Ontology-supported Research on Vaccine Efficacy, Safety, and Integrative Biological Networks

Yongqun He

1. Unit for Laboratory Animal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109, USA, 2. Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109, USA, 3. Center for Computational Medicine and Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA, 4. Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109, USA

Summary

While vaccine efficacy and safety research has dramatically progressed with the methods of *in silico* prediction and data mining, many challenges still exist. A formal ontology is a human- and computer-interpretable set of terms and relations that represent entities in a specific domain and how these terms relate to each other. Several community-based ontologies (including the Vaccine Ontology, Ontology of Adverse Events, and Ontology of Vaccine Adverse Events) have been developed to support vaccine and adverse event representation, classification, data integration, literature mining of host-vaccine interaction networks, and analysis of vaccine adverse events. The author further proposes minimal vaccine information standards and their ontology representations, ontology-based linked open vaccine data and meta-analysis, an integrative One Network ("OneNet") Theory of Life, and ontology-based approaches to study and apply the OneNet theory. In the Big Data era, these proposed strategies provide a novel framework for advanced data integration and analysis of fundamental biological networks including vaccine immune mechanisms.

Keywords

vaccine; vaccine efficacy; vaccine safety; adverse event; ontology; literature mining; data mining; meta-analysis; theory; interaction network

Progresses and Challenges

While vaccines have been used to dramatically improve public health, effective and safe vaccines against many deadly diseases, including HIV, malaria, tuberculosis, and most cancers, still do not exist. Data integration and mining have played an important role in vaccine candidate prediction and causal adverse event (AE) identification. However, with the

ever increasing amount of vaccine-related data available, how to more efficiently integrate and analyze the data has become a huge challenge.

In silico vaccine candidate prediction and bottlenecks

Different bioinformatic methods have been developed to support vaccine research [1]. The first success in bioinformatic prediction of vaccine candidates is through the epitope prediction using immunoinformatic methods. The predicted immune epitopes can be used in different ways, including diagnosis, vaccine immune response studies, and rational development of epitope-based vaccines. Many epitope vaccines have been tested and evaluated [1]. T-cell epitopes are classified into MHC Class I or Class II immune epitopes that are able to stimulate CD8+ cytotoxic T cells or CD4+ T helper cells, respectively. Current T-cell epitope prediction methods are very accurate with a range of 90-95% positive predictive value [1]. B-cell epitopes are able to stimulate the production of specific antibodies. B-cell epitopes include linear (or non-conformational) and discontinuous (or conformational) epitopes. In general, B-cell epitopes, especially discontinuous B-cell epitopes are much more difficult to predict than T-cell epitopes [1]. Approximately 90% of B-cell epitopes are discontinuous epitopes. The challenge of accurately predicting discontinuous B-cell epitopes prevents the successful development of antibody-oriented vaccines.

Protein antigen prediction can be performed using the reverse vaccinology strategy that starts with bioinformatic analysis of genome sequencing data with an aim to predict vaccine protein candidates. The predicted genes and proteins can then be used for efficient development of gene-based vaccines such as subunit protein vaccines, DNA vaccines, and recombinant vector vaccines. Reverse vaccinology has been pioneered by Dr. Rino Rappuoli and proven successful for the development of a vaccine against serogroup B *Neisseria meningitidis* (MenB), the major cause of sepsis and meningitis in children and young adults [2]. Originated from the study, Rappuoli and his colleagues in Novartis have developed Bexsero, a multi-component, broad-coverage MenB vaccine [3]. Currently Bexsero has been approved for use in Europe, Australia, Canada, and USA. The successful story of the Bexsero vaccine development is established as a milestone in reverse vaccinology. Vaxign is the first web-based vaccine design program based on the reverse vaccinology strategy [4]. The Vaxign execution pipeline is able to predict antigen cellular localization, adhesion, epitope binding to MHC class I and class II, and sequence similarities to human, mouse and/or pig proteins. Vaxign has successfully been tested in prediction of vaccine candidates for many different pathogens such as *Brucella* [5-7] , uropathogenic *E. coli* [4], *Francisella* [8], *Streptococcus agalactiae* [9] and human herpes simplex virus [10]. The experimental verification of Vaxign-predicted results [7] confirms the accuracy of the Vaxign vaccine candidate prediction.

While Vaxign has been proven successful, it can be further improved by incorporating other types of vaccine candidate prediction methods such as protein structure-based prediction [11], literature mining [12] and Omics-based gene expression screening [13,14] (Figure 1). For example, since the DNA is not the same as the DNA gene expression, real experimental detection and analysis of gene expression profiles is critical to identify which genes are up-

or down-regulated. The gene expression profiles will support vaccine design. Omics gene expression methods (including transcriptomics and proteomics) and RNA-seq sequencing methods are able to simultaneously monitor the expression profiles of thousands of genes. Transcriptomics and RNA-seq measure the RNA level gene expression. Proteomics measures the protein-level gene expression. Given the complexity and a large amount of data, it is a challenge to efficiently integrate various data into optimal prediction of vaccine antigens (Figure 1).

In addition to the prediction of antigens, it is often challenging to rationalize the selection and usage of other vaccine components, for example, vaccine adjuvants [15] and DNA vaccine plasmids [16]. Bioinformatics programs are also needed to support the rational design of these individual vaccine components.

Vaccine safety-related data mining and future challenges

While licensed vaccines are effective for most vaccinees, they induce adverse events (AE), and sometimes even severe adverse events in specific populations. Different types of vaccines may also be associated with different vaccine adverse events (VAE) [17,18]. For example, many live attenuated vaccines are typically very efficient in stimulating protective immunity. However, live attenuated vaccines are currently difficult to get approval for human use primarily due to their potential VAEs and safety issues.

Many post-licensure vaccine safety surveillance programs exist [19]. One of the most commonly referred to programs is the Vaccine Adverse Event Reporting System (VAERS) [20], a web-based vaccine safety surveillance program co-sponsored by the USA Centers for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA). VAERS collects the data of VAE cases for those vaccines licensed for use in the USA [21,22]. No causal association between the vaccination and reported adverse events is assumed. Since 1990, VAERS has received more than 200,000 case reports. Most of the reported cases describe mild adverse events. Very rarely, some serious adverse events might be reported. By collecting and analyzing the VAERS data, possible causal associations between particular adverse events and administration of particular vaccines could be identified [23].

Since anyone (*e.g.*, vaccine recipients, health care providers and vaccine manufacturers) can file VAERS report, the reported adverse event data can be noisy. For data normalization, the Medical Dictionary for Regulatory Activities (MedDRA) terminology system [24] is used. The MedDRA system, mostly used in pharmacovigilance for AE case report coding, has also been used for drug adverse event reporting [24]. The usage of MedDRA also supports statistical AE data analysis. However, several issues related to the usage of MedDRA exist. For example, MedDRA does not provide specific definitions for individual MedDRA terms. The lack of term definitions may cause confusion when MedDRA terms are assigned to VAERS cases. In addition, the relations between MedDRA terms are not or poorly defined. The poorly defined MedDRA hierarchical structure makes it difficult to use MedDRA for term clustering analysis [18]. The Systematized Nomenclature of Medicine (SNOMED) provides another commonly used terminology for general medical evaluation [25]. However, many issues existing in MedDRA also exist for SNOMED.

In addition to clinical VAE case reports, another focus on vaccine safety is the exploration of basic VAE-related molecular mechanisms. A large number of studies have been published in this area. Progresses and challenges co-exist and will be discussed in detail below.

Data mining of vaccine immune mechanisms and known bottlenecks

The vaccine candidate prediction skills as introduced above are primarily based on microbe-side sequence (DNA/RNA/protein) analysis. Since vaccines are used to improve host (*e.g.*, human) immunity, it is also important to examine the host-side immune mechanisms. The questions to be asked include: (i) What host gene immune responses correlate with vaccine-induced protective immunity? (ii) How does the host respond to individual vaccine components, such as vaccine adjuvant? (iii) What host molecular pathways are critical to the protection? (iv) What genetic susceptibility factors exist that significantly contribute to the induction of various VAEs?

There are different ways to address these questions. A common approach is to raise a hypothesis about how one gene can be regulated or regulate a pathway, followed by experimental testing of the hypothesis with immune response evaluation after the gene knockout or knockdown. While such an approach is highly specific, it is often difficult to identify novel hypothesis and the individual gene testing can be very expensive. Another approach relies on high throughput studies including Omics analysis or next-generation sequencing analysis. The Omics approaches are composed of several methods that ends with the letters "omics", including genomics, transcriptomics, proteomics, and metabolomics. The RNA-seq sequencing technology is able to measure the RNA gene expression profiles by next-generation sequencing [26]. The strategy "systems vaccinology" targets to use high throughput Omics and other systems biology approaches to analyze the interactions between vaccines and host immune systems and identify critical immune signatures and markers that correlate with vaccine-induced protection [27,28]. The systems vaccinology strategy has been successfully demonstrated in different examples, such as Yellow fever vaccine YF-17D [27], and live attenuated and killed inactivated influenza vaccines [29].

To better understand the mechanisms of VAEs, Dr. Gregory A. Poland proposed a strategy called "vaccinomics" that applies high dimensional genetic, proteomic, bioinformatic, and other assays to study vaccine response heterogeneity [17]. After the linkage between vaccine administration and host genetic susceptibility is understood, it will become possible to generate "personalized vaccines" that are specifically targeted for persons or a group of persons and benefit from the prevention of severe adverse events [30,31]. To achieve such a new paradigm of developing "personalized vaccines", more fundamental basic mechanism studies are required.

A critical challenge in "systems vaccinology" and "vaccinomics" is on the difficulty of systematically integrating and analyzing high throughput data. In addition to newly generated experimental studies, a large amount of published literature data are also available. A feasible strategy for uncovering fundamental vaccine immune mechanisms is literature annotation and mining of published experimental results, followed by the storage of curated results in databases, bioinformatic assembly of the results, and statistical analysis and reasoning based on the results. Since manual curation will not be able to catch up with the

speed of paper publication in the era of "Big Data", automated literature mining has become very important. These areas of research will be described below and later in this paper.

Existing vaccine research databases and bottlenecks in further development

Most existing vaccine databases provide administration, safety, or other information about licensed vaccines [32]. Many vaccine databases are generated by governmental agencies and international organizations. These databases usually include information about licensed human or animal vaccines and are typically targeted for clinical usages or for educating the public. For example, the US Food and Drug Administration (FDA) and Department of Agriculture (USDA) provide the information of US-licensed human vaccines [33] and veterinary vaccines [34], respectively. The US Centers for Disease Control and Prevention (CDC) Vaccine Information Statements (VISs) system includes information sheets explaining the benefits and risks of a vaccine [35]. As described earlier in this article, the VAERS database collects the data of randomly reported VAE cases associated with US-licensed vaccines [21,22]. The PATH Vaccine Resource Library (VRL) is a database offering high-quality documents and links to specific diseases and vaccine immunization topics [36,37].

Many research-oriented vaccine-related databases also exist. For example, many immune epitope databases have been developed, including the Immune Epitope Database (IEDB) [38,39], SYFPEITHI [40], and the international ImMunoGeneTics information system (IMGT) [41,42]. Disease-specific vaccine databases, such as the Nonhuman Primate HIV/SIV Vaccine Trials Database [43], also exist. The most comprehensive research-oriented vaccine database is the web-based <u>V</u>accine <u>I</u>nvestigation and <u>O</u>n<u>L</u>ine <u>I</u>nformation <u>N</u>etwork (VIOLIN) [44,45]. The VIOLIN vaccine database has stored over 3,000 experimentally verified vaccines that were manually curated from peer-reviewed publications [46]. The VIOLIN database also includes the information of vaccine components (*e.g.*, vaccine antigen, adjuvant and vector) and the immune responses stimulated by vaccines. Some important components of VIOLIN are introduced below.

