



University of California
San Francisco



Unlocking the Power of FAIR Data Sharing with ImmPort

Sanchita Bhattacharya
Science Program Lead and Outreach Coordinator, ImmPort
Bioinformatics Project Leader
Bakar Computational Health Sciences Institute
University of California, San Francisco

dkNET Webinar Series, April 12, 2024

ImmPort Team



UCSF

Atul Butte, PI
Sanchita Bhattacharya
Reuben Sarwal

Immune System Sciences

Steven H. Kleinstein

NIAID

Anupama Gururaj
Quan Chen
Dawei Lin

ICF

Srinivas Chepuri
Karen Ketchum
Matthew Strub
Olivier Toujas-Bernate
Alicia Williamson

Funding Support

National Institute of Allergy and
Infectious Diseases (NIAID)
National Institutes of Health (NIH)
Health and Human Services (HHS)
Contract #: HHSN316201200036W

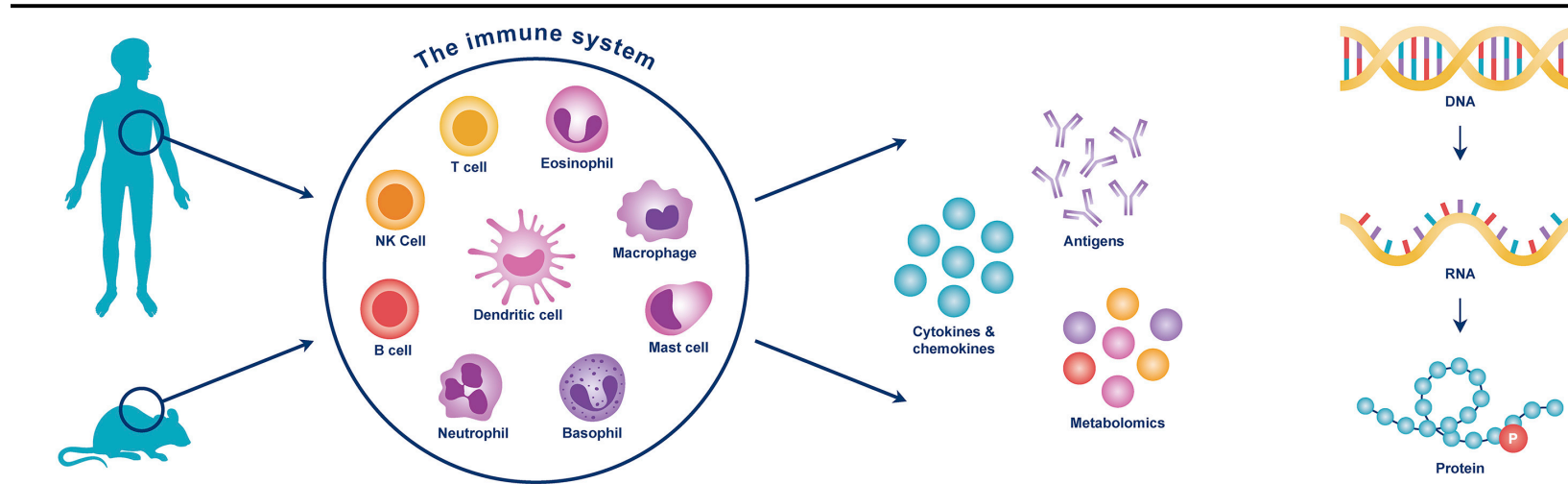
Peraton

Morgan Crafts
Emma Afferton
Sanjiv Desai
John Campbell
Zhiping Gu
Kate Hypes
Jaya Kannan
Ruth Monteiro
Elizabeth Thomson
Zullinel Trilla-Flores
Vilma Thomas
Sammi Smith
Bryan Walters
Shujia Zhou

Outline

- ImmPort – An Overview
- Secondary Data Reuse – Case Studies

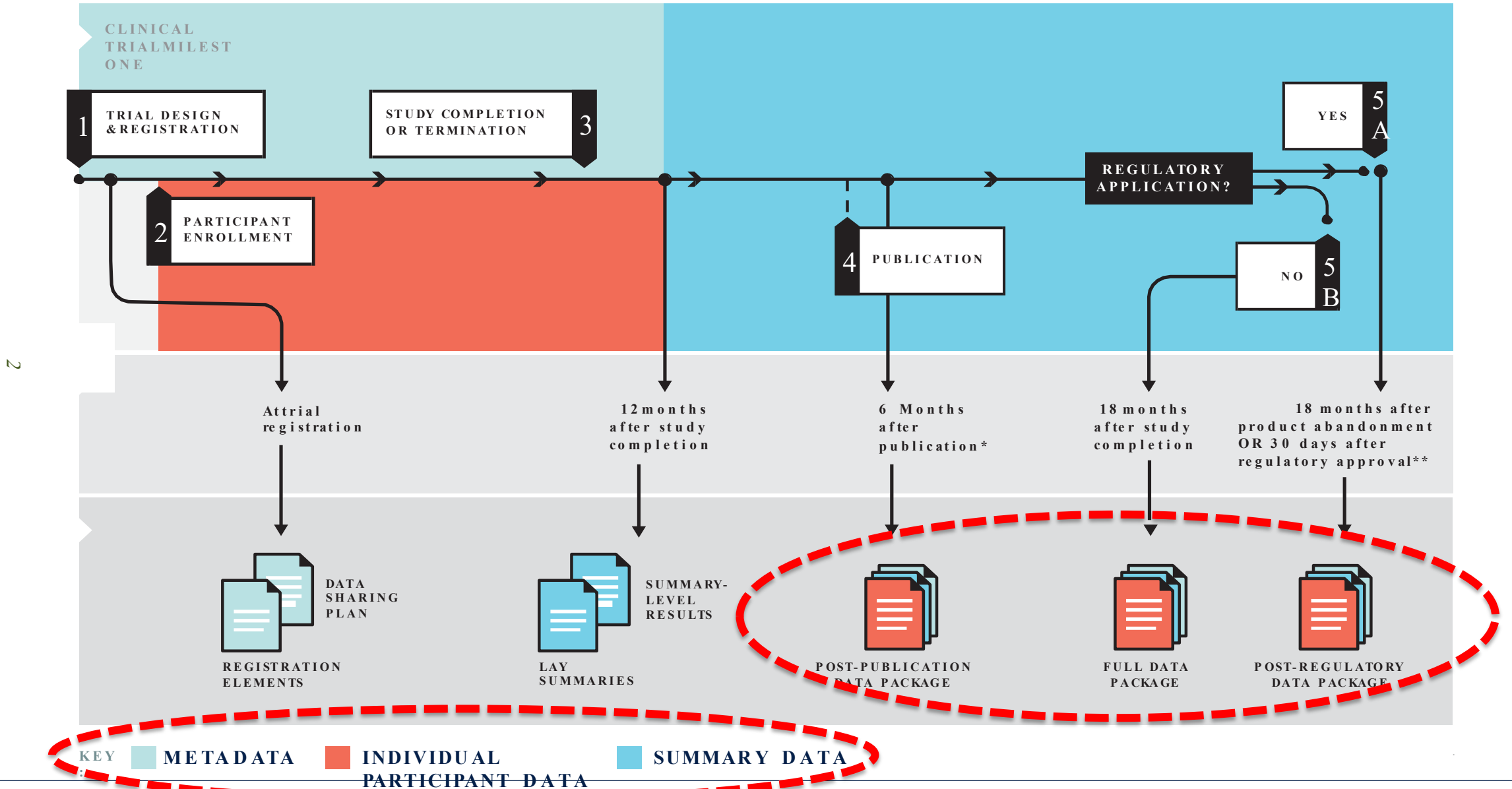
Molecular Portraits of Immune System



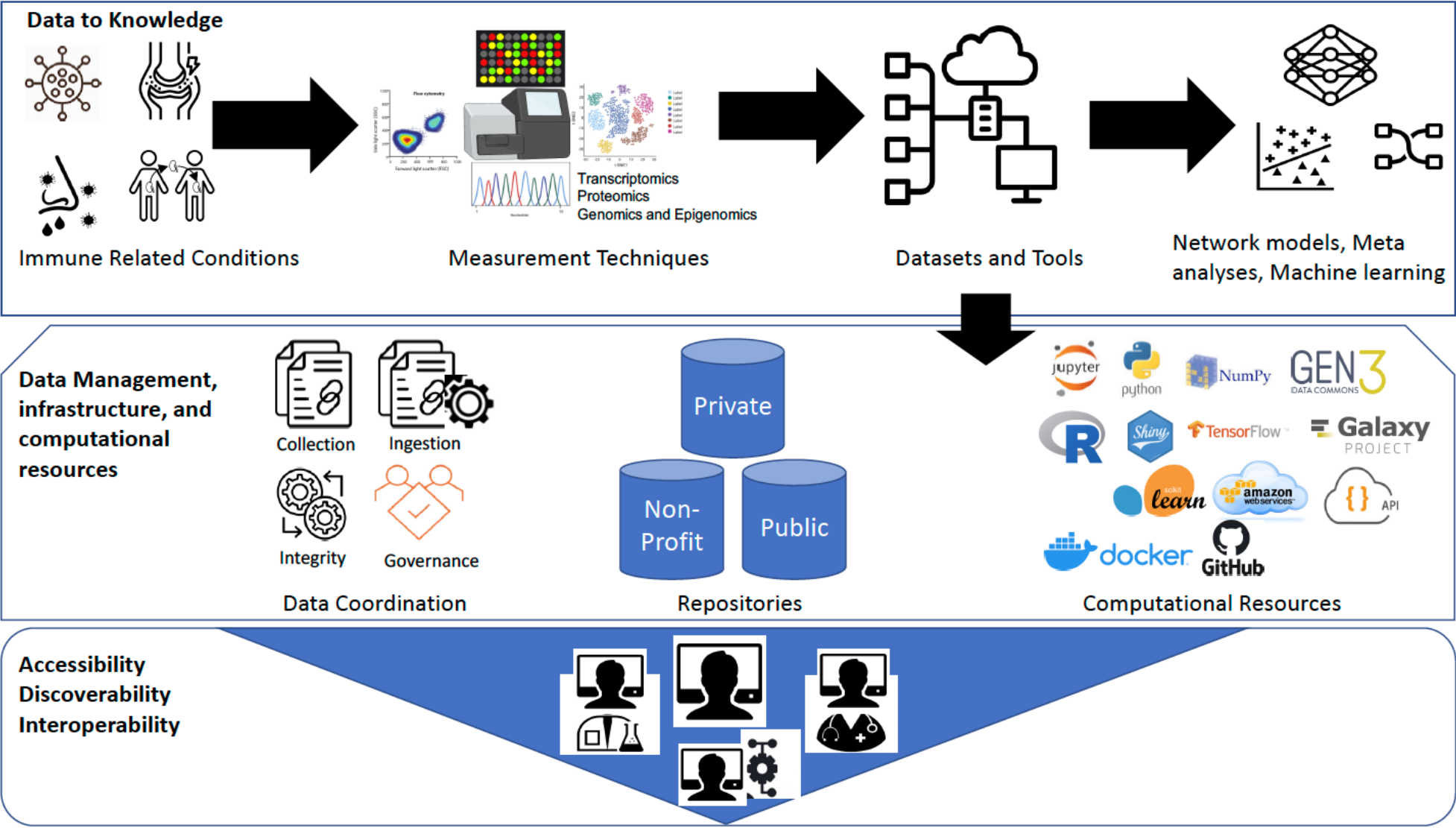
	Proteomics		Transcriptomics		Genomics & epigenomics			
Population level	Flow cytometry 	MS/PTM-MS 	Microarray 	RNA-seq 	DNA-seq 	ChIP-seq 	ATAC-seq 	BS-seq
Single-cell level	CytoF 	scMS 	smFISH 	scRNA-seq 	scDNA-seq 	scChIP-seq 	scATAC-seq 	scBS-seq

Yu et al., Current Opinion in Systems Biology, 2019

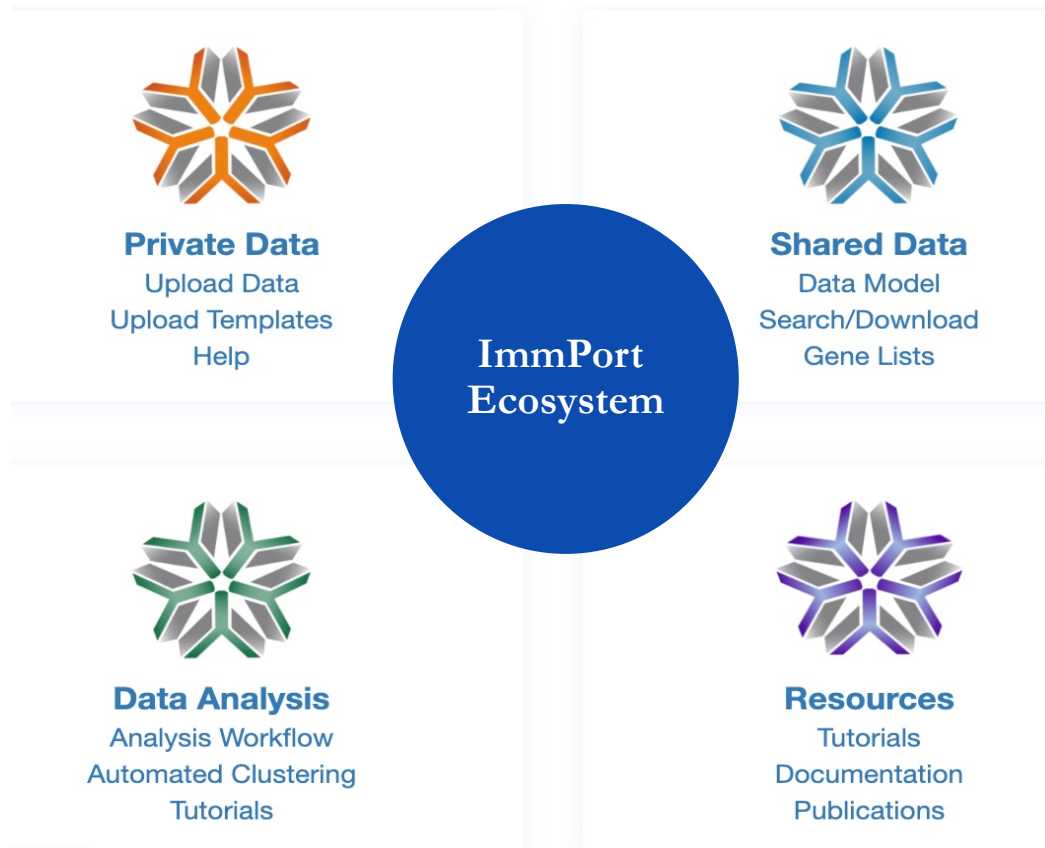
Clinical Trial Life Cycle: When to Share Data



