AgBioData

Toward enhanced genomics, genetics, and breeding research
outcomes through standardization of practices and protocols
across agricultural databases

Developing FAIR open-source interactive web browsers for individual labs to explore single-cell data: Shiny-PIGGI as an example from the AgBioData Single-Cell Working Group.

Muskan Kapoor, Sunita Kumari, Doreen Ware, Christopher K. Tuggle

Plant and Animal Genome, January 2026

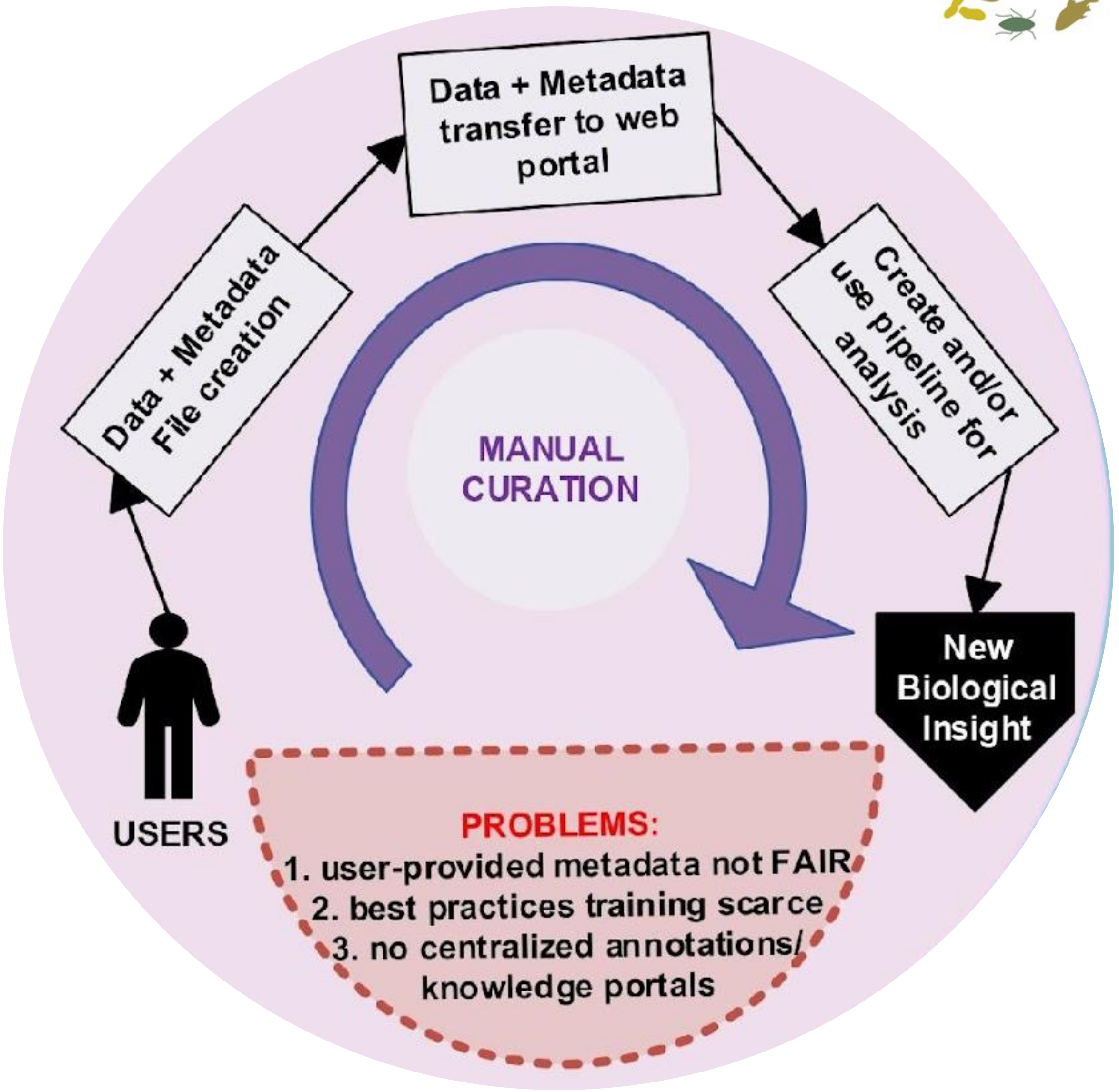
Single-cell Biocuration Working Group– 22 members

	NAME	INSTITUTE
1	Sunita Kumari (Chair)	Cold Spring Harbor Lab
2	Doreen Ware	USDA, Cold Spring Harbor Lab
3	Benjamin Cole	Joint Genome Institute
4	Sharon Greenblum	Join Genome Institute
5	Rex Nelson	USDA
6	Sushma Naithani	Oregon State University
7	Fabian Andres Salinas	University of Florida
8	Teng-Jui(Owen) Lin	UC Berkeley
9	Jennifer Clarke	University of Nebraska-Lincoln
10	Laurel Cooper	Planteome
11	Irene Papatheodorou	Earlham Institute, UK
12	Christopher Tuggle (Chair)	Iowa State University
13	Muskan Kapoor	Iowa State University
14	Mehak Kapoor	Iowa State University
15	Kimberly Davenport	Washington State University
16	Elisabetta Giuffra	INRAE
17	Theros Ng	Western University of Health Sciences
18	Doreen Becker	Farm Animal Biology(FBN)
19	Praveen Krishna	Farm Animal Biology(FBN)
20	Damien Goutte-Gattat	Fly-Base
21	Bjorn Coleis	Friedrich-Loeffler Institut
22	Felix Shaw	Earlham Institute, UK

AgBioData Mission: Why FAIR Resources?

Sample Metadata FAIR Barriers

- Missing, corrupt, incomplete or inaccurate data
- Incomprehensible metadata, unverified results
- Incompatibility with FAIR standards
- Best practices training scarce
- Lack of knowledge portals focusing on centralized annotations
- Requires manual curation on every step
- **Thus, important need to curate and document metadata adhering to FAIR principles**



Data + Metadata File creation

Data + Metadata transfer to web portal

Create and/or use pipeline for analysis

New Biological Insight

MANUAL CURATION

USERS

PROBLEMS:
1. user-provided metadata not FAIR
2. best practices training scarce
3. no centralized annotations/ knowledge portals

Mission of Single cell Biocuration Working Group

1. **Address annotation challenges** of scRNAseq metadata in plants and animals

2. Facilitate **creation of cohesive community** that provides FAIR metadata standards, sc resources and tools

3. **Establish a set of recommendations** for member databases or management of sc datasets

4. Standardize sc datasets for **accurate cell type specific annotations** and other benchmarking studies.

5. Create **FAIR agricultural genomics** user-friendly data ecosystem.

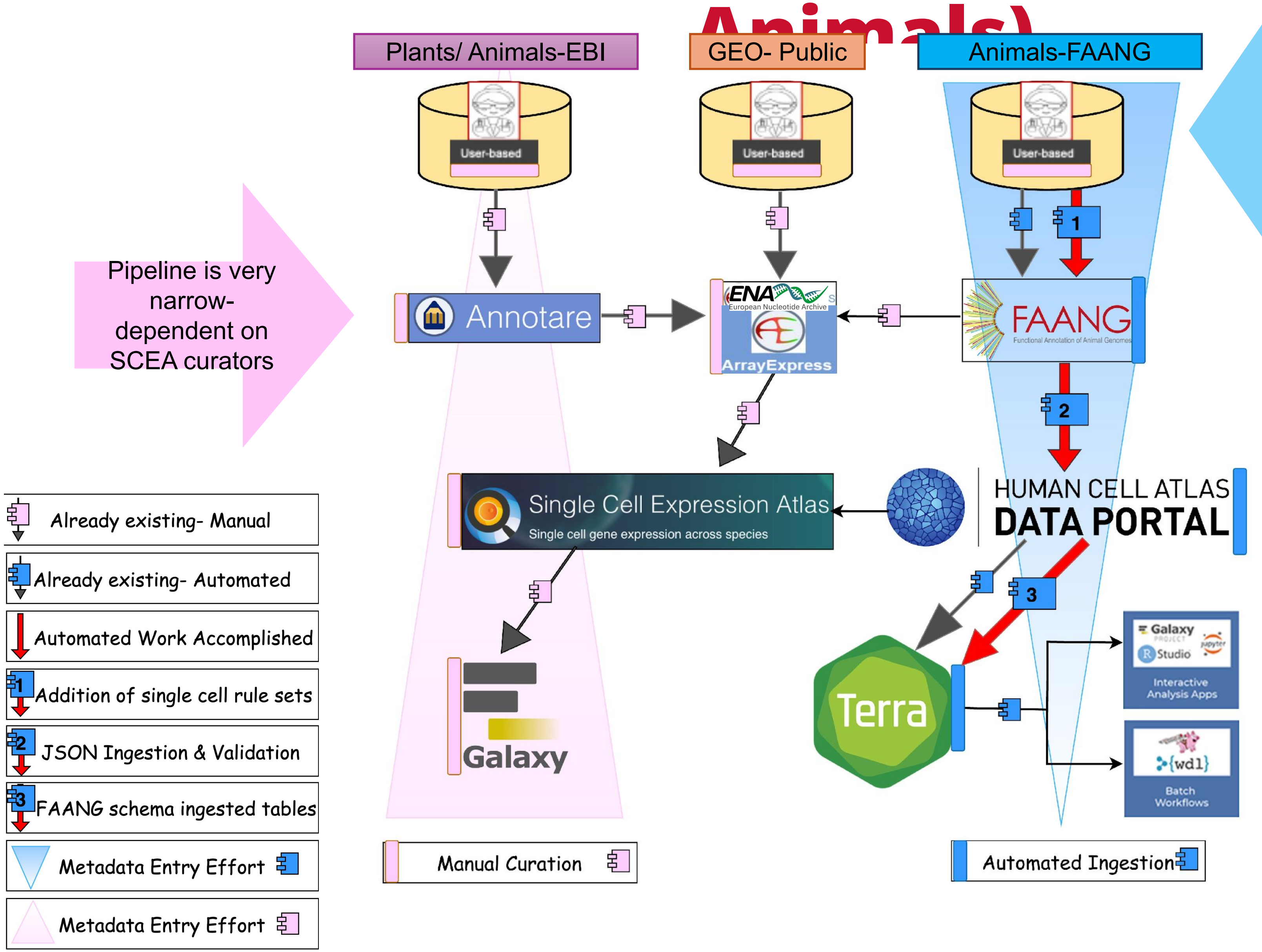
AG2PI
Agricultural Genome to Phenome Initiative

FAANG
Functional Annotation of Animal Genomes

PLANT CELL ATLAS

AgBioData
Toward enhanced genomics, genetics, and breeding research outcomes through standardization of practices and protocols across agricultural databases

Address Annotation Challenges (Plants & Animals)

