

Empowering G2P in Agriculture: Building a Collaborative Environment for Single-Cell Applications

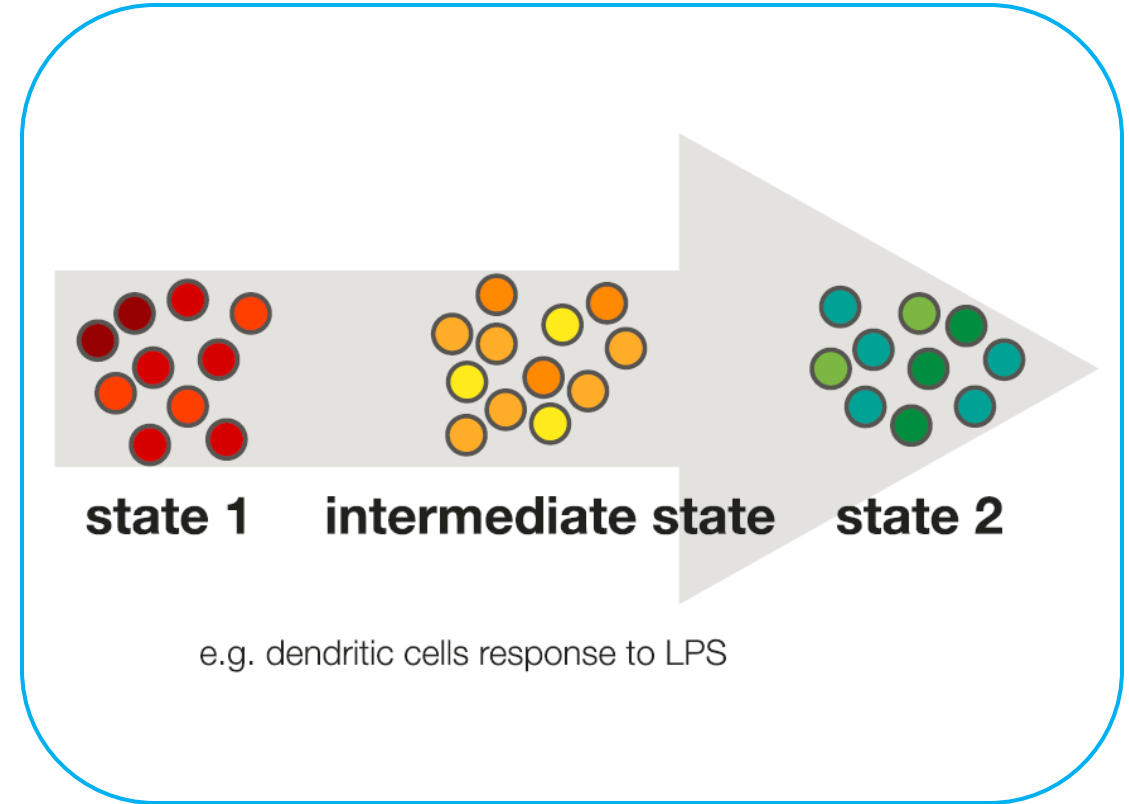
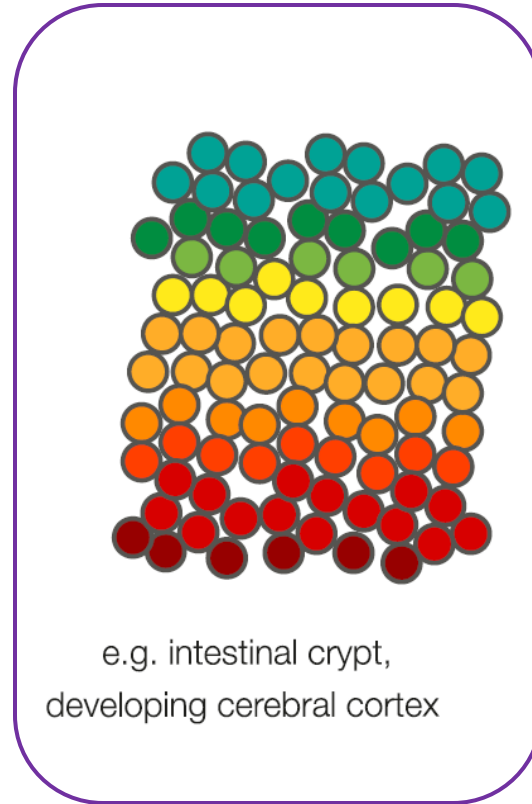
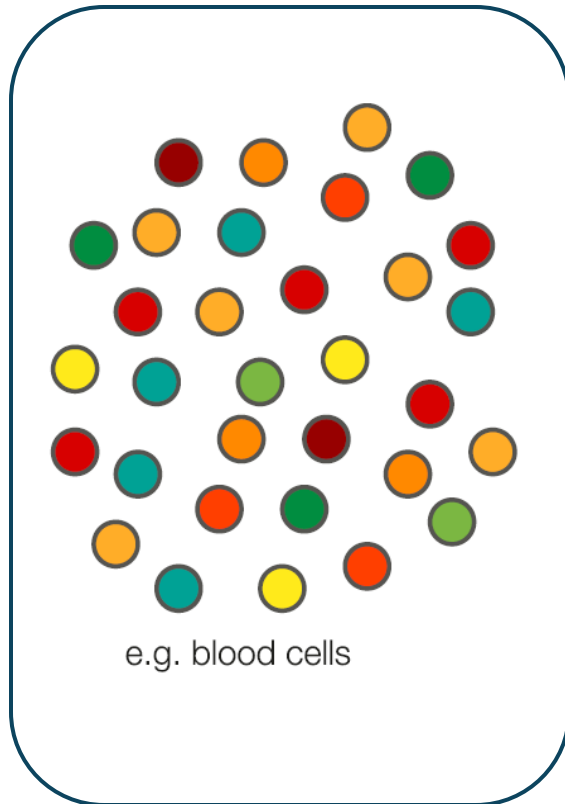
Christopher K Tuggle
Department of Animal Science
Iowa State University

