



ALLEN INSTITUTE *for*
IMMUNOLOGY

Automating Immunology: From Raw Data to Insight

Aishwarya Gogate

Senior Bioinformatics Analyst

January 30, 2025

My background

BE in Biotechnology

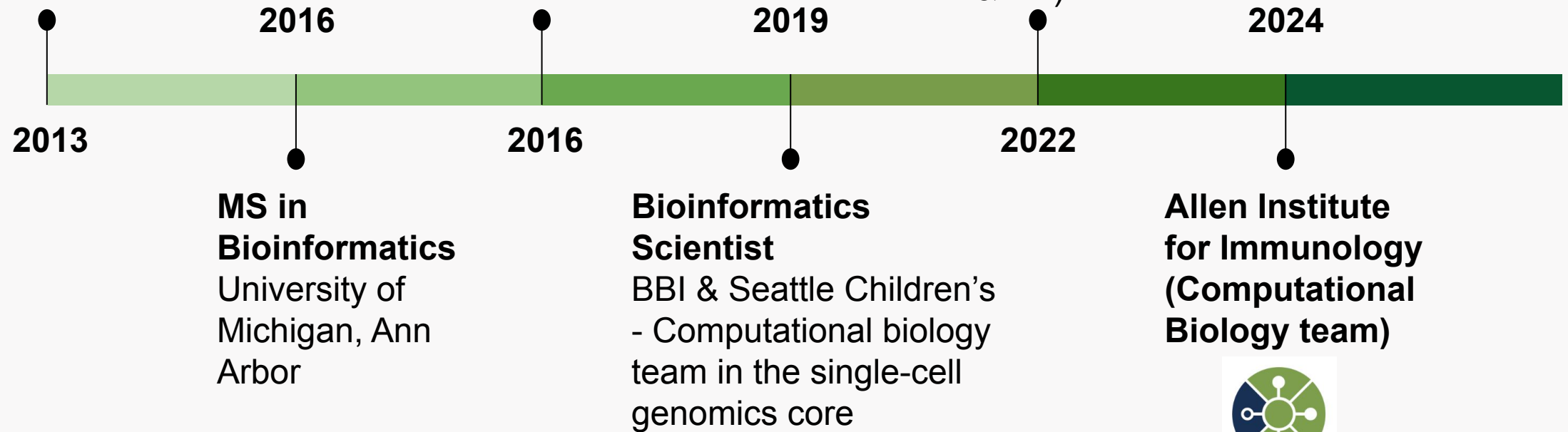
Visvesvaraya
Technological University,
Bangalore India

Computational Biologist

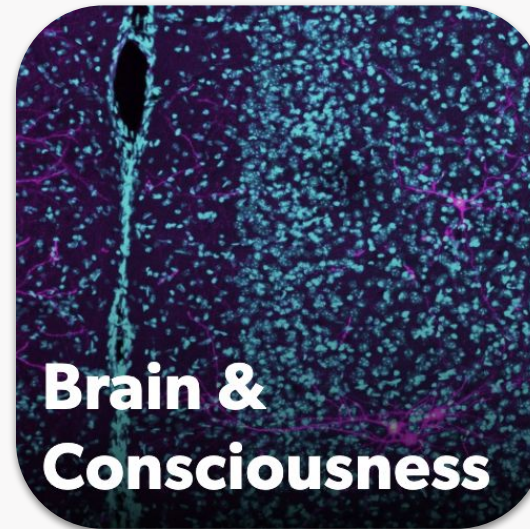
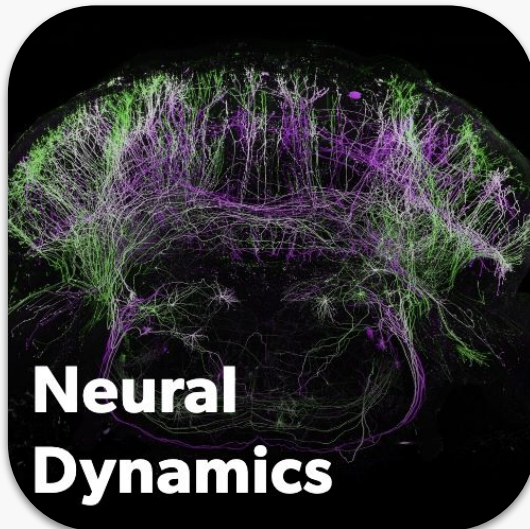
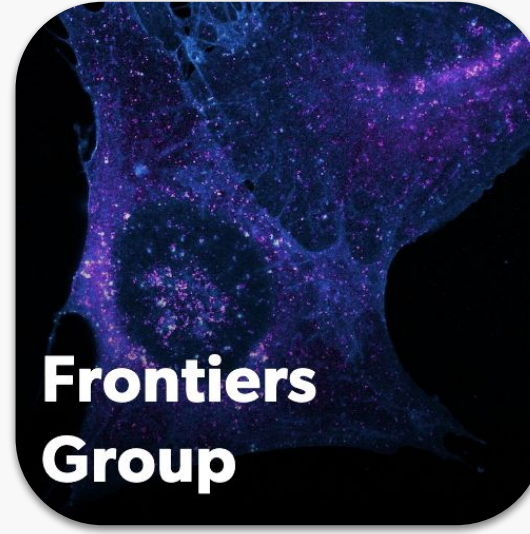
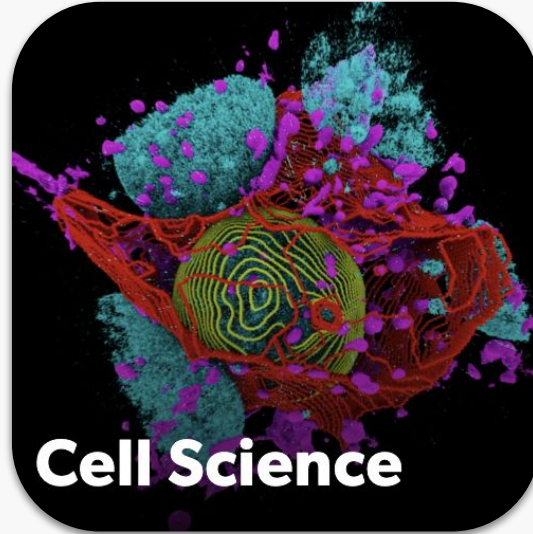
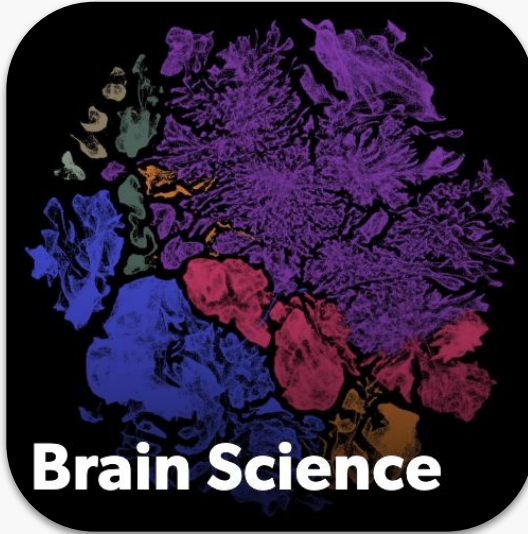
UTSW Med Ctr, Dallas TX
(epigenetics & chromatin
biology lab)

Bioinformatics Scientist

Cajal Neuroscience -
Computational biology team
(target discovery efforts in
PD & AD)



