

Ontologies and Knowledge Graphs in Oncology Research

Marta Contreiras Silva ^{*}, Patrícia Eugénio, Daniel Faria and Catia Pesquita

LASIGE, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal; peugenio@lasige.di.fc.ul.pt (P.E.); dpfaria@fc.ul.pt (D.F.); clpesquita@fc.ul.pt (C.P.)

* Correspondence: mcdsilva@fc.ul.pt

Simple Summary: Cancer is a complex phenomenon and cancer research is increasingly data-rich. Representing this knowledge in a manner that is both human and computer-friendly can help manage and analyze the high volumes of complex cancer data that are created by scientific research and health care. This review looks at the last decade of works on using ontologies—computational representations of knowledge—in cancer, describing their contributions and achievements and charting a path for future research in this area.

Abstract: The complexity of cancer research stems from leaning on several biomedical disciplines for relevant sources of data, many of which are complex in their own right. A holistic view of cancer—which is critical for precision medicine approaches—hinges on integrating a variety of heterogeneous data sources under a cohesive knowledge model, a role which biomedical ontologies can fill. This study reviews the application of ontologies and knowledge graphs in cancer research. In total, our review encompasses 141 published works, which we categorized under 14 hierarchical categories according to their usage of ontologies and knowledge graphs. We also review the most commonly used ontologies and newly developed ones. Our review highlights the growing traction of ontologies in biomedical research in general, and cancer research in particular. Ontologies enable data accessibility, interoperability and integration, support data analysis, facilitate data interpretation and data mining, and more recently, with the emergence of the knowledge graph paradigm, support the application of Artificial Intelligence methods to unlock new knowledge from a holistic view of the available large volumes of heterogeneous data.

Keywords: ontologies; semantic technologies; knowledge graph; cancer; oncology; review



Citation: Silva, M.C.; Eugénio, P.; Faria, D.; Pesquita, C. Ontologies and Knowledge Graphs in Oncology Research. *Cancers* **2022**, *14*, 1906. <https://doi.org/10.3390/cancers14081906>

Academic Editor: David Wong

Received: 8 March 2022

Accepted: 7 April 2022

Published: 10 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Understanding complex phenomena that cannot be modeled purely mathematically is a challenging endeavor transverse to all biomedical research. Ultimately, all boils down to the complex interplay between genes and environment, which manifests in the interactions between the cells in an organism, between host and pathogen, between drug and body. From its genesis, medicine focused on understanding the phenomena which can be generalized between individuals, dating back to the first texts on anatomy by the Ancient Egyptians. Indeed, nomenclature and classification are the first steps towards understanding complex phenomena, and are inextricable from modern medicine, which relies on its precise terminology and its compendium of pathogens, diseases, symptoms, genes and mutations, and drugs and therapies, as well as of the relationships between them.

Over the last three decades, the rise of the digital age and subsequent informatization of clinical records and biomedical research drove the encoding of terminologies, classification schemes and knowledge models into digital machine-readable formats (often captured under the umbrella term ‘ontology’) to promote standardization, support information systems, and enable knowledge discovery. One of the first major efforts to this effect in the biomedical domain was the compilation of the Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT) [1] to support the standardization and interoperability of

clinical information systems and electronic health records. Another major effort was the classification and trans-species standardization of gene functional characteristics under the Gene Ontology (GO) [2]. In the footsteps of these efforts, several hundred other ontologies have been developed for the biomedical domain throughout the years [3], among which we must note the National Cancer Institute Thesaurus (NCIt), a compendium of terminology spanning all aspects of cancer research and health care [4].

More recently, medicine has been witnessing a shift towards the particular, enabled by the decreasing costs of acquiring genetic information, and driven by the understanding that tailored treatments that contemplate the genetic makeup of the patient will likely be more effective and less prone to nefarious side-effects. Cancer is the family of diseases that is benefiting from these precision (or personalized) medicine approaches the most, as despite commonalities, each cancer is genetically unique, and can react very differently to different types of treatment. Moreover, understanding the fine differences between cancer cells and healthy cells can be the key for more successful and less aggressive treatments. Yet the precision medicine paradigm places additional emphasis on having a holistic understanding of the gene–environment interplay in all its manifestations, which requires the integrative analysis of large volumes of heterogeneous data that are individually already complex (e.g., clinical records, medical imaging, transcriptomic data, immunopeptidomic data) [5]. Here too, ontologies have been playing an important role in enabling data integration and facilitating data analysis.

In this article, we review the applications of ontologies in cancer research over the past decade, summarizing published works within this time frame, and categorizing them with respect to their usage of ontologies. Section 2 details core concepts underlying this review article, Section 3 outlines the methodology adopted to conduct the review, Section 4 summarizes both the ontologies reused in the works and the ones created for them, Section 5 reviews and categorizes the aforementioned published works, and Section 6 features our prospects regarding the present and future use of ontologies in cancer research.

2. Background

2.1. Ontologies

The term “ontology” was borrowed from philosophy to computer science to signify a machine-readable formalization of a conceptualization pertaining to a particular domain of knowledge [6]. That is to say, an ontology is a digital artifact that can be interpreted by both humans and computers and which encodes the terminology and the semantic relations between concepts in a given domain. The term “ontology” is often used with some latitude, also encompassing thesauri [7]. While our review of published works adopts the same encompassing perspective, it is important to make a formal distinction between ontologies proper and thesauri due to their different purposes and applications.

Ontologies proper are typically encoded in the Web Ontology Language (OWL), developed by the W3C OWL Working Group [8], which includes various serializations, namely the Open Biomedical Ontologies (OBO) format or the more popular Resource Description Framework (RDF) format in which statements take the form of triples of the form <subject> <predicate> <object>. OWL defines several types of entities which can be used in constructing ontologies, such as: classes, datatypes, object properties, data properties, annotation properties, individuals and literals, among others. All entities in an ontology are identified by an International Resource Identifier (IRI), although in OBO ontologies this is abbreviated to an alphanumeric code. Annotation properties (e.g., *label*) are used to describe the entities in the ontology for human readers, and thus, encode the terminological component of ontologies; they have no semantic value. Individuals (or instances) and literals are data-level entities representing, respectively, concrete objects (e.g., *my heart*) and data values (e.g., “60 beats/min”). The remaining entities are model-level, with classes representing abstract sets of individuals (e.g., *heart*), datatypes representing abstract sets of literals (e.g., *string*), object properties representing relations that can be used to connect individuals (e.g., *part of*) and data properties representing attributes that can be used to describe individuals with

literals (e.g., *has heart beat*). Moreover, OWL defines intrinsic properties that can be used to connect classes (*subclass*, *disjoint*), to assert that individuals belong to a class (*type*), or to constrain object or data properties with respect to the classes that can have them as subjects (*domain*), the classes or datatypes they can take as objects (*range*), or their usage and logic (e.g., *transitive*, *symmetric*). Finally, OWL enables the definition of class expressions, which are classes defined semantically, for example through application of logical operators (*union*, *intersection*, *not*) between classes, or through existential, universal or cardinality restrictions on objects or data properties (e.g., *part of some chest*, which can be applied to class *heart*). OWL ontologies have different degrees of expressiveness depending on which of these features they use, ranging from simple class hierarchies up to semantically intricate knowledge models, which has implications on the possible applications of ontologies. Namely, OWL supports deductive reasoning, that is to say, the use of logical inference to derive non-stated facts from the collection of facts explicitly asserted in the ontology, which will be both harder and more likely to result in non-evident facts the more expressive the ontology is.

Ontologies are often published with only the model-level layer, serving as knowledge models for a given domain, without any data. In some cases, ontologies are used to annotate external data, such as text documents or database entries, without actually instantiating the ontology (e.g., the Gene Ontology is used to annotate genes and proteins, but these are not individuals of the ontology). In other cases, ontologies are developed (or adopted) to serve as the semantic backbone for describing data in a machine interpretable form. When a large number of individuals is represented in a graph that employs an ontology as its schema, we can consider it a Knowledge Graph (KG) [9]. Figure 1 depicts a simplified example of a KG, based on NCI. Classes are represented as circles in a descending hierarchy stemming from the superclass “owl:Thing”, class instances as grey rectangles, and relationships between them are depicted as arrows, corresponding to object properties in an ontology. This KG shows the network around the concepts renal cell carcinoma, MET gene, antineoplastic agent and protein tyrosine kinase, with instances of patient (“Patient X”) and antineoplastic agent (“Sunitinib”).

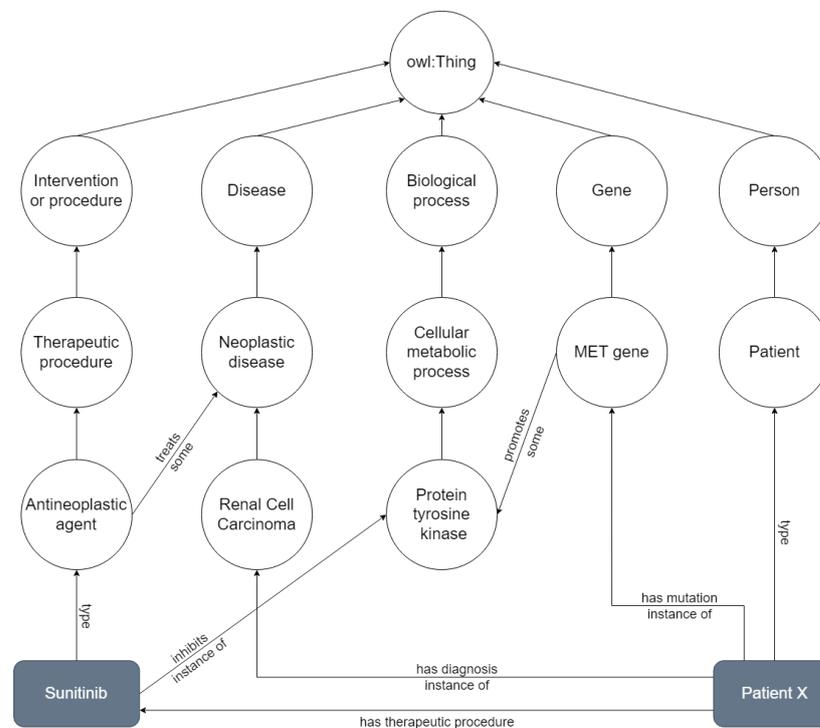


Figure 1. Knowledge graph representing a smaller network that includes renal cell carcinoma, MET gene, antineoplastic agent and protein tyrosine kinase, with instances of a Patient X and the drug Sunitinib. All concepts are derived from the class owl:Thing. Adapted from the NCI.

Thesauri are much simpler than ontologies, and are typically encoded in the Simple Knowledge Organization System (SKOS), which, curiously, is defined on top of OWL. In SKOS, there is no data-level layer, only a model-level layer comprised of *concepts*, their terminological characterization through annotations, and the loose semantic relations between them (*broader, narrower, related*). Thus, thesauri are almost exclusively terminological, and do not enable many of the more sophisticated applications of ontologies proper, namely applications that involve reasoning.

2.2. Ontologies in Cancer Research

The ability to model complex domains is the reason why ontologies are suitable for cancer research and healthcare. For an especially complex disease, such as cancer, that tends towards individual uniqueness and is comprised of various factors and variables, the ability to represent it fully in a manner that can be understood by both clinicians and researchers, and machine algorithms, is invaluable. As such, ontologies represent a unique opportunity to support the domain complexity while allowing for the construction of equally complex solutions that further aid in diagnosing and treating cancer.

At present, there are numerous publicly available biomedical ontologies that have as their principal aim the description of cancer and its characteristics. The National Cancer Institute Thesaurus (NCIt) is perhaps the most often seen. Additionally, there are other biomedical ontologies that, while not directly related to the subject of cancer, are invaluable in its research, for describing fundamental concepts of biology and medicine that form a solid base on which further information stands. Of these, the Gene Ontology (GO) is the most commonly used.

Ontologies in cancer research can be used in varied manners with differing focal objectives. First, despite the fact that cancer-focused ontologies already exist, further conceptualizations of the domain can be developed in the form of new ontologies [10]. These can be reformulations of actual ontologies, updated to include more entities, or even a new, original, ontology to establish a previously less explored section of knowledge. Furthermore, ontologies can be used to annotate data and connect it to the overall context of the domain it pertains to [11]. In this way, for instance, a single value is not simply an isolated value, it is now a single result value from an RNA Sequencing experiment that is placed in a particular section of biomedical knowledge and holds specific relationships to the remaining domain. This annotated value can then be further integrated into developing solutions and their overall context. In addition, ontologies can be directly used as vocabularies to support the organization of data according to known domain information [12]. One objective for this use is, for instance, allowing users to search data that has been annotated using ontologies in a database. Furthermore, NLP methods also need a comprehensive set of terms to use in their application, that then allows for the identification of this information in long-form text, for example [13]. Due to their axiom-based structure, ontologies can support reasoning applications, first to confirm consistency in the ontology and data themselves [14] but also to obtain further inferences from the formal definitions that are established by the ontologies [15]. Lastly, annotation of data with ontologies allows for further use in mining and analyzing this data, for example, with enrichment methods or similarity measures [16,17]. Additionally, there has been an increase in the use of ontology-structured data as input for ML methods, particularly in the biomedical domain with, for example, gene function predictions and clinical decision support systems [18,19].

3. Materials and Methods

3.1. Initial Search and Screening

We carried out an initial search of PubMed [20] on 10 January 2022 with the search query: (“ontology” OR “knowledge graph”) AND “cancer”. We restricted the search to open access articles between 2012 and 2021, setting the search to both *Title* and *Other Term*, and in the case of the “cancer” query, additionally also *MeSH Terms*. We complemented this initial search with a search of Google Scholar [21] on 21 March 2022 with the search

query: (“ontology” OR “knowledge graph”) AND (“cancer” OR “oncology”). The search was constrained to only the title and between 2012 and 2021. The combined results of the two searches were 360 articles.

We screened the resulting lists of articles with the following exclusion criteria: duplicated articles, non-open access articles, and out of scope articles. The latter encompassed articles not related to cancer (misclassification, typos such as oncology/ontology, or mention of only cancer cell lines but not to cancer), articles which did not clearly describe the use of ontologies, and review articles. Additionally, from the Google Scholar results we also excluded theses and non-international and/or non-peer-reviewed publications (which were not an issue in the PubMed search). The screening was conducted in stages, by first examining the title and accessibility of the article, then reading the abstract, and finally reading the article in its entirety. From the initial list of 360 articles, the screening resulted in only 141. A workflow diagram of the whole process can be viewed in Figure 2.