AG2PI SC conference organizing Committee:

Chris Tuggle (Chair)	Iowa State University
Jennifer Clarke	University of Nebraska
Ben Cole	Lawrence Berkeley Lab
Marc Libault	University of Missouri
Fiona McCarthy	University of Arizona
Irene Papatheodorou	Earlham Institute
Sunita Kumari	USDA-ARS/Cold Spring Harbor Lab
Doreen Ware	USDA-ARS/Cold Spring Harbor Lab
Wes Warren	University of Missouri

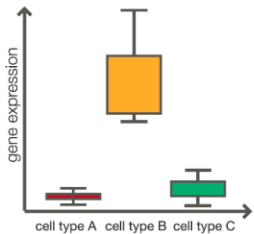
Why single cell RNA-seq?

Heterogeneous cell populations in tissues- analyses of “bulk” tissue misses a lot....

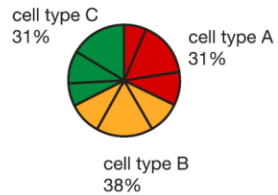


Value of single cell genomics for biology and breeding

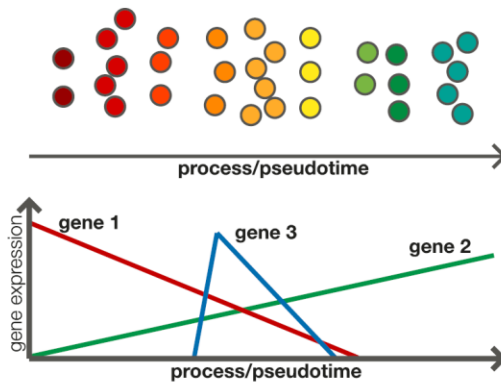
Finding markers of cell type



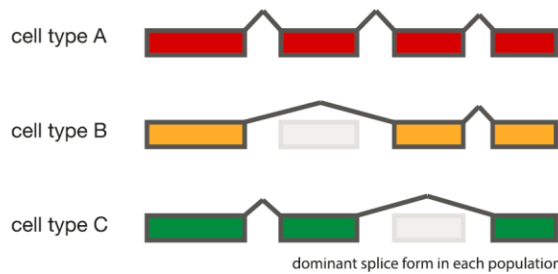
Frequency of cell type in the population



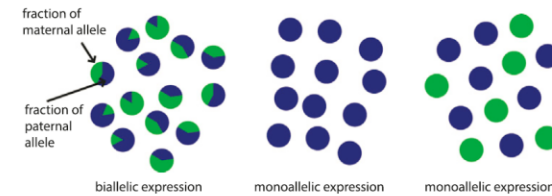
Identification of genes that drive a process