Science Programs at the Allen Institute

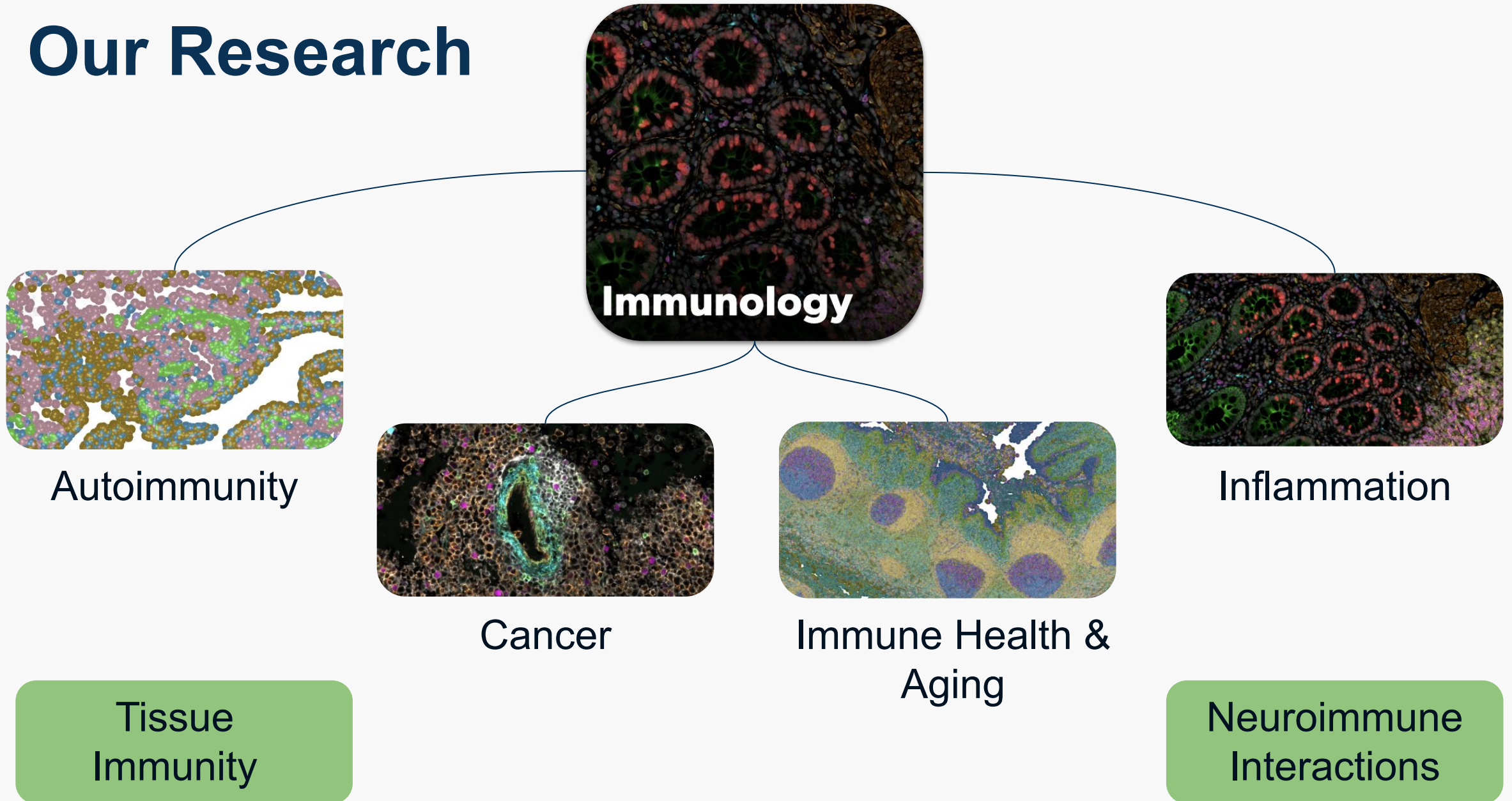


Allen Institute for Immunology (AIFI)



Team of 100
and growing!

Our Research



Our Leadership Team



Ananda Goldrath, Ph.D

Executive Vice President

Experimental Immunology



Troy Torgerson, M.D.,
Ph.D



Ernie Coffey, MBA



Lynne A. Becker, Ph.D.

Scientific Operations &
Program Management

Computational Biology



Xiaojun Li, Ph.D.



Peter Skene, Ph.D.

High Resolution Translational
Immunology

Software Development



Paul Meijer, Ph.D.

Team Science at AIFI!

01

Wealth of expertise to draw from

- Immunologists
- Bioinformaticians & Biostatisticians
- Software Engineers

02

Allen Hour

- Exposure to research from different business units
- Great for networking and early career exploration

03

Lunch and Learn

- Interact with other scientists at all levels
- Meet interns from other science units

“Testimonials”

“My internship at the Allen Institute for Immunology was a great experience. I had the opportunity to work on a **project focused on understanding the healthy immune system at baseline and its responses to ‘dys’regulation in children.**

Additionally, I **collaborated with the spatial transcriptomics team on tasks beyond my primary project**, which was a great learning experience and rewarding. Beyond the work itself, I found the people at the Allen Institute to be welcoming and open to discussing their research. The **supportive and collaborative work culture** made the experience even more enjoyable.”

- **Temi Adewunmi**
(Current Bioinformatics Analyst at AIFI and U of Oregon alumni)

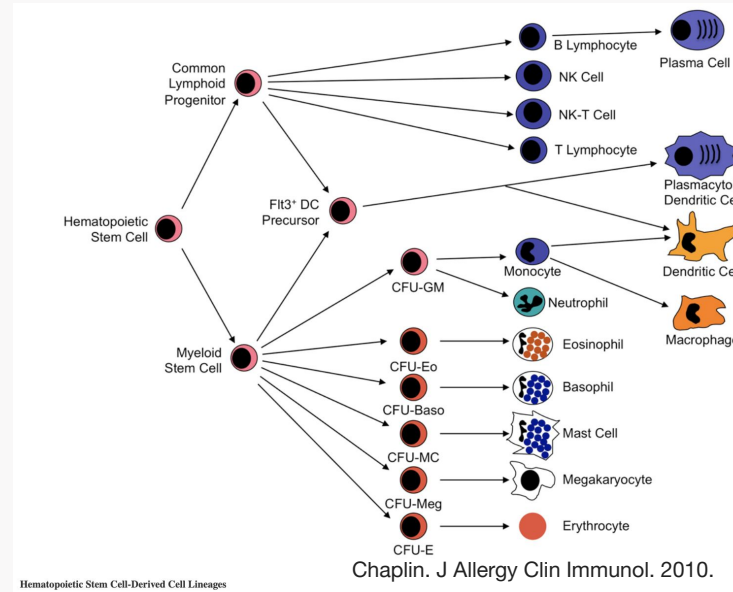
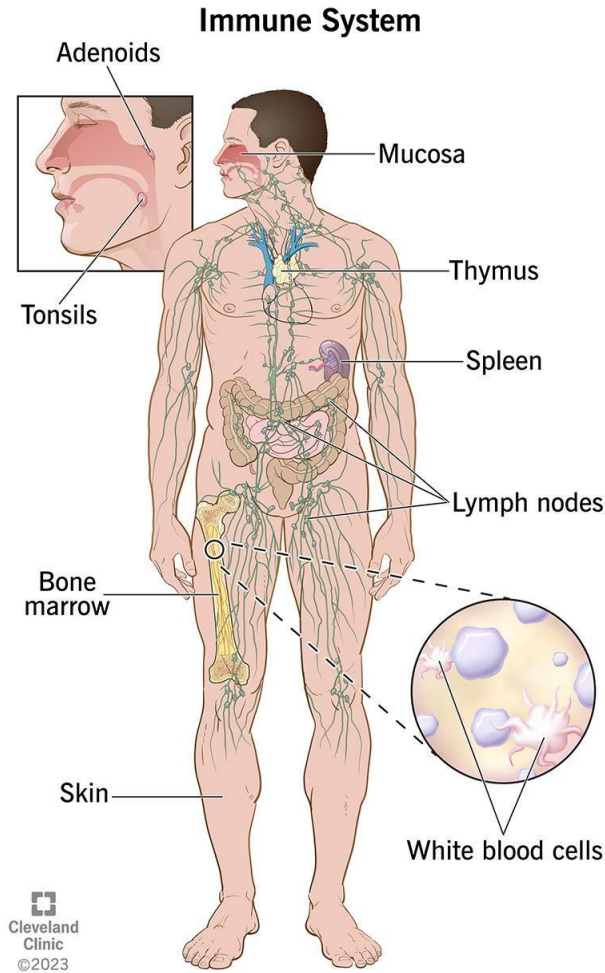
“Allen Institute is really at the **forefront of impactful and highly collaborative research.** It offers a **unique opportunity to learn** from a wide range of experts in immunology, single-cell sequencing, software development, and statistics. People here are really invested in growing your knowledge.”

- **Christian La France**
(Current Bioinformatics Analyst at AIFI and U of Oregon alumni)

“My time as a postbaccalaureate research assistant has been a fantastic learning experience - I've had the **support to grow and develop** new skills along the way, and the **team science focus** of the institute has allowed me to **explore my interests beyond my direct position.** I've always felt welcome to ask even the most basic of questions, and felt welcomed from the beginning even as I was still finding my computational footing.”

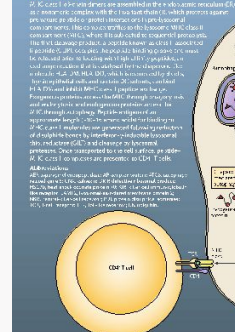
- **Jazmine Castillo**
Postbaccalaureate Researcher

The Immune System Is Complex



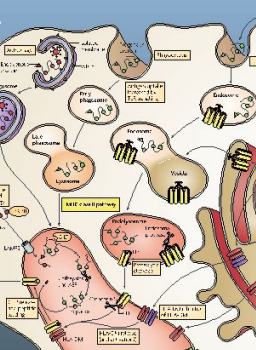
nature
REVIEWS
IMMUNOLOGY

The MHC class II pathway



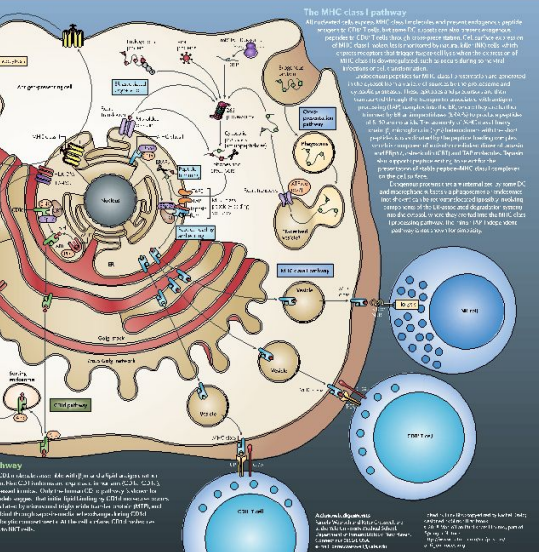
Antigen processing and presentation
 Pamela Woersch and Peter Cresswell

In early-presenting cells (APCs), such as dendritic cells (DCs) and B cells, the immunogenicity of exogenous antigens is determined by the efficiency of their presentation to T cells. This process is regulated by the expression of MHC class II molecules, which are encoded by the DQA and DQB genes. The DQA gene encodes the invariant chain (Ii), which is a non-covalent component of the MHC class II molecule. The Ii chain is cleaved during the maturation of the APC, and the remaining fragments are used for the assembly of the MHC class II molecule. The MHC class II molecule is then loaded with a peptide antigen, which is presented to a T cell. The T cell recognizes the peptide-MHC complex and initiates a response.



