

cNMF on Terra

Employing consensus non-negative matrix factorization on a cloud
based workspace

High-Level Overview

- cNMF overview
 - cNMF paper on eLIFE
 - Our paper on bioRxiv
 - Program identification (cNMF)
 - Cell Identification and annotation (Seurat)
 - Integration and visualization of cells and gene expression programs
- Terra overview
 - Google billing project setup
 - Google cloud cli setup

cNMF github

- <https://github.com/dylkot/cNMF>
- Original paper at:
<https://bit.ly/4dYyDQV>

The screenshot shows two side-by-side web pages. On the left is the GitHub repository for cNMF, featuring a dark-themed interface with a red header bar. The main content area displays the README file, which includes a table of contents, a brief description of the project, and links to the code and documentation. Below the README is a section for "Installation" with instructions for pip installation. On the right is the original research paper from eLife titled "Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq". The paper's title, authors, and abstract are visible, along with a "Full text" link and other navigation options.

Tools and Resources
Computational and Systems Biology, Genetics and Genomics

Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq

Dylan Kotiar, Adrian Veres, M Aurel Nagy, Shervin Tabrizi, Eran Hodis, Douglas A Melton, Pardis C Sabeti
Harvard Medical School, United States; Broad Institute of MIT and Harvard, United States; Massachusetts Institute of Technology, United States; Harvard University, United States; Howard Hughes Medical Institute, United States
Jul 8, 2019 • https://doi.org/10.7554/eLife.43803

Full text Figures and data Peer review Side by side

Abstract

Identifying gene expression programs underlying both cell-type identity and cellular activities (e.g. life-cycle processes, responses to environmental cues) is crucial for understanding the organization of cells and tissues. Although single-cell RNA-Seq (scRNA-Seq) can quantify transcripts in individual cells, each cell's expression profile may be a mixture of both types of programs, making them difficult to disentangle. Here, we benchmark and enhance the use of matrix factorization to solve this problem. We show with simulations that a method we call consensus non-negative matrix factorization (cNMF) accurately infers identity and activity programs, including their relative contributions in each cell. To illustrate the insights this approach enables, we apply it to published brain organoid and visual cortex scRNA-Seq datasets; cNMF refines cell types and identifies both expected (e.g. cell cycle and hypoxia) and novel activity programs, including programs that may underlie a neurosecretory phenotype and synaptogenesis.

Metrics https://doi.org/10.7554/eLife.43803.001

The screenshot shows the "Installation" and "Running cNMF" sections of the cNMF GitHub README. The "Installation" section contains instructions for pip installation and mentions the need for Harmony and scikit-misc for batch correction preprocessing. The "Running cNMF" section provides example command-line commands for running the software. Below these, there is a section for running the steps within a Python environment, showing a sample script.

README MIT license

Installation

cNMF has been tested with Python 3.7 and 3.10 and requires scikit-learn>=1.0, scanpy>=1.8, and AnnData>=0.9

You can install with pip:

```
pip install cnmf
```

If you want to use the batch correction preprocessing, you also need to install the [Python implementation of Harmony](#) and scikit-misc

```
pip install harmonypy  
pip install scikit-misc
```

Running cNMF

cNMF can be run from the command line without any parallelization using the example commands below:

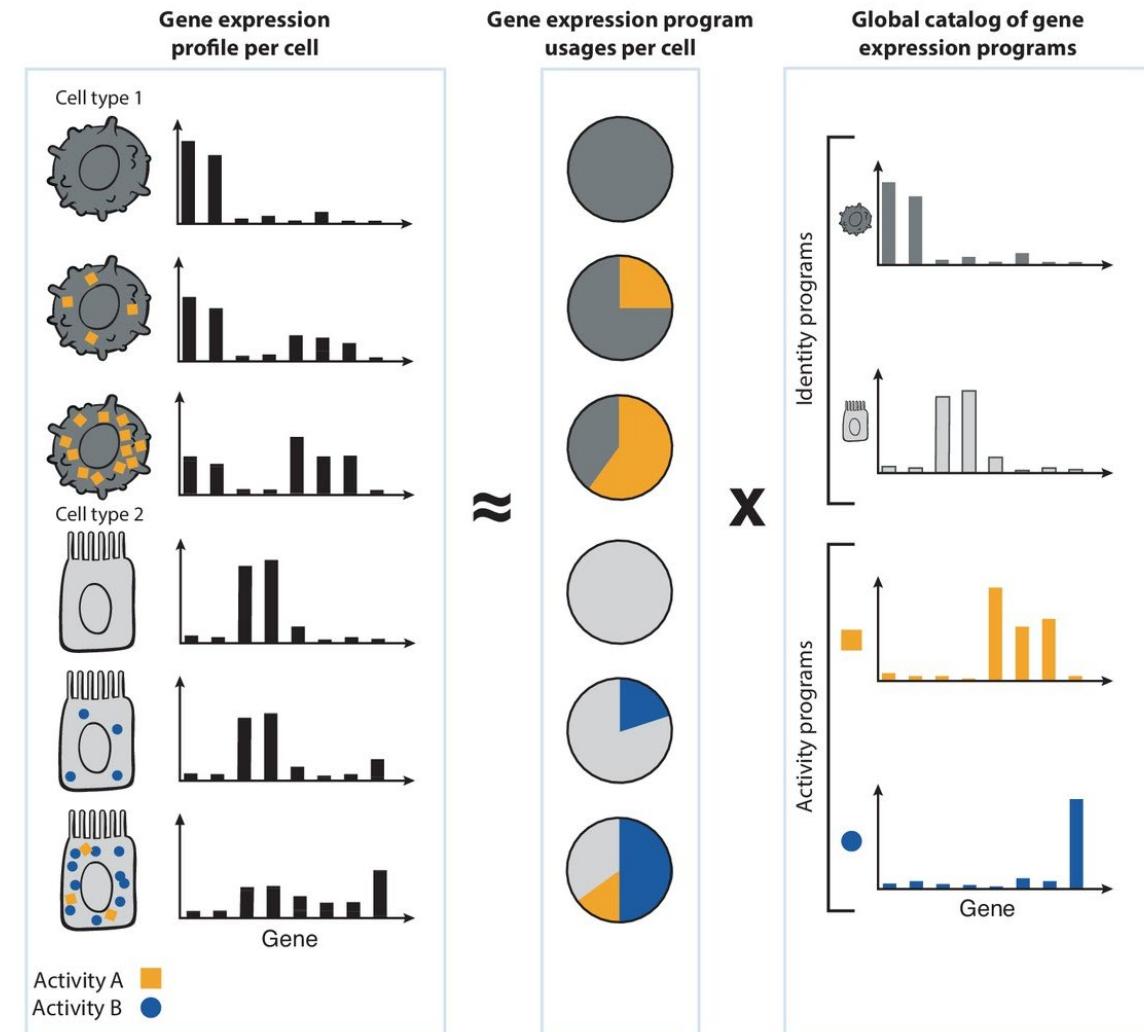
```
cnmf prepare --output-dir ./example_data --name example_cNMF -c ./example_data/counts_prefiltered.txt  
cnmf factorize --output-dir ./example_data --name example_cNMF --worker-index 0 --total-workers 1  
cnmf combine --output-dir ./example_data --name example_cNMF  
cnmf k_selection_plot --output-dir ./example_data --name example_cNMF  
cnmf consensus --output-dir ./example_data --name example_cNMF --components 10 --local-density-threshold
```

Or alternatively, the same steps can be run from within a Python environment using the commands below:

```
from cnmf import CNMF  
cnmf_obj = CNMF(output_dir='./example_data', name='example_cNMF')  
cnmf_obj.prepare(counts_fn='./example_data/counts_prefiltered.txt', components=np.arange(5,14), n_iter=1)  
cnmf_obj.factorize(worker_id=0, total_workers=1)  
cnmf_obj.combine()  
cnmf_obj.k_selection_plot()  
cnmf_obj.consensus(k=10, density_threshold=0.01)  
usage, spectra_scores, spectra_tpm, top_genes = cnmf_obj.load_results(k=10, density_threshold=0.01)
```

Program discovery within cell types (cNMF)

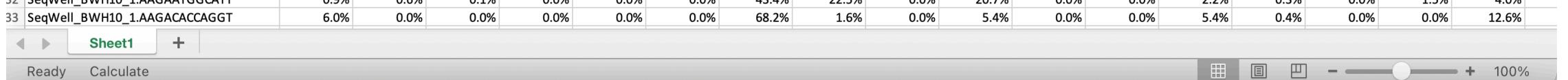
- Variable Genes and Cells.
- Identify patterns of gene expression.
- Sets of genes co-expressed and co-silenced (Programs)
- Activity and Identity



cNMF Outputs (2 Matrices)

