

Integrated analysis of single-cell data across technologies, modalities, and perturbations









Dr. Rahul Satija

Associate Professor of Biology, Genomics & Systems Biology, NYU

Core Faculty Member, NYGC

Associate Faculty, Institute for Systems Genetics,
NYU Langone Medical Center





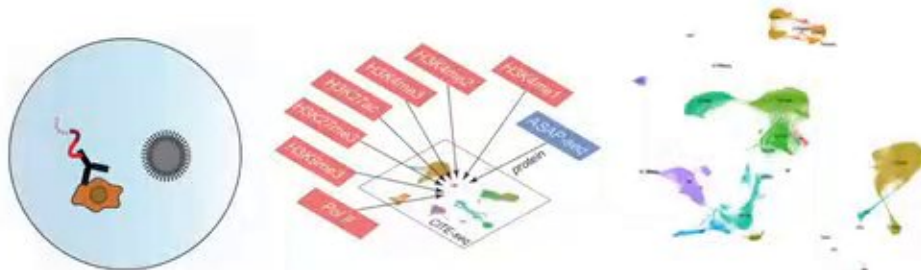








Integrative analysis of single-cell data across modalities and technologies



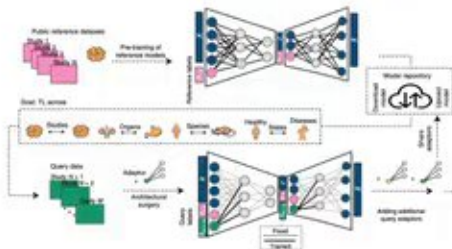
Rahul Satija
Core Faculty Member, New York Genome Center
Associate Professor, Biology, New York University

Reference 'maps' for single-cell genomics



OPEN Mapping single-cell data to reference atlases by transfer learning

Mohammad Lotfollahi^{1,2}, Mohsen Naghipourfar^{1,2}, Malte D. Luecken^{1,2}, Martin Khajavi¹, Maren Büttner¹, Marco Wagenstetter¹, Žiga Avsec^{1,2}, Adam Gayoso^{1,2}, Nir Yosef^{1,2,3,4,5}, Marta Interlandi¹, Sergei Rybakov¹, Alexander V. Misharin^{1,2} and Fabian J. Theis^{1,2}



Lotfollahi et al, NBT 2021 (Theis Lab)

azimuth.hubmapconsortium.org



Azimuth is a web application that helps an untrained clinician or scientist to automate the processing, storage, and interpretation of a new single-cell RNA-seq experiment. Azimuth leverages a reference-based mapping pipeline that maps a batch of cells to a reference atlas of gene expression in single cells, and performs downstream analysis, including cell identity and differential expression. Azimuth can be installed on a laptop, and is available for reference-based analysis.

References

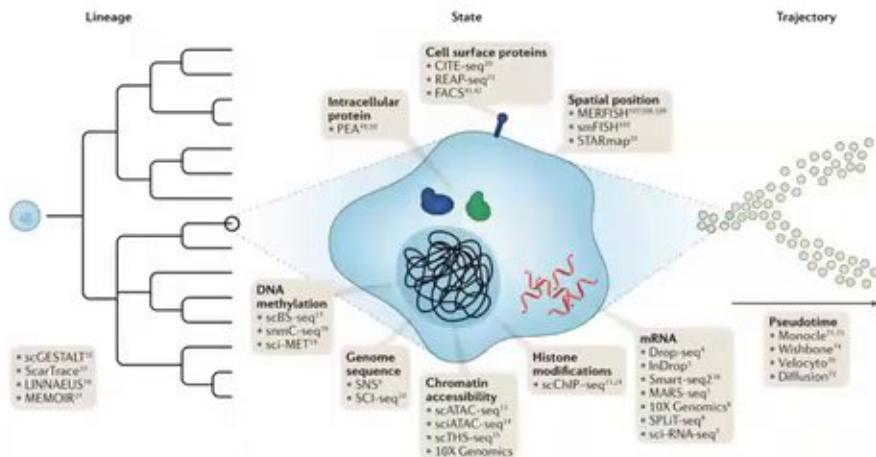


Hao et al, Cell 2021 (Satija Lab)





Multimodal cellular analysis



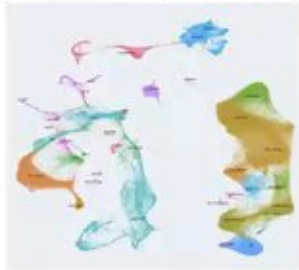
Stuart and Satija, Nature Reviews Genetics, 2020

Goal: fill in the missing gaps (modalities) for cell types in the human body

Bridge Integration via Dictionary Learning

- How can we map datasets when **they measure different things**
- Example: mapping scATAC-seq (peaks) to scRNA-seq (genes)

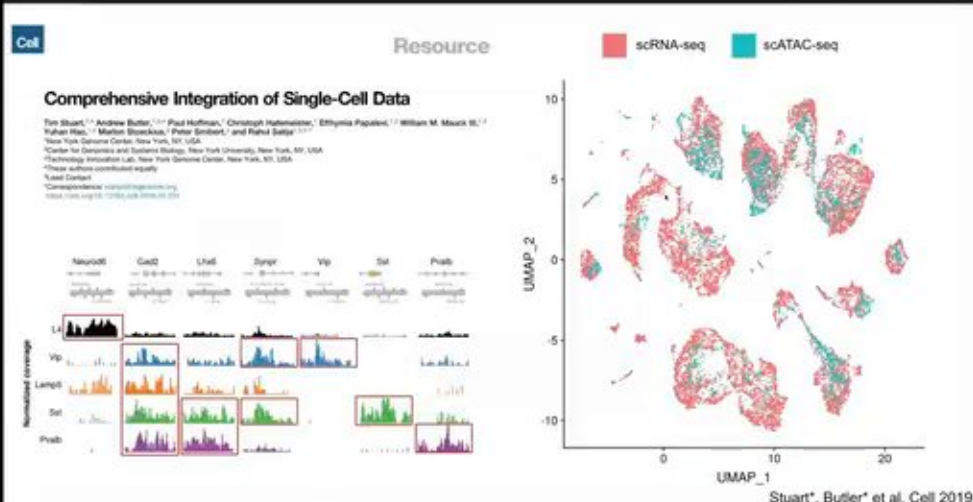
Azimuth scRNA reference
(bone marrow; 320,000 cells)



ATAC-seq query dataset
(bone marrow; 35,000 cells)



Data Integration across Modalities



Integration of ATAC-seq using 'gene activity scores'



Dictionary Learning

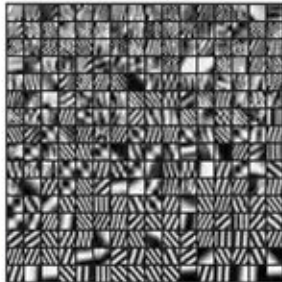
Dictionary learning develops a set (dictionary) of representative elements from the input

Each data point can be represented as a weighted sum of the dictionary elements.

