization and lemmatization. The detection of entity candidates is achieved through a dictionary-based approach along with a moving expanding window technique. Starting with a single word, the algorithm expands the window until a match with a concept in the CDB is found, which allows for identification of biomedical concepts, even when they span multiple words or appear in varied order.

Concept disambiguation is tackled by using a self-supervised learning approach on the Medical Information Mart for Intensive Care (MIMIC-III) [71] database. This involves generating dense vector representations, which capture the semantic meaning of the surrounding text of each concept. When an ambiguous concept is detected, its context embedding is compared to the CDB. If the similarity between these embeddings exceeds a defined threshold, the concept is annotated and linked to the correct CUI.

To efficiently retrieve concepts for any PMID set, BioKGrapher uses an SQLite database. The data is stored in the form `PMID1: ['CUI1', 'CUI2']`, with both an inverted index on CUIs and an index on PMIDs. The indexing allows for quickly aggregating counts of CUIs across the entire PubMed database and enables fast filtering by PMIDs for counting CUIs within subsets of PMIDs for subsequent weighting through KLD.

4.3. Prevalent concept weights

The set of concepts is defined by any UMLS Concept $C = \{c_1, c_2, \ldots, c_n\}$ that appears at least once in PubMed. Let $F(c_i)$ represent the frequency of concept c_i in the PubMed dataset, excluding the documents from the subset and let $f(c_i)$ represent the frequency of the same concept in the subset. The total number of concept occurrences in PubMed is $N = \sum_{i=1}^{n} F(c_i)$, and in the subset is $n = \sum_{i=1}^{n} f(c_i)$. The probability of a concept c_i in the PubMed dataset is $P(c_i) = \frac{F(c_i)}{N}$, and in the subset is $p(c_i) = \frac{f(c_i)}{n}$.

The Kullback-Leibler Divergence for a concept c_i is given by:

$$D_{KL}(c_i) = p(c_i) \cdot \log\left(\frac{p(c_i)}{P(c_i)}\right) \tag{1}$$

The overall divergence measure for all concepts between the subset and the entire dataset is:

$$D_{KL}(\text{Total}) = \sum_{i=1}^{n} p(c_i) \cdot \log\left(\frac{p(c_i)}{P(c_i)}\right) \tag{2}$$

A higher $D_{KL}(c_i)$ indicates that the concept c_i is more prevalent or specific to the subset compared to its general frequency in PubMed. Conversely, a lower $D_{KL}(c_i)$ suggests that the concept c_i is either equally prevalent or less specific to the subset. If a concept c_i does not appear in the subset (i.e., $f(c_i) = 0$), the KLD for that concept is not defined, and a low constant is applied. A justification for choosing KLD is provided in Appendix A.1.

4.4. Frequency re-ranking

To balance the identification of relevant concepts within a subset, a re-ranking process is introduced that integrates local frequency weighting with the initial KLD. This approach addresses the limitation of overlooking relevant but globally common concepts due to their high frequency in the broad PubMed dataset. The final ranking score, $S(c_i)$, for each concept combines normalized KLD and local frequency (LF) scores, applying adjustable weights to balance their influence. The final score is defined as

$$S(c_i) = \alpha \times \widetilde{D_{KL}}(c_i) + \beta \times \widetilde{LF}(c_i),$$

where $\alpha = 0.5$ and $\beta = 0.5$ are the default weights for the normalized KLD and LF scores, respectively. This weighting scheme is designed to account for the uniqueness of concepts within the subset while still considering globally common terms. KLD and LF scores are normalized to a 0 to 1 scale for the final score. The process calculates KLD and LF for all concepts, normalizes them, and computes the final scores using the set weight parameters, and then ranks the concepts, resulting in prevalent terms for that subset.

A threshold t can be set to focus on concepts above a certain score, highlighting the most prevalent concepts that represent a condition.

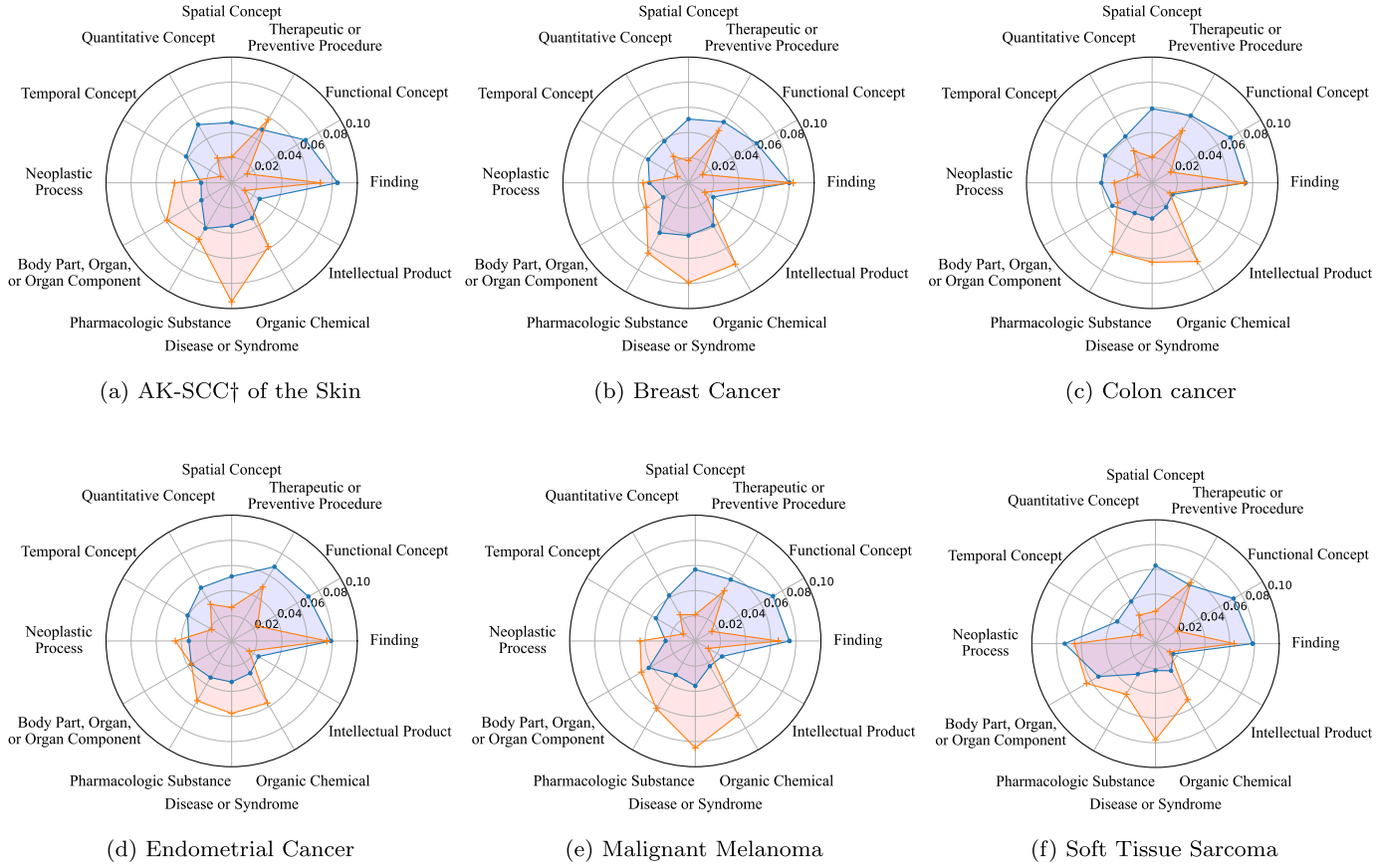
4.5. Hierarchical structuring and semantic triples

To generate the hierarchical structure of the KG, the MRHIER Computable Hierarchies (MRHIER.RRF) file from the UMLS version 2023AB is used. This file contains the hierarchical paths to the root for each atom, represented as a list of AUIs. Nodes for each concept are created along their path to the root, forming the hierarchical structure. The list of prevalent concepts, identified using KLD and re-ranking, guides this process. The resulting hierarchy is dependent on the selected target terminology in the BioKGrapher application, such as SNOMED CT, NCIt, or others.

To establish relationships (triples) within the knowledge graph, the MRREL Related Concepts (MRREL.RRF) file from the same UMLS version is used. This file provides semantic relationships between concepts, represented as subject-predicate-object triples. These triples are applied to the prevalent concepts identified by KLD and re-ranking them to create connections between nodes in the knowledge graph. The specific relationships included also depend on the selected target terminology.

4.6. Guideline alignment for concept evaluation

The relevance of the re-ranked concepts is evaluated by comparing them to concepts extracted from corresponding clinical practice guidelines using the same NER+NEL pipeline. This alignment serves as a benchmark to assess how well the automatically constructed concept sets correspond to those manually curated in guidelines. Precision, Recall, and F_1-Score metrics are calculated to determine the overlap between the top-ranked concepts from the publication sets and the guideline concepts. Precision is defined as the proportion of relevant concepts (those appearing in the guidelines) among the top k concepts from the publication sets. Recall is the proportion of guideline concepts that are also present in the top k publication concepts. The F_1-Score provides a harmonic mean of precision and recall, offering a balanced measure. The top k concepts are retrieved based on the re-ranking algorithm. By



(a) AK-SCC† of the Skin (b) Breast Cancer (c) Colon cancer

(d) Endometrial Cancer (e) Malignant Melanoma (f) Soft Tissue Sarcoma

Fig. 2. Comparison of semantic type distributions in clinical practice guidelines and PubMed publications based on UMLS semantic types. †Actinic keratosis and squamous cell carcinoma.

adjusting k, the metrics show how alignment with guideline concepts changes.

5. Results

This Section is organized as follows. Initially, a descriptive analysis of the distribution of semantic types is presented to show similarities and differences of concepts between literature sourced from PubMed and clinical practice guidelines. Following this, melanoma generated concepts are used to showcase the top k re-ranked concepts across three semantic types. Subsequently, the evaluation focuses on comparing the top k generated and ranked concepts for six medical conditions against those found in their respective clinical practice guidelines. Additionally, visual representations, including a sunburst chart, triples tables for target mappings such as SNOMED CT and NCIt, and a Fruchterman-Reingold [72] layout of relationships illustrated using the melanoma KG, are provided to demonstrate the outputs produced by the BioKGrapher application.

5.1. Semantic type distribution

The radar charts in Fig. 2 show the semantic type distribution overlap between all six guidelines and publication sets. This offers a visual representation of how various medical concepts are emphasized across both sources. The distributions of certain semantic types, such as *Findings (T033)* and *Therapeutic or Preventive Procedures (T061)*, show strong alignment between guidelines and publications, indicating consistency in the emphasis on these concepts across both sources. For example, findings in breast cancer have distributions of 0.080 in guidelines and 0.083 in publications, reflecting close alignment. Similarly, therapeutic procedures in colon cancer show distributions of 0.062 in guidelines and 0.048 in publications, again indicating a consistent focus on treatment modalities.

However, there are notable differences in other categories. The distribution of *Neoplastic Processes (T191)* in melanoma is lower in guidelines, at 0.024, compared to publications, at 0.044. This suggests that publications may place a greater emphasis on detailed aspects of neoplastic processes. This trend is also observed in the distribution of *Diseases or Syndromes (T047)*, where publications show a much higher distribution for melanoma, at 0.085, compared to guidelines, at 0.036. This indicates a stronger focus on disease characterization in publications. The distribution for *Pharmacologic Substances (T121)* is generally higher in publications, such as in melanoma, where it is 0.062 in publications versus 0.031 in guidelines, reflecting the more exploratory nature of academic articles in discussing potential new treatments. Similarly, *Organic Chemicals (T109)* are more prominently represented in publications, as seen in breast cancer where the distribution is 0.075 in publications compared to 0.039 in guidelines.

5.2. Concept re-ranking

Concept re-ranking results for three semantic types in the context of melanoma are presented in Table 2: *Diagnostic Procedure*, *Amino Acid, Peptide, or Protein*, and *Disease or Syndrome*. Rankings are shown as KLD values, frequency counts, and the re-ranking score used to compute the final score for concepts associated with the condition.

In the *Diagnostic Procedure* category, the term `Sentinel Lymph Node Biopsy` has the highest KLD value of 0.0033, indicating a high specificity to melanoma research. Despite a moderate frequency of 2,991 occurrences, it ranks highest after re-ranking with a score of 0.708, underscoring its importance in melanoma diagnostics. Similarly, `Biopsy` exhibits a high KLD value of 0.0021 and a high re-ranking score of 0.477,

Table 2
Top concept ranking before and after KLD re-ranking across different categories for melanoma.

(a) Diagnostic Procedure

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Sentinel Lymph Node Biopsy	0.003334	Magnetic Resonance Imaging	28,081	Sentinel Lymph Node Biopsy	0.708
Biopsy	0.002091	Echocardiography	10,321	Biopsy	0.477
Immunohistochemistry	0.001117	Chest CT	5,496	Immunohistochemistry	0.231
Lymphoscintigraphy	0.000884	Biopsy	4,902	Lymphoscintigraphy	0.180
Examination of skin	0.000530	Sentinel Lymph Node Biopsy	2,991	Examination of skin	0.108
Skin self-examination	0.000381	Biopsy of lymph node	2,375	Skin self-examination	0.078
Excision biopsy	0.000335	Radionuclide Imaging	2,233	Excision biopsy	0.069
Diagnostic Imaging	0.000224	Contrast used	2,088	Fine needle aspiration biopsy	0.039
Fine needle aspiration biopsy	0.000188	X-Ray Computed Tomography	1,896	Contrast used	0.038
PET	0.000180	MRI of abdomen	1,875	PET	0.037
Self-Examination	0.000137	Colonoscopy	1,851	Biopsy of lymph node	0.035
PET/CT scan	0.000133	Radiographic Examination	1,773	Self-Examination	0.030
Fluorescein Angiography	0.000123	CT of abdomen	1,350	Radionuclide Imaging	0.029
Biopsy of skin	0.000123	Ultrasonography	919	PET/CT scan	0.029
Incisional biopsy	0.000102	Palpation	738	Fluorescein Angiography	0.026

(b) Amino Acid, Peptide, or Protein

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Ipilimumab	0.00320	Nivolumab	48,767	Ipilimumab	0.759
Nivolumab	0.00187	Pembrolizumab	18,756	Nivolumab	0.709
Tumor Antigens	0.00137	Ipilimumab	9,553	Pembrolizumab	0.381
Interferons	0.00131	Levothyroxine	4,814	Interferons	0.311
Interleukin-2	0.00127	Interferons	4,013	Tumor Antigens	0.301
Monoclonal Antibodies	0.00126	Troponin	3,221	Interleukin-2	0.284
Pembrolizumab	0.00121	Denosumab	3,095	Monoclonal Antibodies	0.279
Antibodies	0.00104	Antibodies	2,481	Antibodies	0.243
HLA-A2 Antigen	0.00056	Transaminases	1,554	HLA-A2 Antigen	0.126
Growth Factor	0.00055	Atezolizumab	1,514	Growth Factor	0.121
Melphalan	0.00051	Lipase	1,511	Melphalan	0.112
Protein-Serine-Threonine Kinases	0.00043	Fibrin	1,053	Protein-Serine-Threonine Kinases	0.095
Proteins	0.00040	Interleukin-2	917	Proteins	0.090
Cytokine	0.00039	Pregabalin	899	Cytokine	0.085
T-Cell Receptor	0.00036	Insulin	859	T-Cell Receptor	0.080