VIOLIN includes many complementary components that are related to potential vaccine efficacy improvement research. Besides the Vaxign vaccine design program, VIOLIN includes several databases targeting for collecting manually verified data on different vaccine ingredients. Particularly, VIOLIN has the following major databases: (1) Protegen protective antigen database [47]. Protegen collects those antigens that have been experimentally proven to be effective as a protective vaccine antigen component. Protegen has currently included over 800 protective antigens. (2) The VirmugenDB "virmugen" database. The term "virmugen" was coined to define a virulence factor that can be mutated in a wild type virulent pathogen for generation of an effective live attenuated vaccine that stimulates protective immunity against virulent pathogen infections [46]. The VirmugenDB database has collected >200 virmugens and associated live attenuated vaccines [46]. (3) The Vaxjo vaccine adjuvant database [15]. Vaxjo has collected 93 vaccine adjuvants that have been used in 379 vaccines stored in VIOLIN. (4) The DNAVaxDB DNA vaccine database [16]. DNAVaxDB contains >100 DNA vaccine plasmids that have been used in over 400 vaccines [16]. DNAVaxDB also includes bacterial, viral and DNA vaccine plasmid vectors.

In addition, VIOLIN has also started to collect and analyze data for other vaccine components (*e.g.*, vaccine vectors) and for vaccine-induced host immune factors.

The analysis and usage of the data stored in the databases will be able to support rational and efficient development of vaccines against the infections of existing and emerging pathogens. These programs allow systematic analysis of enriched patterns from experimentally verified literature reports. The identified enriched patterns can be used for rational vaccine design. For example, the list of protective antigens in the Protegen database has been used as a gold standard for vaccine design software development [48]. Many specific enriched patterns among the collected virmugens have been identified to support the development of live attenuated vaccines [46].

Although the VIOLIN manually curated databases have been proven successful, many challenges prevent its fast growth. While manual curation and inclusion of the manually curated data in a well-structured database provides high quality data, such an approach cannot catch up with the pace of ever increasing publications. Therefore, computer-assisted, automated literature mining is required to increase the rate of information retrieval. However, the accuracy of current literature mining methods can still not match that of manual curation. How to improve the literature mining and make the data curation better performed is still a big challenge.

A promising solution: ontology-based data mining

In this section, the background of ontology is first introduced. The current status of ontology-based data mining to support the areas of vaccine efficacy and safety improvements is then reviewed.

What is ontology and ontology-based data mining

A formal ontology is a human- and computer-interpretable set of terms and relations that represent entities in a specific domain and how they relate to each other. Most formal biomedical ontologies are represented using the Web Ontology Language (OWL) format [49] or Open Biomedical Ontology (OBO) format [50]. Biomedical ontologies can be incorporated in various software programs and used to support automated reasoning and Semantic Web applications [51].

Current biomedical ontologies are developed usually through international collaborations. The OBO Foundry is a collaborative initiative aimed at establishing a set of ontology development principles and ontologies (such as the VO) following these principles in an evolving non-redundant suite [52]. To facilitate collaborative ontology development, an evolving set of OBO Foundry principles have been designed to encapsulate good practice in ontology development [52]. These principles require that ontologies 1) be developed in a collaborative effort; 2) use common relations that are unambiguously defined; 3) provide procedures for user feedback and for identifying successive versions; and 4) have a clearly bounded subject-matter [52]. These principles have widely been accepted and used as guidance to support the ever increasing efforts of ontology development.

Biomedical ontologies have played important roles in different areas. As shown in Figure 2, ontologies have at least the following roles:

i. **Naming "Things"**. Most ontologies are designed to first represent entities in different scientific domains. For example, the NCBI Taxonomy ontology (NCBITaxon) includes the names and hierarchy of nearly one million taxonomy names [53].

ii. **Knowledgebase construction.** For example, the Foundational Model of Anatomy (FMA) that serves as an explicit declarative knowledge about human anatomy [54]. FMA represents classes (or types) and relations necessary for representing phenotypic structure of the human body in a human- and computer-interpretable form.

iii. **Data exchange**. A typical example is the OWL-based Biological Pathway Exchange (BioPAX) standard language targeted for representing molecular and cellular pathways and facilitating the exchange of biological pathway data [55].

iv. **Data integration**. Ontologies are able to integrate different types of data (*e.g.*, vaccines, vaccine ingredients, vaccination methods, and host immune responses) by representing these types of data and the relations among them. The universal (or called class) level representation also provides a framework of representing particular (or instance) level data [56]. An example of ontology-supported data integration is that the Vaccine Ontology (VO) [57,58] has been used to integrate data available from the VIOLIN vaccine database [46].

v. **Data Analysis**. Many ontologies, such as the Gene Ontology (GO) [59], have played a major role in the integration and analysis of the large amount of heterogeneous biological data available in the post-genomics era. Since its publication in 2000, GO has been cited in over 6,000 peer-reviewed publications.

vi. **Natural Language Processing (NLP)**. For example, both GO and VO have been applied to enhance the performance of literature mining [12,60-64]. The VO-SciMiner is a VO-supported SciMiner program for enhanced literature indexing and analysis [12].

The Vaccine Ontology (VO) and its usage in vaccine data integration and mining

With a large number of vaccines and vaccine-related data collected in VIOLIN, it has become difficult to integrate these data into a more powerful understanding of vaccine knowledge. Towards addressing this issue, the community-based VO development was initiated [57]. Developed by following the OBO Foundry ontology development principles, VO is an ontology member under the OBO Foundry Ontology Library [65]. As of March 1, 2014, VO includes 4,864 ontology terms, including 3976 VO-specific terms with the "VO_" prefix and the other terms imported from 27 other existing ontologies. More specific statistics of VO is available in the Ontobee [66] web page [67].

Developed with the OWL format [49], VO supports both asserted hierarchy and inferred hierarchy. An asserted hierarchy is an ontology hierarchy asserted by the ontology

developers, and an inferred hierarchy is generated by using an ontology reasoner [68-70]. Figure 3 provides an example of how VO supports inferred hierarchies of live attenuated or inactivated vaccines. The ontology-enhanced data mining relies on the fact that an ontology (*e.g.*, VO) not only provides asserted hierarchies of thousands of terms (*e.g.*, different vaccine names), but it also provides the logical definitions that support automated inferencing [12,64,71].

VO can be used in different applications. VO has been used in the VIOLIN system as a way to more logically represent, share, and link vaccine data stored in VIOLIN [15,16,46,47]. For example, the information of virmugens and related vaccines stored in VirmugenDB has been represented in the Vaccine Ontology (VO) [57,64]. The usage of VO provides an efficient way for better data analysis in scientific vaccine research. With extensive vaccine research, the volume of vaccine literature publications has been growing exponentially [1]. The literature query in PubMed is supported by the National Library of Medicine (NLM)'s Medical Subject Headings (MeSH), a controlled vocabulary of medical and scientific terms that is used by biomedical scientists to manually index articles in the PubMed literature database [72]. MeSH has also been used in many other data mining programs [73-75]. The usage of VO in combination with SciMiner, a literature indexing and gene name tagging system [76], has allowed the development of an integrative VO-based SciMiner system VO-SciMiner [12]. VO-SciMiner demonstrated superior performance in indexing/retrieving vaccine papers compared to MeSH with the *Brucella* vaccine use case study [12]. GenoMesh [77] is a novel algorithm and web program able to predict implicit gene-to-gene relationships and networks based on genome-wide gene-MeSH term associations from the PubMed literature database [78]. GenoMesh predicts implicit gene-gene relations by statistically analyzing gene pairs that do not co-exist in any papers but share common MeSH terms in MeSH-indexed PubMed papers. GenoMesh has been successfully evaluated using *E. coli* and *Brucella* use cases [78]. It is anticipated that the GenoMesh algorithm can be implemented using the VO and other OBO Foundry ontologies instead of MeSH to possibly achieve higher performance.

VO-based literature mining has been used to improve the analysis of gene-vaccine interactions. The VO-SciMiner was used to retrieve and analyze bacterial gene-vaccine associations [12]. In addition, VO has been used to improve the discovery of host gene-vaccine interaction networks in natural language processing (NLP)-based literature mining. Using all the abstracts in the PubMed literature repository, a machine learning (particularly, support vector machine or SVM [79])- and centrality-based method was used to identify those host genes that interact with interferon-gamma (IFN-γ) within or without the context of vaccine [80]. By using VO that provides more names and relations of individual vaccines (*e.g.*, tuberculosis vaccine BCG), the literature mining program significantly improved the retrieval of the vaccine-associated IFN-γ-gene interaction network [64]. The inferred VO hierarchy was also found beneficial in the analysis of specific types of vaccines (*e.g.*, live attenuated vaccines). In addition, such a method was applied to detect those vaccine- and fever-associated gene-gene interactions [63].

The Ontology of Adverse Events (OAE) and the Ontology of Vaccine Adverse Events (OVAE)

Two safety-related biomedical ontologies recently developed include: The Ontology of Adverse Events (OAE) [18,81] and the Ontology of Vaccine Adverse Events (OVAE) [82].

OAE was initially developed as a branch of the vaccine adverse events inside the VO. Due to the complexity of AE representation and the shared features among different AEs (*e.g.*, vaccine and drug AEs), the branch of vaccine AEs in VO was singled out to form the basis of the Adverse Event Ontology (AEO). The AEO was focused on representing causal adverse events. Since the namespace AEO was taken by another ontology (*i.e.*, "Anatomical Entity Ontology"), the name of AEO was later changed to OAE, which represents the "Ontology of Adverse Events" [18]. In addition, OAE differs from AEO in that the adverse events represented in OAE do not have to be causally induced by a medical intervention. Such a broader range design makes the OAE 'adverse event' definition consistent with its definition in commonly used clinical scenarios including their uses in the VAERS and the FDA Adverse Event Reporting System (FAERS). Instead of focusing on AEs associated with specific triggers (*e.g.*, vaccine administration), OAE represents host adverse events that are commonly associated with more than one trigger.

OVAE is designed to function as a knowledgebase of the adverse events known to be associated with administration of licensed vaccines [82]. OVAE is an extension of both OAE and VO. OVAE imports all the licensed vaccines from the VO and also imports related adverse event terms from OAE. Currently, OVAE includes all over 1300 adverse events associated with 63 US-licensed human vaccines. Figure 4 demonstrates how OVAE was generated. Examples of OVAE usages are described in the next section.

Ontology-based data mining for more efficient classification and analysis of vaccine adverse events

The OAE has been used to improve the analysis of VAE case reports collected from the VAERS database [18]. In this case study, those VAEs associated with four trivalent (killed) inactivated influenza vaccine (TIV) and one trivalent live attenuated influenza vaccine (LAIV) were extracted from the VAERS database. After classical statistical analysis, 48 TIV-enriched and 68 LAIV-enriched AEs were identified. The MedDRA terms of these AEs were first mapped to OAE terms. The OAE hierarchical structures were generated for these two subsets of terms. The analysis of these two hierarchies allowed better classification of these terms. Different patterns associated with these VAE subsets were identified. Specifically, TIV-enriched AEs include neurological and muscular processing such as paralysis, movement disorders, and muscular weakness. In contrast, LAIV-enriched AEs are inflammatory response and respiratory system disorders. Furthermore, LAIV was found to have lower chance of inducing two severe adverse events (Guillain-Barre Syndrome and paralysis) than TIV [18].

