**ImmPort: Promoting the Finding, Accessibility, Interoperability and Reuse of Immunological Data**

**OR**

**ImmPort: toward repurposing of publicly accessible immunological assay data for clinical research**

**I prefer either:**

**ImmPort: Promoting the Reuse of Immunological Data**

**Or:**

**ImmPort: Repurposing publicly accessible immunological assay data for clinical research**

**Or:**

**ImmPort: Repurposing immunological assay data for clinical research**

**ADD Author names and affiliations**

**Abstract** (100-170 words)

**Main Text** (excluding abstract, Methods, references and figure legends): 3,000 words

**Introduction**

ImmPort is one of the largest open repositories of human immunology data, created in support of the NIH commitment to promoting aggressive data sharing across the broader clinical and translational research community with funding from the National Institute of Allergy and Infectious Diseases Division of Allergy, Immunology and Transplantation (NIAID-DAIT)12. ImmPort contains data from both clinical studies on human subjects and immunology studies on model organisms. For the last thirteen years, ImmPort has been collecting and curating data from NIAID funded researchers. The data and accompanying software tools are made available through the ImmPort portal (http://www.immport.org/immport-open/public/home/home). To date, more than 250 studies have been shared through the repository, with a focus on allergy, vaccine response, autoimmune diseases, and transplantation. These results are made available to all researchers who accept the NIAID DAIT Data Sharing Use Agreement which seeks to balance the need for responsible stewardship of data with the goal of increasing access to data. [INSERT REF for agreement as well as paper that compared].

Recent advances in high-throughput technology coupled with a surge in data collection by biomedical researchers are posing formidable challenges to reproducibility, discoverability and secondary usage of the data collected. It is now widely recognized that to overcome these challenges, standardization of the data formats and terminologies used to describe clinical studies will be necessary [INSERT REF]. For reproducible research knowledge of data provenance plays a crucial role and one of the goals of ImmPort is to promote transparency and reproducibility by enabling access not only to the data but also to documentation of the methods used to generate the data.

All forms of secondary use of clinical data for purposes of meta-analysis will require strict data sharing guidelines resting on standardization of data formats and terminologies [INSERT REF]. ImmPort is in the vanguard of efforts to formulate and implement the necessary standards and guidelines and to demonstrate the potential of meta-analysis of assay data. The ImmPort database architecture is designed to support a wide spectrum of immunological data across multiple modalities, including mechanistic assay data, clinical lab measurements, interventions and concomitant medications, assessment data, adverse event reports, and description of the study methods used at all stages from basic bench biology to clinical trials.

**Results**

**ImmPort Applications for Data Collection, Sharing and Analysis**

The goal is to ensure that the widest possible spectrum of research data and associated findings are accessible to the broader research community in ways that allow effective knowledge and data sharing. To help in achieving this ImmPort provides an ecosystem with four components (Figure 1): *Shared Data*, *Private Data*, *Data Analysis* and *Resources*.

*Shared Data* provides a searchable catalog and distribution portal for publicly available datasets

*Private Data* focuses on the collection and curation of research data generated by contributors ranging in size from single investigators to multi-investigator projects in large consortia. EXPLAIN THAT THE PRIVATE DATE BECOMES PUBLIC DATA AFTER A CERTAIN TIME HAS PASSED, OTHERWISE THERE SEEMS TO BE NO REFERENCE HERE TO THE ‘PUBLICLY ACCESSIBLE DATA’ MENTIONED IN ONE VERSION OF THE TITLE

*ImmPort Galaxy* allows the application of open source analysis tools to both flow cytometry data and also data deriving from mass cytometry (also known as CyTOF or Cytometry by Time-of-Flight). It provides a graphical user interface for tools and pipelines that have a command line interface.

