

Designing FAIR Metadata Workflows for Agricultural Single-Cell Transcriptomics

Sunita Kumari
Cold Spring Harbor Lab, NY

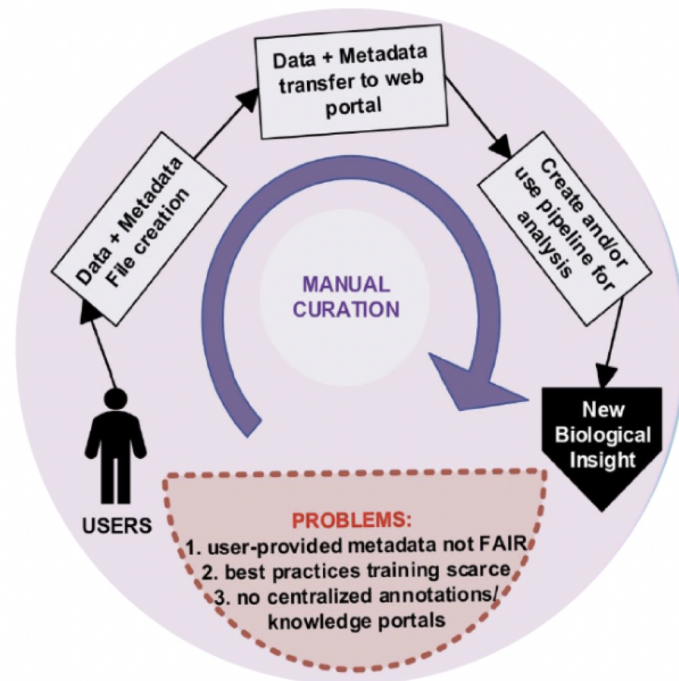
2025 AgBioData Community Workshop
May 14-15, 2025



Sample metadata FAIR Barriers



- Missing, incomplete, or inaccurate data
- Incomprehensible data
- Unverifiable results
- Incompatibility with FAIR
- Best practices training scarce
- No centralized annotations knowledge portals
- Requires manual curation on every step
- Challenge with Cell Type Annotations



Requirement of Single Cell Biocuration Working Group



- Address the annotation challenges of scRNAseq metadata in plants and animals, and establish a set of recommendations for member databases on management of scRNAseq datasets.
- Formulate and facilitate a cohesive community that will provide standards, resources, tools and standardized scRNAseq datasets for FAIR data meta-analysis, leading to initial cell type transcriptome descriptions.