Noisy image



'Dictionary elements'



Denoised image

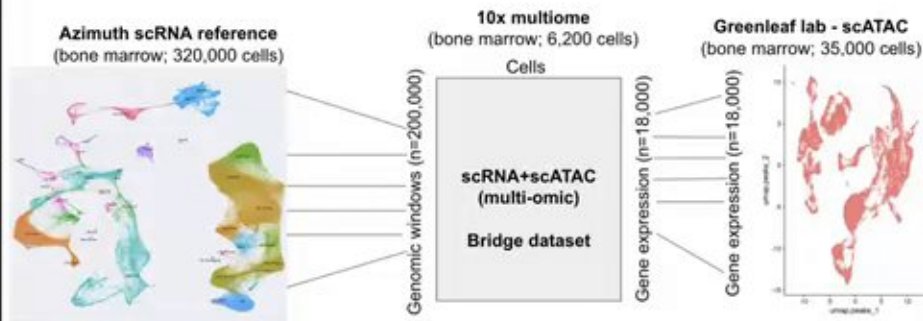


Elad and Iharon, IEEE Transactions on Image Processing, 2006

Dictionary elements can be learned 'within' or 'across' datasets

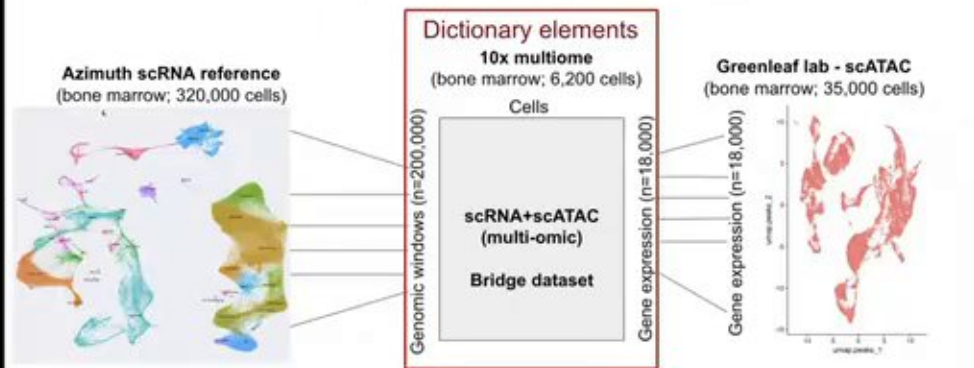
Bridge Integration via Dictionary Learning

Idea: Treat 10x multi-ome dataset (scRNA+scATAC) as a 'bridge'



Bridge Integration via Dictionary Learning

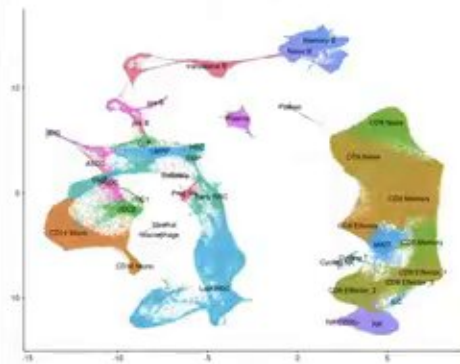
We represent **both** scRNA-seq and scATAC-seq datasets as linear combinations of multi-omic cell profiles (dictionary elements)



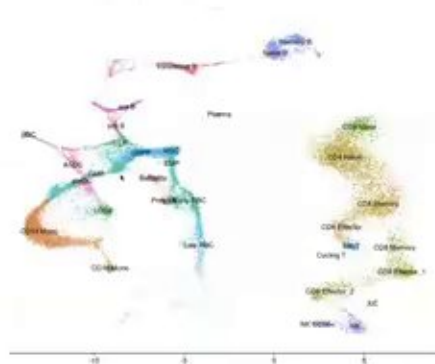
Dictionary learning maps different features into a shared space

Integrating scATAC-seq + scRNA-seq

Azimuth scRNA reference
(bone marrow; 320,000 cells)



ATAC-seq query dataset
(bone marrow; 35,000 cells)



Hao et al. Nat Biotech 2024

Bridge integration enables accurate cross-modality integration

Integrating scATAC-seq + scRNA-seq



RESEARCH ARTICLE SUMMARY

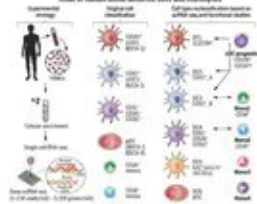
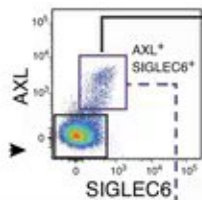
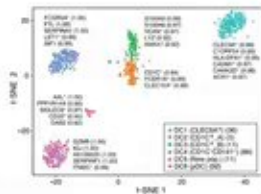
Keywords: child sexual abuse; disclosure; social support

Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors

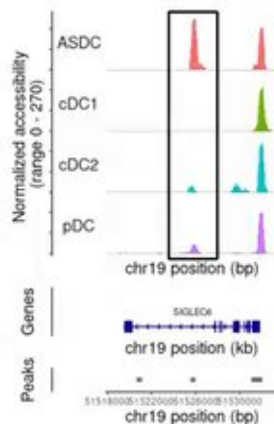
Alexandra (Jill) Viland, "I Belief Is Life," Gary Bernhardt, Wisconsin Architects,
 Kenneth Shickler, James Phillips, Margaret Griesbach, Andrew Butler, Gilbert Chang,
 Susan Lutz, Laura Jordan, David Weiss, Emily Stephenson, Carl Nelson,
 Mike Greenberg, David McManis, Andrew Ellis, Walter L. De Jager,
 Gail Rosenblatt Rosen, Andrew S. Lutz, Mitchell Smith, / John Rogers / My Signature

RESULTS AND DISCUSSION. Inactive wild (WC) and recombinant animal of multiple spontaneous mutant that give a similar role in pathogen clearing, although the mutant was again spontaneous, therefore, the difference between the two mutants can be fully understood, as these predictions have historically been different to contributions of morphology, physical properties, localization, function, developmental origins, and molecular localization of the pathogen.

subtilis cells and monocytes

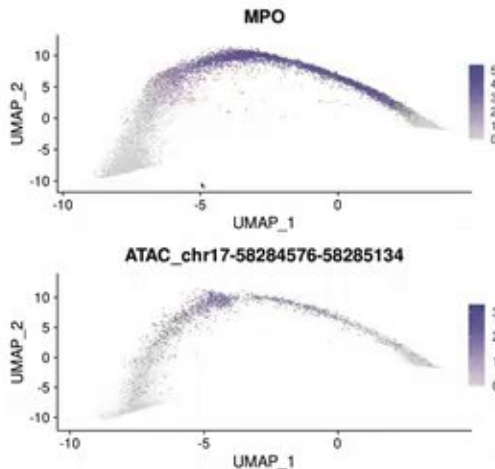
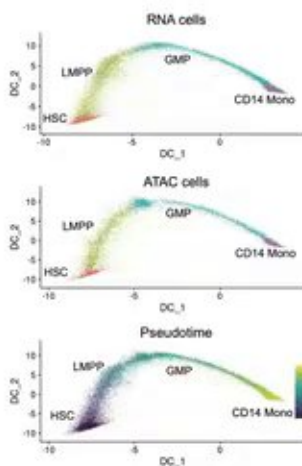
Willard, Satiya et al. *Science* 2017

scATAC-seq dataset (mapped)



Bridge integration can map extremely rare groups (0.04%)