*Resources* includes the tools developed by ImmPort to support data standardization and analyses, for example …

**Collection, Curation and Sharing of Data**

Implementation of the NIH Data Sharing Policy [insert ref] is an ongoing process requiring the concerted efforts of many stakeholder groups to balance the interests of the NIH seeking to ensure access to federally funded data against those of the research community seeking to shepherd the resources they are granted to advance biomedical knowledge. Familiarly, researchers face a series of hurdles at the stage of data collection where sharing of data requires conformity to standard terminologies and formats used in public repositories. The ImmPort data collection, curation and sharing process is designed to address these hurdles for the immunology research community. It is the product of extensive interaction, prototyping, and refinement involving DAIT Program Officers, data providers, ImmPort staff and researchers who use the shared data. Figure 1 is a schematic representation of the ImmPort data flow.

In order to ease the process of data upload and to enhance the standardization of terms and vocabulary, the ImmPort team has developed a set of templates covering key elements of biomedical research for uploading of data (http://www.immport.org/immport-open/public/home/dataTemplates). These templates are informed by community standards, and where standards are lacking, ImmPort participates in upgrading them to meet the needs of the data providers it supports. Immunologists use a variety of non-standard terms when describing their experiments, and the ImmPort data upload pipeline has led to a number of significant improvements in the consistency with which, for example, viral strains represented in vaccine response assays, cell surface markers, cell populations, and lab test panels are described. The ImmPort templates and data upload business rules also serve as initial mechanism of ImmPort data quality assurance.

**Standard Terminology**

ImmPort data is annotated with terms from several ontologies including the Cell Ontology (CL, http://obofoundry.org/ontology/cl.html)3, Disease Ontology (DO, disease-ontology.org)

L. M. Schriml and E. Mitraka, The Disease Ontology: fostering interoperability between biological and clinical human disease-related data, Mammalian Genome, 26:9–10;584–589 (October 2015)

, Ontology for Biomedical Investigations (OBI, obi-ontology.org),

ADD REFERENCE

Anita Bandrowski, *et al*., The Ontology for Biomedical Investigations, PLoS ONE 11:4 (2016),

Protein Ontology (PRO, pir.georgetown.edu/pro/)4, and Vaccine Ontology (VO, www.violinet.org/vaccineontology/)5. MedRA (https://www.meddra.org) is used for adverse event terms and the NCI Thesaurus supplies terms from a variety of sources, including CDISC REF NEEDED.

The Antibody Ontology (AntiO) is a new ontology developed from data curated in ImmPort to provide standardized representation of monoclonal antibodies used in immunology research [INSERT REF, Manuscript in prep] and is an example of ongoing developments in data standardization facilitated by ImmPort. AntiO provides an antidote, in the immunological domain, to the problems created by the partial and often inconsistent information provided in published reports of the antibodies used in experiments. An analogous problem arises in the case of cytokines, where no public domain registry has thus far been available. To fill this gap ImmPort has compiled a registry of cytokines chemokines and receptors for the purpose of collecting, integrating and mapping between entity names and synonyms drawing on resources such as MeSH, the Protein Ontology, EntrezGene, HGNC, MGI, UniProt and others [http://www.immport.org/immport-open/public/reference/cytokineRegistry]. In addition, ImmPort engages with several data standards communities such as the Human Immune Phenotyping Consortium (HIPC) Standards Working Group [INSERT REF], BioSharing, the Patient Derived Tumor Xenograft Minimal Information (PDX-MI) working group [Insert ref to Can Res] and the NIH Big Data to Knowledge (BD2K) initiative (<https://metadatacenter.org/community/nih-bd2k)> through its collaboration with CEDAR (INSERT REF).

**Customized Engagements with Data Providers**

In order to facilitate the data submission process, the ImmPort curation team works closely with its data providers to provide explanations of the steps involved and to assist in the data annotation process to ensure that the most accurate description of the research data is obtained from the domain experts who generate it. Data is uploaded to private workspaces EXPLAIN RELATION OF THESE TO THE Private and Shared Data applications referred to above where the data is embargoed for review prior to sharing. In recognition of the broad spectrum of data management capabilities within its community of data providers, the data collection strategy is highly customized. Lessons learned from successive interactions are used to improve tutorials and other supporting materials for future use. For data are acquired before the relevant standards have been defined or adopted, ImmPort periodically updates previously uploaded data content with recommended standard terms and confirms the accuracy of the updates with the standards groups and data providers. This iterative approach to data collection and curation brings the benefits of encouraging data providers to describe their data with the highest quality annotations when they upload their data with a subsequent review of the content by data curators to ensure the shared data is maintained in a way that is maximally useful for reanalysis in the future.