(c) Disease or Syndrome

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Sentinel node (disorder)	0.00149	Inflammatory disorder	5,565	Lentigo	0.584
Lentigo	0.00144	Hypothyroidism	4,690	Sentinel node (disorder)	0.574
Skin lesion	0.00096	Communicable Diseases	4,231	Skin lesion	0.423
Pigmented skin lesion	0.00076	Pneumonitis	3,856	Pigmented skin lesion	0.295
Vitiligo	0.00041	Pericardial effusion	3,343	Vitiligo	0.191
Melanosis	0.00038	Uveoparotid Fever	3,269	Hyperkeratosis	0.155
Hyperkeratosis	0.00024	Lymphadenopathy	3,240	Melanosis	0.147
Hypopigmentation disorder	0.00021	Obesity	2,922	Lymphadenopathy	0.113
Recurrent disease	0.00018	Hyperthyroidism	2,889	Pneumonitis	0.112
Local disease	0.00018	Colitis	2,758	Colitis	0.107
Glaucoma	0.00017	Hepatitis	2,414	Uveoparotid Fever	0.097
Hypophysitis	0.00013	Lymphocytopenia	2,355	Hypophysitis	0.089
Dermatologic disorders	0.00013	Hyperkeratosis	2,260	Hypopigmentation disorder	0.087
Actinic porokeratosis	0.00010	Skin lesion	2,053	Lymphocytopenia	0.073
Secondary glaucoma	0.00010	Thyroiditis	1,631	Recurrent disease	0.072

emphasizing its importance in the melanoma context, while also having a high frequency of 4,902 occurrences. In contrast, terms like `Magnetic Resonance Imaging` and `Echocardiography`, which appear frequently in melanoma publications with 28,081 and 10,321 occurrences respectively, have low KLD values and do not make it into the top 16 KLD melanoma concepts. This suggests that these procedures are more commonly referenced across various medical fields and are not uniquely tied to melanoma, resulting in low re-ranking positions. Other notable diagnostic procedures with high KLD values, lower occurrences, and high final re-ranking positions include `Immunohistochemistry`, `Lymphoscintigraphy`, `Examination of skin`, and `Excision biopsy`, indicating their specific association with melanoma.

In the *Amino Acid, Peptide, or Protein* category, the drugs `Ipilimumab` and `Nivolumab` have the highest KLD values of 0.0032 and 0.0019, while also having high local frequencies of 9,553 and 48,767 occurrences, respectively. These terms are therefore top-ranked after re-ranking with scores of 0.759 and 0.709, emphasizing their critical role in melanoma treatment. Other terms, such as `Pembrolizumab`, `Interferons`, and `Tumor Antigens`, also achieved high KLD values and high re-ranking positions, indicating their specific relevance to melanoma. Despite lower frequency counts of 18,756 for `Pembrolizumab` and 4,013 for `Interferons` these terms end up with high final re-ranking scores, 0.381 for `Pembrolizumab` and 0.311 for `Interferons`.

Conversely, terms like `Levothyroxine` and `Troponin`, which appear to be frequent in general medical literature, show low KLD values. These terms are classified as less specific to melanoma, as indicated by their absence from the top re-ranked positions.

In the *Disease or Syndrome* category, `Sentinel node (disorder)` shows a high KLD value of 0.0015, highlighting its specificity for this category. Although it has a moderate frequency of 5,565 occurrences, it is top-ranked after re-ranking with a score of 0.574. `Lentigo` similarly shows a high KLD score of 0.0014 and a high re-ranking score of 0.584 despite a lower frequency count.

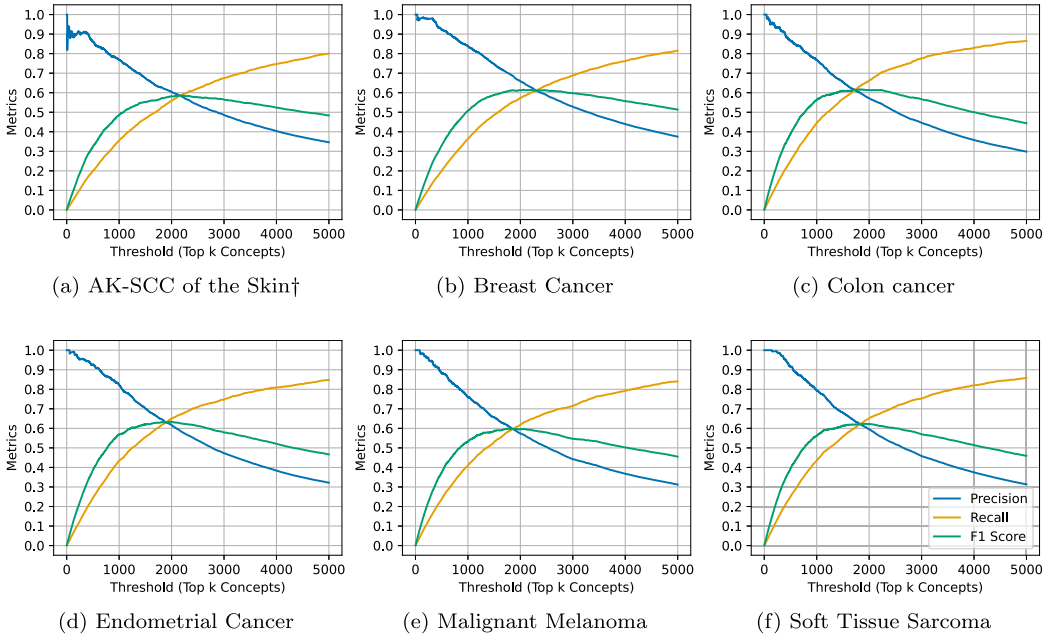


Fig. 3. Precision, recall, and F_1-Score for different thresholds and across multiple conditions. The figure shows the performance of the re-ranking algorithm that balances local frequency (LF) and Kullback-Leibler divergence (KLD) to rank prevalent concepts. Parameter weights: $\alpha = 0.5$, $\beta = 0.5$. †Actinic keratosis and squamous cell carcinoma of the skin.

Other notable concepts with high KLD values and high re-ranking positions include `Skin lesion` and `Pigmented skin lesion`. In contrast, terms like the generic `Inflammatory disorder` and `Hypothyroidism`, which are frequently mentioned in general medical literature with 5,565 and 4,690 occurrences respectively, show lower KLD values. These terms appear to be frequent in other medical fields as well. Immunotherapy-induced undesirable inflammations such as `Hypophysitis`, `Pneumonitis`, and `Colitis` are brought into an important position by the final ranking of the *Disease or Syndrome*-related concepts.

The ranking tables for the same semantic types for other conditions can be found in Appendix A.4.

5.3. Evaluation of reference guideline alignment

Fig. 3 shows Precision, Recall, and F_1-Score metrics for guideline concept overlap at different concept retrieval thresholds from publication sets across multiple conditions (a–f). Top k concepts are retrieved based on the re-ranking algorithm to balance term frequency and KLD. The results show how the metrics change at different concept retrieval amounts. The points where the Precision and Recall curves intersect represent the optimal threshold values for each condition based on the reference guideline alignment. The F_1-Score provides a single score to evaluate the performance, though higher Recall might be preferred at the cost of Precision when evaluating KGs. It can be seen across all conditions that the concepts with the highest ranking often appear within the guidelines from the start. In the conditions *Actinic keratosis and squamous cell carcinoma of the skin*, *Breast Cancer*, and *Soft Tissue Sarcoma*, some of the first concepts show a constant alignment with the guideline, resulting in a short horizontal precision curve before this continuously decreases. The F_1-Scores peak at around 0.6 when approximately 2,000 concepts are retrieved, suggesting that this threshold offers an optimal balance between Precision and Recall for aligning with reference guidelines.

5.4. Hierarchy and triples for knowledge graph construction

For all six conditions, the resulting triples for various specialized target terminologies and classification systems were analyzed, focusing on purpose and domain. The results of the different target terminologies are shown in Table 3. The hierarchical order of the constructed KG at the example of melanoma can be seen in Fig. 4, where *Pharmacologic Substances* are shown. An additional example of *Colon Cancer Operative Surgical Procedures* can be seen in Appendix A.2.

SNOMED CT, covering a comprehensive range of clinical terms, exhibited the highest density of concepts and relationships. *Actinic keratosis and squamous cell carcinoma (AK-SCC) of the Skin* demonstrated the largest dataset with 3,975 concepts and 19,114 relationships. However, *Breast Cancer* showed a high count of orphan nodes without any relationships, totaling 1,975. The NCIt, dedicated to cancer-related clinical and research terms, also showed substantial coverage but with fewer concepts than SNOMED CT. *Soft Tissue Sarcoma* had the highest number of concepts (2,291) and relationships (9,086) within this dataset. *Breast Cancer* had the most orphan nodes, with 1,310 instances, indicating potential gaps in connectivity. LOINC, focusing on laboratory and clinical observations, presented fewer concepts and relationships. *Malignant Melanoma* had the highest figures with 708 concepts and 752 relationships. Orphan nodes were most prevalent in *Soft Tissue Sarcoma*, totaling 400, reflecting potential isolation in specific test-related data. `Foundational Model of Anatomy (FMA)`, which catalogs human anatomical structures, showed lower overall coverage and a higher number of orphan nodes. This indicates that while FMA provides detailed anatomical concepts, these concepts are less interconnected within the KG, resulting in a sparser network. *Soft Tissue Sarcoma* had the highest number of concepts (518) and relationships (280). Orphan nodes were notably high in *Soft Tissue Sarcoma* at 420. `Online Mendelian Inheritance in Man (OMIM)`, cataloging genetic disorders, included fewer concepts but demonstrated high connectivity in diseases with genetic implications, like *Melanoma*, which had 193 concepts and 668 relationships. Orphan nodes were minimal in OMIM, indicating a highly interconnected network of genetic data. This high degree of connectivity reflects the close relationships among genetic concepts within this terminology.

`RxNorm`, focusing on clinical drugs and having a flatter hierarchical structure compared to other terminologies, had very limited data with no depth values recorded, meaning that the concepts did not form hierarchical relationships within the KG. *Breast Cancer* featured the highest

Table 3
Summary of constructed knowledge graph triples statistics for different vocabularies. † Orphan nodes are concepts that have no connections within the graph in addition to their hierarchical classification through `isa` relationships.

Cancer Type	Concepts+PTR	Mean Node Depth (std)	Relationships	Nodes with Relations	Orphan Nodes†
Systematized Nomenclature of Medicine and Clinical Terms (SNOMED CT)					
AK-SCC of the Skin	3,975	7.673 (3.311)	19,114	2,180	1,795
Breast Cancer	3,821	6.875 (2.556)	12,254	1,846	1,975
Colon Cancer	3,705	7.036 (2.914)	12,658	1,720	1,985
Endometrial Cancer	3,744	7.289 (3.355)	14,102	1,842	1,902
Malignant Melanoma	3,763	7.030 (2.910)	14,416	1,913	1,850
Soft Tissue Sarcoma	3,799	7.648 (3.343)	15,046	1,909	1,890
National Cancer Institute Thesaurus (NCIt)					
AK-SCC of the Skin	2,203	6.900 (2.121)	7,732	1,071	1,132
Breast Cancer	2,049	6.676 (2.032)	3,814	739	1,310
Colon Cancer	1,948	6.500 (1.990)	3,510	681	1,267
Endometrial Cancer	1,936	6.888 (2.188)	5,604	779	1,157
Malignant Melanoma	2,054	6.560 (1.948)	4,720	885	1,169
Soft Tissue Sarcoma	2,291	7.117 (2.216)	9,086	1,221	1,070
Logical Observation Identifier Names and Codes (LOINC)					
AK-SCC of the Skin	613	2.008 (3.393)	558	235	378
Breast Cancer	643	2.431 (3.657)	658	278	365
Colon Cancer	697	2.197 (3.497)	664	282	415
Endometrial Cancer	636	2.485 (3.671)	672	282	354
Malignant Melanoma	708	2.483 (3.584)	752	313	395
Soft Tissue Sarcoma	643	2.004 (3.347)	556	243	400
Foundational Model of Anatomy (FMA)					
AK-SCC of the Skin	535	11.0 (2.562)	310	117	418
Breast Cancer	252	10.766 (3.065)	34	19	233
Colon Cancer	287	11.117 (3.162)	196	51	236
Endometrial Cancer	325	10.614 (2.751)	180	66	259
Malignant Melanoma	445	10.466 (2.581)	130	71	374
Soft Tissue Sarcoma	518	10.812 (2.654)	280	98	420
Online Mendelian Inheritance in Man (OMIM)					
AK-SCC of the Skin	178	3.174 (1.180)	540	166	12
Breast Cancer	134	3.202 (1.180)	436	129	5
Colon Cancer	138	3.035 (1.363)	384	127	11
Endometrial Cancer	139	3.093 (1.212)	492	134	5
Melanoma	193	3.378 (0.996)	668	188	5
Soft Tissue Sarcoma	188	2.818 (1.340)	600	168	20
Drugs in RxNorm					
AK-SCC of the Skin	87	-	6	6	81
Breast Cancer	152	-	22	20	132
Colon Cancer	118	-	16	12	106
Endometrial Cancer	120	-	24	21	99
Malignant Melanoma	82	-	10	9	73
Soft Tissue Sarcoma	96	-	14	10	86

number of concepts (152) and relationships (22). Orphan nodes, indicating unlinked drugs, were predominant in *Breast Cancer* with 132 nodes.

The Fruchterman-Reingold [72] layout was applied to visualize the relationships between biomedical concepts extracted from PubMed publications related to melanoma, as generated by BioKGrapher. The resulting KG includes a diverse set of concepts categorized under different semantic types, such as *Diagnostic Procedures*, *Therapeutic Interventions*, and *Pharmacologic Substances*. The relationships in Fig. 5 were mapped using SNOMED CT and reveal clusters of related concepts.

The zoomed-in subgraphs highlight specific clusters focused on diagnostic methods like `Shave biopsy` and `Light Microscopy`, and key therapeutic strategies such as `Immune checkpoint inhibitor therapy`. The visualization allows exploration of concept connections and shows the complexity and depth of relationships within the melanoma KG. The chosen layout helps to represent these relationships as clusters to identify important concept groupings.

6. Downstream applications

Two downstream tasks are evaluated to further evaluate the effectiveness of the proposed BioKGrapher approach. The first task focuses on document multi-label classification using Adapter infusion, while the second task presents a small case study on drug repurposing. These evaluations are conducted in addition to the prior alignment analysis on clinical practice guidelines and descriptive statistical results.