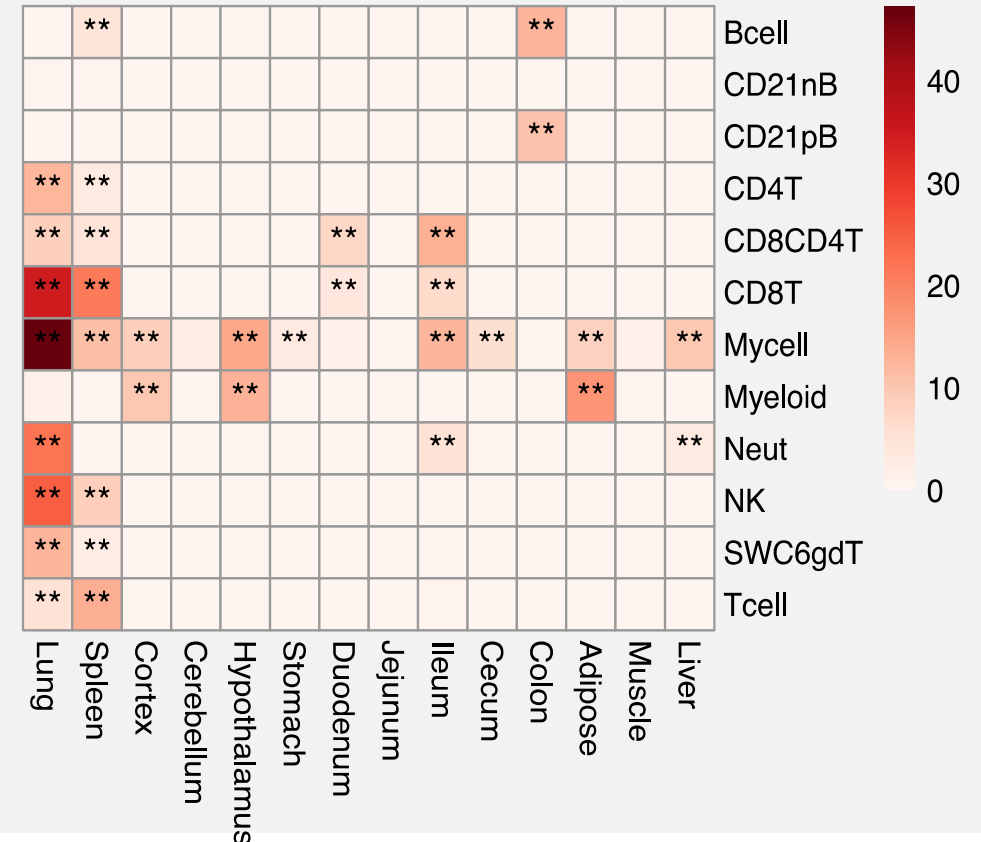
Differential splicing between populations



Allelic expression patterns



Enrichment of pig tissue enhancers (from Pan et al.) among immune-cell specific enhancers (Corbett et al.)



Linking to Breeding Goals: Finding cell-type specific eQTL in tissues/cell mixes

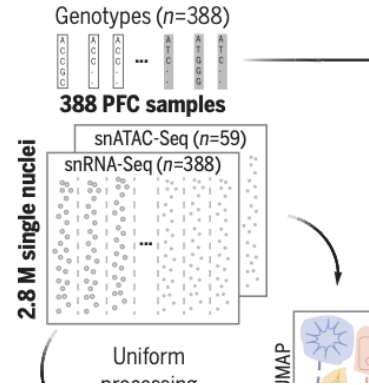
RESEARCH ARTICLE SUMMARY

PSYCHENCODE2

Single-cell genomics and regulatory networks for 388 human brains

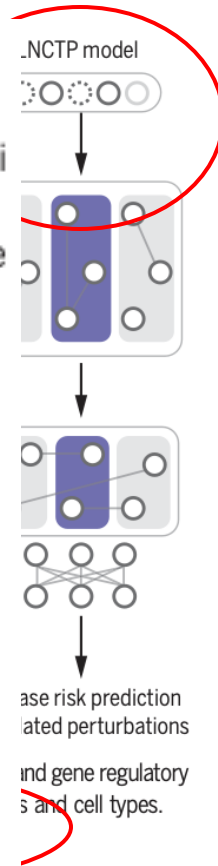
Emani *et al.*, *Science* **384**, 862 (2024) 24 May 2024