In addition to annotating data with community recommended terms, the ImmPort data model also standardizes the formats used to capture experimental results. In this way we address a common lament of bioinformaticians that they need to spend a great deal of effort transforming data into a single consistent set of formats as a prerequisite to data analysis. The wide variety of assay methods employed by immunologists and clinicians has led over time to a wide variety of results reporting methods. ImmPort works with its data providers and standards collaborators to define the appropriate data formats and also the appropriate types of data needed for successful reuse. It is important to determine, for whether data reanalysis requires that the initial results from instruments are needed or only interpreted results or both.<--THIS IS I THINK A POOR EXAMPLE BECAUSE IT IS OBVIOUS (TO ME AT LEAST) THAT BOTH MIGHT BE NEEDED These considerations are generally not a primary concern of data providers, but can be fundamentally important in ensuring the usefulness of shared data.

**De-identification and Federated Data Repository**

One primary responsibility of ImmPort is to de-identify human study participant data using best practices and policies. HIPAA (Health Insurance Portability and Accountability Act of 1996) restricted data elements are intentionally not captured by ImmPort. For data deemed by NIH to be potentially sensitive to re-identification, ImmPort recommends that its data providers upload their data to the appropriate data repositoryNOT CLEAR WHAT THIS MEANS – MORE DETAIL NEEDED. This federated approach to data collection is adopted as a driving design principle in order to avoid the need to reinvent existing capabilities and to serve instead as a means to highlight and integrate data from potentially multiple sources. And again how does this relate to PRIVATE DATA APP, SHARED DATA APP …

Data Sharing practices

When curation (including de-identification?) (or make it clear that in this section you are only talking about de-identified data – and that all the data on ImmPort is de-identified, but some is in Private and some is in Shared – is completed, submitted data submitted resides on a private or shared ImmPort data space depending on the status of the project.

ImmPort has a rigorous policy regarding approval for data sharing built into the process. Detailed outputs of a study data are hidden from anyone except the original investigators until the latter have given their consent to share. Once the data sharing is approved by the investigators pre- or post publication the data gets disseminated through the Shared Data portal EARLIER YOU REFERRED TO SHARED DATA RESOURCES, TO THE SHARED DATA APP, BUT NEVER TO ‘SHARED DATA portal’ YOU NEED TO USE CONSISTENT TERMINOLOGY THROUGHOUT IS . As of May’17 ImmPort has shared 255 studies, 1,137 experiments with 318,943 biosamples on 45,167 subjects. There has been a positive trend for number of data downloads and number of registered users over the time. See Figure 2.

In last few years increasing attention is being paid to the need for improvements in current clinical data sharing practices, especially when it comes to under-reporting failed trials and the need for independent verification of findings on the basis of individual patient-level data. ImmPort is a designated data sharing portal for DAIT-contracted clinical trials, with a focus on clinical and mechanistic assay data related to asthma, allergy, autoimmunity., and transplantation. The ImmPort team works closely to prepare data uploading sets with Rho Inc, the DAIT clinical research organization, to ensure all data sets listed in case report forms, data dictionaries and study protocols are uploaded to ImmPort. These clinical trials are cross-referenced with ClinicalTrials.gov entries and compliment the ClinicalTrials records by providing subject level results. There are XX number of clinical trials and XX number of research studies shared through ImmPort. See Figure 2.