6.1. Multi-label document classification

Recent approaches have focused on leveraging and embedding knowledge from KGs for downstream tasks. One of them is Mixture-of-Partitions (MoP) [29], an infusion approach that can handle large KGs by partitioning it into smaller subgraphs and infusing their knowledge into BERT style models using Adapters [73,30]. Adapters are new initialized modules inserted between the layers of a pre-trained Transformer. For each sub-graph G_k, the tail entity in each triple $(h, r, t) \in G_k$ is removed, and the remaining elements are transformed into a sequence of tokens: `[CLS] h [SEP] r [SEP]`. The sub-graph-specific Adapter is then trained to predict the tail entity using the representation of the `[CLS]` token. The parameters Φ_{G_k} are optimized by minimising the cross-entropy loss [74]. This allows information from the knowledge graph to be embedded in the models, which has led to improved performance on multiple biomedical downstream tasks (Natural Language

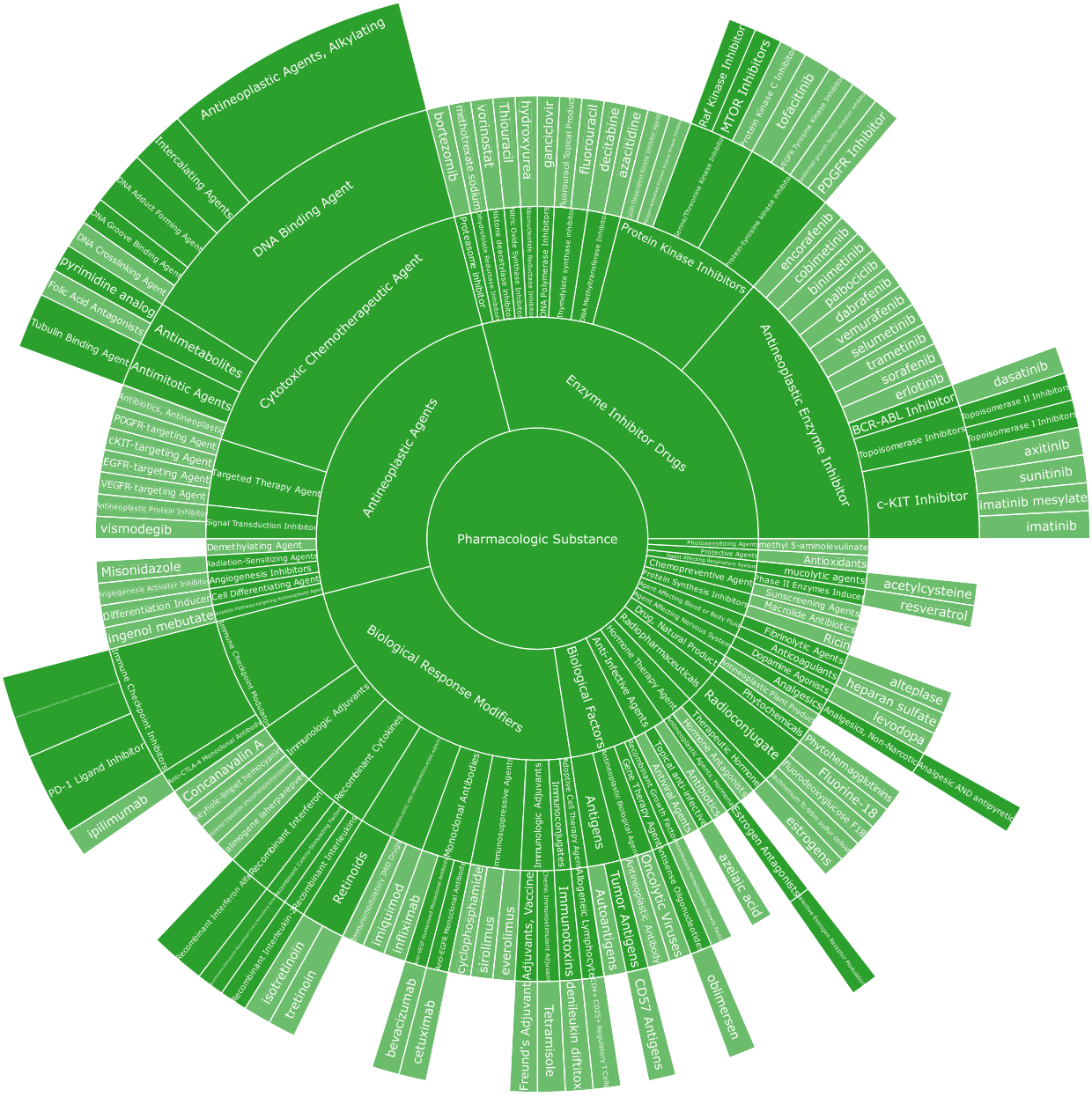


Fig. 4. Hierarchical structure automatically generated from melanoma publications (hierarchy classification by NCIt) with a re-ranking threshold of 2,000 concepts. The figure shows the *Pharmacologic Substances* subcategory visualized in an interactive sunburst chart, displaying a maximum of 5 depths at a time.

Inference (NLI), Question Answering (QA) and Classification) when applying a large SNOMED CT-based knowledge graph [29].

For this evaluation, a large cancer KG is generated using the BioKGrapher application. All publications in which MeSH terms appear under the tree number C04, which represents neoplasms, are used to select cancer related publications. This resulted in 3,596,405 PMIDs as input for the application. The top 15,000 concepts, representing about 20% of all possible concepts in the publication set, were re-ranked and used to build a large cancer relation network using intra-concept triples from NCIt. This network, represented as triples, was then used for pre-training Adapter infusion on three models: BioBERT [26], SciBERT [27], and BiomedBERT (former PubMedBERT) [28] for comparison purpose. The graph was partitioned into five sub-graphs using METIS [75] and pre-training was performed for one epoch using the publicly available implementation from [29]. The resulting configuration used for the infusion is presented in Table 5. In this setup, SFull and S20Rel, which were utilized in prior research, are compared to BioKGrapher Neoplasms, the cancer-specific KG designed to better align with the HoC data domain.

The effectiveness of the infusion is evaluated using a downstream multi-label classification task on the Hallmarks of Cancer (HoC) corpus [76–78]. The corpus is derived from 1,852 PubMed abstracts. In this corpus, class labels are manually annotated by experts according to the HoC taxonomy [79,80]. While the taxonomy comprises 37 classes, organized hierarchically, only the ten top-level classes are used for consistency with previous research. Following the MoP experiments [29], the publicly available train/dev/test splits provided by the Biomedical Language Understanding and Reasoning Benchmark (BLURB) [28] are used. Results are reported as the average micro F_1-Score performance across five runs with standard deviation.

The performance on the multi-label classification downstream task on the HoC corpus can be seen in Table 4. The results show effectiveness of using the BioKGrapher Neoplasms KG for Adapter infusion. The MoP BioKGrapher Neoplasms variant increased the micro F_1-Score across all three models SciBERT, BioBERT, and BiomedBERT compared to the previously reported results for MoP (SFull) and MoP (S20Rel). Notably, the highest score on HoC was observed with BiomedBERT, where the

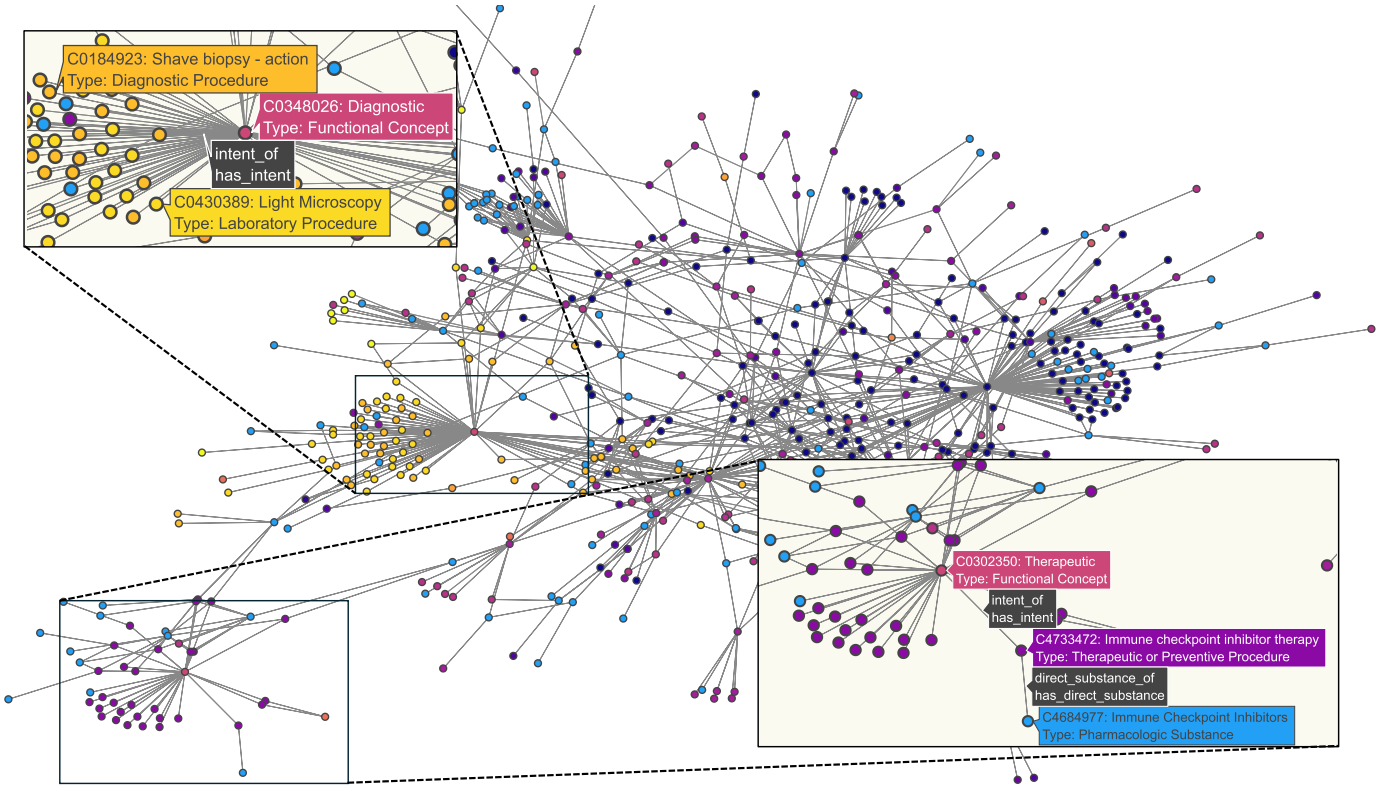


Fig. 5. Fruchterman-Reingold layout of concepts for melanoma, with `isa` relationships omitted. The edges represent bidirectional relationships, i.e., all concepts that have a relation in SNOMED CT or NCIt within a re-ranking threshold of the top 2,000 concepts.

Table 4
Micro F_1-score HoC multi-label classification results grouped by model and knowledge graph adapter infusion variants averaged across 5 runs with standard deviation. † denotes results from MoP authors [29].

Model	F_1-Score (%)
SciBERT†	80.5 $_{\pm 0.6}$
MoP (SFull)†	81.5 $_{\pm 0.4}$
MoP (S20Rel)†	81.8 $_{\pm 0.7}$
MoP (BioKGrapher Neoplasms)	82.7 $_{\pm 0.4}$
BioBERT†	81.4 $_{\pm 0.6}$
MoP (SFull)†	81.5 $_{\pm 0.9}$
MoP (S20Rel)†	82.5 $_{\pm 1.1}$
MoP (BioKGrapher Neoplasms)	82.9 $_{\pm 0.6}$
BiomedBERT (PubMedBERT)†	82.3 $_{\pm 0.5}$
MoP (SFull)†	82.7 $_{\pm 0.6}$
MoP (S20Rel)†	83.3 $_{\pm 0.3}$
MoP (BioKGrapher Neoplasms)	**84.1** $_{\pm 0.3}$

Table 5
Statistics of the knowledge graphs by MoP authors in comparison to BioKGrapher neoplasms knowledge graph.

	Entities	Relations	Triples
SFull	302,332	229	4,129,726
S20Rel	263,808	20	1,750,677
BioKGrapher Neoplasms	15,000	27	68,995

F_1-Score increased by 0.85 percentage points over MoP (S20Rel). Despite the smaller size of the BioKGrapher Neoplasms KG (68,995 triples) relative to SFull and S20Rel, its tailored cancer-specific knowledge appears to enhance on the cancer specific HoC corpus, leading to improved downstream performance.

6.2. Extracting known drug uses from the literature

Drug repurposing involves finding new therapeutic uses for existing drugs [81]. The semantic type *Amino Acid, Peptide, or Protein* (T116) as defined by the UMLS Semantic Network includes a wide range of biological molecules, such as amino acids, peptides, proteins, and protein-based therapeutics like monoclonal antibodies. Table 2 can give an impression of commonly used therapies associated with a disease that fall under this semantic type. Conversely, *Neoplastic Processes* associated with a drug can be determined on the basis of all publications in which the drug appears. In this small downstream case study, the aim is to retrieve already known indications using the BioKGrapher application on the example of `Nivolumab` (MeSH:D000077594) and `Rituximab` (MeSH:D000069283). This approach could be extended to identify novel associations for potential drug repurposing. By analyzing the ranked associations, researchers can explore ABC relations [82] and identify possible applications of existing drugs beyond their current indications.

For `Nivolumab` and `Rituximab`, two large publication sets in which the drugs appear are collected and entered into the BioKGrapher application as a list of PMIDs. Subsequently, re-ranking is performed, and the results are sorted in descending order for the semantic type *Neoplastic Process* (T191) to find strong associated conditions. From the ranking results, the top neoplasms associated with each drug are extracted and manually tested against evidence in literature.

The results of this small case study as shown in Table 6 provide clear evidence supporting the use of `Nivolumab` and `Rituximab` for various neoplastic conditions beyond their main indications.

For `Nivolumab`, the re-ranking results strongly associate it with *melanoma* with a re-ranking score of 0.963, which aligns with its known use as a treatment for advanced melanoma. Additionally, strong associations are found with *non-small cell lung carcinoma* with a re-ranking score of 0.711 and *metastatic renal cell carcinoma* with a re-ranking score of 0.324, which are also consistent with established clinical applications [84,85].

Table 6
Relevance scores for Nivolumab and Rituximab for top associated neoplasms (neoplastic process).