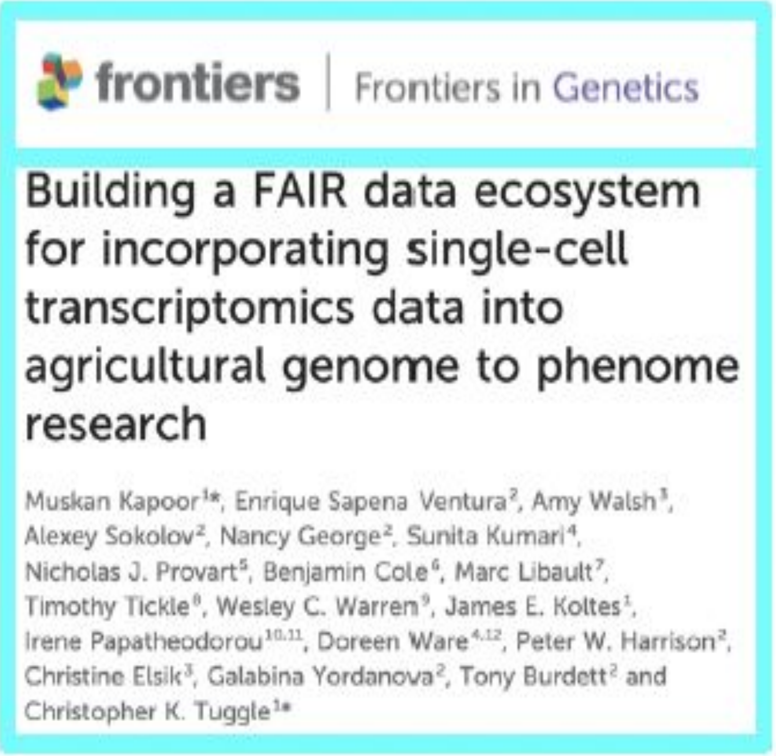


Pipeline is very narrow- dependent on SCEA curators

Data creator provides rich metadata

- Already existing- Manual
- Already existing- Automated
- Automated Work Accomplished
- 1 Addition of single cell rule sets
- 2 JSON Ingestion & Validation
- 3 FAANG schema ingested tables
- Metadata Entry Effort
- Metadata Entry Effort

Scan: GitHub AG2PI


frontiers | Frontiers in Genetics

Building a FAIR data ecosystem for incorporating single-cell transcriptomics data into agricultural genome to phenome research

Muskan Kapoor[1*], Enrique Sapena Ventura[2], Amy Walsh[3], Alexey Sokolov[2], Nancy George[2], Sunita Kumari[4], Nicholas J. Provart[5], Benjamin Cole[6], Marc Libault[7], Timothy Tickle[8], Wesley C. Warren[9], James E. Koltes[1], Irene Papatheodorou[10,11], Doreen Ware[4,12], Peter W. Harrison[2], Christine Elsik[3], Galabina Yordanova[2], Tony Burdett[2] and Christopher K. Tuggle[1*]

Mission of Single cell Biocuration Working Group

1. **Address annotation challenges** of scRNAseq metadata in plants and animals

2. Facilitate **creation of cohesive community** that provides FAIR metadata standards, sc resources and tools

3. **Establish a set of recommendations** for member databases or management of sc datasets

4. Standardize sc datasets for **accurate cell type specific annotations** and other benchmarking studies.

5. Create **FAIR agricultural genomics** user-friendly data ecosystem.

AG2PI
Agricultural Genome to Phenome Initiative

FAANG
Functional Annotation of Animal Genomes

PLANT CELL ATLAS

AgBioData
Toward enhanced genomics, genetics, and breeding research outcomes through standardization of practices and protocols across agricultural databases

Creation of cohesive community

USDA NIFA grant was awarded

Overall Goal - To develop a robust and sustainable network of researchers who are interested in developing SC genomics techniques and best practices to apply to agricultural systems.

Crops and Livestock community together

AgBioData, AG2PI, PCA, FAANG, AnGenMap, Gramene, Earlham Institute

Aim 1. Create a community of researchers interested in applying SC genomics to elucidate genotype to phenotype in agricultural systems.
Organized 4 virtual SC sessions

Aim 2. Determine community needs and potential resources for applying SC genomics techniques to agricultural systems.
Community Survey

Aim 3. Organize and convene a conference on discussion and development of SC resources required for agriculture.
In person Workshop and Community Discussion

Aim 4. Broadly disseminate conference findings and develop resources for community infrastructure.
Ongoing

Organizing Committee

- Chris Tuggle (Iowa State University) AG2PI
- Sunita Kumari, Doreen Ware (CSHL)
- Fiona McCarthy (Univ of Arizona) FAANG
- Jennifer Clarke (Univ of Nebarska)
- Ben Cole, Mark Libault (PCA)
- Wes Warren (Univ of Missouri)
- Irene Papatheodorou (Earlham Institute, UK)

227 people for webinars
48 people for in person workshop
15 key speakers across US & Europe
Plant and animal research community

Single-Cell Workshop held at AGBT (March 29-30), Orlando

Session 1: Building a transdisciplinary community in agricultural SC genomics

- *Lessons learned from the interdisciplinary projects: core values, orientation and organization.*
- *Community Survey Report.*

Session 2: Where are we in plant and animal SC genomics, as well as where we need to be?

- *Data management and infrastructure within the Plant Cell Atlas Community.*
- *Data and metadata management for reusing and integrating cell atlases.*
- *A Single nuclei and cell perspective on the chicken immune system.*

Single-Cell Workshop held at AGBT (March 29-30), Orlando

Small Group Breakout Discussions

Session 3: Identifying best practices for:

- Describing, storing and sharing FAIR data.
- Cell type annotations
- Identifying and benchmarking relevant projects

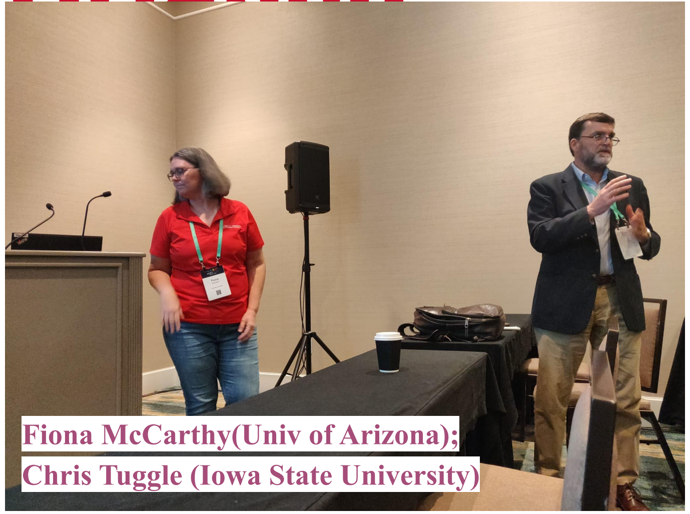
Session 4: Achieving identified best practices for Ag community

- Creating FAIR Data storage/sharing
- Creating common environments for data analysis
- Organizing benchmarking and training needs for data analysis pipelines for agriculture.

Community Discussions

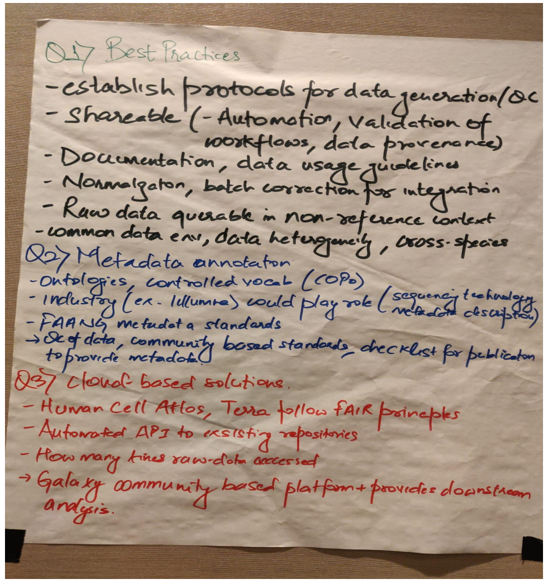
Session 3: Identifying opportunities & mechanisms to generate resources for community needs

- What are additional needs for creating G2P solutions at the cell type level?
- Offer opportunity to govt funding agencies to indicate where these ideas fit within their portfolios.
- Identify writers of white paper and topics to USDA and other agencies



Fiona McCarthy(Univ of Arizona);
Chris Tuggle (Iowa State University)

Irene Papatheodorou (Earlham Institute)



Q1> Best Practices

- establish protocols for data generation/QC
- Shareable (- Automation, Validation of workflows, data provenance)
- Documentation, data usage guidelines
- Normalization, batch correction for integration
- Raw data querable in non-reference context
- common data env, data heterogeneity, cross-species

Q2> Metadata annotation

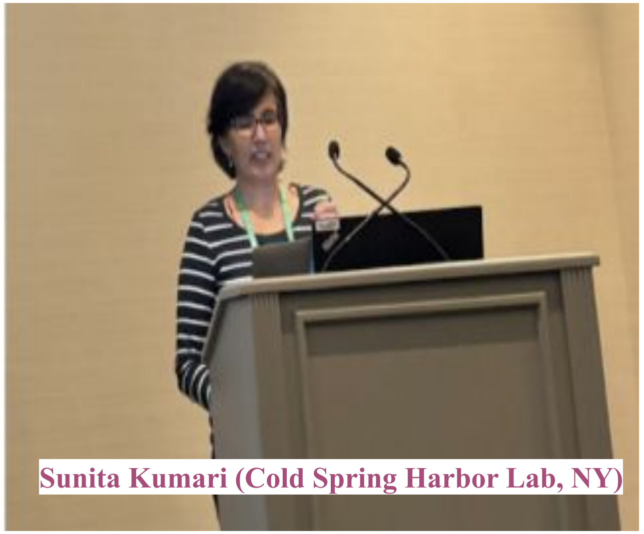
- Ontologies, controlled vocab (COPo)
- Industry (ex. Illumina) could play role (sequencing technology metadata description)
- FAANG metadata standards

→ QC of data, community based standards, checklist for publication to provide metadata

Q3> cloud-based solutions.

- Human Cell Atlas, Terra follow FAIR principles
- Automated API to existing repositories
- How many times raw-data accessed

→ Galaxy community based platform + provides downstream analysis.