**Data Visualization**

Another aspect of communicating data users is powerful visualization which is the key to actionable insights into the complex data in the context of big data. Graphing features helps illustrating patterns, trends and data gaps which might not otherwise be apparent. ImmPort provides study demographics and experiment-centric interactive graphs to facilitate data exploration via a user-configurable display. In the Private Data application, these graphics improve curation efficiency for the ImmPort Team as well as data submitters (Figure 3). In the Shared Data application, they will provide information at-a-glance for study demographics or single experiments BUT NOT DEPLOYED THERE YET SO SHOULD WE EVEN MENTION IT?.

**Data Access Methods**

The research data in ImmPort is available for bulk download in Tab and MySQL format. APIs are also available for programmatic access to the ImmPort data. One can choose between downloading all metadata and database content or a user selected subset. Due to the inherent complexity of immunological studies, and as is true for repositories such as GEO (Gene Expression Omnibus), SRA (Sequence Read Archive) , TCGA ( The Cancer Genome Research Atlas), and GDC (Genomic Data Commons Data Portal), a shared data user gains the most benefit by learning the ImmPort data model and exploring the content of interest.

ImmPort collaborates with biomedical and healthcare Data Discovery Index Ecosystem (bioCADDIE: https://biocaddie.org/) to improve finding and access to shared data respectively. HIPC Immune space team fetch curated data from ImmPort to accelerate their development of interactive queries and analysis tools and for the HIPC Immune Signature identification project. Figure 1 shows schematic representation of the ImmPort Data Flow.

INSERT a short paragraph on API (?)

**Analysis Tools in ImmPort [Rename it as Resources]**

*ImmPort Galaxy* seeks to encourage the use of open source flow cytometry and CyTOF analysis tools. These tools often have a command line interface which may not be familiar to many bench scientists and may limit their adoption [Insert REF to cyto b paper]. *ImmPort Galaxy* provides a graphic interface (i.e. “point and click”) to a is a growing collection of data management open-source resource geared toward helping the scientific community to analyze flow cytometry data. With increasing number of flow analysis tools developed over the years [INSERT REF]. We have developed user-friendly interface on Galaxy platform6 that would allow immunologists with limited bioinformatics skills to rationally combine the available tools and run datasets through different workflows to achieve optimal results7. *ImmPort Galaxy* (immportgalaxy.org) follows the same high standards the Galaxy Project established in making tools and code available to the wider scientific community, and fostering an interactive developer’s community. The adoption of the Galaxy framework to host ImmPort analysis tools, and to a larger extent flow cytometry analysis tools, solved several user-led requirements such as ease of file upload, support for high-throughput analysis, and flexibility to integrate new tools easily (Figure 3).

The study data in ImmPort is structured using a relational data model and is available for bulk download in Tab and MySQL format using ASPERA tool for registered users. However, due to the complex nature of the data model accessing specific data for analysis or integrating data across studies becomes challenging for research scientists. This resulted into the development of *RImmPort* 8, an R package to streamline the accessibility and interoperability of ImmPort data for analysis in the R statistical environment. To aid in the secondary reuse of ImmPort data, *RImmPort* implements a data model that is based on the CDISC clinical trial data standards, and supports a suite of functions that provide access to different types of ImmPort study data.

Meta-analysis of data from multiple studies offers the benefits of increased statistical power and more robust estimation of a quantity of interest. ImmPort harbors increasing number of studies containing cytometry data that can be used in meta-analysis. The heterogeneity of cytometry data across the studies, caused by different antibody-fluorophore/isotope combinations and inconsistent cytometry configurations, makes joint analysis of these data difficult. We have developed a platform-agnostic, user-friendly, flow analysis framework called *MetaCyto* 9 that allows the integration of flow data from multiple sources and establishes a complete picture of all cell types associated to a phenotype in an objective and time-efficient manner as described above.

### To date, many principles of individual cell behavior and inter-cellular circuitry have been identified. To address the deluge of knowledge and to establish a foundation for systematic reasoning over the inter-cellular network of the immune system, we built *immuneXpresso,* a first comprehensive high-resolution knowledgebase of inter-cellular interactions, text-mined from PubMed abstracts. *immuneXpresso* identifies directional relations between more than 300 cell types and 140 signaling molecules across thousands of diseases. This global high-resolution interaction map has been already shown to enable systematic prediction of novel cell-type-specific interactions10. This tool is freely accessible ( http://www.immport-labs.org/immport-immunexpresso/public/immunexpresso/search).