Prashant S. Emani^{1,2†}, Jason J. Liu^{1,2†}, Declan Clarke^{1,2†}, Matthew Jensen^{1,2†}, Jonathan Warrell^{1,2†}, Chirag Gupta^{3,4†}, Ran Meng^{1,2†}, Che Yu Lee^{5†}, Siwei Xu^{5†}, Cagatay Dursun^{1,2†}, Shaoke Lou^{1,2†}, Yuhang Chen^{1,2}, Zhiyuan Chu¹, Timur Galeev^{1,2}, Ahyeon Hwang^{5,6}, Yunyang Li^{2,7}, Pengyu Ni^{1,2}, Xiao Zhou^{1,2}, PsychENCODE Consortium†, Trygve E. Bakken⁸, Jaroslav Bendl^{9,10,11,12}, Lucy Bicks¹³, Tanima Chatterjee^{1,2}, Lijun Cheng¹⁴, Yuyan Cheng^{13,15}, Yi Dai⁵, Ziheng Duan⁵, Mary Flaherty¹⁴, John F. Fullard^{9,10,11,12}, Michael Gancz^{1,2}, Diego Garrido-Martín¹⁶, Sophia Gaynor-Gillett^{14,17}, Jennifer Grundman¹³, Natalie Hawken¹³, Ella Henry^{1,2}, Gabriel E. Hoffman^{9,10,11,12,18,19}, Ao Huang¹, Yunzhe Jiang^{1,2}, Ting Jin^{3,4}, Nikolas L. Jorstad⁸, Riki Kawaguchi^{13,20}, Saniya Khullar^{3,4}, Jianyin Liu¹³, Junhao Liu⁵, Shuang Liu⁴, Shaojie Ma^{21,22}, Michael Margolis¹³, Samantha Mazariegos¹³, Jill Moore²³, Jennifer R. Moran¹⁴, Eric Nguyen^{1,2}, Nishigandha Phalke²³, Milos Pjanic^{9,10,11,12}, Henry Pratt²³, Diana Quintero¹³, Ananya S. Rajagopalan⁷, Tiernon R. Riesenmy²⁴, Nicole Shedd²³, Manman Shi¹⁴, Megan Spector¹⁴, Rosemarie Terwilliger²⁵, Kyle J. Travaglini⁸, Brie Wamsley¹³, Gaoyuan Wang^{1,2}, Yan Xia^{1,2}, Shaohua Xiao¹³, Andrew C. Yang^{1,2}, Suchen Zheng^{1,2}, Michael J. Gandol^{26,27,28,29,30}, Donghoon Lee^{9,10,11,12}, Ed S. Lein^{8,31,32}, Panos Roussos^{9,10,11,12,18,19}, Nenad Sestan²¹, Zhiping Weng²³, Kevin P. White³³, Hyejung Won³⁴, Matthew J. Girgenti^{25,35,36*}, Jing Zhang^{5*}, Daifeng Wang^{3,4,37*}, Daniel Geschwind^{13,20,27,28,38*}, Mark Gerstein^{1,2,7,24,39*}



ACKNOWLEDGMENTS

The authors thank the founder of the Allen Institute, P. G. Allen, for his vision, encouragement, and support. R.T. and M. J. Girgenti thank Keck Microarray Shared Resource (KMSR) and Yale Center for Genome Analysis (YCGA) at Yale University for their assistance with 10x Genomics single-cell RNA-seq services. **Funding:** Data were generated as part of the PsychENCODE Consortium, supported by U01DA048279, U01MH103339, U01MH103340, U01MH103346, U01MH103365, U01MH103392, U01MH116438, U01MH116441, U01MH116442, U01MH116488, U01MH116489, U01MH116492, U01MH122590, U01MH122591, U01MH122592, U01MH122849, U01MH122678, U01MH122681, U01MH116487, U01MH122509, R01MH094714, R01MH105472, R01MH105898, R01MH109677, R01MH109715, R01MH110905, R01MH110920, R01MH110921, R01MH110926, R01MH110927, R01MH110928, R01MH111721, R01MH117291, R01MH117292, R01MH117293, R21MH102791, R21MH103877, R21MH105853, R21MH105881, R21MH109956, R56MH114899, R56MH114901, R56MH114911, R01MH125516, R01MH126459, R01MH129301, R01MH126393, R01MH121521, R01MH116529, R01MH129817, R01MH117406, and P50MH106934 awarded to authors and collaborators A. Abyzov,



use risk prediction
lated perturbations
and gene regulatory
s and cell types.