Sunita Kumari (Cold Spring Harbor Lab, NY)

Mission of Single cell Biocuration Working Group

1. **Address annotation challenges** of scRNAseq metadata in plants and animals

2. Facilitate **creation of cohesive community** that provides FAIR metadata standards, sc resources and tools

3. **Establish a set of recommendations** for member databases or management of sc datasets

4. Standardize sc datasets for **accurate cell type specific annotations** and other benchmarking studies.

5. Create FAIR agricultural genomics user-friendly data ecosystem.

AG2PI
Agricultural Genome to Phenome Initiative

FAANG
Functional Annotation of Animal Genomes

PLANT CELL ATLAS

AgBioData
Toward enhanced genomics, genetics, and breeding research outcomes through standardization of practices and protocols across agricultural databases

Established a set of recommendations

DATA & METADATA STANDARDS

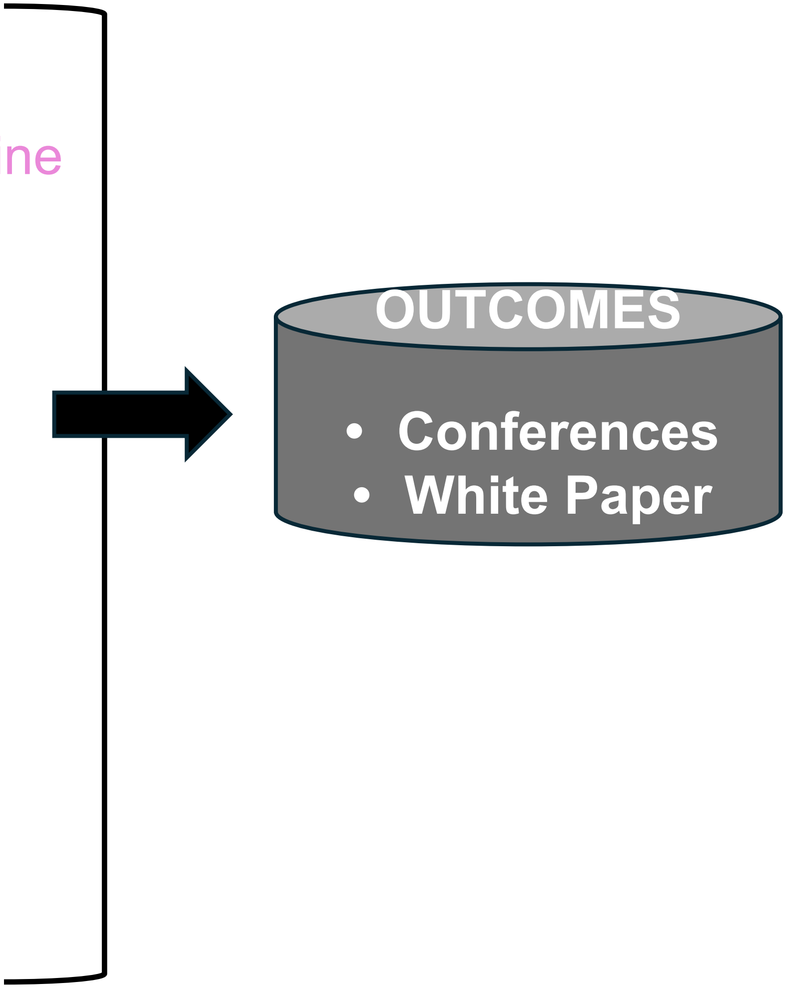
- Define and advertise minimal metadata standards
- Literature curation and meetings with repositories to define metadata standards across ag species
- Survey agencies or other gatekeepers to advertise

CELL TYPE ANNOTATION

- Identify/list high quality publishes SC datasets
- Document cell atlas for use in future cell annotation
- Coordination with PCA members

BENCHMARKING

- Identify datasets and pipelines for specific goals
- Identify metrics for evaluation of pipelines



OUTCOMES

- Conferences
- White Paper

Mission of Single cell Biocuration Working Group

1. **Address annotation challenges** of scRNAseq metadata in plants and animals

2. Facilitate **creation of cohesive community** that provides FAIR metadata standards, sc resources and tools

3. **Establish a set of recommendations** for member databases or management of sc datasets

4. Standardize sc datasets for **accurate cell type specific annotations** and other benchmarking studies.

5. Create **FAIR agricultural genomics** user-friendly data ecosystem.

AG2PI
Agricultural Genome to Phenome Initiative

FAANG
Functional Annotation of Animal Genomes

PLANT CELL ATLAS

AgBioData
Toward enhanced genomics, genetics, and breeding research outcomes through standardization of practices and protocols across agricultural databases

Cell Type Annotation Jamboree

Cell Type Annotation Challenge – Inconsistent cell type annotations

- Coordinated with Plant Cell Atlas members and worked on identifying high-quality published scRNA-seq datasets
- Predicted the plant ontology (PO) annotations using several AI tools
- Held collaborative cell type annotation Jamboree with Single Cell experts at Plant Cell Atlas meeting where five teams manually reviewed and annotated around 300 cell types with PO annotations.
- This hands-on work is also useful in training the AI models that will ultimately automate this process.

Sunita Kumari, CSHL

Ben Cole, JGI

Sharon Greenblum, JGI

Doreen Ware, USDA-ARS

Coordination with PCA experts

Mission of Single cell Biocuration Working Group

1. **Address annotation challenges** of scRNAseq metadata in plants and animals

2. Facilitate **creation of cohesive community** that provides FAIR metadata standards, sc resources and tools

3. **Establish a set of recommendations** for member databases or management of sc datasets

4. Standardize sc datasets for **accurate cell type specific annotations** and other benchmarking studies.

5. Create **FAIR agricultural genomics** user-friendly data ecosystem.

AG2PI
Agricultural Genome to Phenome Initiative

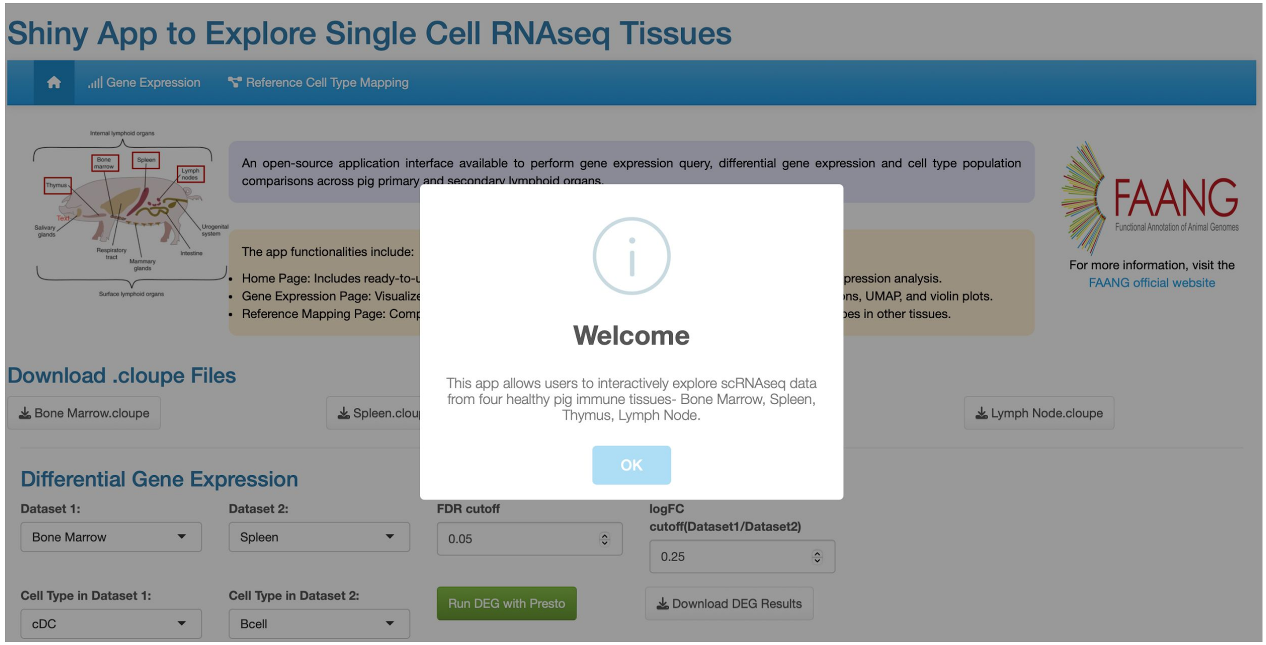
FAANG
Functional Annotation of Animal Genomes

PLANT CELL ATLAS

AgBioData
Toward enhanced genomics, genetics, and breeding research outcomes through standardization of practices and protocols across agricultural databases

Create FAIR agricultural genomics user-friendly data ecosystem – Shiny-PIGGI

- Facilitates easy visualization and reusability across different platforms
- Can be used by both computational/wet-lab scientist for visualization purposes
- Reproducible with easy installation and steps to implement guide on GitHub
- Cell-Type Specific DGE across ag species
- Gene Expression Query across ag species
- Reference Cell type Mapping across ag species



Shiny App to Explore Single Cell RNAseq Tissues

An open-source application interface available to perform gene expression query, differential gene expression and cell type population comparisons across pig primary and secondary lymphoid organs.

The app functionalities include:

- Home Page: Includes ready-to-u … pression analysis.
- Gene Expression Page: Visualize … ns, UMAP, and violin plots.
- Reference Mapping Page: Comp … es in other tissues.

For more information, visit the FAANG official website

Welcome

This app allows users to interactively explore scRNAseq data from four healthy pig immune tissues- Bone Marrow, Spleen, Thymus, Lymph Node.