### *Tutorials* Providing shared data is a way to promote FAIR principles. Another is to support and/or provide material and training opportunities to the scientific community so they can make use of the shared data appropriately. To this end, ImmPort has developed and currently hosts a few data programming and re-analysis tutorial notebooks as well as video tutorials to use ImmPort Galaxy, with plans to increase the collection and work in collaborations with educational organizations to promote re-use of shared data.

**Data Reuse, Reanalysis and Repurposing Resources**

In the field of Immunology we are only now beginning to explore the possibilities of reproducibility, reuse and repurposing of open datasets both for validation of the methods used in the original studies and also for the creation of new knowledge through meta-analysis or through virtual testing of hypotheses not foreseen by original authors. ImmPort team has successfully demonstrated the secondary analysis of ANCA-Associated Vasculitis (RAVE) trial dataset (SDY91) where we identified distinct subsets of granulocytes as novel early markers found at baseline in patients with anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis (AAV) that distinguish those achieving remission at 6 months following Rituximab or cyclophosphamide treatment from those for whom treatment failed in the Rituximab. This analysis led to novel insights to the study and discovery which may lead to more successful trials and therapeutic courses in AAV11.

Here, we describe three projects that we have embarked upon. To date, first, there is no representative reference data resource comparable to what we have in the realm of model organism genomes, for the multitude of immunological assays. The major barriers for understanding human immunological mechanisms include translating observations from model organisms to humans, accounting for the influence of confounding factors, and interpreting analytes--all of which are currently hampered by the lack of a benchmark reference human ‘immunome.’ Using all of the healthy controls from many clinical trials and research studies within ImmPort--which span a range of topics, from studies of allergy, autoimmune diseases, transplantation to immunotherapy--we have identified roughly 10,000 healthy individuals. This reference set will be the largest open dataset of healthy normal immune measurements to date. Over 13 types of measurements will be standardized and harmonized, and these will be made freely available to scientists and clinicians in a web environment for interactive visualization and download. This 10,000 Immunomes resource will enable rapid hypothesis generation and testing, provide a common control population to increase the robustness of human immunology studies, and serve as the foundation for studying immunity across age, sex, and ethnically-diverse populations (provide link to a draft webpage).

Second, the amount of publicly available immunological data, particularly flow cytometry and high-dimensional Mass cytometry, or CyTOF**,** a variation of flow where antibodies are labeled with heavy metal ion tags rather than fluorochromes, is beginning to explode. We believe that in order to understand the variability in the immune system across the human population, a greater understanding of distribution of well-characterized subsets of immune cells across the entire adult lifespan is critical. So, we take advantage of plethora of flow and mass cytometry data across human population in ImmPort. We believe such data will be more powerful if they can be integrated together. In addition, computational tools are now emerging to enable improved statistical and visual analysis of such data. Using MetaCyto, as described above, we performed a meta-analysis of cytometry data from 10 of the human immunology studies available on ImmPort.  The analysis identified multiple novel racial differences in circulating immune cells (INSERT CITATION Manuscript under review).

Third, is the repurposing of ImmPort data relating to living donors in solid organ transplantation. Living donors sacrifice their own health for the benefit of family, friends and strangers. Yet, the lifelong risks of such donation are not fully understood, with many living donors enter the transplant waitlist themselves after they donated. And here again, a principal challenge to living donor risk assessment is the lack of a unified resource resting on a with a large and representative dataset with a sufficiently wide range of clinical and non-clinical information in a form that allows for combination and re-analysis. ImmPort, with 27 clinical trial studies related to transplantation (ImmPort version DR19), provides a unique opportunity to create this resource and to embark on the needed analyses. Our initial curation work has already identified almost 10,000 living donors from just six ImmPort studies, with 21 further studies still to go. We plan to create a resource for transplant living donors from the clinical studies in ImmPort and then perform predictive analytics to assess and quantify the post-transplant risks of living donors of a range of different types on the basis of their clinical information obtained pre-transplant. Taken together, we believe that these three initial efforts at immunological meta-analysis using ImmPort will prove to be an invaluable resource to the burgeoning computational immunology community.

