Ontology-based representation and analysis of conditional vaccine immune responses using Omics data

Anthony Huffman[1], Edison Ong[12], Tim Brunson[3], Nasim Sanati[3], Jie Zheng[4], Anna Maria Masci[5], Guanming Wu[3], Yongqun He[6-8]

[1] *Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, United States.*
[2] *GlaxoSmithKline, Rixensart 1330, Belgium.*
[3] *Oregon Health & Science University, Portland, Oregon, OR, United States.*
[4] *Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States*
[5] *Office of Data Science National Institute of Environmental Health Science, Durham, NC, United States.*
[6] *Unit of Laboratory Animal Medicine, University of Michigan, Ann Arbor, MI, United States.*
[7] *Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, United States.*
[8] *Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, United States.*

Abstract

ImmPort, the world's largest repository of immunology data, includes many vaccine immune response datasets. ImmPort maps the metadata of these studies to ontology and database schema. As of February 28, 2023, our ImmPort data analysis identified 6.258 immune exposures using 47 vaccines in 4,607 human subjects, and 324 cohort studies from the ImmPort. We hypothesized that an integrative ontological representation of the data from these studies would enhance our understanding and analysis of these ImmPort vaccine studies, and with ontological classification and tools such as VIGET, we could further study the effects of different conditions such as vaccine types and host biological sex on the vaccine response gene expression profiles. Our Vaccine Ontology (VO) analysis classified these 37 vaccines into bacterial, viral, and protozoan vaccine types with different vaccine properties. The ImmPort metadata types were modeled with the Vaccine Investigation Ontology (VIO). Our new ontology-based pipeline extracted vaccine response data from the ImmPort database, annotated them based on ontology, obtained corresponding gene expression data from the GEO, and performed consistent omics data analysis. Our use case found gene profiles shared and differed from live and killed inactivated influenza vaccines. Furthermore, our Omics data analysis using the VIGET tool found that female and male human subjects have differential host responses for influenza vaccines. For example, our study showed much stronger early female responses to influenza vaccination than males, and males was able to show active immune responses at a later stage. Interestingly, the female (but not male) human subject group also showed significantly enriched neutrophil degranulation at Day 3 after influenza vaccination; however, males (but not females) displayed significantly enriched neutrophil degranulation at Day 14 after influenza vaccination. These mechanisms have been used to find differences between the gene lists and pathways of host responses to different vaccines conditional to different factors including vaccine types and host biological sex. Moreover, this framework can be expanded to other vaccines and vaccine categories easily.

Keywords

Vaccine Ontology, ImmPort, GEO, Vaccine immune response, Gene expression profiles

Proceedings of the International Conference on Biomedical Ontologies 2023, August 28th-September 1st, 2023, Brasilia, Brazil
EMAIL: huffmaar@umich.edu (A. 1); edong@umich.edu (A. 2); tjbrunson@gmail.com (A. 3), nasim@plenary.org(A.4), jiezhen@pennmedicine.upenn.edu(A.5), mascia2@nih.gov(A.6), wug@ohsu.edu(A.7) yongqunh@med.umich.edu (A.8)
ORCID: 0000-0002-5735-7540 (A. 1); 0000-0002-5159-414X (A. 2); 0000-0001-6681-0418 (A.4) 0000-0002-2999-0103 (A.5), 0000-0001-9189-9661 (A.6), 0000-0001-8196-1177 (A.7), 0000-0001-9189-9661 (A.8)

1. Introduction

As one of the greatest inventions in modern medicine, vaccination has been used to dramatically protect humans against infectious diseases and improve human health. However, infectious diseases are still a major cause of human mortality throughout the world, and effective vaccines that protect against many diseases still do not exist [1,2]. The future success of effective vaccine development relies on deep understanding of the protective immune mechanisms [3].

Vaccine induced host responses depend on various conditions and factors. For example, we have previously developed the VaximmutorDB, a web-based database system that has included over 1,700 experimentally identified vaccine immune effectors (abbreviated as "vaximmutors") induced by 154 vaccines for 46 pathogens [4]. The VaximmutorDB data have been manually annotated from peer-reviewed publications and reliable databases. Our VaximmutorDB data analysis showed that these vaccines induce many common immune factors, for example, Th1 immune factors IFN-gamma and IL-2 and Th2 immune factors IL-4 and IL-6. Responses induced by different vaccine types such as influenza and yellow fever vaccines may also differ [4]. However, more specific mechanisms underlying vaximmutors and vaccine immune responses are still largely unclear, even when we provided a successful use case for this yellow fever and influenzas vaccines for identifying these vaximmutors [4]. Numerous other studies also identified many other factors crucial for vaccine response induction [5-7]. For example, biological sex and age may change the immune responses significantly [7, 8].