AgBioData

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



PLANT CELL ATLAS

AG2P1

Agricultural Genome to Phenome Initiative



FAANG

Functional Annotation of Animal Genomes



Muskan Kapoor

frontiers | Frontiers in Genetics

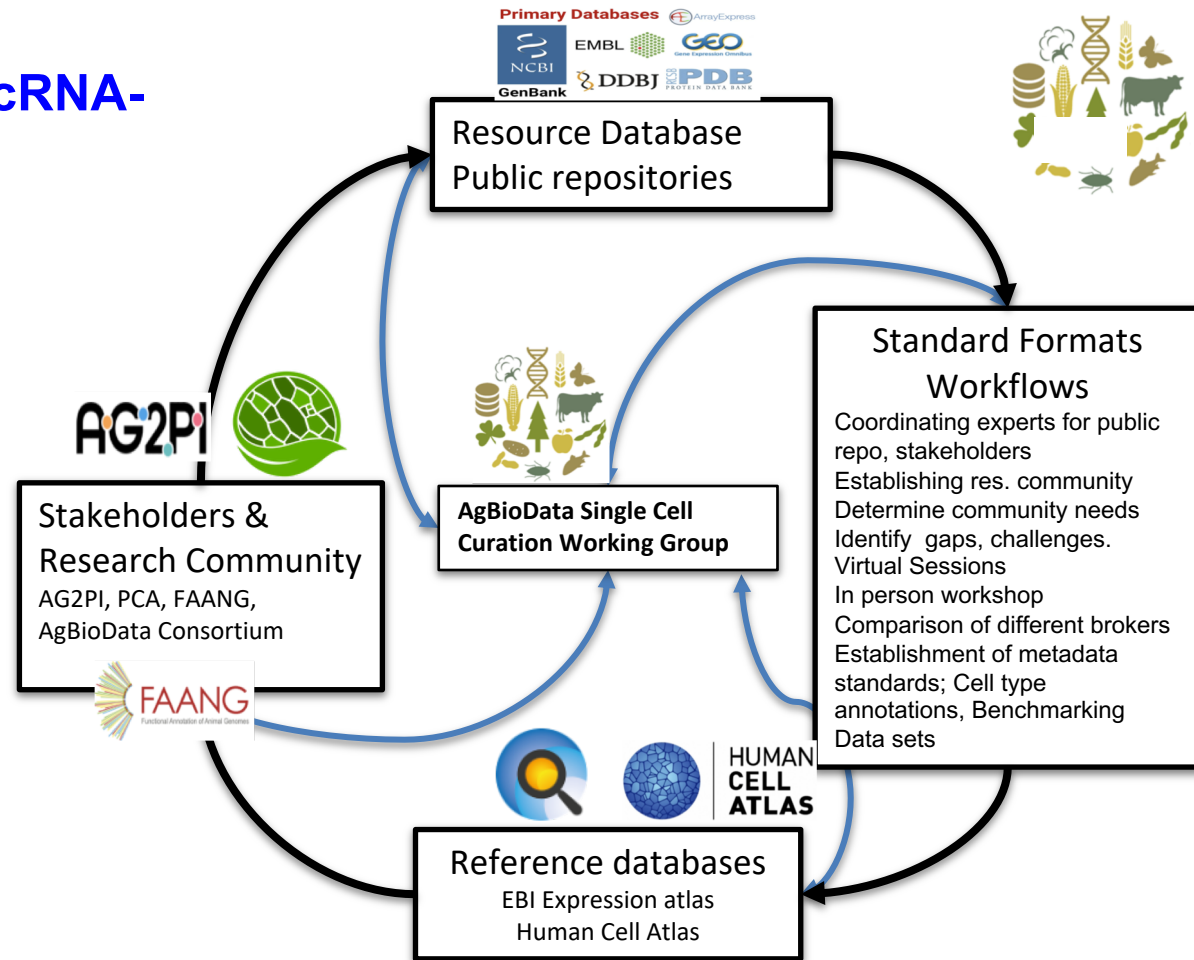
Building a FAIR data ecosystem for incorporating single-cell transcriptomics data into agricultural genome to phenome research

Muskan Kapoor^{1a}, Enrique Sapena Ventura², Amy Walsh³, Alexey Sokolov², Nancy George², Sunita Kumari⁴, Nicholas J. Provatt⁵, Benjamin Cole⁶, Marc Libaut⁷, Timothy Tickle¹, Wesley C. Warren⁸, James E. Klotz⁵, Irene Papatheodorou^{10,11}, Doreen Ware^{4,12}, Peter W. Harrison⁹, Christine Elisk¹, Galabina Yordanova², Tony Burdett¹ and Christopher K. Tuggle^{1a}

AgBioData community international groups on scRNA- seq metadata efforts

Follow FAIR principles

The ultimate goal is to cultivate a community of single cell researchers who are well-versed in single-cell methodologies and prepared to integrate these technologies into their research in a FAIR compliant manner.



Plant and animal single-cell researchers have been discussing means to develop such national/global resources

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>

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

AgBioData+AG2PI (crops and livestock)



AG2PI, AgBioData,
PCA, FAANG, EI
AnGenMap, Gramene

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.

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 (University of Iowa) (Chair)

Fiona McCarthy (University of Arizona)

Wes Warren (University of Missouri).

Sunita Kumari (CSHL)

Doreen Ware (USDA, CSHL)

Ben Cole (Lawrence Berkeley National Laboratory)

Marc Libault (University of Missouri)

Jennifer Clarke (University of Nebraska-Lincoln)

Irene Papatheodorou (Earlham Institute)

Project timeline

Project Start: January 1, 2025

January 1- March 30, 2025

Aim 1 and Aim 2 activities

March 29-30, 2025

Aim 3 conference activities

April - May 2025

Aim 4 White paper org, writing

May 2025- ongoing

Aim 4 Proposal org, writing, sub



Field Day 1 - February 5, 2025

Generating Single Cell Transcriptomic Data on Agricultural Model Systems

Organizers: Marc Libault, Herve Acloque

- Comparing single nucleus sequencing approaches to study maize leaf and floral development.

Samuel Lieboff: Oregon State University, Corvallis, USA

- Which single-cell methods for which biological questions: strengths and weaknesses.