Immunity and Cancer

With the recognition of the importance of immune system involvement in cancer, gathering immunological data for both immune system intact and immunocompromised animal experiments are crucial to elucidate the deep involvement of cancer. ImmPort serves as the basis for the National Cancer Institute (NCI) *Oncology Model Forum* (OMF) and provides a powerful portal for collecting and correlating datasets from multiple studies in the context of cancer. The OMF ImmPort system will unite over 30 studies performed by NCI recent cancer researchers and their animal data for assessing the quality and fidelity to human cancer Syngeneic (genetically identical) Genetically Engineered Mouse Models (GEMMs) and Patient Derived Xenografts (PDXs) provide the basis for preclinical drug testing and investigation into the underlying mechanisms of cancer.

**DISCUSSION**

ImmPort is a data archival and dissemination portal developed for the purpose of promoting re-use of immunological research data and clinical trials generated by NIAID DAIT and DMID funded investigators. ImmPort provides comprehensive patient-level information from each study including metadata associated with each experiment, study protocols, clinical and mechanistic data describing the purpose of the study and detail methods of data generation. The functionality of ImmPort is expanded continuously to accommodate the needs of expanding research communities. ImmPort fosters open-science environment following the data standards and data sharing guidelines from the community. The shared research and clinical data, as well as the analytical tools developed in ImmPort ecosystem are available for public use.

In this article, we described ImmPort’s core functionalities of data submission, curation process, web interface for data sharing; and demonstrated the reuse of clinical trials data for secondary analyses to gain novel insight into early marker discoveries for successful treatment. We also described the projects where we combine multiple studies to repurpose and build reference datasets or resources which could benefit the community. The other big advantage of accessibility of data from large consortiums allows to take a holistic approach to the data instead of looking at individual dataset.

With recent implementation of ImmPort API, we anticipate to cross-reference with other databases seamlessly in future and further integrate the datasets across different disease areas. WRITING IN PROGRESS

**METHODS**

ImmPort Data Submission

ImmPort Galaxy

MetaCyto

The meta-analysis of cytometry data using MetaCyto follows four steps: data aggregation, data pre-processing, identification of common cell subsets across studies, and statistical analysis. First, data  is identified and aggregated from multiple studies. Second, the aggregated data are pre-processed by performing signal compensation (for flow cytometry data only) and asinh transformation. <-the jump to technical details here is ugly Third, the pre-processed cytometry data are clustered using FlowSOM [Van Gassen 2015] to identify cell subsets of interest because of what?. The cell subsets are then labeled on the basis on the which? marker levels. For example, a cluster that expresses high levels of CD4 and CD3 and a low level of CD8 will be labeled as “CD3+ CD4+ CD8-”. <--This does not answer the which? question Cell subsets that have the same label in different studies are considered to be cell populations of the same sort. Finally, for each cell population identified, the effect size of  a factor of interest, such as vaccination, on the cell population will be estimated using a linear regression model within each study. The overall effect size of the factor across studies is estimated using a random effect model.  MetaCyto is available on GitHub (github.com/hzc363/MetaCyto)

RImmPort

RImmPort is an R-driven interface to ImmPort data based on the CDISC standard, which SAY SOMETHING ABOUT WHAT CDISC PROVIDES, thus allowing the data to be used to address WHAT KINDS OF QUESTIONS.

RImmPort comprises of 1) a foundational data model (RImmPort) that encapsulates ImmPort study data in a format that can be loaded into R. 2) A set of data access methods to query and load different types of data from a specific study alongside methods to integrate assay-specific datasets from multiple studies.