As described above, OVAE was used to classify VAEs associated with US-licensed human vaccines [82]. OWL-based OVAE is stored in an RDF triple store, a database system based on the Resource Description Framework (RDF). The RDF data model makes statements

about resources in the form of subject-predicate-object expressions (*i.e.*, triples). SPARQL (a recursive acronym for SPARQL Protocol and RDF Query Language) is an RDF query language that can be used to retrieve and manipulate data stored in RDF format. After the OVAE data were stored in a RDF triple store [66], specific SPARQL queries were generated to retrieve useful information, such as the top 10 vaccines associated with the highest numbers of VAEs and the top 10 VAEs most frequently observed among all human vaccines [82].

Expert commentary

To improving vaccine efficacy and safety, the author argues that ontology-based data mining will play a significant role in predicting vaccine candidates, vaccine adverse events, and the molecular mechanisms of protective immunity and adverse events. Specifically, the author proposes the following ontology-related activities:

Proposing minimum data information about vaccine research publication and ontology-based vaccine data representation, sharing, and meta-analysis

The informatics technology is efficient in processing "big data". However, without standard data format, it is difficult to process the big data. The minimum information standard is a set of guidelines developed for reporting experimental data in a specific domain of biosciences. The standard can be used to ensure the data being easily verified, analyzed and clearly interpreted by the wide scientific community. Examples include the Minimum Information About a Microarray Experiment (MIAME) and Minimum Information about a Flow Cytometry Experiment (MIFlowCyt) [83]. MIAME is a standard that describes the minimum information required to ensure the easy interpretation of microarray data and possible independent verification of the analysis results [84]. MIFlowCyt is the minimum information about a flow cytometry experiment. These recommendations lay out the foundation of structuralized databases, public repositories, and data analysis tool development.

There is also a need to develop the minimum information about vaccine research experiments. For example, a vaccine protection experiment involves three basic processes: vaccination, challenge, and protection efficacy assay. All these three processes are implemented on the same host (*e.g.*, some laboratory animal or human). Different variables are needed to fully represent a vaccine used in the vaccination process and a pathogen used in the challenge process. Different protection efficacy assays may be used. A clear representation of temporal periods between processes is also required. In a recent study [85], nearly 20 variables were identified to logically represent the vaccine protection study. The 20 variables provide a prototype of minimal information for analysis of vaccine protection.

To support data integration, ontology can play a role in standardizing the minimum information into ontology format. For example, the MIAME minimum information for microarray experiments is being represented into the Ontology for Biomedical Investigations (OBI) [86]. Ontologies can also be used for data representation and statistical analysis [87]. Similarly, the author proposes to generate the minimum information about vaccine data publication, and further represent the minimum vaccine data information using an ontology-

based strategy. The combination of VO and OBI can be used as the basis for the ontology representation. A community-based effort of developing and standardizing the minimum information about vaccine data, and applying it with an ontology-based approach would significantly enhance the high throughput vaccine data processing and automated reasoning, supporting rational vaccine design and fundamental vaccine mechanism analysis.

One possible usage of the minimum information standardization and ontology representation is its support on data integration and statistical "meta-analysis" of integrated data. The term "meta-analysis" was coined by Glass in the 1970s to describe the process of gathering and combining information from many studies of the same type [88]. Outcomes from a pooled meta-analysis may include a more precise estimate of the effect of biological factors or outcomes than could be obtained from an individual study. Meta-analysis can also be used to investigate new hypotheses [89,90]. Although meta-analysis has been widely used in biomedical research [88,91,92], how ontologies can be used to advance meta-analysis has not been systematically explored. Different results in different studies may be affected by different experimental variables (or conditions) and specific data of these experimental variables. For better meta-analysis, it is critical to standardize the data as well as the experimental conditions. The ontology-based data standardization supports standard variable representation, data normalization, and advanced meta-analysis.

One example of ontology-based statistical meta-analysis is an ontology modeling and meta-analysis of different variables contributing to the vaccine efficacy of whole organism *Brucella* vaccines using the mouse model [85,87]. In this study, based on the identification of 19 variables (*e.g.*, vaccination route, challenge dose) relevant to the vaccine protection study, 401 experimental groups were manually annotated from 74 peer-reviewed publications. An Analysis of variance (ANOVA) meta-analysis on a whole organism *Brucella* vaccine study based on the collection and analysis of data related to these variables allowed us to statistically compute which variables significantly contributed to the vaccine protection efficacy, and which variables did not. Such a study also found new counter-intuitive results. For example, the meta-analysis has demonstrated that intraperitoneal and subcutaneous vaccinations are much more effective in protecting against aerosol *Brucella* challenge compared to intranasal vaccination [85].

The ontology-based statistical meta-analysis as exemplified in the above *Brucella* vaccine efficacy use case [85,87] can also be applied for complex Omics meta-analysis such as those related to the systems vaccinology studies [27]. Basically, the ontology methods can be used to standardize the variables and relations among variables, normalize related data types, and provide a platform for standardized and efficient statistical data analysis.

Proposing an OBO Ontologies-based Linked Open Data framework (OBO-LOD) and its usage in the vaccine domain (LODV)

To address a major obstacle in "big data" integration, the Linked Data (LD) strategy that extends the Web by publishing and linking various datasets [93] can be used. Dr. Tim Berners-Lee, the inventor of WWW, outlined four principles of linked data [94]: (i) Use uniform resource identifiers (URIs) to denote things. (ii) Use HTTP URIs so that these things can be referred to and looked up ("dereferenced") by people and user agents. (iii)

Provide useful information about the thing when its URI is dereferenced by using standards such as RDF. (iv) Include links to other related things (using their URIs) when publishing data on the Web, so that more things can be discovered. Furthermore, linked data are preferred to be open, leading to the Linked Open Data (LOD).

Many different LD or LOD systems, such as Bio2RDF [95] and DBpedia [96], have been developed. These RDF links between data items can come from different data sources and be accessed anywhere through the web [94]. However, since individual LOD systems tend to use their own ontologies and standards, the difficulty in data exchange and query between different LODs has become a new obstacle. To address this challenge, the author proposes to use the community-based OBO Foundry ontologies as the framework for both "class" (or called "universal") and "instance" (or called "particular") data representation. Such the OBO Foundry-based LOD framework (abbreviated as "OBO-LOD" below) can be applied in the vaccine domain to generate a LOD-Vaccine (LODV) system. Figure 5 lays out the upper class level hierarchy of different components used for the LODV system. The Basic Formal Ontology (BFO) is an upper level ontology [52,97] that has been used by over 100 biological and biomedical ontologies. By aligning all specific ontologies with BFO, all ontologies can be integrated seamlessly. Once the class level hierarchy is generated, the instance level data can be represented as instances of corresponding classes. One advantage of such OBO-LOD representation is that the class level hierarchy will provide the schema (like relational database schema) that defines the relations between different instance data. Various data will be stored in RDF triple stores. Since the VO/OBO-based LODV is developed following the community-based OBO ontologies, the vaccine-related data in LODV can be easily shared with the broader biological and biomedical communities.

Many challenges still exist in fully implementing the OBO-LOD framework. Not all domains of knowledge are represented by ontologies. Many ontologies (including the BFO upper level ontology and specific domain ontologies) are still under constant updating and evaluation. How to use the class-level hierarchy to link and analyze instance level experimental data still needs testing. However, these challenges also pose opportunities. The development of OBO-LOD-based LODV would provide a first evaluation of how such a framework will be useful in novel data integration and analysis in the Big Data era.

Proposing the One Network (OneNet) Theory of Life for studying biological & molecular interaction networks

Historically, many scientific theories, such as the Darwin's Theory of Evolution Theory, have significantly advanced scientific biology research. The Evolutionary Synthesis Theory (or the Synthetic Theory), a synopsis of much of contemporary evolutionary theory, illustrates the relationship between genotype and phenotype [98]. Based on this theory, the genotype is a blueprint for an organism that provides the set of instructions for development, and the phenotype is the physical manifestation of a genotype. However, the environment may influence the phenotype outcome [98]. The Cell Theory states that all organisms are composed of one or more cells; the cell is the most basic unit of structure, function, and organization in all organisms; and all cells arise from pre-existing, living cells [99,100].

Many immune system-specific theories have also been proposed. The "Immune Network Theory" explains how the adaptive immune system works [101]. Based on this theory, the immune system is an interacting network of lymphocytes and molecules with variable (V) regions. These V regions bind not only to those things foreign to the vertebrate, but it also binds to other V regions within the system itself. The immune system is therefore a network with the components connected to each other by V-V interactions. Recently, Dr. Poland proposed a theory or hypothesis of the "immune response gene network" [17]. This theory states that the response to a vaccine is the cumulative result of interactions driven by host genes and their interactions. The immune response includes various responses including innate, humoral, cell-mediated responses, and different adverse events. Basic genetic elements of the theory include key immune-response genes necessary for activation and suppression of immune responses, gene expression, gene polymorphism, epigenetic gene modifications, innate response genes, and gene-gene interactions. Compared to the "Immune Network Theory", the "immune response gene network" theory is more focused on "gene" level. The "immune response gene network theory" has been used as the basis for many "vaccinomics" studies, which establish associations between immune response gene polymorphisms and various antibody and cell-mediated immune responses to many viral vaccines including measles-mumps-rubella vaccine, influenza vaccines, hepatitis B, and smallpox vaccines [17].

While the Evolutionary Synthesis Theory and the Cell Theory introduce the principles of how cells and organisms evolve, they do not emphasize the fundamental interaction network mechanisms under the cellular and organismic evolvements. While the Immune Network Theory and the "immune response gene network" theory propose specific network mechanisms and address specific issues in immune systems, they often ignore other systems (*e.g.*, developmental system) as integrated parts of the whole dynamic life. To integrate and extend these theories, the author here proposes a One Network (abbreviated as "OneNet") Theory (or hypothesis) of Life. Basically, the OneNet Theory of Life includes four tenets:

    i.   The whole process of a life is a single complex and dynamic network (called "OneNet").

    ii.   The OneNet blueprint is stored in the genotype of the organism.

    iii.   The OneNet starts at the moment when the first cell of the organism forms.

    iv.   The OneNet of temporal interactions between the genetic materials and their environments determines the dynamic phenotype of the life.

The OneNet Theory treats the life of an organism as an integrated, complex and dynamic network. For a single-cell organism such as *E. coli*, the dynamic OneNet process starts at the moment when the cell forms. For a human life, the OneNet starts at the moment with the formation of the one-celled zygote out of the process of fertilization [102]. The moment of the first cell formation is the "Big Bang" moment of a new life. After the first cell forms, the cell development process will be dynamically regulated by the interactions between the genetic materials and its environmental factors. The environmental factors of the genetic materials include those inside the cell but outside the genetic materials and those outside the cells. All these environmental factors may change the expression of the genotype leading to

different phenotypes in the development and behavior of the new life. Since the whole *E. coli* organism is a single cell, the environmental factors for the whole bacterial organism is the same as the environmental factors for the single cell. In contrast, a mature multi-cell human is composed of a large number of cells that are all originated from a single fertilized cell. These human cells are well-organized into different tissues and organs as parts of the whole human body in an integrated structure. The framework of how individual cells are organized and interact with each other as a whole is also defined by the genotype. The environmental factors for the life of a human are more complex and may be different at various stages from a single fertilized cell to a multi-cell embryo to a fetus to a more independent life outside the mother's womb. While the granularity and complexity of the environmental factors of human genotype require deeper investigation, all the four tenets of the OneNet theory (or hypothesis) likely hold true for the multi-cell human organism.