ImmPort (the Immunology Database and Analysis Portal; https://ImmPort.niaid.nih.gov/) is the world's largest repository of public-domain de-identified clinical trial data related to immunology [9]. All data derived from clinical trials funded by the Division of Allergy, Immunology and Transplantation (DAIT) of the National Institute of Allergy and Infectious Diseases (NIAID) are required to be published on the ImmPort portal. In addition, the ImmPort portal includes data obtained from the work of the Human Immunology Project Consortium (HIPC, http://www.immuneprofiling.org/) as well as relevant data from several external sources such as the Gates Foundation. ImmPort includes complete clinical and mechanistic study data (e.g., mechanistic assays used, timing of visits, etc.), all of which are publicly available for download in a de-identified form.

Pathway/network based analyses are extremely important in understanding protective immune responses to infectious diseases. The immune systems of two vaccinees with different genetic backgrounds may have different gene expression profiles in response to the same vaccine; however, there may be shared underlying core immunological pathways between these two vaccinees. Various experimental conditions may also affect host immune responses to vaccines. Nevertheless, appropriate tools to support the analysis and visualization of vaccine-induced immune pathways under different given conditions still miss.

We hypothesize that different vaccine types induce differential but coherent host responses, and experimental conditions would significantly change the results of vaccine-induced host responses. More specifically, we wanted to see if by leveraging ontology, that we could integrate multiple vaccine studies to find common patterns for vaccines with the shared experiments. We have previously studied how different factors would affect the host responses to Yellow fever vaccines [10], influenza vaccines [11], and Brucella vaccines [12]. However, the previous studies were still limited with small datasets and deep investigations. In current study, we would investigate further how specific factors such as biological sex and vaccine types would change the vaccine-induced host responses.

Biomedical ontologies have emerged to be critical for the standardization, integration, and analysis of the large amounts of heterogeneous biological data. We previously demonstrated how ontologies, including the Vaccine Ontology (VO) [13], Vaccine Investigation Ontology (VIO) [14], and Ontology of Biological and Clinical Statistics (OBCS) [3, 15] could be used to support vaccine induced host response studies. Vaccine Ontology is a reference ontology focused on vaccines in general while Vaccine Investigation Ontology is an extension that is focused on metadata types in various vaccine investigation studies. Such approaches may be used to study the ImmPort data. Given the ImmPort

database with its own data management system, it would be important to integrate our ontology-based approaches with the ImmPort data system for better studies.

VIGET is our newly developed web-based Vaccine response Gene Expression analysis Tool based on Reactome and ImmPort [16]. VIGET uses the VO to classify various vaccine types and experimental conditions. VIGET allows users to select vaccines with VO classification, choose ImmPort studies and confounding variables, and perform differential gene expression analysis of various vaccine responses. It is possible to apply VIGET to investigate the vaccine response gender differences by comparing patterns for gender-based differences across influenza vaccines. We want to leverage our ontology driven design to identify further patterns contained in multiple vaccine studies.

In this report, we demonstrated how ontology and VIGET can be applied to systematically standardize, classify, integrate, and analyze vaccine-related high-throughput ImmPort and GEO data, and identify vaccine-induced pathways under specific experimental conditions including vaccine types and biological sexes of the human subjects. Three specific use cases were also developed to further illustrate the effectiveness of our ontology-based representation and analysis of vaccine-induced host responses given different vaccine or host conditions.

2. Methods

2.1. ImmPort vaccine metadata extraction and storage

The immune exposure metadata description of various vaccine studies was downloaded from the public ImmPort website (https://www.immport.org/) to a local computer. Such metadata description is stored in different ways in ImmPort. ImmPort provides the annotations to the gene expression data that were deposited in GEO. The raw data can be downloaded from GEO. Our study downloaded the csv text file of such metadata, and converted it to Excel for easy exploration and processing.

The ImmPort database has the immune exposure table that stores the information on how subjects were immune exposed. Figure 1 shows how the vaccine immune exposure data in ImmPort was standardized using the Vaccine Ontology (VO). From the data in ImmPort, we added additional columns to create the metadata file via inter-rater discussion.

	A	B	C	D	E	F
1	EXPOSURE_ ACCESSION	ARM_ ACCESSION	EXPOSURE_ MATERIAL_ID	EXPOSURE _MATERIAL _REPORTED	EXPOSURE _PROCESS _REPORTED	SUBJECT _ACCESSION
2	IM48058	ARM5757	VO_0000867	Influvac	vaccination	SUB233686
3	IM48059	ARM5757	VO_0000867	Influvac	vaccination	SUB233687
4	IM48060	ARM5757	VO_0000867	Influvac	vaccination	SUB233688
5	IM48061	ARM5757	VO_0000867	Influvac	vaccination	SUB233689
6	IM48062	ARM5757	VO_0000867	Influvac	vaccination	SUB233690
7	IM48063	ARM5757	VO_0000867	Influvac	vaccination	SUB233691
8	IM48064	ARM5757	VO_0000867	Influvac	vaccination	SUB233692
9	IM48065	ARM5757	VO_0000867	Influvac	vaccination	SUB233693
10	IM48066	ARM5757	VO_0000867	Influvac	vaccination	SUB233694
11	IM48067	ARM5757	VO_0000867	Influvac	vaccination	SUB233695
12	IM22967	ARM5032	VO_0000644	HBV Vaccine	vaccination	SUB209984
13	IM22969	ARM5033	VO_0000644	HBV Vaccine	vaccination	SUB209985
14	IM22971	ARM5032	VO_0000644	HBV Vaccine	vaccination	SUB209986
15	IM22973	ARM5028	VO_0000644	HBV Vaccine	vaccination	SUB209987