- Genes contribution to each program identified
- Cells usages of the programs

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	Program17	Program18
1	Program1	Program2	Program3	Program4	Program5	Program6	Program7	Program8	Program9	Program10	Program11	Program12	Program13	Program14	Program15	Program16	Program17	Program18	
2	SeqWell_BWH10_1.AAAAACATCATC	0.0%	3.4%	0.0%	0.7%	0.0%	0.0%	66.2%	0.0%	0.0%	0.0%	1.4%	0.0%	1.2%	0.0%	16.1%	4.4%	0.9%	+ ...
3	SeqWell_BWH10_1.AAAAATACGTA	0.0%	0.3%	0.0%	0.0%	0.0%	0.4%	75.6%	0.8%	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%	9.6%	5.9%	0.0%	
4	SeqWell_BWH10_1.AAACCGGGCTT	0.9%	0.2%	0.5%	0.0%	0.0%	0.0%	56.5%	3.2%	0.3%	5.0%	0.0%	0.0%	9.6%	0.0%	7.2%	7.2%	2.9%	
5	SeqWell_BWH10_1.AAACGGCTTTG	2.0%	0.0%	0.4%	0.5%	0.0%	0.0%	49.9%	1.5%	5.0%	14.3%	0.0%	1.0%	5.1%	0.0%	0.0%	1.9%	4.1%	
6	SeqWell_BWH10_1.AAACTAACTCGG	0.0%	0.7%	0.5%	0.0%	0.0%	0.0%	45.9%	1.0%	0.0%	0.7%	0.0%	0.0%	26.3%	0.0%	10.0%	0.0%	0.0%	
7	SeqWell_BWH10_1.AAACTCGTCTCC	0.0%	0.0%	0.0%	2.2%	0.0%	0.0%	80.2%	0.3%	0.0%	1.0%	0.0%	1.5%	7.2%	0.0%	0.0%	2.1%	4.6%	
8	SeqWell_BWH10_1.AAATAGCGCTGA	0.0%	1.5%	0.0%	0.0%	3.1%	0.0%	53.4%	0.0%	0.0%	0.0%	1.3%	0.0%	5.8%	3.7%	10.9%	0.0%	0.0%	
9	SeqWell_BWH10_1.AAATCAGCGTCC	0.0%	0.0%	0.0%	3.1%	0.0%	0.0%	54.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	23.3%	10.2%	3.2%	6.0%	
10	SeqWell_BWH10_1.AAATCTAACGT	6.4%	0.0%	0.6%	0.8%	0.0%	0.0%	53.4%	0.1%	0.0%	1.8%	0.0%	0.0%	18.0%	0.0%	2.7%	0.0%	1.3%	
11	SeqWell_BWH10_1.AAATCGCACTAC	0.0%	0.0%	0.4%	4.7%	0.0%	0.0%	55.1%	12.3%	0.9%	5.1%	0.0%	0.0%	7.4%	0.0%	12.5%	0.0%	0.0%	
12	SeqWell_BWH10_1.AAATGGGACATG	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	66.4%	0.0%	0.0%	3.0%	0.0%	0.0%	12.1%	0.0%	16.4%	0.0%	0.2%	
13	SeqWell_BWH10_1.AAATGTCAAAT	11.0%	3.2%	0.1%	1.3%	0.0%	0.0%	54.2%	0.8%	2.4%	1.4%	0.5%	12.2%	3.7%	0.0%	0.0%	0.0%	0.0%	
14	SeqWell_BWH10_1.AACAACCGGGG	0.0%	0.0%	0.0%	2.8%	0.0%	0.0%	57.3%	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%	0.0%	0.0%	7.5%	0.0%	
15	SeqWell_BWH10_1.AACAAGGCTTAG	5.3%	0.0%	0.0%	2.9%	0.0%	0.0%	45.5%	0.0%	0.0%	0.0%	0.8%	2.1%	0.0%	19.0%	13.8%	0.0%	3.6%	
16	SeqWell_BWH10_1.AACAGCGAAAAC	3.5%	0.2%	0.0%	1.9%	0.0%	0.0%	47.9%	2.4%	0.0%	24.7%	2.0%	0.0%	3.3%	0.0%	5.1%	0.0%	2.5%	
17	SeqWell_BWH10_1.AACAGTCTAGTT	5.8%	0.0%	0.6%	0.0%	0.0%	0.0%	61.1%	4.0%	0.0%	4.3%	0.0%	1.4%	0.0%	0.0%	0.3%	4.7%	1.0%	
18	SeqWell_BWH10_1.AACAGTGTAAA	0.0%	1.7%	0.0%	4.2%	0.0%	0.0%	60.2%	0.3%	4.3%	0.0%	1.9%	1.7%	2.7%	0.0%	0.0%	4.1%	0.0%	
19	SeqWell_BWH10_1.AACATCTGGTTC	4.5%	0.1%	0.0%	0.0%	0.0%	0.0%	60.5%	0.0%	0.0%	0.2%	0.8%	3.4%	0.0%	4.7%	0.0%	0.0%	0.0%	
20	SeqWell_BWH10_1.AACCGAGTCCAG	1.2%	0.3%	0.7%	6.0%	0.0%	0.1%	31.6%	6.0%	0.2%	0.7%	5.4%	4.7%	0.0%	0.0%	0.0%	5.6%	13.0%	
21	SeqWell_BWH10_1.AACCTCATGGTT	0.0%	0.0%	0.5%	0.0%	0.0%	0.5%	61.7%	6.7%	0.0%	6.1%	0.0%	1.4%	14.1%	0.0%	0.0%	3.5%	0.4%	
22	SeqWell_BWH10_1.AACGCAAGAACAT	2.0%	0.0%	0.0%	0.0%	0.0%	0.0%	77.7%	0.0%	0.0%	7.1%	0.0%	0.0%	4.8%	0.9%	1.1%	0.0%	4.2%	
23	SeqWell_BWH10_1.AACGGTTCCATC	4.0%	0.0%	0.0%	1.7%	0.0%	0.0%	76.0%	0.0%	0.0%	9.0%	0.0%	0.0%	2.0%	0.0%	2.3%	0.0%	2.1%	
24	SeqWell_BWH10_1.AACGTGCCCTC	0.0%	0.0%	1.2%	0.0%	0.0%	0.0%	43.6%	2.9%	0.2%	8.9%	0.0%	4.9%	0.0%	5.3%	0.0%	10.1%	0.0%	
25	SeqWell_BWH10_1.AACGTGTGCCAG	7.3%	0.0%	0.1%	0.6%	0.0%	0.2%	47.3%	2.7%	0.0%	2.2%	0.0%	0.6%	0.0%	12.3%	5.9%	7.2%	6.2%	
26	SeqWell_BWH10_1.AACGTGTCTAC	4.4%	0.0%	0.0%	3.4%	5.5%	1.0%	32.5%	4.7%	0.1%	6.4%	0.0%	0.0%	0.2%	0.0%	4.2%	7.5%		
27	SeqWell_BWH10_1.AACTAGTCAGCT	0.0%	0.0%	0.0%	0.0%	1.9%	0.0%	65.1%	0.0%	0.0%	5.1%	0.0%	0.0%	9.2%	0.0%	13.7%	0.0%	0.0%	
28	SeqWell_BWH10_1.AACTATCAGAAA	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	66.5%	0.0%	0.0%	1.8%	1.3%	0.0%	7.2%	0.0%	21.5%	0.0%	1.6%	
29	SeqWell_BWH10_1.AACTCTAACCGG	5.5%	0.0%	0.0%	0.0%	0.0%	0.0%	75.5%	0.0%	0.0%	1.8%	0.0%	0.0%	5.7%	0.0%	0.0%	0.0%	3.4%	
30	SeqWell_BWH10_1.AACTCTTCTAT	6.5%	2.4%	8.0%	0.1%	0.9%	0.0%	25.4%	3.4%	1.3%	9.3%	9.1%	0.0%	13.1%	0.0%	0.0%	2.6%	8.5%	
31	SeqWell_BWH10_1.AACTTCTCCCAA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	71.2%	0.0%	0.0%	1.7%	0.0%	2.3%	1.7%	4.9%	1.5%	8.4%	0.0%	
32	SeqWell_BWH10_1.AAGAATGGCATT	0.9%	0.6%	0.1%	0.0%	0.0%	0.0%	43.4%	22.5%	0.0%	20.7%	0.0%	0.0%	2.2%	0.3%	0.0%	1.5%	4.0%	
33	SeqWell_BWH10_1.AAGACACCAGGT	6.0%	0.0%	0.0%	0.0%	0.0%	0.0%	68.2%	1.6%	0.0%	5.4%	0.0%	0.0%	5.4%	0.4%	0.0%	0.0%	12.6%	