Opportunities and Challenges in Democratizing Clinical Research Datasets



ImmPort data portal was developed to collect and share research and clinical trials data from NIAID/DAIT funded researchers



FAIR Principles

Make your data:

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

Findable

- Descriptive metadata
- Persistent

Accessible

- Determining what to share
- Participant consent and risk management
- Access status

Interoperable

- XML standards
- Data Documentation Initiative
- CDISC

Reusable

- Rights and licence models
- Permitted and non-permitted use

<http://datafairport.org/>

ImmPort Shares Data from Major NIAID-funded Programs and External Organizations



Human Immunology Project Consortium



COVID-19 - ImmPort

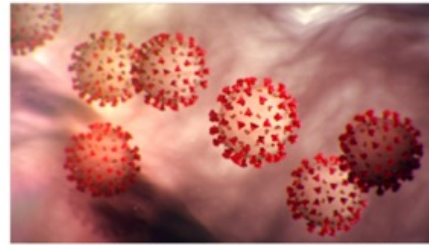


Image CDC Coronavirus Disease 2019 - COVID 19, Alissa Eckert

Coronavirus Disease 19

Powered by ImmPort



BILL & MELINDA GATES foundation

Powered by ImmPort

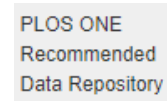
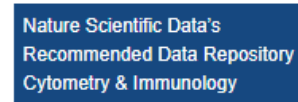


- Immunophenotyping Assessment in a COVID-19 Cohort (IMPACC)
- Serological Sciences Network (SeroNet)
- Multisystem Inflammatory Syndrome in Children (MIS-C)
- Impact of Initial Influenza Exposure on Immunity in Infants (U01)
- Atopic Dermatitis Research Network (ADRN)
- Population Genetics Analysis Program
- Protective Immunity for Special Populations
- HLA Region Genomics in Immune-mediated Diseases
- Modeling Immunity for Biodefense
- Reagent Development for Innate Immune Receptors
- Adjuvant Development Program
- Immunity in Neonates and Infants
- Asthma and Allergic Diseases Cooperative Research Centers
- HLA and KIR Region Genomics in Immune-Mediated Diseases
- Cooperative Study Group for Autoimmune Disease Prevention
- Immunobiology of Xenotransplantation
- Centers for Medical Countermeasures against Radiation Consortium
- Inner City Asthma Consortium
- Systems Approach to Immunity and Inflammation
- Innate Immune Receptors and Adjuvant Discovery Program
- Maintenance of Macaque Specific Pathogen-Free Breeding Colonies
- Non-human Primate Transplantation Tolerance Cooperative Study Group
- Consortium for Food Allergy Research
- Development of Sample Sparing Assays for Monitoring Immune Responses (U24)
- Asthma and Allergic Diseases Clinical Research Consortium (AACRC)
- The Clinical Islet Transplantation (CIT) Consortium
- Autoimmunity Centers of Excellence (ACE)
- Clinical Trials in Organ Transplantation (CTOC)
- Human Immunology Project Consortium (HIPC)
- Collaborative Influenza Vaccine Innovation Centers (CIVICS)
- Centers for Research in Emerging and Infectious Diseases (CREID)
- Cooperative Centers on Human Immunology
- Impact of Initial Influenza Exposure on Immunity in Infants (U01)
- A Multidisciplinary Approach to Study Vaccine-elicited Immunity and Efficacy Against Malaria (MVIE)

20 Years of FAIR Data Sharing



<http://doi.org/10.17616/R30J3F>

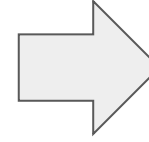
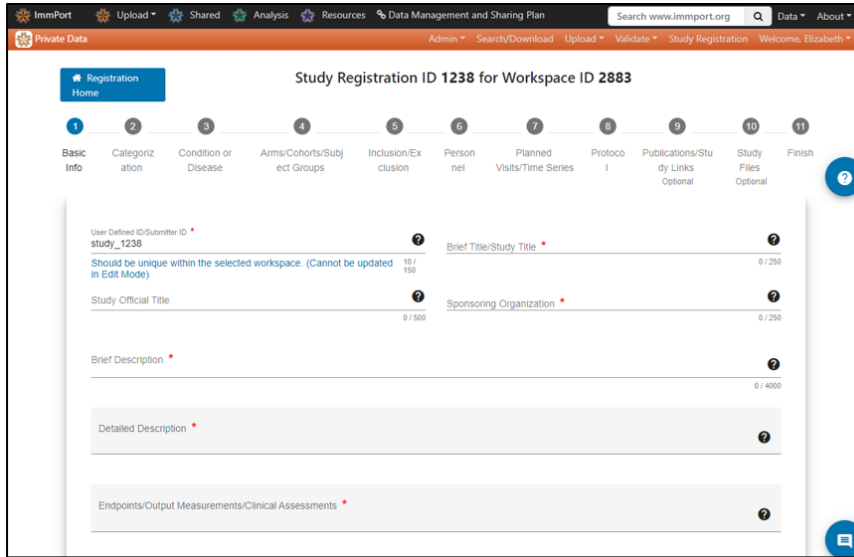
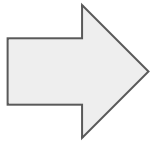


Data Submission Process Promotes FAIR Data

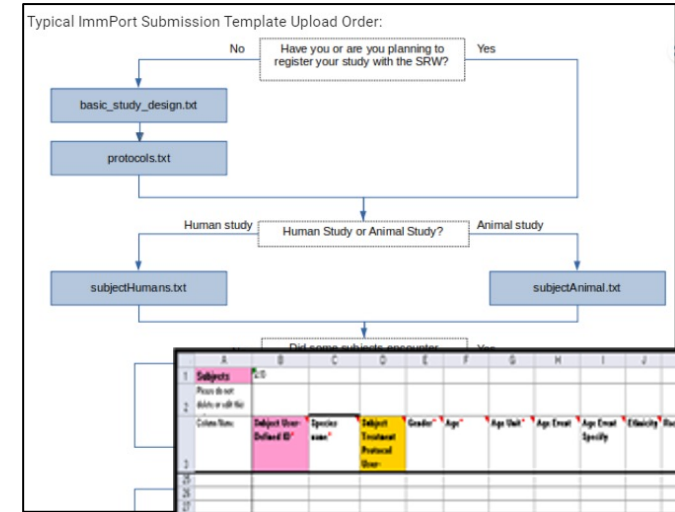
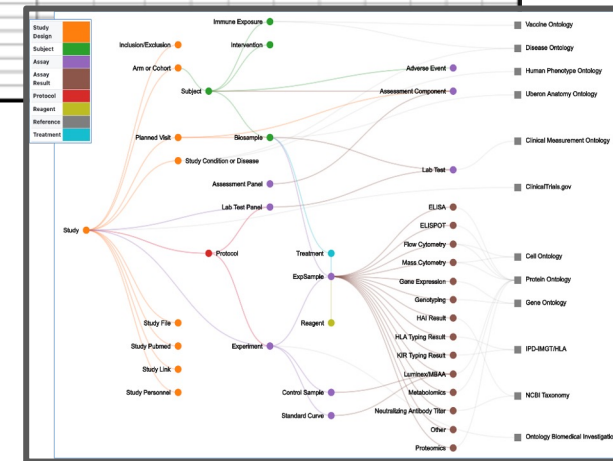
Major Steps in Data Submission for Data Submitters:



The Study Registration Wizard (SRW) kick-starts the data upload process and captures initial metadata associated with the study

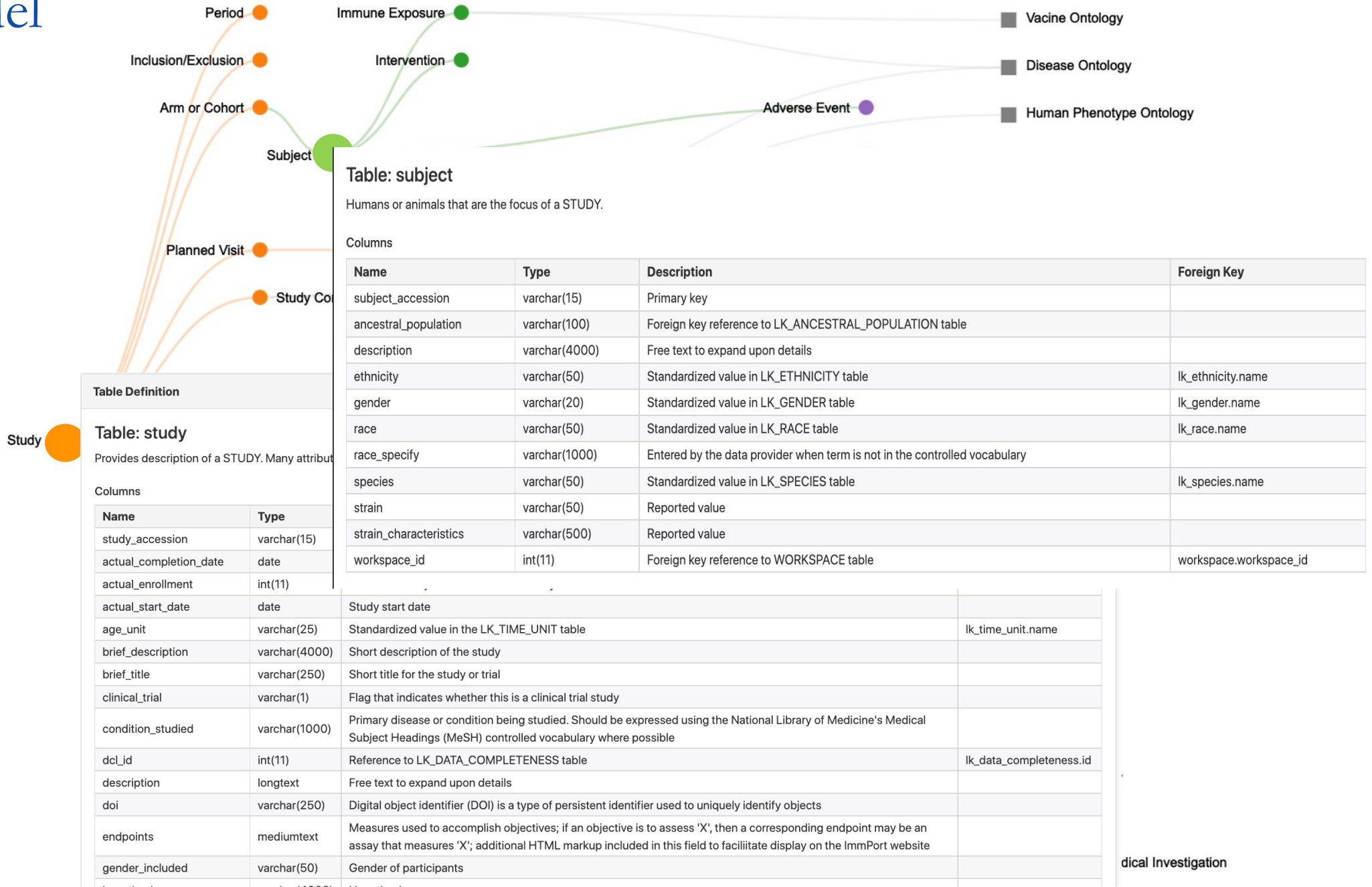


Data Submission Templates capture associated data and metadata based on study design

Submission templates incorporate controlled vocabulary terms from clinical and research ontologies.

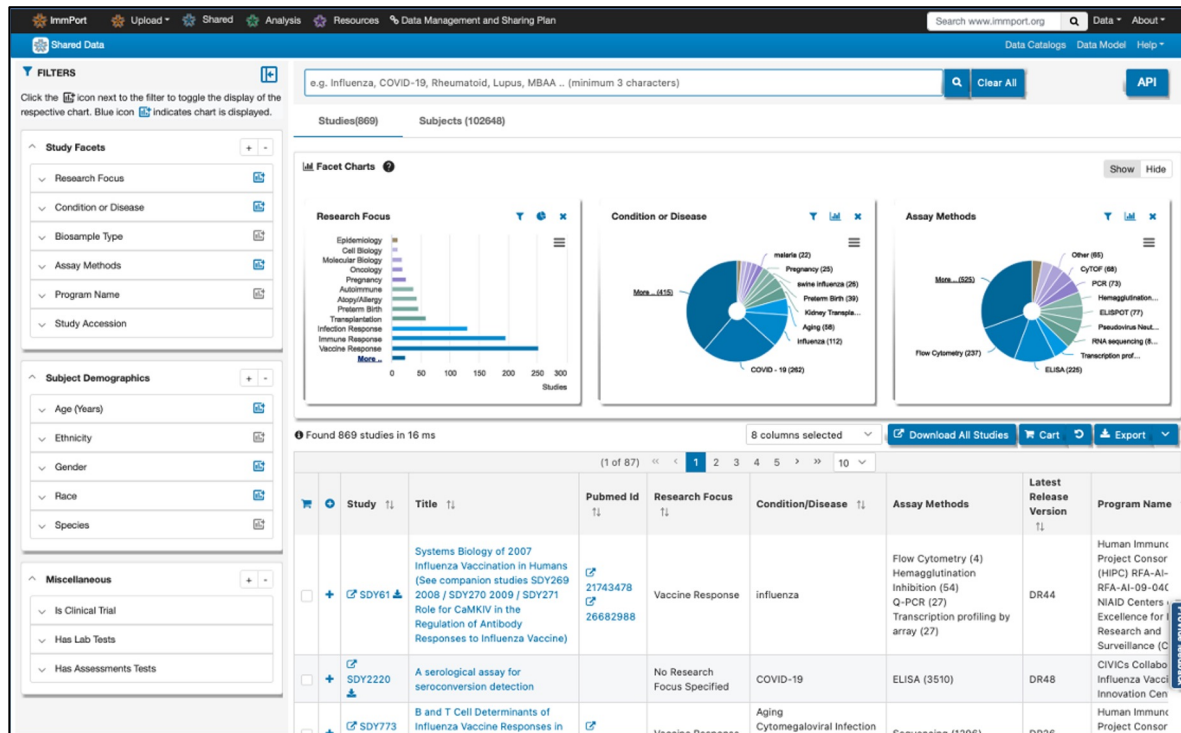
Data Model



Adherence to FAIR principles increases the visibility of your data!