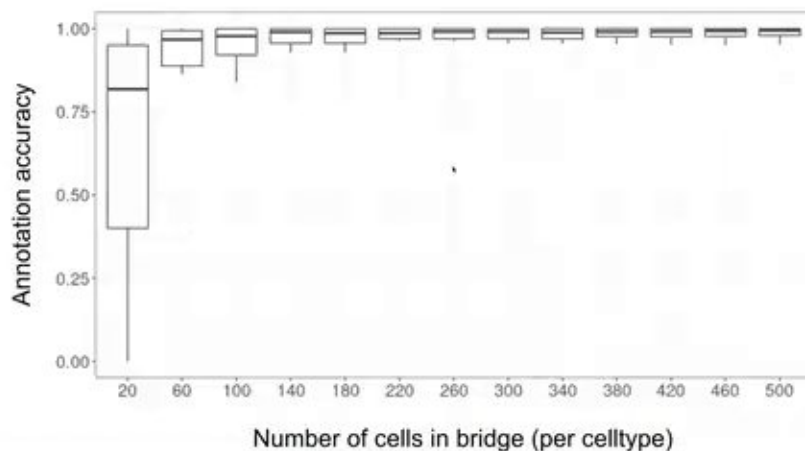
Bridge Integration via Dictionary Learning



Identifying genes whose expression 'lags' behind chromatin state



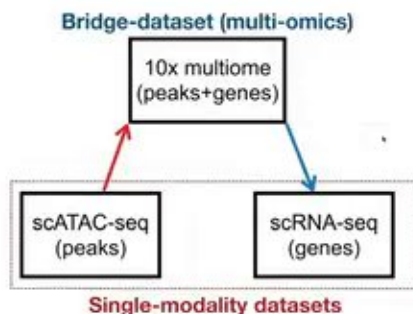
Integrating scATAC-seq + scRNA-seq



Accurate annotations when the bridge dataset is 'sufficient'



Integrating scATAC-seq + scRNA-seq



Example bridge technologies

Modality	Bridge Technology
Chromatin accessibility	10x Genomics 'multiome' kit (ATAC+RNA)
Chromatin modifications	scCUT&Tag-pro ³⁰ (chromatin mods+protein)
DNA methylation	snmCT-seq ³¹ (methylation + RNA)
Genome organization	scMethyl-HiC ³² (HiC+methylation)
Protein levels	CITE-seq ³³ , DOGMA-seq ³⁴ (RNA+ATAC+protein)
Lipid metabolism	scLipid ³⁵ (MALDI-MSI + RNA)

Bridge integration can be broadly applied to multiple modalities



Integrating scDNAm + scRNA-seq



RESEARCH

NEURAL EPIDEMIOLOGY

Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex

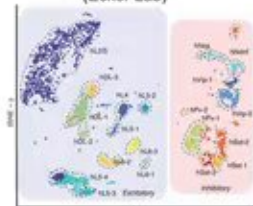
Chengnan Luo,^{1,2} Christopher J. Kwon,² Leslie Korfman,² Jinglin Zhou,^{1,2} Yiqing Jin,^{1,2} Jonathan Li,² Anna Carstens,² Jorjelle Lamm,² Joseph B. Nery,² Justin F. Sandness,² Brian Bell,² Terence J. Sejnowski,^{1,2} Timothy S. Blackton,² Eric A. Mikkelsen,^{1,2} M. Margarita Betancur,² Joseph R. Ecker^{1,2}

The mammalian brain contains diverse neuronal types, yet we lack single-cell epigenomic assays that are able to identify and characterize them. DNA methylation is a stable epigenetic mark that distinguishes cell types and marks regulatory elements. We generated ~6000 methylomes from single neuronal nuclei and used them to identify 35 mouse and 23 human neuronal subpopulations in the frontal cortex, CG and non-CG methylation exhibited cell type-specific distributions, and we identified regulatory elements with differential methylation across neuron types. Methylation signatures identified a type 6 excitatory neuron subtype and a unique human parvalbumin-expressing inhibitory neuron subtype. We observed stronger cross-species conservation of regulatory elements in inhibitory neurons than in excitatory neurons. Single-nucleus methylomes expand the atlas of brain cell types and identify regulatory elements that drive expanded brain cell diversity.

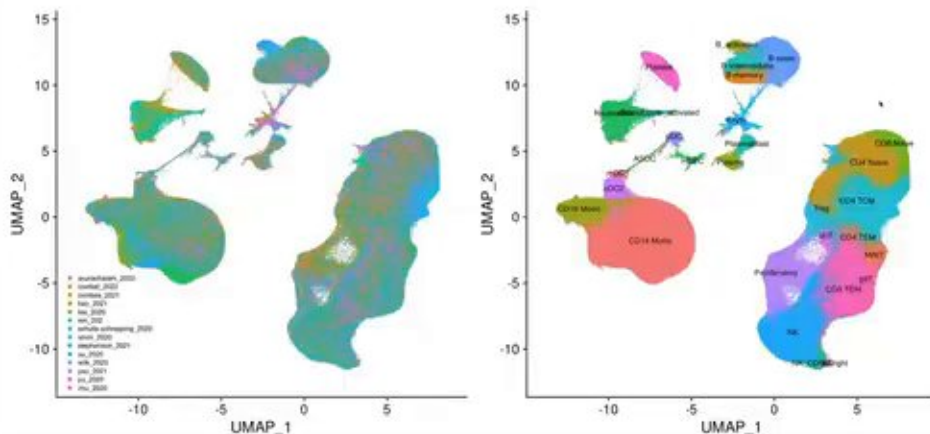
Mammalian neuron types are identified by their structure, electrophysiology, and connectivity (1). The difficulty of making traditional cellular and molecular assays in whole animal populations has prevented comprehensive analyses of brain cell types. Sequencing mRNA transcripts from single cells or nuclei has identified cell types with unique transcriptional profiles in the mouse brain (2,3) and human brain (4). However, these methods are restricted to RNA expression, which are influenced by the environment, epigenetic marks, and

DNA methylation (mC), are cell type-specific and developmentally regulated, yet stable across sub-cultures and over the life span (5-7). We showed that epigenomic profiles using single-cell DNA methylomes could enable the classification of new cell subtypes in the mammalian brain. During perinatal neurogenesis, neurons are normally reprogrammed (DNA methylation at non-CG sites (mCG) and non-CG methylation of CpG methylation (mCG) (5). Patterns of mCG and mCG at gene bodies, promoters, and enhancers are specific to neuronal types (8-10). Gene body mCG is

scDNAm dataset (Ecker Lab)



Robert Lott



Integrated dataset of 3.46M human PBMC (19 studies; 693 individuals)

Bridge Integration FAQ

Question: What are the requirements of the bridge dataset?

- Needs to be representative (20-50 cells per cell type/state)
- Can be (substantially) lower quality than reference or query

Question: What if cell types are missing in the bridge datasets

- Dictionary representation detects high reconstruction error
- Confidence scores for mapping sharply decrease

Question: Why not always process with multi-omic technologies?