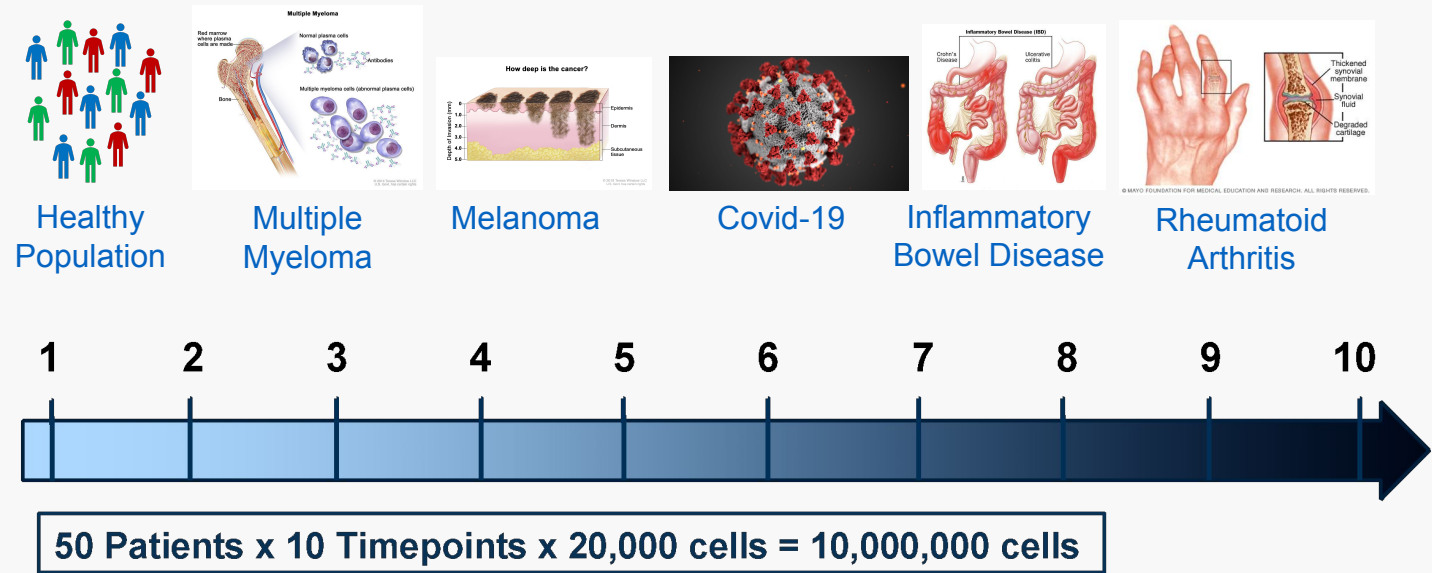
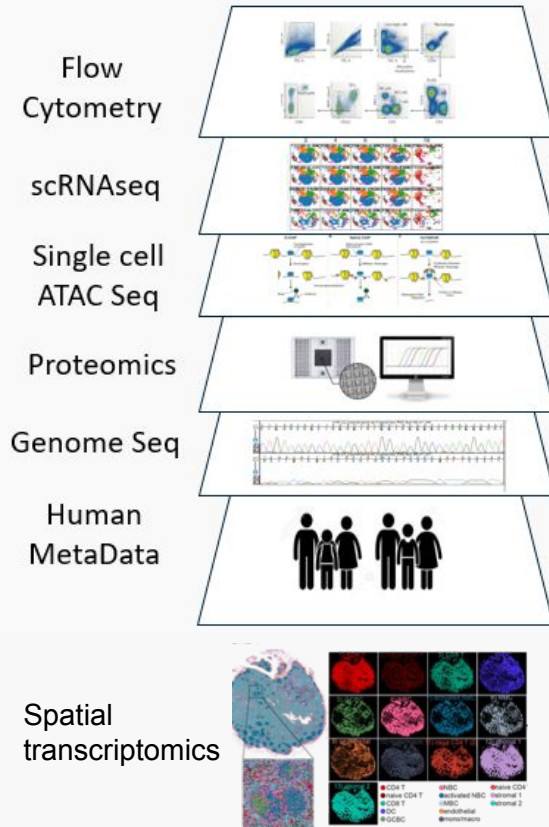
The MHC class II pathway

The MHC class II pathway is a complex process that involves the degradation of exogenous antigens into peptides, which are then loaded onto MHC class II molecules for presentation to T cells. The pathway is regulated by the expression of MHC class II molecules, which are encoded by the DQA and DQB genes. The DQA gene encodes the invariant chain (Ii), which is a non-covalent component of the MHC class II molecule. The Ii chain is cleaved during the maturation of the APC, and the remaining fragments are used for the assembly of the MHC class II molecule. The MHC class II molecule is then loaded with a peptide antigen, which is presented to a T cell. The T cell recognizes the peptide-MHC complex and initiates a response.



And Plays Essential Roles in Health & Disease

Longitudinal Multi-Omics Studies in Human & Mouse - Immune States or Diseases



Goals

- To understand disease mechanisms
- To identify disease diagnostic biomarkers and/or therapeutic targets
- To better understand biology/immunology
- To disseminate our huge amount of data to the public

A Recent Study on Immune Health and Aging



bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

HOME | SUBMIT

New Results

 Follow this preprint

Longitudinal Multi-omic Immune Profiling Reveals Age-Related Immune Cell Dynamics in Healthy Adults

 Qiuyu Gong,  Mehul Sharma,  Emma L. Kuan,  Marla C. Glass,  Aishwarya Chander,  Mansi Singh,  Lucas T. Graybuck,  Zachary J. Thomson,  Christian M. LaFrance,  Samir Rachid Zaim,  Tao Peng,  Lauren Y. Okada,  Palak C Genge,  Katherine E. Henderson, Elisabeth M. Dornisch, Erik D. Layton, Peter J. Wittig, Alexander T. Heubeck, Nelson M. Mukuka, Julian Reading, Charles R. Roll, Veronica Hernandez, Vaishnavi Parthasarathy, Tyanna J. Stuckey, Blessing Musgrove, Elliott Swanson, Cara Lord, Morgan D.A. Weiss, Cole G. Phalen, Regina R. Mettey, Kevin J. Lee, John B. Johanneson, Erin K. Kawelo, Jessica Garber, Upaasana Krishnan, Megan Smithmyer, E. John Wherry, Laura Vella, Sarah E. Henrickson, Mackenzie S. Kopp, Adam K. Savage, Lynne A. Becker, Paul Meijer, Ernest M. Coffey, Jorg J. Goronzy, Cate Speake, Thomas F. Bumol, Ananda W. Goldrath, Troy R. Torgerson, Xiao-jun Li, Peter J. Skene, Jane H. Buckner, Claire E. Gustafson

doi: <https://doi.org/10.1101/2024.09.10.612119>

This article is a preprint and has not been certified by peer review [what does this mean?].