Therefore, the OneNet theory provides a basis for systematically studying what happens once a life has formed. The systematical and dynamic OneNet network integrates different development and growth stages (or processes) of the whole life. The OneNet theory can be used to explain and study various genotype-phenotype linkages, including the association between vaccination, host genetic variation, and host immune response heterogeneity. The OneNet theory provides a clear systematic framework that may eventually explain vaccine-induced protective immune responses or adverse events by uncovering detailed fundamental mechanisms of dynamic molecular interaction networks. The OneNet theory may also be used to study scientific questions beyond vaccinology and immunology. However, it is noted that there is still a long way to go before we can fully understand how every interaction of the OneNet works from the moment when a cell is fertilized or formed.

While it might be relatively easy to understand the OneNet Theory, how to study and apply this theory can be challenging. The study of the OneNet can rely on computational and statistical methods such as Bayesian network [103-105] and other causality and network analysis methods [106]. The computational or statistical analysis of the OneNet Theory is not discussed here since it goes beyond the scope of this manuscript. Here, the author suggests that ontologies are a powerful approach to studying the OneNet dynamic network. Indeed, the OBO-LOD system introduced in the above section provides an ideal platform for the OneNet data integration and analysis. In the next section, the author proposes an example strategy to study the dynamic interaction networks and how it can be used to study the interactions between vaccine administration and the host immune system.

Proposing ontology-based integrative representation and analysis of the "OneNet" interaction networks

While hundreds of databases of biological interactions and pathways (such as Reactome [107], Kyoto Encyclopedia of Genes and Genomes (KEGG) [108], BioCyc [109], and BioCarta [110]) have been developed [111], the knowledge and data in these databases are not integrated. One major reason for the lack of integration is different representation formats used by different databases. Some common formats, such as BioPAX [55], are used for data exchange. The BioPAX ontology format represents basic high level pathways, associated interactions, and entities participating in the interactions and pathways. Individual

Reactome interactions and pathways are then represented as instances of BioPAX ontology classes. While the approach of listing individual pathways as instances in OWL is reasonable for data exchange, different levels of pathways (*e.g.*, apoptosis pathways), reactions (*e.g.*, phosphorylation), and entities (*e.g.*, protein) are universals (or classes) instead of particulars (or instances). Given verified experimental conditions, these pathways always occur. Therefore, they are more appropriately represented as classes instead of instances in the scope of ontology. The BioPAX representation of individual interactions/pathways as instances does not solve the issue of data integration among different databases. Therefore, the author proposes the development of the Human Interaction Network Ontology (HINO) [112] that integrates data from different interaction pathway databases (*e.g.*, Reactome and KEGG) and represents individual genes, molecules, interactions, and pathways as non-redundant classes instead of instances [112]. Such an ontology can eventually be expanded to fully represent the OneNet of a human being. The same strategy can also be applied to study other organisms (*e.g.*, *E. coli*). Similar to the Periodic Table of Chemical Elements that systematically organizes chemical elements, the ontology logically represents known interactions and pathways and may eventually be used to predict unknown elements in the interaction network.

In addition to integrating annotated knowledge from existing databases, we may also need to incorporate the experimental results published in the vast amount of literature. Ontologies as knowledge bases of annotated interaction networks (including vaccine-induced host response pathways) can be used as a prior for further improvements. Ontologies can also be applied to enhance the performance of literature mining in the steps of Reading, Assembly and Explaining [113] (Figure 6). The advantages of using ontologies in literature mining come from formal logical representation of asserted and inferred hierarchies by ontologies. All the ontologies (including HINO) are aligned with the upper-level Basic Formal Ontology (BFO) (Figure 5). Different domain ontologies can be used for name tagging of specific types of text in the literature. For example, chemical entities can be tagged in literature text with the Chemical Entities of Biological Interest (ChEBI) [114], gene names tagged with the Ontology of Genes and Genomes (OGG) [115], and cell types and cell line cells tagged with the Cell Ontology (CL) [116] and Cell Line Ontology (CLO) [117]. The Ontology for Biomedical Investigations (OBI) [86] can be used for tagging those terms related to experimental assays and conditions. In addition to the noun name tagging, those verbs or keywords indicating the interaction types can be tagged using the Interaction Network Ontology (INO) [64]. The literature extracted molecular interactions can then be assembled into the HINO basis using bioinformatic and statistical methods. Ontology-based inferences can be used for different applications, such as resolving inconsistency and predicting hypotheses. An ontology-based integrated software system can be generated to integrate different technical methods into a pipeline for efficient literature mining and reasoning. Newly predicted hypotheses can be selectively verified using wet-lab experimental methods. Such a systematic system can thus result in the generation of new knowledge (Figure 6). However, it may take a long time to have a comprehensive understanding of human OneNet interactions and pathways.

Ontology-based software tool development and applications

To increase the development and applications of ontologies, different software programs are needed. Many ontology-oriented tools have been developed for various activities in ontology development and applications [51,66,118,119]. Using the VO development and applications as model, many tools have been developed to support ontology development and applications [51,66,118,119] (Figure 7). Some of these tools, including Ontobee [66], OntoFox [51], and the Hegroup triple store [66], have been widely used beyond the vaccine community. For example, during the past three years, OntoFox and Ontobee have been visited by >3,800 and >21,000 unique visitors, respectively.

Expert commentary

In summary, in the section of Expert Commentary, the author envisions and proposes new strategies and their ontology-based representations and implementations. To support better vaccine data analysis, the author proposes the development of minimal vaccine information standards for reporting vaccine studies. For computer-assisted automated processing, the minimal vaccine information standards can be represented using ontology formats. Different ontology-based meta-analysis can then be developed for advanced statistical data analysis. Ontologies are also the foundation of linked (open) data. To support data integration in the big data era, the OBO Foundry ontologies can be integrated under the Basic Formal Ontology (BFO) to provide an ontological framework for constructing and integrating various LODs, including the LOD in the vaccine domain (LODV) (Figure 5). For a better understanding of vaccine immune mechanisms, the author proposes an integrative One Network (OneNet) theory. The OneNet theory can be used as guidance for generating an ontology-based representation of the OneNet interactions and networks of an integrative and dynamic life. The ontology-based data representation and mining based on the theory will be critical to better predict vaccine candidates and understand the molecular mechanisms of protective immunity and adverse event responses (Figure 6). To support ontology development and applications, ontology-oriented software programs can be generated (Figure 7). Since these proposed strategies are formed under a general integrative view, these strategies are likely applicable for studies in scientific areas beyond vaccine immunology.

Five-year view

As demonstrated in this article, new technologies such as Omics and systems biology have been used to generate a large amount of data in the field of vaccinology. The publication rate of scientific articles has also dramatically increased in the last decade. The usage of computers plays an important role in big data analysis. The computational analysis of the big high-throughput and literature data has resulted in rational design of new or improved vaccines or counter measurements of adverse events. However, it is often challenging to analyze big data since the big data can be very noisy. To more efficiently use computational technologies, data are required to be integrated and organized in a standardized way. Since ontologies logically represent terms and relations in a human- and computer-interpretable format, ontology-based data mining provides a new way for advanced computer-assisted data mining.

The biomedical ontology research is a booming research area. In the vaccine domain, the community-based Vaccine ontology (VO) and the Ontology of Vaccine Adverse Events (OVAE) have been developed to support various vaccine research and applications. These studies are still at its early stage, and more efforts are required to make the ontologies better and more useful. In five years, the author anticipates that much progress will be made in generation of improved and new ontologies, development of ontology-based standards and software programs, and their applications in scientific discoveries. More use cases in the vaccinology or other research areas are expected to be studied and published to demonstrate the advantages and usages of ontology-based data mining strategies.

The OneNet theory considers the whole life of an organism as a single complex and dynamic network. Such a theory and its associated ontology representations provide a feasible framework for systematically organizing, integrating and analyzing fundamental mechanisms of biological interactions and networks. For the vaccine case, a human vaccination process initiates an interaction between a foreign antigen and the human host, and it induces a series of human responses including innate and adoptive immune responses. Studies with monozygotic twins (*i.e.*, identical twins) that share the same genotypes can be useful in investigating how different vaccination exposures contribute to the differences in immune responses. Similarly, given the same vaccination exposures, the genetic differences between dizygotic twins (*i.e.*, non-identical twins) may cause differences in positive protective immunity or negative adverse event responses [120]. New methods such as Omics and next-generation sequencing can be used to explore how individual vaccines work when they are administered into a human body *in vivo* or a population of cells *in vitro*. However, given that we just start to use new systems biology technologies and ontology-based data integration/analysis approaches, to understand every detail of the OneNet interaction network mechanism starting from the single cell formation is expected to remain a goal for a long period of time in scientific research.

Conclusion

In the big data era, how to use big data to improve vaccine efficacy and vaccine safety has become critical in vaccine research. In this article, the author summarizes existing progress and challenges in these areas. Basic introduction of ontologies and ontology applications in the vaccine research is provided. The author has also proposed many ontology-based strategies and approaches to better study vaccine-related immune network and other biological mechanisms. In summary, the following key issues are introduced and discussed.

Acknowledgments

References

Papers of special note have been highlighted as:

• of interest

•• of considerable interest

1••. He Y, Rappuoli R, De Groot AS, Chen RT. Emerging vaccine informatics. J Biomed Biotechnol. 2010; 2010:218590. Vaccine informatics has become an emerging field of research. Different informatical and computational methods have been developed to promote *in silico* vaccine candidate prediction and vaccine safety analysis. [PubMed: 21772787]

2. Pizza M, Scarlato V, Masignani V, et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science. 2000; 287(5459):1816–1820. [PubMed: 10710308]

3. Althoff, E.; Gassenbach, N. Novartis submits Bexsero®, a multi-component meningococcal B vaccine, for regulatory review in Europe. Novartis Media releases. 2010. Available from:www.novartis.com/newsroom/media-releases/en/2010/1475256.shtml

4. He Y, Xiang Z, Mobley HL. Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. J Biomed Biotechnol. 2010; 2010:297505. [PubMed: 20671958]

5. He Y, Xiang Z. Bioinformatics analysis of *Brucella* vaccines and vaccine targets using VIOLIN. Immunome Res. 2010; 6(Suppl 1):S5. [PubMed: 20875156]

6. Xiang Z, He Y. Vaxign: a web-based vaccine target design program for reverse vaccinology. Procedia in Vaccinology. 2009; 1(1):23–29.

7. Gomez G, Pei J, Mwangi W, Adams LG, Rice-Ficht A, Ficht TA. Immunogenic and invasive properties of *Brucella melitensis* 16M outer membrane protein vaccine candidates identified via a reverse vaccinology approach. PloS one. 2013; 8(3):e59751. [PubMed: 23533646]

8. He, Y. Genome-based computational vaccine discovery by reverse vaccinology. In: Flower, editor. Immunomic Discovery of Adjuvants and Candidate Subunit Vaccines. Springer; New York: 2013.