Download .cloupe Files

Differential Gene Expression

Dataset 1: Bone Marrow

Dataset 2: Spleen

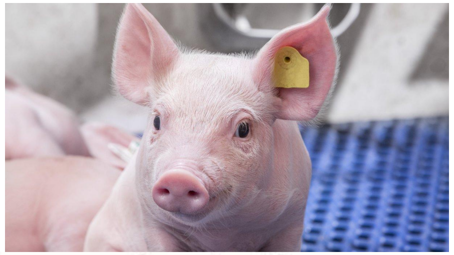
FDR cutoff: 0.05

logFC cutoff(Dataset1/Dataset2): 0.25

Cell Type in Dataset 1: cDC

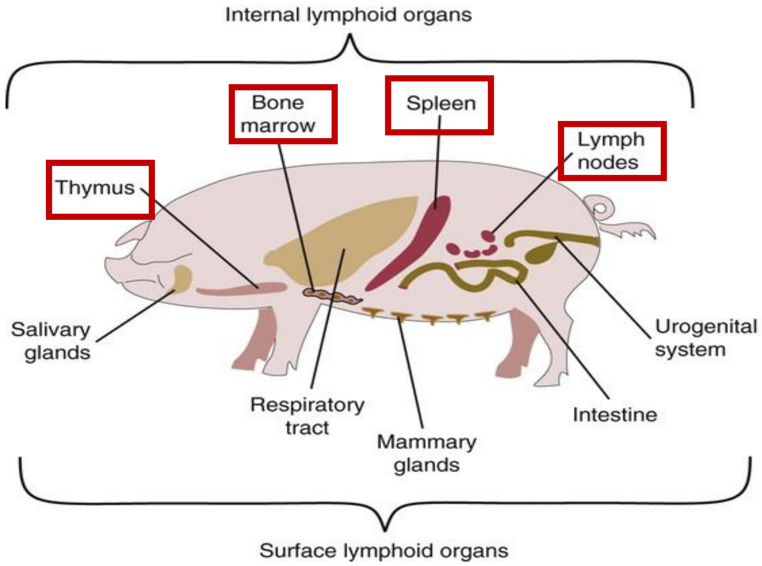
Cell Type in Dataset 2: Bcell

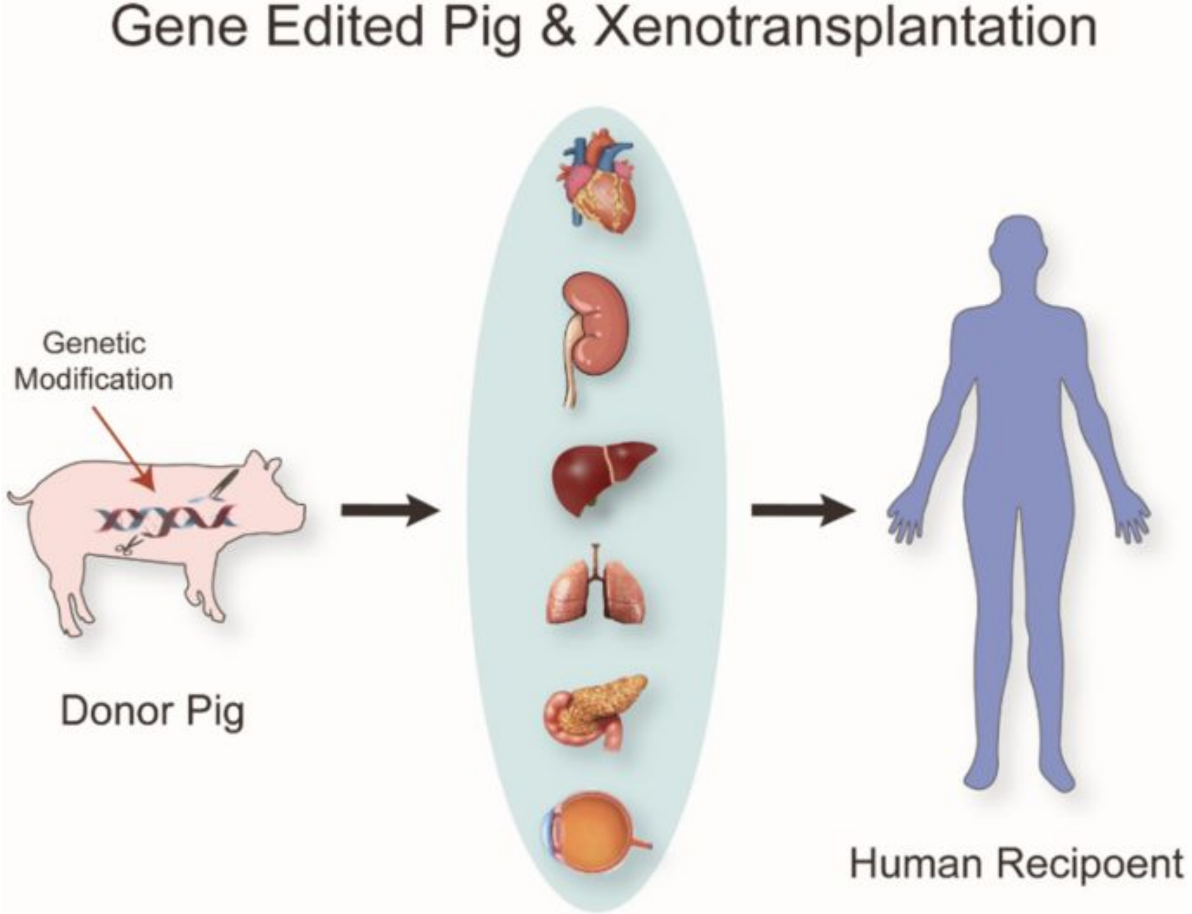
Motivation: Pigs as a valuable biomedical model



- Pigs (*Sus Scrofa*) used in variety of research areas including agricultural and biomedical studies
- Ideal animal model for human health and disease as genomically, physiologically and immunologically similar to humans

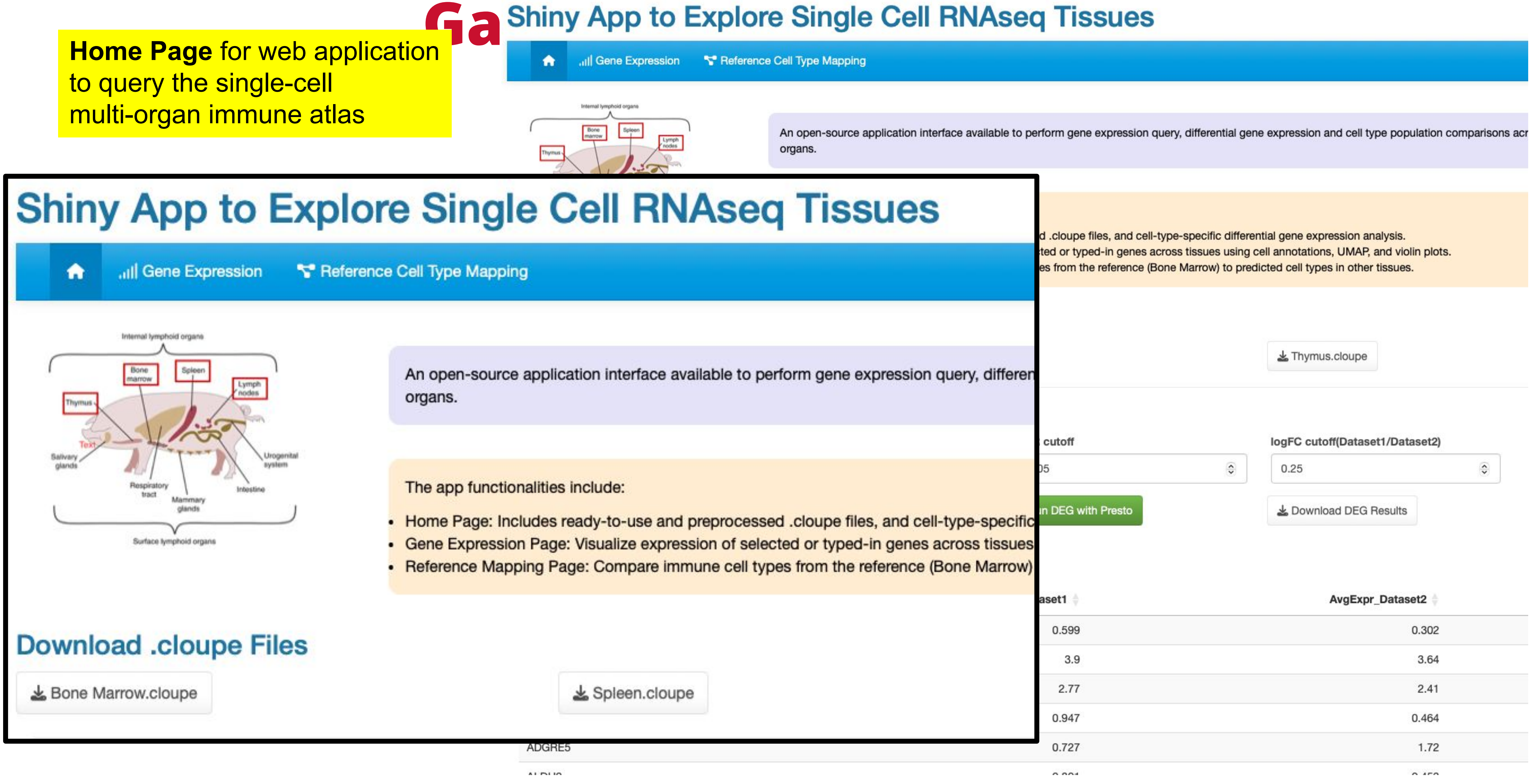
Overall goal if the study: Create a porcine single-cell immune cell transcriptional atlas





Shiny-PIGGI: Porcine Immune Genomics Ga

Home Page for web application to query the single-cell multi-organ immune atlas



Shiny App to Explore Single Cell RNAseq Tissues

An open-source application interface available to perform gene expression query, differential gene expression and cell type population comparisons acr
organs.

The app functionalities include:

- Home Page: Includes ready-to-use and preprocessed .cloupe files, and cell-type-specific
- Gene Expression Page: Visualize expression of selected or typed-in genes across tissues
- Reference Mapping Page: Compare immune cell types from the reference (Bone Marrow)

d .cloupe files, and cell-type-specific differential gene expression analysis.
ted or typed-in genes across tissues using cell annotations, UMAP, and violin plots.
es from the reference (Bone Marrow) to predicted cell types in other tissues.

Download .cloupe Files

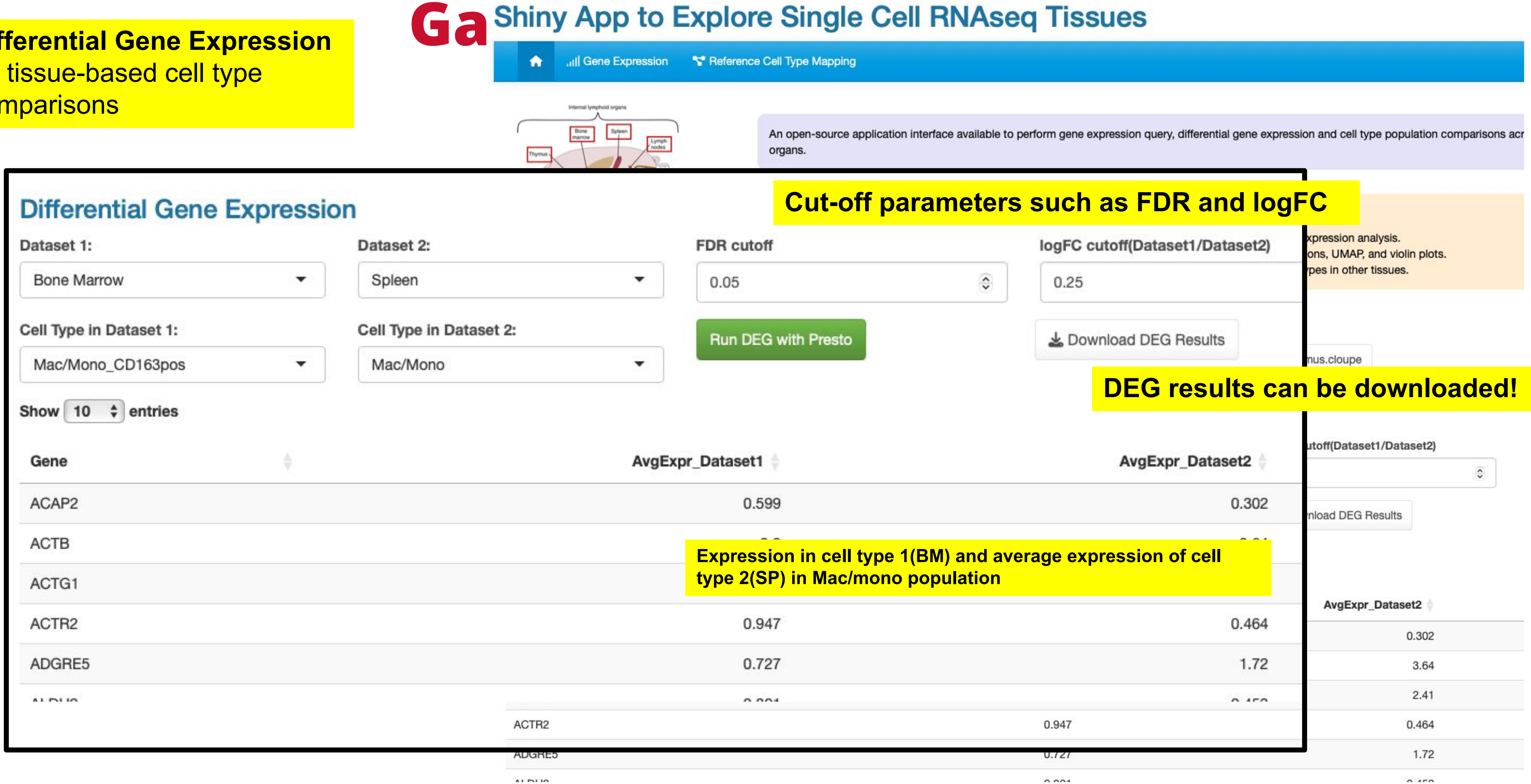
cutoff: 05

logFC cutoff(Dataset1/Dataset2): 0.25

	aset1	AvgExpr_Dataset2
	0.599	0.302
	3.9	3.64
	2.77	2.41
	0.947	0.464
ADGRE5	0.727	1.72

Shiny-PIGGI: Porcine Immune Genomics Ga

Differential Gene Expression for tissue-based cell type comparisons



Shiny App to Explore Single Cell RNAseq Tissues

An open-source application interface available to perform gene expression query, differential gene expression and cell type population comparisons acr organs.

xpression analysis.
ons, UMAP, and violin plots.
pes in other tissues.