VO IDs

Vaccine names

Figure 1: ImmPort immune exposure metadata illustration. The immune_exposure.txt file was downloaded on March 28, 2023, from ImmPort website, opened in Excel format, and the vaccine related information extracted. Note that it is only a portion of the contents in the file, and some columns of the file are not shown here to save space. We added The Vaccine Ontology (VO) is used here for vaccine information standardization.

2.2. Vaccine response data analysis using ontology-annotated ImmPort and GEO

We used Ontofox [17] to generate a hierarchy of vaccine terms from VO that mapped to vaccine terms studied in different experiments from ImmPort. We used the Vaccine Investigation Ontology (VIO) to map and model metadata types for the vaccine experiments. Protégé-OWL editor was used for ontology display and editing. SPARQL and DL-Query were used for ontology knowledge query.

2.3. Vaccine response data analysis using ontology-annotated ImmPort and GEO

ImmPort provided vaccine information and the metadata for different vaccine response studies. The vaccine types and metadata were annotated using VO and VIO. Based on the metadata of different studies, we extracted the normalized microarray gene expression data from the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo/). Software programs and APIs were generated to automatically query the final integrative ontology and ImmPort data. As a use-case, influenza vaccine studies were selected through ontology-based queries along with ontology-defined vaccine and condition (e.g., health and age) classification. For GEO data analysis, log2 transformation was applied if the expression values of a given GEO dataset were not in log space. For each comparison pair, expression data collected from individuals vaccinated on Day 0 was considered as the control group, while data collected on Day 3 to Day 7 after vaccination were considered as the vaccine exposure group. In addition, each pair was defined to have the same ontology-defined vaccine and condition within the same cohort in the same ImmPort study.

2.4. Omics data analysis using VIGET

The web-based Reactome pathway analysis tool (https://reactome.org/) was used to support our gene expression profile analyses. Adjusted FDR p-value cutoff of 0.05 was used for statistical data analysis [18].

3. Results

3.1. Vaccine response data extraction from ImmPort

The ImmPort website maintains a publicly available data model schema (https://immport.org/shared/dataModel) and a corresponding relational database. ImmPort does not store raw Omics gene expression data. However, ImmPort provides the annotations to the gene expression data that were deposited in GEO. The raw data can be downloaded from GEO.

An important table in ImmPort database is the immune exposure table that stores the information on how subjects were immune exposed. Figure 1 shows how the vaccine immune exposure data in ImmPort was standardized using the Vaccine Ontology (VO). Basically, the metadata file has many columns including Exposure Accession, ARM Accession, Exposure Material ID (which uses VO IDs for vaccines), Exposure Merial Reported (which is the label of the material), Exposure Process Reported (which is “vaccination” for vaccine immune exposure studies), and Subject Accession. With the VO IDs used in ImmPort metadata file, it became easy and efficient for us to standardize and process the vaccine data and categorization.

Based on our analysis, as of February 28, 2023, the ImmPort database included 36,140 immune records. Our analysis of the database identified 6,258 vaccine-related immune exposure records that used 37 vaccine terms with unique VO IDs in 4,607 human subjects, 324 ARMs (a single group or cohort for a specific study purpose).