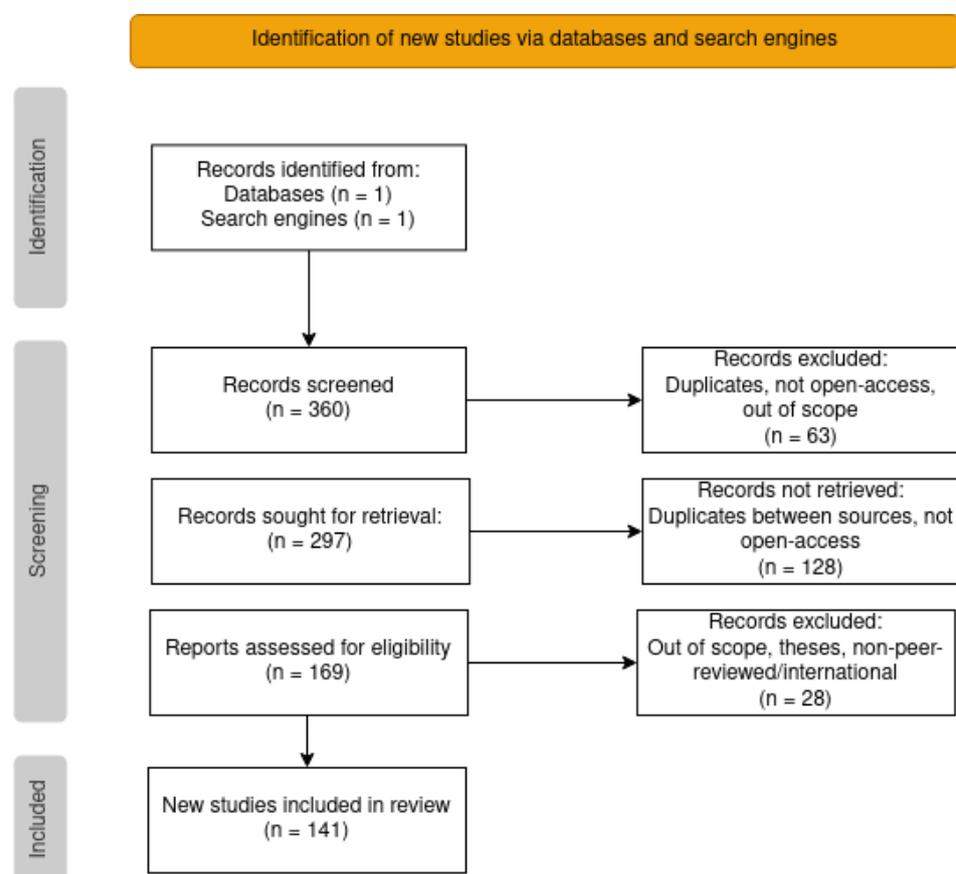


Figure 2. PRISMA flowchart with the steps taken to reach the final list of articles for categorization.

3.2. Categorization

We developed a novel categorization scheme composed of 14 hierarchical categories that describe how the reviewed works employ ontologies and knowledge graphs. These categories fall into two main branches: *Terminology-focused applications* and *Semantic-focused applications*.

The original purpose of clinical and biomedical ontologies was to serve as a source of controlled terminology to tackle the challenges of data-intensive research and clinical practice. As biomedical data production increases and the further it spreads across databases and repositories, there is a reinforced need to connect it to the overall context and to assign the same “meaning” to data that is saved in different and independent places. Ontologies represent the domain concepts in a standardized manner—using a unique identifier for each concept—and placing data into this context increases its own individual reusability by ensuring that it will be understood by anyone, but also, it allows for data from different

sources to be easily matched in their relation to a specific entity.

We have organized **Terminology-focused applications** under four categories:

- **Data Annotation:** ontologies are used to describe data under a common schema, linking data objects to ontology classes that describe them.
- **Data Integration:** ontologies support the integration of different data sets or databases.
- **Database Interface:** ontologies are used to support user interfaces for databases, where labels of ontology classes and relations allow text annotation. These interfaces are notably useful in dealing with medical data, for integration and querying of different knowledge resources.
- **NLP:** ontologies are used as the vocabulary source for Natural Language Processing (NLP) methods, where entities, events or relations in a text are identified through the corresponding ontology labels.

Semantic-focused applications fall under two sub-categories, which are further subdivided:

- **Reasoning:** Automatic reasoners process ontologies' axioms and their formal definitions.
 - **Inference of New Knowledge:** complex reasoning-based queries can reveal novel biological knowledge based on the already defined axioms.
 - **Error Detection:** reasoning applied to check for consistency (or contradictions) in the ontology.
- **Data Mining and Analytics:** ontologies are used to support data mining and analytics tasks.
 - **Semantic Filtering:** ontology-based annotations are used to filter and process data.
 - **Semantic Similarity:** ontology-based annotations are used to compare data entities.
 - **Machine Learning:** ontologies and KGs are explored by machine learning algorithms.
 - **Gene Set Enrichment:** statistical analysis of gene set ontology-annotations.

From the final list, articles were sorted into one or more of the 10 leaf categories according to how the work uses ontologies.

The schema of classification is shown in Figure 3, outlining all the categories and their hierarchical organization used in the following sections.

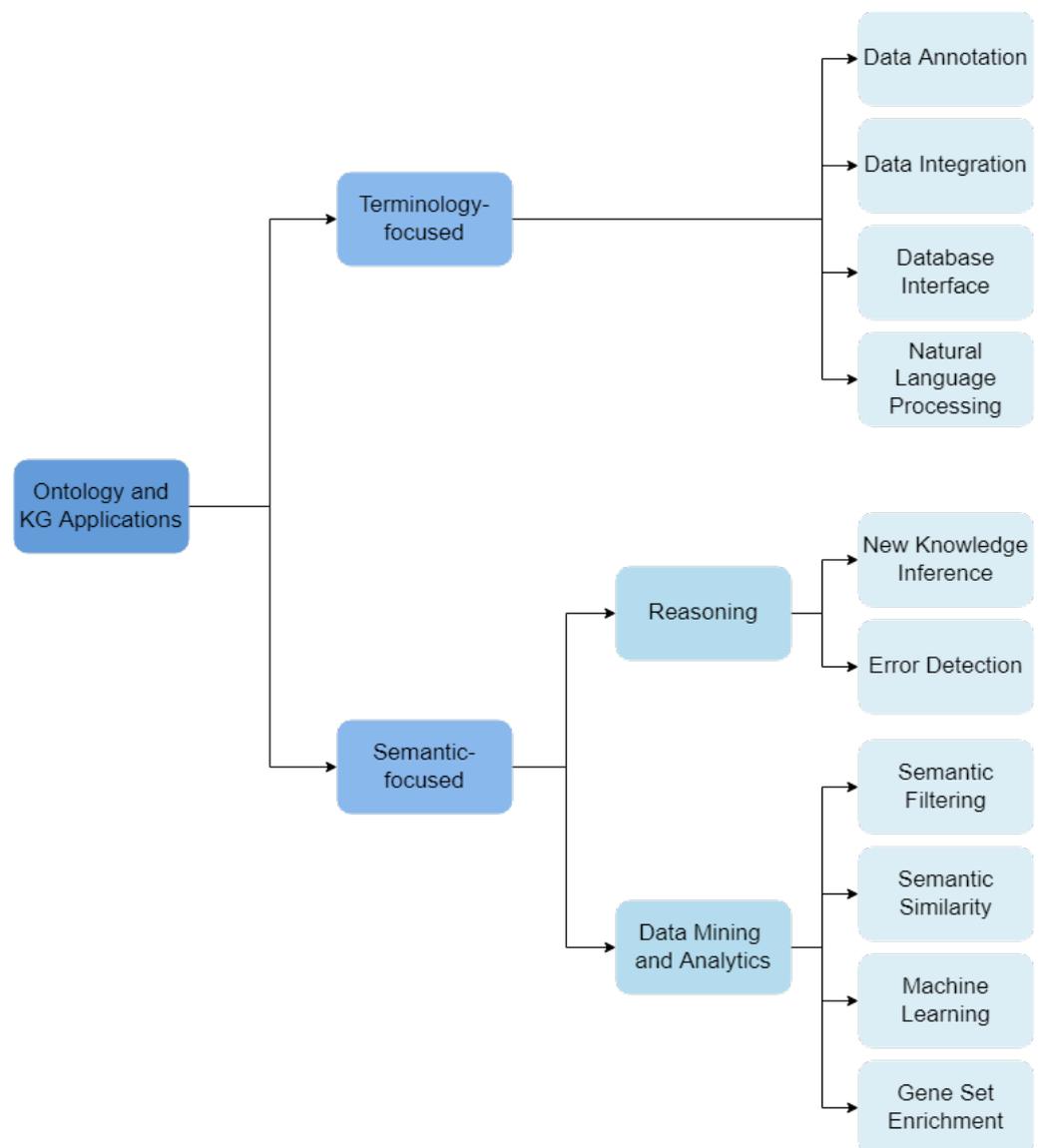


Figure 3. Classification schema for the works included in this articles.

4. Ontologies in Oncology

4.1. Ontologies Used in the Reviewed Applications

One of the ontologies most commonly used in cancer research is, as expected, the National Cancer Institute thesaurus (NCIt) [22], which is a comprehensive ontology devoted specifically to cancer and encompassing both the clinical and research aspects. The SNOMED-CT [1], a broad scope healthcare ontology that has played a key role in systematizing electronic health records, has been used in applications involving clinical data. UMLS [23] is also popular, and is the largest compendium of biomedical terminology, aggregating several healthcare ontologies and vocabularies (namely NCIt and SNOMED-CT) and including mappings between them to enable interoperability.

The Medical Subject Headings (MeSH) thesaurus [24], which are used to index scientific publications, have often been used for bibliographic searches and natural language processing applications. The Disease Ontology (DO) [25] is narrower in scope than the UMLS, focusing only on diseases, but also includes extensive mappings to other healthcare vocabularies (namely MeSH, NCIt and SNOMED-CT).

Other ontologies with narrower scope nevertheless describe aspects that are critical for cancer research. Among them, we include the oncology subset (ICD-O) of the International Classification of Disease (ICD) [26], which categorizes tumors; the Ontology for Biomedical

Investigations (OBI) [27], which aims to describe the terms related to biological and medical investigations; the Cell Line Ontology [28] which classifies cell lines; the Time Event Ontology (TEO) [29], which models temporal expressions and is especially useful when dealing with timed occurrences as healthcare often includes; and the Gene Ontology (GO) [2], which describes gene functions. The latter is the most used ontology of the works reviewed, as it is employed in almost all Gene Set Enrichment applications.

4.2. Ontologies Created for the Reviewed Applications

Several works pertaining to ontologies in cancer research reported on the creation of new ontologies, as summarized in Table 1. The fact that multiple ontologies have been developed in this domain reflects the fact that an ontology is a conceptualization formalized for a particular objective, which represent a given point of view of the underlying domain. As such, despite the existence of several ontologies within the domain, it is often necessary to develop new ontologies for different purposes or to model novel datasets. This is also a testament to the complexity of the cancer research domain, and the several biomedical disciplines it traverses.

One common reason why new ontologies have been developed was to semantically formalize already existing standards. Within this category, Nicholson et al. [30] derived the ENCR core-data ontology from the European Network of Cancer Registries (ENCR) data-validation rules to further support the validation of cancer datasets through an unambiguous formalization and ensure coherence through automatic reasoning logic. Similarly, Zhang et al. [31] also developed the Ontology for the Documentation of vAriable selecTion and daTa sourcE Selection and inTegration (OD-ATTEST) based on a set of reporting guidelines for cancer risk factor variable and data source selection to serve as a standardization of data models. With the aim of describing cancer cells and capturing the properties of tumorigenesis, Rasmussen et al. [32] created the OncoCL. Jusoh et al. [33] built a breast cancer ontology using a hybrid approach to help integrate cancer data from different sources into a single database. Furthermore, in the breast cancer domain, Myneni et al. [34] created OntoMama to assist medical students and professionals. Maly et al. [35] created an ontology of standardized cancer treatments that maps to standard nomenclatures based on HemOnc. Dinakarandian et al. [36] created the Temporal Ontology for Comparing the Survival Outcomes (TOCSOC), a temporal ontology of survival outcome measures of clinical trials in oncology, reusing numerous ontologies. PCLiON is a new standardized lifestyle ontology created by Chen et al. [37], reusing multiple ontologies to harmonize the different data types related to prostate cancer. Looking to generalize the pattern of definitions to correctly classify all gastrointestinal tumor configurations, Herrmann et al. [38] developed their ontology based on BioTopLite2.

Another common reason for ontology development is to create a semantic model for existing datasets. For example, Esteban-Gil et al. [39] used data from a cancer registry relational database to develop a semantic model that can then be queried to analyze patient data through ontology-driven search. The NeoMark European project [11] also developed a specialized ontology for their data content, the NeoMark Ontology, built from its existing database. Amith et al. [40] used a lightweight Open Information Extraction (OIE) tool to extract semantic information from MedlinePlus and seed a knowledge-base. To represent obesity related cancer information, organize and allow data querying, Elhefny et al. [41] reused DOID to develop the Fuzzy Ontology for Obesity-Related Cancer (FOORC).

Ontologies have also been developed to harmonize the communication between clinicians and patients, namely by exploiting social media. Tapi Nzali et al. [42] built a Consumer Health Vocabulary (CHV) in french for breast cancer by mapping terms from forum messages and standardized medical terms. Lee et al. [43] created an ontology to understand information needs and emotions regarding cancer from social media. Myneni et al. [34] developed the Profile Ontology for Cancer Survivors (POCS) to facilitate the fast development of patient-engaging mobile apps.