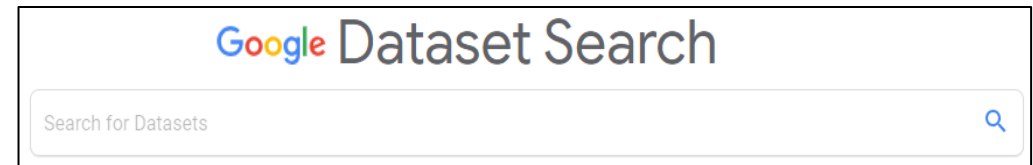
ImmPort Search – Cohort Discovery Tool (CDT)

Additional Repositories and Search Engines



The screenshot shows the ImmPort Search interface. On the left, there are filter panels for 'Study Facets' (Research Focus, Condition or Disease, Biosample Type, Assay Methods, Program Name, Study Accession) and 'Subject Demographics' (Age, Ethnicity, Gender, Race, Species). Below these are 'Miscellaneous' filters. The main area features a search bar with the text 'e.g. influenza, COVID-19, Rheumatoid, Lupus, MBAA .. (minimum 3 characters)'. Below the search bar are two facet charts: 'Research Focus' (a horizontal bar chart) and 'Condition or Disease' (a pie chart). To the right of these is another pie chart for 'Assay Methods'. Below the charts is a table of search results with columns: Study, Title, Pubmed Id, Research Focus, Condition/Disease, Assay Methods, Latest Release Version, and Program Name. The table shows three results, with the first one expanded to show details like 'Systems Biology of 2007 Influenza Vaccination in Humans' and 'A serological assay for seroconversion detection'.

<https://immport.org/shared/search>



ImmPort Shared Data Browser (Cohort Discovery Tool)

- ImmPort currently shares over 900 studies encompassing a range of research areas, species & assay types including 181 Clinical Trials data.


The screenshot shows the ImmPort Shared Data Browser interface. On the left, there are filter sections for 'Study Facets' (Research Focus, Condition or Disease, Biosample Type, Assay Methods, Program Name, Study Accession) and 'Subject Demographics' (Age (Years), Ethnicity, Gender, Race, Species). Below these is a 'Miscellaneous' section with checkboxes for 'Is Clinical Trial', 'Has Lab Tests', and 'Has Assessments Tests'. The main area features a search bar with the text 'e.g. Influenza, COVID-19, Rheumatoid, Lupus, MBAA .. (minimum 3 characters)'. Below the search bar, there are two tabs: 'Studies(869)' and 'Subjects (102648)'. The 'Facet Charts' section displays three charts: 'Research Focus' (a horizontal bar chart), 'Condition or Disease' (a pie chart), and 'Assay Methods' (a pie chart). Below the charts, a table lists studies with columns for 'Study', 'Title', 'Pubmed Id', 'Research Focus', 'Condition/Disease', 'Assay Methods', 'Latest Release Version', and 'Program Name'. The table shows three rows of study data.






Studies 927	Subjects 100062	Diseases 161
Experiments 3552	Total Results 7016441	Lab Tests 1285908



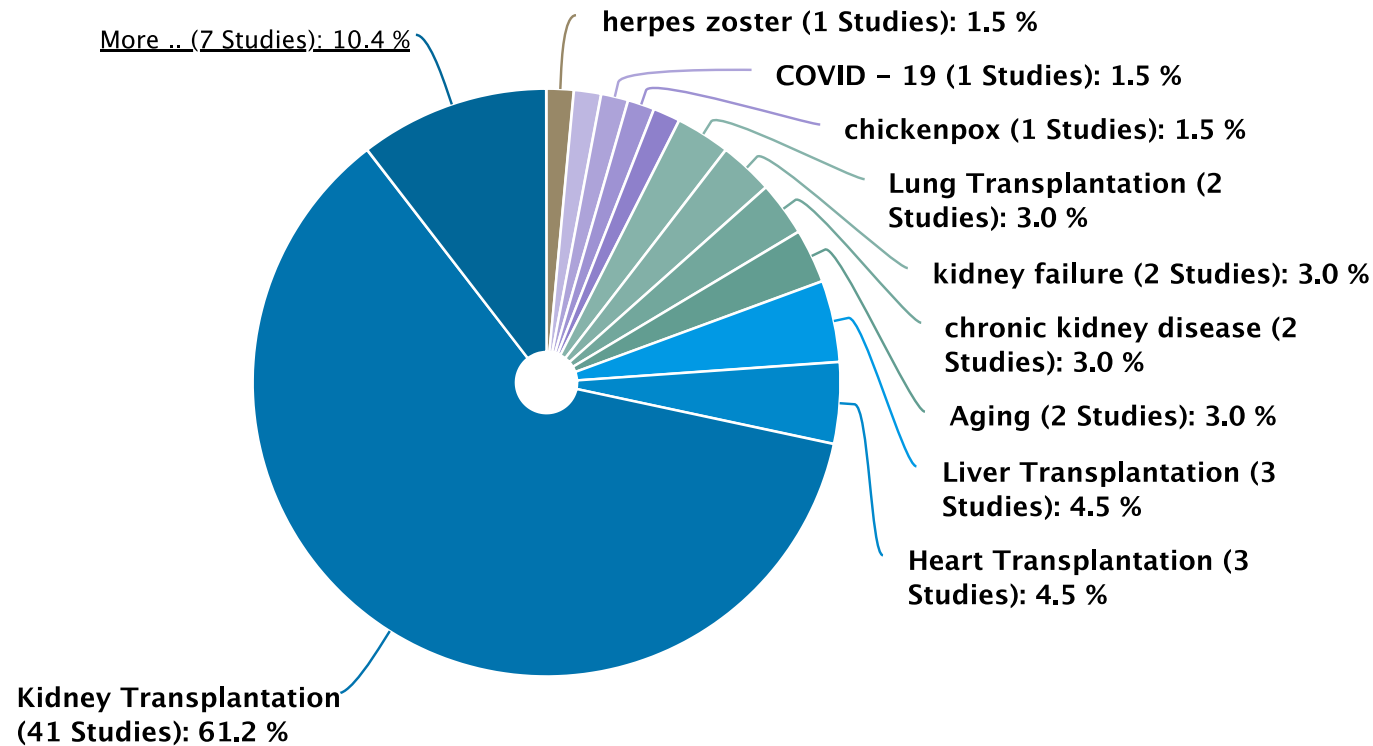
<https://immport.org/shared/search>

Data Summary: Release 51, March 2024

Click on the counts with  icon to visualize the count breakdown

Studies	 927
Subjects	 100062
Diseases	 161
Protocols	2246
Experiments	 3552
Total Results	 7016441
Elisa Results	382362
Elispot Results	71132
Flow Cytometry Results	5164196
PCR Results	127676
HAI Results	29297
HLA Typing Results	149342
Luminex/MBAA Results	1006232
Neutralizing Antibody Results	33523
GEO/SRA/Other	52681
Lab Tests	1285908
Total Test Panels	432

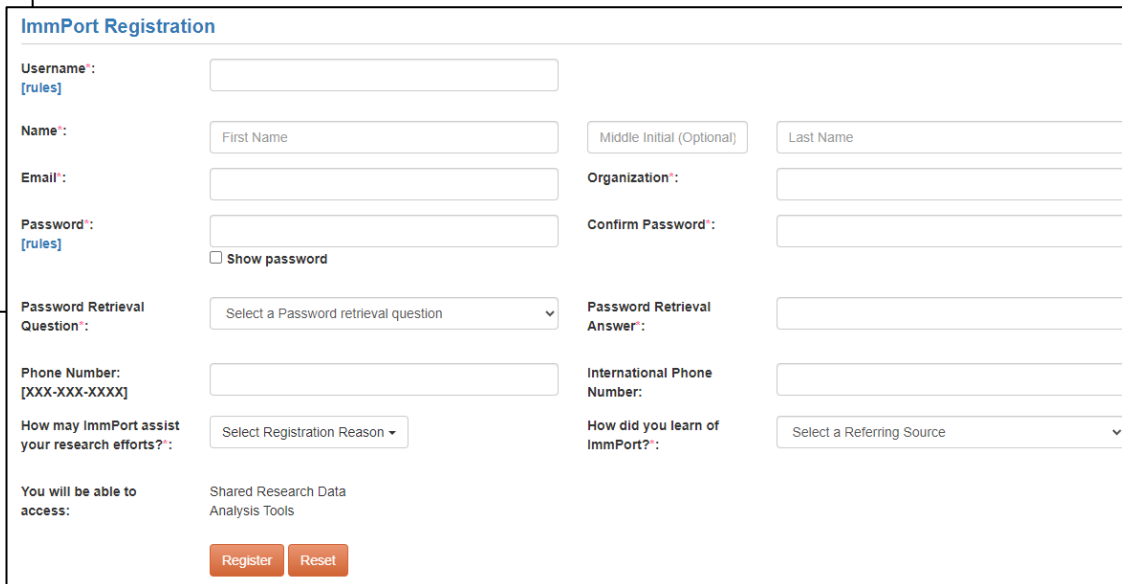
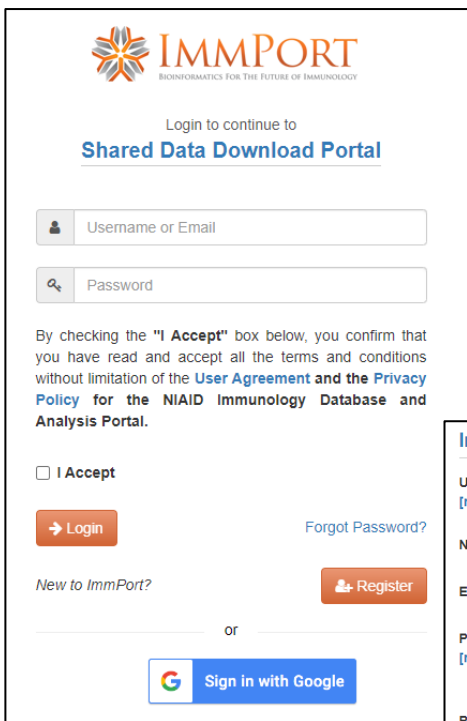
Condition or Disease



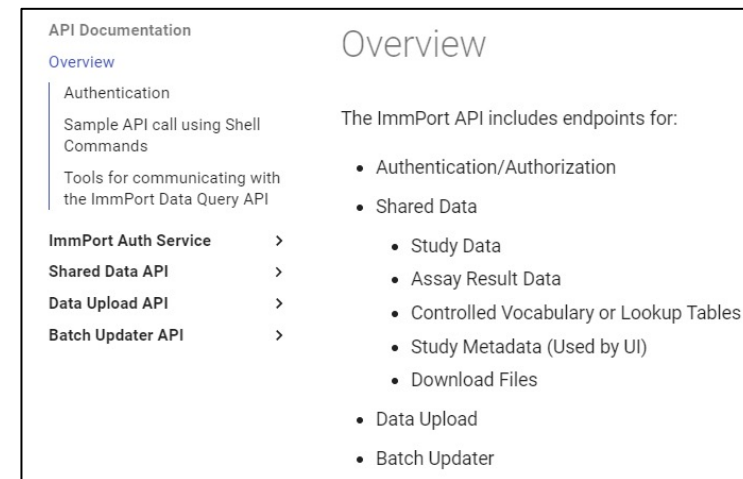
Accessible

ImmPort Registration & Login

- ImmPort study metadata (CDT Search) is browsable without login
- Registration and acceptance of Data Use Agreement is required to upload or download data
- Registration is free, simple, and immediate



ImmPort Application Programming Interfaces (APIs)



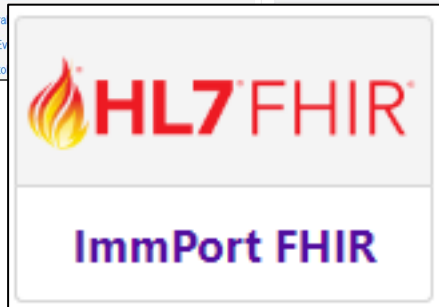
<https://docs.immport.org/apidocumentation/>

- ImmPort offers several APIs with detailed documentaiton for use

<https://www.immport.org/auth/login>

Interoperability with Other Resources

FHIR Resources (R5)	ImmPort Templates and Data Elements
Observation	Lab Test Component, Assessments Component
Specimen	Subject, Biosample, type
Patient	Subject (Human), gender, race, ethnicity
ResearchSubject	Linking Identifiers of Study, Subject, and Arm/cohort
DiagnosticReport	Lab Tests Panel, Assessments Panel, and link to the components
PractitionerRole	Study Personnel
	Subject: species, arm name, arm type
	Study Personnel