Qiuyu
Gong



Mehul
Sharma



Peter J.
Skene



Claire E.
Gustafson

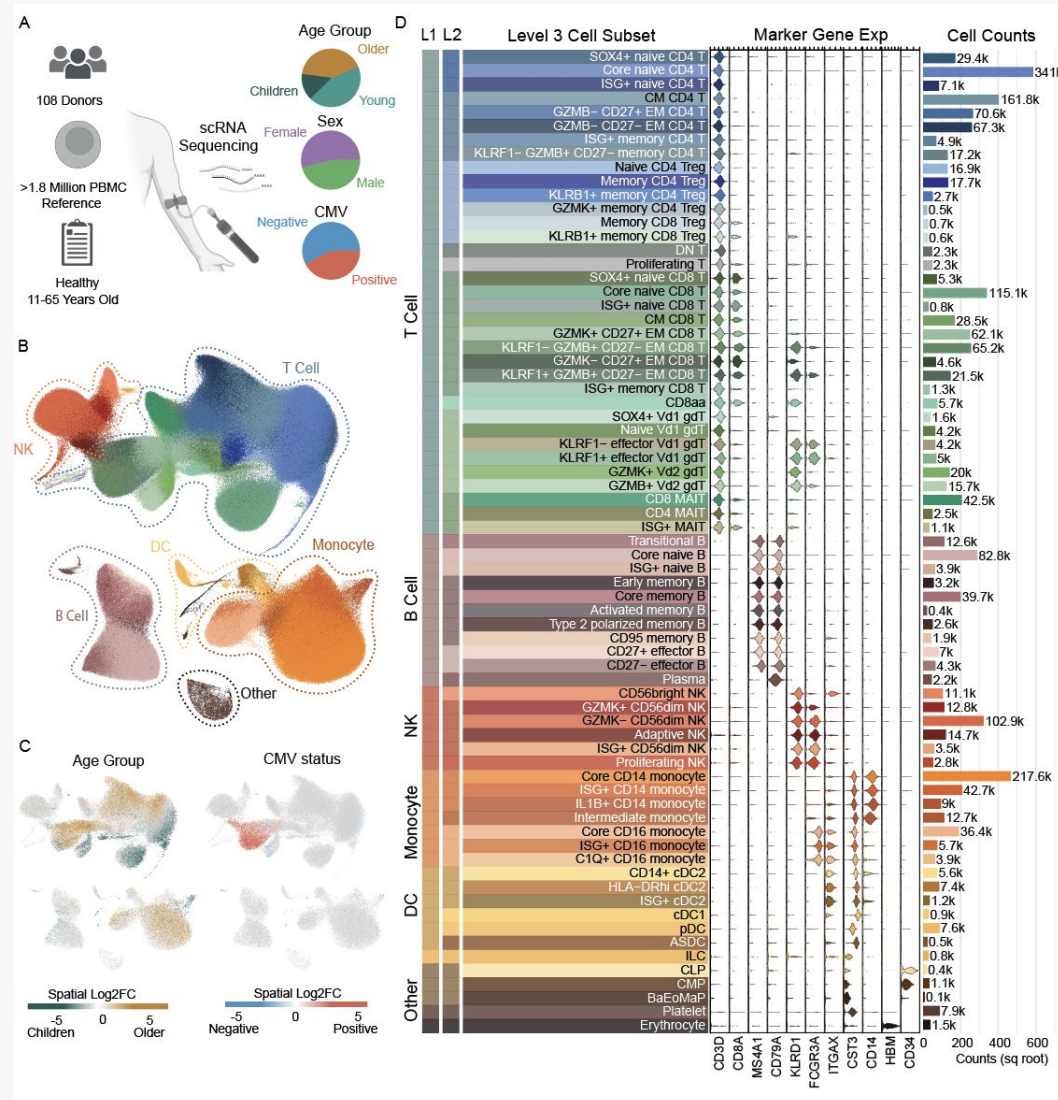


Jane
Buckner



Feature	Count
Cohort 1	
Donors	96
Samples	868
Cells (scRNA)	>13.7 million
Plasma proteomic measurements	>1.3 million
Cohort 2	
Donors	234
Samples	234
Cells (scRNA)	3.2 million
Serum proteomic measurements	>0.77 million

In-Depth Labeling of PBMCs



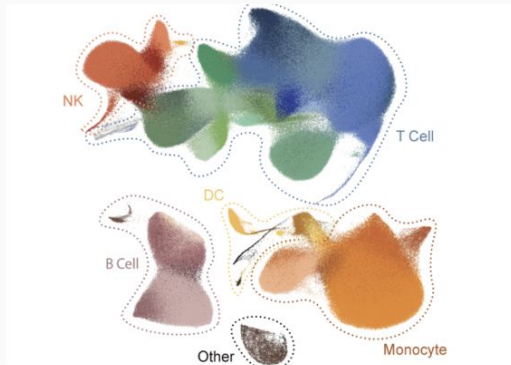
Cell Type	Subsets
T cells	35
B cells	11
Monocytes	7
Natural killer cells	6
Dendritic cells	6
Others	6
Total	71

- scRNA-seq data with >16M PBMCs
- 71 immune cell subsets

Large data releases with Data Apps

Data Apps

Large data releases by scientists at the Allen Institute for Immunology



AIFI Immune Health Atlas

Claire E Gustafson, Peter J Skene, Ananda W Goldrath, Xiao-jun Li, Troy R Torgerson, Lynne A Becker, Thomas F Bumol, Aishwarya Chander, Ernest M Coffey, Elisabeth M Dornisch, Jessica Garber, Palak C Genge, Marla Glass, Qiuyu Gong, Katherine E Henderson, Veronica Hernandez, Alexander T Heubeck, John B Johanneson, Erin K Kawelo, Mackenzie S Kopp, Upaasana Krishnan, Emma L Kuan, Christian M LaFrance, Erik D Layton, Kevin J Lee, Cara Lord, Regina R Mettey, Nelson M Makuka, Blessing Musgrove, Lauren Y Okada, Vaish Parthasarathy, Tao Peng, Cole G Phalen, Samir Rachid Zaim, Julian Reading, Charles R Roll, Mehul Sharma, Mansi Singh, Tyanna J Stuckey, Elliott Swanson, Zachary J Thomson, Morgan DA Weiss, Peter J Wittig, Jane H Buckner, Megan Smithmeyer, Cate Speake, Sarah Henrickson, Laura Vella, E John Wherry, Yousef Aggoune, Madeline Ambrose, Aldan Beaubien, James Harvey, Nicole Howard, Neelima Inala, Ed Johnson, Autumn Kelsey, Melissa Kinsey, Jessica Liang, Paul Mariz, Stark Pister, Sathya Subramanian, Vitalii Tereschenko, Anne Vetto, Paul Meyer and Lucas T Graybuck ▲

[File Set](#) [Visualization](#)

An atlas of peripheral immune cells collected from 108 healthy adults and children

- Quickly look up gene expression
- Visualize and inspect PBMCs on a UMAP
- Investigate sample metadata & clinical lab results

A Recent Study on Rheumatoid Arthritis

Systemic inflammation and lymphocyte activation precede rheumatoid arthritis

 Ziyuan He,  Marla C. Glass,  Pravina Venkatesan,  Marie L. Feser, Leander Lazaro,  Lauren Y. Okada,  Nhung T.T. Tran,  Yudong D. He,  Samir Rachid Zaim,  Christy E. Bennett,  Padmapriyadarshini Ravisankar,  Elisabeth M. Dornisch, Najeeb A. Arishi, Ashley G. Asamoah, Saman Barzideh,  Lynne A. Becker, Elizabeth A. Bemis,  Jane H. Buckner, Christopher E. Collora, Megan A. L. Criley,  M. Kristen Demoruelle, Chelsie L. Fleischer,  Jessica Garber,  Palak C. Genge,  Qiuyu Gong,  Lucas T. Graybuck,  Claire E. Gustafson, Brian C. Hattel,  Veronica Hernandez,  Alexander T. Heubeck,  Erin K. Kawelo,  Upasana Krishnan,  Emma L. Kuan, Kristine A. Kuhn, Christian M. LaFrance,  Kevin J. Lee,  Ruoxin Li,  Cara Lord, Regina R. Mettey, Laura Moss,  Blessing Musgrove, Kathryn Nguyen, Andrea Ochoa,  Vaishnavi Parthasarathy,  Mark-Phillip Pebworth, Chong Pedrick,  Tao Peng,  Cole G. Phalen,  Julian Reading,  Charles R. Roll, Jennifer A. Seifert, Marguerite D. Siedschlag,  Cate Speake, Christopher C. Striebich,  Tyanna J. Stuckey,  Elliott G. Swanson, Hideto Takada, Tylor Thai,  Zachary J. Thomson, Nguyen Trieu, Vlad Tsaltskan, Wei Wang,  Morgan D.A. Weiss, Amy Westermann,  Fan Zhang, David L. Boyle,  Ananda W. Goldrath,  Thomas F. Bumol,  Xiao-jun Li,  V. Michael Holers,  Peter J. Skene,  Adam K. Savage,  Gary S. Firestein,  Kevin D. Deane,  Troy R. Torgerson,  Mark A. Gillespie

doi: <https://doi.org/10.1101/2024.10.25.620344>

Feature	Count
<i>Cohort 1</i>	
Donors	94
Samples	141
Cells (scRNA)	>2.1 million
Cells (flow cytometry)	>200 million
Plasma proteomic measurements	>0.22 million
<i>Cohort 2</i>	
Donors	12
Samples	12
Cells (TEA-seq)	~90k