(a) Nivolumab

Condition	Re-rank↓	Evidence
Melanoma	0.963	[83]
Non-Small Cell Lung Carcinoma	0.711	[84]
Metastatic Renal Cell Carcinoma	0.324	[85]
Clear Cell Renal Cell Carcinoma	0.173	[86]
Renal Cell Carcinoma	0.155	[87]
Renal Carcinoma	0.153	[87]
Malignant Neoplasm Of Stomach	0.104	[88]
SCC Of The Head And Neck	0.104	[89]
Secondary Neoplasm	0.100	-
Metastatic Malignant Neoplasm To Brain	0.093	[90]
Hodgkin Disease	0.078	[91]
Cutaneous Melanoma	0.060	[83,92]
Liver Carcinoma	0.060	[93]
Secondary Malignant Neoplasm Of Liver	0.058	-
Adenocarcinoma Of Lung (Disorder)	0.057	[94]
Malignant Head And Neck Neoplasm	0.052	[89]
Classical Hodgkin's Lymphoma	0.048	[91]
Adenocarcinoma	0.046	[94]
Secondary Malignant Neoplasm Of Lung	0.045	-

(b) Rituximab

Condition	Re-rank↓	Evidence
THRLBCL	1.000	[95]
Lymphoma, Non-Hodgkin	0.691	[96]
Lymphoma, Follicular	0.496	[97]
Chronic Lymphocytic Leukemia	0.369	[98]
B-Cell Lymphomas	0.362	[99]
Mantle Cell Lymphoma	0.236	[100]
Lymphoproliferative Disorders	0.225	[101]
Marginal Zone B-Cell Lymphoma	0.127	[102]
Waldenstrom Macroglobulinemia	0.102	[103]
Residual Tumor	0.071	[104,105]
Microglioma	0.062	[106]
Burkitt-Like Lymphoma	0.049	[107]
Hodgkin Disease	0.047	[108]
Small Lymphocytic Lymphoma	0.046	[109]
Hematologic Neoplasms	0.034	[110]
Hairy Cell Leukemia	0.034	[111]
Monoclonal Gammapathies	0.033	[112]
Mediastinal large B-cell lymphoma	0.029	[113]
Primary Cutaneous B-Cell Lymphoma	0.029	[114]

In the case of `Rituximab`, the top-ranked condition is *THRLBCL* (*T-cell/histiocyte-rich large B-cell lymphoma*) with a perfect re-ranking score of 1.000, indicating a strong association [95]. Other highly relevant neoplasms include *non-Hodgkin lymphoma* with a re-ranking score of 0.691 and *follicular lymphoma* with a re-ranking score of 0.496, which are well-documented indications for `Rituximab` [96,97].

7. Discussion

The semantic type distributions between clinical practice guidelines and biomedical publications reveal both alignments and differences. For example, categories such as *Findings* and *Therapeutic Procedures* show a high degree of consistency, indicating similar priorities in both sources. However, differences are noticeable in areas such as *Neoplastic Processes* and *Pharmacologic Substances*, where publications tend to focus more on exploratory topics that may not yet be fully reflected in clinical practice guidelines. This divergence could highlight the different roles of research literature and clinical practice guidelines where the former may explore new frontiers, and the latter consolidates established knowledge. The translation from literature to clinical practice also reflects into the unrestricted selection of literature and therefore impacts the alignment evaluation.

The re-ranking process effectively prioritizes concepts specific to certain conditions, such as melanoma and breast cancer. For instance, concepts like `Sentinel Lymph Node Biopsy` and `Ipilimumab` are ranked highly in their respective categories, reflecting their relevance in the diagnosis and treatment of melanoma. This indicates that the re-ranking mechanism, which balances local frequency and KLD, successfully identifies key concepts that are prevalent and frequent in the subset. However, common concepts with high global frequency, such as `Magnetic Resonance Imaging`, are down-weighted unless they have specific relevance to the condition, which underscores the importance of context-specific ranking parameters in KG construction. As the weighting of KLD and LF can be adjusted via the parameters, the final re-ranking of the BioKGrapher concepts can be fine-tuned to emphasize either specificity to the condition (as indicated by higher KLD values) or prevalence in the literature (as indicated by higher LF counts). For instance, in the *Diagnostic Procedure* category of melanoma, increasing the weight of KLD would push terms like `Sentinel Lymph Node Biopsy` and `Biopsy` even higher in the rankings, due to their strong association with melanoma. Conversely, prioritizing LF might elevate more commonly referenced procedures such as `Magnetic Resonance Imaging` and `Echocardiography`, which are frequently mentioned across various medical fields. The focus during re-ranking may vary depending on the intended application of the KG and the user. Therefore, the weighting is made adjustable in the software, allowing for modification from the default values of 0.5 for both frequency and KLD.

The evaluation metrics demonstrate the effectiveness of the re-ranking algorithm in aligning automatically extracted concepts with those in clinical practice guidelines. The precision curves start high, indicating that the top-ranked concepts are highly relevant to the guidelines. However, precision generally decreases as more concepts are included, reflecting the broader scope of biomedical publications compared to the more focused nature of clinical practice guidelines. The F_1-Score peaks suggest that a balance between precision and recall can be optimized at specific thresholds, providing practical guidance for KG construction with a focus on conditions.

The analysis of constructed KGs across various vocabularies reveals differences in node depth, relationship density, and the presence of orphan nodes. These variations reflect the different focuses of each vocabulary, with SNOMED CT offering broader clinical coverage, while NCIt is more specialized in oncology. The presence of orphan nodes, particularly in more focused ontologies like RxNorm and FMA, suggests potential gaps in connectivity that could impact the usability of the graphs.

The resulting KGs are versatile, enabling a range of potential use cases in intelligent medicine and healthcare. One possible key application is in clinical decision support systems, where KGs can assist healthcare providers in diagnosing conditions and suggesting evidence-based treatments that may align with the conceptual view on recent literature. Due to the nature of the proposed KLD re-ranking approach, the KGs not only represent strong associations but also strong disassociations, which could be employed for abnormality checks, flagging unusual or inconsistent data in patient records. For instance, if symptoms and treatments for a patient do not typically align, the KG could alert the clinician to this anomaly, prompting further investigation.

7.1. Limitations

The appropriate evaluation of ontologies and KGs across multiple conditions is difficult to implement. In this work, it was decided to leverage extracted concepts from consensus-based clinical practice guidelines as reference terminology. In addition to advantages such as a large data basis for a variety of conditions, this also has some disadvantages. For example, it does not achieve the evaluation quality that would be possible with a manual evaluation by domain experts and may lag behind the latest scientific discoveries due to the time required for validation and consensus-building. The initial extraction of concepts is based on

the same extraction pipeline that produces references of silver-standard quality, which therefore may contain biases in entity extraction and linking that leads to degradation or improvement of the reported results. Until now, the relations have been obtained from existing terminologies and classification systems such as SNOMED CT or NCIt, which on the one hand guarantees correctness, but on the other hand means that no novel relations, for example from the latest publications, can be added to the KG. This principle also pertains to concepts. However, because BioKGrapher is open-source, users can enhance or replace UMLS MRREL relationships or concepts with NLP-mined relations or other data sources to capture novel concepts and emerging relations from the latest literature. It is also possible to recognize new concepts by separating the entity recognition process from the entity linking process during indexing.

8. Conclusion

This study introduced BioKGrapher, a tool for automatically constructing biomedical KGs from unstructured text sources, such as PubMed abstracts. The evaluation of BioKGrapher shows that it effectively identifies and ranks biomedical concepts, achieving a high degree of overlap with guideline-recommended terms. Across multiple conditions, the tool maintains F_1-Scores around 0.6 when retrieving the top 2,000 concepts, balancing precision and recall. Semantic differences between publications and guidelines indicate that publications may focus on different aspects compared to guidelines. Overall, BioKGrapher proves to be an effective tool for generating and applying KGs in biomedical research, showing particular promise in enhancing information retrieval and improving downstream applications. The level of concept overlap observed in the guideline literature suggests that restricting the literature based on common characteristics found in guidelines is not needed. However, future work could explore using bibliometrics such as study phase, level of evidence, and publication recency to enhance the initial publication selection, potentially improving the alignment of biomedical concepts in the KG compared to the guidelines. BioKGrapher allows for the adjustment of metadata and bibliometrics when providing custom PMID lists. By using a list of only the most recent PMIDs on a given topic, one can construct a KG that captures recent trends in the field. Regularly updating guidelines with bibliometric criteria could enhance the precision of predicting which concepts should be included in future guideline revisions. Beyond traditional structure-based KG evaluations, task-based methods now enhance applications like information retrieval, LLM integration, and machine learning. Combining graph neural networks with RAG for tasks like question answering is a promising direction, and future work could improve relation extraction from texts to capture new insights.

CRediT authorship contribution statement

Henning Schäfer: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Ahmad Idrissi-Yaghir:** Writing – review & editing, Visualization, Software. **Kamyar Arzideh:** Writing – review & editing, Visualization, Software. **Hendrik Damm:** Writing – review & editing, Data curation. **Tabea M.G. Pakull:** Writing – review & editing, Data curation. **Cynthia S. Schmidt:** Writing – review & editing, Validation. **Mikel Bahn:** Writing – review & editing, Validation. **Georg Lodde:** Writing – review & editing, Software, Investigation. **Elisabeth Livingstone:** Writing – review & editing, Validation, Investigation. **Dirk Schadendorf:** Supervision, Resources. **Felix Nensa:** Supervision, Resources. **Peter A. Horn:** Writing – review & editing, Supervision, Conceptualization. **Christoph M. Friedrich:** Writing – review & editing, Supervision, Conceptualization.

Table A.7
Table of TUIs, frequencies, and definitions.

TUI	Occurrence	Label
T080	170,038,624	Qualitative Concept
T169	166,162,256	Functional Concept
T033	96,556,917	Finding
T081	88,006,624	Quantitative Concept
T082	67,416,124	Spatial Concept
T079	59,801,115	Temporal Concept
T073	47,342,685	Manufactured Object
T061	44,965,961	Therapeutic or Preventive Procedure
T047	41,035,310	Disease or Syndrome
T121	40,029,182	Pharmacologic Substance

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used DeepL Write and Grammarly in order to enhance language quality, grammar, and style. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors have no relevant interests to disclose.

Acknowledgements

The work of Henning Schäfer, Ahmad Idrissi-Yaghir, Hendrik Damm, Tabea M. G. Pakull and Mikel Bahn was funded by a PhD grant from the DFG Research Training Group 2535 *Knowledge- and data-based personalisation of medicine at the point of care (WisPerMed)*. The graphical abstract was created with BioRender.

We also acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen.

Appendix A

This appendix includes additional tables and figure examples to support the main content (Figs. A.6, A.7 and Tables A.7–A.12).

A.1. Justification for KLD

Kullback-Leibler Divergence (KLD) is used to compare concept distributions between two corpora, making it suitable for identifying and weighting biomedical concepts that are more prevalent in a condition-specific subset of literature compared to the general PubMed corpus. Unlike TF-IDF, which focuses on term importance within individual documents relative to a single corpus, or BM25, which ranks documents based on query relevance, KLD measures the divergence between two probability distributions, allowing for a direct comparison of concept frequencies between the two corpora. This made KLD an appropriate choice for the task of highlighting how specific concepts are disproportionately represented in the condition-specific literature relative to the general biomedical corpus.

A.2. Knowledge graphs

This subsection presents additional constructed KGs as examples.

A.3. Semantic type distribution

This subsection shows the distribution of the 10 most frequent concept classes in PubMed.

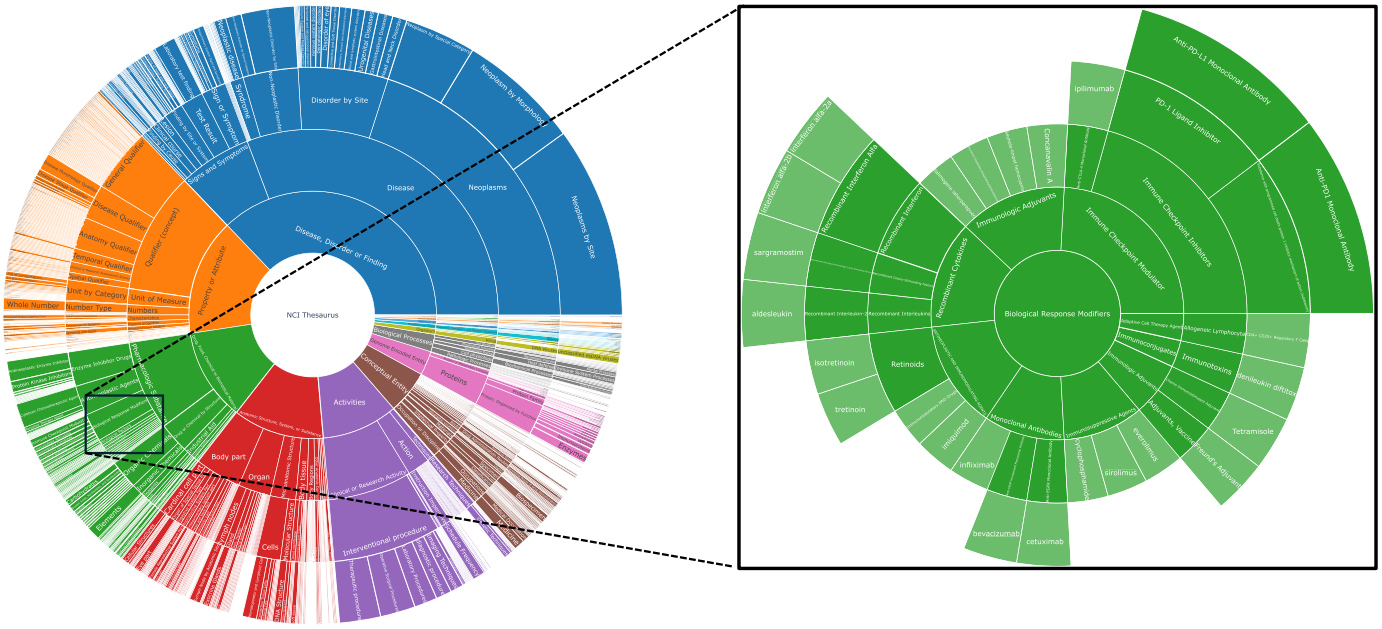


Fig. A.6. Sunburst chart of the hierarchical KG generated from melanoma-related publications, highlighting the "biological response modifiers" category and its associated pharmacological substances.

Table A.8
Top concept ranking before and after KLD re-ranking across different categories for AK-SCC of the skin.