<https://fhir.immport.org/>

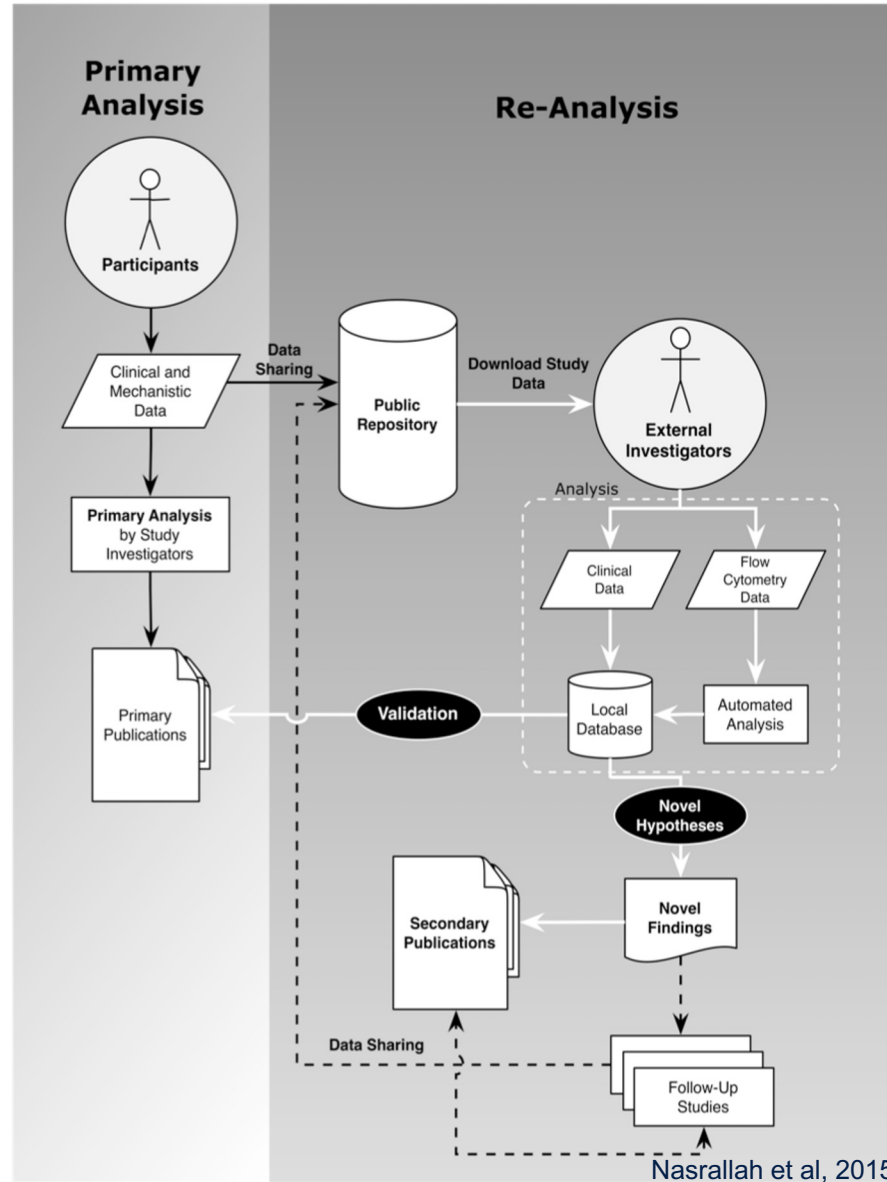
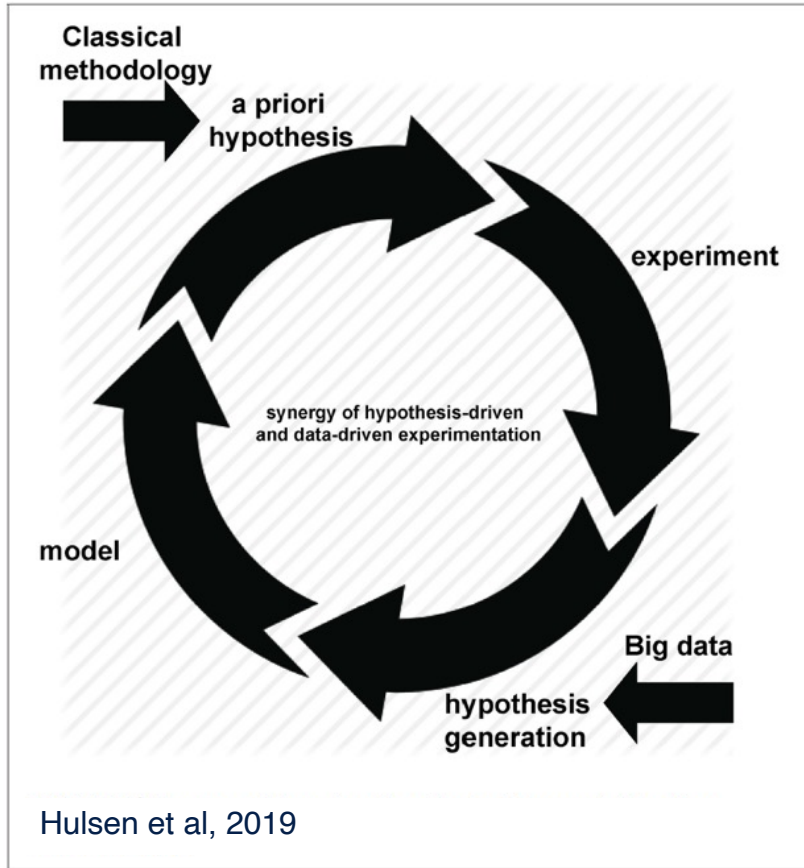
- To further interoperability, ImmPort data is being mapped to Fast Healthcare Interoperability Resources (FHIR) format
- Users can explore ImmPort data in FHIR format using the ImmPort HAPI FHIR server

Name	Value
ClinicalTrials.gov	https://clinicaltrials.gov/ct2/show/NCT00114881
SRA SRP102104	https://trace.ncbi.nlm.nih.gov/Traces/?view=study&acc=SRP102104
GEO GSE96783	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE96783
BioProject	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA379624
SRA SRP249918	https://trace.ncbi.nlm.nih.gov/Traces/?view=study&acc=SRP249918
GEO GSE145505	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145505
BioProject	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA607333
dbGaP	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002921.v2.p1
GEO GSE132181	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132181
BioProject	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA546267
Related study - GSE145505	https://www.immport.org/shared/search?text=sdyl025

Series GSE145505	Query DataSets for GSE145505
Status	Public on Apr 01, 2021
Title	Endotype of Allergic Asthma with Airway Obstruction in Urban Children
Organism	Homo sapiens
Experiment type	Expression profiling by high throughput sequencing

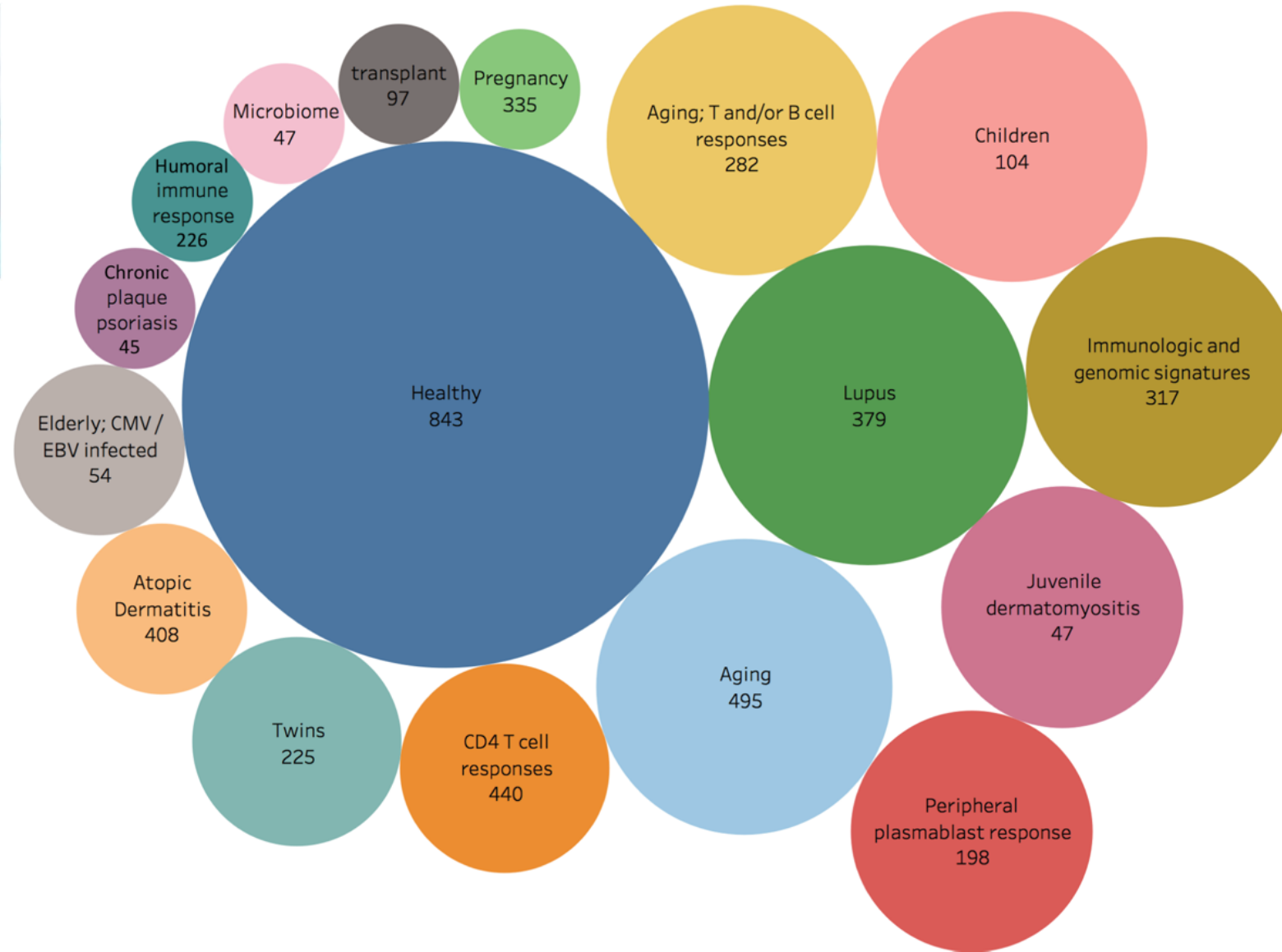
- ImmPort subject and sample metadata can be mapped to GEO subject metadata, creating a larger dataset for studies that have data in both repositories

Benefits of Open-Access Immunological Data





Reproducibility
Re-Analyze
Repurpose

Crowdsourcing: Influenza Vaccination Cohorts in ImmPort Database



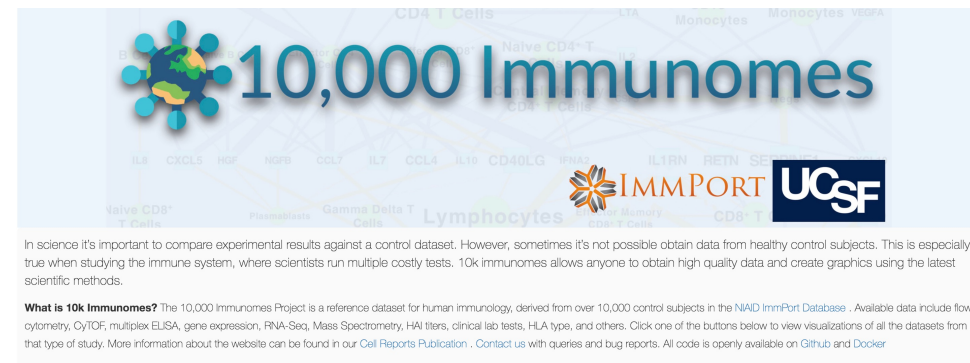
Resource

The 10,000 Immunomes Project: Building a Resource for Human Immunology

Kelly A. Zalocusky^{1, 2}, Matthew J. Kan^{1, 2}, Zicheng Hu^{1, 2}, Patrick Dunn³, Elizabeth Thomson³, Jeffrey Wiser³, Sanchita Bhattacharya^{1, 2, 4}, Atul J. Butte^{1, 2, 4, 5}  

10kimmunomes.org

- Large, diverse, cleaned reference dataset for human immunology
- Interactive data visualization
- Custom control cohorts and standardized data download



The banner features a central graphic of a globe with colorful dots representing data points. To the right, the text "10,000 Immunomes" is prominently displayed. Below this, the "IMMPORT" logo and the "UCSF" logo are visible. The background is a light blue grid with various immunological terms like "CD4+ T Cells", "Monocytes", "Lymphocytes", and "CD8+ T Cells" scattered across it.

In science it's important to compare experimental results against a control dataset. However, sometimes it's not possible obtain data from healthy control subjects. This is especially true when studying the immune system, where scientists run multiple costly tests. 10k Immunomes allows anyone to obtain high quality data and create graphics using the latest scientific methods.