Ziyuan He



Marla Glass



Adam Savage



Troy Torgerson



Mark Gillespie



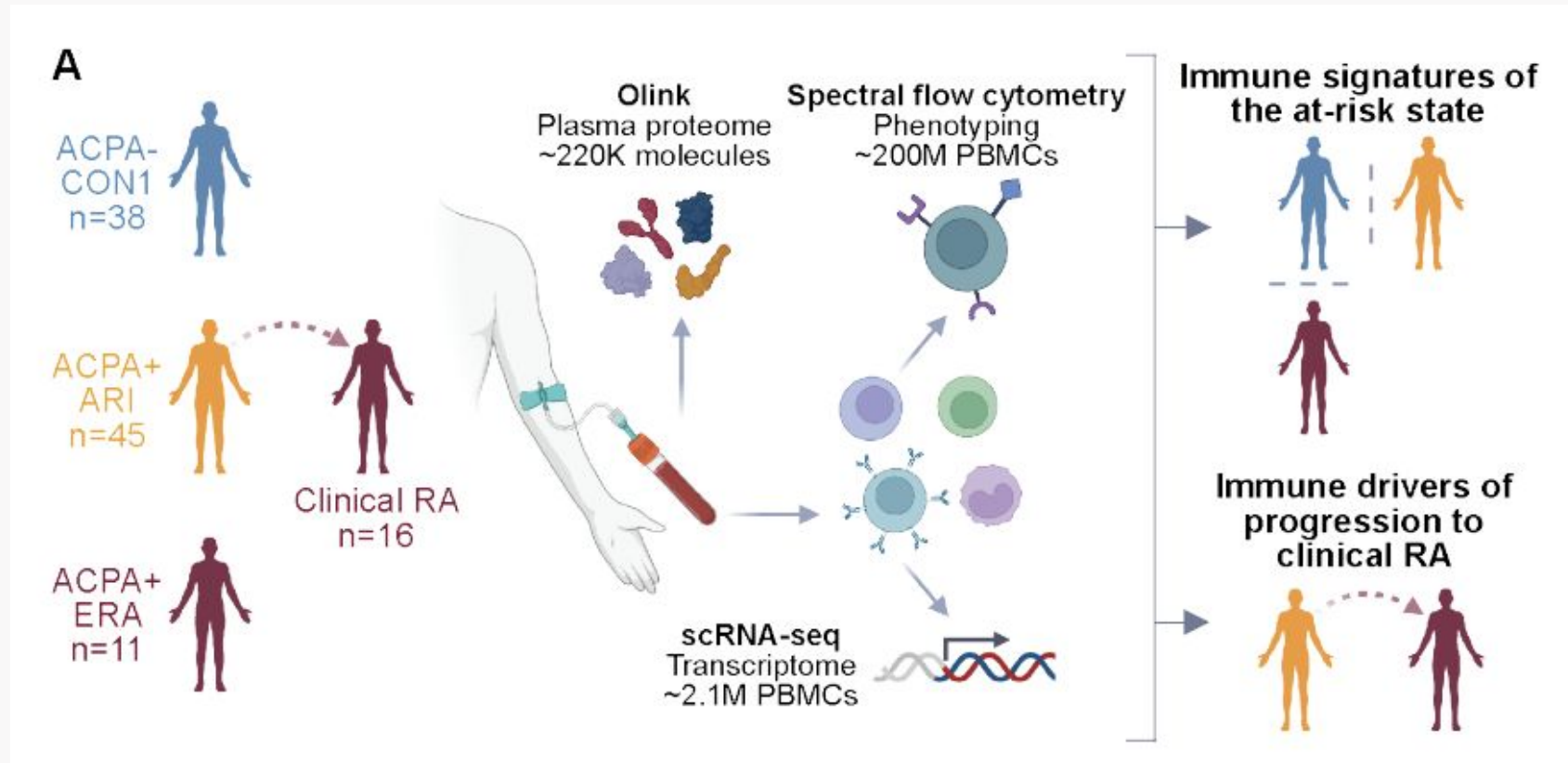
Gary Firestein



Kevin Deane

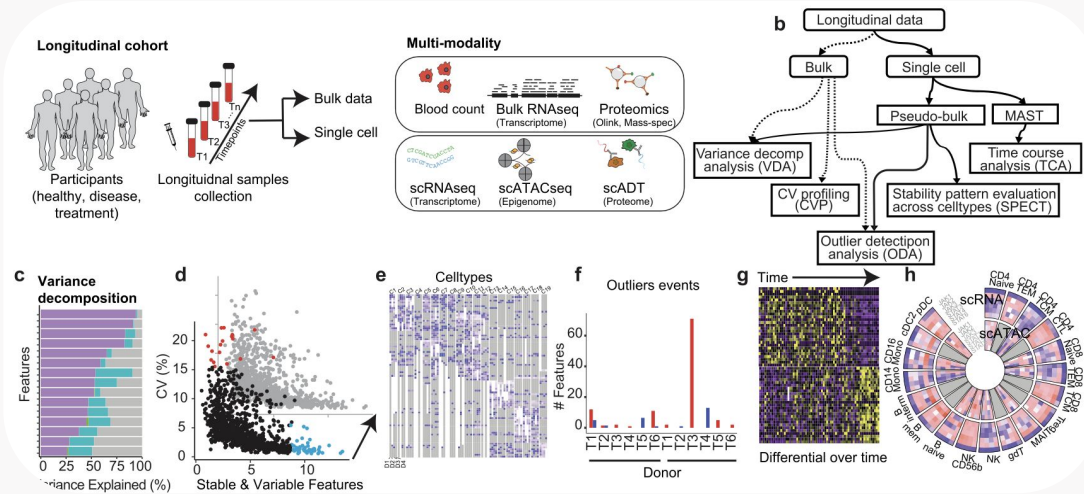


At-Risk-Individuals exhibit signs of systemic inflammation prior to the onset of clinical RA

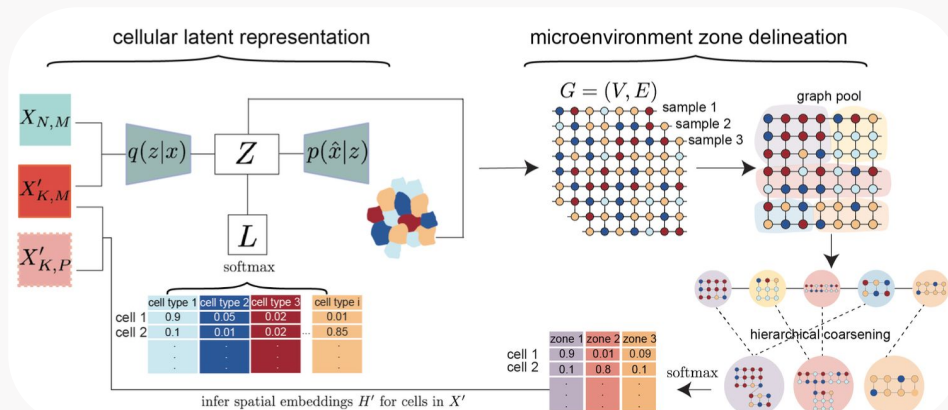


- RA is a chronic autoimmune disease
 - autoantibodies
 - joint inflammation
- Elevated ACPA are a predictor of disease
 - identify “at-risk” for developing RA
- Data
 - Hi-Res protein abundance
 - Genomics Assays
 - Multiple time points

We Develop and Disseminate Novel Methodologies



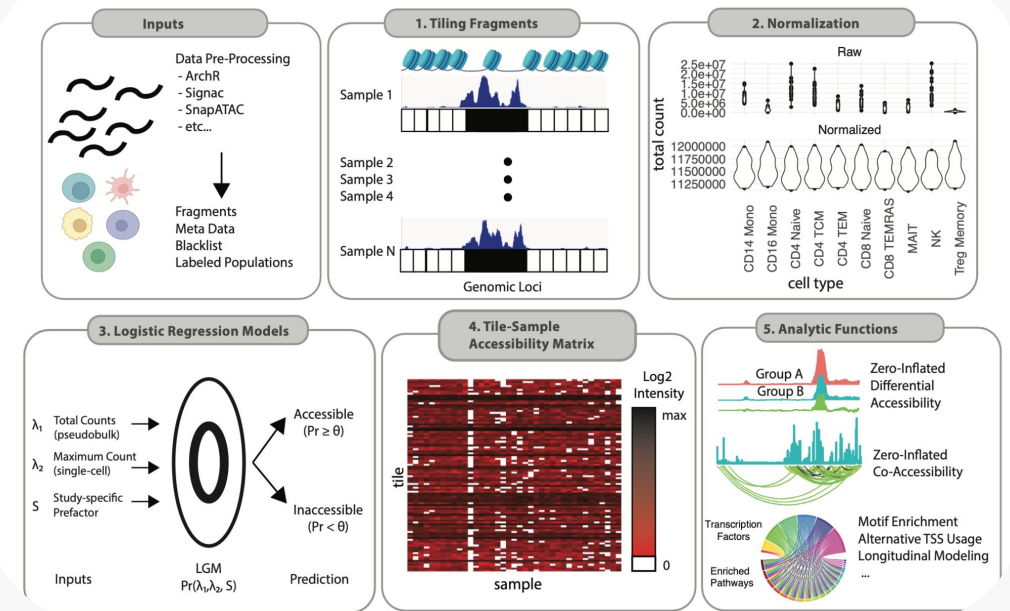
Vasaikar et al. A Comprehensive Platform to Analyze Multimodal Longitudinal Omics Data. *Nature Communications* (2023) 14:1684.



Zhao, et al. SPARROW reveals cell states and functions influenced by microenvironment zones in complex tissues.

(Now accepted at Cell Systems)

Figure 1. General workflow of MOCHA.

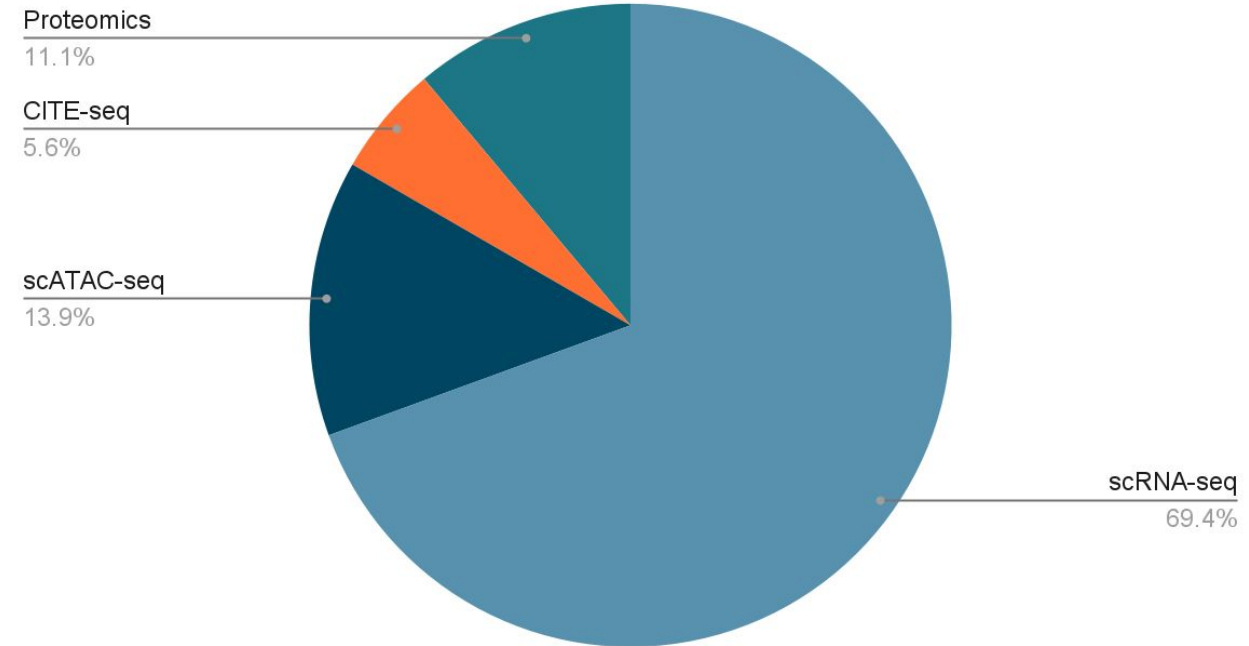


Rachid Zaim & Pebworth, et al. MOCHA: advanced statistical modeling of scATAC-seq data enables functional genomic inference in large human disease cohorts. *Nature Communications* (2024) 15: 6828.