Supporting the development of applications to aid diagnosis and treatment by providing a semantic representation of existing knowledge has been another major motivation for the development of new ontologies. For hepatocellular carcinoma (HCC), Messaoudi et al. [44] developed the Ontology of Hepatocellular Carcinoma (OntHCC) to support their application in the detection of nodules in medical imaging, while Gurcan et al. [45] created the Quantitative Histopathological Imaging Ontology (QHIO) to represent both data and methods used in clinical imaging and analysis. Boeker et al. [46] developed TNM-O to represent the Tumor–Node–Metastasis (TNM) classification of malignant tumors and Tagliaferri et al. [47] developed the ENT COBRA (COnsortium for BRachytherapy data Analysis) ontology to standardize data collection for head and neck cancer patients that have been specifically treated with interventional radiotherapy, while SKIN-COBRA has a similar objective for non-melanoma skin cancer patients with the same treatment [48]. With a very focused aim, Oyelade et al. [14] proposed Breast Cancer Fuzzy Ontology (BCFO) to address vagueness in the domain of this specific cancer. Mahmoodi et al. [49] manually created the Gastric Cancer Ontology (GCO) with experts to support the extraction of association rules. Gao et al. [50] constructed a treatment-based cancer ontology using a Bayesian derivation that focuses on cancer reclassification and drug inference. For lung cancer, Sesen et al. [51] constructed the LUCADA ontology to use with the clinical decision support application Lung Cancer Assistant. In the domain of bladder cancer, Barki et al. [52] developed an ontology to predict side effects caused by treatments.

Finally, ontologies have been developed for enabling data interoperability and integration, a pressing demand given the increasing volume of heterogeneous data sources available for cancer research. To study the connection between various risk factors and cancer survival, Zhang et al. [53] created the Ontology for Cancer Research Variables (OCRV) reusing some existing resources, and then linked it to a data integration pipeline. Lin et al. [10] developed the Cancer Care Treatment Outcome Ontology (CCTOO) that organizes high-level oncology treatment end points into four domains: cancer treatment, health services, physical, and psycho-social health-related concepts. To aid drug target prediction, Tao et al. [54] created the CRC ontology, reusing PharmGKB. Balasubramanian et al. [55] reused BFO and created the Ontology of Cancer Related Social-Ecological Variables (OCRSEV) to enable data integration and posterior association between Social-Economical Factors and health outcomes in cancer. Aiming to increase interoperability between data sources to allow the creation of Big Data studies that involve several treatment centers, Bibault et al. [56] created the Radiation Oncology Structures (ROS) based on FMA. To also support integrative data analysis in cancer outcomes research, Zhang et al. [57] created the Ontology for Documentation of Variable and Data Source (ODVDS) reusing BFO. Divakar et al. [58] developed CCOWL in order to analyze patient’s cytological tissue images of cervical cancer. Additionally, RiskExplorer was created by Daowd et al. [59] to represent causal associations between the incidence of breast cancer and risk factors.

Some works also report on updates or extensions to existing ontologies, motivated by some of the same objectives for creating new ontologies. Serra et al. [60] developed the Cancer Cell Ontology (CCL) as an extension of the Cell Ontology (CL), to serve as a formal representation of immunophenotyping cell types from hematologic malignancies. The Cell Line Ontology (CLO) was updated and extended by Ong et al. [61] to include NIH Common Fund Library of Integrated Network-based Cellular Signatures (LINCS) cell lines, with a subset LINCS-CLOview being generated. Campbell et al. [62] created additional concepts for Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) that unify it with Logical Observation Identifier Names and Codes (LOINC) for colorectal and breast cancer.

Table 1. New ontologies.

Ref	Objective	Ontology Name	Domain	Reused Ontologies	Language
[51]	Model lung cancer for the clinical decision support application Lung Cancer Assistant	LUCADA ontology	Clinical	SNOMED-CT	OWL
[33]	Use a hybrid approach to build a breast cancer ontology	N/A	Breast Cancer	N/A	OWL
[32]	Describe cancer cells and capture the properties of tumorigenesis	OncoCL	Cell Lines	CL, UBERON, BTO, Pathway Ontology, PATO, CPO, SO	OWL
[11]	Represent the project domain and link the NeoMark data to other domains	NeoMark ontology	Clinical	BFO, RO	OWL
[50]	Cancer reclassification and drug inference	N/A	Farmacology	N/A	N/A
[54]	Drug target prediction	CRC ontology	Colorectal Cancer	PharmGKB	OWL
[63]	Assist medical students and professionals in the breast cancer domain	OntoMama	Clinical	N/A	N/A
[34]	Development of an ontology-driven survivor engagement framework for mobile apps	POCS	Social	FOAF	OWL
[46]	Creation of TNM-O	TNM-O	Anatomical	FMA, BioTopLite 2	OWL
[41]	Represent obesity-related cancer (ORC) ontology to organize information and allow data querying	FOORC	Obesity Related Cancer	DOID	OWL
[49]	Extraction of association rules from large datasets on gastric cancer patients	Gastric cancer ontology	Clinical	N/A	N/A
[55]	Aid data integration; enable association between SE variables and health outcomes	OCRSEV	Social-Ecological Factors	BFO	OWL
[45]	Interoperability across quantitative histopathological imaging data sets	QHIO	Imaging	OBI	OWL
[39]	Design of a semantic model for local cancer registries	N/A	Epidemiology	SIO, OBI	OWL
[40]	Development of ontologies for the public health domain	N/A	Public Health	N/A	OWL
[61]	Understand cellular responses to different perturbations	LINCS-CLOview	Cell Lines	CLO	OWL
[53]	Integrate heterogeneous datasets	OCRV	Cancer Outcomes	BFO, NCIt, TEO	OWL
[47]	Define a specific terminological system to standardized data collection for head and neck cancer patients	ENT COBRA ontology	Clinical	N/A	N/A
[10]	Use structured knowledge representation with concepts of treatment end points	CCTOO	Clinical	NCIt, CTCAE	OBO
[62]	Represent the data elements identified by the synoptic worksheets of College of American Pathologists	SNOMED CT observable ontology	Clinical	SNOMED CT, LOINC	N/A
[35]	Create a standardized hierarchic ontology of cancer treatments, mapped to standard nomenclatures	N/A	Cancer Treatments	HemOnc	OWL
[56]	Increase interoperability between data sources to allow the creation of Big Data studies involving several treatment centers	ROS	Radiation Oncology	FMA	OWL
[36]	Create temporal ontology of survival outcome measures of clinical trials in oncology	TOCSOC	Clinical	EFO, CCTOO, IOBC, NCIT	OWL
[60]	Provide an ontological representation of immunophenotyping cell types found in hematologic malignancies	CCL	Hematologic Malignancies	CL	OWL
[42]	Semi-automatic development of CHV for breast cancer	MuEVo	Clinical	MeSH, MedDRA, SNOMEDint	SKOS
[44]	Offer ontology-based approach modeling HCC tumors	OntHCC	Liver Cancer	N/A	OWL

Table 1. Cont.

Ref	Objective	Ontology Name	Domain	Reused Ontologies	Language
[57]	Support integrative data analysis in cancer outcomes research	ODVDS	Risk Factors	BFO	OWL
[58]	Cytological tissue image analysis of cervical cancer	CCOWL	Cervical Cancer	N/A	OWL
[31]	Standardize the terminology used in the selection and integration steps of RF variables and data sources	OD-ATTEST	Risk Factors	BFO, others in NCBO (not specified)	OWL
[48]	Standardize data collection for non-melanoma skin cancer patients treated with brachytherapy	SKIN-COBRA ontology	Clinical	N/A	N/A
[43]	Analyze social media data to identify information needs and emotions related to cancer	N/A	Social	LCO, BCO, GCO, SOSW	N/A
[37]	Solve the heterogeneity and diversity of different data types related to prostate cancer by establishing a standardized lifestyle ontology	PCLiON	Risk Factors	NCIT, WordNet, SNOMED CT, The Cochrane Library, FooDB, CheBI	OWL
[59]	Build a knowledge graph that represents causal associations between incidence of breast cancer and risk factors	RiskExplorer	Clinical	UMLS	N/A
[30]	Facilitate the integrity and maintenance of ENCR core data set.	ENCR core-data	Epidemiology	N/A	OWL
[14]	Minimizing vagueness in the formalization of medical knowledge	BCFO	Clinical	DO	OWL
[52]	Predict side effects of bladder cancer treatments	N/A	Bladder Cancer	N/A	OWL
[38]	Provide a generalizing pattern of more concise definitions to correctly classify all tumor configurations	N/A	Gastrointestinal Tumors	BioTopLite2	N/A

5. Ontologies and Knowledge Graph Applications in Cancer Research

The categorization of the reviewed works relied exclusively on the information presented in the article and no additional searches were conducted to obtain further details. The information gathered in the process of categorization is presented in Tables 2–4 organized into columns relevant to each category.

5.1. Terminology-Focused Applications

Table 2 describes the articles from these categories, according to the ontologies and data employed and cancer type.

5.1.1. Data Annotation

Most *Data Annotation* works use existing ontologies, such as NCIt, Medical Subject Headings (MeSH), and GO, among others, but there are quite a few instances where new ontologies were created to address specific needs.

In breast cancer, Zhu et al. [15] used the semantic modeling of drugs from PharmGKB to infer repositioning. As cancer care is a continuum, Myneni et al. [34] developed an ontology-driven adolescent and young adult survivor engagement framework, to aid the development of mobile apps for information dissemination about treatments and effects of cancer therapies provided through Survivorship Care Plans. Esteban-Gil et al. [39] created a semantic representation of data from a cancer registry database, that results in a model that can be reused and extended to other registries and is capable of supporting further semantic queries on patient profiles that are crucial to research. Yan et al. [13] used NLP tools and an enriched ontology from the MeSH graph to develop UDT-RF, aiming to categorize literature into the corresponding cancer hallmarks through text annotation by estimating the information of interest contained. Using the Time Event Ontology (TEO),

Chen et al. [64] semantically modeled the time component of Common Data Elements (CDEs) that, in capturing clinical research data, highly benefit from a temporal dimension. For HCC, in addition to developing OntHCC, Messaoudi et al. [44] used it to help in the classification of the staging of tumors that are detected in medical imaging.

5.1.2. Data Integration

A vital part of having large amounts of data in differing repositories and/or originating from various sources is integrating them into a single cohesive semantic representation.

Salvi et al. [11] used a focused ontology to annotate their data from various sources that they have compiled in their relational database concerning Oral Squamous Cell Carcinoma (OSCC). The web-based application LncRNA Ontology was developed by Li et al. [65] from the results of their approach to predict probable functions of most human long non-coding RNAs (lncRNAs). Focusing on reusability and comparison of different sources, Milian et al. [66] developed a method that automatically structures clinical trial eligibility criteria from text. Kim et al. [67] used a graph-based framework that integrates multi-omics data with genomic knowledge in order to improve predictions of clinical outcomes. Wu et al. [68] developed a focused view of the DO from a variety of cancer datasets of various sources in order to enable pan-cancer analysis across datasets. Bona et al. [69] focused on accessibility of non-image data from the Cancer Imaging Archive (TCIA) by using ontologies to integrate it into semantic representations. In their two papers, refs. [53,70] also created a focused ontology, OCRV, but then used it with a data integration pipeline for data in relational databases with the aim of making the semantic relationships explicit and clear across different sources. Hasan et al. [71] developed a prototype of a KG that semantically encodes cancer registry data with the expressed aim of enabling the connection to third-party data to further enable new research. Li et al. [72], on the other hand, constructed a KG by first extracting knowledge triples from available data and then using these to construct a network for healthcare professionals that allows them to traverse this contextualized knowledge. Tao et al. [12] developed a web-based system called Interactive Mapping Interface (IMI) to first map the data dictionary in use by the North American Association of Central Cancer Registries (NAACCR) to the NCI with the final goal of facilitating the dissemination and reuse of North American cancer registries data. Chen et al. [73] established a consensus knowledge for cancer hallmarks using functional annotations and gene set overlap, again aiming towards enabling the ability to compare data from different sources.

5.1.3. Database Interfaces

One application reported in the articles lies on ontology-based annotations to create user interfaces for databases, where labels of ontology classes and relations allow text annotation. These interfaces are notably useful in dealing with medical data, for integration and querying of different knowledge resources.

Works within this category that have already been mentioned before are Myneni et al. [34] and Esteban-Gil et al. [39] from data annotation, and Milian et al. [66], Hasan et al. [71], and Tao et al. [12] from data integration. Sesen et al. [51] used a lung ontology with the clinical decision support application Lung Cancer Assistant to categorize patients and produce treatment recommendations. González-Beltrán et al. [74] aimed to ease queries over cancer research data, by extending an existing tool, caGrid [75], with additional services, its domain metadata consisting of ontology-based annotations associated with the structural information of each incorporated data source. In lung cancer, circ2GO is a database developed by Lyu et al. [76] that holds information about the functional annotation of circular RNAs by integrating GO information for all genes in their dataset.

5.1.4. Natural Language Processing

Natural Language Processing (NLP) is also a field that can benefit from the use of a standardized organization of knowledge and terms. The works by Milian et al. [66] and Yan et al. [13] have been mentioned in previous sub-categories. In the case of Tapi

Nzali et al. [42], the goal was to use their own french CHV of non-experts' expressions for breast cancer and compare them to biomedical terms used by health care professionals. Directed toward a social scope, Lee et al. [43] created an ontology from a social media crawler and NLP, to evaluate social media data and understand information needs and emotions related to cancer.

Table 2. Terminology-focused applications.

Ref	Summary	Ontologies	Data	Tag	Cancer Type
[51]	Ontology for a clinical decision support system to produce treatment recommendations	SNOMED-CT, New ontology	N/A	Database Interface	Lung
[74]	Ontology-based querying for cancer research data	NCIt	N/A	Database Interface	Various
[77]	Mining of genetic marker data in a journal	SNOMED-CT, HUGO	NEJM	NLP	Various
[11]	Automatic translation of NeoMark relational database	BFO, RO, OBI, OGMS, HDO	NeoMark database	Data Integration	OSCC
[15]	Manual identification and inference of associations between breast cancer drugs	New ontology	PharmGKB, NCI	Data Annotation	Breast
[65]	Genome-wide functional predictions of lncRNAs	GO	Gencode, Ensembl, ENCODE project, LncRNA Ontology	Data Integration	Various
[66]	Extraction of semantic entities in eligibility criteria and annotation	UMLS	CTG	Data Integration, Database Interface, NLP	Breast
[34]	Development of an ontology-driven survivor engagement framework for mobile apps	FOAF	N/A	Database Interface, Data Annotation	POCS
[67]	Prediction of clinical outcomes from a graph-based approach with multi-omics and genetic data	GO	TCGA	Data Integration	Ovarian
[68]	Development of a focused view within the DO from cancer datasets	DO	COSMIC, TCGA, ICGC, TARGET, IO, EDRN	Data Integration	Various
[39]	Development of a platform for analysis and visualization of data	ICD10, ICD-O-3, TNM staging, SIO, OBI, OQuaRE	NCRI	Data Annotation, Database Interface	Various
[13]	Automatic annotation of cancer hallmarks on biomedical literature	MeSH	N/A	Data Annotation, NLP	Various
[70]	Connection of predictors with cancer survival with a use-case ontology	OCRV	FCDS 2000 U.S. census, BRFS	Data Integration	Various
[69]	Data integration of several databases with ontologies to enable querying of patient data	DO, UBERON	TCIA, TCGA, LIDC-IDRI, Head-Neck-PET-CT	Data Integration	Various
[78]	Construction of OCRV based on data analysis needs	NCIt, TEO, ICD-O-3, ICD-9-CM	UF Health CCCA, FCDS, ATSDR, USCB, BRFS, County Health Ranking & Roadmaps	Data Integration	Various
[64]	Manual representation of semantic temporal components of CDEs	TEO	NCI, caDSR	Data Annotation	Various
[44]	Ontology built following the MethOntology methodology [79]	DICOM	University Hospital of Clermont-Ferrand	Data Annotation	HCC
[42]	Semi-automatic development of CHV for breast cancer	INDC dictionary	N/A	NLP	Various
[71]	KG of cancer registry data, with data analysis and visualization	New ontology	LTR	Data Integration, Database Interface	Various
[43]	Development of an ontology to understand information needs and emotions	LCO, BCO, GCO, SOSW	N/A	NLP	Various

Table 2. Cont.