What is 10k Immunomes? The 10,000 Immunomes Project is a reference dataset for human immunology, derived from over 10,000 control subjects in the NAID ImmPort Database. Available data include flow cytometry, CyTOF, multiplex ELISA, gene expression, RNA-Seq, Mass Spectrometry, HAI titers, clinical lab tests, HLA type, and others. Click one of the buttons below to view visualizations of all the datasets from that type of study. More information about the website can be found in our Cell Reports Publication. Contact us with queries and bug reports. All code is openly available on GitHub and Docker

 Transcriptomics

 Proteomics

 Immunoassays

 Lab Tests



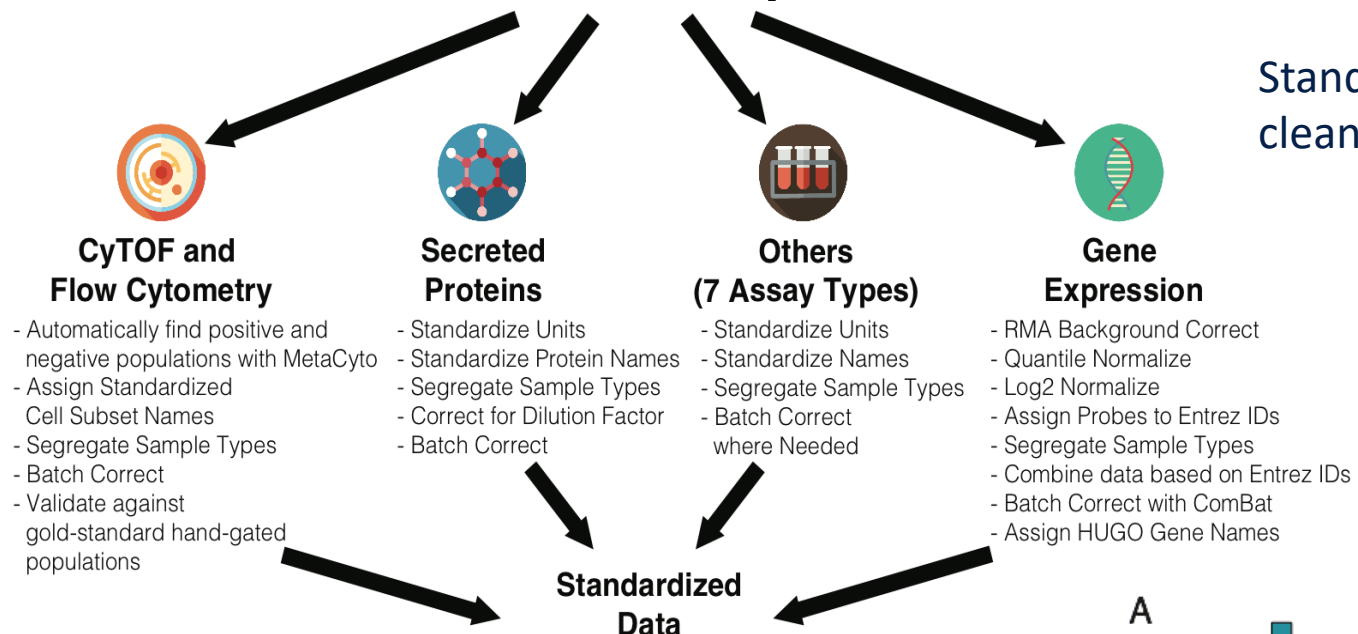
Data from 242 studies and 44,775 subjects was collected from the NAID Immunology Data and Analysis Portal, ImmPort. It includes flow cytometry, CyTOF, mRNA expression, secreted protein levels, clinical lab tests, HAI titers, HLA type, and others. We hand curated the entire contents of ImmPort to filter for normal

Data available in the 10,000 Immunomes Project

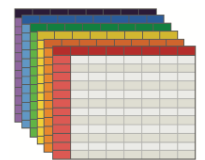
Total Samples	42117
Total Distinct Subjects	10344

MEASUREMENT	SUBJECTS
Secreted Proteins	4835
<i>ELISA</i>	4035
<i>Multiplex ELISA</i>	1286
Virus Titer	3609
<i>Virus Neutralization Titer</i>	2265
<i>HAI Titer</i>	1344
Clinical Lab Tests	2639
<i>Complete Blood Count</i>	1684
<i>Comprehensive Metabolic Panel</i>	664
<i>Fasting Lipid Profile</i>	664
Questionnaire	1422
Cytometry	1415
<i>Flow Cytometry (PBMC)</i>	907
<i>CytoF (PBMC)</i>	583
<i>Flow Cytometry (Whole Blood)</i>	164
HLA Type	1093
Gene Expression Array	476
<i>Whole Blood</i>	311
<i>PBMC</i>	165

**85 Studies
10,344 Subjects
42,000+ Samples**



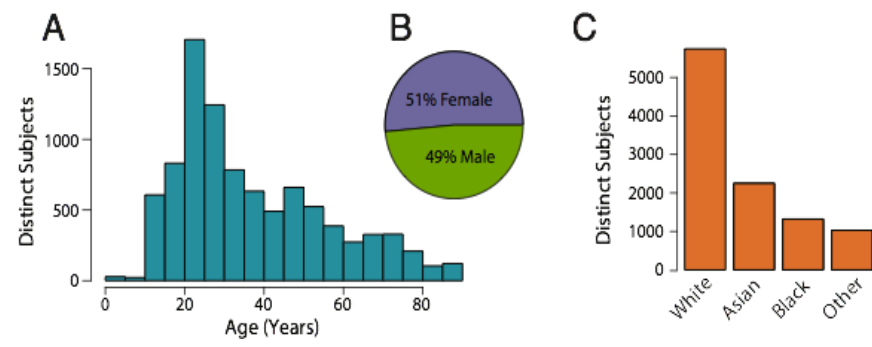
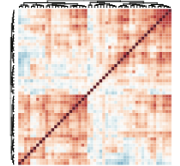
Standardized Data



Data Dissemination
Web Interface for Data Visualization and Download



Data Repurposing and Meta-analysis



Immunoassays

Immunoassays are a useful tool for obtaining precise measurements on the concentration of molecules, bacteria, and viruses inside of a sample. This technique relies on the strong bonding affinity between an antibody and its corresponding antigen. Specific antibodies are chosen to 'grab' the particle of interest. Immunoassays can also be used to measure the relative effectiveness of an antibody on a specific virus of interest. We provide immunoassays from ELISA, Multiplex ELISA, HAI Titer, and Virus Neutralization studies.

 Transcriptomics

 Proteomics

 Immunoassays

 Lab Tests

Select Data Type:

Normalized Outliers

Select Analyte:

Age Range:

Ethnicities:
 White
 Black or African American
 Asian
 Other

Sex:
 Female Male

Plot By:
 Age & Sex
 Ethnicity
 Study



Download

 Image  Plot Data  All Data

* All Data is dataset's raw and formatted files

Example of AI-ready ImmPort Data: Re-analysis of 10K Immunomes CyTOF Data Using GPT4

ChatGPT Prompt:

SA Find if there are any age related differences across cell types

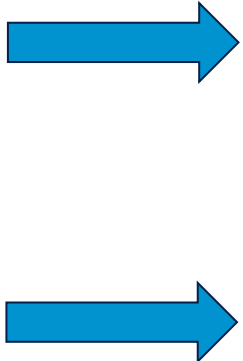
AI To determine if there are age-related differences across cell types, we can employ a few strategies:

- Visualization:** We can create scatter plots for each cell type against age to visually assess any trends.
- Correlation Analysis:** Compute the correlation coefficient between age and each cell type to quantify the strength and direction of the relationship.
- Statistical Testing:** Perform regression analysis or other statistical tests to determine if the observed relationships are statistically significant.

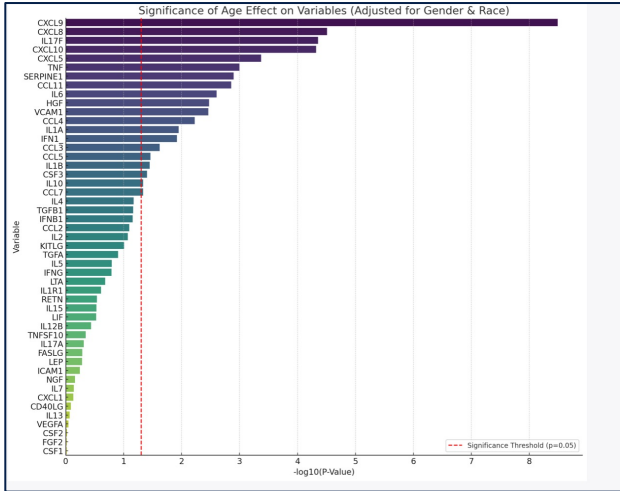
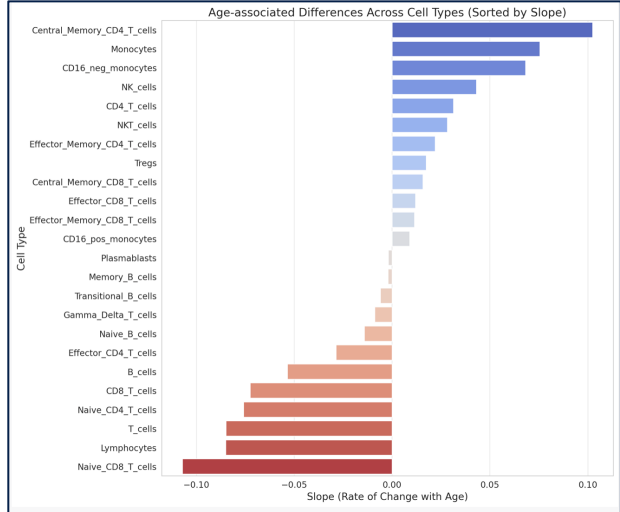
Let's start with the visualization to see if we can visually observe any patterns. We'll create scatter plots for each cell type against age.

Finished working Show work

ChatGPT Response:



Additional ChatGPT Response:



AI can analyze large scale cytometry datasets with ease, even adjusting for confounding variables

- Age-associated differences in cell types
- Age- and gender-associated effects on cytokines

ImmPort powered AI-Ready Datasets

Coming Soon!!



NAIRR AI-Ready Datasets





Powered by  ImmPort

In partnership with the National Artificial Intelligence Research Resource (NAIRR), ImmPort has prepared two AI-ready datasets for the [NAIRR Pilot](#). The NAIRR Pilot aims to connect U.S. researchers and educators to computational, data, and training resources needed to advance AI research and research that employs AI. Federal agencies are collaborating with government-supported and non-governmental partners to implement the Pilot as a preparatory step toward an eventual full NAIRR implementation.

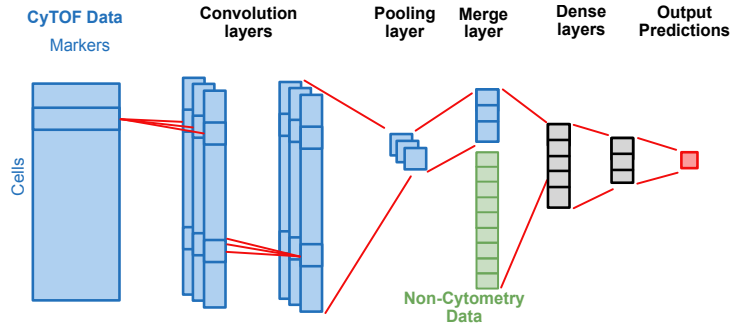
The two datasets will be listed on the NAIRR Pilot [Available Resources](#) page, along with other resources aligned with the NAIRR Pilot goals, such as pre-trained models, additional AI ready datasets, and relevant platforms. This resource will also provide high-quality immunology information for NAIRR researchers.

ImmPort Provided AI-Ready Datasets:

Dataset	Description	AI-Readiness Assessment
10k Immunomes Download 	The 10,000 Immunomes Project is a reference dataset for human immunology, derived from over 10,000 control subjects in the NIAID ImmPort Database. The dataset provides comprehensive profiles of the human immune system within healthy cohorts, representing various races. Read the publication in Cell Reports Browse ImmPort Studies in Shared Data Review the source code on GitHub or Docker	10k Immunomes Assessment
COVID-19 Compendium (Coming Soon) Download 	The COVID-19 compendium is a harmonized collection of curated COVID-19 studies, with a large number being part of NCI's SeroNet program. The dataset is compiled to enable meta-analysis of COVID-19 data by SeroNet and other communities.	COVID-19 Compendium Assessment (Coming Soon)

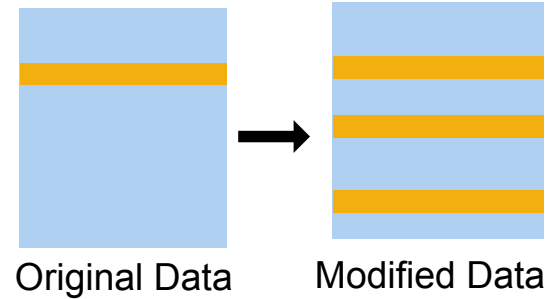
A robust and interpretable end-to-end deep learning model for cytometry data

A convolutional neural network (CNN) for cytometry data

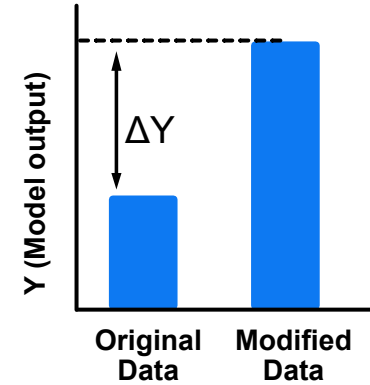


Goal: To diagnose the latent cytomegalovirus (CMV) in healthy individuals

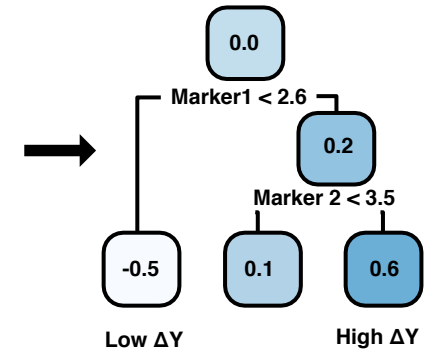
Step 1: up-sample each cell in the data



Step 2: Calculate the changes in model output (ΔY)