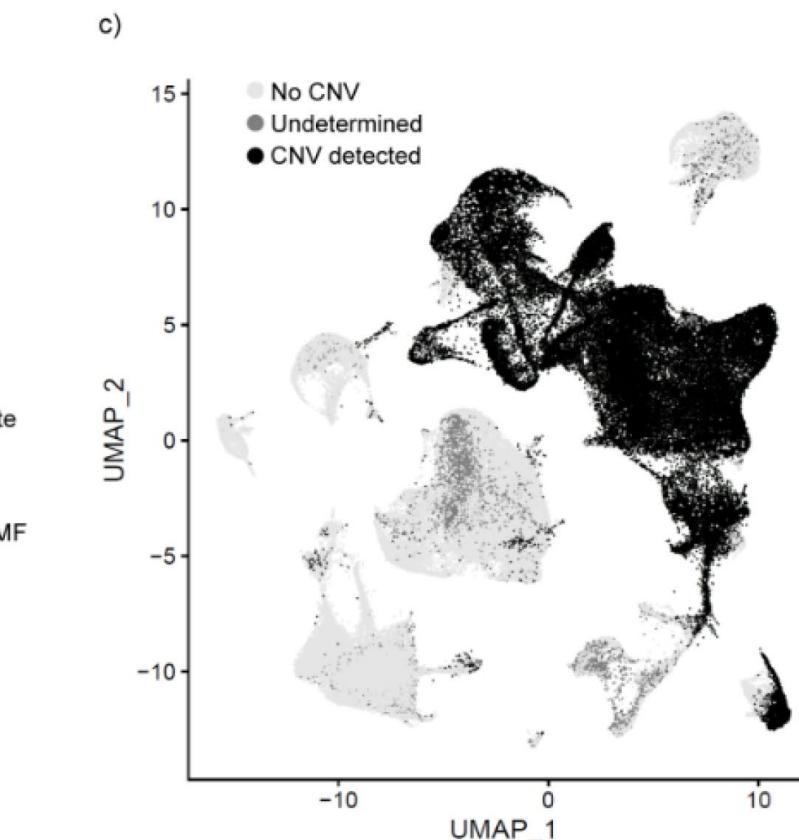
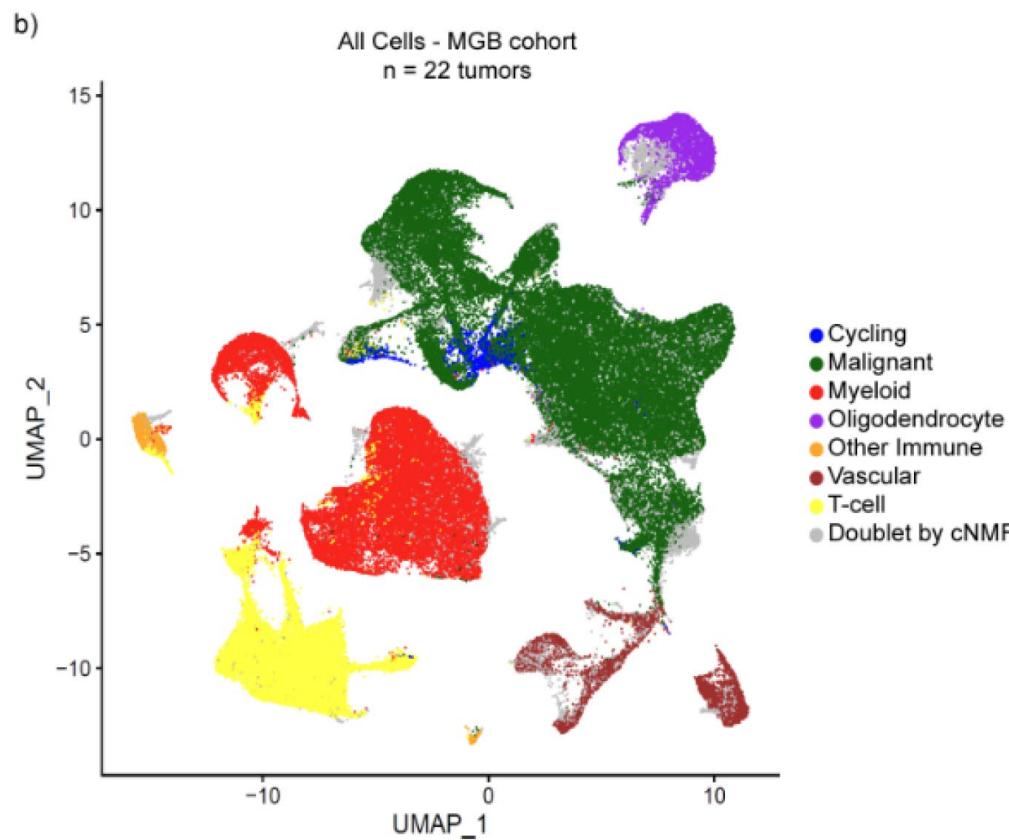
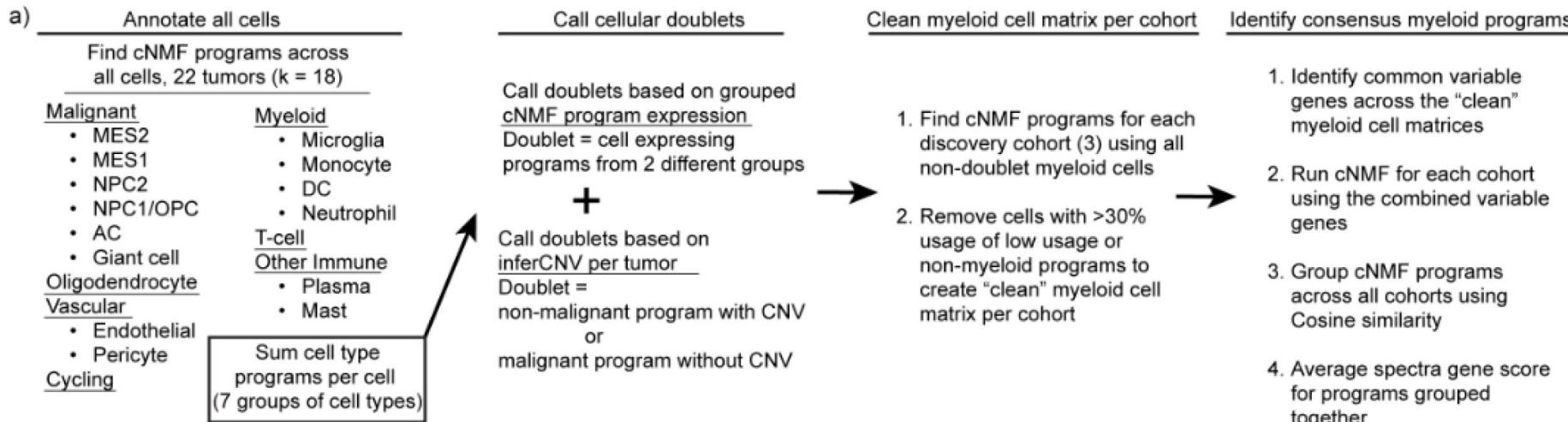
The project:

1. Combine single-cell genomics and proteomics to define myeloid states, functional markers, and tissue origins.
2. Develop interventions to reprogram or neutralize pro-tumor macrophages.

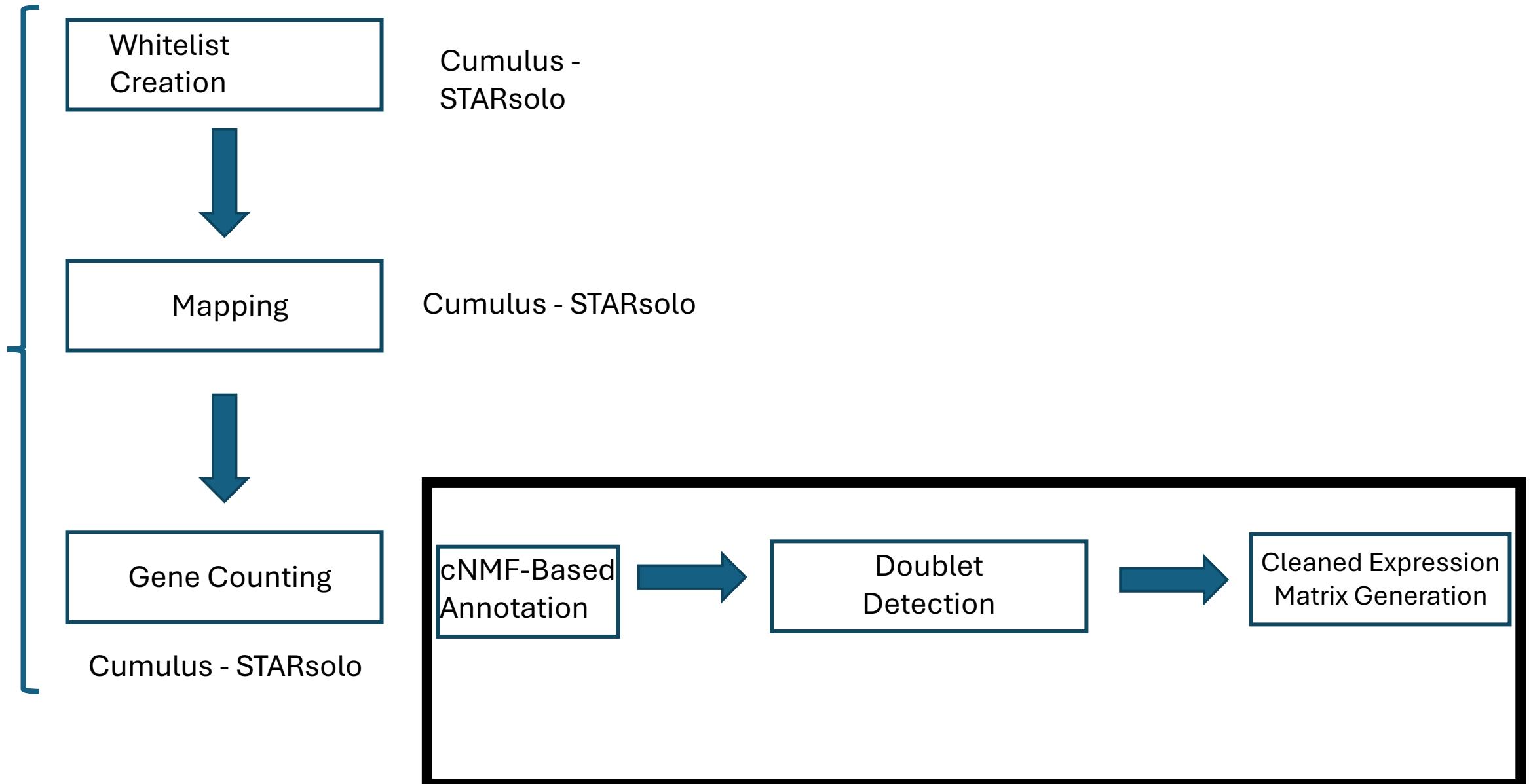
The payoff:

Therapeutic approaches that reprogram myeloid cells or re-engineer T-cells cells for effective immunotherapy in glioblastoma.

What cells are present in
Gliomas?



Bioinformatic Pipeline (Version 4.0)



cNMF

- Top 4000 Variable genes
- All Cells in the matrix

Annotate Cells – Based on Program Usage - Minimum Usage 25% and Max

- T-cells
- Mast Cells
- Plasma Cells
- DC
- Microglia
- Monocytes
- Neutrophils
- AC
- NPC1/OPC
- NPC2
- MES1
- MES2
- Giant Glioma Cell
- Oligo
- Pericytes
- Endothelial Cells
- Cycling

Cell Identification performed in Seurat

- <https://satijalab.org/seurat/>
- R toolkit for single-cell genomics



Seurat v5

We are excited to release Seurat v5! To install, please follow the instructions in our [install page](#). This update brings the following new features and functionality:

- **Integrative multimodal analysis:** The cellular transcriptome is just one aspect of cellular identity, and recent technologies enable routine profiling of chromatin accessibility, histone modifications, and protein levels from single cells. In Seurat v5, we introduce 'bridge integration', a statistical method to integrate experiments measuring different modalities (i.e. separate scRNA-seq and scATAC-seq datasets), using a separate multiomic dataset as a molecular 'bridge'. For example, we demonstrate how to map scATAC-seq datasets onto scRNA-seq datasets, to assist users in interpreting and annotating data from new modalities.