9. Pereira UP, Soares SC, Blom J, et al. In silico prediction of conserved vaccine targets in Streptococcus agalactiae strains isolated from fish, cattle, and human samples. Genet Mol Res. 2013; 12(3):2902–2912. [PubMed: 24065646]

10•. Xiang Z, He Y. Genome-wide prediction of vaccine targets for human herpes simplex viruses using Vaxign reverse vaccinology. BMC bioinformatics. 2013; 14(Suppl 4):S2. Vaxign is the first web-based vaccine candidate prediction program based on the reverse vaccinology strategy. [PubMed: 23514126]

11. Van Regenmortel MH. Basic research in HIV vaccinology is hampered by reductionist thinking. Front Immunol. 2012; 3:194. [PubMed: 22787464]

12. Hur J, Xiang Z, Feldman EL, He Y. Ontology-based *Brucella* vaccine literature indexing and systematic analysis of gene-vaccine association network. BMC immunology. 2011; 12:49. [PubMed: 21871085]

13. Bagnoli F, Baudner B, Mishra RP, et al. Designing the next generation of vaccines for global public health. OMICS. 2011; 15(9):545–566. [PubMed: 21682594]

14. He Y. Analyses of *Brucella* Pathogenesis, Host Immunity, and Vaccine Targets using Systems Biology and Bioinformatics. Frontiers in cellular and infection microbiology. 2012; 2:2. [PubMed: 22919594]

15. Sayers S, Ulysse G, Xiang Z, He Y. Vaxjo: a web-based vaccine adjuvant database and its application for analysis of vaccine adjuvants and their uses in vaccine development. J Biomed Biotechnol. 2012; 2012:831486. [PubMed: 22505817]

16. Racz R, Li X, Patel M, Xiang Z, He Y. DNAVaxDB: the first web-based DNA vaccine database and its data analysis. BMC bioinformatics. 2014; 15(Suppl 4):S2. [PubMed: 25104313]

17•. Poland GA, Ovsyannikova IG, Jacobson RM. Application of pharmacogenomics to vaccines. Pharmacogenomics. 2009; 10(5):837–852. The "immune response gene network" hypothsis (or theory) and vaccinomics were discussed and used to the genetic association between immune response gene variations and host immune responses to vaccines. [PubMed: 19450131]

18••. Sarntivijai S, Xiang Z, Shedden KA, et al. Ontology-based combinatorial comparative analysis of adverse events associated with killed and live influenza vaccines. PloS one. 2012; 7(11):e49941. The Ontology of Adverse Events (OAE) was used to classify adverse events associated with killed or live attenuated influenza vaccines. [PubMed: 23209624]

19. Bonhoeffer J, Black S, Izurieta H, Zuber P, Sturkenboom M. Current status and future directions of post-marketing vaccine safety monitoring with focus on USA and Europe. Biologicals. 2012; 40(5):393–397. [PubMed: 22902972]

20. Vaccine adverse event reporting system. Available from: www.vaers.hhs.gov

21. Chen RT, Rastogi SC, Mullen JR, et al. The Vaccine Adverse Event Reporting System (VAERS). Vaccine. 1994; 12(6):542–550. [PubMed: 8036829]

22. Varricchio F, Iskander J, Destefano F, et al. Understanding vaccine safety information from the Vaccine Adverse Event Reporting System. The Pediatric infectious disease journal. 2004; 23(4): 287–294. [PubMed: 15071280]

23. VAERS frequently asked questions. Available from: www.vaers.hhs.gov/about/faqs

24. Brown EG. Methods and pitfalls in searching drug safety databases utilising the Medical Dictionary for Regulatory Activities (MedDRA). Drug safety. 2003; 26(3):145–158. [PubMed: 12580645]

25. Brown SH, Elkin PL, Bauer BA, et al. SNOMED CT: utility for a general medical evaluation template. AMIA Annual Symposium proceedings. 2006:101–105. [PubMed: 17238311]

26. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10(1):57–63. [PubMed: 19015660]

27. Nakaya HI, Li S, Pulendran B. Systems vaccinology: learning to compute the behavior of vaccine induced immunity. Wiley Interdiscip Rev Syst Biol Med. 2012; 4(2):193–205. [PubMed: 22012654]

28. Pulendran B, Li S, Nakaya HI. Systems vaccinology. Immunity. 2010; 33(4):516–529. [PubMed: 21029962]

29•. Nakaya HI, Wrammert J, Lee EK, et al. Systems biology of vaccination for seasonal influenza in humans. Nat Immunol. 2011; 12(8):786–795. The systems vaccinology strategy was used to study human cell responses to seasonal influenza vaccination. [PubMed: 21743478]

30. Poland GA, Ovsyannikova IG, Jacobson RM. Personalized vaccines: the emerging field of vaccinomics. Expert Opin Biol Ther. 2008; 8(11):1659–1667. [PubMed: 18847302]

31. Poland GA, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Jacobson RM. Vaccinomics and a new paradigm for the development of preventive vaccines against viral infections. OMICS. 2011; 15(9):625–636. [PubMed: 21732819]

32. He Y, Xiang Z. Databases and in silico tools for vaccine design. Methods in molecular biology. 2013; 993:115–127. [PubMed: 23568467]

33. US FDA vaccines. Available from: www.fda.gov/BiologicsBloodVaccines/Vaccines/default.htm

34. USDA Veterinary Biological Products, Licensees and Permittees, Prepared April 9, 2014. Available from: www.aphis.usda.gov/animal_health/vet_biologics/publications/CurrentProdCodeBook.pdf

35. Edlich RF, Martin ML, Foley ML, et al. Vaccine information statements. Revolutionary but neglected educational advances in healthcare in the United States. J Long Term Eff Med Implants. 2005; 15(1):91–114. [PubMed: 15715520]

36. Vaccine resource library. Available from: www.path.org/vaccineresources/

37. Saleh JA, Yusuph H, Zailani SB, Aji B. Malaria vaccine: the pros and cons. Niger J Med. 2010; 19(1):8–13. [PubMed: 20232750]

38. Vita R, Zarebski L, Greenbaum JA, et al. The immune epitope database 2.0. Nucleic Acids Res. 2010; 38(Database issue):D854–862. [PubMed: 19906713]

39. Immune epitope database and analysis resource. Available from: www.iedb.org

40. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics. 1999; 50(3-4):213–219. [PubMed: 10602881]

41. the international ImMunoGeneTics information system. Available from: www.imgt.org

42. Lefranc MP. Immunoglobulin and T Cell Receptor Genes: IMGT((R)) and the Birth and Rise of Immunoinformatics. Front Immunol. 2014; 5:22. [PubMed: 24600447]

43. The Nonhuman Primate HIV/SIV Vaccine Trials Database. Available from: www.hiv.lanl.gov/content/vaccine/home.html

44. VIOLIN. Available from: www.violinet.org

45. Xiang Z, Todd T, Ku KP, et al. VIOLIN: vaccine investigation and online information network. Nucleic Acids Res. 2008; 36(Database issue):D923–928. [PubMed: 18025042]

46•. He Y, Racz R, Sayers S, et al. Updates on the web-based VIOLIN vaccine database and analysis system. Nucleic Acids Res. 2014; 42(1):D1124–1132. VIOLIN is the largest online database with a focus on vaccine research. [PubMed: 24259431]

47. Yang B, Sayers S, Xiang Z, He Y. Protegen: a web-based protective antigen database and analysis system. Nucleic Acids Res. 2011; 39(Database issue):D1073–1078. [PubMed: 20959289]

48. Jaiswal V, Chanumolu SK, Gupta A, Chauhan RS, Rout C. Jenner-predict server: prediction of protein vaccine candidates (PVCs) in bacteria based on host-pathogen interactions. BMC bioinformatics. 2013; 14:211. [PubMed: 23815072]

49. W3C. OWL 2 Web Ontology Language document overview (Second edition). 2012. www.w3.org/TR/owl2-overview/

50. Tirmizi SH, Aitken S, Moreira DA, et al. Mapping between the OBO and OWL ontology languages. J Biomed Semantics. 2011; 2(Suppl 1):S3. [PubMed: 21388572]

51. Xiang Z, Courtot M, Brinkman RR, Ruttenberg A, He Y. OntoFox: web-based support for ontology reuse. BMC Res Notes. 2010; 3:175, 1–12. [PubMed: 20569493]

52••. Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature biotechnology. 2007; 25(11):1251–1255. OBO Foundry and its ontology principles lay out a foundation for open and collaborative ontology development.

53. NCBITaxon: An ontology representation of the NCBI organismal taxonomy. Available from: obofoundry.org/ontology/ncbitaxon.html

54. Rosse C, Mejino JL Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. Journal of biomedical informatics. 2003; 36(6):478–500. [PubMed: 14759820]

55. Demir E, Cary MP, Paley S, et al. The BioPAX community standard for pathway data sharing. Nature biotechnology. 2010; 28(9):935–942.

56. Smith B, Ceusters W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. Applied Ontology. 2010; 5:139–188. [PubMed: 21637730]

57. He, Y.; Cowell, L.; Diehl, AD., et al. The 1st International Conference on Biomedical Ontology (ICBO-2009). Buffalo, NY, USA: 2009. VO: Vaccine Ontology. Nature Precedings: precedings.nature.com/documents/3552/version/1

58. Lin Y, He Y. Ontology representation and analysis of vaccine formulation and administration and their effects on vaccine immune responses. J Biomed Semantics. 2012; 3(1):17. [PubMed: 23256535]

59. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000; 25(1):25–29. [PubMed: 10802651]

60. Bettembourg C, Diot C, Burgun A, Dameron O. GO2PUB: Querying PubMed with semantic expansion of gene ontology terms. J Biomed Semantics. 2012; 3(1):7. [PubMed: 22958570]

61. Doms A, Schroeder M. GoPubMed: exploring PubMed with the Gene Ontology. Nucleic Acids Res. 2005; 33(Web Server issue):W783–786. [PubMed: 15980585]

62. Plake C, Royer L, Winnenburg R, Hakenberg J, Schroeder M. GoGene: gene annotation in the fast lane. Nucleic Acids Res. 2009; 37(Web Server issue):W300–304. [PubMed: 19465383]

63. Hur J, Ozgur A, Xiang Z, He Y. Identification of fever and vaccine-associated gene interaction networks using ontology-based literature mining. J Biomed Semantics. 2012; 3(1):18. [PubMed: 23256563]

64••. Ozgur A, Xiang Z, Radev DR, He Y. Mining of vaccine-associated IFN-gamma gene interaction networks using the Vaccine Ontology. J Biomed Semantics. 2011; 2(Suppl 2):S8. Vaccine Ontology provides support for enhanced literature mining of genetic interaction networks. [PubMed: 21624163]

65. The VO in the OBO Foundry ontology library. Available from: www.obofoundry.org/ontology/vo.html

66. Xiang, Z.; Mungall, C.; Ruttenberg, A.; He, Y. Ontobee: A linked data server and browser for ontology terms. The 2nd International Conference on Biomedical Ontologies (ICBO); CEUR Workshop Proceedings; Buffalo, NY, USA. 2011; p. 279-281.