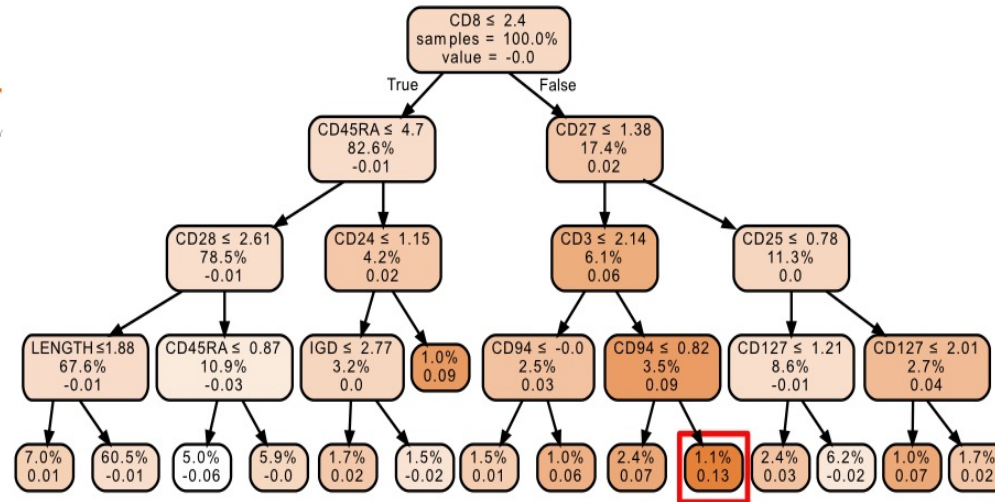
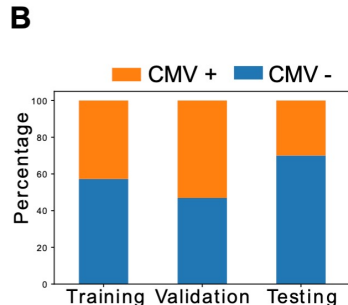
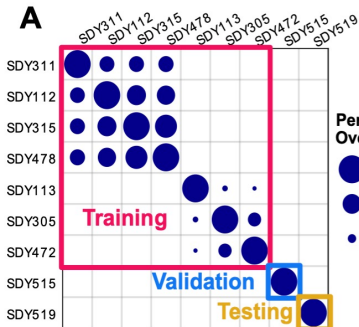


Step 3: Identify cell associated with high ΔY

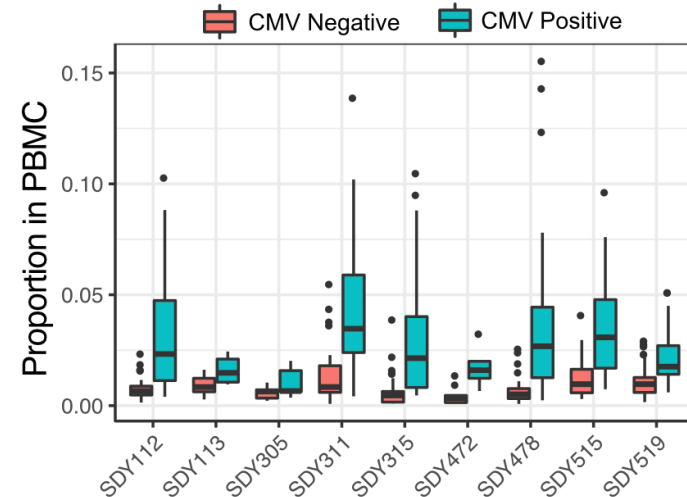


The dataset contains

- CyTOF and CMV serological data
- 472 healthy individuals
- 9 studies



CD3+ CD8+ CD27- CD94+ cells



Visualizing Open-Access Living Donor Transplant Data

Organ Transplant @

 27 clinical trials

Study Curation:

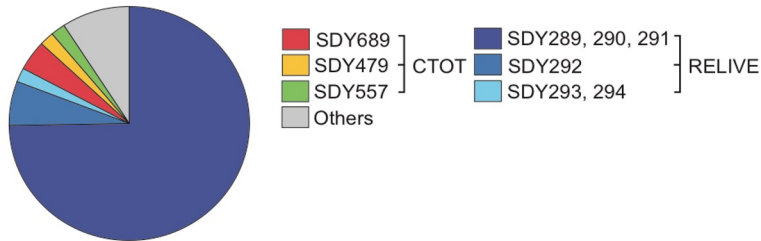
- clinical trials that excluded or have insufficient living donation information

20 Curated clinical trials

Subject Curation:

- deceased donors
- living donors with >95% missing data

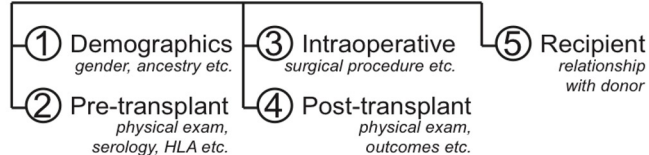
11,263 Subjects



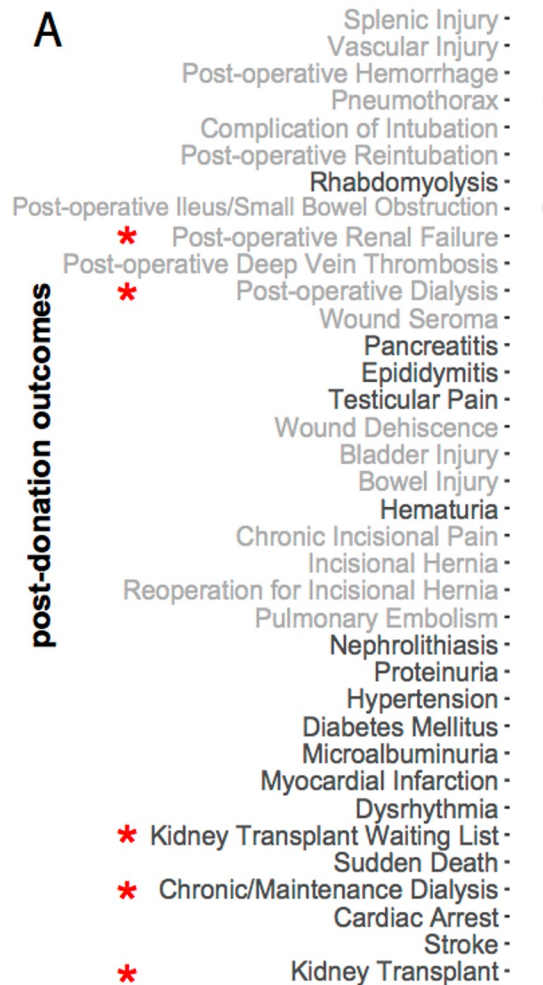
Data Curation:

- Data compilation
- Data standardization

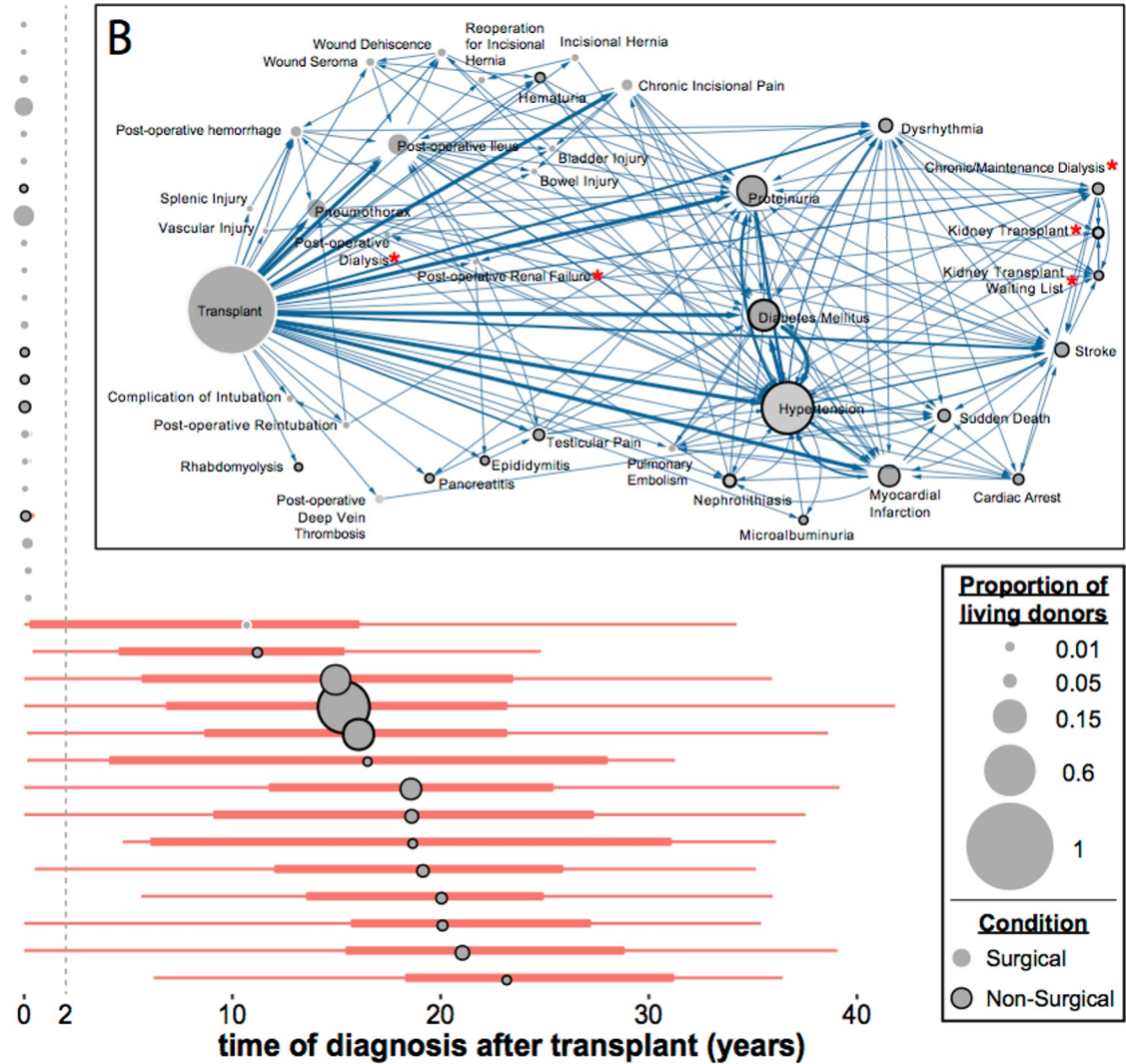
Curated Data



A



B



Article | Published: 12 July 2021

Systems vaccinology of the BNT162b2 mRNA vaccine in humans

Prabhu S. Arunachalam, Madeleine K. D. Scott, [...] Bali Pulendran ✉

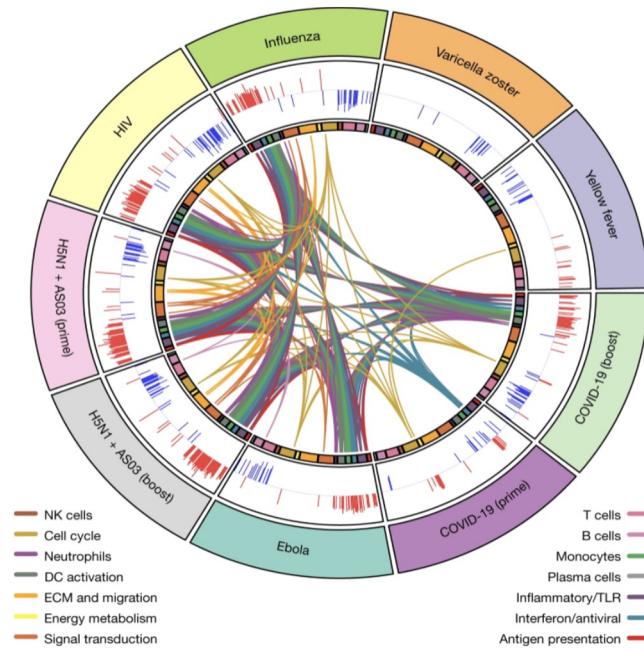
Nature 596, 410–416 (2021) | Cite this article

61k Accesses | 1109 Altmetric | Metrics

PMID: 34252919 PMCID: PMC8761119 DOI: 10.1038/s41586-021-03791-x

Fig. 5: Comparison of transcriptional responses with other vaccines.