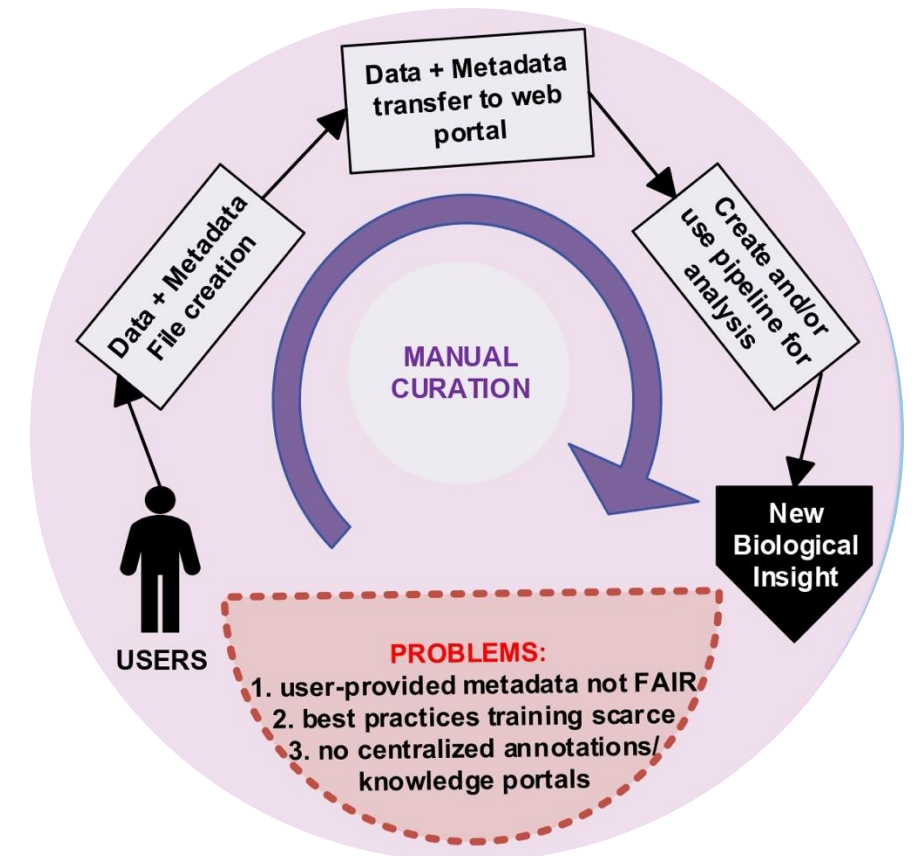
Agricultural SC research resource needs

- Excellent research can be done at a single lab level, with significant effort, but democratization would benefit agriculture.