(a) Diagnostic Procedure

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Diagnosis	0.00305	Diagnosis	34,735	Diagnosis	1.0
Immunohistochemistry	0.00228	Biopsy	9,895	Immunohistochemistry	0.595
Biopsy	0.00214	Immunohistochemistry	8,516	Biopsy	0.576
Endoscopy (procedure)	0.00081	Diagnostic Imaging	6,742	Endoscopy (procedure)	0.234
Sentinel Lymph Node Biopsy	0.00052	Contrast used	5,676	X-Ray Computed Tomography	0.149
Positron-Emission Tomography	0.00051	Endoscopy (procedure)	5,557	Positron-Emission Tomography	0.138
X-Ray Computed Tomography	0.00046	Magnetic Resonance Imaging	5,404	Sentinel Lymph Node Biopsy	0.132
Endoscopic Ultrasound	0.00036	X-Ray Computed Tomography	5,123	Endoscopic Ultrasound	0.098
Colposcopy	0.00034	Radionuclide Imaging	2,539	Colposcopy	0.088
PET/CT scan	0.00024	Positron-Emission Tomography	2,499	Diagnostic Imaging	0.07
Diagnostic Neoplasm Staging	0.00021	Ultrasonography	2,144	PET/CT scan	0.063
Differential Diagnosis	0.00017	Endoscopic Ultrasound	1,719	Magnetic Resonance Imaging	0.062
Bronchoscopy	0.00016	Differential Diagnosis	1,664	Diagnostic Neoplasm Staging	0.053
Lymphoscintigraphy	0.00016	Sentinel Lymph Node Biopsy	1,484	Differential Diagnosis	0.053
Cervical biopsy (procedure)	0.00014	Radiographic imaging procedure	1245	Bronchoscopy	0.043
Mediastinoscopy	0.00012	Autopsy	1,013	Lymphoscintigraphy	0.04
Diffusion weighted imaging	0.00011	Colposcopy	993	Cervical biopsy (procedure)	0.034
Incisional biopsy - action	0.0001	Sampling - Surgical action	897	Diffusion weighted imaging	0.034

(b) Amino Acid, Peptide, or Protein

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Cetuximab	0.00186	Proteins	25279	Cetuximab	0.746
Epidermal Growth Factor Receptor	0.00169	Receptor	8460	Epidermal Growth Factor Receptor	0.697
Cytokeratin	0.00136	Antibodies	6227	Cytokeratin	0.559
Squamous Cell Carcinoma Antigen	0.00131	Epidermal Growth Factor Receptor	4971	Squamous Cell Carcinoma Antigen	0.516
Oncogene Protein Bcl-1	0.0011	Cytokeratin	3956	Oncogene Protein Bcl-1	0.451
Bleomycin	0.00075	Cetuximab	3863	Bleomycin	0.308
Matrix Metalloproteinase 2	0.0005	Monoclonal Antibodies	3241	Matrix Metalloproteinase 2	0.213
Matrix Metalloproteinase 9	0.0004	Oncogene Protein Bcl-1	3082	Monoclonal Antibodies	0.177
Monoclonal Antibodies	0.00037	Growth Factor	2689	Matrix Metalloproteinase 9	0.176
PTHrP	0.00033	Cytokine	2558	VEGF-A	0.148
VEGF-A	0.00033	Enzymes	2554	PTHrP	0.139
Vimentin	0.0003	Bleomycin	2227	Vimentin	0.131
Epidermal Growth Factor	0.00028	Matrix Metalloproteinase 2	2048	Epidermal Growth Factor	0.119
Nivolumab	0.00025	Squamous Cell Carcinoma Antigen	2017	Nivolumab	0.105
Cemiplimab	0.00024	Matrix Metalloproteinase 9	1997	Protein Tyrosine Kinase	0.103
Protein Tyrosine Kinase	0.00022	VEGF-A	1898	Cemiplimab	0.095
Carcinoembryonic Antigen	0.00021	Peptides	1871	Carcinoembryonic Antigen	0.09
Oncogene Proteins	0.0002	Protein Tyrosine Kinase	1735	Oncogene Proteins	0.086

Table A.8 (*continued*)

(c) Disease or Syndrome

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Esophageal Diseases	0.00249	Syndrome	42097	Syndrome	0.999
Syndrome	0.00249	Esophageal Diseases	6526	Esophageal Diseases	0.746
Human papilloma virus infection	0.00164	Communicable Diseases	4689	Human papilloma virus infection	0.488
Hyperkeratosis	0.00148	Human papilloma virus infection	3703	Hyperkeratosis	0.436
Actinic porokeratosis	0.0011	Hyperkeratosis	2788	Actinic porokeratosis	0.321
Deglutition Disorders	0.00063	Inflammatory disorder	2701	Deglutition Disorders	0.193
Recurrent disease	0.00049	Deglutition Disorders	2369	Recurrent disease	0.147
Disorder of neck	0.00035	Actinic porokeratosis	1732	Disorder of neck	0.104
Condyloma	0.00034	Recurrent disease	1324	Condyloma	0.102
Sentinel node (disorder)	0.00032	Skin lesion	1117	Sentinel node (disorder)	0.095
Skin lesion	0.00024	Hypercalcemia	877	Skin lesion	0.077
Lichen Planus, Oral	0.00023	Absolute anemia	873	Lichen Planus, Oral	0.069
Lichen Sclerosus et Atrophicus	0.00022	Sentinel node (disorder)	857	Xerostomia	0.066
Xerostomia	0.00022	Lymphadenopathy	831	Hypercalcemia	0.065
Hypercalcemia	0.00021	Condyloma	816	Lichen Sclerosus et Atrophicus	0.064
Lymphadenopathy	0.00019	Xerostomia	782	Lymphadenopathy	0.058
Leukopenia	0.00016	Leukopenia	669	Leukopenia	0.051

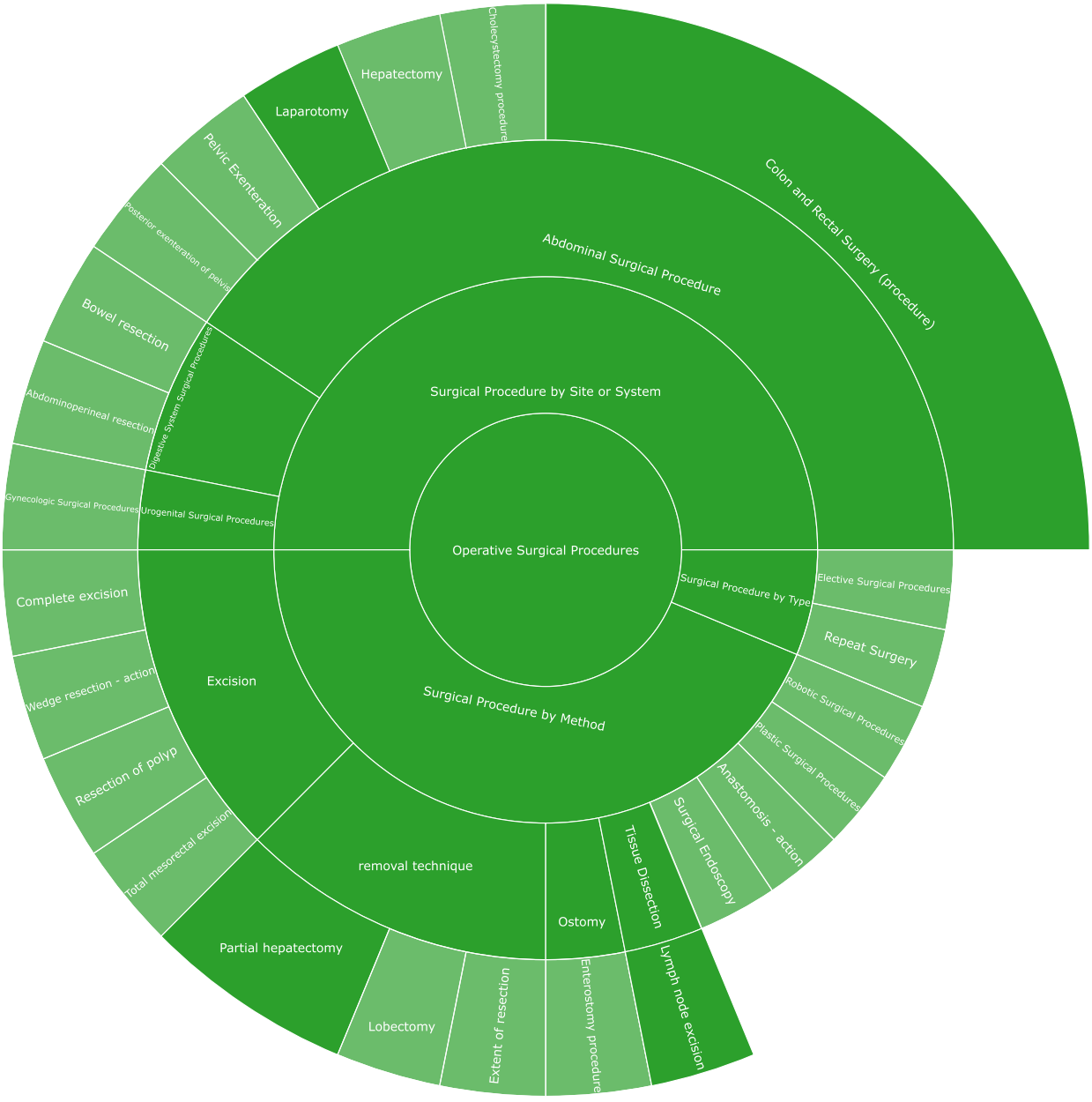


Fig. A.7. Sunburst chart of the hierarchical KG generated from colon cancer-related publications, highlighting the “operative surgical procedures” category.

Table A.9
Top concept ranking before and after KLD re-ranking across different categories for breast cancer.

(a) Diagnostic Procedure

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Mammography	0.01054	Diagnosis	92,089	Mammography	0.853
Diagnosis	0.00357	Mammography	46,922	Diagnosis	0.537
Screening mammography	0.00296	Diagnostic Imaging	32,352	Screening mammography	0.238
Sentinel Lymph Node Biopsy	0.00221	MRI	28,002	Biopsy	0.199
Biopsy	0.00188	Biopsy	22,776	Diagnostic Imaging	0.19
MRI	0.00127	Contrast used	19,526	Sentinel Lymph Node Biopsy	0.182
Diagnostic Imaging	0.00127	Ultrasonography	16,093	MRI	0.176
Core needle biopsy	0.00122	Screening mammography	12,730	Core needle biopsy	0.103
MRI of breast	0.00099	Immunohistochemistry	11,885	Immunohistochemistry	0.098
Biopsy of breast	0.0009	Sentinel Lymph Node Biopsy	10,991	Ultrasonography	0.094
Immunohistochemistry	0.00089	Radionuclide Imaging	6,669	MRI of breast	0.08
Screening for cancer	0.0008	Core needle biopsy	6,609	Screening for cancer	0.073
Ultrasonography	0.00063	Screening for cancer	6,105	Biopsy of breast	0.073
Breast Self-Examination	0.0006	Radiographic imaging procedure	5,544	Contrast used	0.073
Examination of breast	0.00055	X-Ray Computed Tomography	4,892	Breast Self-Examination	0.048
Ultrasonography, Mammary	0.00054	MRI of breast	4,260	Examination of breast	0.045
Digital Breast Tomosynthesis	0.00038	Sampling - Surgical action	4,002	Ultrasonography, Mammary	0.044
Lymphoscintigraphy	0.00034	Biopsy of breast	3,952	Digital Breast Tomosynthesis	0.031

(b) Amino Acid, Peptide, or Protein

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Estrogen Receptors	0.00668	Proteins	56438	Estrogen Receptors	0.888
Trastuzumab	0.00544	Receptor	48155	Trastuzumab	0.702
Receptors, Progesterone	0.00246	Estrogen Receptors	35401	Receptor	0.366
Erbb-2 Receptor	0.00193	Trastuzumab	24893	Receptors, Progesterone	0.327
EGFR	0.00164	Antibodies	15160	Erbb-2 Receptor	0.256
Receptor	0.00105	Receptors, Progesterone	13061	EGFR	0.236
Pertuzumab	0.00061	EGFR	12107	Growth Factor	0.091
Oncogene Protein Bcl-1	0.00052	Erbb-2 Receptor	10111	Pertuzumab	0.079
Growth Factor	0.00038	Growth Factor	9549	Oncogene Protein Bcl-1	0.078
Mucinous CA Antigen	0.00037	Enzymes	8914	Monoclonal Antibodies	0.064
Cytokeratin	0.00028	Peptides	8788	Mucinous CA Antigen	0.052
G-CSF	0.00028	Monoclonal Antibodies	6972	G-CSF	0.048
Receptors, Steroid	0.00027	Cytokine	5706	Cytokeratin	0.048
Monoclonal Antibodies	0.00026	Oncogene Protein Bcl-1	4537	Insulin-Like Growth Factor I	0.044
Cathepsin D	0.00023	Protein Tyrosine Kinase	4156	Protein Tyrosine Kinase	0.043
Bevacizumab	0.00023	Insulin-Like Growth Factor I	4120	Bevacizumab	0.04
Ado-Trastuzumab Emtansine	0.00022	Chromatin	3812	Matrix Metalloproteinase 9	0.04

(c) Disease or Syndrome

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Sentinel Node (Disorder)	0.001	Syndrome	82535	Syndrome	0.902
Breast Diseases	0.00089	Sentinel Node (Disorder)	5087	Sentinel Node (Disorder)	0.718
Syndrome	0.00086	Inflammatory Disorder	4969	Breast Diseases	0.638
Cystic Breast Disease	0.00071	Diabetes Mellitus	4370	Cystic Breast Disease	0.511
Febrile Neutropenia	0.00021	Communicable Diseases	4118	Febrile Neutropenia	0.15
Adenosis	0.00018	Breast Diseases	4051	Adenosis	0.128
Comedone	0.00018	Coronary Artery Disease	3843	Comedone	0.127
Alopecia	0.00013	Auditory Recruitment	3267	Alopecia	0.094
IGF-I	0.0001	Cystic Breast Disease	3264	IGF-I	0.074
Osteitis	0.0001	Obesity	3221	Osteitis	0.07
Leukopenia	0.00009	Covid19 (Disease)	2500	Leukopenia	0.065
Disorder Of Axilla	0.00007	Disorder Of Circulatory System	1957	Recurrent Disease	0.049
Recurrent Disease	0.00007	Diabetes Mellitus, Insulin-Dependent	1857	Disorder Of Axilla	0.048
Sickle Cell Dactylitis	0.00006	Osteoporosis	1811	Sickle Cell Dactylitis	0.046
TRPS I	0.00006	Alopecia	1757	Thrombotic Microangiopathies	0.044
Thrombotic Microangiopathies	0.00006	Febrile Neutropenia	1659	TRPS I	0.042
Local Disease	0.00005	Drug Resistant Epilepsy	1436	Pleural Effusion Disorder	0.041
Pleural Effusion Disorder	0.00005	Osteopenia	1429	Local Disease	0.041

A.4. Concept ranking for remaining conditions

This subsection presents concept rankings for conditions such as AK-SCC of the skin, breast cancer, colon cancer, endometrial cancer, and soft tissue sarcoma. The tables compare the top concepts before and after KLD re-ranking, focusing on diagnostic procedures, amino acids, peptides or proteins, and diseases or syndromes, highlighting key diagnostic and therapeutic targets.

Table A.10
Top concept ranking before and after KLD re-ranking across different categories for colon cancer.