67. VO statistics shown in Ontobee. Available from: www.ontobee.org/ontostat/VO

68. Kazakov, Y.; Krötzsch, M.; Simancík, F. ELK reasoner: architecture and evaluation. Proceedings of the OWL Reasoner Evaluation Workshop (ORE 2012); Manchester, UK. 2012; 12 pages

69. Sirin E, Parsia B, Grau BC, Kalyanpur A, Katz Y. Pellet: A practical OWL-DL reasoner. Web Semant. 2007; 5:51–53.

70. Tsarkov D, Horrocks I. FaCT++ Description Logic Reasoner: System Description. Proc. of the Int. Joint Conf. on Automated Reasoning (IJCAR 2006). 4130 of Lecture Notes in Artificial Intelligence. 2006:292–297.

71. Zhang Y, Tao C, He Y, Kanjamala P, Liu H. Network-based analysis of vaccine-related associations reveals consistent knowledge with the vaccine ontology. J Biomed Semantics. 2013; 4(1):33. [PubMed: 24209834]

72. Lipscomb CE. Medical Subject Headings (MeSH). Bull Med Libr Assoc. 2000; 88(3):265–266. [PubMed: 10928714]

73. Srinivasan P. MeSHmap: a text mining tool. for MEDLINE Proc AMIA Symp. 2001:642–646. [PubMed: 11825264]

74. Djebbari A, Karamycheva S, Howe E, Quackenbush J. MeSHer: identifying biological concepts in microarray assays based on PubMed references and MeSH terms. Bioinformatics. 2005; 21(15): 3324–3326. [PubMed: 15919728]

75. Osborne JD, Zhu LJ, Lin SM, Kibbe WA. Interpreting microarray results with gene ontology and MeSH. Methods in molecular biology. 2007; 377:223–242. [PubMed: 17634620]

76. Hur J, Schuyler AD, States DJ, Feldman EL. SciMiner: web-based literature mining tool for target identification and functional enrichment analysis. Bioinformatics. 2009; 25(6):838–840. [PubMed: 19188191]

77. Genomesh. Available from: www.hegroup.org

78. Xiang Z, Qin T, Qin Z, He Y. A genome-wide MeSH-based literature mining system predicts implicit gene-to-gene relationships and networks. BMC systems biology. 2013; 7(Suppl 3):S9. [PubMed: 24555475]

79. Joachims, T. Making large-scale support vector machine learning practical. In: Schölkopf, B.; B, CJ.; Smola, AJ., editors. Advances in Kernel Methods: Support Vector Learning. MIT Press; Cambridge, MA: 1999. p. 169-184.

80. Ozgur A, Xiang Z, Radev D, He Y. Literature-based discovery of IFN-γ and vaccine-mediated gene interaction networks. Journal of Biomedicine and Biotechnology. 2010 2010:Article ID 426479 (426413 pages).

81. He, Y.; Xiang, Z.; Sarntivijai, S.; Toldo, L.; Ceusters, W. AEO: a realism-based biomedical ontology for the representation of adverse events. Adverse Event Representation Workshop, International Conference on Biomedical Ontologies (ICBO-2011); CEUR Workshop Proceedings; Buffalo, NY, USA. 2011; p. 309-315.available from: ceur-ws.org/Vol-833/paper359.pdf

82••. Marcos E, Zhao B, He Y. The Ontology of Vaccine Adverse Events (OVAE) and its usage in representing and analyzing adverse events associated with US-licensed human vaccines. J Biomed Semantics. 2013; 4:40. The OVAE provides a novel way to annotate and classify adverse events associated with those licensed vaccines. [PubMed: 24279920]

83. Lee JA, Spidlen J, Boyce K, et al. MIFlowCyt: the minimum information about a Flow Cytometry Experiment. Cytometry A. 2008; 73(10):926–930. [PubMed: 18752282]

84. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nature genetics. 2001; 29(4):365–371. [PubMed: 11726920]

85. Todd TE, Tibi O, Lin Y, et al. Meta-analysis of variables affecting mouse protection efficacy of whole organism *Brucella* vaccines and vaccine candidates. BMC bioinformatics. 2013; 14(Suppl 6):S3. [PubMed: 23735014]

86. Brinkman RR, Courtot M, Derom D, et al. Modeling biomedical experimental processes with OBI. J Biomed Semantics. 2010; 1(Suppl 1):S7. [PubMed: 20626927]

87. He, Y.; Xiang, Z.; Todd, T., et al. Ontology representation and ANOVA analysis of vaccine protection investigation. Bio-Ontologies 2010: Semantic Applications in Life Sciences; CEUR

Workshop Proceedings; 11-13 July 2010; Boston, MA, USA. 2010. p. 1-8. Available from: ceur-ws.org/Vol-754/he_krmed2010.pdf

88. Glass GV. Primary, secondary, and meta-analysis of research. Educational Research. 1976; 5:3–8.

89. Altman DG. Statistics in medical journals: some recent trends. Statistics in medicine. 2000; 19(23): 3275–3289. [PubMed: 11113959]

90. Lean IJ, Rabiee AR, Duffield TF, Dohoo IR. Invited review: Use of meta-analysis in animal health and reproduction: methods and applications. J Dairy Sci. 2009; 92(8):3545–3565. [PubMed: 19620636]

91. Lukk M, Kapushesky M, Nikkila J, et al. A global map of human gene expression. Nature biotechnology. 2010; 28(4):322–324.

92. Zheng-Bradley X, Rung J, Parkinson H, Brazma A. Large scale comparison of global gene expression patterns in human and mouse. Genome biology. 2010; 11(12):R124. [PubMed: 21182765]

93. Bizer C, Heath T, Berners-Lee T. Linked Data - The Story So Far. International Journal of Semantic Web and Information Systems. 2009; 5(3):22.

94. Design Issues: Linked Data. Available from: www.w3.org/DesignIssues/LinkedData

95. Belleau F, Nolin MA, Tourigny N, Rigault P, Morissette J. Bio2RDF: towards a mashup to build bioinformatics knowledge systems. Journal of biomedical informatics. 2008; 41(5):706–716. [PubMed: 18472304]

96. Yamamoto Y, Yamaguchi A, Yonezawa A. Building Linked Open Data towards integration of biomedical scientific literature with DBpedia. J Biomed Semantics. 2013; 4(1):8. [PubMed: 23497538]

97. Grenon P, Smith B. SNAP and SPAN: Towards Dynamic Spatial Ontology. Spatial Cognition and Computation. 2004; 4(1):69–103.

98. Futuyma, DJ. Evolutionary Biology. 3rd. Sinauer Associates; USA: 1997.

99. Baluska F, Volkmann D, Barlow PW. Eukaryotic cells and their cell bodies: Cell Theory revised. Ann Bot. 2004; 94(1):9–32. [PubMed: 15155376]

100. Mazzarello P. A unifying concept: the history of cell theory. Nat Cell Biol. 1999; 1(1):E13–15. [PubMed: 10559875]

101. Perelson AS. Immune network theory. Immunol Rev. 1989; 110:5–36. [PubMed: 2477327]

102. Moore, KL. Essentials of Human Embryology. B.C. Decker, Inc.; Toronto, Canada: 1988.

103. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. Journal of computational biology : a journal of computational molecular cell biology. 2000; 7(3-4):601–620. [PubMed: 11108481]

104. He, Y.; Xiang, Z. miniTUBA: a web-based dynamic Bayesian network analysis system and an application in host-pathogen interaction analysis. In: Rebai, A., editor. Bayesian Network. Sciyo; 2010. Available fro: www.intechopen.com/articles/show/title/minituba-a-web-based-dynamic-bayesian-network-analysis-system-

105. Hodges AP, Dai D, Xiang Z, Woolf P, Xi C, He Y. Bayesian network expansion identifies new ROS and biofilm regulators. PloS one. 2010; 5(3):e9513. [PubMed: 20209085]

106. Pearl, J. Causality. 2nd. Cambridge University Press; 2009.

107. Croft D, O'Kelly G, Wu G, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011; 39(Database issue):D691–697. [PubMed: 21067998]

108. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004; 32(Database issue):D277–280. [PubMed: 14681412]

109. Latendresse M, Paley S, Karp PD. Browsing metabolic and regulatory networks with BioCyc. Methods in molecular biology. 2012; 804:197–216. [PubMed: 22144155]

110. BioCarta Apoptosis pathway. Available from: www.biocarta.com/genes/Apoptosis.asp

111. HarshaRani GV, Vayttaden SJ, Bhalla US. Electronic data sources for kinetic models of cell signaling. J Biochem. 2005; 137(6):653–657. [PubMed: 16002985]

112••. He Y, Xiang Z. HINO: a BFO-aligned ontology representing human molecular interactions and pathways. arXiv. 2013 doi:arXiv:1311.3355. HINO represents a first effort to integrate OBO

Foundry ontologies and develop new ontologies to logically represent interactions and pathways in a non-redundant and systematic ontology approach.

113. DARPA. Big mechanism. DARPA Broad Agency Announcment. 2014 DARPA-BAA-14-14.

114. Hastings J, de Matos P, Dekker A, et al. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res. 2013; 41(Database issue):D456–463. [PubMed: 23180789]

115. Announcement of the Ontology of Genes and Genomes (OGG). Available from: groups.google.com/forum/#!topic/ogg-discuss/wy0132CCdNA

116. Diehl AD, Augustine AD, Blake JA, et al. Hematopoietic cell types: prototype for a revised cell ontology. Journal of biomedical informatics. 2011; 44(1):75–79. [PubMed: 20123131]

117. Sarntivijai S, Lin Y, Xiang Z, et al. CLO: The Cell Line Ontology. J Biomed Semantics. 2014; 5:37. [PubMed: 25852852]

118. Xiang, Z.; Lin, Y.; He, Y. Ontorat web server for automatic generation and annotations of new ontology terms. The 3rd International Conference on Biomedical Ontologies (ICBO); CEUR Workshop Proceedings; Graz, Graz, Austria. 2012. Available from: ceur-ws.org/Vol-897/poster_12.pdf

119. Zheng J, Xiang Z, J SJC, He Y. Ontodog: A web-based ontology community view generator. Bioinformatics. 2013; 30(9):1340–1342. [PubMed: 24413522]

120. Tan PL, Jacobson RM, Poland GA, Jacobsen SJ, Pankratz VS. Twin studies of immunogenicity--determining the genetic contribution to vaccine failure. Vaccine. 2001; 19(17-19):2434–2439. [PubMed: 11257374]

121. The protege ontology editor. Available from: protege.stanford.edu/

Key issues

- Vaccine efficacy and safety research has dramatically progressed with the methods of in silico prediction, literature mining, and bioinformatic analysis

- The community-based Vaccine Ontology (VO) supports vaccine data integration and enhanced vaccine literature mining.

- The Ontology of Adverse Events (OAE) represents various adverse events and supports classification and data analysis of clinically reported adverse events.

- The Ontology of Vaccine Adverse Events (OVAE) represents known vaccine adverse events as a knowledge base and supports automated query and analysis of know vaccine adverse events.

- To support vaccine data integration, minimal information about vaccine research can be generated and formulated with ontology support.

- Ontology-based statistical meta-analysis will allow more advanced meta-analysis of the relations between different variables using data from different studies.

- An OBO Foundry ontologies-based linked open data (LOD) (OBO-LOD) is proposed for biological data representation and integration, and its application in the vaccine domain (LODV) will support better vaccine data standardization and its integration and sharing with other research domains.

- The author proposes to integrate and extend existing theories to form the "OneNet" Theory of Life for better understanding of fundamental biological interaction networks including vaccine immune mechanism.

- To apply the "OneNet", the author proposes to develop ontologies (e.g., the human interaction network ontology or HINO) to represent non-redundant interaction pathways by integrating knowledge collected in databases and reading/assembling experimentally verified data from published literature reports.