We recognize that while the goal of matching shared cell types across datasets may be important for many problems, users may also be concerned about which method to use, or that integration could result in a loss of biological resolution. In Seurat v5, we also introduce flexible and streamlined workflows for the integration of multiple scRNA-seq datasets. This makes it easier to explore the results of different integration methods, and to compare these results to a workflow that excludes integration steps.

- Paper: [Dictionary learning for integrative, multimodal, and scalable single-cell analysis](#)
- Vignette: [Streamlined integration of scRNA-seq data](#)
- Vignette: [Cross-modality bridge integration](#)
- Website: [Azimuth-ATAC, reference-mapping for scATAC-seq datasets](#)

- **Flexible, interactive, and highly scalable analysis:** The size and scale of single-cell sequencing datasets is rapidly increasing, outpacing even Moore's law. In Seurat v5, we introduce new infrastructure and methods to analyze, interpret, and explore exciting datasets spanning millions of cells, even if they cannot be fully loaded into memory. We introduce support for 'sketch'-based analysis, where representative subsamples of a large dataset are stored in-memory to enable rapid and iterative analysis - while the full dataset remains accessible via on-disk storage.

We enable high-performance via the BPCells package, developed by Ben Parks in the Greenleaf Lab. The BPCells package enables high-performance analysis via innovative bit-packing compression techniques, optimized C++ code, and use of streamlined and lazy operations.

- Vignette: [Sketch-based clustering of 1.3M brain cells \(10x Genomics\)](#)
- Vignette: [Sketch-based integration of 1M healthy and diabetic PBMC \(Parse Biosciences\)](#)

Links

[View on CRAN](#)

[Browse source code](#)

[Report a bug](#)

License

[Full license](#)

[MIT + file LICENSE](#)

Community

[Code of conduct](#)

Citation

[Citing Seurat](#)

Developers

Rahul Satija

Author, maintainer 

Satija Lab and Collaborators

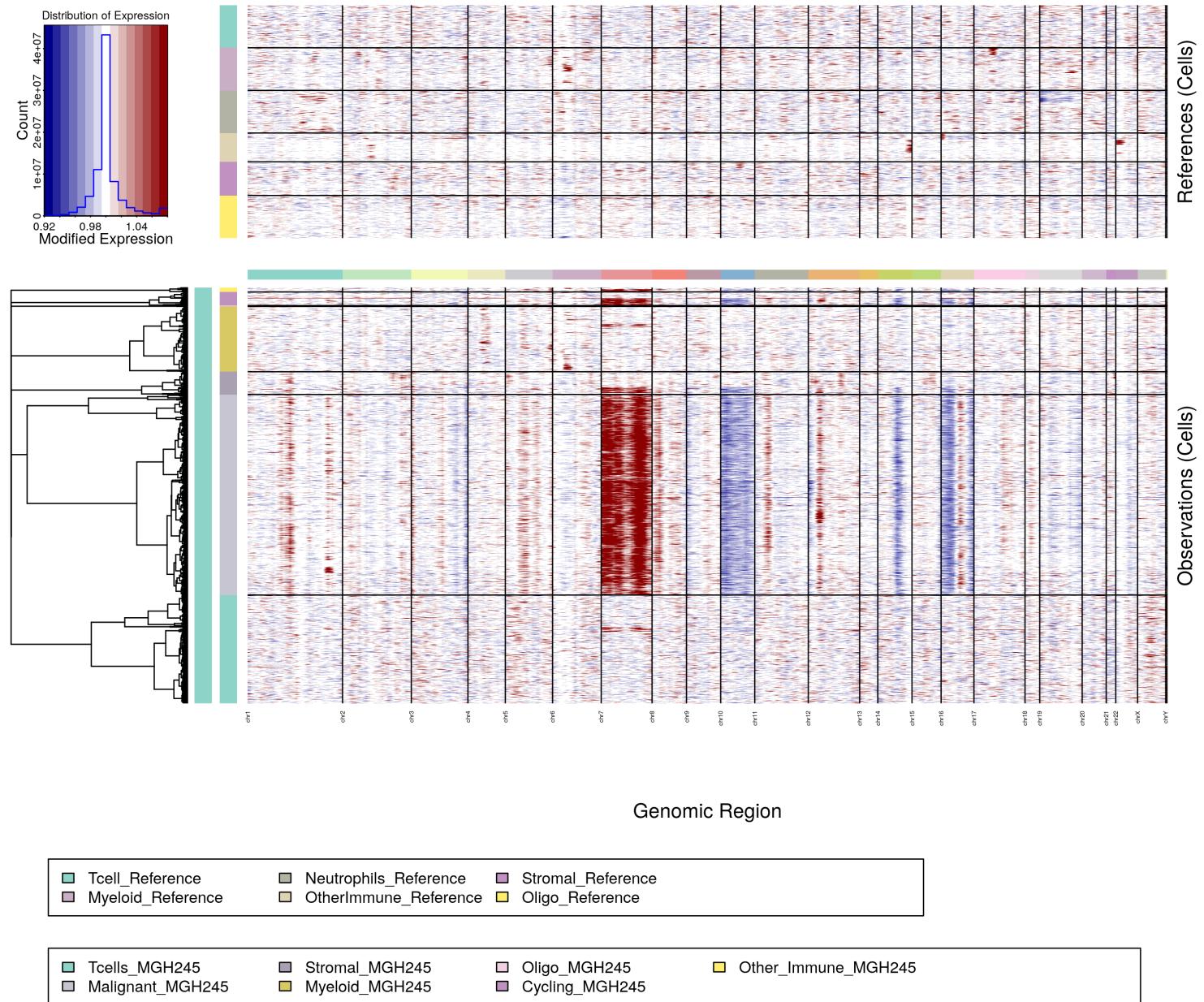
Funder

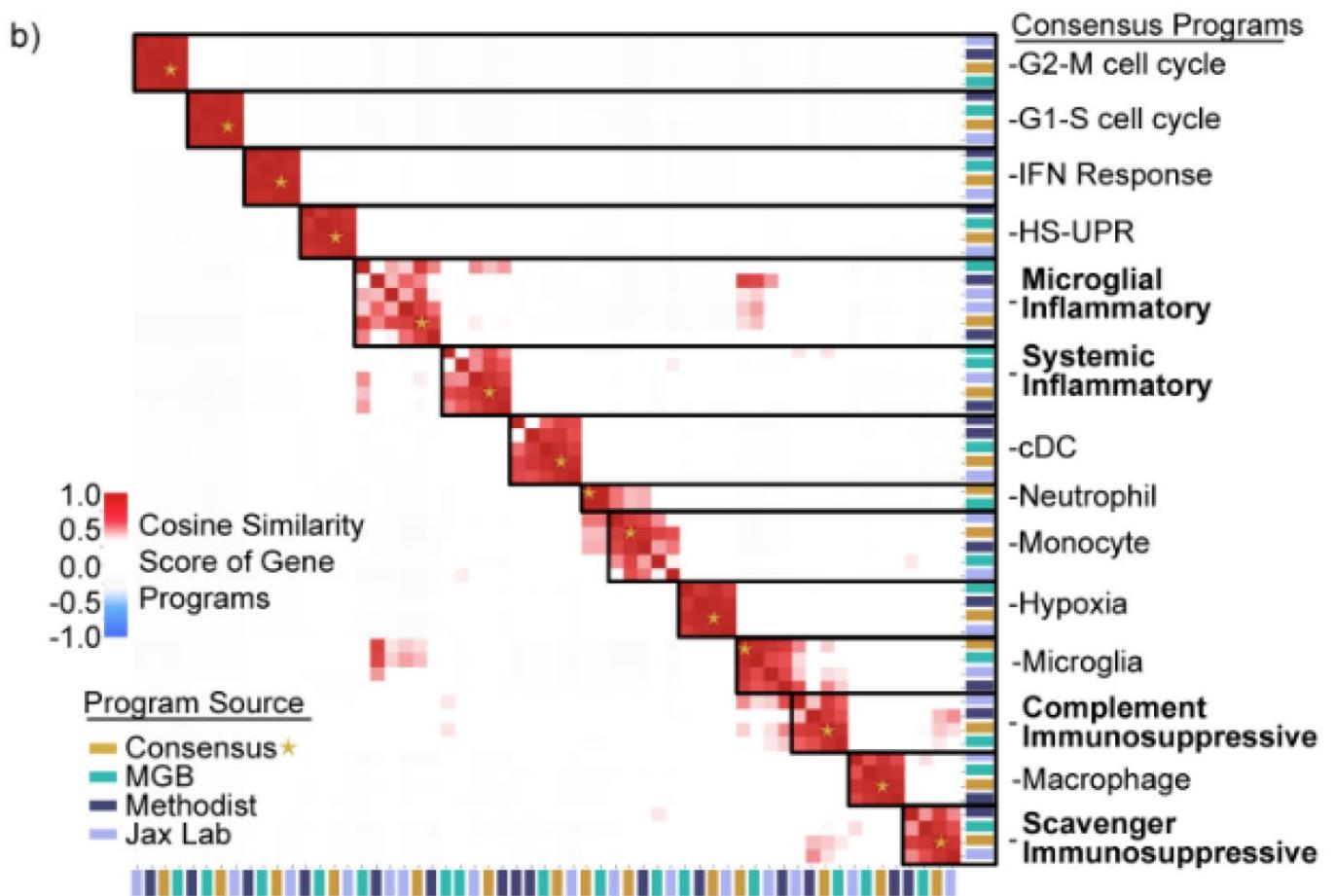
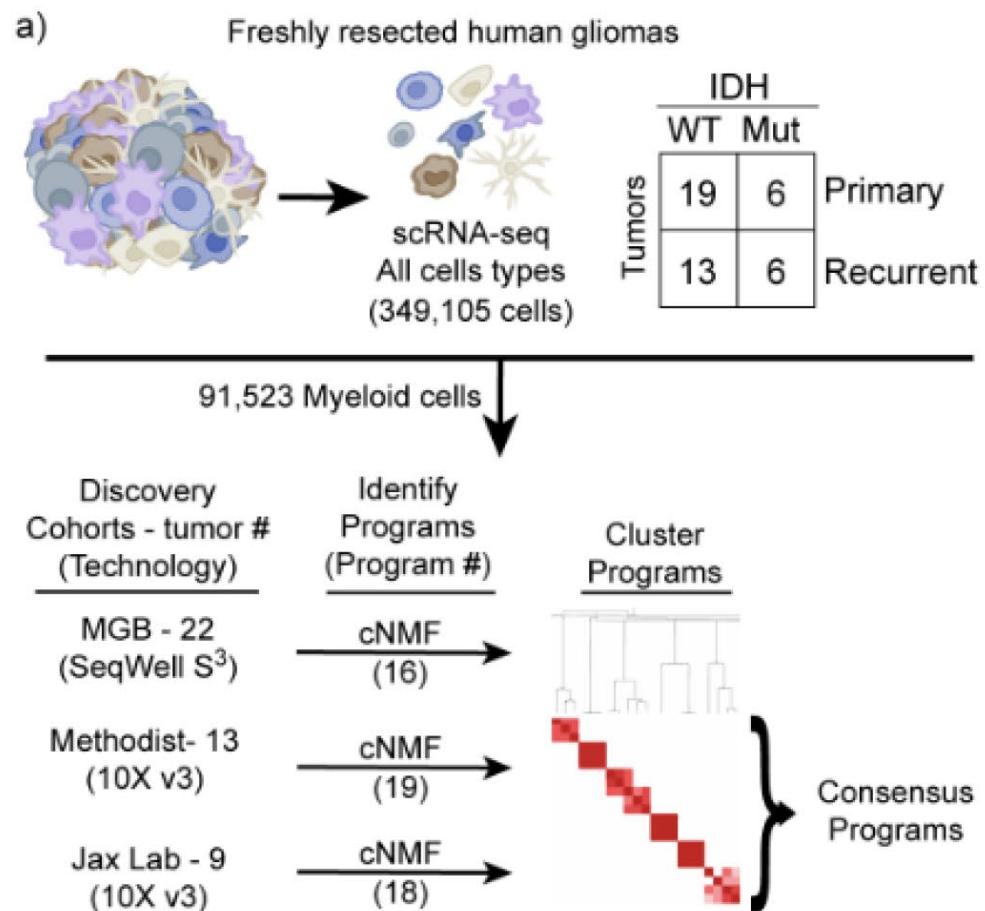
[More about authors...](#)