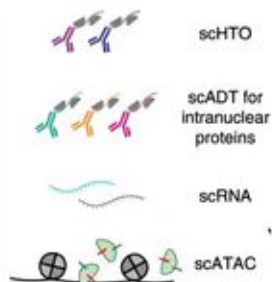
- Compatibility, data quality, and throughput
- Bridge integration enables more flexible experimental design



Multimodal profiling technologies

NEAT-seq

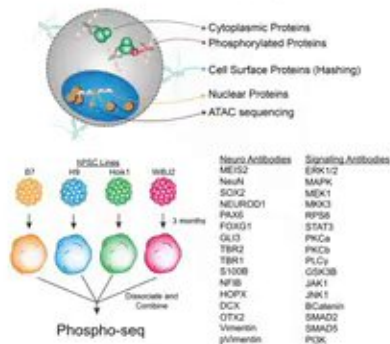
snATAC + snRNA + intra-nuclear ADT
10x multiome kit



Chen et al, Nature Methods

Phospho-seq

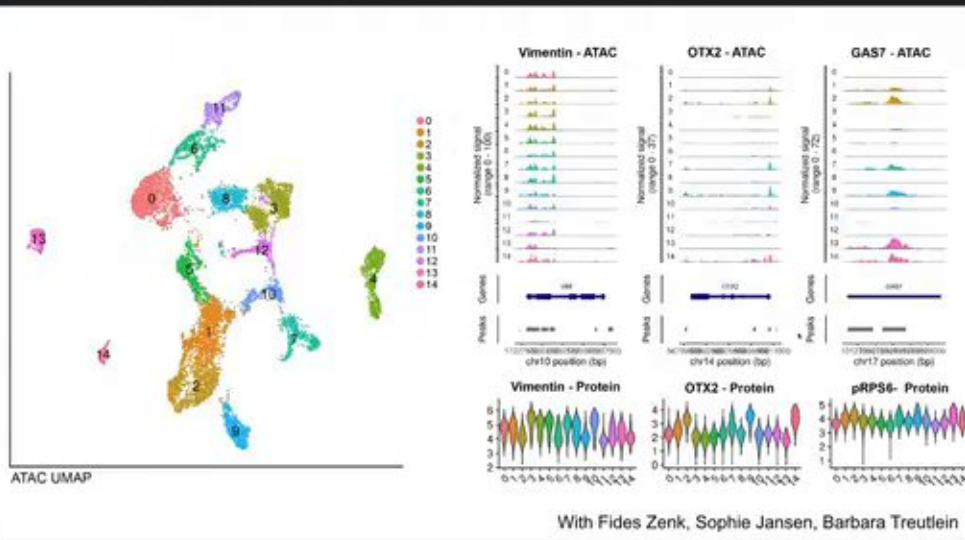
scRNA-seq + intra-cellular/nuclear ADT
10x scATAC-seq kit



Blair et al, In Revision

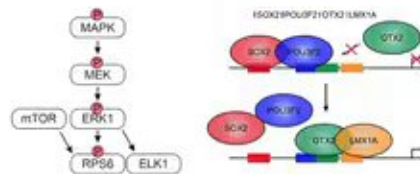
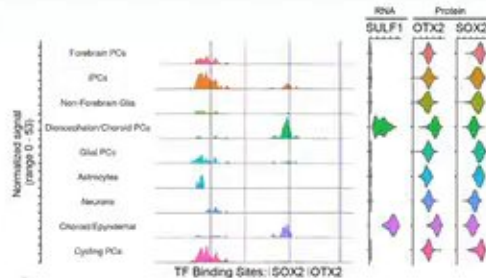
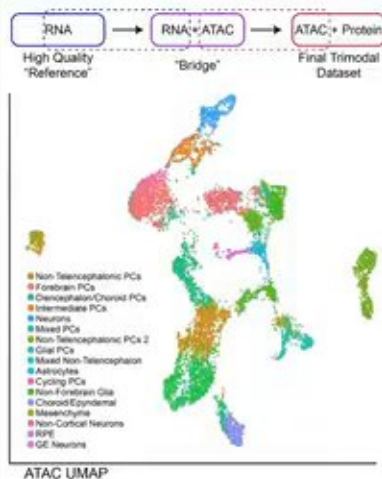


Intracellular protein profiling + scATAC



Phospho-seq datasets from human cerebral organoids

Intracellular protein profiling + scATAC



Bridging provides high-quality, integrated, multimodal profiles

COVID vaccine-induced T cell responses



Bingjie Zhang
Postdoctoral Fellow
Satija / Littman Labs



Rabi Upadhyay
Heme/Onc Fellow
Littman Lab



Dan Littman
Professor
NYU SoM / HHMI



Mark Mulligan
Professor
NYU Vaccine Center

Zhang*, Upadhyay* et al, Nat Immunology 2024



COVID vaccine-induced T cell responses

NIH RESEARCH MATTERS

June 1, 2022

T cells protect against COVID-19 in absence of antibody response

□□□□□

At a Glance

- In studies of mice, immune cells called T cells protected against COVID-19 even in the absence of antibodies targeting the SARS-CoV-2 virus.
- Mice that developed no protective T cell response may be able to help predict against current and future variants of the virus that causes COVID-19.

SCIENCE IMMUNOLOGY | VIEWPOINT

CORONAVIRUS

Understanding T cell responses to COVID-19 is essential for informing public health strategies

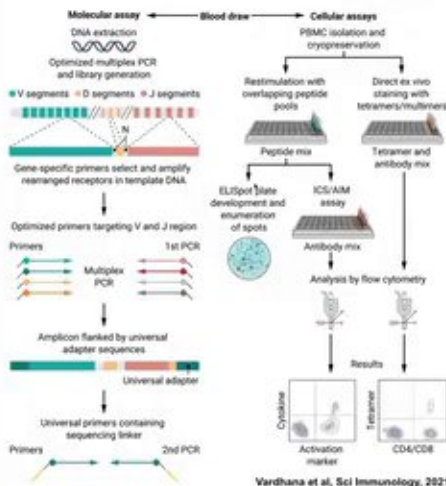
Santosha Vardhana^{1,2,3}, Lance Baldo^{4*}, William G. Morice II⁵, E. John Wherry^{6,7,8}

Durable T cell responses to SARS-CoV-2 antigens after infection or vaccination improve immune-mediated viral clearance. To date, population-based surveys of COVID-19 adaptive immunity have focused on testing for IgG antibodies that bind spike protein and/or neutralize the virus. Deployment of existing methods for measuring T cell immunity could provide a more complete profile of immune status, informing public health policies and interventions.

1. T cell contributions are required for early and durable SARS-Cov-2 protection
2. T cell responses are linked to disease severity and progression
3. New technologies are needed to identify, characterize, and track T cell responses during infection and vaccination



Assays for characterizing T cell responses



Cellular assays

- Cytokine production after antigen staining
 - Requires in-vitro restimulation
 - Low-cost and broadly applicable
- Tetramer-based assays
 - Direct ex-vivo characterization
 - Restricted for individual haplotypes

Molecular assays

- Immune repertoire analysis
 - Use of NGS provides flexibility when analyzing new variants

Existing assays each have individual strengths and limitations

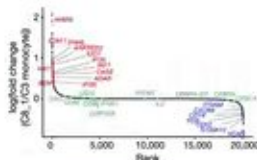
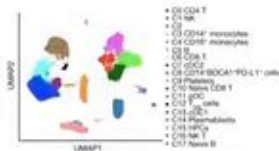


COVID vaccine-induced T cell responses



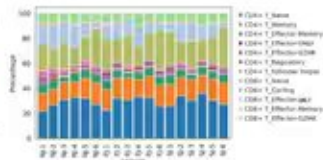
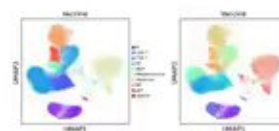
Systems vaccinology of the BNT162b2 mRNA vaccine in humans

Page 1 of 1 Page 2 of 1 Page 3 of 1	Page 1 of 1 Page 2 of 1 Page 3 of 1
--	--

medRxiv preprint doi: <https://doi.org/10.1101/2020.04.29.20076100>; this version posted May 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Single-cell transcriptomic atlas of individuals receiving inactivated COVID-19 vaccines reveals distinct immunological responses between vaccine and natural SARS-CoV-2 infection

© 2008 Wiley Periodicals, Inc. *Journal of the Philosophy of Education Society of Great Britain* 38(1): 1–12

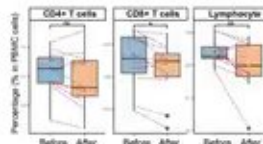
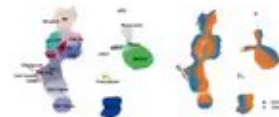


Copyright © 2007 John Wiley & Sons, Ltd.