From: Systems vaccinology of the BNT162b2 mRNA vaccine in humans



Online attention



- 2556 tweeters
- 6 blogs
- 2 Facebook pages
- 64 news outlets
- 11 Redditors
- 1 F1000
- 1 Video uploaders
- 396 Mendeley

Extended Data Table 3 Vaccine meta-analysis datasets

From: Systems vaccinology of the BNT162b2 mRNA vaccine in humans

Vaccine	Pathogen	Vaccine Type	Adjuvant/ Vector	Timepoints used	N	GEO/ ImmPort
BNT162b2	SARS-CoV-2	mRNA	mRNA- LNP	0,1,7, 21,22,28	31	GSE169159
TIV	Seasonal Influenza	Inactivated	None	0,1,7	19	GSE74813/ SDY56
VZV	Varicella zoster	Live attenuated	VZV	0,1,7	31	GSE79396/ SDY984
YF17D	Yellow fever	Live attenuated	YF17D	0,1,7	25	GSE13486/ SDY1264
rVSV-ZEBOV (high dose)	Ebola	Recombinant viral vector	VSV	0,1,7	7	GSE97590/ SDY1373
MRKAd5/HIV	HIV	Recombinant viral vector	Ad5	0,1,7	10	GSE22768/ SDY1291
H5N1+AS03	H5N1 Influenza	Inactivated	AS03	0,1,7, 21,22,28	33	GSE102012
MPSV4	Meningococcus	Polysaccharide	None	0,7	13	GSE52245/ SDY1260
MCV4	Meningococcus	Conjugate	None	0,7	17	GSE52245/ SDY1260

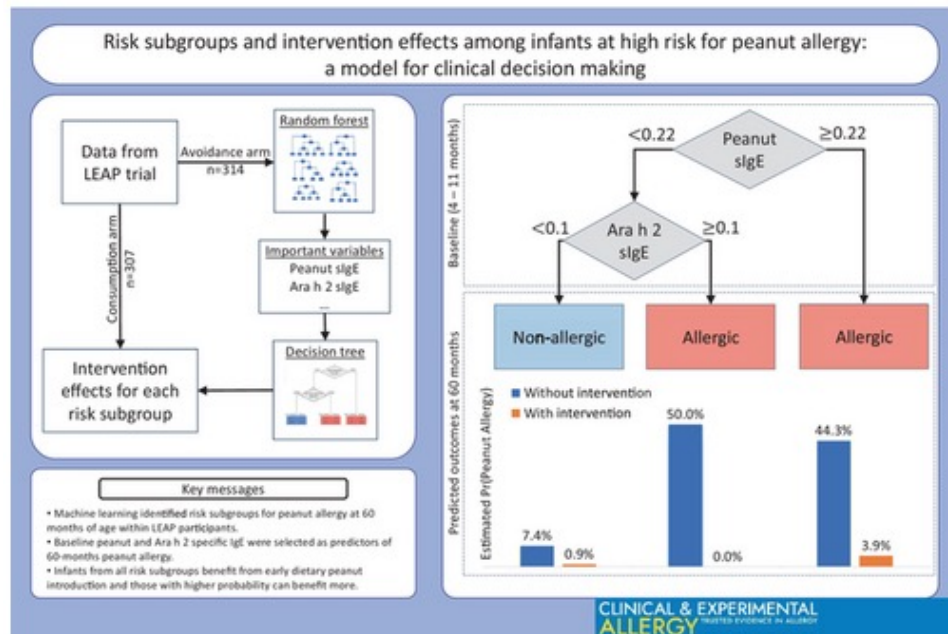
> Clin Exp Allergy. 2024 Mar;54(3):185-194. doi: 10.1111/cea.14452. Epub 2024 Jan 19.

Risk subgroups and intervention effects among infants at high risk for peanut allergy: A model for clinical decision making

Yuxiang Li^{1,2}, Ashley Devonshire^{3,4}, Bin Huang^{1,3}, Sandra Andorf^{1,3,4,5}

Affiliations + expand

PMID: 38243616 PMID: PMC10932885 (available on 2025-03-01) DOI: 10.1111/cea.14452



Background

The Learning Early About Peanut Allergy (LEAP) trial showed that early dietary introduction of peanut reduced the risk of developing peanut allergy by age 60 months in infants at high risk for peanut allergy. In this secondary analysis of LEAP data, we aimed to determine risk subgroups within these infants and estimate their respective intervention effects of early peanut introduction.

Key messages

- Machine learning identified risk subgroups for peanut allergy at 60 months within LEAP participants.
- Baseline peanut and Ara h 2-specific-IgE were selected as predictors of 60-months peanut allergy.
- Infants with higher predicted probability of peanut allergy benefit more from the early introduction intervention.

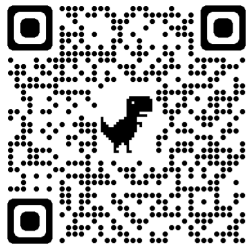
The individual participant-level data of LEAP are made available through ImmPort (SDY660) and ITN TrialShare (ITN032AD, www.itntrialshare.org) from where the data were retrieved.²³⁻²⁵ More details about the study population and data preprocessing can be found in the Supplemental methods and Table S1.

ImmuneSpace

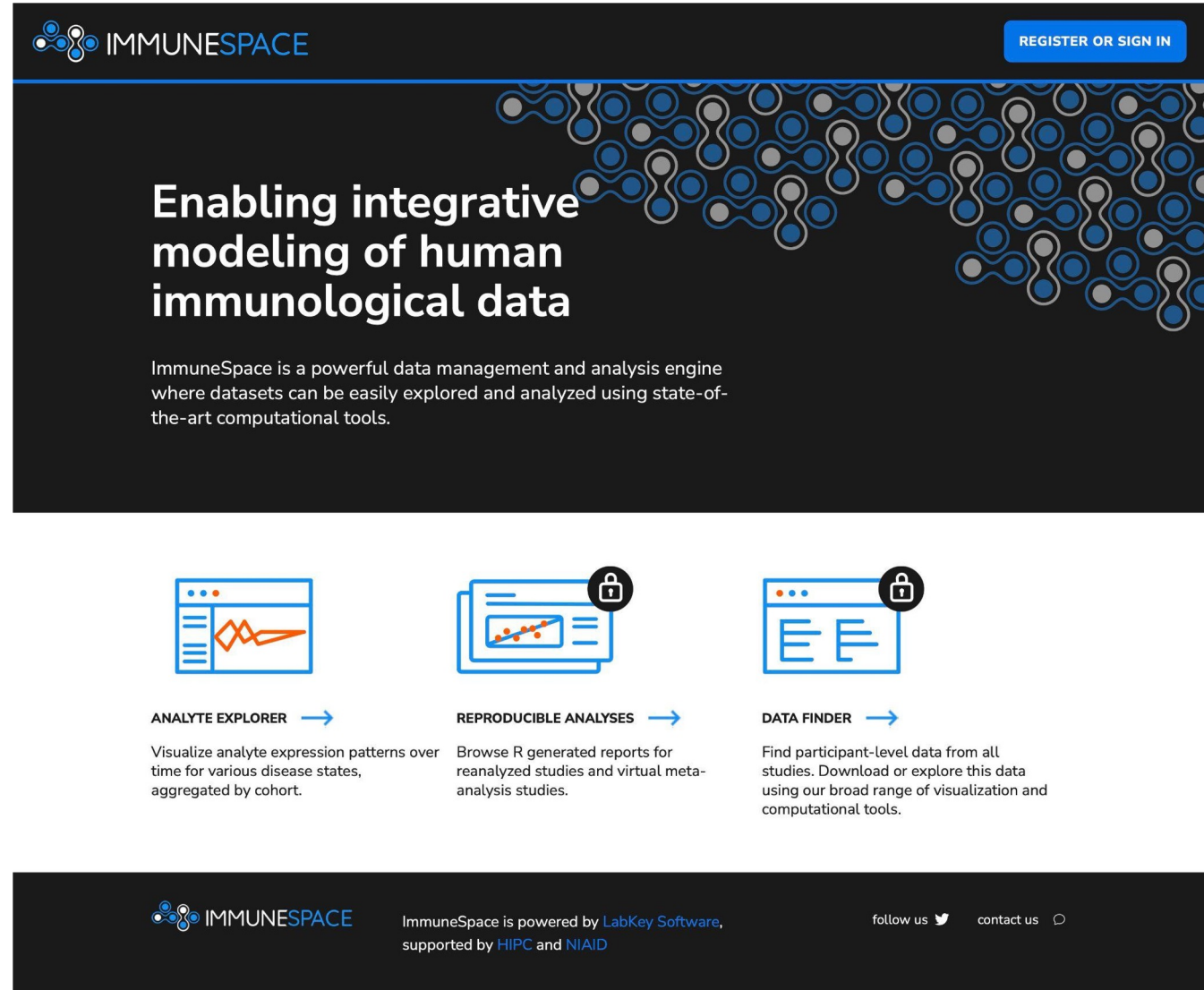
Human Immunology Project Consortium (HIPC)

HIPC's ImmuneSpace extends ImmPort, providing access to additional data (e.g., standardized gene expression matrices) and web-based R tools for data accession, analysis, and reporting.

Studies in the Immune Signatures Data Resource are archived through the Shared Data Portal on ImmPort and ImmuneSpace repositories and may be updated over time.



<https://immunespace.org>





The banner features the ImmuneSpace logo in the top left, a 'REGISTER OR SIGN IN' button in the top right, and a decorative pattern of blue and white circles on the right side. The main text reads 'Enabling integrative modeling of human immunological data'. Below this, a paragraph describes ImmuneSpace as a powerful data management and analysis engine. At the bottom, three feature cards are displayed: 'ANALYTE EXPLORER' with a line graph icon, 'REPRODUCIBLE ANALYSES' with a document and lock icon, and 'DATA FINDER' with a search icon. Each card includes a brief description of the feature.

Enabling integrative modeling of human immunological data

ImmuneSpace is a powerful data management and analysis engine where datasets can be easily explored and analyzed using state-of-the-art computational tools.

- ANALYTE EXPLORER** → Visualize analyte expression patterns over time for various disease states, aggregated by cohort.
- REPRODUCIBLE ANALYSES** → Browse R generated reports for reanalyzed studies and virtual meta-analysis studies.
- DATA FINDER** → Find participant-level data from all studies. Download or explore this data using our broad range of visualization and computational tools.

ImmuneSpace is powered by [LabKey Software](#), supported by [HIPC](#) and [NIAID](#)

follow us  contact us 

Education : Analysis Tutorial

Analysis of HAI data using SDY212

This tutorial is based on the original R Markdown Tutorial written by Dr. Sandra Andorf, The original work is available at the [ImmPort Resources Tutorials](#). We would like to thank Dr. Andorf for her dedication and hard work in developing this tutorial.

SDY212

[SDY212](#) is one of the many public data sets available from [ImmPort](#).

Title: Apoptosis and other immune biomarkers predict influenza vaccine (TIV 2008) responsiveness

Principal Investigator: Mark M. Davis

Description: In an effort to identify benchmarks of immunological health, influenza vaccination was used in 30 young (20 to 30 years) and 59 older subjects (60 to 89 years) as models for strong and weak immune responses, respectively.

Download the MetaData Files Representing SDY212

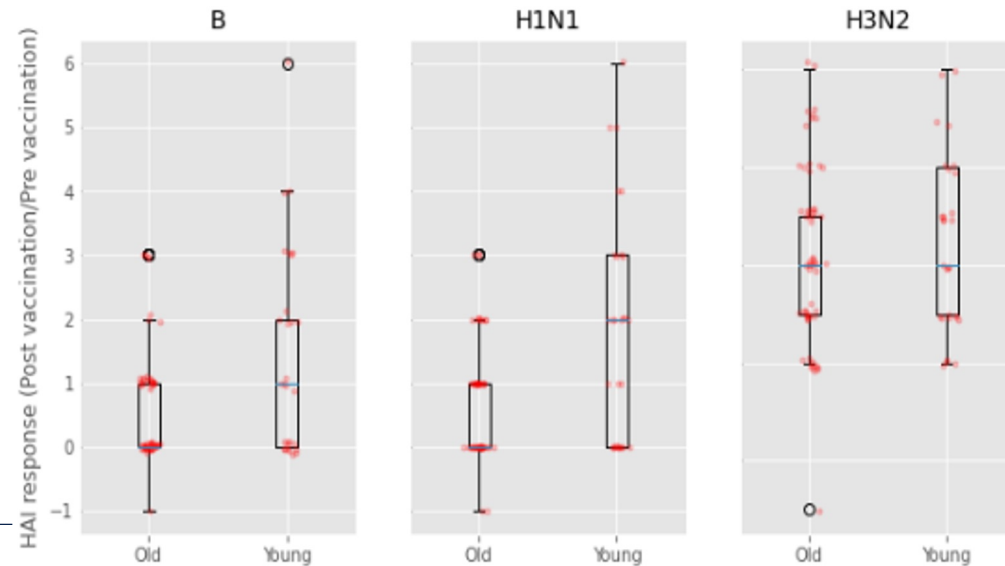
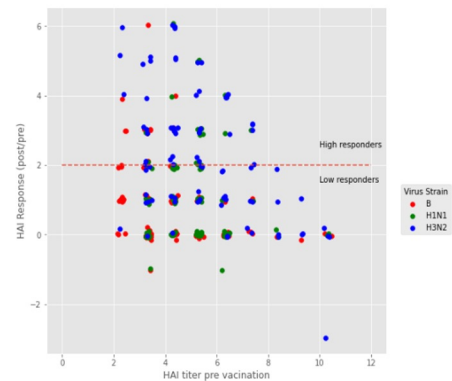
In this case we will be using either the preloaded SDY212 metadata files or the ImmPort [File Download API](#) to download files for this analysis. ImmPort data can also be downloaded using the ImmPort Data Browser or the ImmPort Data Query API. Additional information about the Data Query API is available [here](#).