3.2. VO-based classification of the ImmPort-reported vaccine studies

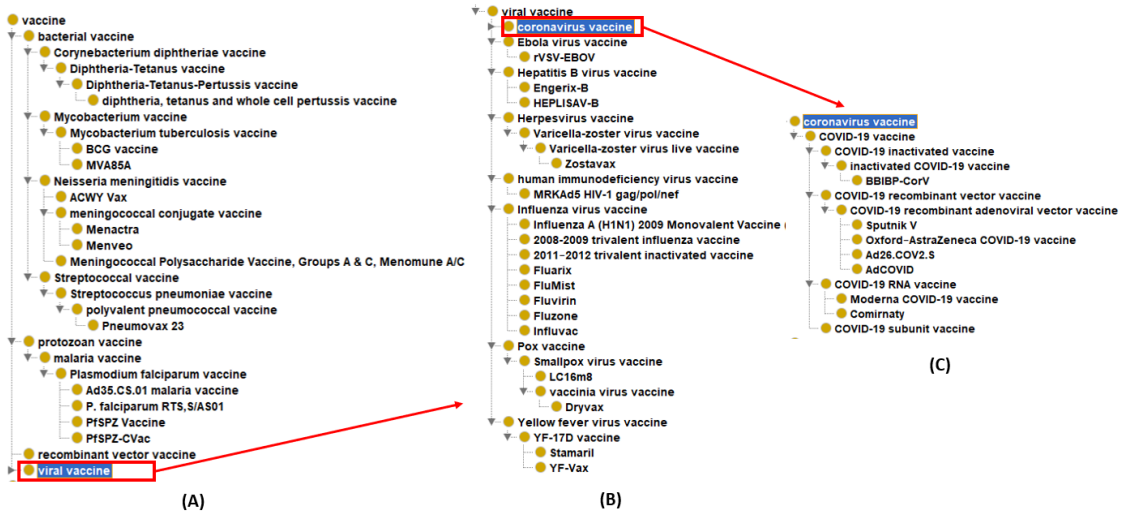

Figure 2: Ontological representation of 37 vaccines in ImmPort and their relations using VO. Vaccines in VO are asserted by the pathogen they vaccinate against. Column A shows that ImmPort contains a small selection of bacterial vaccines (7) and parasite vaccines (4). 3 of the 37 vaccines were mapped to 2008-2009 influenza vaccine. The red boxes represent parent terms shown later in the hierarchy.

3.3. VIO-based of the vaccine investigation conditions

Figure 3 illustrates the basic flow to extract GEO expression data accession in the ImmPort database schema by following six primary accessions. The flow chart represents the annotations under each GEO GSM ID. Each study has one or more ARM (or cohort). Each ARM has one or more subjects. Each subject has one or more biosample. Each biosample has one or more experimental sample (i.e., expsample). Meanwhile, each subject also has an immune exposure such as vaccination, and each vaccination exposure points to a VO ID. Finally, each experimental sample corresponds to a specific GSM ID. Each table also has more detailed annotations in the ImmPort database. Eventually, the GSM ID can be used to get the data from the GEO database (Figure 3).

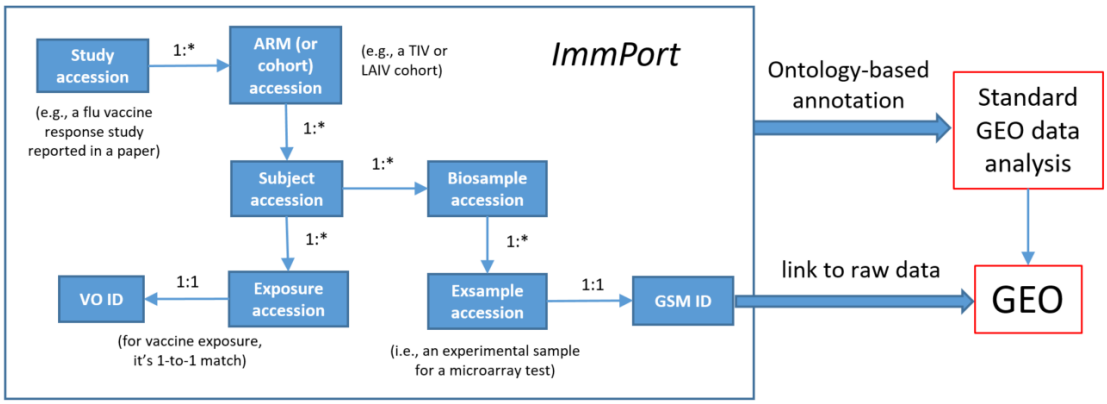

Figure 3: Vaccine response Omics data analysis using ontology-annotated ImmPort and GEO. Six IDs in the boxes represent the primary keys of corresponding tables of the ImmPort database. These six

primary accessions are IDs for the study (study accession), cohort of a study (ARM accession), subject of a cohort (subject accession), vaccine used on a subject (exposure accesion), biological test used on a subject (biosample accession), the biological assay utilized on the subject (biosample accession), and the experimental sample that was part of the biological assay (exsample accesion). The schema illustrated here explains the components in Figure 1 and the relations among them, and links the ImmPort data to GEO data.

3.4. Use case 1: Ontology-based query of vaccines and vaccine investigation data

Based on the logically defined hierarchies and semantic relations in ontology, we can use our ontology system to perform computer-assisted queries and analysis.

For example, based on the VO ontology hierarchy as shown in Figure 2, we can easily identify which vaccines are bacterial vaccines, and which are viral vaccines. We can also detect which are influenza vaccines or yellow fever vaccines. VO classification clearly lays out the relations of these vaccines. Since the VO is computer-interpretable, the results can be automatically parsed by a semantic query such as SPARQL or DL-query for various follow-up analyses or to aid retrieval of experimental studies as part of use case 2.

Not shown in Figure 2, VO includes axioms that illustrate different properties of vaccines. For example, the FluMist vaccine is defined to have live attenuated feature as seen in the following axiom:

'has quality' some 'vaccine organism live attenuated'

In contrast, the Fluarix vaccine is defined to be an inactivated whole organism vaccine based on the following axiom:

'has quality' some 'vaccine organism inactivated'

Such logical axioms defined in VO provide us a logical way to automatically identify different features. For example, we can use SPARQL or DL-query to easily find which vaccines are live attenuated vaccines and which are inactivated vaccines. This was done to retrieve experimental studies for further analysis of the vaccines within ImmPort.