Cell Discovery

ARTICLE Open Access

Comprehensive investigations revealed consistent pathophysiological alterations after vaccination with COVID-19 vaccines

[illegible]

Previous scRNA-seq studies have not identified antigen-specific T cells

Characterizing molecular responses to BNT162b2

Initial dataset (6 individuals)



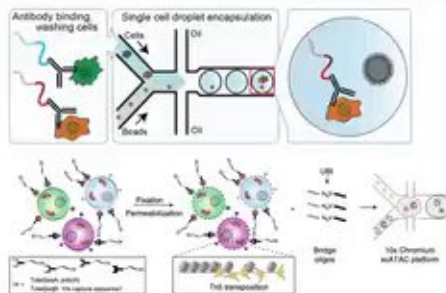
Modalities (initial dataset)

3' RNA (CITE-seq)

ATAC (ASAP-seq)

175 Surface protein (CITE/ASAP-seq)

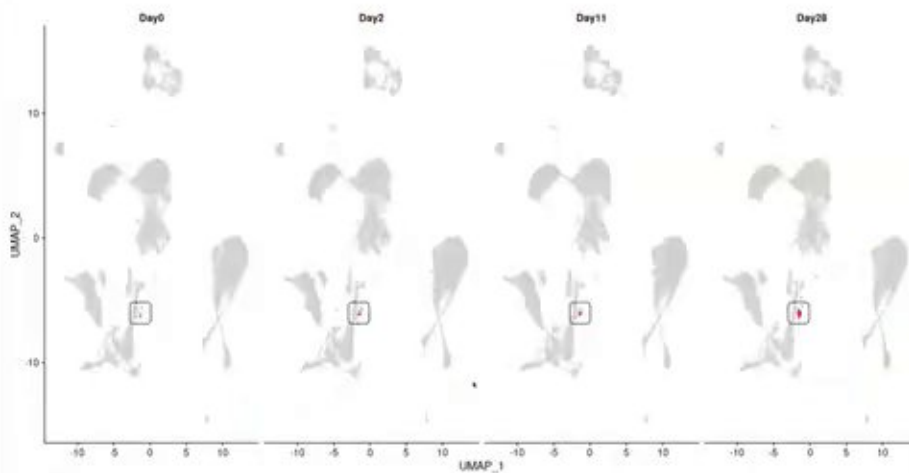
Total: 195,790 single cells



Rafael Lertze

CITE-seq data reveals high-resolution immune subpopulations

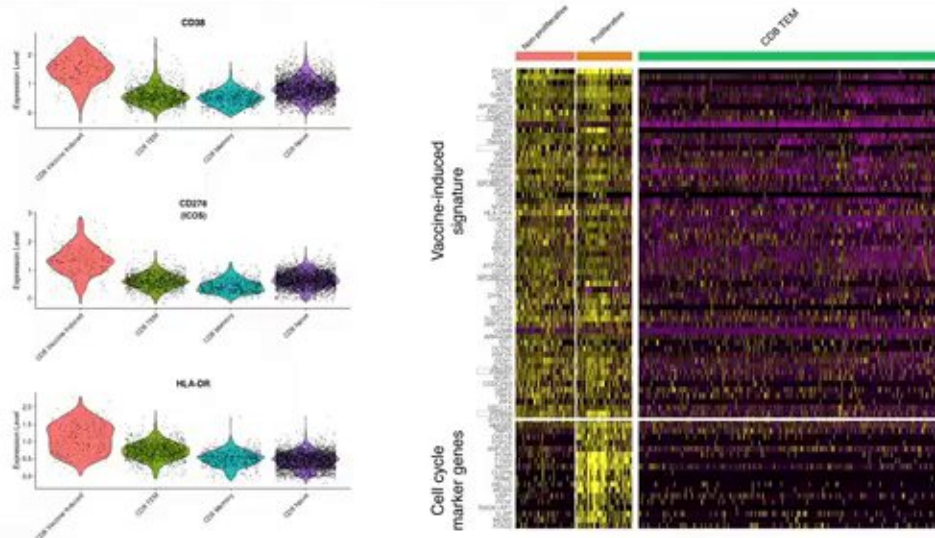
Vaccine-induced CD8 T cells



CITE-seq data reveals high-resolution immune subpopulations



Vaccine-induced CD8 T cells



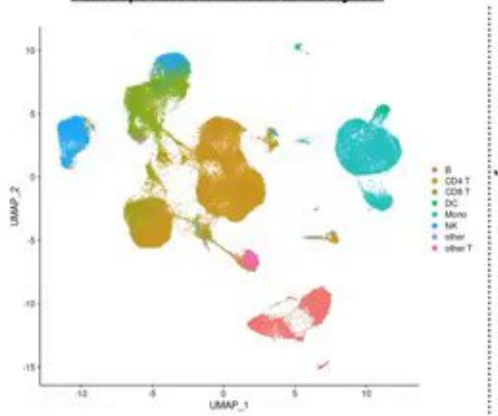
RNA and protein molecular signatures of vaccine-induced CD8 cells



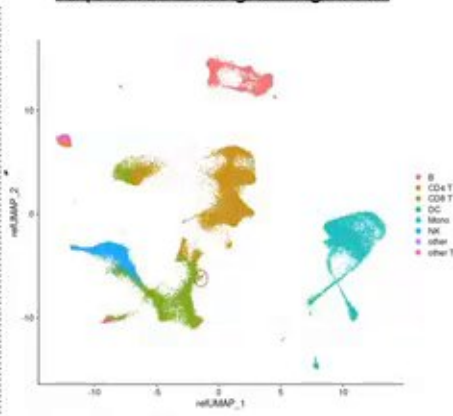
Vaccine-induced CD8 T cells

ASAP-seq dataset (chromatin accessibility)

Unsupervised ATAC analysis



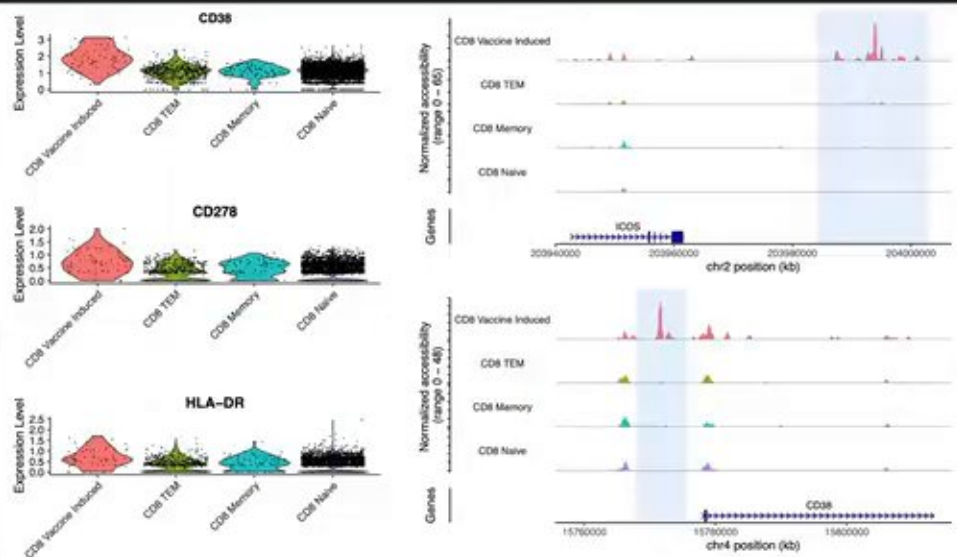
Supervised 'Bridge Integration'