(a) Diagnostic Procedure

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Colonoscopy	0.01668	Diagnosis	29484	Colonoscopy	0.971
Endoscopy	0.00285	Colonoscopy	26636	Diagnosis	0.396
Diagnosis	0.00227	Endoscopy	10931	Endoscopy	0.231
Screening For Cancer	0.00178	Immunohistochemistry	6095	Immunohistochemistry	0.124
Screening Colonoscopy	0.00174	Diagnostic Imaging	5965	Screening For Cancer	0.118
Immunohistochemistry	0.00147	Contrast Used	5187	Screening Colonoscopy	0.099
Flexible Sigmoidoscopy	0.0011	Screening For Cancer	4156	Flexible Sigmoidoscopy	0.063
Sigmoidoscopy	0.00099	Biopsy	3831	Diagnostic Imaging	0.061
Intraoperative Ultrasound	0.00039	X-Ray Computed Tomography	3068	Sigmoidoscopy	0.058
Biopsy	0.00038	Magnetic Resonance Imaging	2974	Biopsy	0.055
Diagnostic Neoplasm Staging	0.00035	Screening Colonoscopy	2501	X-Ray Computed Tomography	0.035
Total Colonoscopy	0.00031	Flexible Fiberoptic Sigmoidoscopy	1632	Intraoperative Ultrasound	0.025
Barium Enema	0.00021	Radionuclide Imaging	1587	Diagnostic Neoplasm Staging	0.024
Double Contrast Barium Enema	0.0002	Sigmoidoscopy	1548	Positron-Emission Tomography	0.019
Endoscopic Examination On Colon	0.00019	Ultrasonography	1246	Total Colonoscopy	0.019
Laparoscopy	0.00018	Positron-Emission Tomography	1215	Laparoscopy	0.018
Malignant Neoplasm Screening	0.00017	Sampling - Surgical Action	1185	Barium Enema	0.014

(b) Amino Acid, Peptide, or Protein

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Cetuximab	0.00469	Proteins	21940	Cetuximab	0.804
Bevacizumab	0.00444	Bevacizumab	8936	Bevacizumab	0.784
Carcinoembryonic Antigen	0.00221	Receptor	8033	Carcinoembryonic Antigen	0.39
Panitumumab	0.00175	Cetuximab	7636	Panitumumab	0.296
EGFR	0.00074	Antibodies	6206	EGFR	0.148
Monoclonal Antibodies	0.00058	Carcinoembryonic Antigen	4380	Monoclonal Antibodies	0.14
VEGF-A	0.00058	Monoclonal Antibodies	3813	VEGF-A	0.122
Thymidylate Synthase	0.00044	Enzymes	3444	Antibodies	0.096
Oncogene Protein Bcl-1	0.00035	Cytokine	3185	Thymidylate Synthase	0.081
Cyclooxygenase 2	0.00035	EGFR	2672	Oncogene Protein Bcl-1	0.073
Matrix Metalloproteinase 7	0.0003	Panitumumab	2531	Cyclooxygenase 2	0.071
Mucins	0.00027	Peptides	2497	Mucins	0.059
Aflibercept	0.00026	VEGF-A	2491	Matrix Metalloproteinase 7	0.057
Matrix Metalloproteinase 9	0.00024	Growth Factor	2431	Matrix Metalloproteinase 9	0.056
Ramucirumab	0.00023	Matrix Metalloproteinase 9	1406	Growth Factor	0.052
Dihydropyrimidine Dehydrogenase	0.00022	Oncogene Protein Bcl-1	1343	Aflibercept	0.049
Matrix Metalloproteinase 2	0.00019	Protein-Serine-Threonine Kinases	1264	Matrix Metalloproteinase 2	0.045
Tissue-Inhibitor Of Metalloproteinase-1	0.00018	Mucins	1260	Ramucirumab	0.042

(c) Disease or Syndrome

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Lynch Syndrome	0.00244	Syndrome	33788	Lynch Syndrome	0.738
Inflammatory Bowel Diseases	0.00167	Inflammatory Bowel Diseases	5889	Syndrome	0.673
Syndrome	0.00129	Inflammatory Disorder	4334	Inflammatory Bowel Diseases	0.532
Chronic Ulcerative Colitis	0.00084	Lynch Syndrome	4265	Chronic Ulcerative Colitis	0.269
Colitis	0.00082	Diabetes Mellitus	2995	Colitis	0.263
Turcot Syndrome (Disorder)	0.00068	Chronic Ulcerative Colitis	2844	Turcot Syndrome (Disorder)	0.208
Hematochezia	0.0003	Colitis	2755	Hematochezia	0.096
Primary Sclerosing Cholangitis	0.00029	Communicable Diseases	1744	Primary Sclerosing Cholangitis	0.094
Ileus	0.00021	Crohn Disease	1202	Ileus	0.072
Recurrent Disease	0.00021	Obesity	1164	Recurrent Disease	0.07
Sarcopenia	0.00017	Turcot Syndrome (Disorder)	1089	Sarcopenia	0.061
Sickle Cell Dactylitis	0.00017	Absolute Anemia	1058	Crohn Disease	0.059
Crohn Disease	0.00016	Covid19 (Disease)	1040	Sickle Cell Dactylitis	0.054
Obstruction Of Colon	0.00016	Sarcopenia	1016	Obstruction Of Colon	0.051
Diverticulitis	0.00013	Auditory Recruitment	1001	Diverticulitis	0.045
Diverticular Diseases	0.00013	Primary Sclerosing Cholangitis	943	Diverticular Diseases	0.044
CMMRD	0.0001	Ileus	875	CMMRD	0.034
Stomatitis	0.0001	Hematochezia	764	Stomatitis	0.034

Table A.11
Top concept ranking before and after KLD re-ranking across different categories for endometrial cancer.

(a) Diagnostic Procedure

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Endometrial Biopsy	0.00689	Diagnosis	9843	Diagnosis	0.929
Diagnosis	0.00618	Magnetic Resonance Imaging	2542	Endometrial Biopsy	0.759
Hysteroscopy	0.00543	Immunohistochemistry	2198	Hysteroscopy	0.606
Transvaginal Echography	0.00368	Endometrial Biopsy	1928	Transvaginal Echography	0.417
Immunohistochemistry	0.00314	Hysteroscopy	1698	Immunohistochemistry	0.393
Laparoscopy	0.00214	Ultrasonography	1675	Laparoscopy	0.261
Magnetic Resonance Imaging	0.00153	Biopsy	1673	Magnetic Resonance Imaging	0.243
Biopsy	0.00151	Diagnostic Imaging	1536	Biopsy	0.215
Infertility Study	0.00148	Contrast Used	1369	Infertility Study	0.196
Ultrasonography	0.00103	Transvaginal Echography	1213	Ultrasonography	0.168
Diagnostic Neoplasm Staging	0.0008	Infertility Study	1173	Diagnostic Neoplasm Staging	0.104
Sentinel Lymph Node Biopsy	0.00072	Laparoscopy	1132	Diffusion Weighted Imaging	0.1
Diffusion Weighted Imaging	0.0007	Sampling - Surgical Action	685	Sentinel Lymph Node Biopsy	0.097
Sampling Of Lymph Node	0.00032	Diffusion Weighted Imaging	522	Diagnostic Imaging	0.075
Differential Diagnosis	0.00026	X-Ray Computed Tomography	483	Differential Diagnosis	0.052
Screening For Cancer	0.00022	Differential Diagnosis	422	Contrast Used	0.051
Magnetic Resonance Imaging (Mri) Of Pelvis	0.0002	Sentinel Lymph Node Biopsy	390	Sampling - Surgical Action	0.049
Diagnostic Imaging	0.00015	Diagnostic Neoplasm Staging	366	Sampling Of Lymph Node	0.048

(b) Amino Acid, Peptide, or Protein

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Estrogen Receptors	0.0029	Proteins	5101	Estrogen Receptors	0.793
Receptors, Progesterone	0.00281	Receptor	2568	Receptors, Progesterone	0.75
Oncogene Protein Bcl-1	0.00077	Estrogen Receptors	1574	Proteins	0.359
Pembrolizumab	0.00054	Receptors, Progesterone	1206	Oncogene Protein Bcl-1	0.279
Receptors, Steroid	0.00044	Antibodies	703	Receptor	0.218
Vimentin	0.00042	Growth Factor	549	Pembrolizumab	0.218
Matrix Metalloproteinase 2	0.00041	Enzymes	518	Vimentin	0.195
Cytokeratin	0.00033	Oncogene Protein Bcl-1	499	Matrix Metalloproteinase 2	0.194
GnRH	0.0003	Insulin	402	Receptors, Steroid	0.192
Cytochrome P-450 Cyp1B1	0.00028	Leptin	396	Cytokeratin	0.175
VEGF-A	0.00027	Matrix Metalloproteinase 2	374	VEGF-A	0.163
Erbb-2 Receptor	0.00025	VEGF-A	350	GnRH	0.159
Adiponectin	0.00025	Vimentin	345	Adiponectin	0.156
Thymidine Phosphorylase	0.00025	Monoclonal Antibodies	342	Cytochrome P-450 Cyp1B1	0.155
Wfdc2 Protein, Human	0.00024	Cytokeratin	322	Leptin	0.154
Cytochrome P-450 Cyp1A1	0.00024	Insulin-Like Growth Factor I	319	Erbb-2 Receptor	0.15
Inhibin	0.00023	Pembrolizumab	305	Cytochrome P-450 Cyp1A1	0.146
Gonadorelin	0.00022	Adiponectin	302	Thymidine Phosphorylase	0.145

(c) Disease or Syndrome

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Endometrial Hyperplasia	0.00808	Syndrome	9585	Endometrial Hyperplasia	0.765
Lynch Syndrome	0.00617	Endometrial Hyperplasia	2063	Lynch Syndrome	0.602
Syndrome	0.0033	Lynch Syndrome	1986	Syndrome	0.601
Atypical Endometrial Hyperplasia	0.00289	Obesity	1353	Atypical Endometrial Hyperplasia	0.287
Endometriosis	0.00195	Endometriosis	1139	Endometriosis	0.223
Obesity	0.00116	Diabetes Mellitus	1064	Obesity	0.164
Recurrent Disease	0.00093	Atypical Endometrial Hyperplasia	679	Recurrent Disease	0.117
Turcot Syndrome (Disorder)	0.00092	Recurrent Disease	445	Sentinel Node (Disorder)	0.113
Sentinel Node (Disorder)	0.00091	Sentinel Node (Disorder)	394	Turcot Syndrome (Disorder)	0.112
Lymphocele	0.00052	Inflammatory Disorder	326	Lymphocele	0.075
Benign Endometrial Hyperplasia	0.00033	Turcot Syndrome (Disorder)	306	Benign Endometrial Hyperplasia	0.055
Complex Endometrial Hyperplasia	0.00029	Absolute Anemia	213	Polycystic Ovary Syndrome	0.053
Uterine Diseases	0.00026	Lymphocele	211	Complex Endometrial Hyperplasia	0.052
Polycystic Ovary Syndrome	0.00025	Polycystic Ovary Syndrome	208	Uterine Diseases	0.049
Hyperestrogenism	0.00018	Communicable Diseases	169	Hyperestrogenism	0.042
Lymphedema Of Lower Extremity	0.00015	Metabolic Syndrome X	161	Lymphedema Of Lower Extremity	0.039
Endometrial Disorder	0.00014	Auditory Recruitment	138	Female Genital Diseases	0.039

Table A.12
Top concept ranking before and after KLD re-ranking across different categories for soft tissue sarcoma.

(a) Diagnostic Procedure

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Biopsy	0.00201	Diagnostic Imaging	2561	Biopsy	0.213
Differential Diagnosis	0.00174	Magnetic Resonance Imaging	2358	Magnetic Resonance Imaging	0.202
Magnetic Resonance Imaging	0.00165	Biopsy	1767	Diagnostic Imaging	0.194
Immunohistochemistry	0.00153	Immunohistochemistry	1227	Differential Diagnosis	0.175
Diagnostic Imaging	0.00147	X-Ray Computed Tomography	1202	Immunohistochemistry	0.161
X-Ray Computed Tomography	0.00079	Differential Diagnosis	1148	X-Ray Computed Tomography	0.103
Incisional Biopsy - Action	0.00057	Contrast Used	1133	Positron-Emission Tomography	0.062
Core Needle Biopsy	0.00052	Radionuclide Imaging	663	Incisional Biopsy - Action	0.059
Positron-Emission Tomography	0.00052	Ultrasonography	550	Core Needle Biopsy	0.057
Autopsy	0.0003	Positron-Emission Tomography	470	Autopsy	0.043
Fine Needle Aspiration Biopsy	0.00021	Autopsy	374	Radionuclide Imaging	0.036
Diffusion Weighted Imaging	0.00019	Radiographic Imaging Procedure	332	Contrast Used	0.033
Pet/Ct Scan	0.00018	Echocardiography	261	Diffusion Weighted Imaging	0.029
Excision Biopsy	0.00017	Core Needle Biopsy	261	Fine Needle Aspiration Biopsy	0.029
Surgical Pathology Procedure	0.00017	Incisional Biopsy - Action	210	Pet/Ct Scan	0.026
Needle Biopsy Procedure	0.00015	Diffusion Weighted Imaging	208	Excision Biopsy	0.025
Chest Ct	0.00015	Sampling - Surgical Action	202	Surgical Pathology Procedure	0.025

(b) Amino Acid, Peptide, or Protein

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Cytokeratin	0.00159	Proteins	2172	Cytokeratin	0.814
Vimentin	0.00139	Receptor	1114	Vimentin	0.747
Desmin	0.00114	Antibodies	1016	Desmin	0.654
Olaratumab	0.00096	Cytokeratin	824	Melphalan	0.589
Melphalan	0.00092	Vimentin	703	Olaratumab	0.573
Cd34 Antigens	0.00054	Tumor Necrosis Factor-Alpha	543	Cd34 Antigens	0.495
Mucinous CA Antigen	0.00045	Actins	486	Proto-Oncogene C-Kit	0.448
Proto-Oncogene C-Kit	0.00042	Desmin	462	Mucinous CA Antigen	0.448
Protein Tyrosine Kinase	0.00032	Melphalan	396	Actins	0.444
Dactinomycin	0.00029	Protein Tyrosine Kinase	390	Protein Tyrosine Kinase	0.441
Actins	0.00028	Cd34 Antigens	390	Antibodies	0.428
Granulocyte Colony-Stimulating Factor	0.00023	Monoclonal Antibodies	389	Tumor Necrosis Factor-Alpha	0.427
Tumor Necrosis Factor-Alpha	0.00018	Cytokine	308	Dactinomycin	0.404
Tumor Antigens	0.00016	Enzymes	302	GM-CSF	0.398
GM-CSF	0.00014	Glycoproteins	296	Monoclonal Antibodies	0.39
Gamma-Enolase	0.00012	Growth Factor	276	Tumor Antigens	0.366
Pembrolizumab	0.00011	Proto-Oncogene C-Kit	270	Glycoproteins	0.364

(c) Disease or Syndrome

Concept	KLD↓	Concept	Freq.↓	Concept	Re-rank↓
Fasciitis	0.00061	Inflammatory Disorder	599	Recurrent Disease	0.176
Recurrent Disease	0.00061	Local Disease	331	Fasciitis	0.173
Nodular Fasciitis	0.00054	Recurrent Disease	290	Nodular Fasciitis	0.163
Osteitis	0.00041	Nodular Fasciitis	216	Osteitis	0.145
Marfan Syndrome	0.00023	Gigantism	214	Marfan Syndrome	0.118
Febrile Neutropenia	0.00023	Absolute Anemia	208	Febrile Neutropenia	0.117
Li-Fraumeni Syndrome	0.00019	Osteitis Deformans	201	Li-Fraumeni Syndrome	0.11
Infection By Spirocerca	0.00017	Fasciitis	201	Gigantism	0.108
Leukopenia	0.00016	Marfan Syndrome	152	Leukopenia	0.107
Gigantism	0.00014	Febrile Neutropenia	132	Infection By Spirocerca	0.107
Pulmonary Thromboembolisms	0.00014	Pleural Effusion Disorder	126	Pulmonary Thromboembolisms	0.103
Myositis Ossificans	0.00012	Leukopenia	121	Lymphadenopathy	0.101
Lymphadenopathy	0.00011	Lymphadenopathy	115	Myositis Ossificans	0.1
Gross' Disease	0.00011	Diabetes Mellitus	100	Pleural Effusion Disorder	0.1
Bone Lesion	0.00011	Primary Sclerosing Cholangitis	98	Primary Sclerosing Cholangitis	0.1
Myositis, Proliferative	0.00011	Uveoparotid Fever	87	Bone Lesion	0.099

References

[1] Dash S, Shakyawar SK, Sharma M, Kaushik SR. Big data in healthcare: management, analysis and future prospects. J Big Data 2019;6.