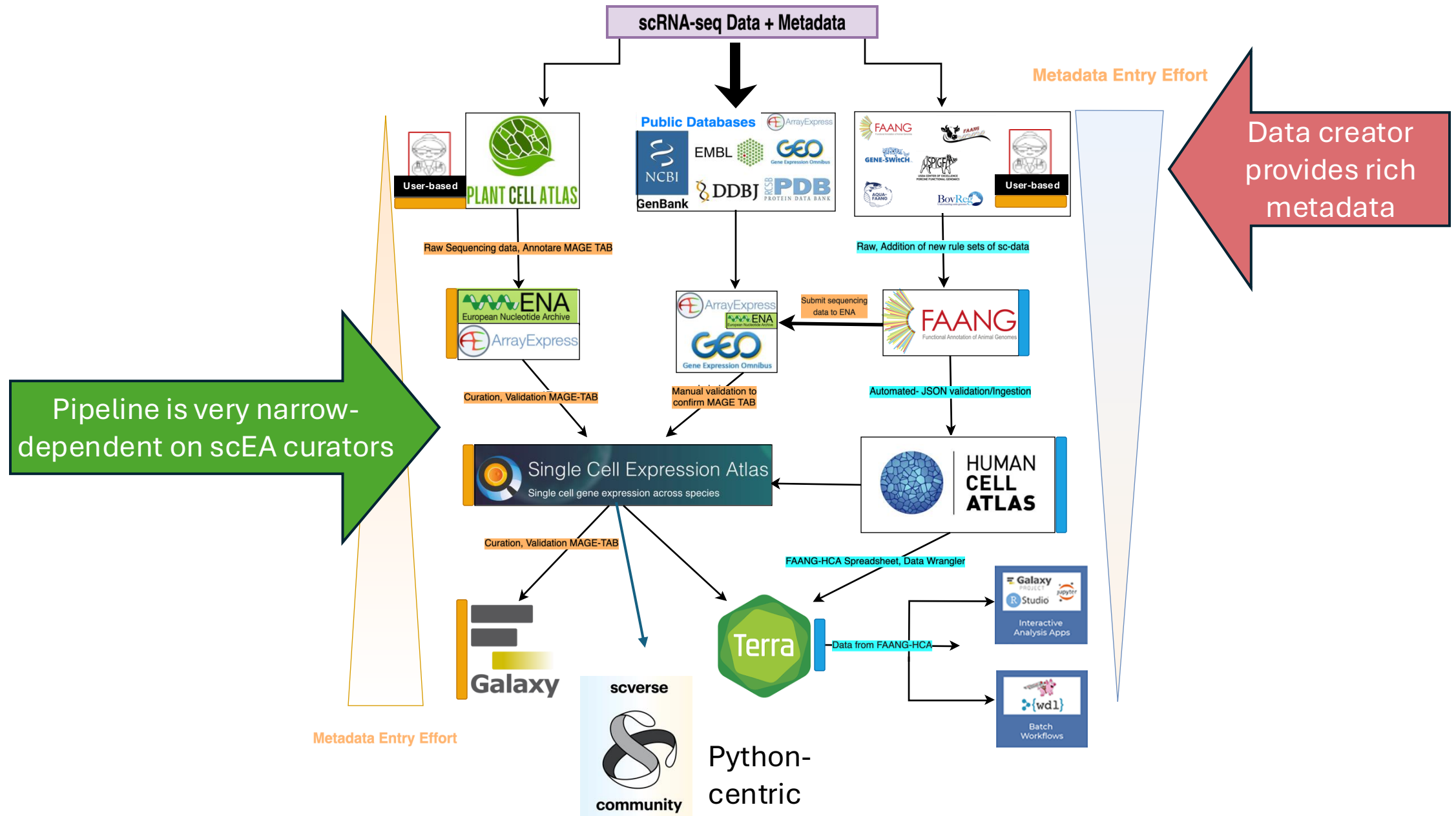
Difficult to impossible to do within a single lab:

- SC data and meta-data storage and sharing → efficacy and accuracy
- Utilization of organism-wide data for optimal cell annotation and gene regulatory networks
- Improved annotation of reference assemblies (Pan-genome) with SC data
- Analytical data pipelines – development and validation
- Comparative analysis of SC data across species

Agricultural research output –quantity, quality AND reproducibility could be significantly improved if such community-wide resources were developed!