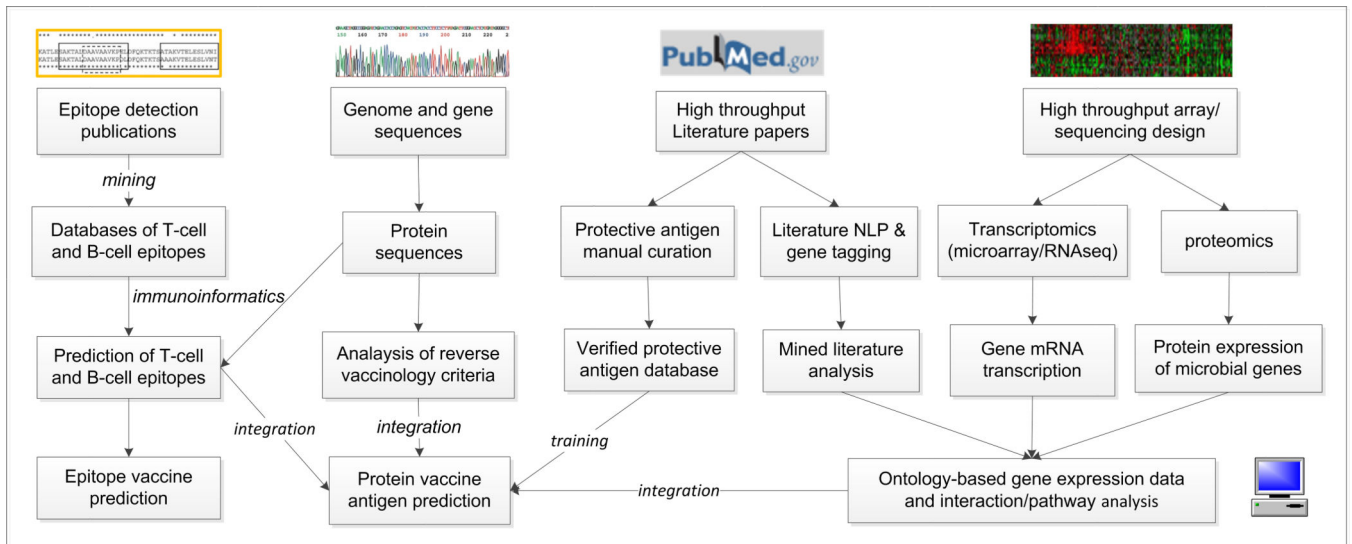- Ontology-based software programs are being developed to support ontology generation and applications.

Figure 1. *in silico* protective vaccine antigen prediction
Different methods can be used for predicting protective antigens for vaccine development.
Specifically, T- and B-cell epitopes can be predicted using protein sequences as input based
on training data available from existing immune epitope databases. Bioinformatic analysis of
microbial genome sequences allows *in silico* prediction of protective vaccine antigens using
the reverse vaccinology strategy. The database of verified protective antigens (*e.g.*, Protegen)
provides gold standard positive antigens for various programs to evaluate. The gene
expression data, coming from transcriptomics (DNA microarray or RNA-seq), proteomics,
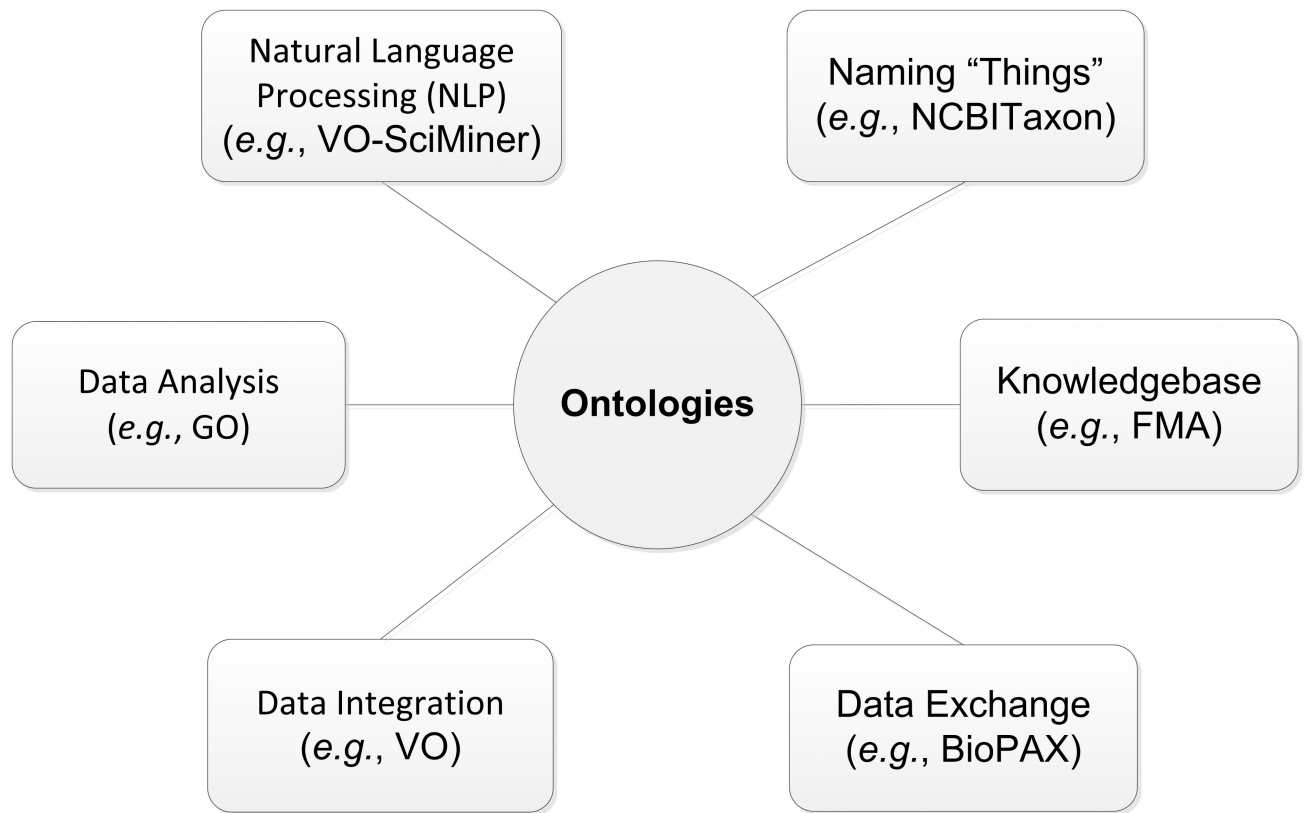or literature mining can be analyzed and used for further protein vaccine antigen predictions.

Figure 2. Various ontology-based applications
The details of different applications are described in the text. Abbreviations: NCBITaxon: NCBI Taxonomy ontology; FMA: Foundational Model of Anatomy; BioPAX: Biological Pathway Exchange; VO: Vaccine Ontology; GO: Gene Ontology; VO-SciMiner: VO-based SciMiner program.
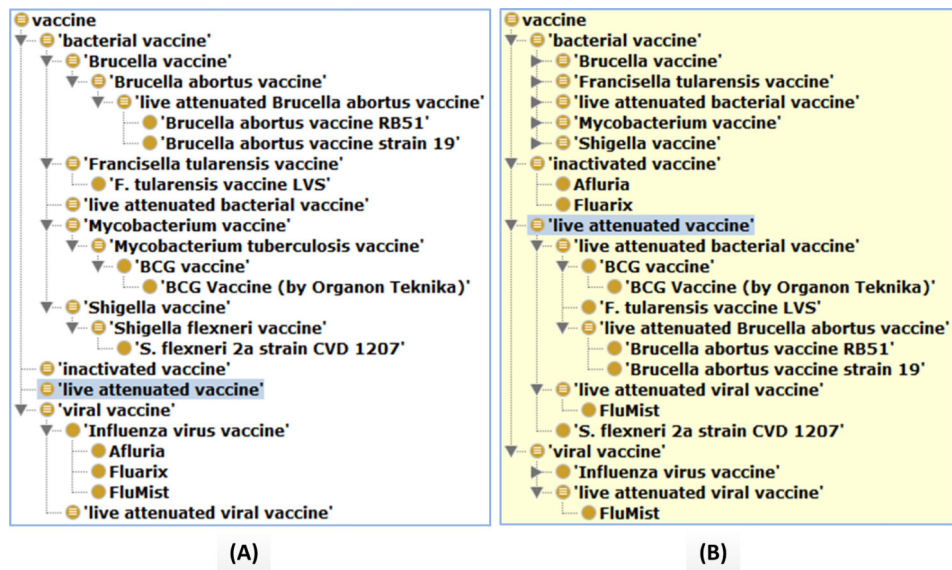
Figure 3. Ontology VO supports asserted and inferred hierarchies