Ref	Summary	Ontologies	Data	Tag	Cancer Type
[72]	KGHC is a KG constructed from clinical data available publicly	UMLS	PubMed, UpToDate, CTG, SemMedDB	Data Integration	HCC
[76]	Functional annotation of circRNAs obtained from sequencing lung cell lines	GO	Lung cell lines sequencing data	Database Interface	Lung
[12]	IMI is a web-based system that creates mappings from the NAACCR data dictionary to NCI	NAACCR data dictionary, NCI	KCR	Data Integration, Database Interface	Various
[73]	Comparative analysis of cancer hallmark mapping strategies	GO	MSigDB, KEGG, cancer hallmark mapping schemes, TCGA	Data Integration	Various

5.2. Semantic-Focused Applications

5.2.1. Formalized Definitions and Axioms: Reasoning with Ontologies

In the works collected, reasoning is applied to the inference of new knowledge from ontologies or error detection is also reported, as summarized in Table 3. The most common way to access and use reasoners in the reviewed papers consisted of using Protégé, an ontology editor, while creating or editing ontologies, due to ease of access [80].

There are works that use reasoners to infer new knowledge from semantically annotated data and/or established rules. Alfonse et al. [81] used FaCT++ to determine the type and stage of a patient's cancer in order to recommend treatments. Zhu et al. [15] used a rule-based Description Logic (DL) unnamed OWL reasoner to infer additional associations in pathways, drugs, genes and diseases for 18 breast cancer drugs from the ontological representation of the PharmGKB pathway data file. Moreover, using the same ontological representation of PharmGKB, Tao et al. [82] used Pellet to predict new targets for therapy development. Mahmoodi et al. [49] derived association rules from the GCO and patient data using a modified version of an Apriori algorithm, to establish system-wide associations between events in text through large-scale text mining. Barki et al. [52] predicted side effects of treatments for bladder cancer with Pellet. Nicholson et al. [83] used reasoners to signal rule violations in the validation of international rules for multiple primary tumors.

Reasoners can also be used to detect errors in the ontologies or models that have been built. Works by Barki et al. [52], and Nicholson et al. [30,83] were described above. Herrmann et al. [38] aimed at providing a generalizing pattern to classify tumors. Boeker et al. [46] used HermIT DL in their TNM Ontology to evaluate its soundness. Oyelade et al. [14] focused on addressing the issue of vagueness in breast cancer ontology (BCO).

Table 3. Semantic-focused applications: reasoning with ontologies.

Ref	Objective	Input Ontologies	Reasoner	Tag	Cancer Type
[81]	Determine cancer type and stage of the patient to recommend treatments	LuCO, BCO, LCO	FaCT++	New Knowledge Inference	Various
[15]	Identification of new indications for existing drugs	New ontology	Automated semantic inference (Protégé)	New knowledge Inference	Breast
[82]	Prediction of new drug targets	New ontology	Pellet (Protégé)	New knowledge Inference	Colorectal
[49]	Extraction of association rules from large datasets on gastric cancer patients	GCO	Apriori algorithm	New Knowledge Inference	Gastric

Table 3. Cont.

Ref	Objective	Input Ontologies	Reasoner	Tag	Cancer Type
[38]	Provide a generalizing pattern of more concise definitions to correctly classify all tumor configurations	New ontology	HermiT DL (Protégé)	Error Detection	Various
[46]	Creation of TNM-O	FMA, BioTopLite 2	HermiT DL	Error Detection	Various
[52]	Predict side effects of bladder cancer treatments	New ontology	Pellet (Protégé)	New knowledge Inference + Error Detection	Bladder
[83]	Signal rule violations in a validation process of multiple primary tumors international rules	ICD-O-3	FaCT++, HermiT	New knowledge Inference + Error Detection	Multiple primary tumors
[30]	Facilitate the integrity and maintenance of ENCR core data set	New ontology	FaCT++ (Protégé)	Error Detection	Various
[14]	Minimizing vagueness in the formalization of medical knowledge	DO	Fuzzy DL, HermiT/Pellet (Protégé)	Error Detection	Breast

5.2.2. Mining and Analyzing Multimodal Data with Ontologies

By far the majority of the works reviewed, fall into the category of mining and analyzing, as can be partially observed by Table 4 and the additional 72 gene set enrichment articles not present in it that belong to this category. The use of ontologies in cancer research has undoubtedly opened a new avenue in data analysis, where different methodologies (or combinations of) are used to achieve the most varied goals to derive meaning from large quantities of data.

One of the applications reported in data analysis and mining is semantic filtering [84]. The annotation of data with its semantic concepts enables the use of those same concepts to filter data. Chen et al. [85] used biomedical ontologies to guide a set of sequential filtering steps with the objective of predicting microRNAs related to the regulation of glucocorticoid resistance in the specific case of pediatric acute lymphoblastic leukemia (ALL). In another case, users can use the Semantic Web platform developed by Esteban-Gil et al. [39] to run semantic queries over the annotated data and visualize the results in different ways.

An additional use is similarity measuring [86], where the distance between items is measured by the overlap in meaning, to discern what concepts (and therefore their data) are closer or further apart. For example, Modules and Gene Ontology-based Gene Prioritization, developed by Su et al. [17], uses fuzzy similarity for cancer-related gene prioritization.

One of the main approaches used to analyze large amounts of biomedical data is the employment of ML techniques on data that has been semantically annotated. With the evolution of AI algorithms, researchers have been increasingly able to pose more complex questions and use various methodologies to obtain their answers, which is easily observed from the variety of methods and objectives in the articles reviewed. UMVMO-select is a Unsupervised Multi-View Multi-Objective clustering-based gene selection approach developed by Acharya et al. [87] that uses functional annotation to identify gene markers. Su et al. [88] used an ML method over functionally annotated genetic information to look into the immunofunctionomes of ovarian clear cell carcinoma (OCCC). Chen et al. [64] predicted drug synergy using a deep belief network over genetic expression and an ontological profile of genes built from literature (Ontology Fingerprints). For clinical decision support, Shen et al. [19] outlined an architecture that combines Case-Based Reasoning (CBR) with a Multi-Agent System (MAS) to provide treatment suggestions. [77] used the Multi-threaded Clinical Vocabulary Server (MCVS) NLP engine to mine data related to genetic markers from the New England Journal of Medicine (NEJM), with the aim of further supporting the role of inflammation in cancer. To predict drug targets, Tao et al. [82] used a combination of ontology reasoning with network-assisted gene ranking over an ontology that represents

PharmGKB data. Althubaiti et al. [18] used neuro-symbolic feature learning over several ontologies to predict cancer driver genes. Deep GONet, developed by Bourgeais et al. [89], is a self-explainable deep learning model where each biological function is represented by a neuron, that can be used to predict phenotypes. Gao et al. [50] obtained drug inference results from a treatment-based cancer ontology obtained by Bayesian derivation. Comparing the same method with and without ontologies, Min et al. [90] used a rule learning system to predict patients' ability to perform activities of daily living. Furthermore, to predict cervical cancer cells from cytological tissue images, Divakar et al. [58] used deep neural networks (DNN) on their developed ontology. Salvi et al. [11] used a variety of classifiers—Bayesian networks, artificial neural networks (ANN), support vector machines (SVMs), decision trees and random forests—in a data analysis model of their NeoMark system that holds its own semantic model. By comparing several different models, Yan et al. [13] reached an approach that outperforms the others that uses ontological features with a combined use of United Decision Trees and Random Forest algorithms. González-Beltrán et al. [74] developed a system for ontology-based queries over the caGrid infrastructure than can be reused with other service-oriented and model-driven infrastructures. Xi et al. [91] leverages KG embeddings for tolerating missing data from breast cancer clinical ultrasound reports. Using graph attention networks (GAT), Zhang et al. [92] developed a method for real-time inference on a lung KG, using a new ontology.

However, in the end, the most common approach to the use of ontologies in the analysis of biomedical data was the application of GO in Gene Set Enrichment Analysis (GSEA) [16,53,73,93–159]. GSEA statistically compares set of genes that share biological characteristics and interprets their expression data in light of on whether they differ across defined phenotypes [160] and as such is commonly used in biomedical research to, for example, establish candidate genes for further studies.

Table 4. Semantic-focused applications: mining and analyzing multimodal data with ontologies.

Ref	Objective	Method	Input Ontologies	Input Data	Tag	Cancer Type
[77]	Mining of genetic marker data in a journal	MCVS NLP engine	SNOMED CT, HUGO	NEJM	ML	Various
[74]	Ontology-based querying for cancer research data	Construction of a OWL Generation facility	NCIt	caGrid	ML	Various
[11]	Represent the project domain and link the NeoMark data to other domains	Bayesian Networks, ANN, SVMs, Decision Trees, Random Forests	BFO, RO, OBI, OGMS, HDO	N/A	ML	OSCC
[50]	Cancer reclassification and drug inference	Vazquez Bayesian clustering algorithm	N/A	HemOnc.org	ML	Various
[19]	Ontological application in Clinical Decision Support	CBR and MAS	UML	Patient Health Records	ML	Gastric
[82]	Prediction of new drug targets	KEGG functional PharmGKB drug annotation. Network neighborhood modeling ranking	New ontology, ATC	PharmGKB, GAD, CGC, OMIM, NCI, DrugBank, TTD	ML	Colorectal
[39]	Design of a semantic model for local cancer registries	Ontology-driven search filters and aggregates properties of interest	ICD10, ICD-O-3, TNM staging, SIO, OBI, OQuARE	NCRI	Filtering	Various
[90]	Discover patterns related to the patients' ability to perform daily living activities	AQ21—multi-task ML and data mining system	UMLS	Surveillance, Epidemiology, and End Results—Medicare HOS	ML	Various

Table 4. Cont.

Ref	Objective	Method	Input Ontologies	Input Data	Tag	Cancer Type
[13]	Automatic annotation of cancer hallmarks on biomedical literature	United Decision Tree and Random Forest	MeSH	Pubmed abstracts	ML	Various
[85]	Prediction of microRNA related to glucocorticoid resistance	Manual background literature search. Semantic searches in resulting subset	OMIT, NCRO, MeSH	PubMed	Filtering	Pediatric ALL
[17]	Cancer-related gene prioritization	Fuzzy similarity	GO	GSEA website, TCGA, SNP4Disease AstraZeneca-Sanger	Similarity	PAC, Breast
[161]	Predict drug synergy in cancer treatment	Stacked Restricted Boltzmann machine	GO, Ontology Fingerprints	Drug Combination Prediction Challenge, GDSC, KEGG Uniprot, MGI	ML	Various
[18]	Identification of cancer driver genes with role distinction	Neuro-symbolic deep learning on semantic knowledge representation on genetic information	CMPO, GO, MP	database, Mutational Cancer Drivers Database, CPD	ML	Naso-pharyngeal, Colorectal
[87]	Identification of relevant, expression data non-redundant cancer gene markers	Unsupervised Multi-View Multi-Objective clustering	GO	Gene expression datasets from own lab	ML	Prostate, DLBCL, FL
[58]	Predict cervical cancer cells from cytological tissue images	DNN	New ontology	hospital cervical cancer data, kaggle data repository	ML	Cervical
[88]	Complement system role inference from immunofunctionome analysis	SVMs	GO	GEO database	ML	OCCC
[89]	Cancer detection based on gene expression data	Multilayer Perceptrons	GO	Affymetrix HG-U133Plus2 chip arrays, TCGA	ML	Various
[91]	Tolerating data missing in breast cancer diagnosis from clinical ultrasound reports	KG embeddings	BI-RADS	Ultrasound reports	ML	Breast
[92]	Real-time inference on a lung KG	GAT	New ontology	KEGG, Uniprot, DrugBank, TCGA	ML	Lung

Of the 141 papers selected in this systematic review, 72 employed gene set enrichment in some manner. Of these, 21 only used GO, and 48 used it in conjunction with other resources, of which Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database was more common with 45 articles, followed by REACTOME pathway database with 3. Of this application, we have the example of Tian et al. [131] that profiled the transcriptome of gastric cancer patients and used the enrichment to confirm the annotation of genes with digestive system process, secretion and digestion. She et al. [109] used GO and KEGG in an enrichment analysis with the overall objective of finding the importance of C reactive protein and its interactors in HCC. Moreover, developing research in the same cancer, Agioutantis et al. [16] also used enrichment with both GO and REACTOME in their pursuit of deciphering molecular heterogeneity and drug responsiveness by exploring the molecular diversity of tumors and drug sensitivity. No table is provided for this type of use since the methodology is standardized.

6. Conclusions

Over the last two decades, ontologies gained traction in biomedical research in general, and cancer research in particular, enabling FAIR data (findability, accessibility, interoper-

ability and reusability) [162], supporting data integration and analysis, and facilitating data interpretation and data mining. Presently, we are witnessing the emergence of the knowledge graph paradigm, whereby large volumes of heterogeneous data are brought together under a single holistic ontological knowledge model. Yet, there are still a number of open challenges to the development and application of ontologies and knowledge graphs for cancer research.