CNV (Copy number Variation) Analysis

- inferCNV
- Reference Cells (Used to create median expression level)
- Concentration of Genes in chromosomes
- Selection of reference cells is very crucial

inferCNV

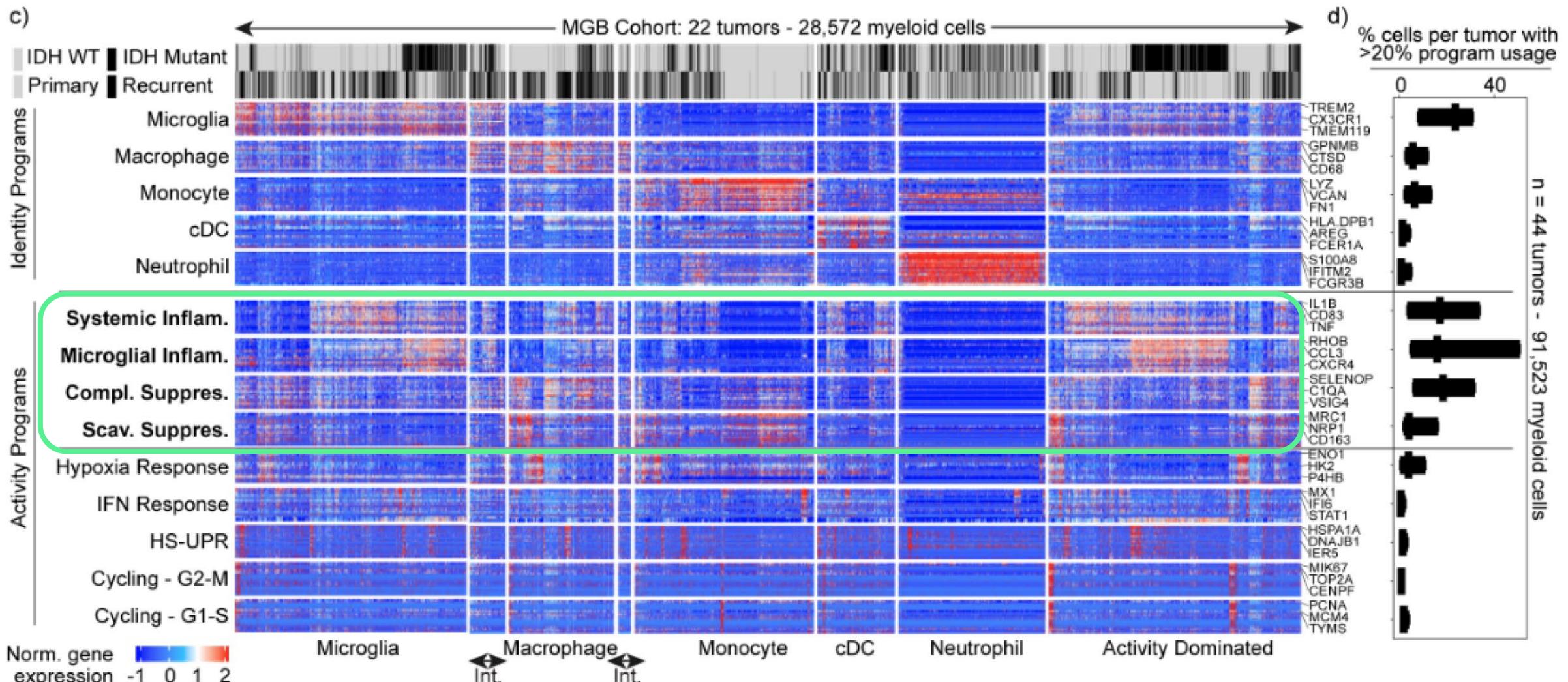




Miller et al, 2023

doi: <https://doi.org/10.1101/2023.10.24.563466>

Gene Expression Programs identified by cNMF



Miller et al, 2023

<https://bit.ly/4dOjhOD>

doi: <https://doi.org/10.1101/2023.10.24.563466>

Programs, Origins, and Niches of Immunomodulatory Myeloid Cells in Gliomas

- <https://bit.ly/4dOjhOD>
 - Miller et al, 2023
 - Github for this paper: <https://bit.ly/3XnEE3o>

New Results

Follow this preprint

Previous

Next

Programs, Origins, and Niches of Immunomodulatory Myeloid Cells in Gliomas

Tyler E. Miller, Chadi A. El Farran, Charles P. Couturier, Zeyu Chen, Joshua P. D'Antonio, Julia Verga, Martin A. Villanueva, L. Nicolas Gonzalez Castro, Yuzhou Evelyn Tong, Tariq Al Saadi, Andrew N. Chiocca, David S. Fischer, Dieter Henrik Heiland, Jennifer L. Guerrero, Kevin Petrecca, Mario L. Suva, Alex K. Shalek, Bradley E. Bernstein

doi: <https://doi.org/10.1101/2023.10.24.563466>

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

0 0 0 0 45

Abstract Full Text Info/History Metrics Preview PDF

Download PDF Print/Save Options Supplementary Material Data/Code

Email Share Citation Tools Get QR code

COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv

Subject Area

Cancer Biology

Subject Areas

All Articles

Animal Behavior and Cognition
Biochemistry
Bioengineering
Bioinformatics
Biophysics
Cancer Biology
Cell Biology
Clinical Trials®
Developmental Biology
Ecology
Epidemiology®
Evolutionary Biology
Genetics
Genomics
Immunology

Programs, Origins, and Niches of Immunomodulatory Myeloid Cells in Gliomas

This GitHub repository includes all scripts used to perform the analyses to study the Biology of Myeloid Cells in Glioma Microenvironments as per the methods section of: The steps are numbered in each folder according to the order in which they must be implemented.

Processing of the Single-cell RNA-Seq libraries (Related to Figure 1):

The folder "Processing of scRNA-Seq Files (Related to Figure 1)" includes the necessary steps to analyze the scRNA-Seq libraries. This folder contains all the codes required to align the fastq files, generate expression matrices, process the cells in the gene expression matrices, annotate the cells, determine CNVs, identify doublets, perform batch correction for the myeloid cells in the MGB cohort, and generate UMAPs. The "Required Files" folder includes additional files to run these codes.

Identifying recurrent expression programs in Glioma-associated myeloid cells (Related to Figure 1):

The folder "Identifying recurrent programs in Myeloid Cells in Gliomas (Related to Figure 1)" includes the necessary scripts and instructions to identify the recurrent expression programs in Tumor-associated myeloid cells using cNMF. This folder contains all the codes required to identify variable genes in the myeloid cells expression matrix in each cohort, run cNMF to clean out these expression matrices, unify the genes to be included for the second round of cNMF, identify the gene expression programs in the myeloid cells of each discovery cohort and identifying the consensus cNMF programs across the discovery cohorts. The "Required Files" folder includes additional files to run these codes.

Generating Gene Expression heatmaps for the myeloid programs in MGB and McGill Cohorts (Related to Figure 1):

The folder "Figure 1 Visualizations" includes the necessary scripts and instructions to generate the gene expression heatmaps shown in Figure 1. This folder contains all the codes and instructions required to determine the frequency of the expression of the genes associated with the consensus programs and generate the heatmaps. The "Required Files" folder includes additional files to run these codes.