Note that the ImmPort database does not provide the semantic information described above. We extracted the VO IDs annotated in ImmPort and generated a VO subset that contains these VO IDs and their corresponding vaccines and related vaccine attributes. This subset of VO was then applied to support our semantic queries and analysis.

3.5. Use case 2: Detecting the effect of biological sex on gene expression profiles stimulated by influenza vaccines

After we identify which vaccines or vaccine groups to evaluate, we can come back to query the ImmPort database to obtain detailed vaccine study information (Figure 2). For example, using the ontology strategy, we identified many vaccines grouped as inactivated influenza vaccines as shown in Figure 2. Using ImmPort database, we identified three studies involving Fluzone and Fluarix, two inactivated influenza vaccines which will be linked to GEO data as part of use case 2. It is known that males and females may have different responses to vaccination. To see if using VIGET can recapitulate this vaccine response gender difference, we compared patterns for gender-based differences across influenza vaccines. We expanded our analysis to include non-immune related pathways related to cell transcription or cell signaling. For influenza vaccines, we utilized all 16 influenza studies covering one attenuated influenza vaccine, FluMist (VO_000044), and five inactivated influenza vaccine, Fluarix (VO_000044), Fluzone (VO_000047), Fluvirin (VO_000046), the 2008-2009 trivalent inactivated vaccine (VO_0004808) and the 2011-2012 trivalent inactivated vaccine (VO_0004810). These include the same vaccines as part of use case 2. Our VIGET study included 210 male and 260 female subject samples collected from either whole blood or peripheral blood mononuclear cells (PBMCs) [16]. These

subjects ranged from 0 years old to 90 years old and included the same ethnicities as the previous use case. Finally, due to reduced sample sizes after 14 days, our analysis is focused on days 3, 7, and 14 in comparison to day 0. All results were adjusted by age, race, and vaccine type. Using the VIGET tool, we analyzed the effect on pathways using differences in log10-fold expression greater than 0.1, 0.2, 0.5, 1.0. These pathways can be found as part of Supplemental Table 1.

Our study showed that female vaccine responses tended to exhibit greater fold change than males earlier to the time response. Using the default 0.2 pathway reveals that males exhibited less significant pathways in females at all 3 time points (Table 1). Day 7 had the highest amount of significant genes and pathways for both males and females. For males, significant pathways only emerged at log-fold changes of 0.5 or greater; with the 1 log-fold expression showing neutrophil degranulation (FDR = 6.48 e-7) and innate immune system (FDR = 1.02e-2) as the only significant pathways. Females, in contrast, had 10 pathways at the same time point (Table 1), including multiple immune response pathways, including neutrophil degranulation (FDR = 5.06 e-11), IL-4 and IL-13 signaling (FDR = 2.31e-5), IL-10 signaling (FDR = 4.01e-2), and the CLEC7A/inflammasome pathway (FDR = 4.77e-2). Additionally, females at day 7 also exhibited cellular response to stress (FDR = 3.19e-4), with all genes in this pathway being up-expressed (Supplemental Table 1, Flu-F-07-1.0-Pos.) Otherwise, there is a consistent pattern of males having fewer significant pathways, and immune related pathways than females for influenza vaccines.

Table 1. Summary of gender differences in number significant genes and pathways in influenza vaccine response. Differential gene response for influenza vaccines across males and females. The initial timepoint for comparison was Day 0 for all entries. All genes that had the magnitude of their log fold change greater than the threshold were counted. For # of significant pathways, the numbers indicate the number of pathways with FDR < 0.05. The full list of pathways can be found as part of the Supplemental Table 1. Day 14 Males had no pathways related to immune response. All log-fold changes are base 10. Results were corrected for age, race, batch, platform.

Time Comparison	Log-Fold Change Threshold	# of Significant Flu M Genes	# of Significant Flu M Pathways	# of Significant Flu F Genes	# of Significant Flu F Pathways
Day 3	0.1	176	17	969	73
	0.2	3	35	56	126
	0.3	0	0	9	126
	0.5	0	0	0	0
Day 7	0.1	1388	0	1427	0
	0.2	1329	0	1052	3
	0.3	1133	0	1048	0
	0.5	1119	0	690	8
	1.0	626	2	515	10
Day 14	0.1	446	39	969	73
	0.2	44	36	98	70

	0.3	4	68	11	34
	0.5	0	0	0	0

Figure 4 shows the Reactome pathway enrichment analysis results at three days, seven days, and fourteen days post-vaccination for males and females. Overall, females showed a significantly earlier immune responses at Day 3 post influenza vaccination than males, and then later males caught up with active immune responses. As shown in this figure, females exhibit an earlier vaccine response as shown in pathways that are linked to signaling of interleukins 4, 13 and 10. An interesting finding is the difference in neutrophil degranulation pathway expression in influenza-vaccinated female and male groups. At Day 3 neutrophil degranulation was significantly enriched in the influenza-vaccinated female group but not in the male group. At Day 7, both female and male groups showed significant enrichment of neutrophil degranulation. However, in Day 14, only the male (but not female) group showed significant enrichment of neutrophil degranulation (Figure 4).