Using RImmPort, an entire study can be loaded into R with a single command. For example, a researcher can run the RImmPort instruction *getStudy(‘SDY1’)* to load the dataset associated with ImmPort Study: *SDY1* into R. Internally, the functions query for specific study data from the ImmPort data source, organize the data into the RImmPort model, and then load the data into the R environment. The RImmPort package has been released in R/Bioconductor (bioconductor.org/packages/RImmPort ).

AntiO

AntiO provides metadata that links antibody products not only to the vendors and vendor catalog information but also to the clones that the antibodies contain, their protein targets, and their fluorescent conjugations. This information is provided in an ontology form that allows for advanced querying for antibodies and their targets, so that one can easily identify, for instance, all PE-labeled anti-CD25 antibodies their vendors and catalog numbers, and the ImmPort studies in which they were employed.

**Acknowledgements**

This work was supported by the National Institute of Allergy and Infectious Diseases (Bioinformatics Support Contract HHSN272201200028C) and by NIH/NIGMS R01GM080646 The authors thank Quan Chen, Gabriel Rosenfeld, Dawei Lin at the National Institute of Allergy and Infectious Diseases for their oversight of the program. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Data Citations (optional)- cite datasets or repositories**

**Competing interests statement**

The authors declare no competing financial interests.

CHANGE Natale reference (4) TO: Natale, D. A. et al. Protein Ontology (PRO): Enhancing and scaling up the representation of protein entities, Nucleic Acids Res 45 (D1), 2017, D339-D346, [doi.org/10.1093/nar/gkw1075](https://doi.org/10.1093/nar/gkw1075) (2017)

CHANGE Meehan (3) to

Diehl A. D., et al., The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability, J biomed semantics 7 (1), 44 [doi.org/10.1186/s13326-016-0088-7](https://doi.org/10.1186/s13326-016-0088-7" \t "_blank) (2016)

**REFERENCES**

1 Bhattacharya, S. *et al.* ImmPort: disseminating data to the public for the future of immunology. *Immunol Res* **58**, 234-239, doi:10.1007/s12026-014-8516-1 (2014).

2 bioRxiv 130948; doi: <https://doi.org/10.1101/130948>

3 ~~Meehan, T. F.~~ *~~et al.~~* ~~Logical development of the cell ontology.~~ *~~BMC Bioinformatics~~***~~12~~**~~, 6, doi:10.1186/1471-2105-12-6 (2011).~~

4 ~~Natale, D. A.~~ *~~et al.~~* ~~Protein Ontology: a controlled structured network of protein entities.~~ *~~Nucleic Acids Res~~***~~42~~**~~, D415-421, doi:10.1093/nar/gkt1173 (2014).~~

5 Lin, Y. & He, Y. Ontology representation and analysis of vaccine formulation and administration and their effects on vaccine immune responses. *J Biomed Semantics* **3**, 17, doi:10.1186/2041-1480-3-17 (2012).

6 Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**, W3-W10, doi:10.1093/nar/gkw343 (2016).

7 Kvistborg, P. *et al.* Thinking outside the gate: single-cell assessments in multiple dimensions. *Immunity* **42**, 591-592, doi:10.1016/j.immuni.2015.04.006 (2015).

8 Shankar, R. D. *et al.* RImmPort: an R/Bioconductor package that enables ready-for-analysis immunology research data. *Bioinformatics* **33**, 1101-1103, doi:10.1093/bioinformatics/btw719 (2017).

9 bioRxiv 130948; doi: <https://doi.org/10.1101/130948>

10 Rieckmann, J. C. *et al.* Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat Immunol* **18**, 583-593, doi:10.1038/ni.3693 (2017).

11 Nasrallah, M. *et al.* Reanalysis of the Rituximab in ANCA-Associated Vasculitis trial identifies granulocyte subsets as a novel early marker of successful treatment. *Arthritis Res Ther* **17**, 262, doi:10.1186/s13075-015-0778-z (2015).