Bridge integration maps integrated CITE-seq and scATAC-seq data



COVID vaccine-induced T cell responses



ASAP-seq reveals vaccine-induced enhancers in CD8 T cells



COVID vaccine-induced T cell responses



BATF3 programs CD8⁺ T cell memory

Marco A. Altalid^{1,2,3,4}, Karl Komander^{1,2}, Konrad Knöpper^{1,2}, Annika E. Peters¹, Hao Wu¹, Sarah Eickhoff¹, Tea Gogishvili¹, Justus Weber^{1,2}, Anika Grafen^{1,2}, Axel Kallies^{1,2}, Natalio Garbi¹, Hermann Einsele^{1,2}, Michael Hudecek^{1,2}, Georg Gasteiger^{1,2}, Michael Hölzel¹, Martin Voeth¹ and Wolfgang Kastnermüller^{1,2,3,4}

Antiviral CD8⁺ T cell responses are characterized by an initial activation/priming of T lymphocytes followed by a massive proliferation, robust differentiation, population contraction and the development of a stable memory pool. The transcription factor BATF3 has been shown to play a central role in the development of conventional dendritic cells, which in turn are critical for optimal priming of CD8⁺ T cells. Here we show that BATF3 was expressed transiently within the first days after T cell priming and had long-lasting effects. T cells that lacked Batf3 showed impaired expansion and differentiation, yet succumbed to an exaggerated contraction and had a diminished memory response. Vice versa, BATF3 overexpression in CD8⁺ T cells promoted their survival and transition to memory. Mechanistically, BATF3 regulated T cell apoptosis and longevity via the proapoptotic factor BIM. By programming CD8⁺ T cell survival and memory, BATF3 is a promising candidate to optimize adoptive T cell therapy in patients.

Upon encounter with dendritic cells (DCs) in the context of infection, naive antigen-specific CD8⁺ T cells undergo initial activation (priming) and consecutively a massive expansion to generate the required effector-to-target ratio to effectively eliminate infected cells in the body. At the peak of the response, expanded CD8⁺ T cells comprise a spectrum of differentiation states ranging from short-lived effector cells to long-lived memory cells. Strikingly, the quality of the developing memory T cell response is already determined during the first days of activation, among other factors by cognate CD8⁺ T cells, which transmit their signals via a subset of conventional DCs (cDCs)^{1–3}. Revealing the underlying programs and the transcriptional elements that promote CD8⁺ T cell fitness, functionality and, ultimately, determine the efficacy of successful immunotherapy in humans.

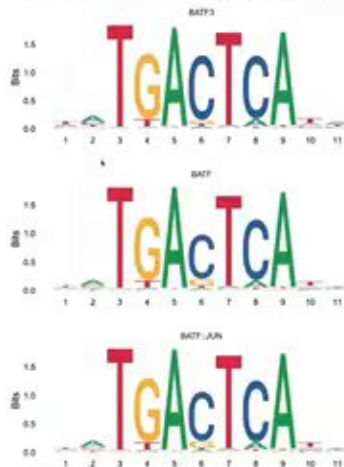
The family of AP-1 (activator protein 1) transcription factors comprises numerous basic region leucine zipper (BRLZ) proteins that belong to different families (JUN, FOS, ATF, basic leucine zipper transcription factor 125-like (BATF) and MAZ) and serve intermediaries to execute transcriptional activity. AP-1 transcription factors play a central role in many tissue and cell type and regulate cell death, survival and proliferation. The overall regulation of AP-1 activity and its molecular elements is highly complex and involves transcriptional coactivators, modulation by MAP kinase and transcriptional inhibitors such as the BATF family of proteins. BATF proteins play important roles in the development and function of both myeloid and lymphoid cell populations. They act as negative regulators of the AP-1 complex, but also promote transcriptional activity by interacting with interferon (IFN) regulatory factor (IRF) family members⁴. BATF3 was first identified in humans

(CD8⁺ versus CD8⁺), the conventional conventional and the relative latency⁵. Thus, Batf3-deficient mice became a key model for elucidating the *in vivo* functions of cDCs.

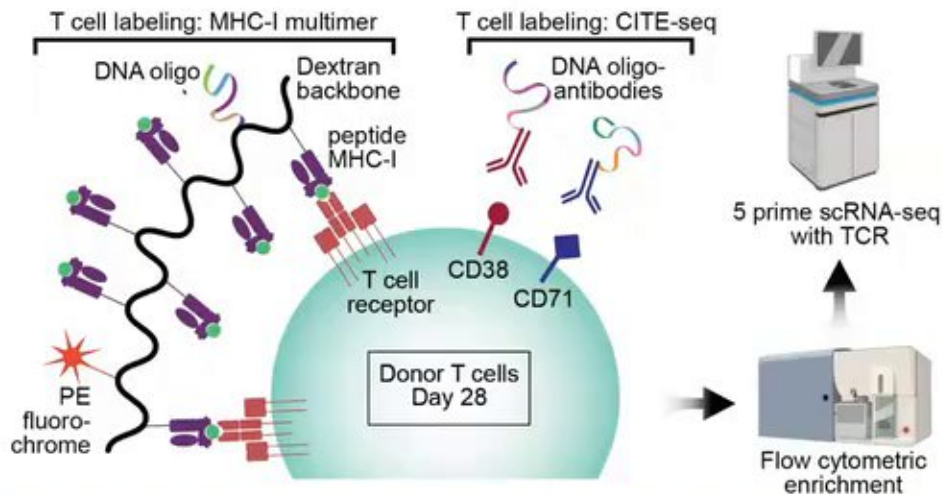
In murine lymphoid cells, Batf rather than Batf3 plays a central role in their function, development and differentiation. In particular, BATF regulates various CD8⁺ T cell subsets such as intrahepatic (IL-17) expressing helper T cells, follicular helper T cells and T regulatory type 1 cells^{6–8}. In conventional CD8⁺ T cells, BATF is required for their expansion and differentiation, but during chronic antigen exposure, BATF also drives functional exhaustion^{9–11}. Mechanistically, BATF functions within the AP-1 complex, interacts with cJUN and exerts transcriptional activity together with IRF4, leading to reduced metabolic metabolism and a lowered transcriptional membrane potential. Thus, BATF regulates a large network of key transcription factors for CD8⁺ T cell differentiation, expansion, metabolism, T cell contraction and survival¹².

In this study we identified a specific and CD8⁺ T cell-intrinsic role for BATF3. In contrast to Batf-deficient mice of Batf3 did not impact CD8⁺ T cell differentiation or their capacity to produce inflammatory cytokines. Instead, Batf3-deficient CD8⁺ T cells showed a specific defect in T cell memory formation following viral or bacterial infection. The underlying molecular mechanisms were due to exaggerated apoptosis caused by a dysregulated expression of the proapoptotic factor BIM (BCL2L1) in Batf3-deficient T cells. Conversely, overexpression of BATF3 regulated CD8⁺ T cell metabolism, inhibited cell death via suppression of BIM and promoted their memory development. By increasing the lifespan of both murine and human CD8⁺ T cells, BATF3 is a promising candidate to enhance adoptive immunotherapy against cancer and infection.