(A) Flu vaccines, Females, Day 3 over Day 0

(B) Flu vaccines, Males, Day 3 over Day 0

(C) Flu vaccines, Females, Day 7 over Day 0

(D) Flu vaccines, Males, Day 7 over Day 0

(E) Flu vaccines, Females, Day 14 over Day 0

(F) Flu vaccines, Males, Day 14 over Day 0

Figure 4: Comparison of sex-based differences in immune response to influenza vaccines. All Reactome subfigures were compared to Day 0 of vaccine administration. The conditions for subfigures were labeled with text in the figure. Subfigures were generated by Reactome's Reacfoam tool (Release 82, September 2022). The detailed information is provided in Supplemental Table 1.

We also investigated for sex differences that are common to live attenuated and inactivated influenza vaccines. Due to a small data set for female inactivated influenza vaccine users, analysis for Day 0 to Day 3 females could not be done using VIGET (Supplemental Table 2). However, inactivated influenza

vaccines revealed patterns of increased ribosomal translation as being enriched at Day 3 with fewer immune pathways for males and females. Day 7 for the influenza vaccines failed to find pathways that were found significant in males. Day 14 revealed 44 enriched immune related pathways. Males had uniquely enriched pathways linked to the general immune system and interferon alpha/beta signaling (Supplemental Table 2). Females, in contrast, had enriched pathways linked to innate immunity enriched but not the general immune system. Intriguingly, day 7 and day 14 both reveal that females have enriched pathways related to programmed cellular death and other cellular functions. Further investigation of yellow fever vaccines revealed a similar pattern of additional enrichment of cell death and cell signaling. Taken together, this shows that females have a common, unique response to the vaccines we have tested.

4. Discussion

4.1. Application of Ontologies to ImmPort

The contributions of this study are multiple. First, we demonstrate our application of ontologies and semantic relations to annotate the ImmPort data and link to GEO database. Second, we studied the effect of vaccine type as a vaccine factor on the vaccine immune response profiles. Third, we detected the effect of biological sex as an important host factor that affects the vaccine immune responses.

Our study shows that ontology can be applied with the relational ImmPort database to better support vaccine immune response data analysis. There were initial difficulties with finding an appropriate mapping between ImmPort and GEO due to the multiple lookup tables from each study to the list of GEO IDs. While this was aided by ImmPort documentation, it still took significant effort trawling through the database to find the matching IDs. There were also issues resolving if time matching definitions for day 0 meant that data was taken before or after vaccination. However, once done, we can now use this framework to expand to other vaccines or vaccine categories in ImmPort. As such, this approach satisfies and aids in data FAIRness (Findable, Accessible, Interoperable, and Reusable). While adaption of new axioms or ontology terms do require manual annotation, these can be easily done following eXtensible Ontology Design (XOD) principles [19] and reusing a suite of Ontozoo tools designed for this task [20, 21].Two database systems can be applied simultaneously. The ontology triple store can be used to store ontology knowledge and metadata, and the relational database can be used to store instance data. We can then use different query languages to link them together seamlessly. This can be adapted to other databases, albeit after looking to find the best mapping for data. The use of ontology can then be used to aid further intersections of different vaccination procedures (vaccine types, vaccine timing, or different vaccination routes) and patient phenotypes (sex or age-based differences between vaccine response). Ontology standardization has already been suggested to ImmPort to standardize different cell types [22]. However, this was focused on mapping terms to biological entity terms from Gene Ontology, Cell Ontology, Protein Ontology and the Ontology of Biomedical Investigation. The use of ontology standardization for vaccines on these datasets is novel outside of our lab.

It is to the best of our knowledge, that this is the first time that both databases were linked together to gain information on the gene expression profile differences of TIV and LAIV. As GEO and ImmPort have different focuses, this linkage will help facilitate greater understanding of the data. According to a PubMed search using “GEO ImmPort database” (8/16/2021), GEO and ImmPort, have been used as data sources to identifying genes related to cancer development and survival [23, 24].