Generating the tumor-associated myeloid cells Quadrant Plots (Related to Figure 1):

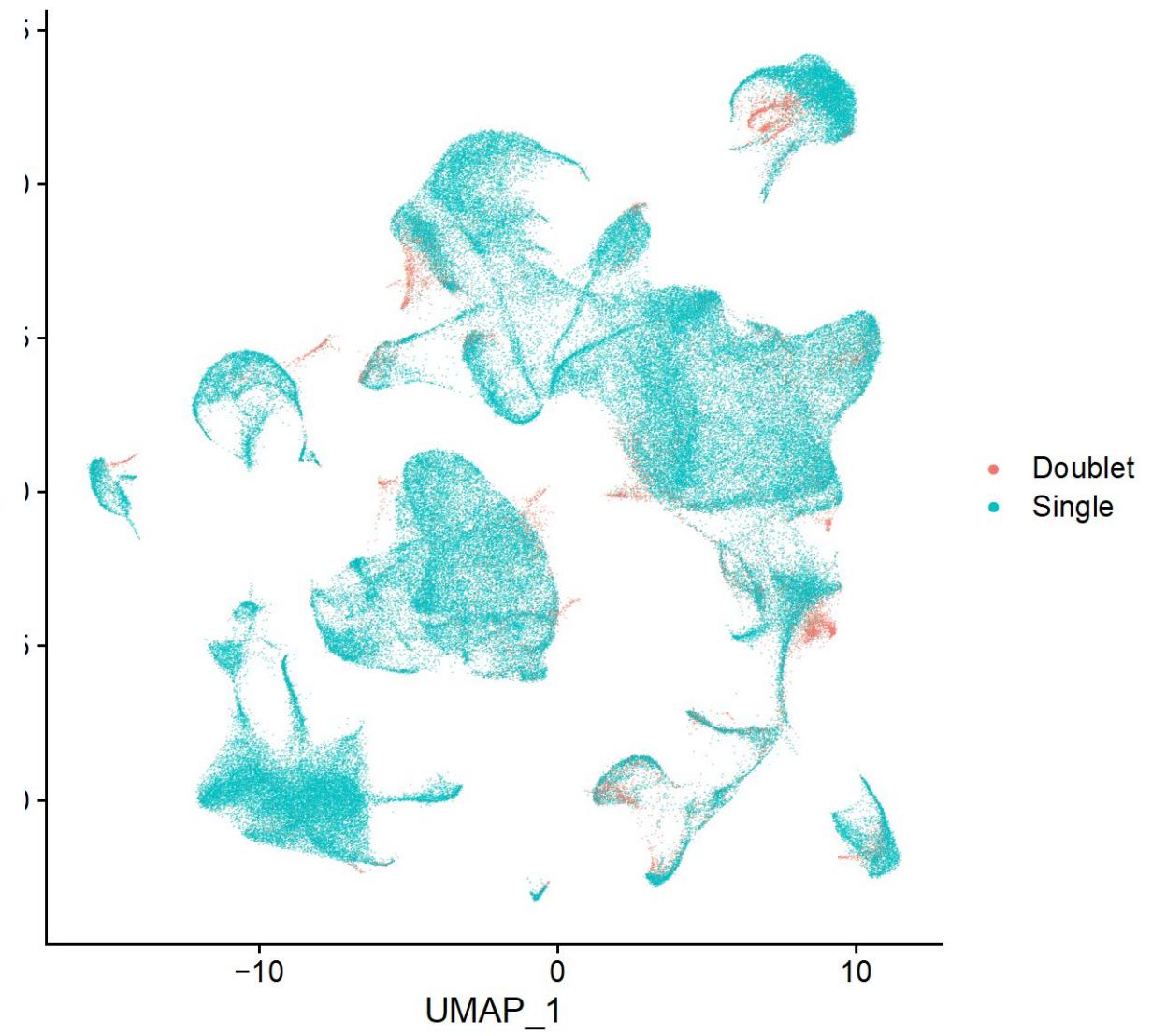
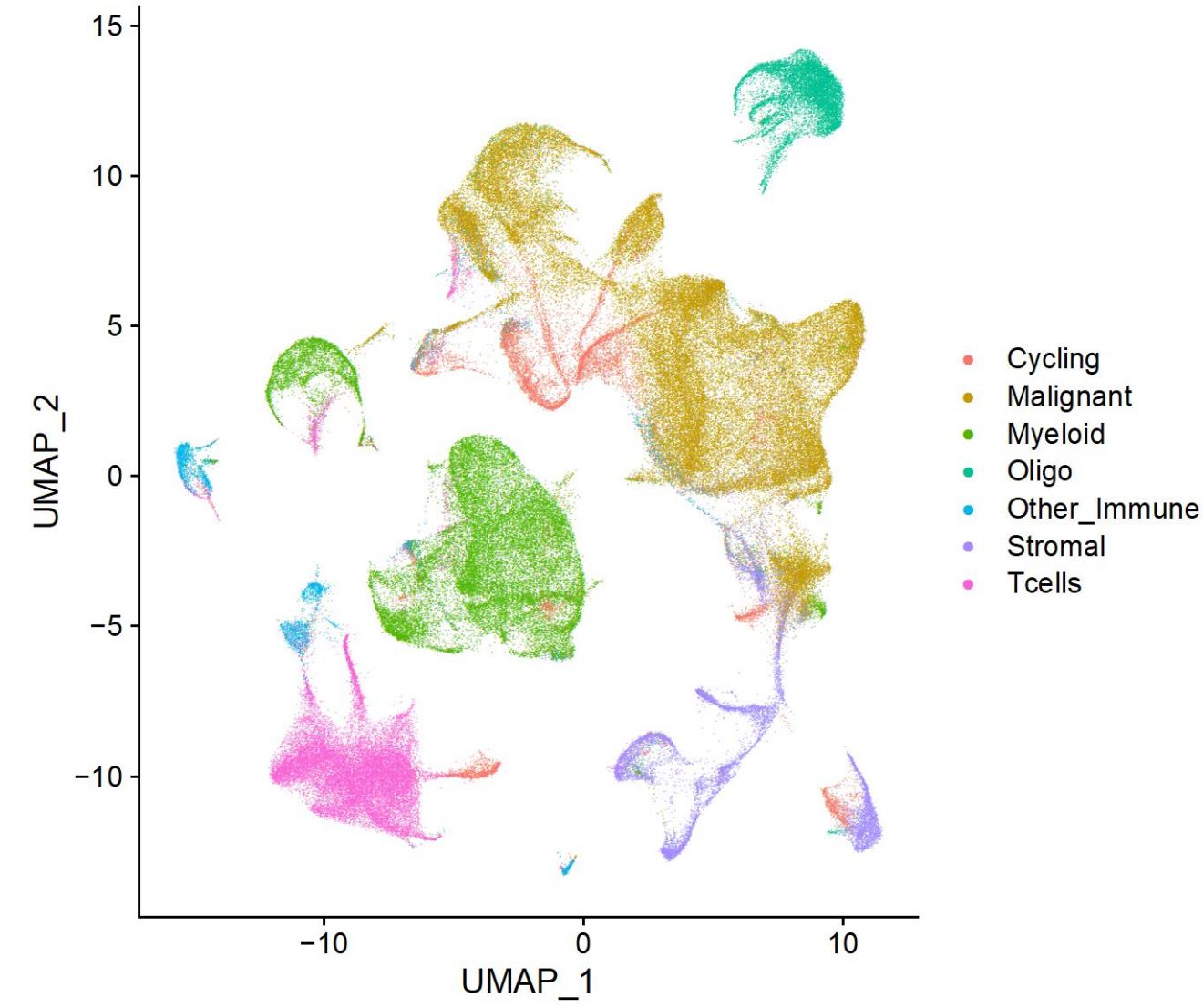
Doublet detection

Doublet Detection

- A cell should not show high usage for programs belonging to different categories
- Immune cells should not have CNV.

Secondary programs

Annotation	Malignant	Myeloid	Stromal	T cells	Oligo	Other_Immune
Malignant		30%	25%	25%	20%	20%
Myeloid	20%		20%	30%	20%	20%
Stromal	20%	20%		20%	20%	20%
T cells	20%	30%	20%		20%	20%
Oligo	40%	30%	20%	20%		20%
Other_Immune	30%	30%	30%	40%	30%	



- <https://bit.ly/3AJ24rb> - Bernstein Lab cNMF shinyApp github

- Standalone cNMF back-calculation tool implemented in R as a standalone web interface app
- Utilizes our existing gene expression program table, not de novo cNMF
- Also implementable as a cli tool using python

cNMF Program Usage Calculator

Choose Mode
 Annotation Mode
 Myeloid Program Calculation Mode

Choose Input Format
h5ad

Choose File
Browse... No file selected

[Download Output Data](#) [Download Log File](#)

Welcome to the cNMF program usage calculator for human glioma expression data

Instructions:

1. Choose the mode. Annotation Mode calculates the enrichment of cell types NMF programs to help you annotate the Cells. Myeloid Program Calculation Mode calculates the usages of the consensus cNMF programs in glioma-associated myeloid cells. Upload gene expression matrix of myeloid cells for Myeloid Program Calculation Mode

a. Choose input format after choosing the mode. The matrix should be in h5ad, csv, or mtx. For CSV, genes should be in rows and cells in columns. For mtx, you should upload the associated barcodes and feature files, which are outputs of CellRanger or STARSolo. Feature file has to have three columns, with the second column including gene symbols and the third column having the words Gene Expression in all rows

b. The matrix can be normalized or raw. mtx and associated files can be gz compressed or uncompressed. To make the process faster, upload gzipped mtx file

c. You can view an example of generating a myeloid h5ad matrix by clicking [here](#) and following the steps until you annotate the clusters. Then type the following:

```
adata_myeloid = adata[adata.obs['leiden']=='myeloid',:]  
adata.write_h5ad('Myeloid_Matrix.h5ad')
```

2. Once the file(s) is/are loaded, usage calculation will begin automatically.

3. Once the 'Calculating Usages' process is completed, the progress bar will disappear, and you can download the output file by clicking 'Download Output Data'. In case of an error, you can download the log file which can help in troubleshooting.

a. The output is automatically normalized by dividing each program usage score per cell by the sum of all usage scores for that cell and converting the values into percentages

If you would like to run this program locally (may be needed for large matrices), you can download the program by clicking [here](#) to visit the GitHub page of the tool.