Wealth of Data Generated

- Human immune states or immune diseases studied: **18**
- Human participants studied: **2422**
- Human samples analyzed: **4699**
- Flow cytometry: 3672 (~4 trillion cells)
- Spatial Biology (CellDive, Xenium, Visium, scRNA): 252
- Human datasets submitted to GEO/dbGaP:
 - **12** studies, **1896** samples

Number of Cells/Measurements



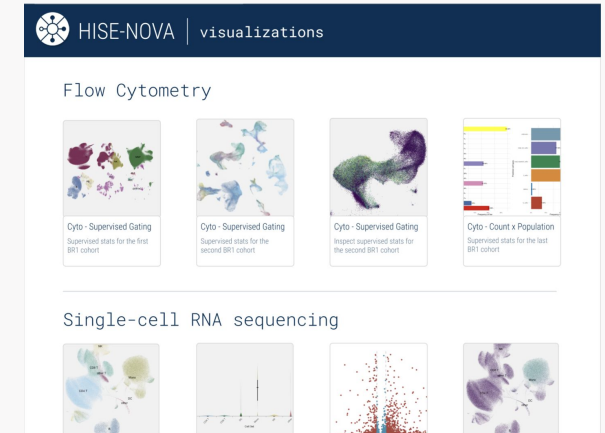
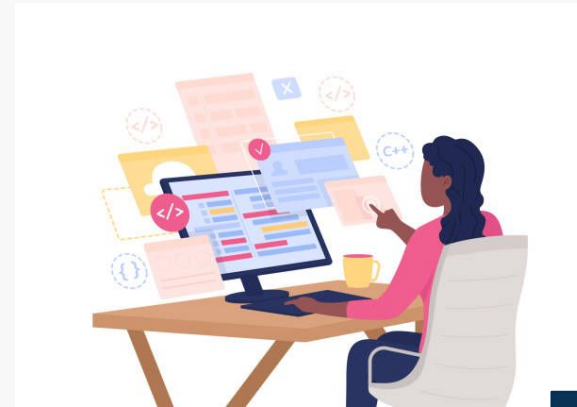
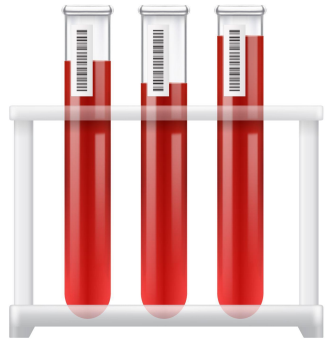
Human Immune System Explorer (HISE)



Paul Meijer

A platform with tools enabling scientists to work with large, complex data

Data Collection → Analysis Pipelines → Secondary Analysis → Results



We Are Actively Working on New Frontiers

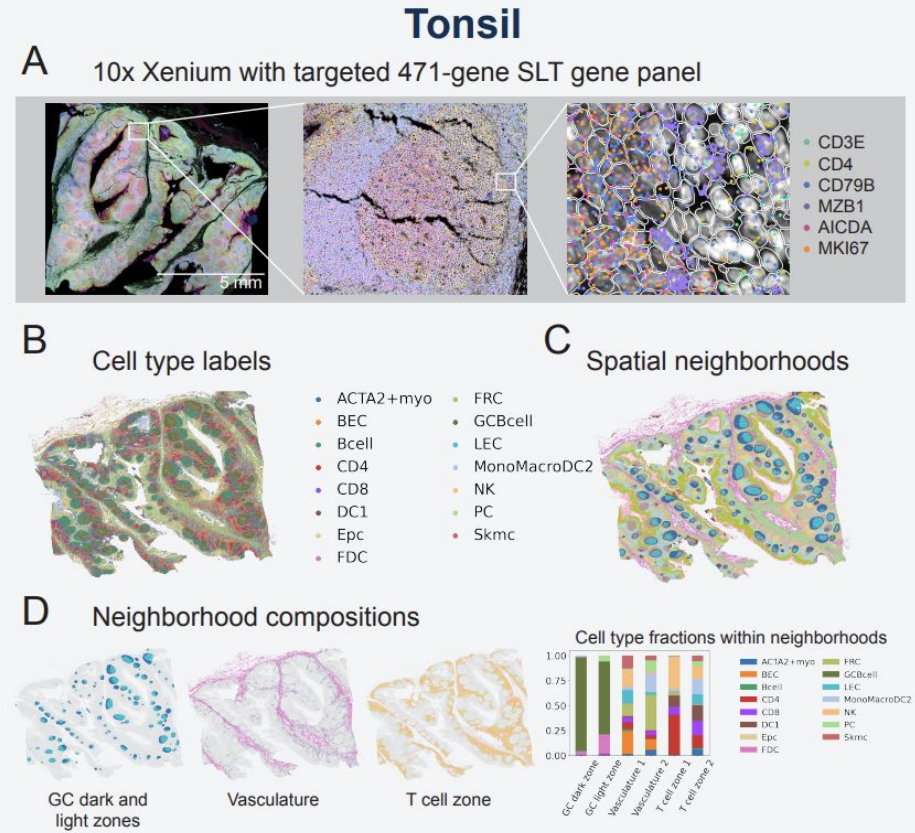
What is spatial biology?

- Spatial biology is the study of how cells are organized within a tissue and how they interact/behave in this context.
- Combination of microscopy and scRNA-seq, opening up totally new questions that we can ask

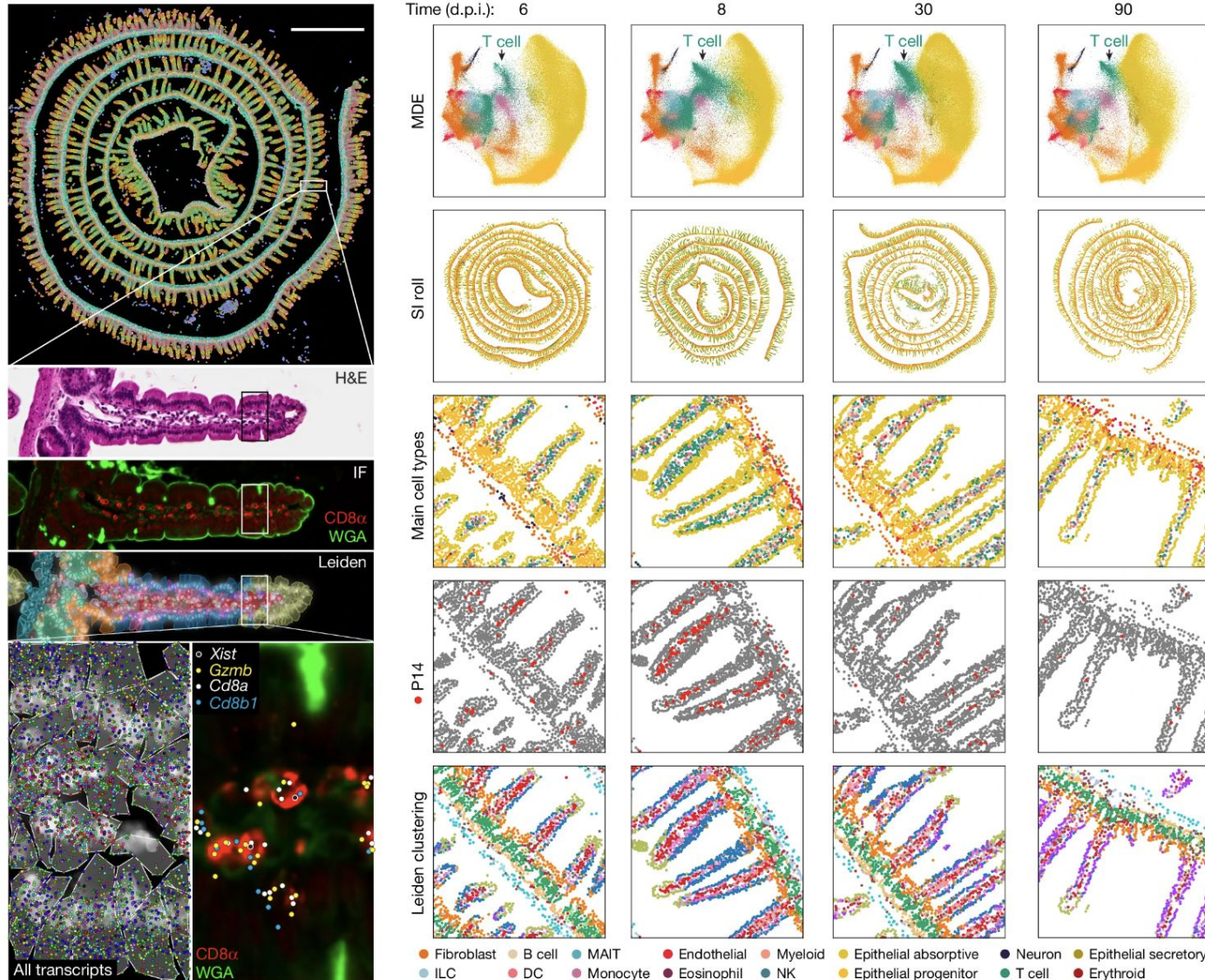
Bioinformatics for Spatial Biology

- Develop and implement data analysis pipelines for analysis across different tissues in human & mouse
 - SLT - LN and tonsil
 - Gut
 - Lung
 - Bone marrow