4.2. The Effect of Biological Sex on Vaccine Immune Response for ImmPort vaccines

Our study illustrates the significant effect of biological sex in the vaccine immune response generation. The analysis of influenza vaccines showed females having an earlier and much stronger

activation of immune related pathways than males during the first week of vaccination. Moreover, female immune responses at Day 7 uniquely had all genes part of cellular response to stress be upregulated. It has been reported that females have a stronger early immune response than males to vaccines [25], with females more likely to experience adverse events caused from an autoimmune response, which may explain why these traits were only found when looking only at females. As this pattern was found across live attenuated and dead influenza vaccines, this may be the result of vaccines not being as optimized for human females and represents an avenue for further vaccine improvement and mechanism research. It is interesting to find the significant differences in neutrophil degranulation pathway expression in influenza-vaccinated female and male groups (Figure 6). The influenza-vaccinated females induced fast neutrophil degranulation at the early stage (Day 3) than males. At Day 7, both female and male groups showed similar enrichment of neutrophil degranulation. However, only males (but not females) had significant enrichment of neutrophil degranulation at Day 14 post vaccination (Figure 6). Neutrophil degranulation is the regulated exocytosis of secretory granules containing cellular mediators such as proteases and inflammatory mediators. Many recent studies infer various roles of neutrophil degranulation in inflammatory disorder (e.g. septic shock and asthma) [26], bacterial virulence strategy [27], and COVID-19 infection [28]. However, the role of neutrophil degranulation in vaccine immune response induction is still unclear and worth further investigation.

Still, there are other studies that validate our results. A prior article that also identified increases of ZAP-70 influenza immune response [29], suggesting that the role of ZAP-70 is duplicated in effective vaccines and is part of gaining vaccine immunity. Another study has investigated differential gene expression differences between older males and females from quadrivalent inactivated influenza virus vaccines [7]. Quadrivalent influenza vaccines can be incorporated into this framework to help understand how much of their differentially expressed genes are influenced by their vaccine type. Readers can validate our results by going to the VIGET website (https://viget.violinet.org/).

In the future, we plan to further examine how ontology combined with tools such as VIGET [16] can be used to further enhance our study of conditional vaccine immune responses given various experimental and clinical conditions. For example, by linking the VaximmutorDB data with the ImmPort data and Vaccine Ontology knowledge, we might be able to better understand the fundamental mechanisms of protective vaccine immune responses. Different ontology-based machine learning methods can also be explored to improve our prediction of new vaccine immune correlates and mechanisms.

5. Conclusions

In summary, the development of this ontology-based framework is able to aid data standardization and help guide further novel insights from the Omics data including the Omics metadata stored in the ImmPort database and Omics raw data from GEO. These mechanisms have been used to find differences between the pathways from different vaccines given different vaccine types and biological sexes. Moreover, this framework can also be expanded to other vaccines and vaccine categories.

6. Acknowledgements

This project was funded by a NIH-NIAID UH2 grant (1UH2AI132931-01A1) and a NIH-NIAID U24 grant (1U24AI171008). We appreciate the support provided by the ImmPort consortium. EO, YH and GW contributed to the overall study design; JZ, AMM and YH provided manual verification and ontology editing; EO, GW contributed to programming and data analysis; YH prepared the initial manuscript, AH prepared the updated manuscript for the paper and all authors contributed to the manuscript writing and reviews. All authors read and approved the final manuscript.

7. References

[1] **Standing up to infectious disease.** *Nat Microbiol.* 2019;**4**:1.

[2] Stutzer C, Richards SA, Ferreira M, Baron S, Maritz-Olivier C. **Metazoan Parasite Vaccines: Present Status and Future Prospects.** *Front Cell Infect Microbiol* 2018, **8**:67.

[3] Zhang C, Maruggi G, Shan H, Li J. **Advances in mRNA Vaccines for Infectious Diseases.** *Front Immunol* 2019, **10**:594.

[4] Berke K, Sun P, Ong E, Sanati N, Huffman A, Brunson T, Loney F, Ostrow J, Racz R, Zhao B *et al*: **VaximmutorDB: A Web-Based Vaccine Immune Factor Database and Its Application for Understanding Vaccine-Induced Immune Mechanisms**. *Front Immunol* 2021, **12**:639491.

[5] Beijnen EMS, Odumade OA, Haren SDV: **Molecular Determinants of the Early Life Immune Response to COVID-19 Infection and Immunization**. *Vaccines (Basel)* 2023, **11**(3).

[6] Zheng J, Li H, Liu Q, He Y: **The Ontology of Biological and Clinical Statistics (OBCS)-based statistical method standardization and meta-analysis of host responses to yellow fever vaccines**. *Quant Biol* 2017, **5**(4):291-301.

[7] Yang J, Huang X, Zhang J, Fan R, Zhao W, Han T, Duan K, Li X, Zeng P, Deng J *et al*: **Sex-specific differences of humoral immunity and transcriptome diversification in older adults vaccinated with inactivated quadrivalent influenza vaccines**. *Aging (Albany NY)* 2021, **13**(7):9801-9819.

[8] Hilleman MR, Carlson AJ, Jr., McLean AA, Vella PP, Weibel RE, Woodhour AF: **Streptococcus pneumoniae polysaccharide vaccine: age and dose responses, safety, persistence of antibody, revaccination, and simultaneous administration of pneumococcal and influenza vaccines**. *Rev Infect Dis* 1981, **3 Suppl**:S31-42.

[9] Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, Hu Z, Zalocusky KA, Shankar RD, Shen-Orr SS *et al*: **ImmPort, toward repurposing of open access immunological assay data for translational and clinical research**. *Scientific data* 2018, **5**:180015.