For additional information we recommend these tutorials:

1. [Downloading ImmPort Data](#)
2. [Download_FCS_Files_File_Download_API](#)
3. [Download HAI Results_DataQuery_API](#)

Plot of HAI response vs pre (Day 0) vaccination HAI titer

```
[36]: fig, (ax) = plt.subplots(1,1,figsize=(8, 8))
ax.scatter(rand_jitter_log(np.log2(B_strain['Day0'])),rand_jitter_log(B_strain['FOLD_CHANGE_LOG2']),label="B",c="red")
ax.scatter(rand_jitter_log(np.log2(H1N1_strain['Day0'])),rand_jitter_log(H1N1_strain['FOLD_CHANGE_LOG2']),label="H1N1",c="green")
ax.scatter(rand_jitter_log(np.log2(H3N2_strain['Day0'])),rand_jitter_log(H3N2_strain['FOLD_CHANGE_LOG2']),label="H3N2",c="blue")
ax.set_ylabel('HAI Response (post/pre)')
ax.set_xlabel('HAI titer pre vaccination')
ax.hlines(2,0,12,linestyle='dashed')
ax.text(10,2.5,"High responders")
ax.text(10,1.5,"Low responders")
ax.legend(bbox_to_anchor=(1.2, .5),scatterpoints=1,title="Virus Strain")
plt.show()
```

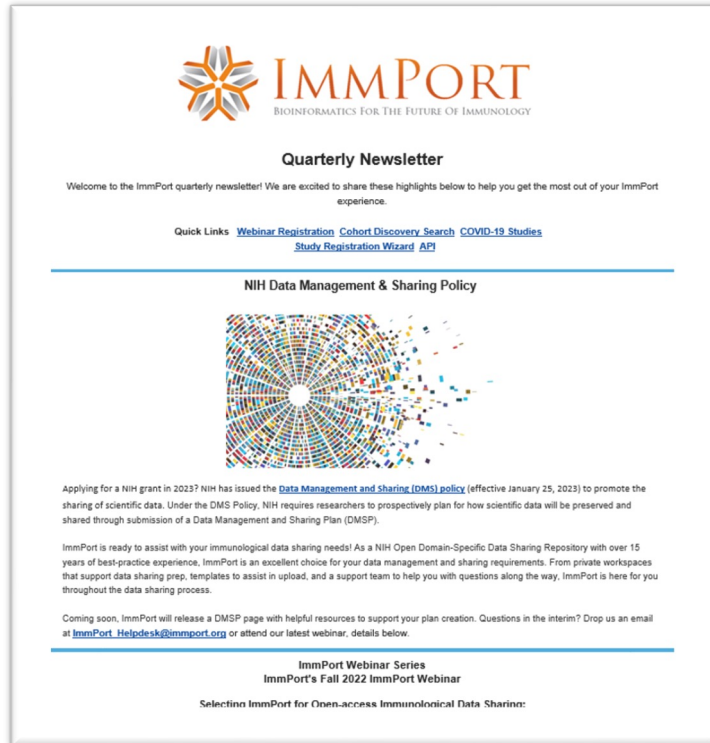


Take Home Messages

- Open-access immunological studies are a valuable resource to evaluate new in silico hypotheses testing, gain novel insights, and a productive starting point for informing the design of future experiments
- Holistic approach to analyzing clinical research data
- 10,000 Immunomes Project- a framework for growing a diverse human immunology reference, from ImmPort, a publicly available resource of subject-level immunology.
 - Allows us to learn from the features and candidates we already know.
 - Enables us to explore new factors to be discovered.
- Deep convolutional neural network model can accurately diagnose the latent cytomegalovirus (CMV) in healthy individuals.
- Expanded uses of crowdsourcing in immunology will allow for more efficient large-scale data collection and analysis. It will also involve, inspire, educate, and engage the community in a variety of meaningful ways.

Embrace open-access datasets!

Ways to Stay Updated on ImmPort Activities



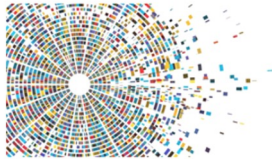
IMMPORT
BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY

Quarterly Newsletter

Welcome to the ImmPort quarterly newsletter! We are excited to share these highlights below to help you get the most out of your ImmPort experience.

Quick Links [Webinar Registration](#) [Cohort Discovery Search](#) [COVID-19 Studies](#)
[Study Registration Wizard](#) [API](#)

NIH Data Management & Sharing Policy



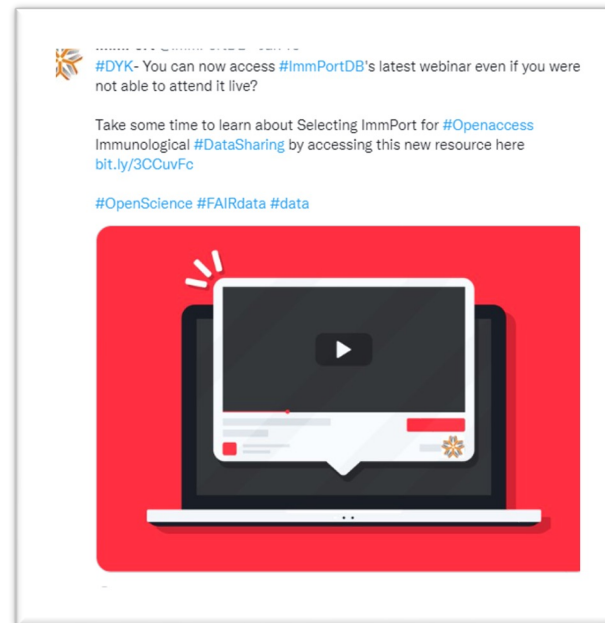
Applying for a NIH grant in 2023? NIH has issued the [Data Management and Sharing \(DMS\) policy](#) (effective January 25, 2023) to promote the sharing of scientific data. Under the DMS Policy, NIH requires researchers to prospectively plan for how scientific data will be preserved and shared through submission of a Data Management and Sharing Plan (DMSP).

ImmPort is ready to assist with your immunological data sharing needs! As a NIH Open Domain-Specific Data Sharing Repository with over 15 years of best-practice experience, ImmPort is an excellent choice for your data management and sharing requirements. From private workspaces that support data sharing prep, templates to assist in upload, and a support team to help you with questions along the way, ImmPort is here for you throughout the data sharing process.

Coming soon, ImmPort will release a DMSP page with helpful resources to support your plan creation. Questions in the interim? Drop us an email at ImmPort_Helpdesk@import.org or attend our latest webinar, details below.

ImmPort Webinar Series
ImmPort's Fall 2022 ImmPort Webinar

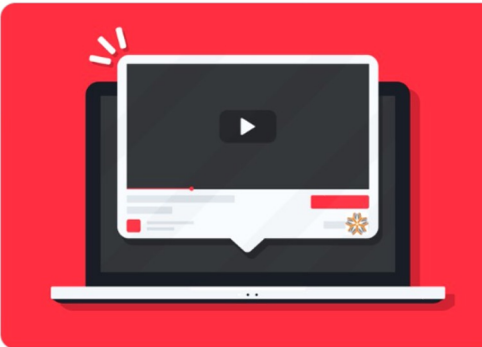
Selecting ImmPort for Open-access Immunological Data Sharing:



#DYK- You can now access **#ImmPortDB**'s latest webinar even if you were not able to attend it live!

Take some time to learn about Selecting ImmPort for **#Openaccess** Immunological **#DataSharing** by accessing this new resource here bit.ly/3CCuvFc

#OpenScience #FAIRdata #data



Monthly Newsletter

[Subscribe to our newsletter](#)

[Add us to your address book](#)

<https://docs.import.org/home/newsletter/>

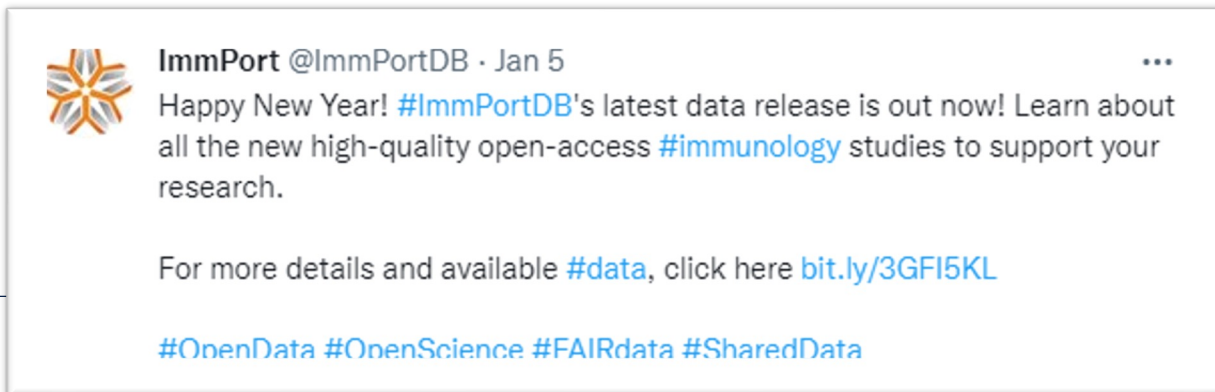
X [@ImmPortDB](#)

[LinkedIn](#)

<https://www.linkedin.com/company/import/>



ImmPort_Helpdesk@import.org



ImmPort @ImmPortDB · Jan 5

Happy New Year! **#ImmPortDB**'s latest data release is out now! Learn about all the new high-quality open-access **#immunology** studies to support your research.

For more details and available **#data**, click here bit.ly/3GFI5KL

#OpenData #OpenScience #FAIRdata #SharedData



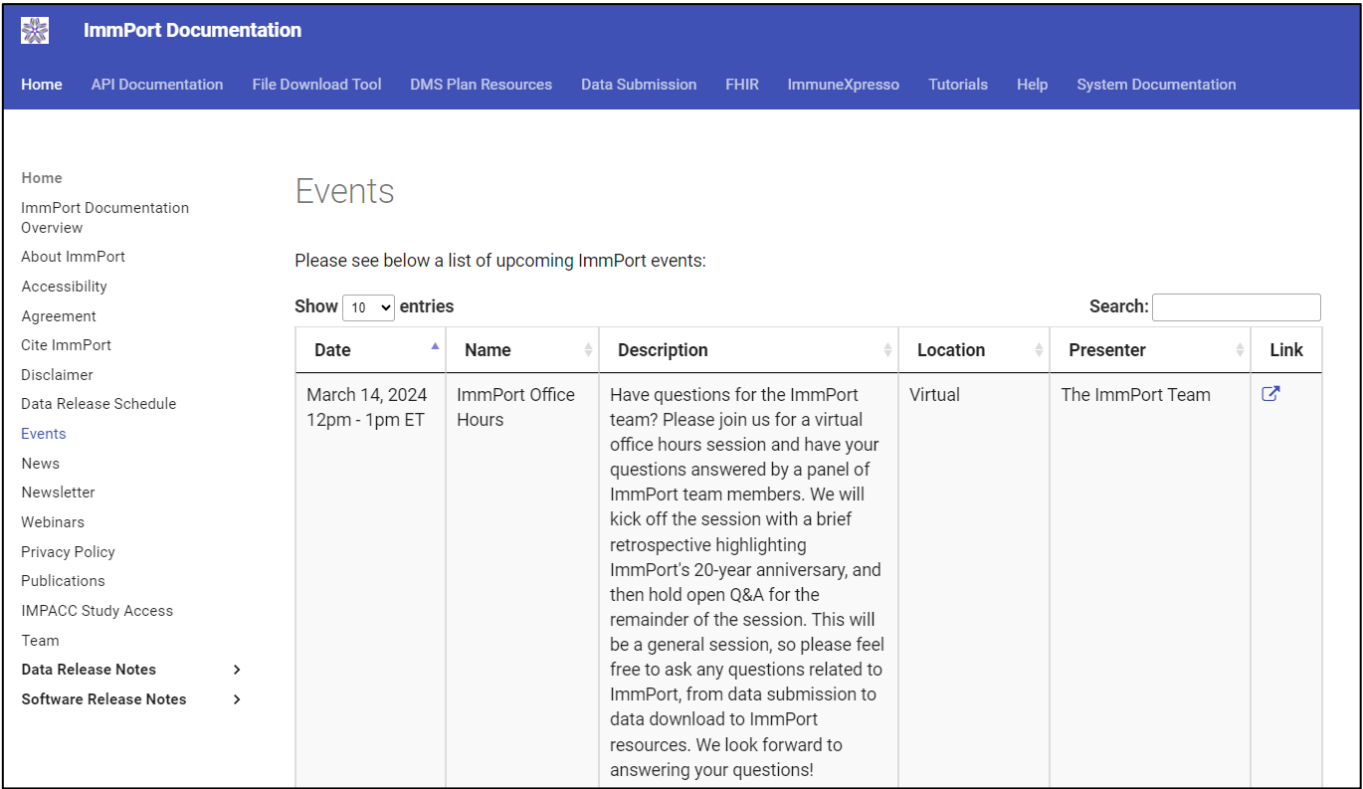
IMMPORT
Resources
Your site for documentation,
training, and reference materials

Visit the ImmPort Events page to add ImmPort Office Hours to your calendar

<https://docs.immport.org/home/events/>



- ImmPort holds open office hours sessions on the first Thursday of each month from 2 PM – 3 PM ET
- Office Hours are a great opportunity to discuss your questions directly with the ImmPort team and learn more about ImmPort
- All user levels are welcome, whether new to ImmPort or an experienced user



ImmPort Documentation

Home API Documentation File Download Tool DMS Plan Resources Data Submission FHIR ImmuneXpresso Tutorials Help System Documentation

Home
ImmPort Documentation Overview
About ImmPort
Accessibility
Agreement
Cite ImmPort
Disclaimer
Data Release Schedule
Events
News
Newsletter
Webinars
Privacy Policy
Publications
IMPACC Study Access
Team
Data Release Notes >
Software Release Notes >

Events

Please see below a list of upcoming ImmPort events:

Show 10 entries Search:

Date	Name	Description	Location	Presenter	Link
March 14, 2024 12pm - 1pm ET	ImmPort Office Hours	Have questions for the ImmPort team? Please join us for a virtual office hours session and have your questions answered by a panel of ImmPort team members. We will kick off the session with a brief retrospective highlighting ImmPort's 20-year anniversary, and then hold open Q&A for the remainder of the session. This will be a general session, so please feel free to ask any questions related to ImmPort, from data submission to data download to ImmPort resources. We look forward to answering your questions!	Virtual	The ImmPort Team	🔗

HOME / EDUCATION / BIG DATA IN IMMUNOLOGY

Big Data in Immunology

Tuesday, June 18, 2024

San Francisco Marriott Marquis
780 Mission Street, San Francisco, California 94103



Overview ▾

Program ▾

Registration ▾

Overview

Overview

Are you interested in exploring public datasets? Or using shared immunology study results to further facilitate your research and gain insights into discoveries? Strengthen your funding requests? Want to develop analytical skills and hear from those who have done it?

The open-access [ImmPort database](#) provides publicly available clinical trial and immunology research datasets and bioinformatics tools for analyses. ImmPort is an NIAID-funded data repository created to share immunology-related research data with the scientific community. ImmPort currently shares over 900 studies with new data released monthly.

Attend the ImmPort workshop to hear from scientists who have published on repurposed data. Plus, attend for a hands-on introduction to the ImmPort database and become familiar with mining the ImmPort data. Experts from the [Immunology Database and Analysis Portal \(ImmPort\)](#) will teach the workshop.

Thanks!

Sanchita.Bhattacharya@ucsf.edu

Bakar Institute of Computational Health Sciences, UCSF

X @sanchitab