Motifs differentiating vaccine-induced populations

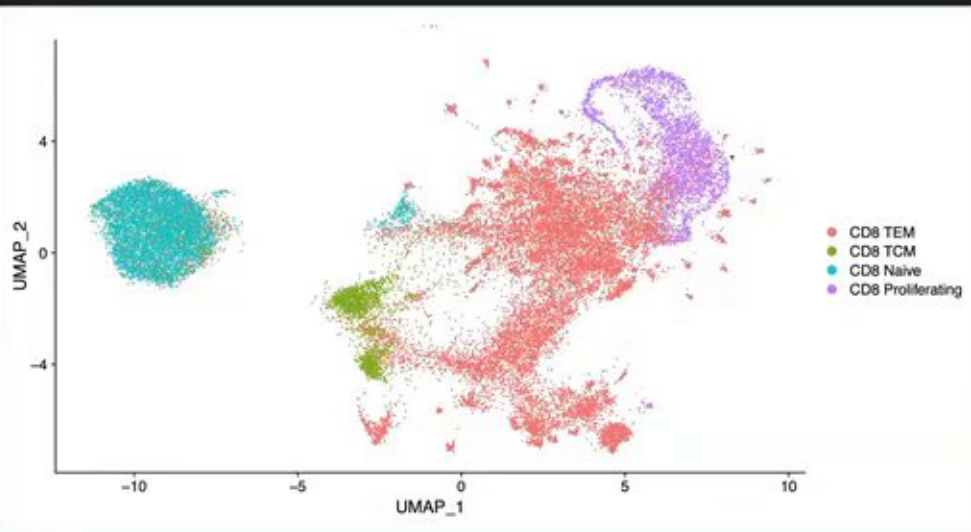


Antigen-specific or passenger response?



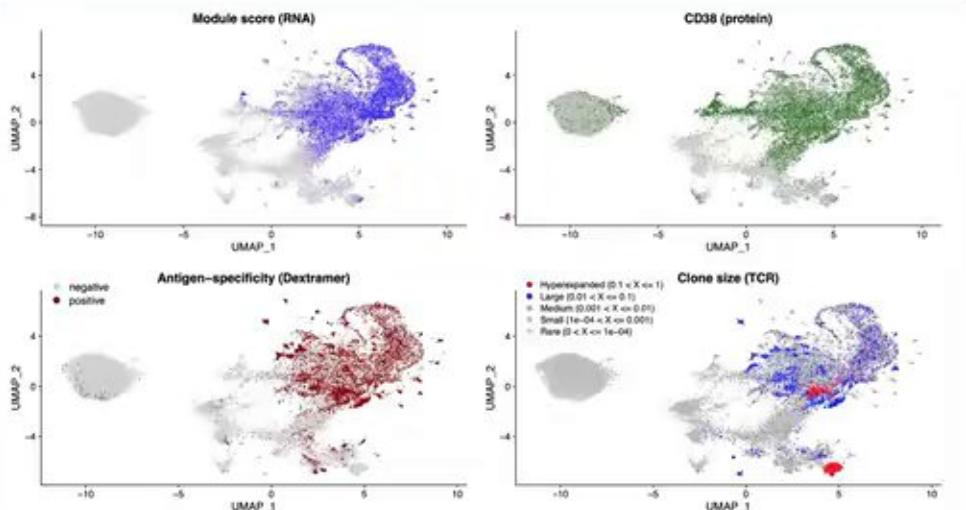
ECCITE-seq can capture antigen specificity and immune repertoires

Antigen-specific or passenger response?



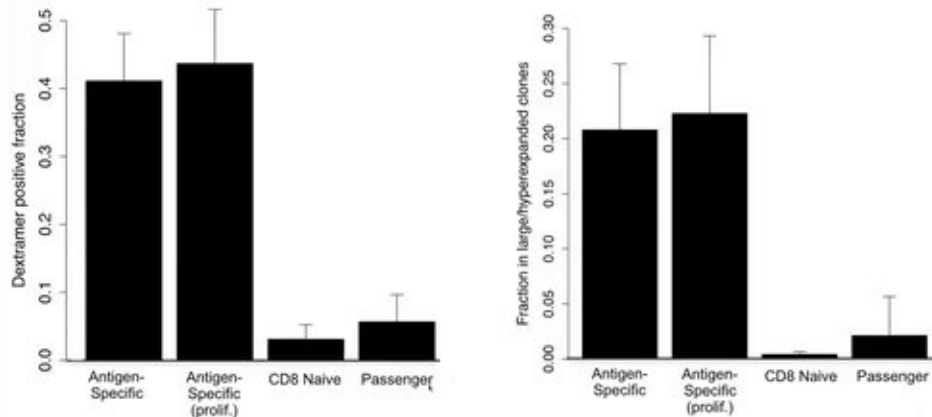
ECCITE-seq can capture antigen specificity and immune repertoires

Antigen-specific or passenger response?



ECCITE-seq can capture antigen specificity and immune repertoires

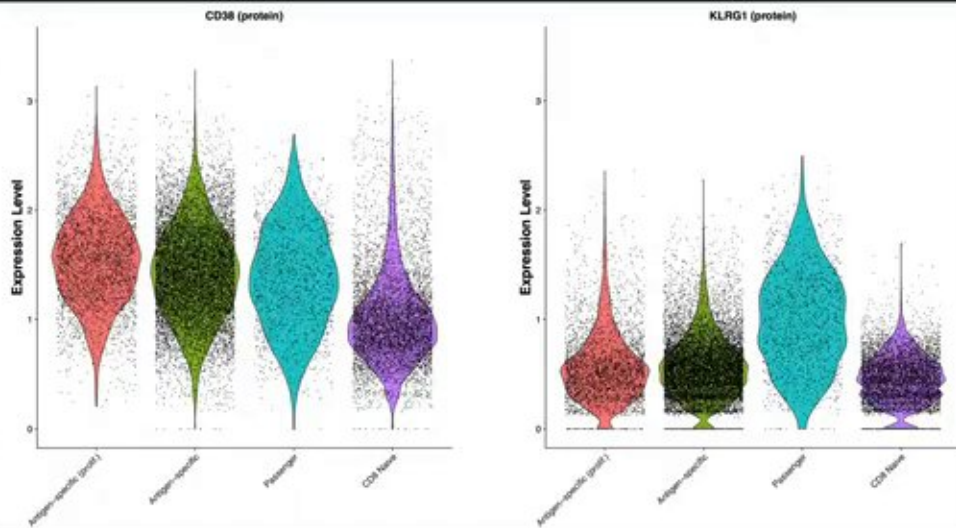
Antigen-specific or passenger response?