This shiny app is developed and maintained by Chadi A. El Farran, M.Sc., Ph.D. (ChadiA_ElFarran@dfci.harvard.edu), with significant contributions from Charles P. Couturier, MD, Ph.D., and Tyler E. Miller, MD, Ph.D.

If you use this tool, please cite: Tyler E Miller, Chadi Abdul Kader El Farran, Charles P Couturier, et al., Programs, Origins, and Niches of Immunomodulatory Myeloid Cells in Gliomas. bioRxiv 2023.10.24.563466; doi: <https://doi.org/10.1101/2023.10.24.563466>

Terra

- <https://app.terra.bio>
- Cloud-based workspace environment for biomedical and bioinformatics analyses, workflows, and pipelines
- Developed by The Broad Institute of MIT and Harvard
- Utilizes WDL (workflow development language)
- Cloud compute performed by Google, or a private instance of Azure
- Expansive tutorial pages

Terra WORKSPACES Workspaces > TyMillerLab0/TyMillerLab0 > Dashboard

DASHBOARD DATA ANALYSES WORKFLOWS JOB HISTORY

ABOUT THE WORKSPACE 

The first Ty Miller Workspace (testing)

WORKSPACE INFORMATION

Last Updated	8/29/2024
Creation Date	8/13/2024
Access Level	Project Owner

CLOUD INFORMATION

Cloud Name	 Google Cloud
Location	 us-central1 (Iowa)
Google Project ID	
Bucket Name	
Estimated Storage Cost	
Updated on 8/30/2024	
Bucket Size	
Updated on 8/30/2024	
Open bucket in browser	
Open project in Google Cloud Console	

OWNERS

TAGS

NOTIFICATIONS

Rate: **\$0.04** per hour





WORKSPACES Workspaces > TyMillerLab0/TyMillerLab0 > Data

SHBOARD DATA ANALYSES WORKFLOWS JOB HISTORY ⋮

BLES Name Size Last modified
tables have been uploaded.
load TSV

ERENCE DATA All_SeqWell_220818/
references have been added.
d reference data

HER DATA All_SeqWell_220818_MetaData.csv 62 MB Aug 27, 2024
 All_SeqWell_220818_Raw_Expression.txt 20 GB Aug 27, 2024
 All_SeqWell_MetaData.csv 60 MB Aug 27, 2024

Workspace Data

Files

		Size	Last modified
<input type="checkbox"/>	All_Yun_220826_Raw_Expression.txt	20 GB	Aug 27, 2024
<input type="checkbox"/>	BWH10_Cumulus.csv	864 B	Aug 27, 2024
<input type="checkbox"/>	BWH10_Samples.txt	882 B	Aug 27, 2024
<input type="checkbox"/>	BWH10_WT_MetaData.csv	4 MB	Aug 27, 2024
<input type="checkbox"/>	BWH10_WT_Myeloid_Raw_Expression.txt	357 MB	Aug 27, 2024
<input type="checkbox"/>	BWH11_Cumulus.csv	864 B	Aug 27, 2024
<input type="checkbox"/>	BWH11_Mono_Cumulus.csv	510 B	Aug 27, 2024
<input type="checkbox"/>	BWH11_Mono_Samples.txt	2 KB	Aug 27, 2024
<input type="checkbox"/>	BWH11_Samples.txt	882 B	Aug 27, 2024

Rate: \$0.04 per hour

Upload New folder Delete Analyze in Seqr

Cloud icon

Jupyter icon

Logout icon



DASHBOARD

DATA

ANALYSES

WORKFLOWS

JOB HISTORY



Your Analyses

+ Start

Search analyses

Rate:
\$0.00
per hour

Application	Name	Last Modified	
Jupyter	MGH252_WT_Version2.ipynb	Aug 22, 2024	
Jupyter	Saving Breast Cancer Myeloid Expression matrix as h5ad Round 2.ipynb	Aug 20, 2024	
Jupyter	Saving Breast Cancer Myeloid Expression matrix as h5ad.ipynb	Aug 20, 2024	
Jupyter	BWH27_Classified_Myeloid_RNA_Velocity.ipynb	Aug 13, 2024	
Jupyter	Saving TA ASC SSL SSC data as h5ad.ipynb	Aug 9, 2024	
Jupyter	Saving MOuse Osteo data as h5ad-Copy1.ipynb	Jul 18, 2024	
Jupyter	Saving Myeloid Mouse Osteo data as h5ad.ipynb	Jul 18, 2024	
Jupyter	MGH240_GBO_Annotation_Trial_Final2_Oligo_Edit2-Copy1.ipynb	Jul 15, 2024	

WORKSPACES

Workspaces > gbm-chadi/cumulus > Workflows

DASHBOARD DATA ANALYSES WORKFLOWS JOB HISTORY

SEARCH WORKFLOWS Sort By: Alphabetical

Rate: \$0.00 per hour

WORKFLOWS

Workflow Name	Version	Source	Action
BuildBamIndex	V. 2	Source: Terra	(⋮)
cellphonedb	V. 5	Source: Terra	(⋮)
Cellranger	V. 2.4.1	Source: Dockstore	(⋮)
cibersortx_fractions	V. 1	Source: Terra	(⋮)
cNMF	V. 20	Source: Terra	(⋮)
cnmf_parallel	V. 7	Source: Terra	(⋮)
Cumulus	V. 2.4.1	Source: Dockstore	(⋮)
fractions	V. 11	Source: Terra	(⋮)
infercnv	V. 2	Source: Terra	(⋮)
SCENIC	V. 13	Source: Terra	(⋮)
SCENIC_R4_Part1	V. 1	Source: Terra	(⋮)
SortSam	V. 1	Source: Terra	(⋮)
STARsolo	V. 2.4.0	Source: Dockstore	(⋮)
STARsolo_create_reference	V. 2.1.1	Source: Dockstore	(⋮)



0

DASHBOARD

DATA

ANALYSES

WORKFLOWS

JOB HISTORY



Search

Rate:
\$0.00
per hour

Submission (click for details)	Data entity	No. of Workflows	Status	Submitted ↑	Submission ID	Comment	Actions
mparikh/cnmf_parallel Submitted by celfarr@broadinstitute.org		1	✓ Done	Aug 20, 2024 4:47 PM	883d18ea-f9e2-425d-9261-ca347980f473		⋮
mparikh/cnmf_parallel_CvF920elJ4o Submitted by celfarr@broadinstitute.org		1	✓ Done	Aug 20, 2024 4:46 PM	36111f30-f22e-482c-8alc-891bb0d4791c		⋮
mparikh/cnmf_parallel_kVHq7Ot0Co4 Submitted by celfarr@broadinstitute.org		1	✓ Done	Aug 20, 2024 3:27 PM	dc8201bb-ca45-4d1e-94b9-45700052ab07		⋮
mparikh/cnmf_parallel_hAw6_FNiZaY Submitted by celfarr@broadinstitute.org		1	✓ Done	Aug 12, 2024 11:28 AM	3c26eca2-6b70-42e3-8d07-21f091c8c91c		⋮
mparikh/cnmf_parallel_nrF7dE8a25E Submitted by celfarr@broadinstitute.org		1	✓ Done	Aug 9, 2024 4:34 PM	25d0c4bb-5e51-4191-a1bf-cc2756ca880e		⋮
mparikh/cnmf_parallel_YeOpghErAal Submitted by celfarr@broadinstitute.org		1	✓ Done	Aug 7, 2024 4:10 PM	03d9a89f-d8a6-4e93-85ae-007e84d4990c		⋮
mparikh/cnmf_parallel_TT6boN3E4Q Submitted by celfarr@broadinstitute.org		1	✓ Done	Aug 7, 2024 1:32 PM	583ea16b-b774-4e8c-bdbd-881f38e58364		⋮
mparikh/cnmf_parallel_ExOK8TOW7IM Submitted by celfarr@broadinstitute.org		1	✓ Done	Aug 2, 2024 4:35 PM	539c30f2-5ee4-42a3-a056-e9af1f7a4d8d		⋮
mparikh/cnmf_parallel_879YISZlqZE Submitted by celfarr@broadinstitute.org		1	✓ Done	Aug 2, 2024 3:24 PM	b9a9c9c7-c292-4d39-804b-7f00ce781337		⋮
mparikh/cnmf_parallel_SLPZ-VZpdZo Submitted by celfarr@broadinstitute.org		1	⚠ Done	Aug 2, 2024 1:54 PM	f86a31d3-d4a9-466b-b797-185968d53c35		⋮
mparikh/cnmf_parallel_yKXoR1_kTb4 Submitted by celfarr@broadinstitute.org		1	✓ Done	Aug 2, 2024 8:58 AM	7db49515-a95e-4a02-91d0-950570356f23		⋮
mparikh/cnmf_parallel_YV8ALyiHHYU Submitted by celfarr@broadinstitute.org		1	✓ Done	Aug 1, 2024 2:52 PM	ce74e8fe-cecf-4cb2-8f71-3d5ccb6aa80f		⋮
mparikh/cnmf_parallel_N0mWKNyty4Y Submitted by celfarr@broadinstitute.org		1	✓ Done	Jul 23, 2024 1:34 PM	9a6b4ae5-ac51-4819-8d10-0e02f2cb2498		⋮
mparikh/cnmf_parallel_p5wIELAEyrM Submitted by celfarr@broadinstitute.org		1	✓ Done	Jul 19, 2024 11:49 AM	53c9d08c-909c-4d76-945f-44b46edff1e0		⋮