Differential Gene Expression

Dataset 1: Bone Marrow

Dataset 2: Spleen

FDR cutoff: 0.05

logFC cutoff(Dataset1/Dataset2): 0.25

Cell Type in Dataset 1: Mac/Mono_CD163pos

Cell Type in Dataset 2: Mac/Mono

Gene	AvgExpr_Dataset1	AvgExpr_Dataset2
ACAP2	0.599	0.302
ACTB		
ACTG1		
ACTR2	0.947	0.464
ADGRE5	0.727	1.72

		AvgExpr_Dataset2
		0.302
		3.64
		2.41
ACTR2	0.947	0.464
ADGRE5	0.727	1.72

Cut-off parameters such as FDR and logFC

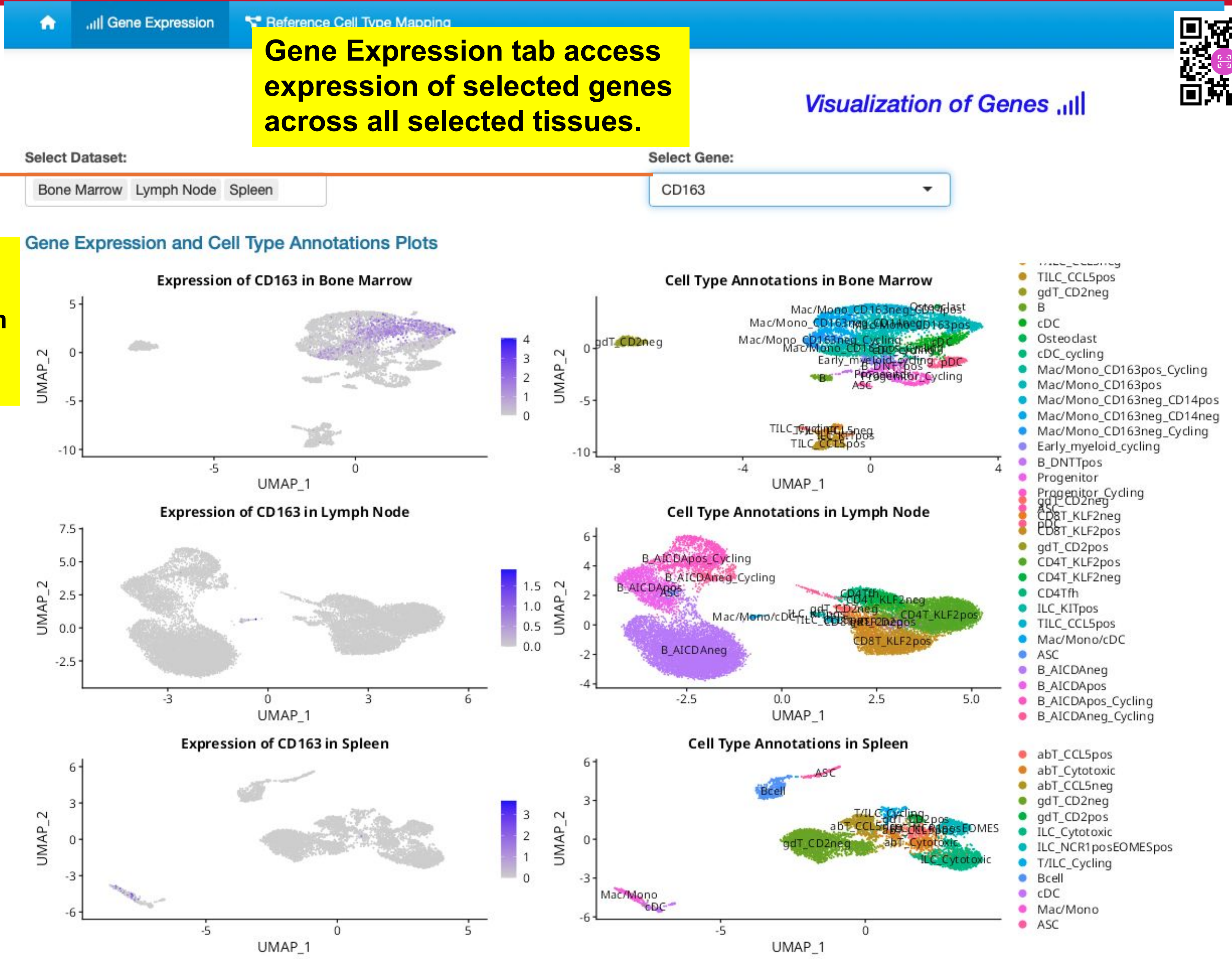
DEG results can be downloaded!

Expression in cell type 1(BM) and average expression of cell type 2(SP) in Mac/mono population

Shiny-PIGGI: Gene Expression Page

Gene Expression tab access expression of selected genes across all selected tissues.

Example shown, a key marker gene CD163 used to annotate unique monocyte/macrophage populations in bone marrow compared to other tissues!



Visualization of Genes

Select Dataset: Bone Marrow, Lymph Node, Spleen

Select Gene: CD163

Gene Expression and Cell Type Annotations Plots

Shiny-PIGGI: Gene

Gene Expression tab access expression of selected genes across all selected tissues.

Violin plots for CD163 indicate it was not highly expressed in lymph node. But splenic monocytes and other cycling monocytes in BM are also expressed

Select Dataset: Bone Marrow Lymph Node Spleen

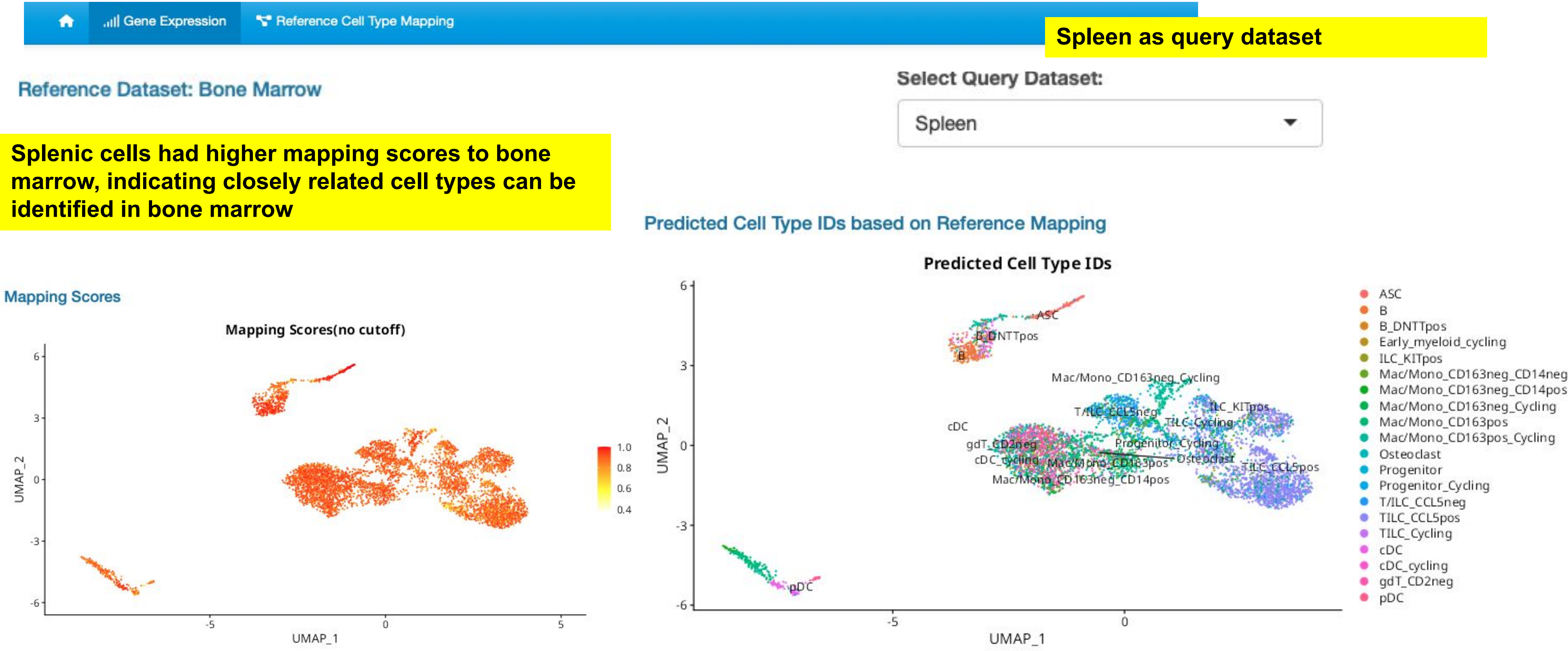
Select Gene: CD163

Violin Plot by Cell Type

Shiny-PIGGI: Reference Mapp

Bone marrow as a reference dataset

Comparisons of cell types recovered from bone marrow and spleen datasets performed using reference cell type mapping



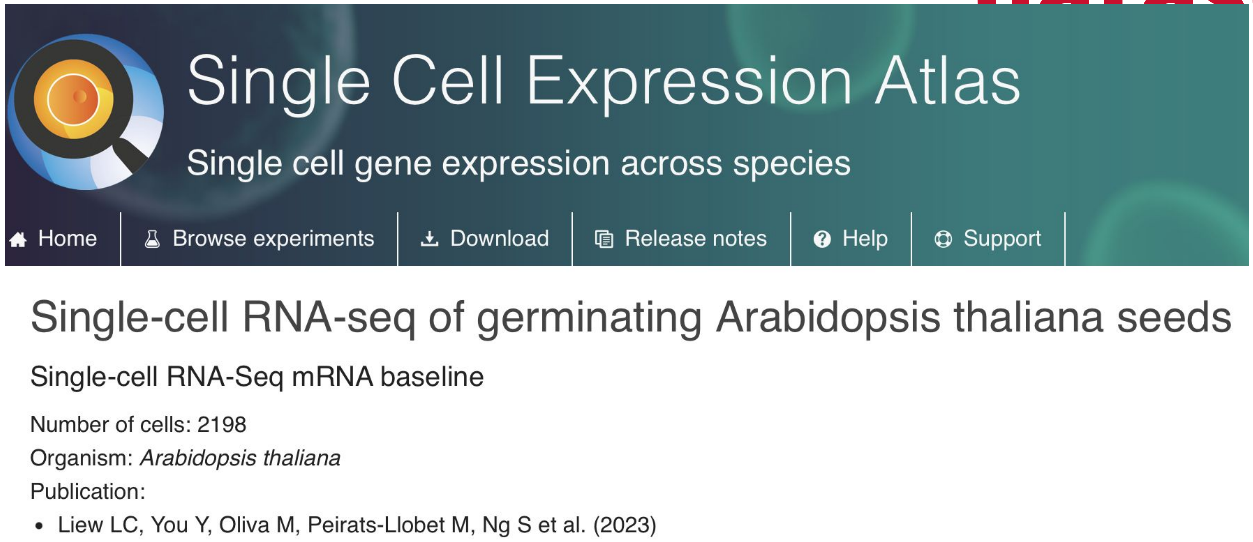
Reference Dataset: Bone Marrow

Select Query Dataset: Spleen

Spleen as query dataset

Splenic cells had higher mapping scores to bone marrow, indicating closely related cell types can be identified in bone marrow