[10] Zheng J, Li H, Liu Q, He Y: **The Ontology of Biological and Clinical Statistics (OBCS)-based statistical method standardization and meta-analysis of host responses to yellow fever vaccines**. *Quantitative Biology* 2017, **5**(4):291-301.

[11] Zheng J, Harris MR, Masci AM, Lin Y, Hero A, Smith B, He Y: **OBCS: The Ontology of Biological and Clinical Statistics**. In: *The 2014 International Conference on Biomedical Ontologies (ICBO 2014): 2014; Houston, TX, USA*. 2014: 1-6.

[12] Todd TE, Tibi O, Lin Y, Sayers S, Bronner DN, Xiang Z, He Y: **Meta-analysis of variables affecting mouse protection efficacy of whole organism Brucella vaccines and vaccine candidates**. *BMC bioinformatics* 2013, **14 Suppl 6**:S3.

[13] Ozgur A, Xiang Z, Radev DR, He Y: **Mining of vaccine-associated IFN-gamma gene interaction networks using the Vaccine Ontology**. *Journal of biomedical semantics* 2011, **2 Suppl 2**:S8.

[14] Ong E, Sun P, Berke K, Zheng J, Wu G, He Y: **VIO: ontology classification and study of vaccine responses given various experimental and analytical conditions**. *BMC Bioinformatics* 2019, **20**(Suppl 21):704.

[15] Zheng J, Harris MR, Masci AM, Lin Y, Hero A, Smith B, He Y: **The Ontology of Biological and Clinical Statistics (OBCS) for standardized and reproducible statistical analysis**. *J Biomed Semantics* 2016, **7**(1):53.

[16] Brunson T, Sanati N, Huffman A, Masci AM, Zheng J, Cooke MF, Conley P, He Y, Wu G: **VIGET: A web portal for study of vaccine-induced host responses based on Reactome pathways and ImmPort data**. *Front Immunol* 2023, **14**:1141030.

[17] Xiang Z, Courtot M, Brinkman RR, Ruttenberg A, He Y: **OntoFox: web-based support for ontology reuse**. *BMC research notes* 2010, **3:175**:1-12.

[18] Storey, J. D., & Tibshirani, R. (2003). **Statistical significance for genomewide studies.** *Proc. Natl. Acad. Sci.* 2003, **100**(16), 9440–9445.

[19] He Y, Xiang Z, Zheng J, Lin Y, Overton JA, Ong E: **The eXtensible ontology development (XOD) principles and tool implementation to support ontology interoperability**. *J Biomed Semantics* 2018, **9**(1):3.

[20] Xiang Z, Courtot M, Brinkman RR, Ruttenberg A, He Y: **OntoFox: web-based support for ontology reuse**. *BMC Res Notes* 2010, **3**:175.

[21] Xiang Z, Zheng J, Lin Y, He Y: **Ontorat: automatic generation of new ontology terms, annotations, and axioms based on ontology design patterns**. *J Biomed Semantics* 2015, **6**:4.

[22] Overton JA, Vita R, Dunn P, Burel JG, Bukhari SAC, Cheung KH, Kleinstein SH, Diehl AD, Peters B: **Reporting and connecting cell type names and gating definitions through ontologies**. *BMC Bioinformatics* 2019, **20**(Suppl 5):182.

[23] Wu C, Hu Q, Ma D: **Development of an immune-related gene pairs signature for predicting clinical outcome in lung adenocarcinoma**. *Sci Rep* 2021, **11**(1):3611

[24] Guo L, Wu Q, Ma Z, Yuan M, Zhao S: **Identification of immune-related genes that predict prognosis and risk of bladder cancer: bioinformatics analysis of TCGA database**. *Aging (Albany NY)* 2021, **13**(15):19352-19374.

[25] Ciarambino T, Para O, Giordano M: **Immune system and COVID-19 by sex differences and age**. *Womens Health (Lond)* 2021, **17**:17455065211022262.

[26] Lacy P: **Mechanisms of degranulation in neutrophils**. *Allergy Asthma Clin Immunol* 2006, **2**(3):98-108.

[27] Eichelberger KR, Goldman WE: **Manipulating neutrophil degranulation as a bacterial virulence strategy**. *PLoS Pathog* 2020, **16**(12):e1009054.

[28] Petito E, Franco L, Falcinelli E, Guglielmini G, Conti C, Vaudo G, Paliani U, Becattini C, Mencacci A, Tondi F *et al*: **COVID-19 infection-associated platelet and neutrophil activation is blunted by previous anti-SARS-CoV-2 vaccination**. *Br J Haematol* 2023.

[29] Nguyen THO, Sant S, Bird NL, Grant EJ, Clemens EB, Koutsakos M, Valkenburg SA, Gras S, Lappas M, Jaworowski A *et al*: **Perturbed CD8(+) T cell immunity across universal influenza epitopes in the elderly**. *J Leukoc Biol* 2018, **103**(2):321-339.