One major challenge lies in reusing existing ontologies. With over 800 biomedical ontologies publicly available in BioPortal [3], most biomedical subjects are covered by one or more ontologies, and it might seem foolish not to reuse them. However, the fact that there are so many ontologies and many overlap in domain makes it difficult to navigate the ontology landscape and select which ones to reuse. Moreover, many ontologies were typically developed with a singular purpose in mind, and have a particular perspective on the domain they model which may be unsuited for other purposes. This means that additional care is needed when selecting ontologies to reuse, to make sure that their perspective on the domain is compatible with the new use case. Last but not least, it may be the case that existing ontologies are no longer actively maintained and kept up to date, which in a dynamic domain like biomedicine, will render them useless in a short time span. Ultimately, it may very well be that no existing ontology is compatible with or usable in the new use case, and that a new ontology must be developed, which indeed is the main reason why there are presently so many ontologies. Thus, to avoid perpetuating the problem, new ontologies should be designed circumspectly, taking into account possible other applications within their specific domain [30].

Another challenge lies in the disconnection between data and ontologies, due to the fact that, in the large majority of cases, biomedical ontologies do not include data. In fact, few biomedical ontologies were designed with the prospect of directly encoding data, as the biomedical research community has, for the most part, viewed ontologies merely as abstract knowledge models used for classification or at best annotation of data, with the data kept in relational databases or even data files. This is tied to the reusability challenge, as existing ontologies may not be reusable for use cases such as constructing knowledge graphs if they are unsuited to being instantiated. Furthermore, it means that constructing biomedical knowledge graphs to support cancer research requires (semi-)automated approaches to integrating the data with the knowledge model, which, considering the variety and heterogeneity of relevant biomedical data sources, can be burdensome [163]. However, as the knowledge graph paradigm becomes more popular, we may witness a shift in the biomedical community towards storing data in graph databases rather than relational databases.

Tied to the two previous challenges is the challenge of integrating multiple ontologies, a necessity for constructing holistic knowledge graphs for cancer research, due to the multidisciplinary nature of the domain. Although there are comprehensive ontologies on cancer (e.g., NCI), available data is often connected to more specialized ontologies (e.g., GO, MeSH), eliciting the need to integrate them. The problem is that, due to their different perspectives, overlapping ontologies may be semantically irreconcilable [164], which may impede their joint use. Thus, the costs of reusing existing ontologies may outweigh their benefits, prompting the development of an independent ontological knowledge model for a knowledge graph, ideally with mappings to existing ontologies to ensure interoperability and facilitate data integration.

The benefits of developing holistic knowledge graphs that integrate all the data relevant for cancer research are deeply tied to the potential of AI approaches to unlock knowledge conducive to better diagnostics or treatments. Knowledge graphs can serve as sources of background knowledge to AI approaches, compensating for missing values in the data, they can support image classification and NLP approaches to enrich image or textual data, which in turn can improve the performance of AI approaches relying on that data, and they provide a means to afford explainability to AI approaches [165], tackling the black-box problem of state-of-the-art AI methods.

The immense potential of ontologies and the knowledge graph paradigm to support cancer research data management and analysis is increasingly recognized by the oncology research community as an essential building block of the P4 medicine vision (preventative, predictive, personalized and participatory).

Author Contributions: Formal analysis, data curation and writing—original draft preparation, M.C.S. and P.E.; conceptualization, methodology and writing—reviewing and editing, D.F. and C.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by FCT through the LASIGE Research Unit (UIDB/00408/2020 and UIDP/00408/2020). It was also partially supported by the KATY project which has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 101017453.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ATC	Anatomical Therapeutic Chemical
ATSDR	Agency for Toxic Substances and Disease Registry
ALL	Acute Lymphoblastic Leukemia
ANN	Artificial Neural Network
BCFO	Breast Cancer Fuzzy Ontology
BCO	Breast Cancer Ontology
BFO	Basic Formal Ontology
BRFSS	Behavioral Risk Factor Surveillance System
BTL2	BioTopLite 2
caDSR	Cancer Data Standards Repository
CBR	Case-Based Reasoning
CCL	Cancer Cell Ontology
CCTOO	Cancer Care Treatment Outcome Ontology
CDEs	Common Data Elements
CGC	Cancer Gene Census
CHV	Consumer Health Vocabulary
CL	Cell Ontology
CLO	Cell Line Ontology
CMPO	Cellular Microscopy Phenotype Ontology
COBRA	COnsortium for BRachytherapy data Analysis
COntQueSt	Cancer Ontology Querying System
CPD	Cellular Phenotype Database
CTCAE	Common Terminology Criteria for Adverse Events
CTG	ClinicalTrials.gov
DICOM	Digital Imaging and Communications in Medicine
DL	Description Logic
DLBCL	Diffuse Large B Cell Lymphoma
DO	Disease Ontology
EFO	Experimental Factor Ontology
ENCR	European Network of Cancer Registries
ENCR core-data	European Cancer-Registry core-data ontology
FCDS	Florida Cancer Data System

FL	Follicular Lymphoma
FMA	Foundational Model of Anatomy
FOAF	Friend of a Friend ontology
FOORC	Fuzzy Ontology for Obesity-Related Cancer
GAD	Genetic Association Database
GCO	Gastric Cancer Ontology
GDSC	Genomics of Drug Sensitivity in Cancer
GO	Gene Ontology
HDO	Human Disease Ontology
HCC	Hepatocellular Carcinoma
HOS	Health Outcomes Survey
HUGO Gene Nomenclature	Human Genome Organization Gene Nomenclature
ICD-9-CM	International Classification of Diseases Ninth Revision Clinical Modification
ICD-O-3	International Classification of Disease for Oncology 3rd edition
IMI	Interactive Mapping Interface
IOBC	Interlinking Ontology for Biological Concepts
KCR	Kentucky Cancer Registry
KEGG	Kyoto Encyclopedia of Genes and Genomes
KG	Knowledge Graph
LCO	Liver Cancer Ontology
LCKGO	Lung Cancer Knowledge Graph Ontology
LINCS	Library of Integrated Network-based Cellular Signatures
lncRNAs	long non-coding RNAs
LOINC	Logical Observation Identifier Names and Codes
LTR	Louisiana Tumor Registry
LuCO	Lung Cancer Ontology
MAS	Multi-Agent System
MCVS	Multi-threaded Clinical Vocabulary Server
MedDRA	Medical Dictionary for Regulatory Activities
MeSH	Medical Subject Headings
MGI	Mouse Genome Informatics
ML	Machine Learning
MP	Mammalian Phenotype ontology
MuEVo	Multi-Expertise Vocabulary
NAACCR	North American Association of Central Cancer Registries
NCI	National Cancer Institute
NCIt	National Cancer Institute Thesaurus
NCRI	National Cancer Registry Ireland
NCRO	Non-Coding RNA Ontology
NEJM	New England Journal of Medicine
NLP	Natural Language Processing
OBDA	Ontology-Based Data Access
OBI	Ontology for Biomedical Investigators
OCCC	Ovarian clear cell carcinoma
OCRV	Ontology for Cancer Research Variables
OCRSEV	Ontology of Cancer Related Social-Ecological Variables
OD-ATTEST	Ontology for the Documentation of vAriable selecTion and daTa source Selection and inTegration
ODVDS	Ontology for Documentation of Variable and Data Source
OGMS	Ontology of General Medical Science
OIE	Open Information Extraction
OMIM	Online Mendelian Inheritance in Man
OMIT	Ontology for MicroRNA Target
OntHCC	Ontology of Hepatocellular Carcinoma
OQuaRE	Ontology Quality Evaluation Framework
OSCC	Oral Squamous Cell Carcinoma
OWL	Web Ontology Language

PAC	Prostatic Adenocarcinoma
POCS	Profile Ontology for Cancer Survivors
QHIO	Quantitative Histopathological Imaging Ontology
RO	Relation Ontology
ROS	Radiation Oncology Structures
SCRS	Semantic Cancer Registry System
SEER-MHOS	Surveillance, Epidemiology, and End Results—Medicare Health Outcomes Survey
SIO	Semanticscience Integrated Ontology
SKOS	Simple Knowledge Organization System
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms
SNOMEDint	SNOMED International
SOSW	Sentiment Ontology for Social Web
SVMs	Support Vector Machines
SWIT	Semantic Web Integration Tool
TCGA	The Cancer Genome Atlas
TEO	Time Event Ontology
TNM	Tumor–Node–Metastasis
TNM-O	Tumor–Node–Metastasis Ontology
TOCSOC	Temporal Ontology for Comparing the Survival Outcomes
TTD	Therapeutic Target Database
UMLS	Unified Medical Language System
USCB	United States Census Bureau

References

- SNOMED International. Available online: <https://www.snomed.org/> (accessed on 25 March 2022).
- Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)] [[PubMed](#)]
- Whetzel, P.L.; Noy, N.F.; Shah, N.H.; Alexander, P.R.; Nyulas, C.; Tudorache, T.; Musen, M.A. 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*, W541–W545. [[CrossRef](#)] [[PubMed](#)]
- Golbeck, J.; Fragoso, G.; Hartel, F.; Hendler, J.; Oberthaler, J.; Parsia, B. The National Cancer Institute’s thesaurus and ontology. *J. Web Semant. First Look* **2003**, *1*, 4.
- Chin, L.; Andersen, J.N.; Futreal, P.A. Cancer genomics: From discovery science to personalized medicine. *Nat. Med.* **2011**, *17*, 297–303. [[CrossRef](#)]
- Gruber, T.R. Toward principles for the design of ontologies used for knowledge sharing? *Int. J. Hum.-Comput. Stud.* **1995**, *43*, 907–928. [[CrossRef](#)]
- McGuinness, D.L. Ontologies come of age. In *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*; MIT Press: Cambridge, MA, USA, 2002; pp. 171–194.
- OWL 2 Web Ontology Language Document Overview (Second Edition). Available online: <https://www.w3.org/TR/owl2-overview/> (accessed on 25 March 2022).
- Gutiérrez, C.; Sequeda, J.F. Knowledge graphs. *Commun. ACM* **2021**, *64*, 96–104. [[CrossRef](#)]
- Lin, F.P.; Groza, T.; Kocbek, S.; Antezana, E.; Epstein, R.J. Cancer Care Treatment Outcome Ontology: A novel computable ontology for profiling treatment outcomes in patients with solid tumors. *JCO Clin. Cancer Inform.* **2018**, *2*, 1–14. [[CrossRef](#)]
- Salvi, D.; Picone, M.; Arredondo, M.T.; Cabrera-Umpierrez, M.F.; Esteban, Á.; Steger, S.; Poli, T. Merging person-specific bio-markers for predicting oral cancer recurrence through an ontology. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 216–220. [[CrossRef](#)]
- Tao, S.; Zeng, N.; Hands, I.; Hurt-Mueller, J.; Durbin, E.B.; Cui, L.; Zhang, G.Q. Web-based interactive mapping from data dictionaries to ontologies, with an application to cancer registry. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 271. [[CrossRef](#)]
- Yan, S.; Wong, K. Elucidating high-dimensional cancer hallmark annotation via enriched ontology. *J. Biomed. Inform.* **2017**, *73*, 84–94. [[CrossRef](#)]
- Oyelade, O.N.; Ezugwu, A.E.; Adewuyi, S.A. Enhancing reasoning through reduction of vagueness using fuzzy OWL-2 for representation of breast cancer ontologies. *Neural Comput. Appl.* **2021**, *34*, 1–26. [[CrossRef](#)] [[PubMed](#)]
- Zhu, Q.; Tao, C.; Shen, F.; Chute, C.G. Exploring the pharmacogenomics knowledge base (PharmGKB) for repositioning breast cancer drugs by leveraging Web ontology language (OWL) and cheminformatics approaches. *Pac. Symp. Biocomput.* **2014**, *2014*, 172–182.
- Agioutantis, P.C.; Loutrari, H.; Kolis, F.N. Computational analysis of transcriptomic and proteomic data for deciphering molecular heterogeneity and drug responsiveness in model human hepatocellular carcinoma cell lines. *Genes* **2020**, *11*, 623. [[CrossRef](#)] [[PubMed](#)]