[2] Ehrlinger L, Wöß W. Towards a definition of knowledge graphs. SEMANTiCS (Posters, Demos, SuCCESS) 2016;48(1–4):2.

[3] Suchanek FM, Kasneci G, Weikum G. Yago: a core of semantic knowledge. In: Proceedings of the 16th international conference on world wide web. New York, NY, USA: Association for Computing Machinery; 2007. p. 697–706.

[4] Sheth A, Padhee S, Gyrard A. Knowledge graphs and knowledge networks: the story in brief. IEEE Internet Comput 2019;23(4):67–75. https://doi.org/10.1109/MIC.2019.2928449.

[5] Nicholson DN, Greene CS. Constructing knowledge graphs and their biomedical applications. Comput Struct Biotechnol J 2020;18:1414–28. https://doi.org/10.1016/j.csbj.2020.05.017.

[6] Larsen PO, von Ins M. The rate of growth in scientific publication and the decline in coverage provided by science citation index. Scientometrics 2010;84:575–603.

[7] Li L, Wang P, Yan J, Wang Y, Li S, Jiang J, et al. Real-world data medical knowledge graph: construction and applications. Artif Intell Med 2020;103:101817. https://doi.org/10.1016/j.artmed.2020.101817.

[8] Hoyt CT, Domingo-Fernández D, Aldisi R, Xu L, Kolpeja K, Spalek S, et al. Re-curation and rational enrichment of knowledge graphs in biological expression language. Database 2019;2019:baz068. https://doi.org/10.1093/database/baz068.

[9] Cejuela JM, Vinchurkar S, Goldberg T, Prabhu Shankar MS, Baghudana A, Bojchevski A, et al. LocText: relation extraction of protein localizations to assist database curation. BMC Bioinform 2018;19(1):15. https://doi.org/10.1186/s12859-018-2021-9.

[10] Yuan J, Jin Z, Guo H, Jin H, Zhang X, Smith T, et al. Constructing biomedical domain-specific knowledge graph with minimum supervision. Knowl Inf Syst 2020;62(1):317–36. https://doi.org/10.1007/s10115-019-01351-4.

[11] Wang Q, Mao Z, Wang B, Guo L. Knowledge graph embedding: a survey of approaches and applications. IEEE Trans Knowl Data Eng 2017;29(12):2724–43. https://doi.org/10.1109/TKDE.2017.2754499.
[12] Ong E, Wang LL, Schaub J, O'Toole JF, Steck B, Rosenberg AZ, et al. Kidney precision medicine project, modelling kidney disease using ontology: insights from the kidney precision medicine project. Nat Rev Nephrol 2020;16(11):686–96. https://doi.org/10.1038/s41581-020-00335-w.
[13] Gao Z, Ding P, Xu R. KG-predict: a knowledge graph computational framework for drug repurposing. J Biomed Inform 2022;132:104133. https://doi.org/10.1016/j.jbi.2022.104133.
[14] Zhao C, Jiang J, Guan Y, Guo X, He B. EMR-based medical knowledge representation and inference via Markov random fields and distributed representation learning. Artif Intell Med 2018;87:49–59. https://doi.org/10.1016/j.artmed.2018.03.005.
[15] Li X, Wang Y, Wang D, Yuan W, Peng D, Mei Q. Improving rare disease classification using imperfect knowledge graph. BMC Med Inform Decis Mak 2019;19(5):238. https://doi.org/10.1186/s12911-019-0938-1.
[16] Vilela J, Asif M, Marques AR, Santos JX, Rasga C, Vicente A, et al. Biomedical knowledge graph embeddings for personalized medicine: predicting disease-gene associations. Expert Syst 2023;40(5):e13181. https://doi.org/10.1111/exsy.13181.
[17] Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. Sci Data 2023;10(1):67. https://doi.org/10.1038/s41597-023-01960-3.
[18] van der Velden D, van Herpen C, van Laarhoven H, Smit E, Groen H, Willems S, et al. Molecular tumor boards: current practice and future needs. Ann Oncol 2017;28(12):3070–5. https://doi.org/10.1093/annonc/mdx528. The antibody-drug conjugate target landscape.
[19] Zhang N, Deng S, Sun Z, Wang G, Chen X, Zhang W, et al. Long-tail relation extraction via knowledge graph embeddings and graph convolution networks. In: Burstein J, Doran C, Solorio T, editors. Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: human language technologies, volume 1 (long and short papers). Minneapolis, Minnesota: Association for Computational Linguistics; 2019. p. 3016–25.
[20] Huang X, Zhang J, Li D, Li P. Knowledge graph embedding based question answering. In: Proceedings of the twelfth ACM international conference on web search and data mining. New York, NY, USA: Association for Computing Machinery; 2019. p. 105–13.
[21] Pan S, Luo L, Wang Y, Chen C, Wang J, Wu X. Unifying large language models and knowledge graphs: a roadmap. IEEE Trans Knowl Data Eng 2024;36(7):3580–99. https://doi.org/10.1109/TKDE.2024.3352100.
[22] Fei H, Ren Y, Zhang Y, Ji D, Liang X. Enriching contextualized language model from knowledge graph for biomedical information extraction. Brief Bioinform 2020;22(3):bbaa110. https://doi.org/10.1093/bib/bbaa110.
[23] Yuan Z, Liu Y, Tan C, Huang S, Huang F. Improving biomedical pretrained language models with knowledge. In: Demner-Fushman D, Cohen KB, Ananiadou S, Tsujii J, editors. Proceedings of the 20th workshop on biomedical language processing. Association for Computational Linguistics, Online; 2021. p. 180–90.
[24] Hu L, Liu Z, Zhao Z, Hou L, Nie L, Li J. A survey of knowledge enhanced pre-trained language models. IEEE Trans Knowl Data Eng 2024;36(4):1413–30. https://doi.org/10.1109/TKDE.2023.3310002.
[25] Gong F, Wang M, Wang H, Wang S, Liu M. SMR: medical knowledge graph embedding for safe medicine recommendation. Big Data Res 2021;23:100174. https://doi.org/10.1016/j.bdr.2020.100174.
[26] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2019;36(4):1234–40. https://doi.org/10.1093/bioinformatics/btz682.
[27] Beltagy I, Lo K, Cohan A. SciBERT: a pretrained language model for scientific text. In: Inui K, Jiang J, Ng V, Wan X, editors. Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing. Hong Kong, China: Association for Computational Linguistics; 2019. p. 3615–20.
[28] Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, et al. Domain-specific language model pretraining for biomedical natural language processing. ACM Trans Comput Healthc oct 2021;3(1). https://doi.org/10.1145/3458754.
[29] Meng Z, Liu F, Clark T, Shareghi E, Collier N. Mixture-of-partitions: infusing large biomedical knowledge graphs into BERT. In: Moens M-F, Huang X, Specia L, Yih SW-t, editors. Proceedings of the 2021 conference on empirical methods in natural language processing. Punta Cana, Dominican Republic: Association for Computational Linguistics, Online; 2021. p. 4672–81.
[30] Pfeiffer J, Kamath A, Rücklé A, Cho K, Gurevych I. AdapterFusion: non-destructive task composition for transfer learning. In: Merlo P, Tiedemann J, Tsarfaty R, editors. Proceedings of the 16th conference of the European chapter of the Association for Computational Linguistics: main volume. Association for Computational Linguistics, Online; 2021. p. 487–503.
[31] Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32(suppl_1):D267–70. https://doi.org/10.1093/nar/gkh061.
[32] Stearns MQ, Price C, Spackman KA, Wang AY. SNOMED clinical terms: overview of the development process and project status. In: AMIA 2001, American medical informatics association annual symposium; November, 2001. p. 3–7.
[33] Zhong L, Wu J, Li Q, Peng H, Wu X. A comprehensive survey on automatic knowledge graph construction. ACM Comput Surv nov 2023;56(4). https://doi.org/10.1145/3618295.
[34] Abu-Salih B, AL-Qurishi M, Alweshah M, AL-Smadi M, Alfayez R, Saadeh H. Healthcare knowledge graph construction: a systematic review of the state-of-the-art, open issues, and opportunities. J Big Data 2023;10(1):81. https://doi.org/10.1186/s40537-023-00774-9.
[35] Al-Moslmi T, Gallofré Ocaña M, Opdahl AL, Veres C. Named entity extraction for knowledge graphs: a literature overview. IEEE Access 2020;8:32862–81. https://doi.org/10.1109/ACCESS.2020.2973928.
[36] Nayak T, Majumder N, Goyal P, Poria S. Deep neural approaches to relation triplets extraction: a comprehensive survey. Cogn Comput 2021;13:1215–32. https://doi.org/10.1007/s12559-021-09917-7.
[37] Cimiano P, Paulheim H. Knowledge graph refinement: a survey of approaches and evaluation methods. Semant Web 2017;8(3):489–508. https://doi.org/10.3233/SW-160218.
[38] Aronson AR, Lang F-M. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229–36. https://doi.org/10.1136/jamia.2009.002733.
[39] Wu H, Toti G, Morley KI, Ibrahim ZM, Folarin A, Jackson R, et al. SemEHR: a general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research. J Am Med Inform Assoc 2018;25(5):530–7. https://doi.org/10.1093/jamia/ocx160.
[40] Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507–13. https://doi.org/10.1136/jamia.2009.001560.
[41] Soysal E, Wang J, Jiang M, Wu Y, Pakhomov S, Liu H, et al. CLAMP – a toolkit for efficiently building customized clinical natural language processing pipelines. J Am Med Inform Assoc 2017;25(3):331–6. https://doi.org/10.1093/jamia/ocx132.
[42] Neumann M, King D, Beltagy I, Ammar W. ScispaCy: fast and robust models for biomedical natural language processing. In: Demner-Fushman D, Cohen KB, Ananiadou S, Tsujii J, editors. Proceedings of the 18th BioNLP workshop and shared task. Florence, Italy: Association for Computational Linguistics; 2019. p. 319–27.
[43] Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res 2009;37(suppl_2):W170–3. https://doi.org/10.1093/nar/gkp440.
[44] Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, et al. BioPortal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. Nucleic Acids Res 2011;39(suppl_2):W541–5. https://doi.org/10.1093/nar/gkr469.
[45] Kraljevic Z, Searle T, Shek A, Roguski L, Noor K, Bean D, et al. Multi-domain clinical natural language processing with MedCAT: the medical concept annotation toolkit. Artif Intell Med 2021;117:102083. https://doi.org/10.1016/j.artmed.2021.102083.
[46] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: Burges CJC, Bottou L, Ghahramani Z, Weinberger KQ, editors. Advances in neural information processing systems 26: 27th annual conference on neural information processing systems 2013, proceedings of a meeting held December 5–8, 2013, Lake Tahoe, Nevada, United States; 2013. p. 3111–9.
[47] Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of tricks for efficient text classification. In: Lapata M, Blunsom P, Koller A, editors. Proceedings of the 15th conference of the European chapter of the Association for Computational Linguistics: volume 2, short papers. Valencia, Spain: Association for Computational Linguistics; 2017. p. 427–31.
[48] Pennington J, Socher R, Manning C. GloVe: global vectors for word representation. In: Moschitti A, Pang B, Daelemans W, editors. Proceedings of the 2014 conference on empirical methods in natural language processing. Doha, Qatar: Association for Computational Linguistics; 2014. p. 1532–43.
[49] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Guyon I, von Luxburg U, Bengio S, Wallach HM, Fergus R, Vishwanathan SVN, et al., editors. Advances in neural information processing systems 30: annual conference on neural information processing systems 2017; 2017. p. 5998–6008.
[50] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Burstein J, Doran C, Solorio T, editors. Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: human language technologies, volume 1 (long and short papers). Minneapolis, Minnesota: Association for Computational Linguistics; 2019. p. 4171–86.
[51] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.
[52] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw 2005;18(5):602–10. https://doi.org/10.1016/j.neunet.2005.06.042.
[53] Alsentzer E, Murphy J, Boag W, Weng W-H, Jindi D, Naumann T, et al. Publicly available clinical BERT embeddings. In: Rumshisky A, Roberts K, Bethard S, Naumann T, editors. Proceedings of the 2nd clinical natural language processing workshop. Minneapolis, Minnesota, USA: Association for Computational Linguistics; 2019. p. 72–8.