FAIR Design to adapt to other agricultural SC datasets



Single-cell RNA-seq of germinating Arabidopsis thaliana seeds

Single-cell RNA-Seq mRNA baseline

Number of cells: 2198

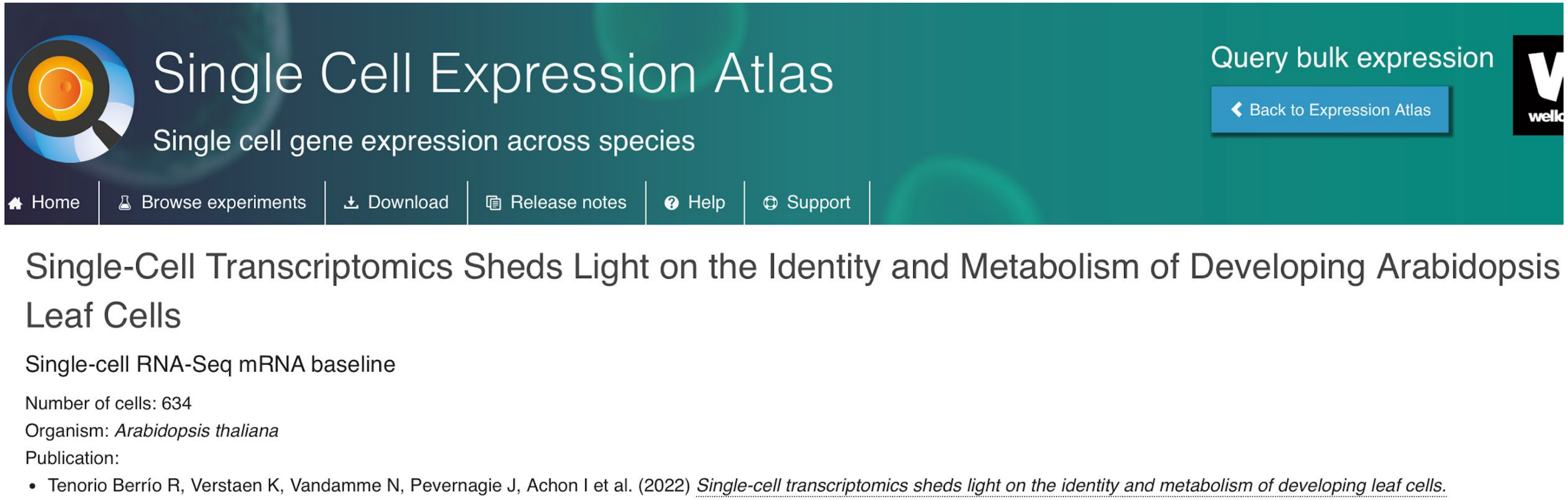
Organism: *Arabidopsis thaliana*

Publication:

- Liew LC, You Y, Oliva M, Peirats-Llobet M, Ng S et al. (2023)

ARABIDOPSIS - A Model Organism

Sunita Kumari, CSHL



Single-Cell Transcriptomics Sheds Light on the Identity and Metabolism of Developing Arabidopsis Leaf Cells

Single-cell RNA-Seq mRNA baseline

Number of cells: 634

Organism: *Arabidopsis thaliana*

Publication:

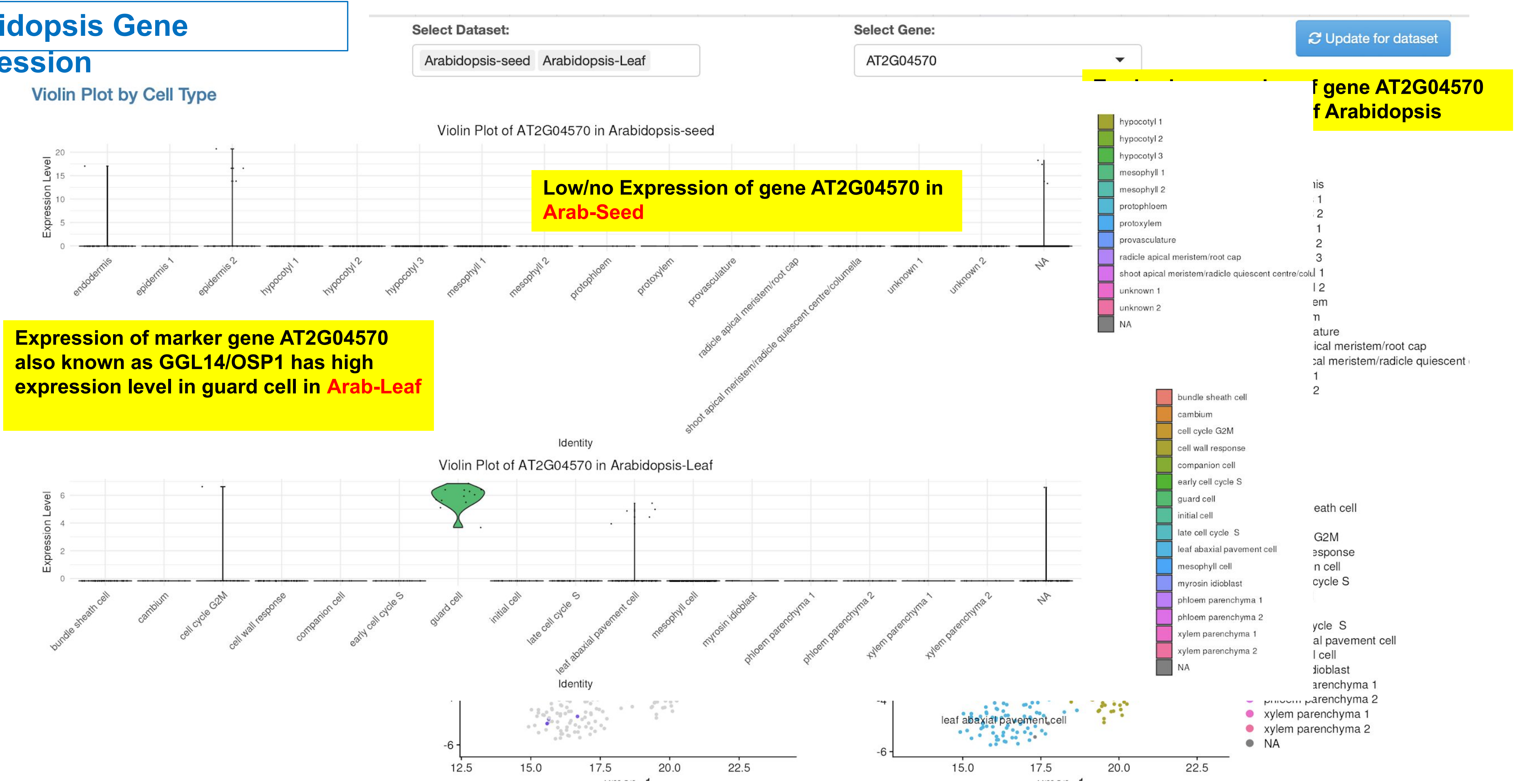
- Tenorio Berrío R, Verstaen K, Vandamme N, Pevernagie J, Achon I et al. (2022) *Single-cell transcriptomics sheds light on the identity and metabolism of developing leaf cells.*

Goal:

- Violin plot by cell type in Seeds and Leaf Cells.
- Differential expresion of genes in different cell types of leaf cells such as cell wall response and guard cell

Shiny-PIGGI adopted for agricultural SC datasets

Arabidopsis Gene Expression



Select Dataset: Arabidopsis-seed Arabidopsis-Leaf

Select Gene: AT2G04570

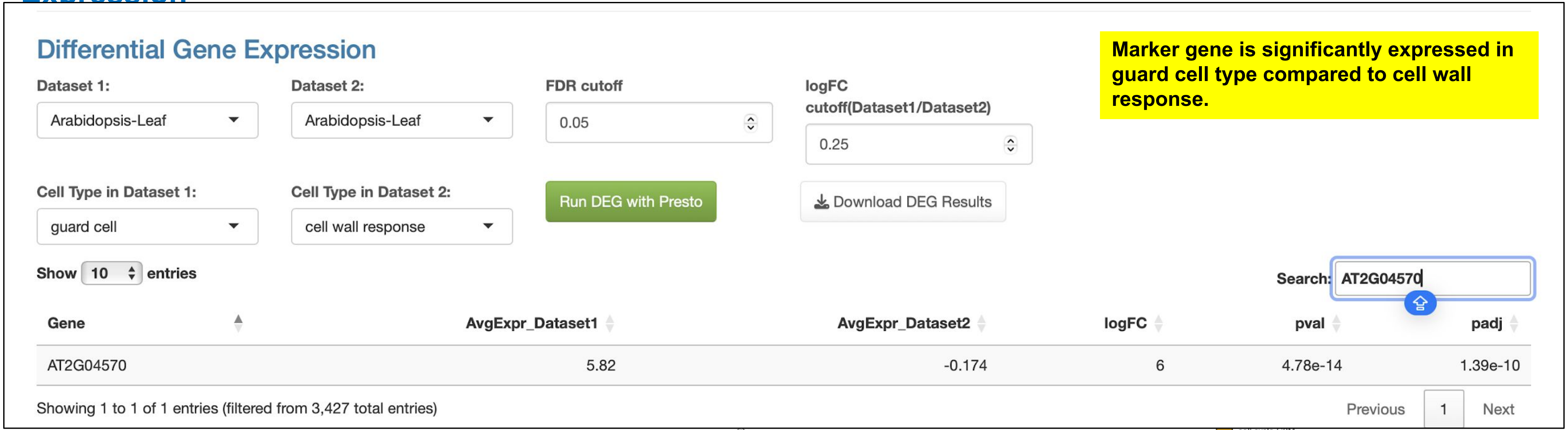
f gene AT2G04570
f Arabidopsis

Low/no Expression of gene AT2G04570 in Arab-Seed

Expression of marker gene AT2G04570 also known as GGL14/OSP1 has high expression level in guard cell in Arab-Leaf