A spatial analysis framework for secondary lymphoid tissues (SLT)



Beautiful spatial data of the gut



- Spatial transcriptomics data of the small intestine
- Capture cellular locations along three anatomical axes of the SI
- Visualize the spatiotemporal distribution of cell types and gene expression
- T cell location and functional state are fundamentally intertwined

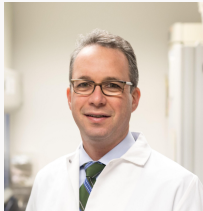
Our Collaborators



Evan Newell



Phil Greenberg



Damian Green



Julie McElrath



Julie Czartoski



Raphael Gottardo



Mariano Gabbito



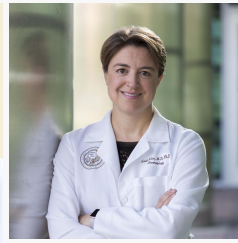
Michael Holers



Kevin Deane



K. Demoruelle



Kristine Kuhne



Jane Buckner



Cate Speake



John Wherry



A. Huang



S. Apostolidis



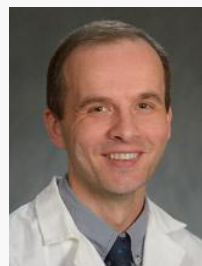
L. Vella



M. Bewtra



S. Henrickson



V. Tomov



Gary Firestein



David Boyle

CompBio Operations Team



Ananda Goldrath, Ph.D.
Executive Vice President,
Director of Allen Institute for
Immunology



Lynne A. Becker, Ph.D.
Associate Director of Scientific
Program Management



Ernie Coffey, MBA
Executive Director, Scientific
Operations



Xiaojun Li, Ph.D.
Director, Informatics &
Computational Biology



Alina Ott, Ph.D.
Manager, Bioinformatics
Operations



Paul Meijer, Ph.D.
Director, Software
Development, Database, and
Pipelines



Peter Skene, Ph.D.
Director, High Resolution
Translational Immunology



**Troy Torgerson, M.D.,
Ph.D**
Director, Experimental
Immunology

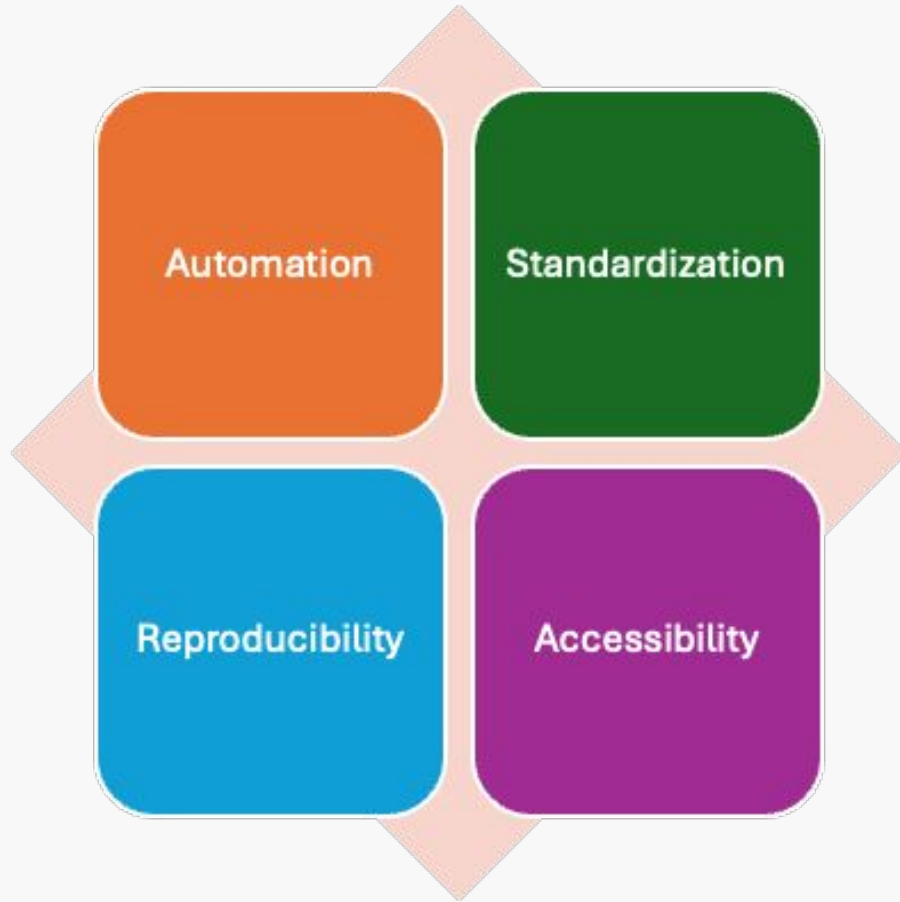


Aishwarya Gogate, M.S.
Sr. Bioinformatics Analyst



Upaasana Krishnan, M.S.
Bioinformatics Analyst I

CompBio Ops Vision



Objectives:

- Uphold Data Integrity
- Accelerate Insights Through Automation and Standardization
- Streamline Data Access
- Bridge Knowledge Gaps

Objective: Uphold Data Integrity

Generate Data



Ingest Data into HISE



Store files



Ingest



Automated QC and Processing

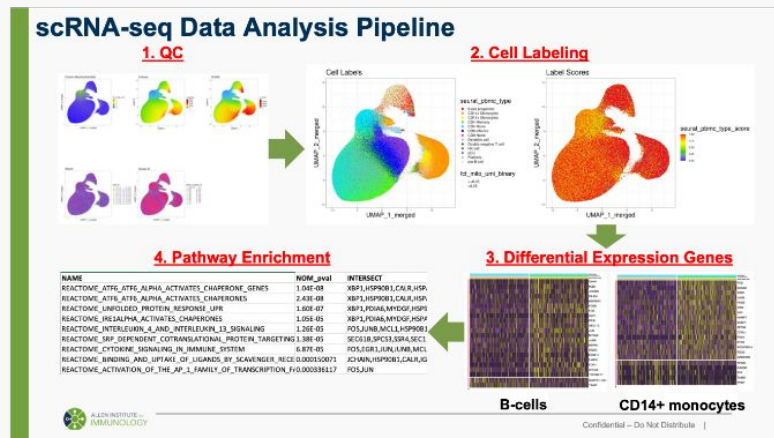
QC Reports

Review
QC

Data
searchable in
HISE

Data Integrity is Critical for Downstream Applications

Analysis



Open Science



Publication

Journal of Translational Medicine

CryoSCAPE: Scalable immune profiling using cryopreserved whole blood for multi-omic single cell and functional assays
Jan 03, 2025

Alexander T. Heubeck, Cole Phalen, Neel Kaul, Peter J. Wittig, Jessica Garber, Morgan Weiss, Palak C. Genge, Zachary Thomson, Claire...

[READ MORE →](#)

Nature Communications

MOCHA's advanced statistical modeling of scATAC-seq data enables functional genomic inference in large human cohorts
Aug 09, 2024

Samir Rachid Zaim, Mark-Phillip Pebworth, Imran McGrath, Lauren Okada, Morgan Weiss, Julian Reading, Julie L. Czartoski, Troy R....

[READ MORE →](#)

Journal of Immunological Methods

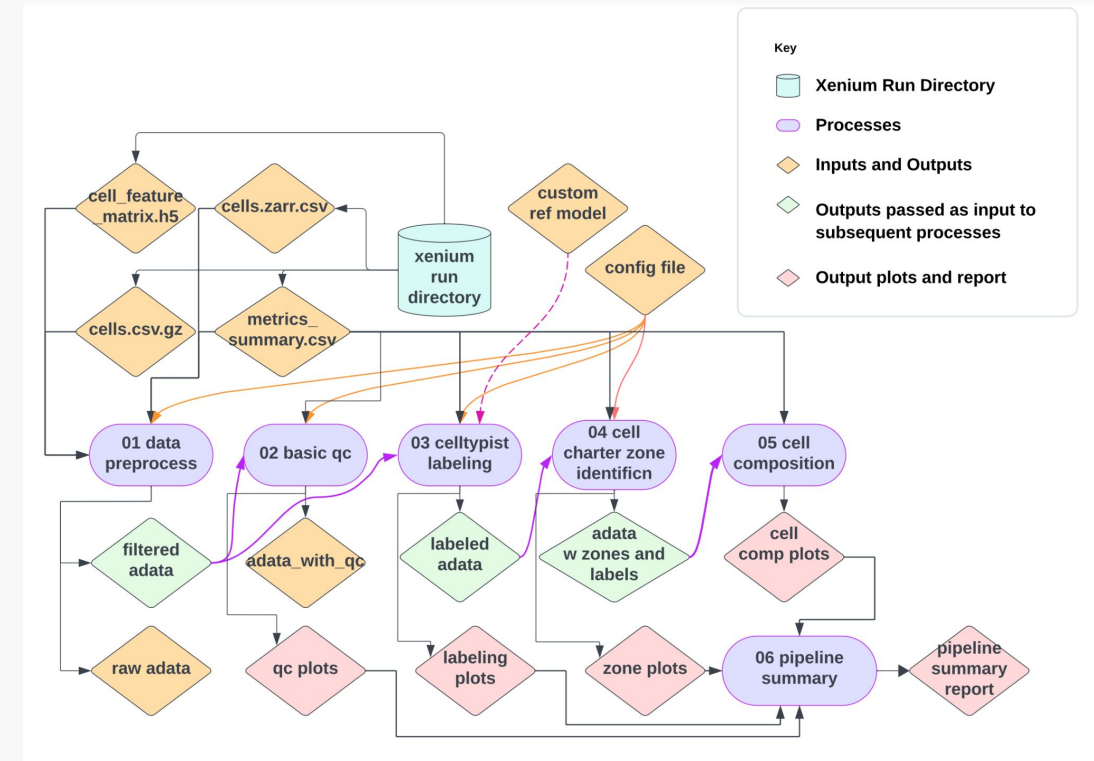
Development of a high-throughput image cytometric screening method as a research tool for immunophenotypic characterization of patient samples from clinical studies
Jan 01, 2024

Samir Patel, James I. McDonald, Hamza Mohammed, Vaishnavi Parthasarathy, Veronica Hernandez, Tyanna Stuckey, Allen H. Lin,...