17. Su, L.; Liu, G.; Bai, T.; Meng, X.; Ma, Q. MGOGP: A gene module-based heuristic algorithm for cancer-related gene prioritization. *BMC Bioinform.* **2018**, *19*, 215. [[CrossRef](#)] [[PubMed](#)]
18. Althubaiti, S.; Karwath, A.; Dallol, A.; Noor, A.; Alkhayyat, S.S.; Alwassia, R.; Mineta, K.; Gojobori, T.; Beggs, A.D.; Schofield, P.N.; et al. Ontology-based prediction of cancer driver genes. *Sci. Rep.* **2019**, *9*, 17405. [[CrossRef](#)] [[PubMed](#)]
19. Shen, Y.; Colloc, J.; Jacquet-Andrieu, A.; Lei, K. Emerging medical informatics with case-based reasoning for aiding clinical decision in multi-agent system. *J. Biomed. Inform.* **2015**, *56*, 307–317. [[CrossRef](#)]
20. PubMed. Available online: <https://pubmed.ncbi.nlm.nih.gov/> (accessed on 10 January 2022).
21. Google Scholar. Available online: <https://ncit.nci.nih.gov/ncitbrowser/> (accessed on 21 March 2022).
22. NCI Thesaurus. Available online: <https://scholar.google.com/> (accessed on 25 March 2022).
23. Bodenreider, O. The unified medical language system (UMLS): Integrating biomedical terminology. *Nucleic Acids Res.* **2004**, *32*, D267–D270. [[CrossRef](#)]
24. Medical Subject Headings. Available online: <https://www.nlm.nih.gov/mesh/meshhome.html> (accessed on 25 March 2022).
25. Schriml, L.M.; Mitraka, E.; Munro, J.; Tauber, B.; Schor, M.; Nickle, L.; Felix, V.; Jeng, L.; Bearer, C.; Lichenstein, R.; et al. Human Disease Ontology 2018 update: Classification, content and workflow expansion. *Nucleic Acids Res.* **2019**, *47*, D955–D962. [[CrossRef](#)] [[PubMed](#)]
26. World Health Organization (WHO). *International Classification of Diseases for Oncology (ICD-O)*, 3rd ed.; 1st Revision; World Health Organization (WHO): Geneva, Switzerland, 2013.
27. Bandrowski, A.; Brinkman, R.; Brochhausen, M.; Brush, M.H.; Bug, B.; Chibucos, M.C.; Clancy, K.; Courtot, M.; Derom, D.; Dumontier, M.; et al. The ontology for biomedical investigations. *PLoS ONE* **2016**, *11*, e0154556. [[CrossRef](#)]
28. Sarntivijai, S.; Lin, Y.; Xiang, Z.; Meehan, T.F.; Diehl, A.D.; Vempati, U.D.; Schürer, S.C.; Pang, C.; Malone, J.; Parkinson, H.; et al. CLO: The cell line ontology. *J. Biomed. Semant.* **2014**, *5*, 1–10. [[CrossRef](#)]
29. Li, F.; Du, J.; He, Y.; Song, H.Y.; Madkour, M.; Rao, G.; Xiang, Y.; Luo, Y.; Chen, H.W.; Liu, S.; et al. Time event ontology (TEO): To support semantic representation and reasoning of complex temporal relations of clinical events. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 1046–1056. [[CrossRef](#)] [[PubMed](#)]
30. Nicholson, N.C.; Giusti, F.; Bettio, M.; Negrao Carvalho, R.; Dimitrova, N.; Dyba, T.; Flego, M.; Neamtui, L.; Randi, G.; Martos, C. An ontology-based approach for developing a harmonised data-validation tool for European cancer registration. *J. Biomed. Semant.* **2021**, *12*, 1. [[CrossRef](#)] [[PubMed](#)]
31. Zhang, H.; Guo, Y.; Prospero, M.; Bian, J. An ontology-based documentation of data discovery and integration process in cancer outcomes research. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 292. [[CrossRef](#)] [[PubMed](#)]
32. Rasmussen, K.E.; Dolan, M.E. *OncoCL: A Cancer Cell Ontology*; ICBO: Lansing, MI, USA, 2013; p. 126.
33. Jusoh, F.; Ibrahim, R.; Othman, M.S.; Omar, N. Development of breast cancer ontology based on hybrid approach. *Int. J. Innov. Comput.* **2013**, *3*, 1.
34. Myneni, S.; Amith, M.; Geng, Y.; Tao, C. Towards an ontology-driven framework to enable development of personalized mHealth solutions for cancer survivors' engagement in healthy living. *Stud. Health Technol. Inform.* **2015**, *216*, 113–117.
35. Maly, A.M.; Jain, S.K.; Yang, P.C.; Harvey, K.; Warner, J.L. Computerized approach to creating a systematic ontology of hematology/oncology regimens. *JCO Clin. Cancer Inform.* **2018**, *2*, 1–11. [[CrossRef](#)]
36. Dinakarpanthian, D.; Liedtke, M.; Musen, M.A.; Dinakar, B. *TOCSOC: A Temporal Ontology for Comparing the Survival Outcomes of Clinical Trials in Oncology*; ICBO: Lansing, MI, USA, 2018.
37. Chen, Y.; Yu, C.; Liu, X.; Xi, T.; Xu, G.; Sun, Y.; Zhu, F.; Shen, B. PCLiON: An ontology for data standardization and sharing of prostate cancer associated lifestyles. *Int. J. Med Inform.* **2021**, *145*, 104332. [[CrossRef](#)]
38. Herrmann, J.; Zabka, S.; Boeker, M.; Schulz, S. Ontology Patterns for Tubular or Spherical Layered Structures. A Case Study from Oncology. In Proceedings of the Joint Ontology Workshop, Graz, Austria, 23–25 September 2019; Volume 2518.
39. Esteban-Gil, A.; Fernández-Breis, J.T.; Boeker, M. Analysis and visualization of disease courses in a semantically-enabled cancer registry. *J. Biomed. Semant.* **2017**, *8*, 46. [[CrossRef](#)]
40. Amith, M.; Song, H.Y.; Zhang, Y.; Xu, H.; Tao, C. Lightweight predicate extraction for patient-level cancer information and ontology development. *BMC Med. Inform. Decis. Mak.* **2017**, *17*, 73. [[CrossRef](#)]
41. Elhefny, M.; Elmogy, M.; Elfetouh, A.; Badria, F. FOORC: A Fuzzy Ontology-Based Representation for Obesity Related Cancer Knowledge. *Int. J. Intell. Comput. Inf. Sci.* **2016**, *16*, 15–36. [[CrossRef](#)]
42. Tapi Nzali, M.D.; Aze, J.; Bringay, S.; Lavergne, C.; Mollevi, C.; Optiz, T. Reconciliation of patient/doctor vocabulary in a structured resource. *Health Inform. J.* **2019**, *25*, 1219–1231. [[CrossRef](#)] [[PubMed](#)]
43. Lee, J.; Park, H.A.; Park, S.K.; Song, T.M. Using social media data to understand consumers' information needs and emotions regarding cancer: Ontology-based data analysis study. *J. Med. Internet Res.* **2020**, *22*, e18767. [[CrossRef](#)] [[PubMed](#)]
44. Messaoudi, R.; Jaziri, F.; Mtibaa, A.; Grand-Brochier, M.; Ali, H.M.; Amouri, A.; Fourati, H.; Chabrot, P.; Gargouri, F.; Vacavant, A. Ontology-based approach for liver cancer diagnosis and treatment. *J. Digit. Imaging* **2019**, *32*, 116–130. [[CrossRef](#)] [[PubMed](#)]
45. Gurcan, M.N.; Tomaszewski, J.; Overton, J.A.; Doyle, S.; Ruttenberg, A.; Smith, B. Developing the Quantitative Histopathology Image Ontology (QHIO): A case study using the hot spot detection problem. *J. Biomed. Inform.* **2017**, *66*, 129–135. [[CrossRef](#)]
46. Boeker, M.; França, F.; Bronsert, P.; Schulz, S. TNM-O: Ontology support for staging of malignant tumors. *J. Biomed. Semant.* **2016**, *7*, 64. [[CrossRef](#)]

47. Tagliaferri, L.; Budrukkar, A.; Lenkiewicz, J.; Cambeiro, M.; Bussu, F.; Guinot, J.L.; Hildebrandt, G.; Johansson, B.; Meyer, J.E.; Niehoff, P.; et al. ENT COBRA ONTOLOGY: The covariates classification system proposed by the Head & Neck and Skin GEC-ESTRO Working Group for interdisciplinary standardized data collection in head and neck patient cohorts treated with interventional radiotherapy (brachytherapy). *J. Contemp. Brachyther.* **2018**, *10*, 260–266.
48. Lancellotta, V.; Guinot, J.L.; Fionda, B.; Rembielak, A.; Di Stefani, A.; Gentileschi, S.; Federico, F.; Rossi, E.; Guix, B.; Chyrek, A.J.; et al. SKIN-COBRA (Consortium for Brachytherapy data Analysis) ontology: The first step towards interdisciplinary standardized data collection for personalized oncology in skin cancer. *J. Contemp. Brachyther.* **2020**, *12*, 105–110. [[CrossRef](#)]
49. Mahmoodi, S.A.; Mirzaie, K.; Mahmoudi, S.M. A new algorithm to extract hidden rules of gastric cancer data based on ontology. *Springerplus* **2016**, *5*, 312. [[CrossRef](#)]
50. Gao, M.; Warner, J.; Yang, P.; Alterovitz, G. On the Bayesian derivation of a treatment-based cancer ontology. *AMIA Summits Transl. Sci. Proc.* **2014**, *2014*, 209–217.
51. Sesen, M.B.; Banares-Alcántara, R.; Fox, J.; Kadir, T.; Brady, J.M. Lung Cancer Assistant: An ontology-driven, online decision support prototype for lung cancer treatment selection. In Proceedings of the OWL: Experiences and Directions Workshop (OWLED), Heraklion, Greece, 27–28 May 2012.
52. Barki, C.; Rahmouni, H.B.; Labidi, S. Prediction of Bladder Cancer Treatment Side Effects Using an Ontology-Based Reasoning for Enhanced Patient Health Safety. *Informatics* **2021**, *8*, 55. [[CrossRef](#)]
53. Zhang, L.; Geng, Z.; Meng, X.; Meng, F.; Wang, L. Screening for key lncRNAs in the progression of gallbladder cancer using bioinformatics analyses. *Mol. Med. Rep.* **2018**, *17*, 6449–6455. [[CrossRef](#)] [[PubMed](#)]
54. Tao, C.; Sun, J.; Zheng, W.J.; Chen, J.; Xu, H. *Drug Target Prediction for Colorectal Cancer by Combining Ontology and Network Approaches*; ICBO: Lansing, MI, USA, 2014; p. 67.
55. Balasubramanian, D.K.; Khan, J.Z.; Bian, J.; Guo, Y.; Hogan, W.R.; Hicks, A. *Ontology of Cancer Related Social-Ecological Variables*; ICBO: Lansing, MI, USA, 2017.
56. Bibault, J.E.; Zapletal, E.; Rance, B.; Giraud, P.; Burgun, A. Labeling for Big Data in radiation oncology: The Radiation Oncology Structures ontology. *PLoS ONE* **2018**, *13*, e0191263. [[CrossRef](#)] [[PubMed](#)]
57. Zhang, H.; Guo, Y.; Bian, J. Ontology for Documentation of Variable and Data Source Selection Process to Support Integrative Data Analysis in Cancer Outcomes Research. In Proceedings of the SEPDA@ ISWC, Auckland, New Zealand, 27 October 2019; pp. 63–67.
58. Divakar, H.; Ramesh, D.; Prakash, B.; Tumkur, M.T. Prediction of Cervical Cancer with Ontology Based Deep Learning Approach. *Int. J. Comput. Sci. Commun.* **2020**, 60–66.
59. Daowd, A.; Barrett, M.; Abidi, S.; Abidi, S.S.R. Building a Knowledge Graph Representing Causal Associations Between Risk Factors and Incidence of Breast Cancer. In *Public Health and Informatics*; IOS Press: Amsterdam, The Netherlands, **2021**; pp. 724–728.
60. Serra, L.M.; Duncan, W.D.; Diehl, A.D. An ontology for representing hematologic malignancies: The cancer cell ontology. *BMC Bioinform.* **2019**, *20*, 181. [[CrossRef](#)] [[PubMed](#)]
61. Ong, E.; Xie, J.; Ni, Z.; Liu, Q.; Sarntivijai, S.; Lin, Y.; Cooper, D.; Terry, R.; Stathias, V.; Chung, C.; et al. Ontological representation, integration, and analysis of LINCS cell line cells and their cellular responses. *BMC Bioinform.* **2017**, *18*, 556. [[CrossRef](#)]
62. Campbell, W.S.; Karlsson, D.; Vreeman, D.J.; Lazenby, A.J.; Talmon, G.A.; Campbell, J.R. A computable pathology report for precision medicine: Extending an observables ontology unifying SNOMED CT and LOINC. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 259–266. [[CrossRef](#)] [[PubMed](#)]
63. Melo, M.T.D.; Gonçalves, V.; Costa, H.; Braga, D.; Gomide, L.; Alves, C.; Brasil, L.M. OntoMama: An Ontology Applied to Breast Cancer. In *MEDINFO 2015: eHealth-Enabled Health*; IOS Press: Amsterdam, The Netherlands, 2015; p. 1104.
64. Chen, H.W.; Du, J.; Song, H.Y.; Liu, X.; Jiang, G.; Tao, C. Representation of time-relevant common data elements in the Cancer Data Standards Repository: Statistical evaluation of an ontological approach. *JMIR Med. Inform.* **2018**, *6*, e7. [[CrossRef](#)]
65. Li, Y.; Chen, H.; Pan, T.; Jiang, C.; Zhao, Z.; Wang, Z.; Zhang, J.; Xu, J.; Li, X. LncRNA ontology: Inferring lncRNA functions based on chromatin states and expression patterns. *Oncotarget* **2015**, *6*, 39793–39805. [[CrossRef](#)]
66. Milian, K.; Hoekstra, R.; Bucur, A.; Ten Teije, A.; van Harmelen, F.; Paulissen, J. Enhancing reuse of structured eligibility criteria and supporting their relaxation. *J. Biomed. Inform.* **2015**, *56*, 205–219. [[CrossRef](#)]
67. Kim, D.; Joung, J.G.; Sohn, K.A.; Shin, H.; Park, Y.R.; Ritchie, M.D.; Kim, J.H. Knowledge boosting: A graph-based integration approach with multi-omics data and genomic knowledge for cancer clinical outcome prediction. *J. Am. Med. Inform. Assoc.* **2015**, *22*, 109–120. [[CrossRef](#)]
68. Wu, T.J.; Schriml, L.M.; Chen, Q.R.; Colbert, M.; Crichton, D.J.; Finney, R.; Hu, Y.; Kibbe, W.A.; Kincaid, H.; Meerzaman, D.; et al. Generating a focused view of disease ontology cancer terms for pan-cancer data integration and analysis. *Database* **2015**, *2015*, bav032. [[CrossRef](#)] [[PubMed](#)]
69. Bona, J.P.; Nolan, T.S.; Brochhausen, M. Ontology-enhanced representations of non-image data in The Cancer Imaging Archive. In Proceedings of the 9th International Conference on Biological Ontology (ICBO 2018), Corvallis, OR, USA, 7–10 August 2018.
70. Zhang, L.; Hao, C.; Li, J.; Qu, Y.; Bao, L.; Li, Y.; Yue, Z.; Zhang, M.; Yu, X.; Chen, H.; et al. Bioinformatics methods for identifying differentially expressed genes and signaling pathways in nano-silica stimulated macrophages. *Tumour Biol.* **2017**, *39*, 1010428317709284. [[CrossRef](#)] [[PubMed](#)]