Adrien Dufour: NeuroDiderot, INSERM, Universite Paris Cite, Paris, France

- Cell-type specific gene expression and function in Maize

Sunil Kenchanmane Raju: University of California, Riverside, USA



Field Day 2 - February 12, 2025

Submitting, sharing and accessing single cell data and metadata

Organizers: Sunita Kumari, Irene Papatheodorou, Doreen Ware

- Single-cell metadata challenges and AgBioData SC WG
Sunita Kumari: Cold Spring Harbor Lab, NY
- Single-cell data ingestion and metadata through Annotare
Silvie Fexova: European Bioinformatics Institute, Hinxton, UK
- COPO for submitting data and metadata to ENA
Felix Shaw: Earlham Institute, Norwich, UK



Field Day 3 - March 5, 2025

Strategies for assigning cell identity in single-cell datasets

Organizers: Fiona McCarthy, Ben Cole

- Orthologous marker groups reveal broad cell identity conservation across plant single-cell transcriptomes.
Song Li: Virginia Polytechnic Institute and State University, USA
- CellTypist and CellHint: towards automated annotation and integration of single-cell data.
Chaun Xu: Cambridge Stem Cell Institute, Cambridge, UK



Field Day 4 - March 12, 2025

The use of single cell data for advancing Genome 2 Phenome

Organizers: Chris Tuggle, Wes Warren

- The use of single cell data for advancing Genome2Phenome
Chris Tuggle: Iowa State University, Ames, USA
- The sc-eQTLGen consortium framework: a federated pipeline for genotype- and phenotype-based association analyses
Monique van der Wijst: Groningen University Medical Center, Groningen, The Netherlands

In-Person Single Cell Workshop (March 29-30), Orlando



Purpose: Develop working relationships, talks from experts working with medical and agricultural single cell data, metadata challenges, how to address the community's needs, breakout sessions and community discussions.

March 29, 2025: Session 1: Building a transdisciplinary community in agricultural SC genomics

- *Lessons learned from the interdisciplinary projects: core values, orientation and organization.* – Jennifer Clark, Univ of Nebraska
- *Community Survey Report* – Caleb Stull, Univ of Missouri

March 30, 2025: Session 2: Where we are in plant and animal SC genomics, as well as where we need to be?

- *Data Analysis and QC for Plant Single-nuclei Atlases*
- Sharon Greenblum, JGI
- *Data and meta-data management for re-using and integrating cell atlases*
- Irene Papatheodorou, Earlham Institute.
- *A Single nucleus and cell perspective on the chicken immune system*
- Wes Warren, Univ of Missouri
- *A student-driven approach for developing community workflows*
- Fiona McCarthy, Univ of Arizona



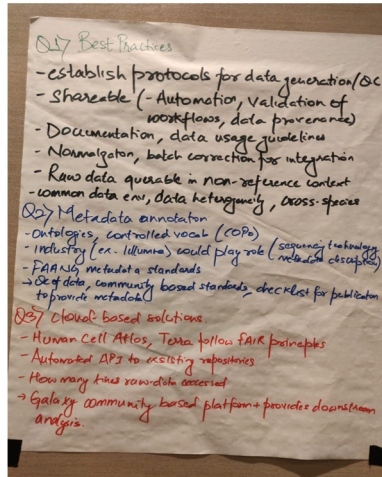
Small Group Breakout Discussions

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 5: Identifying opportunities and mechanisms to generate resources for community needs

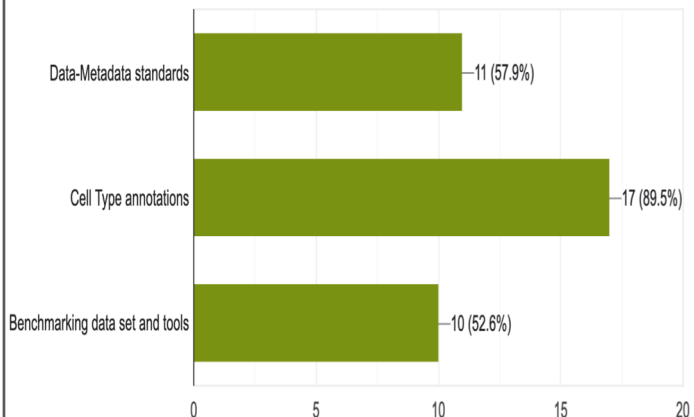
- What are additional needs for creating G2P solutions at the cell type level?
- Offer opportunity to govt funding agencies to indicate where these ideas fit within their portfolios.
- Identify writers of white paper and topics to USDA and other agencies
- Identify proposal organizers/leaders for applying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