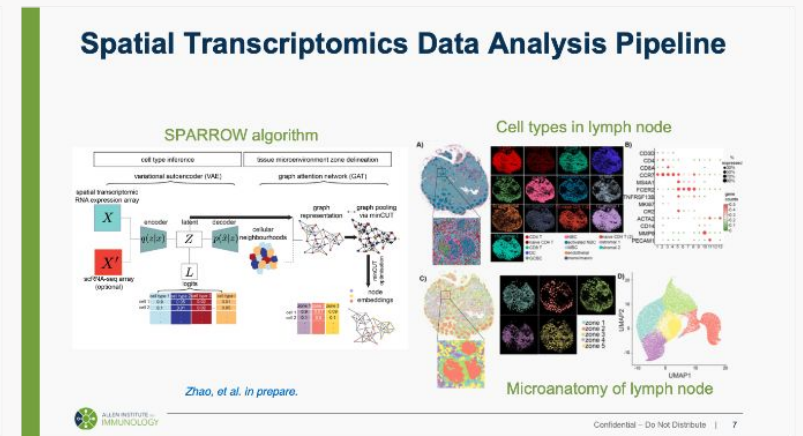
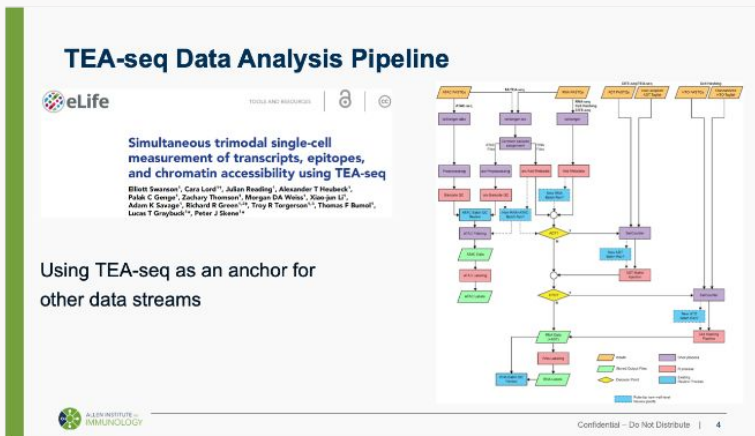
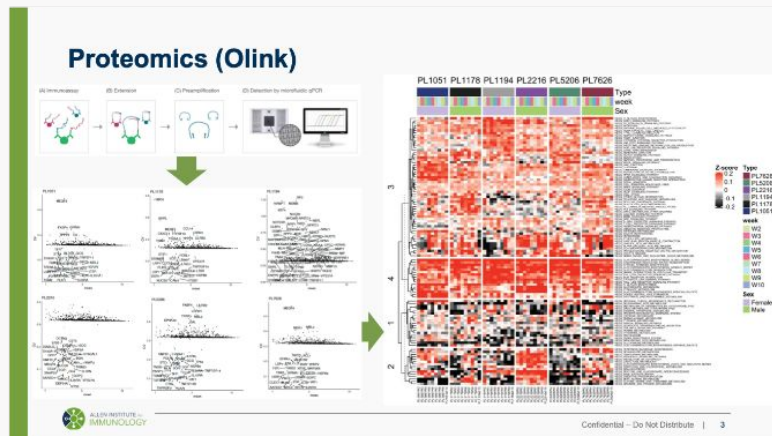
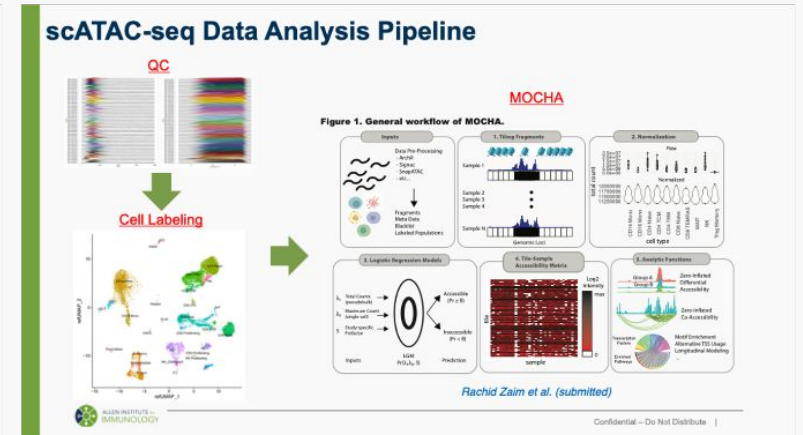
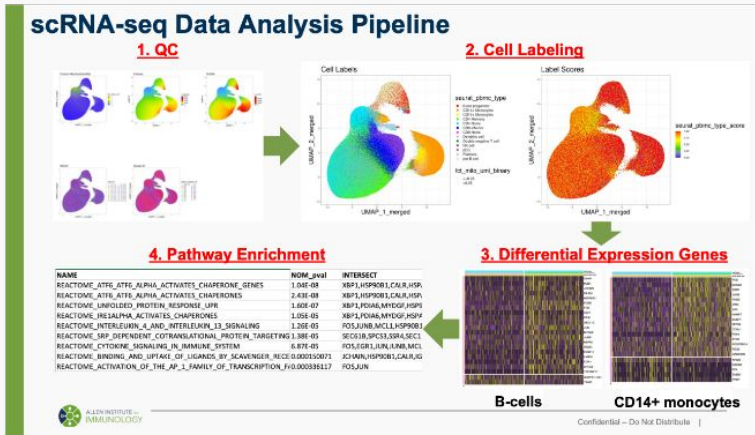
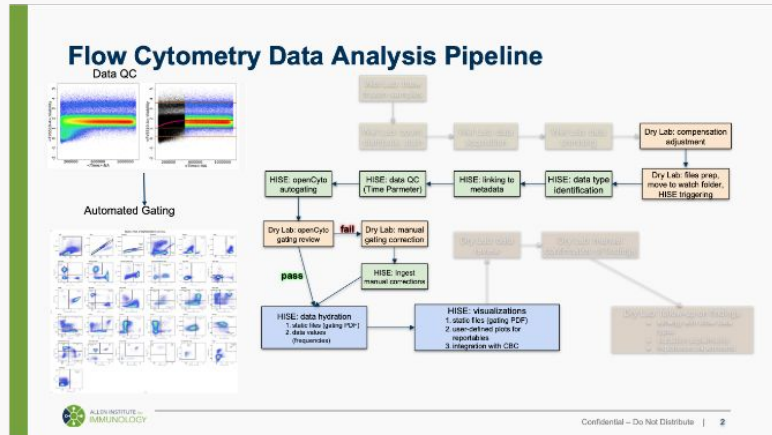
[READ MORE →](#)

Objective: Accelerate Insights Through Automation and Standardization

- Implementation of new pipelines
- Pipeline updates to expand scope of automation, increase speed/efficiency, and incorporate the latest tools
- Create templates, functions, and/or packages for common analyses

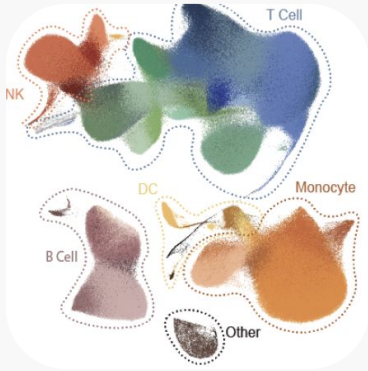


Maintain and Enhance Pipelines



Potential projects for the intern

Engage in deep biological interrogation and computational tool development



Revamp single-cell data processing pipelines

- Improve and standardize cell type labeling methods
- Improve upon metrics captured within QC reports
- ***What you will learn: Work on new data modalities, identify anomalies in seq data, build robust pipelines, communicate findings***



Assist in implementing a new spatial (Xenium) data processing pipeline

- Build data dashboards to visually track QC metrics for spatial data over time
- This will assist wet lab scientists in making the platform more efficient
- ***What you will learn: End user survey, clean coding, code review, engage in meaningful scientific discussions, literature review***



Relevant data and links:
<https://bit.ly/allen-gia-2025>

THANK YOU

We wish to thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support.



ALLEN INSTITUTE *for*
IMMUNOLOGY