A subpopulation of CD38+ CD8+ T cells are not antigen specific



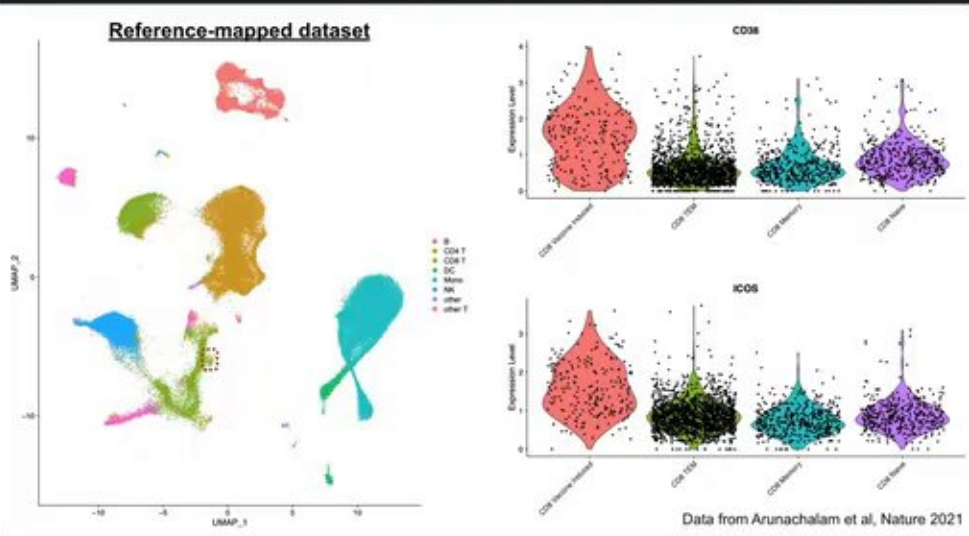
Antigen-specific or passenger response?



A subpopulation of CD38+ CD8+ T cells are not antigen specific



Identification of vaccine-induced cell populations



Vaccine-induced cells are reproducibly identified across datasets

Antigen-specific cells in COVID datasets

nature

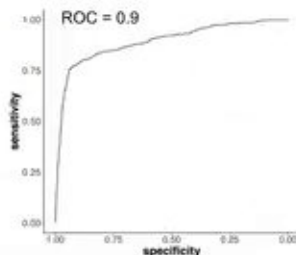
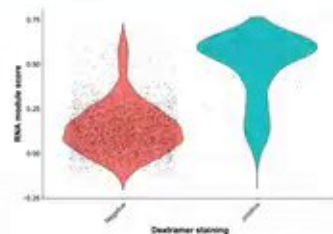
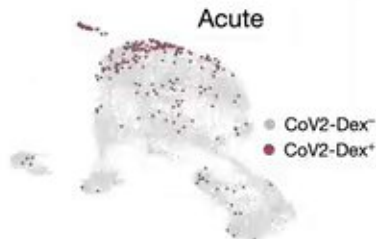
Explore content ▾ About the journal ▾ Publish with us ▾

nature > articles > article

Article | Open Access | Published: 07 December 2021

Signature of long-lived memory CD8⁺ T cells in acute SARS-CoV-2 infection

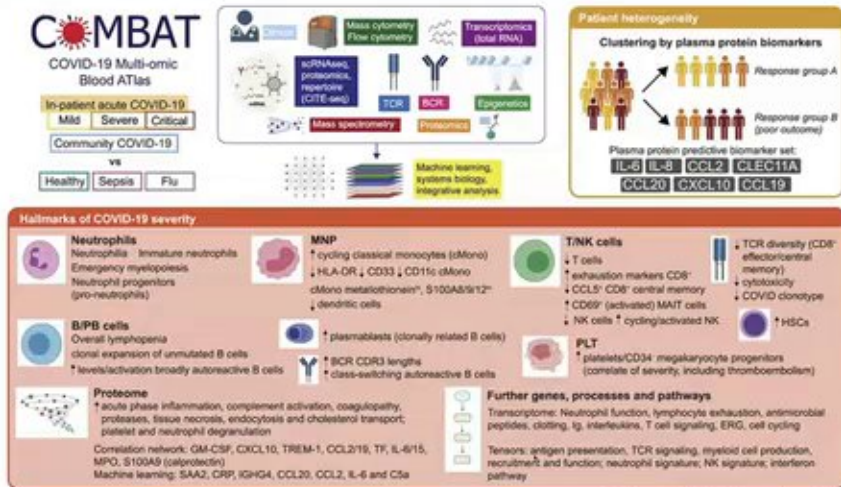
Sirab Adams, Jari Mitter, Yves Zurbuchen, Carlo Serevi, Patrick Tetschler, Mira E. Raebler, Simone Beuthelein, Jakob Nilsson, Andreas E. Moor & Omar Bouman 



Module score accurately classifies antigen specificity during acute COVID

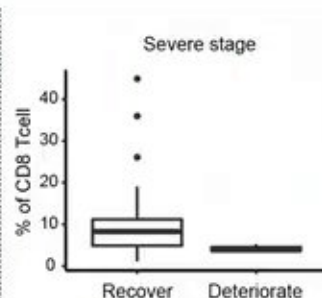
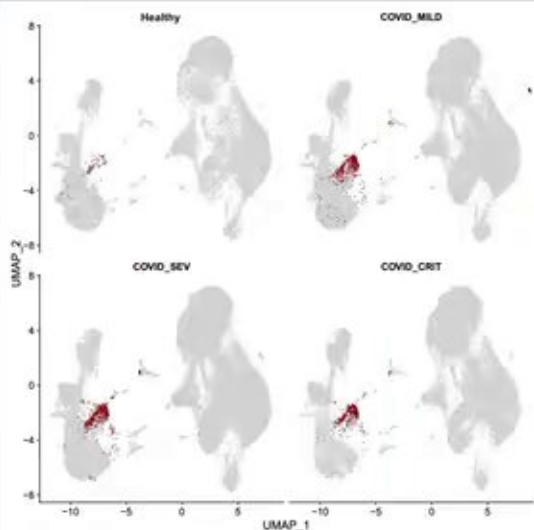


Antigen-specific cells in COVID datasets



Data from COMBAT Consortium, Cell 2022

Antigen-specific cells in COVID datasets

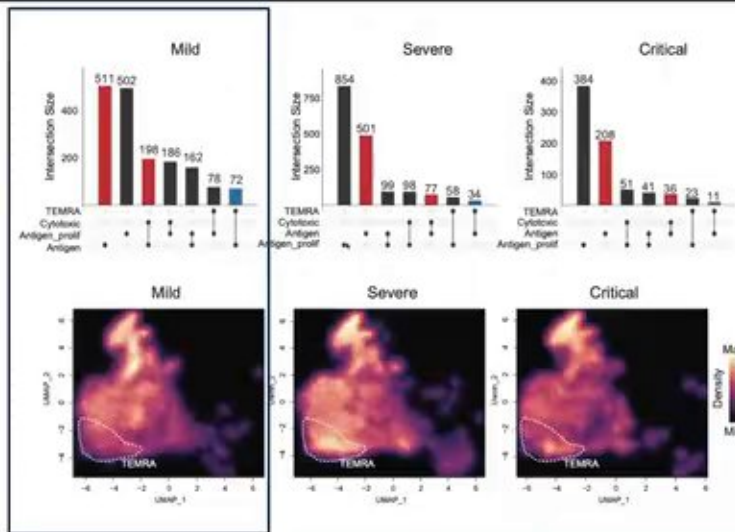


Patients with minimal CD8 T responses have a substantially reduced rate of recovery from severe disease

Data from COMBAT Consortium, Cell 2022

CD8 T cell responses correlate with patient outcomes

Antigen-specific cells in COVID datasets



In mild disease, antigen-specific cells differentiate into cytotoxic states

Summary

Identification of vaccine-induced CD8 T cells in single-cell datasets

- Multimodal data and analysis is key for identification
- Clear separation of proliferating and non-proliferating populations
- Leads to identification of rich gene signature



Acknowledgements



Bingjie Zhang
Gesmira Molla
Jimmy Yun
Madeline Kowalski
Carol Dalgarno
Alex Bradu
Saket Choudhary

John Blair
Tracy Smith
Isabella Mascio
Nathan Nakatsuka
John Blair
Izzy Grabski
Sarah Curtiss



















































































































