DASHBOARD

DATA

ANALYSES

WORKFLOWS

JOB HISTORY



Search

Rate:
\$0.00
per hour

Submission (click for details)	Data entity	No. of Workflows	Status	Submitted ↑	Submission ID	Comment	Actions
mparikh/cnmf_parallel Submitted by celfarra@broadinstitute.org		1	✓ Done	Aug 20, 2024 4:47 PM	883d18ea-f9e2-425d-9261-ca347980f473		⋮
mparikh/cnmf_parallel_CvF920elJ4o Submitted by celfarra@broadinstitute.org		1	✓ Done	Aug 20, 2024 4:46 PM	36111f30-f22e-482c-8alc-891bb0d4791c		⋮
mparikh/cnmf_parallel_kVHq7Ot0Co4 Submitted by celfarra@broadinstitute.org		1	✓ Done	Aug 20, 2024 3:27 PM	dc8201bb-ca45-4d1e-94b9-45700052ab07		⋮
mparikh/cnmf_parallel_hAw6_FNiZaY Submitted by celfarra@broadinstitute.org		1	✓ Done	Aug 12, 2024 11:28 AM	3c26eca2-6b70-42e3-8d07-21f091c8c91c		⋮
mparikh/cnmf_parallel_nrF7dE8a25E Submitted by celfarra@broadinstitute.org		1	✓ Done	Aug 9, 2024 4:34 PM	25d0c4bb-5e51-4191-a1bf-cc2756ca880e		⋮
mparikh/cnmf_parallel_YeOpghErAal Submitted by celfarra@broadinstitute.org		1	✓ Done	Aug 7, 2024 4:10 PM	03d9a89f-d8a6-4e93-85ae-007e84d4990c		⋮
mparikh/cnmf_parallel_TT6boN3E4Q Submitted by celfarra@broadinstitute.org		1	✓ Done	Aug 7, 2024 1:32 PM	583ea16b-b774-4e8c-bdbd-881f38e58364		⋮
mparikh/cnmf_parallel_ExOK8TOW7IM Submitted by celfarra@broadinstitute.org		1	✓ Done	Aug 2, 2024 4:35 PM	539c30f2-5ee4-42a3-a056-e9af1f7a4d8d		⋮
mparikh/cnmf_parallel_879YISZlqZE Submitted by celfarra@broadinstitute.org		1	✓ Done	Aug 2, 2024 3:24 PM	b9a9c9c7-c292-4d39-804b-7f00ce781337		⋮
mparikh/cnmf_parallel_SLPZ-VZpdZo Submitted by celfarra@broadinstitute.org		1	⚠ Done	Aug 2, 2024 1:54 PM	f86a31d3-d4a9-466b-b797-185968d53c35		⋮
mparikh/cnmf_parallel_yKXoR1_kTb4 Submitted by celfarra@broadinstitute.org		1	✓ Done	Aug 2, 2024 8:58 AM	7db49515-a95e-4a02-91d0-950570356f23		⋮
mparikh/cnmf_parallel_YV8ALyiHHYU Submitted by celfarra@broadinstitute.org		1	✓ Done	Aug 1, 2024 2:52 PM	ce74e8fe-cecf-4cb2-8f71-3d5ccb6aa80f		⋮
mparikh/cnmf_parallel_N0mWKNyty4Y Submitted by celfarra@broadinstitute.org		1	✓ Done	Jul 23, 2024 1:34 PM	9a6b4ae5-ac51-4819-8d10-0e02f2cb2498		⋮
mparikh/cnmf_parallel_p5wIELAEyrM Submitted by celfarra@broadinstitute.org		1	✓ Done	Jul 19, 2024 11:49 AM	53c9d08c-909c-4d76-945f-44b46edff1e0		⋮

[← Terra on Google Cloud](#) Enter a question, topic or keyword...[Terra Support](#) / [Terra on Google Cloud](#) / [Getting Started \(GCP\)](#) / [Working in Terra on GCP](#) / [Quickstart tutorials](#)

Categories

Platform-wide
(security, status and
updates)

Terra on Google
Cloud

[Getting Started \(GCP\)](#)

[Account and billing](#)

[Managing Access/
Billing](#)

[Managing Cloud
costs](#)

[Workspaces](#)

[Data](#)

[Terra Data Repository](#)

[Workflows](#)

Quickstart tutorials

Curated workspaces to help get you started with hands-on practice. Learn about data tables, workflows, and notebooks in the Terra on GCP Quickstart workspace.

[Follow](#)

New to Terra? Try the T101 Quickstart tutorials

Walk through the steps in a mock research study to learn the basics of Terra's functionality. Three Quickstart tutorials build on each other. Each takes about half an hour or so and costs less than a dollar to complete.

Hands-on practice to learn how to

- **Access and organize data** using workspace data tables
- Run a **workflow** for bulk analysis
- Run an **interactive Jupyter Notebook** to plot data

Quickstart 1:
[Data Tables](#)



Quickstart 2:
[Workflows](#)



Quickstart 3:
[Notebooks](#)

Tutorial Guides

<http://bit.ly/3AKLQ0Q>

 [Terra on Google Cloud](#) Enter a question, topic or keyword...[Terra Support](#) / [Terra on Google Cloud](#) / [Workflows](#) / [WDL Resources](#)

Categories

Platform-wide
(security, status and
updates)

Terra on Google Cloud

[Getting Started \(GCP\)](#)[Account and billing](#)[Managing Access/
Billing](#)[Managing Cloud
costs](#)[Workspaces](#)[Data](#)[Terra Data Repository](#)[Workflows](#)[Interactive Analysis](#)[Working with
Containers \(Docker\)](#)

Overview: Getting started with WDL



Kate Herman
Updated 23 days ago

[Follow](#)

In this article

[WDL: A community-driven data processing language](#)[WDL and Terra: Partners in pipelining](#)[Developing WDL workflows: "Hello, learn-wdl!"](#)[Navigating the learn-wdl GitHub](#)[New to computing? No problem! Try these high-level overviews](#)[Additional resources](#)[Next steps](#)

The [Workflow Description Language](#) (WDL, pronounced 'widdle') is a community-driven, human-readable language for data processing workflows. Whether you want to use WDL in Terra or develop your own WDL workflows, we have the resources to get you started!

Learn to use WDL in Terra with our online course

Check out our [free online course](#) on [Writing WDL Workflows in Terra](#) to learn more about how to write Terra workflows in WDL. The course includes hands-on exercises that guide you through writing basic WDL scripts, pre-configuring and running workflows, and customizing your workflow's Docker containers.

<https://bit.ly/3Z6rZTV>

Google cloud billing

- <https://bit.ly/3yVSiBD>
- New google cloud accounts get \$300 to test with
- Requires link to credit card or bank account, or third-party vendor to handle billing
- Can use NIH STRIDES credits

Categories

Platform-wide
(security, status and
updates) ▾

Terra on Google Cloud ^

[Getting Started \(GCP\)](#)

[Account and billing](#)

[Managing Access/
Billing](#)

[Managing Cloud
costs](#)

[Workspaces](#)

[Data](#)

[Terra Data Repository](#)

[Workflows](#)

[Interactive Analysis](#)

[Working with
Containers \(Docker\)](#)

[Troubleshooting](#)

[About](#)

How to set up billing in Terra (GCP)



Anton Kovalsky
Updated 1 month ago



Follow

In this article

[Step 1. Set up a Google Cloud Billing account](#)

[Step 2. Link the Cloud Billing account to Terra](#)

[Step 3. Create a Terra Billing project](#)

[Next steps and billing resources](#)

Read on for step-by-step instructions to set up billing so you can work in Terra GCP. Once you set up your billing, Terra takes care of interfacing with Google Cloud, so many funding and resources management tasks can be done in Terra.

Other options for billing

Do you have STRIDES credits?

See [How to access STRIDES](#) for step-by-step instructions.

\$300 in Google Cloud credits for new Google Cloud Billing accounts

If you've never logged into the Google Cloud console or set up a Google Cloud billing account, you are eligible for \$300 in free Google Cloud credits you can use for working in Terra.

Go to [Claim \\$300 Google credits to explore Terra](#). Down the line, you convert

gcloud cli installation

- <https://cloud.google.com/sdk/docs/install>

The screenshot shows the Google Cloud SDK Guides page. The left sidebar has a 'Cloud SDK' tab selected, followed by 'Guides' (which is underlined), 'Reference', 'Support', and 'Resources'. Under 'Guides', there's a 'Filter' button and a list of topics: 'gcloud CLI' (with sub-links 'Product overview', 'gcloud CLI overview', 'gcloud CLI cheat sheet'), 'Quickstart' (with sub-link 'Install the Google Cloud CLI'), 'How-to guides' (with sub-links 'All how-to guides', 'Installing the gcloud CLI' (which is expanded to show 'Recommended installation' and sub-links 'Other installation methods', 'Setting up the gcloud CLI', 'Managing gcloud CLI components', 'Scripting gcloud CLI commands', 'Enabling accessibility features', 'Using gcloud interactive shell', 'Uninstalling the gcloud CLI')), and 'Cloud SDK' (with sub-links 'Documentation', 'Technology areas', 'Cross-product tools', 'Related sites'). A search bar is at the top right. The main content area shows the 'Install the gcloud CLI' guide. It includes a breadcrumb trail 'Cloud SDK > Documentation > Guides', a 'Was this helpful?' button, a 'Send feedback' button, and a 'Install the gcloud CLI' title with a bookmark icon. The text explains the page contains instructions for choosing and maintaining a Google Cloud CLI installation, mentioning the `gcloud`, `gsutil`, and `bq` command-line tools. It also links to 'All features' and 'Cloud Client Libraries'. Below this is a section titled 'Installation instructions' with a note about proxy settings. At the bottom, tabs for 'Linux', 'Debian/Ubuntu', 'Red Hat/Fedora/CentOS', 'macOS', and 'Windows' are shown, with 'Windows' being the active tab. The Windows section notes that the Google Cloud CLI works on Windows 8.1 and later and Windows Server 2012 and later. It provides steps to download the installer and run PowerShell commands, including a code block:

```
(New-Object Net.WebClient).DownloadFile("https://dl.google.com/dl/cloudsdk/channels/rapid/rapid-release/google-cloud-sdk.exe")  
& $env:Temp\GoogleCloudSDKInstaller.exe
```

On the right side, a 'On this page' sidebar lists 'Installation instructions', 'Other installation options', 'Manage an installation', 'Earlier versions of the gcloud CLI', and 'Supported Python versions'.