[54] Rossanez A, Dos Reis JC, Torres RdS, de Ribaupierre H. KGen: a knowledge graph generator from biomedical scientific literature. BMC Med Inform Decis Mak 2020;20(Suppl 4):314. https://doi.org/10.1186/s12911-020-01341-5.
[55] Zhang Y, Sheng M, Zhou R, Wang Y, Han G, Zhang H, et al. HKGB: an inclusive, extensible, intelligent, semi-auto-constructed knowledge graph framework for healthcare with clinicians' expertise incorporated. Inf Process Manag 2020;57(6):102324. https://doi.org/10.1016/j.ipm.2020.102324.
[56] Xu J, Kim S, Song M, Jeong M, Kim D, Kang J, et al. Building a PubMed knowledge graph. Sci Data 2020;7(1):205. https://doi.org/10.1038/s41597-020-0543-2.
[57] Maghawry N, Ghoniemy S, Shaaban E, Emara K. An automatic generation of heterogeneous knowledge graph for global disease support: a demonstration of a cancer use case. Big Data Cogn Comput 2023;7(1). https://doi.org/10.3390/bdcc7010021.
[58] German Guideline Program in Oncology (German Cancer Society, German Cancer Aid, AWMF), Breast Cancer, English Long Version 4.4, https://www.leitlinienprogramm-onkologie.de/german-guideline-program-in-oncology, May 2021. aWMF Registration Number: 032/045OL.
[59] German Guideline Program in Oncology (German Cancer Society, German Cancer Aid, AWMF), Colorectal Cancer, English Long Version 2.1, https://www.leitlinienprogramm-onkologie.de/german-guideline-program-in-oncology, Jan. 2016. aWMF Registration Number: 021/007OL.
[60] German Guideline Program in Oncology (German Cancer Society, German Cancer Aid, AWMF), Endometrial Cancer, English Long Version 2.0, https://www.leitlinienprogramm-onkologie.de/german-guideline-program-in-oncology, Sep. 2022. aWMF Registration Number: 032/034OL.
[61] German Guideline Program in Oncology (German Cancer Society, German Cancer Aid, AWMF), Actinic Keratosis and SCC of the Skin, English Long Version 2.0, https://www.leitlinienprogramm-onkologie.de/german-guideline-program-in-oncology, Dec. 2022. aWMF Registration Number: 032/022OL.
[62] German Guideline Program in Oncology (German Cancer Society, German Cancer Aid, AWMF), Malignant Melanoma, German Long Version 3.3, https://www.leitlinienprogramm-onkologie.de/german-guideline-program-in-oncology, Mar. 2023. aWMF Registration Number: 032/024OL.
[63] German Guideline Program in Oncology (German Cancer Society, German Cancer Aid, AWMF), Adult Soft Tissue Sarcomas, English Long Version 1.1, https://www.leitlinienprogramm-onkologie.de/german-guideline-program-in-oncology, May 2022. aWMF Registration Number: 032/044OL.
[64] Borchert F, Lohr C, Modersohn L, Langer T, Follmann M, Sachs JP, et al. GGPONC: a corpus of German medical text with rich metadata based on clinical practice guidelines. In: Holderness E, Jimeno Yepes A, Lavelli A, Minard A-L, Pustejovsky J, Rinaldi F, editors. Proceedings of the 11th international workshop on health text mining and information analysis. Association for Computational Linguistics, Online; 2020. p. 38–48.
[65] Borchert F, Lohr C, Modersohn L, Witt J, Langer T, Follmann M, et al. GGPONC 2.0 - the German clinical guideline corpus for oncology: curation workflow, annotation policy, baseline NER taggers. In: Calzolari N, Béchet F, Blache P, Choukri K, Cieri C, Declerck T, et al., editors. Proceedings of the thirteenth language resources and evaluation conference. Marseille, France: European Language Resources Association; 2022. p. 3650–60.
[66] Ng N, Yee K, Baevski A, Ott M, Auli M, Edunov S. Facebook FAIR's WMT19 news translation task submission. In: Proceedings of the fourth conference on machine translation (volume 2: shared task papers, day 1). Florence, Italy: Association for Computational Linguistics; 2019. p. 314–9.
[67] Schäfer H, Idrissi-Yaghir A, Horn P, Friedrich C. Cross-language transfer of high-quality annotations: combining neural machine translation with cross-linguistic span alignment to apply NER to clinical texts in a low-resource language. In: Naumann T, Bethard S, Roberts K, Rumshisky A, editors. Proceedings of the 4th clinical natural language processing workshop. Seattle, WA: Association for Computational Linguistics; 2022. p. 53–62.
[68] Idrissi-Yaghir A, Dada A, Schäfer H, Arzideh K, Baldini G, Trienes J, et al. Comprehensive study on German language models for clinical and biomedical text understanding. In: Calzolari N, Kan M-Y, Hoste V, Lenci A, Sakti S, Xue N, editors. Proceedings of the 2024 joint international conference on computational linguistics, language resources and evaluation. Torino, Italia: ELRA and ICCL; 2024. p. 3654–65.
[69] Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat 1951;22(1):79–86. https://doi.org/10.1214/aoms/1177729694.
[70] McGuinness DL, Van Harmelen F. OWL web ontology language overview. W3C Recomm 2004;10:1–19.
[71] Johnson AE, Pollard TJ, Shen L, Lehman L-wH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016;3(1):160035. https://doi.org/10.1038/sdata.2016.35.
[72] Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw Pract Exp 1991;21(11):1129–64. https://doi.org/10.1002/spe.4380211102.
[73] Houlsby N, Giurgiu A, Jastrzebski S, Morrone B, De Laroussilhe Q, Gesmundo A, et al. Parameter-efficient transfer learning for NLP. In: Chaudhuri K, Salakhutdinov R, editors. Proceedings of the 36th international conference on machine learning. Proceedings of machine learning research, vol. 97. PMLR; 2019. p. 2790–9.
[74] Zhang Z, Sabuncu M. Generalized cross entropy loss for training deep neural networks with noisy labels. In: Advances in neural information processing systems, vol. 31. Curran Associates, Inc.; 2018. p. 8792–802.
[75] Karypis G, Kumar V. A fast and high quality multilevel scheme for partitioning irregular graphs. SIAM J Sci Comput 1998;20(1):359–92. https://doi.org/10.1137/S1064827595287997.
[76] Baker S, Silins I, Guo Y, Ali I, Högberg J, Stenius U, et al. Automatic semantic classification of scientific literature according to the hallmarks of cancer. Bioinformatics 2015;32(3):432–40. https://doi.org/10.1093/bioinformatics/btv585.
[77] Baker S, Ali I, Silins I, Pyysalo S, Guo Y, Högberg J, et al. Cancer hallmarks analytics tool (CHAT): a text mining approach to organize and evaluate scientific literature on cancer. Bioinformatics 2017;33(24):3973–81. https://doi.org/10.1093/bioinformatics/btx454.
[78] Baker S, Korhonen A. Initializing neural networks for hierarchical multi-label text classification. In: Cohen KB, Demner-Fushman D, Ananiadou S, Tsujii J, editors. BioNLP 2017. Vancouver, Canada: Association for Computational Linguistics; 2017. p. 307–15.
[79] Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100(1):57–70. https://doi.org/10.1016/S0092-8674(00)81683-9.
[80] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144(5):646–74. https://doi.org/10.1016/j.cell.2011.02.013.
[81] Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov 2019;18(1):41–58. https://doi.org/10.1038/nrd.2018.168.
[82] Swanson DR, Smalheiser NR. An interactive system for finding complementary literatures: a stimulus to scientific discovery. Artif Intell 1997;91(2):183–203. https://doi.org/10.1016/S0004-3702(97)00008-8. Scientific Discovery.
[83] Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Cowey CL, Lao CD, et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. N Engl J Med 2015;373(1):23–34. https://doi.org/10.1056/NEJMoa1504030.
[84] Hellmann MD, Paz-Ares L, Caro RB, Zurawski B, Kim S-W, Costa EC, et al. Nivolumab plus ipilimumab in advanced non-small-cell lung cancer. N Engl J Med 2019;381(21):2020–31. https://doi.org/10.1056/NEJMoa1910231.
[85] Motzer RJ, Rini BI, McDermott DF, Redman BG, Kuzel TM, Harrison MR, et al. Nivolumab for metastatic renal cell carcinoma: results of a randomized phase II trial. J Clin Oncol 2015;33(13):1430–7. https://doi.org/10.1200/JCO.2014.59.0703.
[86] McFarlane JJ, Kochenderfer MD, Olsen MR, Bauer TM, Molina A, Hauke RJ, et al. Safety and efficacy of nivolumab in patients with advanced clear cell renal cell carcinoma: results from the phase IIIb/IV CheckMate 374 study. Clin Genitourin Cancer 2020;18(6):469–4764.e. https://doi.org/10.1016/j.clgc.2020.06.002.
[87] Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. N Engl J Med 2015;373(19):1803–13. https://doi.org/10.1056/NEJMoa1510665.
[88] Kang Y-K, Boku N, Satoh T, Ryu M-H, Chao Y, Kato K, et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2017;390(10111):2461–71. https://doi.org/10.1016/S0140-6736(17)31827-5.
[89] Ferris RL, Blumenschein G, Fayette J, Guigay J, Colevas AD, Licitra L, et al. Nivolumab for recurrent squamous-cell carcinoma of the head and neck. N Engl J Med 2016;375(19):1856–67. https://doi.org/10.1056/NEJMoa1602252.
[90] Tawbi HA, Forsyth PA, Algazi A, Hamid O, Hodi FS, Moschos SJ, et al. Combined nivolumab and ipilimumab in melanoma metastatic to the brain. N Engl J Med 2018;379(8):722–30. https://doi.org/10.1056/NEJMoa1805453.
[91] Ansell SM, Lesokhin AM, Borrello I, Halwani A, Scott EC, Gutierrez M, et al. PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. N Engl J Med 2015;372(4):311–9. https://doi.org/10.1056/NEJMoa1411087.
[92] Dimitriou F, Long G, Menzies A. Novel adjuvant options for cutaneous melanoma. Ann Oncol 2021;32(7):854–65. https://doi.org/10.1016/j.annonc.2021.03.198.
[93] Yau T, Park J-W, Finn RS, Cheng A-L, Mathurin P, Edeline J, et al. Nivolumab versus sorafenib in advanced hepatocellular carcinoma (CheckMate 459): a randomised, multicentre, open-label, phase 3 trial. Lancet Oncol 2022;23(1):77–90. https://doi.org/10.1016/S1470-2045(21)00604-5.
[94] Janjigian YY, Shitara K, Moehler M, Garrido M, Salman P, Shen L, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. Lancet 2021;398(10294):27–40. https://doi.org/10.1016/S0140-6736(21)00797-2.
[95] Salles G, Barrett M, Foà R, Maurer J, O'Brien S, Valente N, et al. Rituximab in B-cell hematologic malignancies: a review of 20 years of clinical experience. Adv Ther 2017;34(10):2232–73. https://doi.org/10.1007/s12325-017-0612-x.
[96] Maverakis E, Kim K, Shimoda M, Gershwin ME, Patel F, Wilken R, et al. Glycans in the immune system and the altered glycan theory of autoimmunity: a critical review. J Autoimmun 2015;57:1–13. https://doi.org/10.1016/j.jaut.2014.12.002.
[97] Vidal L, Gafter-Gvili A, Salles G, Dreyling MH, Ghielmini M, Hsu Schmitz S-F, et al. Rituximab maintenance for the treatment of patients with follicular lymphoma: an updated systematic review and meta-analysis of randomized trials. J Natl Cancer Inst 2011;103(23):1799–806. https://doi.org/10.1093/jnci/djr418.
[98] Seymour JF, Kipps TJ, Eichhorst B, Hillmen P, D'Rozario J, Assouline S, et al. Venetoclax–rituximab in relapsed or refractory chronic lymphocytic leukemia. N Engl J Med 2018;378(12):1107–20. https://doi.org/10.1056/NEJMoa1713976.

[99] Coiffier B. Rituximab therapy in malignant lymphoma. Oncogene 2007;26(25):3603–13. https://doi.org/10.1038/sj.onc.1210376.
[100] Kluin-Nelemans H, Hoster E, Hermine O, Walewski J, Trneny M, Geisler C, et al. Treatment of older patients with mantle-cell lymphoma. N Engl J Med 2012;367(6):520–31. https://doi.org/10.1056/NEJMoa1200920.
[101] Choquet S, Leblond V, Herbrecht R, Socié G, Stoppa A-M, Vandenberghe P, et al. Efficacy and safety of rituximab in B-cell post-transplantation lymphoproliferative disorders: results of a prospective multicenter phase 2 study. Blood 2006;107(8):3053–7. https://doi.org/10.1182/blood-2005-01-0377.
[102] Conconi A, Martinelli G, Thiéblemont C, Ferreri AJM, Devizzi L, Peccatori F, et al. Clinical activity of rituximab in extranodal marginal zone B-cell lymphoma of MALT type. Blood 2003;102(8):2741–5. https://doi.org/10.1182/blood-2002-11-3496.
[103] Treon SP, Ioakimidis L, Soumerai JD, Patterson CJ, Sheehy P, Nelson M, et al. Primary therapy of Waldenström macroglobulinemia with bortezomib, dexamethasone, and rituximab: WMCTG clinical trial 05-180. J Clin Oncol 2009;27(23):3830–5. https://doi.org/10.1200/JCO.2008.20.4677.
[104] Ladetto M, Magni M, Pagliano G, De Marco F, Drandi D, Ricca I, et al. Rituximab induces effective clearance of minimal residual disease in molecular relapses of mantle cell lymphoma. Biol Blood Marrow Transplant 2006;12(12):1270–6. https://doi.org/10.1016/j.bbmt.2006.07.007.
[105] Kater AP, Seymour JF, Hillmen P, Eichhorst B, Langerak AW, Owen C, et al. Fixed duration of venetoclax-rituximab in relapsed/refractory chronic lymphocytic leukemia eradicates minimal residual disease and prolongs survival: post-treatment follow-up of the MURANO phase III study. J Clin Oncol 2019;37(4):269–77. https://doi.org/10.1200/JCO.18.01580.
[106] Fritsch K, Kasenda B, Hader C, Nikkhah G, Prinz M, Haug V, et al. Immunochemotherapy with rituximab, methotrexate, procarbazine, and lomustine for primary CNS lymphoma (PCNSL) in the elderly. Ann Oncol 2011;22(9):2080–5. https://doi.org/10.1093/annonc/mdq712.
[107] Hoelzer D, Walewski J, Döhner H, Viardot A, Hiddemann W, Spiekermann K, et al. Improved outcome of adult Burkitt lymphoma/leukemia with rituximab and chemotherapy: report of a large prospective multicenter trial. Blood 2014;124(26):3870–9. https://doi.org/10.1182/blood-2014-03-563627.
[108] Younes A, Romaguera J, Hagemeister F, McLaughlin P, Rodriguez MA, Fiumara P, et al. A pilot study of rituximab in patients with recurrent, classic Hodgkin disease. Cancer 2003;98(2):310–4. https://doi.org/10.1002/cncr.11511.
[109] Hainsworth JD, Litchy S, Barton JH, Houston GA, Hermann RC, Bradof JE, et al. Single-agent rituximab as first-line and maintenance treatment for patients with chronic lymphocytic leukemia or small lymphocytic lymphoma: a phase II trial of the minnie pearl cancer research network. J Clin Oncol 2003;21(9):1746–51. https://doi.org/10.1200/JCO.2003.09.027.
[110] Byrd JC, Waselenko JK, Maneatis TJ, Murphy T, Ward FT, Monahan BP, et al. Rituximab therapy in hematologic malignancy patients with circulating blood tumor cells: association with increased infusion-related side effects and rapid blood tumor clearance. J Clin Oncol 1999;17(3):791. https://doi.org/10.1200/JCO.1999.17.3.791.
[111] Thomas DA, O'Brien S, Bueso-Ramos C, Faderl S, Keating MJ, Giles FJ, et al. Rituximab in relapsed or refractory hairy cell leukemia. Blood 2003;102(12):3906–11. https://doi.org/10.1182/blood-2003-02-0630.
[112] Kilidireas C, Anagnostopoulos A, Karandreas N, Mouselimi L, Dimopoulos M-A. Rituximab therapy in monoclonal IgM-related neuropathies. Leuk Lymphoma 2006;47(5):859–64. https://doi.org/10.1080/14786410500441664.
[113] Rieger M, Österborg A, Pettengell R, White D, Gill D, Walewski J, et al. Primary mediastinal B-cell lymphoma treated with CHOP-like chemotherapy with or without rituximab: results of the Mabthera International Trial Group study. Ann Oncol 2011;22(3):664–70. https://doi.org/10.1093/annonc/mdq418.
[114] Fink-Puches R, Wolf IH, Zalaudek I, Kerl H, Cerroni L. Treatment of primary cutaneous B-cell lymphoma with rituximab. J Am Acad Dermatol 2005;52(5):847–53. https://doi.org/10.1016/j.jaad.2005.01.093.