Reorganization of the Working Group - 19 members



Please select the AgBioData Single Cell Sub-Groups of your interest

19 responses



Sunita Kumari	Cold Spring Harbor Lab
Benjamin Cole	Joint Genome Institute
Doreen Ware	USDA,Cold Spring Harbor Lab
Rex Nelson	USDA
Sharon Greenblum	Joint Genome Institute
Sushma Naithani	Oregon State University
Fabian Andres Reyes Salinas	University of Florida
Teng-Jui (Owen) Lin	UC Berkeley
Jennifer Clarke	University of Nebraska-Lincoln
Chris Tuggle	Iowa State University
Muskan Kapoor	Iowa State University
Mehak Kapoor	Iowa State University
Kimberly Davenport	Washington State University
Elisabetta Giuffra	INRAE
Theros Ng	Western University of Health Sciences
Doreen Becker	Farm Animal Biology (FBN)
Björn Corleis	Friedrich-Loeffler Institut
Felix Shaw	Earlham Institute, UK
Irene Papatheodorou	Earlham Institute, UK

Future Directions: Three Sub-Working groups



1. Data/Metadata standards

- Define and advertise minimal metadata standards
- Literature curation and meetings with repositories to define metadata standards for all Ag species.
- Survey agencies/other gatekeepers to advertise/incentivize standards.

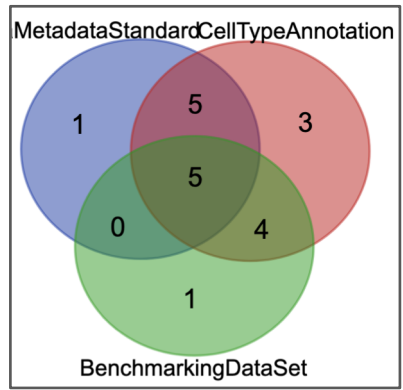
2. Cell type Annotation

- Identify and list of published/high quality SC datasets.
- Document Cell Atlases that can be used for future cell type annotation.
- Coordination with PCA members.

3. Benchmarking

- Identify datasets and pipelines for specific goals.
- Identify metrics for evaluating those pipelines.

Product: Conferences; White Paper etc.



Data Netadata Standards	Cell Type Annotations	Benchmrking Data Sets & Tools
Sunita Kumari	Sunita Kumari	Sunita Kumari
Doreen Ware	Doreen Ware	Doreen Ware
Chris Tuggle	Chris Tuggle	Elisabetta Giuffra
Benjamin Cole	Benjamin Cole	Muskan Kapoor
Elisabetta Giuffra	Muskan Kapoor	Fabian Andres
Rex Nelson	Fabian Andres	Mehak Kapoor
Kimberly Davenport	Mehak Kapoor	Theros Ng
Teng-Jui	Theros Ng	Jennifer Clarke
Sushma Naithani	Rex Nelson	Felix Shaw
Felix Shaw	Kimberly Davenport	Irene Papatheodorou
Irene Papatheodorou	Teng-Jui	
	Doreen Becker	
	Björn Corleis	
	Sharon Greenblum	
	Felix Shaw	
	Irene Papatheodorou	

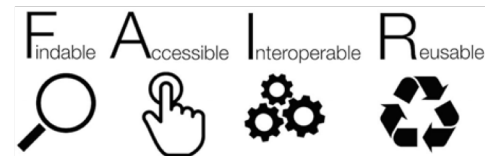


Aim 1:

Recommendations,
standards, and
implementation plans
for FAIR data.

Aim 3:

Educational and training
materials for
researchers.



Aim 2:

Expand the network
to include key
stakeholders.

Aim 4:

Roadmap for a sustainable
GGB data/database
ecosystem.



Special Thanks To Key Leaders and Experts

EBI Atlas - Christine Ernst, Silvie Fexova
Earlham Institute on COPO - Felix Shaw, Irene Papatheodorou
NCBI GEO - Emily Clough
AG2PI - Chris Tuggle, Wes Warren, Marc Libault, Jennifer Clarke
FAANG - Fiona McCarthy
PCA - Ben Cole, Sue Rhee
Gramene/SorghumBase - Doreen Ware

Upcoming Presentations

- ASPB - July 2025
- PCA - GRC Maine, August 2025
- PAG - Jan, 2026





Break Out Questions

Q1: How can we lower the barriers for users to share their data in FAIR compliant manner to repositories?

Q2: How can we motivate the researchers to share their data in FAIR compliant manner to repositories?

- Short term benefit
- Long term benefit

Q3: *How can we establish FAIR compliant data repositories for single cell data for agricultural species.*