Shiny-PIGGI adopted for agricultural SC datasets

Arabidopsis Differential Gene Expression



Differential Gene Expression

Dataset 1: Arabidopsis-Leaf

Dataset 2: Arabidopsis-Leaf

FDR cutoff: 0.05

logFC cutoff(Dataset1/Dataset2): 0.25

Cell Type in Dataset 1: guard cell

Cell Type in Dataset 2: cell wall response

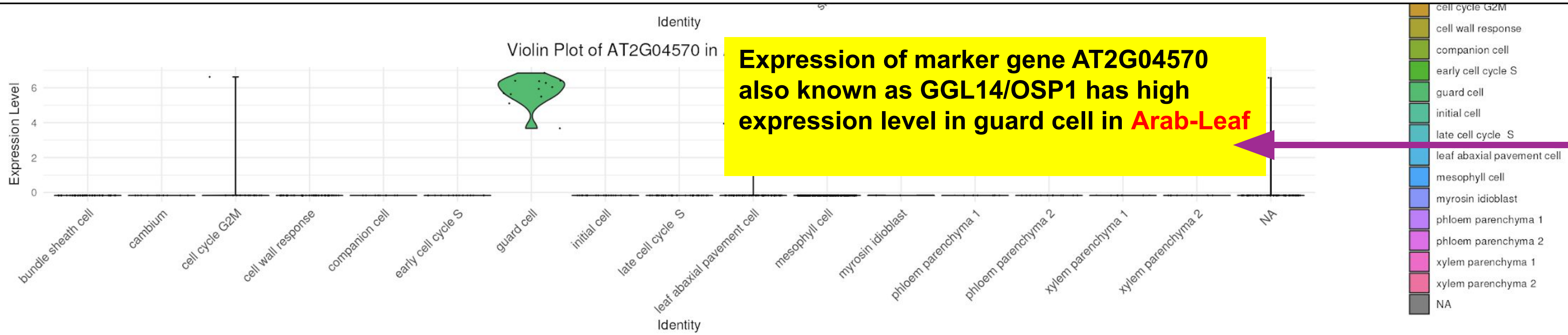
Show 10 entries

Search: AT2G04570

Gene	AvgExpr_Dataset1	AvgExpr_Dataset2	logFC	pval	padj
AT2G04570	5.82	-0.174	6	4.78e-14	1.39e-10

Showing 1 to 1 of 1 entries (filtered from 3,427 total entries)

Marker gene is significantly expressed in guard cell type compared to cell wall response.



Expression of marker gene AT2G04570 also known as GGL14/OSP1 has high expression level in guard cell in Arab-Leaf

Vision for this year

We have successfully created a cohesive community to work on standards, resources, tools and the benchmarked data set last year.

- We are planning to extend this work with the community on metadata standards, cell type annotations and benchmarking the data sets and tools especially on identifying datasets and pipelines for specific research goals.
- We are also planning to work on plant and animal cell-type ontologies and mappings this year, exploring AI more extensively for metadata curation and cell type annotation.

Thanks! Join our Working Group and Chat with us…!


Sunita Kumari


Chris Tuggle


Crystal Loving


Ben Cole


Theros Ng

Western University of Health Sciences

AgBioData Booth #223

AG2PI


Doreen Ware


Muskan Kapoor


Jayne Wiarda


Sharon Greenblum


Praveen Krishna

FAANG
Functional Annotation of Animal Genomes

CSH Cold Spring Harbor Laboratory

Iowa State University

USDA

JGI Joint Genome Institute

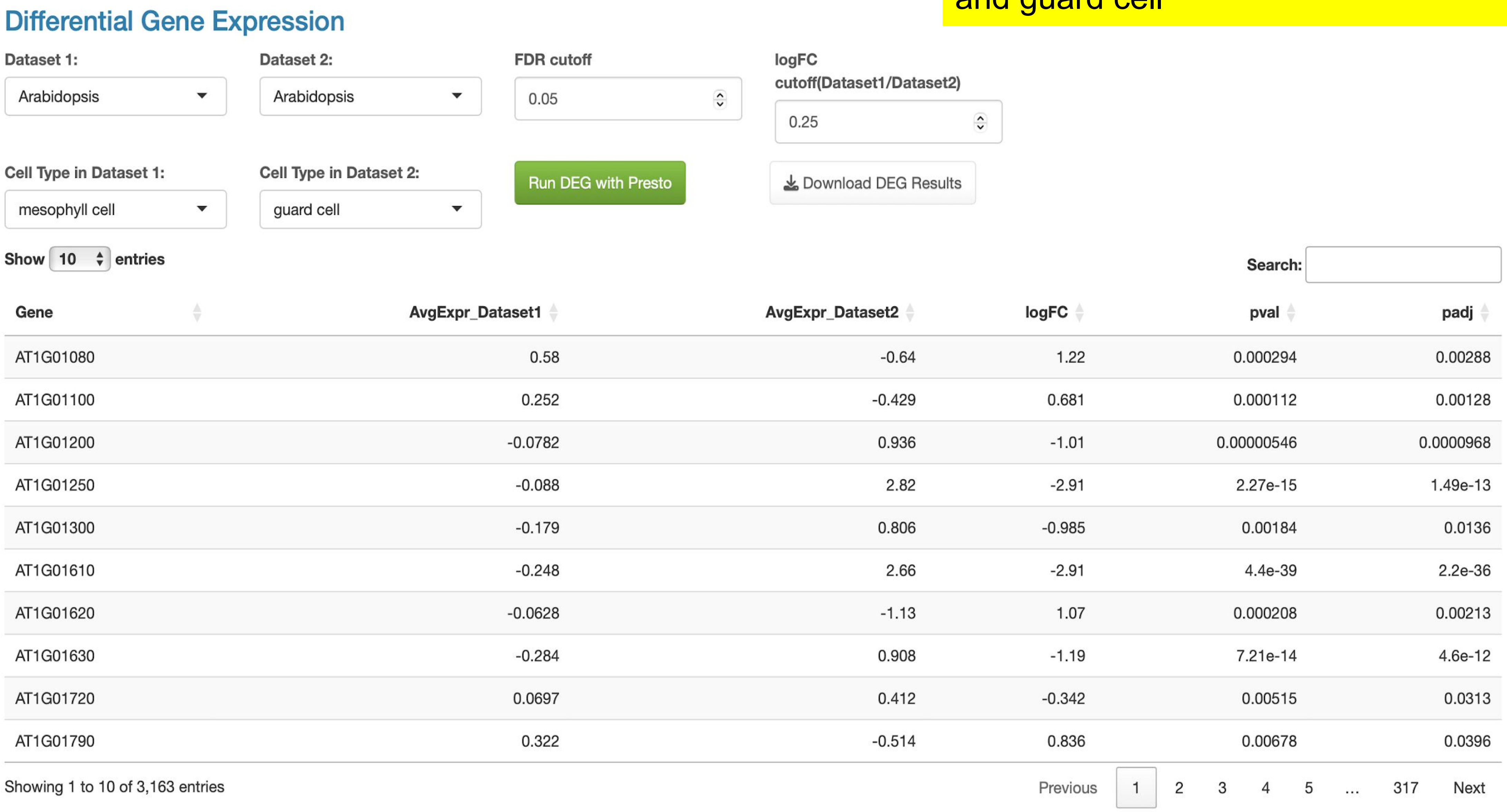
Research Institute for Farm Animal Biology

PLANT CELL ATLAS

Supported by a grant from USDA NIFA AG2PI # 2024-70412-43649 and AgBioData Consortium NSF-RCN Award # 2126334

Arabidopsis Cell-Type Specific Differential Gene Expression

Differential expresion of genes in different cell types of leaf cells such as mesophyll and guard cell



Differential Gene Expression

Dataset 1: Arabidopsis

Dataset 2: Arabidopsis

FDR cutoff: 0.05

logFC cutoff(Dataset1/Dataset2): 0.25

Cell Type in Dataset 1: mesophyll cell

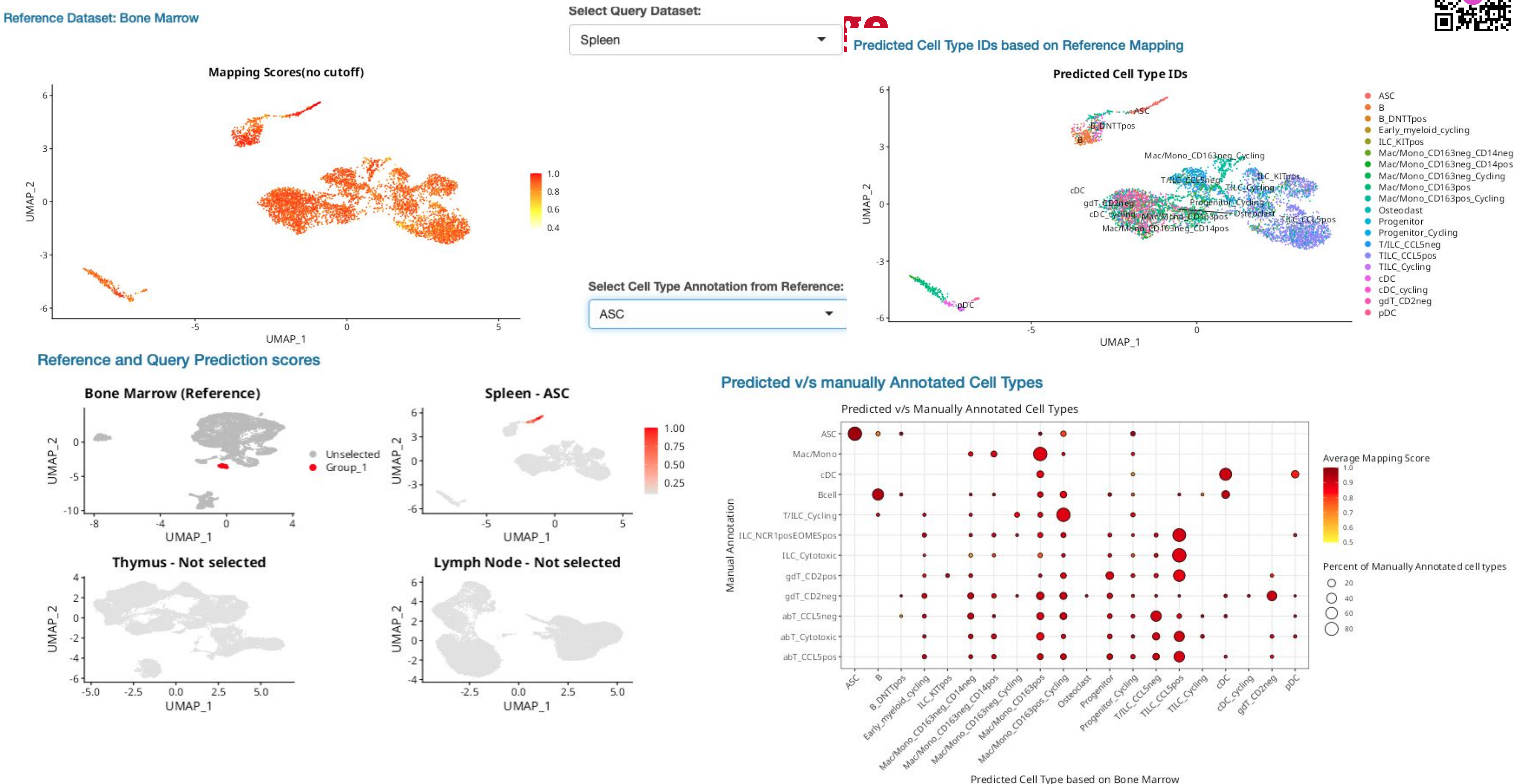
Cell Type in Dataset 2: guard cell

Show 10 entries

Gene	AvgExpr_Dataset1	AvgExpr_Dataset2	logFC	pval	padj
AT1G01080	0.58	-0.64	1.22	0.000294	0.00288
AT1G01100	0.252	-0.429	0.681	0.000112	0.00128
AT1G01200	-0.0782	0.936	-1.01	0.00000546	0.0000968
AT1G01250	-0.088	2.82	-2.91	2.27e-15	1.49e-13
AT1G01300	-0.179	0.806	-0.985	0.00184	0.0136
AT1G01610	-0.248	2.66	-2.91	4.4e-39	2.2e-36
AT1G01620	-0.0628	-1.13	1.07	0.000208	0.00213
AT1G01630	-0.284	0.908	-1.19	7.21e-14	4.6e-12
AT1G01720	0.0697	0.412	-0.342	0.00515	0.0313
AT1G01790	0.322	-0.514	0.836	0.00678	0.0396