(A) Asserted VO hierarchy of 8 bacterial and viral vaccines. Note that the term 'live attenuated vaccine' has no asserted child term. (B) Inferred hierarchy after using Fact++ description logic reasoner [70]. The results were extracted from the VO using OntoFox [51] and displayed using the Protégé ontology editor [121]. No specific vaccines were asserted under live attenuated or inactivated vaccines. However, after reasoning, 6 vaccines were inferred to be different types of live attenuated vaccines and 2 other vaccines were inferred as inactivated vaccines. VO: Vaccine Ontology. Data taken from VO.
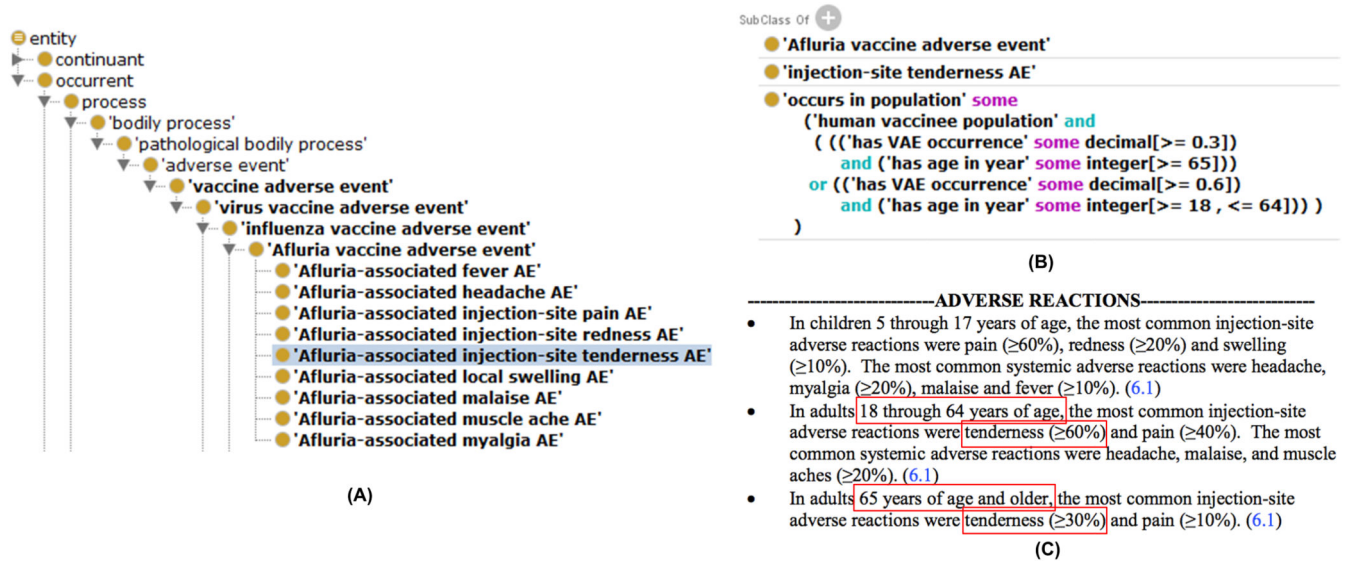
Figure 4. OVAE representing Afluria VAEs reported in FDA vaccine package insert
(A) The hierarchical structure of Afluria VAEs represented in OVAE. (B) OVAE ontological representation of 'Afluria-associated injection-site tenderness AE' based on two age groups. (C) Afluria adverse reactions recorded in the FDA package insert document. The (A) and (B) subfigures were screenshots of OVAE using the Protégé OWL editor. The text from (C) was retrieved from the FDA package insert document of the Afluria vaccine. The red color box highlighted terms in (C) represents different age groups and percentage rate (*i.e.*, occurrence) of vaccinees who are likely have to the tenderness adverse event (AE). The occurrence rates of tenderness AE under two different age groups are ontologically represented in OVAE (B). OVAE also represents other Afluria-associated AEs using the same strategy (A). AE: Adverse Event; OWL: Web Ontology Language; OVAE: Ontology of Vaccine AEs.
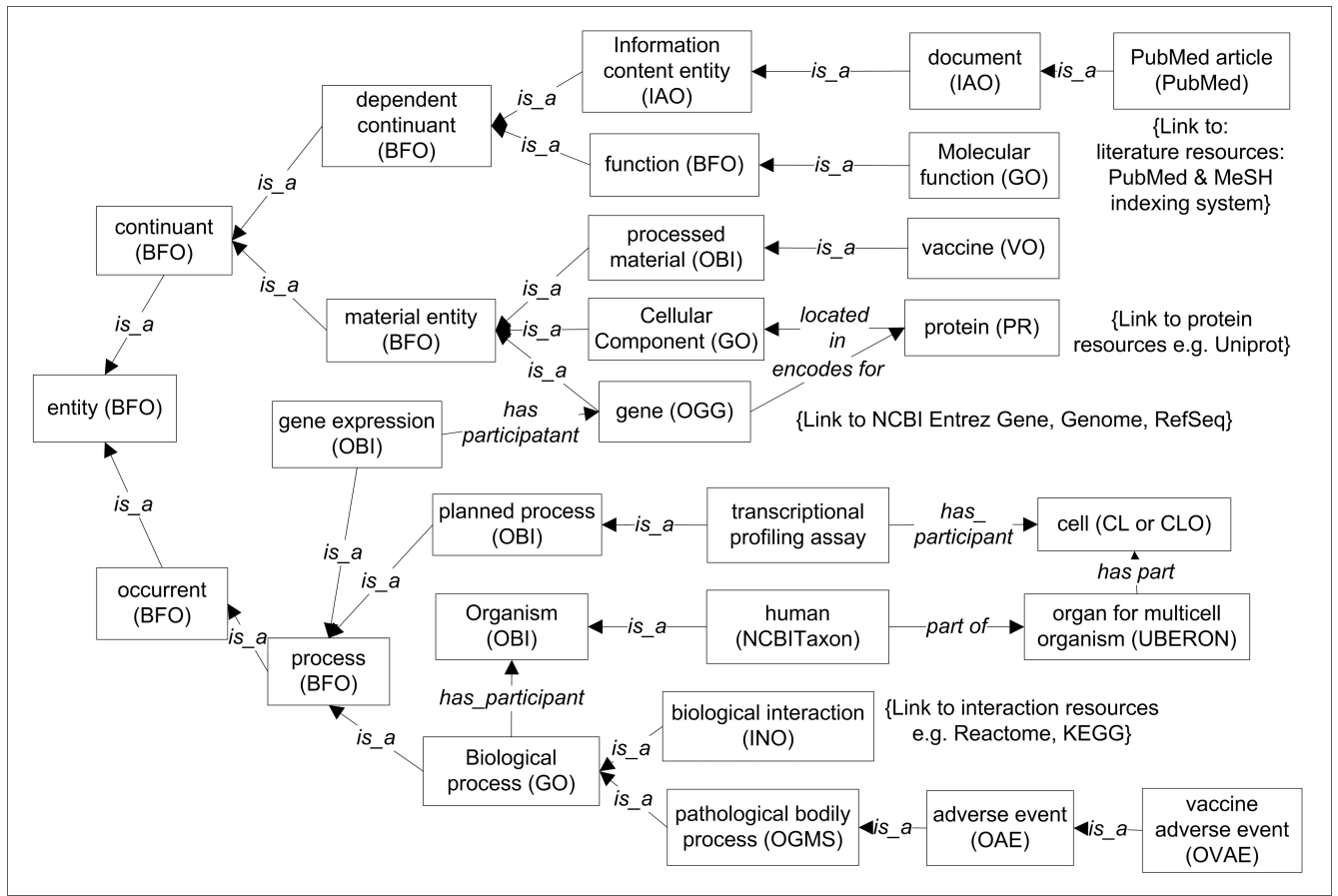
Figure 5. The top level ontology framework of linked open vaccine data (LODV) to support vaccine data integration

This framework represents seamless integration of different data types represented by different ontologies (*e.g.*, VO, OAE, and OVAE). All related ontologies are aligned with the Basic Formal Ontology (BFO). Vaccine-related experimental data can be represented as instances of VO, OBI, and other ontology classes. A new Ontology of Genes and Genomes (OGG) [115] can be used to represent individual genes and genomes from different organisms. A new Interaction Network Ontology (INO) and specific INO extensions (*e.g.*, human INO or HINO) to represent interaction data stored in databases such as Reactome [107]. HINO: Human Interaction Network Ontology; INO: Interaction Network Ontology; LOD: Linked Open Data; LODV: LOD-vaccine; OAE: Ontology of Adverse Events; OBI: Ontology for Biomedical Investigations; OGG: Ontology of Genes and Genomes; PR: Protein Ontology; VO: Vaccine Ontology; OVAE: Ontology of Vaccine Adverse Events.
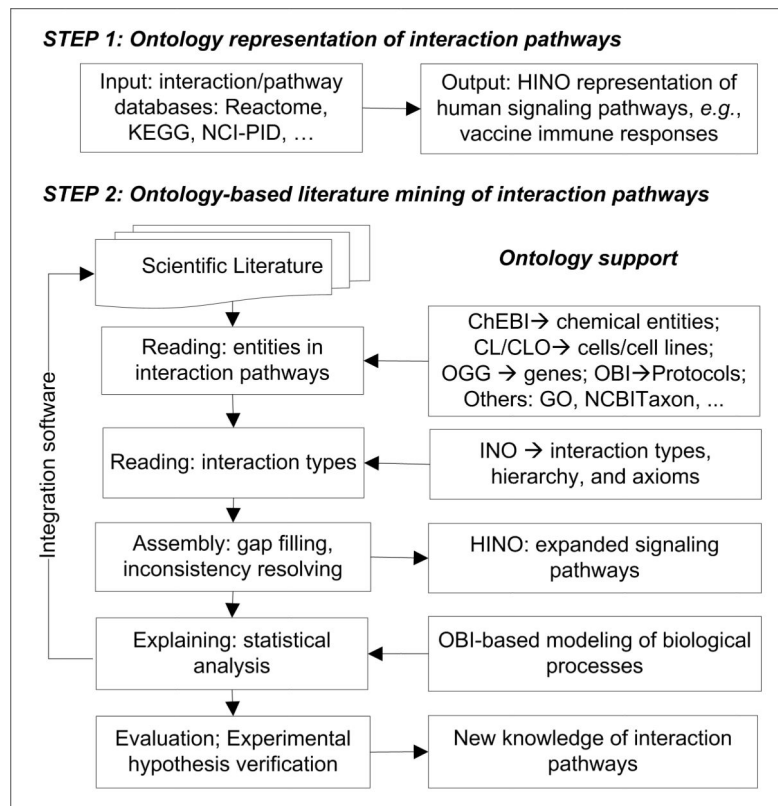
STEP 1: Ontology representation of interaction pathways

Input: interaction/pathway databases: Reactome, KEGG, NCI-PID, … → Output: HINO representation of human signaling pathways, *e.g.*, vaccine immune responses

STEP 2: Ontology-based literature mining of interaction pathways

Integration software

Scientific Literature → Reading: entities in interaction pathways → Reading: interaction types → Assembly: gap filling, inconsistency resolving → Explaining: statistical analysis → Evaluation; Experimental hypothesis verification

Ontology support

ChEBI→ chemical entities; CL/CLO→ cells/cell lines; OGG → genes; OBI→Protocols; Others: GO, NCBITaxon, ...

INO → interaction types, hierarchy, and axioms

HINO: expanded signaling pathways

OBI-based modeling of biological processes

New knowledge of interaction pathways

Figure 6. Proposal of ontology-based pathway representation and literature mining of gene interaction networks
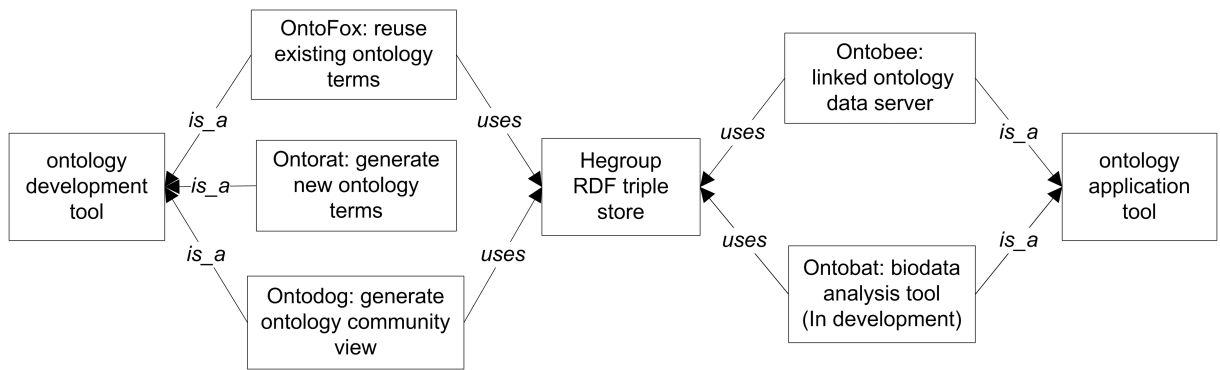
See text for detailed explanation.

Figure 7. Tools available to develop, query, and apply ontologies

OntoFox is a web-based tool that supports ontology term reuse and sharing by automatically fetching ontology axioms from existing ontologies [51]. To simultaneously generate and annotate a large number of new ontology terms based on specified design patterns, Ontorat was generated [118]. The tool Ontodog generates ontology community views from a large ontology [119]. The Hegroup RDF triple store has been used as the default triple store that includes all OBO Foundry ontologies [66]. Ontobee is a linked data server that supports ontology term, dereferencing, browsing and querying. Ontobee currently supports over 130 ontologies with over two million ontology terms [66]. Ontobee has become the default linked ontology data server for publishing and browsing biomedical ontologies in the OBO foundry library. Ontobat, a tool still under development, will specifically be designed for ontology-based LOD data analysis. LOD: Linked open data; OBO: Open Biomedical Ontology; RDF: Resource Description Framework.