71. Hasan, S.M.S.; Rivera, D.; Wu, X.C.; Durbin, E.B.; Christian, J.B.; Tourassi, G. Knowledge graph-enabled cancer data analytics. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 1952–1967. [[CrossRef](#)] [[PubMed](#)]
72. Li, N.; Yang, Z.; Luo, L.; Wang, L.; Zhang, Y.; Lin, H.; Wang, J. KGHC: A knowledge graph for hepatocellular carcinoma. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 135. [[CrossRef](#)]
73. Chen, Y.; Verbeek, F.J.; Wolstencroft, K. Establishing a consensus for the hallmarks of cancer based on gene ontology and pathway annotations. *BMC Bioinform.* **2021**, *22*, 178. [[CrossRef](#)]
74. González-Beltrán, A.; Tagger, B.; Finkelstein, A. Federated ontology-based queries over cancer data. *BMC Bioinform.* **2012**, *13* (Suppl. 1), S9. [[CrossRef](#)]
75. Oster, S.; Langella, S.; Hastings, S.; Ervin, D.; Madduri, R.; Phillips, J.; Kurc, T.; Siebenlist, F.; Covitz, P.; Shanbhag, K.; et al. caGrid 1.0: An enterprise Grid infrastructure for biomedical research. *J. Am. Med. Inform. Assoc.* **2008**, *15*, 138–149. [[CrossRef](#)]
76. Lyu, Y.; Caudron-Herger, M.; Diederichs, S. Circ2GO: A database linking circular RNAs to gene function. *Cancers* **2020**, *12*, 2975. [[CrossRef](#)]
77. Elkin, P.L.; Frankel, A.; Liebow-Liebling, E.H.; Elkin, J.R.; Tuttle, M.S.; Brown, S.H. Bioprospecting the bibleome: Adding evidence to support the inflammatory basis of cancer. *Metabolomics* **2012**, *2*, 6451. [[CrossRef](#)]
78. Zhang, H.; Guo, Y.; Li, Q.; George, T.J.; Shenkman, E.; Modave, F.; Bian, J. An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival. *BMC Med. Inform. Decis. Mak.* **2018**, *18*, 41. [[CrossRef](#)]
79. Fernández-López, M.; Gómez-Pérez, A.; Juristo, N. Methontology: From ontological art towards ontological engineering. In Proceedings of the Ontological Engineering AAAI-97 Spring Symposium, Stanford, CA, USA, 24–26 March 1997.
80. Musen, M.; Protégé Team. The protégé project: A look back and a look forward. *AI Matters* **2015**, *1*, 4–12.
81. Alfonse, M.; Aref, M.M.; Salem, A.B.M. An ontology-based system for cancer diseases knowledge management. *Int. J. Inf. Eng. Electron. Bus.* **2014**, *6*, 55–63. [[CrossRef](#)]
82. Tao, C.; Sun, J.; Zheng, W.J.; Chen, J.; Xu, H. Colorectal cancer drug target prediction using ontology-based inference and network analysis. *Database* **2015**, *2015*, bav015. [[CrossRef](#)] [[PubMed](#)]
83. Nicholson, N.C.; Giusti, F.; Bettio, M.; Negro Carvalho, R.; Dimitrova, N.; Dyba, T.; Flego, M.; Neamtii, L.; Randi, G.; Martos, C. An ontology to model the international rules for multiple primary malignant tumours in cancer registration. *Appl. Sci.* **2021**, *11*, 7233. [[CrossRef](#)]
84. Rebholz-Schuhmann, D.; Oellrich, A.; Hoehndorf, R. Text-mining solutions for biomedical research: Enabling integrative biology. *Nat. Rev. Genet.* **2012**, *13*, 829–839. [[CrossRef](#)] [[PubMed](#)]
85. Chen, H.; Zhang, D.; Zhang, G.; Li, X.; Liang, Y.; Kasukurthi, M.V.; Li, S.; Borchert, G.M.; Huang, J. A semantics-oriented computational approach to investigate microRNA regulation on glucocorticoid resistance in pediatric acute lymphoblastic leukemia. *BMC Med. Inform. Decis. Mak.* **2018**, *18*, 57. [[CrossRef](#)] [[PubMed](#)]
86. Pesquita, C.; Faria, D.; Falcao, A.O.; Lord, P.; Couto, F.M. Semantic similarity in biomedical ontologies. *PLoS Comput. Biol.* **2009**, *5*, e1000443. [[CrossRef](#)]
87. Acharya, S.; Cui, L.; Pan, Y. Multi-view feature selection for identifying gene markers: A diversified biological data driven approach. *BMC Bioinform.* **2020**, *21*, 483. [[CrossRef](#)]
88. Su, K.M.; Lin, T.W.; Liu, L.C.; Yang, Y.P.; Wang, M.L.; Tsai, P.H.; Wang, P.H.; Yu, M.H.; Chang, C.M.; Chang, C.C. The potential role of complement system in the progression of ovarian clear cell carcinoma inferred from the Gene Ontology-based immunofunctionome analysis. *Int. J. Mol. Sci.* **2020**, *21*, 2824. [[CrossRef](#)]
89. Bourgeais, V.; Zehraoui, F.; Ben Hamdoune, M.; Hanczar, B. Deep GONet: Self-explainable deep neural network based on Gene Ontology for phenotype prediction from gene expression data. *BMC Bioinform.* **2021**, *22*, 455. [[CrossRef](#)]
90. Min, H.; Mobahi, H.; Irvin, K.; Avramovic, S.; Wojtusiak, J. Predicting activities of daily living for cancer patients using an ontology-guided machine learning methodology. *J. Biomed. Semant.* **2017**, *8*, 39. [[CrossRef](#)]
91. Xi, J.; Ye, L.; Huang, Q.; Li, X. Tolerating data missing in breast cancer diagnosis from clinical ultrasound reports via knowledge graph inference. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Singapore, 14–18 August 2021; pp. 3756–3764.
92. Zhang, M.Y.; Du, R.Z. A Real-time Inference Method of Graph Attention Network Based on Knowledge Graph for Lung Cancer. In Proceedings of the 5th International Conference on Digital Signal Processing, Chengdu, China, 26–28 February 2021; pp. 326–331.
93. Kim, J. In silico analysis of differentially expressed genesets in metastatic breast cancer identifies potential prognostic biomarkers. *World J. Surg. Oncol.* **2021**, *19*, 188. [[CrossRef](#)] [[PubMed](#)]
94. Sun, X.; Liu, Y.; Gao, X.; Du, M.; Gao, M.; Zhong, X.; Wei, X. Analysis of LncRNA-mRNA co-expression profiles in patients with polycystic ovary syndrome: A pilot study. *Front. Immunol.* **2021**, *12*, 669819. [[CrossRef](#)] [[PubMed](#)]
95. Yu, C.; Chen, J.; Ma, J.; Zang, L.; Dong, F.; Sun, J.; Zheng, M. Identification of key genes and signaling pathways associated with the progression of gastric cancer. *Pathol. Oncol. Res.* **2020**, *26*, 1903–1919. [[CrossRef](#)] [[PubMed](#)]
96. Wang, L.; Wang, B.; Quan, Z. Identification of aberrantly methylated-differentially expressed genes and gene ontology in prostate cancer. *Mol. Med. Rep.* **2020**, *21*, 744–758. [[CrossRef](#)] [[PubMed](#)]
97. Zhang, X.; Yin, S.; Ma, K. Bioinformatics analysis of different candidate genes involved in hepatocellular carcinoma induced by HepG2 cells or tumor cells of patients. *J. Int. Med. Res.* **2020**, *48*, 300060520932112. [[CrossRef](#)]
98. Tang, M.; Dai, W.; Wu, H.; Xu, X.; Jiang, B.; Wei, Y.; Qian, H.; Han, L. Transcriptome analysis of tongue cancer based on high-throughput sequencing. *Oncol. Rep.* **2020**, *43*, 2004–2016. [[CrossRef](#)] [[PubMed](#)]

99. Wei, S.; Chen, J.; Huang, Y.; Sun, Q.; Wang, H.; Liang, X.; Hu, Z.; Li, X. Identification of hub genes and construction of transcriptional regulatory network for the progression of colon adenocarcinoma hub genes and TF regulatory network of colon adenocarcinoma. *J. Cell. Physiol.* **2020**, *235*, 2037–2048. [[CrossRef](#)]
100. Anukriti; Dhasmana, A.; Uniyal, S.; Somvanshi, P.; Bhardwaj, U.; Gupta, M.; Haque, S.; Lohani, M.; Kumar, D.; Ruokolainen, J.; et al. Investigation of precise molecular mechanistic action of tobacco-associated carcinogen ‘NNK’ induced carcinogenesis: A system biology approach. *Genes* **2019**, *10*, 564. [[CrossRef](#)]
101. Rendleman, M.C.; Buatti, J.M.; Braun, T.A.; Smith, B.J.; Nwakama, C.; Beichel, R.R.; Brown, B.; Casavant, T.L. Machine learning with the TCGA-HNSC dataset: Improving usability by addressing inconsistency, sparsity, and high-dimensionality. *BMC Bioinform.* **2019**, *20*, 339. [[CrossRef](#)]
102. Yang, H.; Zhou, L.; Chen, J.; Su, J.; Shen, W.; Liu, B.; Zhou, J.; Yu, S.; Qian, J. A four-gene signature for prognosis in breast cancer patients with hypermethylated IL15RA. *Oncol. Lett.* **2019**, *17*, 4245–4254. [[CrossRef](#)]
103. Guo, F.; Wang, C.Y.; Wang, S.; Zhang, J.; Yan, Y.J.; Guan, Z.Y.; Meng, F.J. Alteration in gene expression profile of thymomas with or without myasthenia gravis linked with the nuclear factor-kappaB/autoimmune regulator pathway to myasthenia gravis pathogenesis. *Thorac. Cancer* **2019**, *10*, 564–570. [[CrossRef](#)] [[PubMed](#)]
104. Ren, F.H.; Yang, H.; He, R.Q.; Lu, J.N.; Lin, X.G.; Liang, H.W.; Dang, Y.W.; Feng, Z.B.; Chen, G.; Luo, D.Z. Analysis of microarrays of miR-34a and its identification of prospective target gene signature in hepatocellular carcinoma. *BMC Cancer* **2018**, *18*, 12. [[CrossRef](#)] [[PubMed](#)]
105. Zhang, G.; Bi, M.; Li, S.; Wang, Q.; Teng, D. Determination of core pathways for oral squamous cell carcinoma via the method of attract. *J. Cancer Res. Ther.* **2018**, *14*, S1029–S1034. [[PubMed](#)]
106. Xu, X.; Li, M.; Hu, J.; Chen, Z.; Yu, J.; Dong, Y.; Sun, C.; Han, J. Expression profile analysis identifies a two-gene signature for prediction of head and neck squamous cell carcinoma patient survival. *J. Cancer Res. Ther.* **2018**, *14*, 1525–1534.
107. Shen, Y.; Feng, Y.; Chen, H.; Huang, L.; Wang, F.; Bai, J.; Yang, Y.; Wang, J.; Zhao, W.; Jia, Y.; et al. Focusing on long non-coding RNA dysregulation in newly diagnosed multiple myeloma. *Life Sci.* **2018**, *196*, 133–142. [[CrossRef](#)]
108. Yang, M.; Li, H.; Li, Y.; Ruan, Y.; Quan, C. Identification of genes and pathways associated with MDR in MCF-7/MDR breast cancer cells by RNA-seq analysis. *Mol. Med. Rep.* **2018**, *17*, 6211–6226. [[CrossRef](#)]
109. She, S.; Jiang, L.; Zhang, Z.; Yang, M.; Hu, H.; Hu, P.; Liao, Y.; Yang, Y.; Ren, H. Identification of the C-reactive protein interaction network using a bioinformatics approach provides insights into the molecular pathogenesis of hepatocellular carcinoma. *Cell. Physiol. Biochem.* **2018**, *48*, 741–752. [[CrossRef](#)]
110. Wang, S.; Cai, Y. Identification of the functional alteration signatures across different cancer types with support vector machine and feature analysis. *Biochim. Biophys. Acta Mol. Basis Dis.* **2018**, *1864*, 2218–2227. [[CrossRef](#)]
111. Yang, Z.; Li, H.; Wang, Z.; Yang, Y.; Niu, J.; Liu, Y.; Sun, Z.; Yin, C. Microarray expression profile of long non-coding RNAs in human lung adenocarcinoma. *Thorac. Cancer* **2018**, *9*, 1312–1322. [[CrossRef](#)]
112. Yu, C.; Xue, P.; Zhang, L.; Pan, R.; Cai, Z.; He, Z.; Sun, J.; Zheng, M. Prediction of key genes and pathways involved in trastuzumab-resistant gastric cancer. *World J. Surg. Oncol.* **2018**, *16*, 174. [[CrossRef](#)]
113. Chang, C.M.; Yang, Y.P.; Chuang, J.H.; Chuang, C.M.; Lin, T.W.; Wang, P.H.; Yu, M.H.; Chang, C.C. Discovering the deregulated molecular functions involved in malignant transformation of endometriosis to endometriosis-associated ovarian carcinoma using a data-driven, function-based analysis. *Int. J. Mol. Sci.* **2017**, *18*, 2345. [[CrossRef](#)]
114. Xu, K.; Zhang, Y.Y.; Han, B.; Bai, Y.; Xiong, Y.; Song, Y.; Zhou, L.M. Suppression subtractive hybridization identified differentially expressed genes in colorectal cancer: microRNA-451a as a novel colorectal cancer-related gene. *Tumour Biol.* **2017**, *39*, 1010428317705504. [[CrossRef](#)]
115. Zhang, T.; Fan, X.; Song, L.; Ren, L.; Ma, E.; Zhang, S.; Ren, L.; Zheng, Y.; Zhang, J. c-Fos is involved in inhibition of human bladder carcinoma T24 cells by brazilin. *IUBMB Life* **2015**, *67*, 175–181. [[CrossRef](#)]
116. Vashisht, S.; Bagler, G. An approach for the identification of targets specific to bone metastasis using cancer genes interactome and gene ontology analysis. *PLoS ONE* **2012**, *7*, e49401. [[CrossRef](#)]
117. Kwon, Y.K.; Lee, S.Y.; Kang, H.S.; Sung, J.S.; Cho, C.K.; Yoo, H.S.; Shin, S.; Choi, J.S.; Lee, Y.W.; Jang, I.S. Differential expression of gene profiles in MRGX-treated lung cancer. *J. Pharmacopunct.* **2013**, *16*, 30–38. [[CrossRef](#)]
118. Yang, L.; Zhang, J.; Jiang, A.; Liu, Q.; Li, C.; Yang, C.; Xiu, J. Expression profile of long non-coding RNAs is altered in endometrial cancer. *Int. J. Clin. Exp. Med.* **2015**, *8*, 5010–5021.
119. Valavanis, I.; Pilalis, E.; Georgiadis, P.; Kyrtopoulos, S.; Chatziioannou, A. Cancer biomarkers from genome-scale DNA methylation: Comparison of evolutionary and semantic analysis methods. *Microarrays* **2015**, *4*, 647–670. [[CrossRef](#)]
120. Lo, Y.H.; Chung, E.; Li, Z.; Wan, Y.W.; Mahe, M.M.; Chen, M.S.; Noah, T.K.; Bell, K.N.; Yalamanchili, H.K.; Klisch, T.J.; et al. Transcriptional regulation by ATOH1 and its target SPDEF in the intestine. *Cell. Mol. Gastroenterol. Hepatol.* **2017**, *3*, 51–71. [[CrossRef](#)]
121. Liu, M.Y.; Zhang, H.; Hu, Y.J.; Chen, Y.W.; Zhao, X.N. Identification of key genes associated with cervical cancer by comprehensive analysis of transcriptome microarray and methylation microarray. *Oncol. Lett.* **2016**, *12*, 473–478. [[CrossRef](#)]
122. Yang, F.; Lyu, S.; Dong, S.; Liu, Y.; Zhang, X.; Wang, O. Expression profile analysis of long noncoding RNA in HER-2-enriched subtype breast cancer by next-generation sequencing and bioinformatics. *Onco. Targets Ther.* **2016**, *9*, 761–772. [[CrossRef](#)]
123. Yin, H.; Wang, S.; Zhang, Y.H.; Cai, Y.D.; Liu, H. Analysis of important gene ontology terms and biological pathways related to pancreatic cancer. *Biomed Res. Int.* **2016**, *2016*, 7861274. [[CrossRef](#)]