Showing 1 to 10 of 3,163 entries

Shiny-PIGGI: Reference



Reference Dataset: Bone Marrow

Select Query Dataset: Spleen

Select Cell Type Annotation from Reference: ASC

Predicted Cell Type IDs based on Reference Mapping

Reference and Query Prediction scores

Predicted v/s manually Annotated Cell Types

Shiny-PIGGI: Reference Mapping Page

Bone marrow as a reference dataset

Reference Dataset: Bone Marrow

Spleen as query dataset

Select Query Dataset:

Spleen ▾

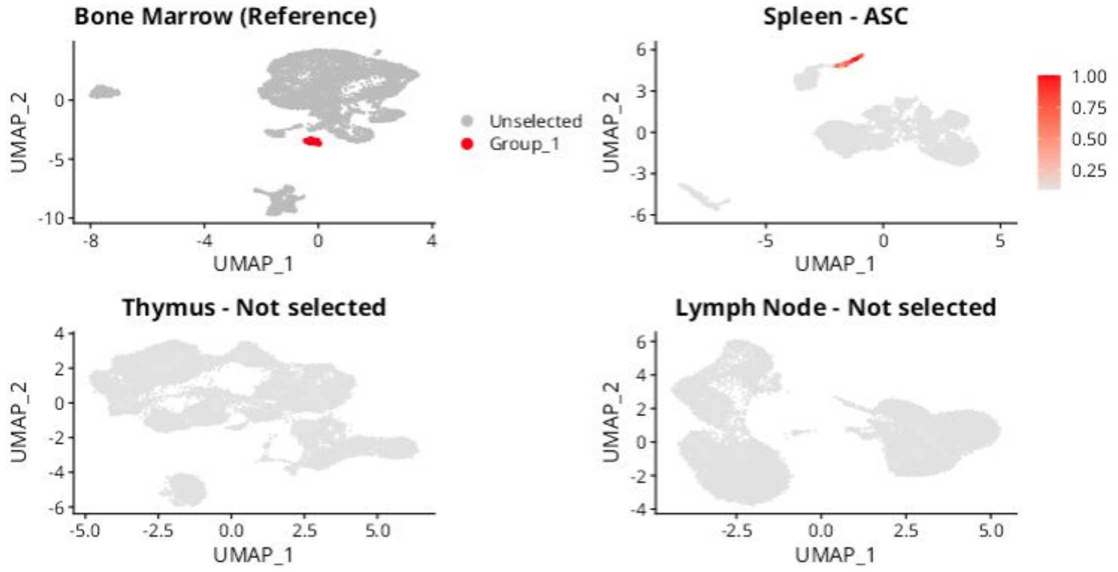
Select Cell Type Annotation from Reference:

ASC ▾

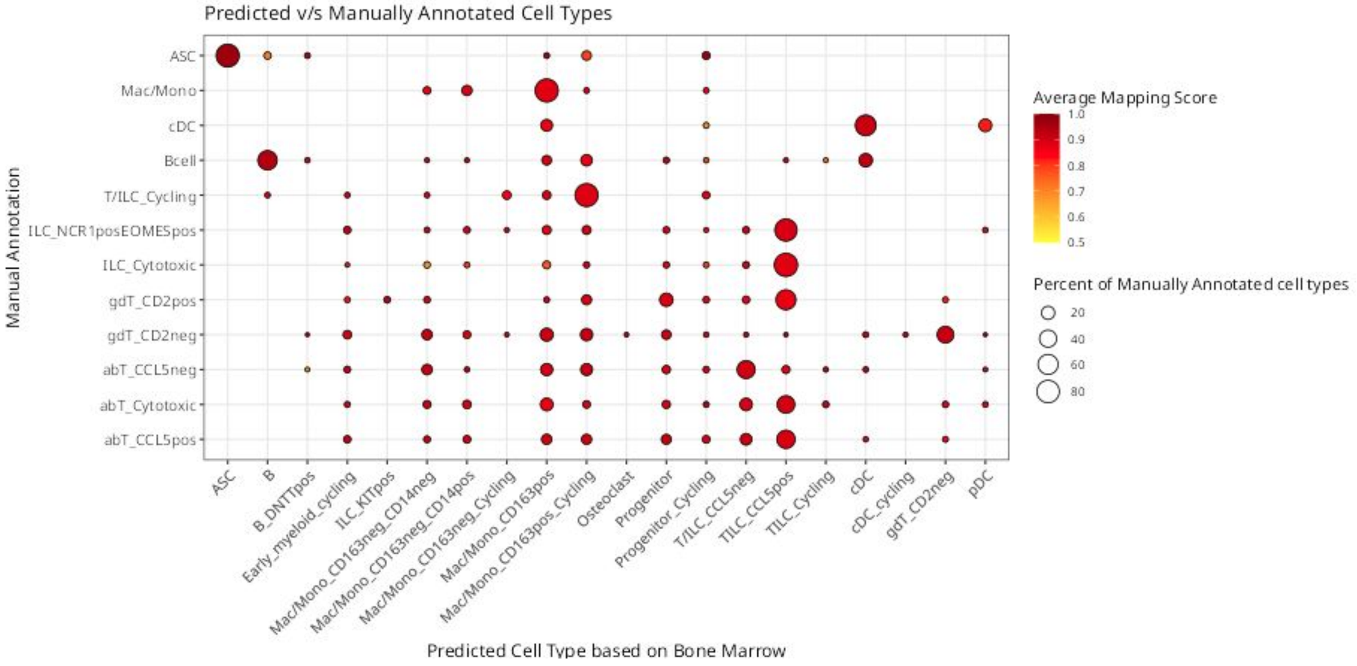
ASC shows high prediction score in spleen cells

Similarly annotated cell types were in consensus based on cell predictions including B, ASC, mac/mono and cDC

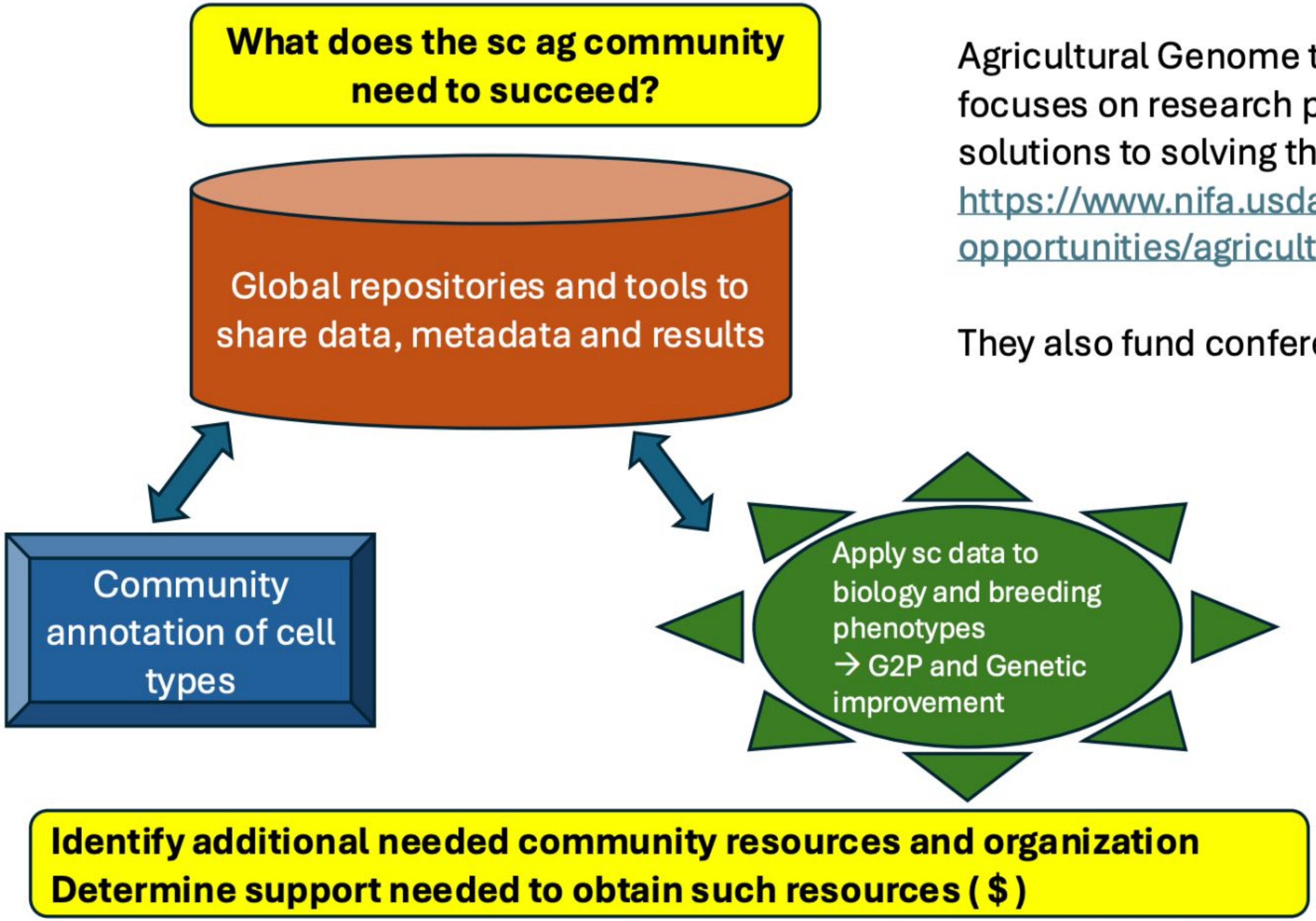
Reference and Query Prediction scores



Predicted v/s manually Annotated Cell Types



Plants and Animal single-cell researchers have been discussing means to develop such



What does the sc ag community need to succeed?

Global repositories and tools to share data, metadata and results

Community annotation of cell types

Apply sc data to biology and breeding phenotypes
→ G2P and Genetic improvement

Identify additional needed community resources and organization
Determine support needed to obtain such resources ($)

Agricultural Genome to Phenome Initiative (AG2PI), which focuses on research projects to develop community solutions to solving the Genome to Phenome problems. https://www.nifa.usda.gov/grants/funding-opportunities/agricultural-genome-phenome-initiative

They also fund conferences!

Supported by a grant from USDA NIFA AG2PI #2024-70412-43649