124. Shangkuan, W.C.; Lin, H.C.; Chang, Y.T.; Jian, C.E.; Fan, H.C.; Chen, K.H.; Liu, Y.F.; Hsu, H.M.; Chou, H.L.; Yao, C.T.; et al. Risk analysis of colorectal cancer incidence by gene expression analysis. *PeerJ* **2017**, *5*, e3003. [\[CrossRef\]](#)
125. Khayer, N.; Zamanian-Azodi, M.; Mansouri, V.; Ghassemi-Broumand, M.; Rezaei-Tavirani, M.; Heidari, M.H.; Rezaei Tavirani, M. Oral squamous cell cancer protein-protein interaction network interpretation in comparison to esophageal adenocarcinoma. *Gastroenterol. Hepatol. Bed Bench* **2017**, *10*, 118–124.
126. Vaseghi Maghvan, P.; Rezaei-Tavirani, M.; Zali, H.; Nikzamir, A.; Abdi, S.; Khodadoostan, M.; Asadzadeh-Aghdaei, H. Network analysis of common genes related to esophageal, gastric, and colon cancers. *Gastroenterol. Hepatol. Bed Bench* **2017**, *10*, 295–302.
127. Ding, Y.; Yang, D.Z.; Zhai, Y.N.; Xue, K.; Xu, F.; Gu, X.Y.; Wang, S.M. Microarray expression profiling of long non-coding RNAs in epithelial ovarian cancer. *Oncol. Lett.* **2017**, *14*, 2523–2530. [\[CrossRef\]](#)
128. Kumar, R.; Samal, S.K.; Routray, S.; Dash, R.; Dixit, A. Identification of oral cancer related candidate genes by integrating protein-protein interactions, gene ontology, pathway analysis and immunohistochemistry. *Sci. Rep.* **2017**, *7*, 2472. [\[CrossRef\]](#)
129. Valizadeh, R.; Bahadorimonfared, A.; Rezaei-Tavirani, M.; Norouzinia, M.; Ehsani Ardakani, M.I. Evaluation of involved proteins in colon adenocarcinoma: An interactome analysis. *Gastroenterol. Hepatol. Bed Bench* **2017**, *10*, S129–S138.
130. Attar, R.; Cincin, Z.B.; Bireller, E.S.; Cakmakoglu, B. Apoptotic and genomic effects of corilagin on SKOV3 ovarian cancer cell line. *Onco. Targets Ther.* **2017**, *10*, 1941–1946. [\[CrossRef\]](#)
131. Tian, P.; Liang, C. Transcriptome profiling of cancer tissues in Chinese patients with gastric cancer by high-throughput sequencing. *Oncol. Lett.* **2018**, *15*, 2057–2064. [\[CrossRef\]](#)
132. Deng, Y.; He, R.; Zhang, R.; Gan, B.; Zhang, Y.; Chen, G.; Hu, X. The expression of HOXA13 in lung adenocarcinoma and its clinical significance: A study based on The Cancer Genome Atlas, Oncomine and reverse transcription-quantitative polymerase chain reaction. *Oncol. Lett.* **2018**, *15*, 8556–8572. [\[CrossRef\]](#)
133. Li, H.; Gong, M.; Zhao, M.; Wang, X.; Cheng, W.; Xia, Y. LncRNAs KB-1836B5, LINC00566 and FAM27L are associated with the survival time of patients with ovarian cancer. *Oncol. Lett.* **2018**, *16*, 3735–3745. [\[CrossRef\]](#) [\[PubMed\]](#)
134. Wu, C.; Zhao, Y.; Liu, Y.; Yang, X.; Yan, M.; Min, Y.; Pan, Z.; Qiu, S.; Xia, S.; Yu, J.; et al. Identifying miRNA-mRNA regulation network of major depressive disorder in ovarian cancer patients. *Oncol. Lett.* **2018**, *16*, 5375–5382. [\[CrossRef\]](#) [\[PubMed\]](#)
135. Zhang, Y.; Luo, J.; Wang, X.; Wang, H.L.; Zhang, X.L.; Gan, T.Q.; Chen, G.; Luo, D.Z. A comprehensive analysis of the predicted targets of miR-642b-3p associated with the long non-coding RNA HOXA11-AS in NSCLC cells. *Oncol. Lett.* **2018**, *15*, 6147–6160. [\[CrossRef\]](#) [\[PubMed\]](#)
136. Liu, Y.; Hua, T.; Chi, S.; Wang, H. Identification of key pathways and genes in endometrial cancer using bioinformatics analyses. *Oncol. Lett.* **2019**, *17*, 897–906. [\[CrossRef\]](#)
137. Qi, F.; Qin, W.X.; Zang, Y.S. Molecular mechanism of triple-negative breast cancer-associated BRCA1 and the identification of signaling pathways. *Oncol. Lett.* **2019**, *17*, 2905–2914. [\[CrossRef\]](#)
138. Wang, X.; Yang, Y.; Tan, X.; Mao, X.; Wei, D.; Yao, Y.; Jiang, P.; Mo, D.; Wang, T.; Yan, F. Identification of tRNA-derived fragments expression profile in breast cancer tissues. *Curr. Genom.* **2019**, *20*, 199–213. [\[CrossRef\]](#)
139. Jin, L.; Zhu, C.; Qin, X. Expression profile of tRNA-derived fragments in pancreatic cancer. *Oncol. Lett.* **2019**, *18*, 3104–3114. [\[CrossRef\]](#)
140. Guo, W.; Yu, H.; Zhang, L.; Chen, X.; Liu, Y.; Wang, Y.; Zhang, Y. Effect of hyperoside on cervical cancer cells and transcriptome analysis of differentially expressed genes. *Cancer Cell Int.* **2019**, *19*, 235. [\[CrossRef\]](#)
141. Asadzadeh-Aghdaei, H.; Okhovatian, F.; Razzaghi, Z.; Heidari, M.; Vafae, R.; Nikzamir, A. Radiation therapy in patients with brain cancer: Post-proteomics interpretation. *J. Lasers Med. Sci.* **2019**, *10*, S59–S63. [\[CrossRef\]](#)
142. Han, B.; Wang, H.; Zhang, J.; Tian, J. FNDC3B is associated with ER stress and poor prognosis in cervical cancer. *Oncol. Lett.* **2020**, *19*, 406–414. [\[CrossRef\]](#)
143. Vallino, L.; Ferraresi, A.; Vidoni, C.; Secomandi, E.; Esposito, A.; Dhanasekaran, D.N.; Isidoro, C. Modulation of non-coding RNAs by resveratrol in ovarian cancer cells: In silico analysis and literature review of the anti-cancer pathways involved. *J. Tradit. Complement. Med.* **2020**, *10*, 217–229. [\[CrossRef\]](#) [\[PubMed\]](#)
144. Sarkar, J.P.; Saha, I.; Lancucki, A.; Ghosh, N.; Wlasnowolski, M.; Bokota, G.; Dey, A.; Lipinski, P.; Plewczynski, D. Identification of miRNA biomarkers for diverse cancer types using statistical learning methods at the whole-genome scale. *Front. Genet.* **2020**, *11*, 982. [\[CrossRef\]](#) [\[PubMed\]](#)
145. Zhu, L.; Yang, X.; Zhu, R.; Yu, L. Identifying discriminative biological function features and rules for cancer-related long non-coding RNAs. *Front. Genet.* **2020**, *11*, 598773. [\[CrossRef\]](#)
146. Hermawan, A.; Ikawati, M.; Jenie, R.I.; Khumaira, A.; Putri, H.; Nurhayati, I.P.; Angraini, S.M.; Muflikhasari, H.A. Identification of potential therapeutic target of naringenin in breast cancer stem cells inhibition by bioinformatics and in vitro studies. *Saudi Pharm. J.* **2021**, *29*, 12–26. [\[CrossRef\]](#)
147. Liu, W.Q.; Li, W.L.; Ma, S.M.; Liang, L.; Kou, Z.Y.; Yang, J. Discovery of core gene families associated with liver metastasis in colorectal cancer and regulatory roles in tumor cell immune infiltration. *Transl. Oncol.* **2021**, *14*, 101011. [\[CrossRef\]](#)
148. Abeni, E.; Grossi, I.; Marchina, E.; Coniglio, A.; Incardona, P.; Cavalli, P.; Zorzi, F.; Chiodera, P.L.; Paties, C.T.; Crosatti, M.; et al. DNA methylation variations in familial female and male breast cancer. *Oncol. Lett.* **2021**, *21*, 468. [\[CrossRef\]](#)
149. Pedroza, D.A.; Ramirez, M.; Rajamanickam, V.; Subramani, R.; Margolis, V.; Gurbuz, T.; Estrada, A.; Lakshmanaswamy, R. MiRNome and functional network analysis of PGRMC1 regulated miRNA target genes identify pathways and biological functions associated with triple negative breast cancer. *Front. Oncol.* **2021**, *11*, 710337. [\[CrossRef\]](#)

150. Wu, S.; Lv, X.; Zhang, Y.; Xu, X.; Zhao, F.; Zhang, Y.; Chen, L.; Ou-Yang, H.; Ti, X. Microarray analysis of genes with differential expression of m6A methylation in lung cancer. *Biosci. Rep.* **2021**, *41*, BSR20210523. [[CrossRef](#)]
151. Siavoshi, A.; Taghizadeh, M.; Dookhe, E.; Piran, M. Gene expression profiles and pathway enrichment analysis to identification of differentially expressed gene and signaling pathways in epithelial ovarian cancer based on high-throughput RNA-seq data. *Genomics* **2021**, *114*, 161–170. [[CrossRef](#)]
152. Ai, X.; Jia, Z.M.; Wang, J.; Di, G.P.; Zhang, X.U.; Sun, F.; Zang, T.; Liao, X. Bioinformatics analysis of the target gene of fibroblast growth factor receptor 3 in bladder cancer and associated molecular mechanisms. *Oncol. Lett.* **2015**, *10*, 543–549. [[CrossRef](#)]
153. Ung, T.H.; Madsen, H.J.; Hellwinkel, J.E.; Lencioni, A.M.; Graner, M.W. Exosome proteomics reveals transcriptional regulator proteins with potential to mediate downstream pathways. *Cancer Sci.* **2014**, *105*, 1384–1392. [[CrossRef](#)] [[PubMed](#)]
154. Heo, S.G.; Koh, Y.; Kim, J.K.; Jung, J.; Kim, H.L.; Yoon, S.S.; Park, J.W. Identification of somatic mutations using whole-exome sequencing in Korean patients with acute myeloid leukemia. *BMC Med. Genet.* **2017**, *18*, 23. [[CrossRef](#)] [[PubMed](#)]
155. Makler, A.; Narayanan, R. Mining exosomal genes for pancreatic cancer targets. *Cancer Genom. Proteom.* **2017**, *14*, 161–172. [[CrossRef](#)]
156. Yao, H.; Wu, C.; Chen, Y.; Guo, L.; Chen, W.; Pan, Y.; Fu, X.; Wang, G.; Ding, Y. Spectrum of gene mutations identified by targeted next-generation sequencing in Chinese leukemia patients. *Mol. Genet. Genom. Med.* **2020**, *8*, e1369. [[CrossRef](#)]
157. Hindmathi, V.; Kranthi, T.; Rao, S.; Manimaran, P. The prediction of candidate genes for cervix related cancer through gene ontology and graph theoretical approach. *Mol. BioSyst.* **2014**, *10*, 1450–1460. [[CrossRef](#)] [[PubMed](#)]
158. Simjanoska, M.; Madevska Bogdanova, A.; Panov, S. Gene ontology analysis of colorectal cancer biomarkers probed with affymetrix and illumina microarrays. In Proceedings of the 5th International Joint Conference on Computational Intelligence, Algarve, Portugal, 25–27 October 2013.
159. Chen, L.; Zhang, Y.H.; Lu, G.; Huang, T.; Cai, Y.D. Analysis of cancer-related lncRNAs using gene ontology and KEGG pathways. *Artif. Intell. Med.* **2017**, *76*, 27–36. [[CrossRef](#)] [[PubMed](#)]
160. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)]
161. Chen, G.; Tsoi, A.; Xu, H.; Zheng, W.J. Predict effective drug combination by deep belief network and ontology fingerprints. *J. Biomed. Inform.* **2018**, *85*, 149–154. [[CrossRef](#)]
162. Vesteghem, C.; Brøndum, R.F.; Sønderkær, M.; Sommer, M.; Schmitz, A.; Bødker, J.S.; Dybkær, K.; El-Galaly, T.C.; Bøgsted, M. Implementing the FAIR Data Principles in precision oncology: Review of supporting initiatives. *Brief. Bioinform.* **2020**, *21*, 936–945. [[CrossRef](#)]
163. Seneviratne, O.; Rashid, S.M.; Chari, S.; McCusker, J.P.; Bennett, K.P.; Hendler, J.A.; McGuinness, D.L. Knowledge integration for disease characterization: A breast cancer example. In Proceedings of the International Semantic Web Conference, Monterey, CA, USA, 8–12 October 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 223–238.
164. Pesquita, C.; Faria, D.; Santos, E.; Couto, F.M. To repair or not to repair: Reconciling correctness and coherence in ontology reference alignments. In Proceedings of the 8th ISWC Ontology Matching Workshop (OM), Sydney, Australia, 25 October 2013; Volume 3.
165. Lecue, F. On the role of knowledge graphs in explainable AI. *Semant. Web* **2020**, *11*, 41–51. [[CrossRef](#)]