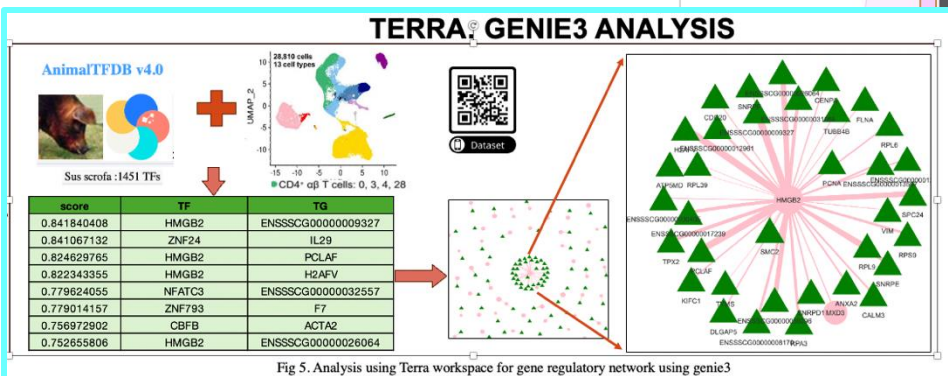
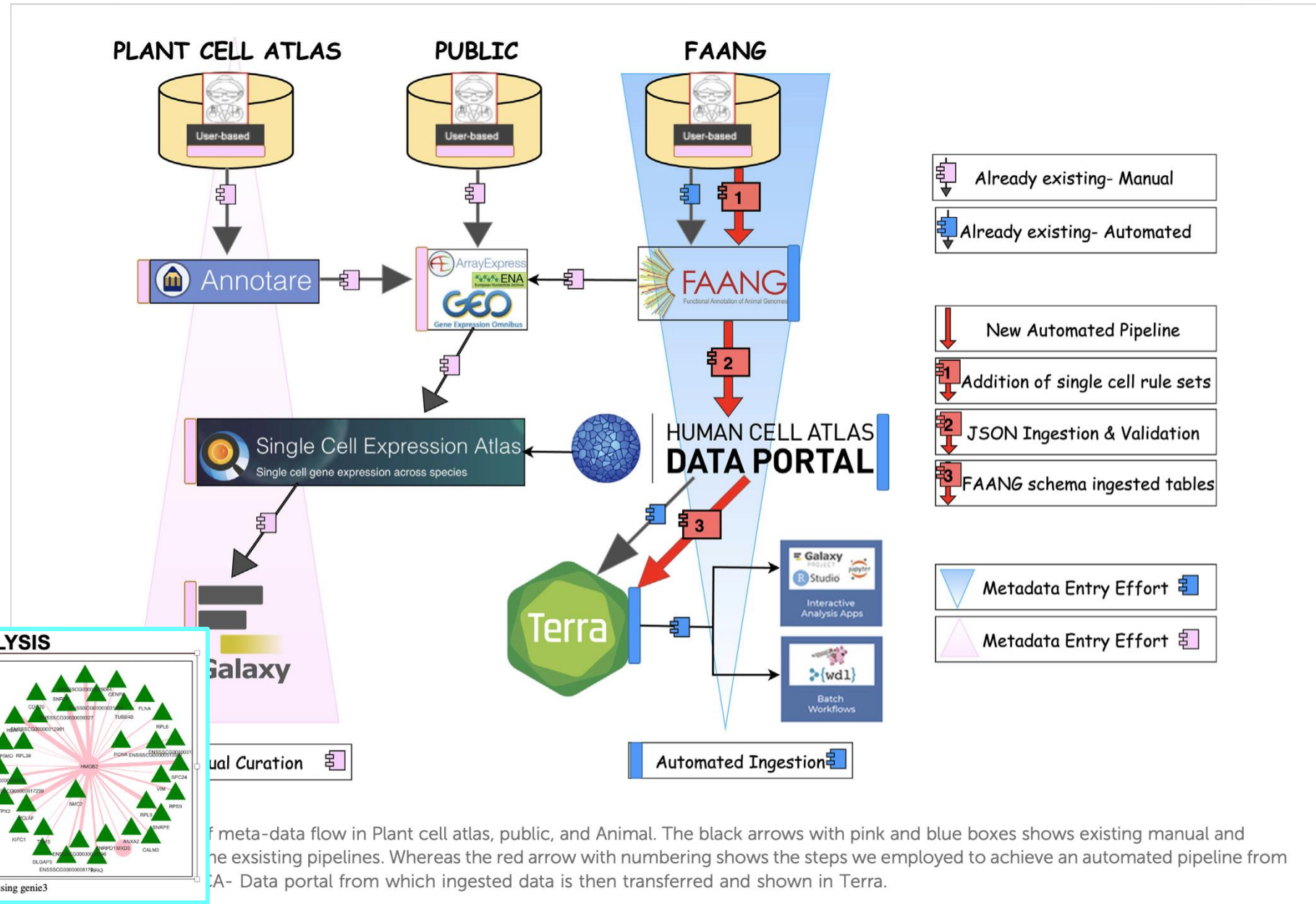
Example: Differences in plant and animal SC data/metadata flow



Improving SC data/metadata flow

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

Muskan Kapoor^{1*}, Enrique Sapena Ventura², Amy Walsh³, Alexey Sokolov², Nancy George², Sunita Kumari⁴, Nicholas J. Provart⁵, Benjamin Cole⁶, Marc Libault⁷, Timothy Tickle⁸, Wesley C. Warren⁹, James E. Koltes¹, Irene Papatheodorou^{10,11}, Doreen Ware^{4,12}, Peter W. Harrison², Christine Elisk³, Galabina Yordanova², Tony Burdett² and Christopher K. Tuggle^{1*}



of meta-data flow in Plant cell atlas, public, and Animal. The black arrows with pink and blue boxes shows existing manual and automated pipelines. Whereas the red arrow with numbering shows the steps we employed to achieve an automated pipeline from FAANG- Data portal from which ingested data is then transferred and shown in Terra.

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>

They also fund conferences!

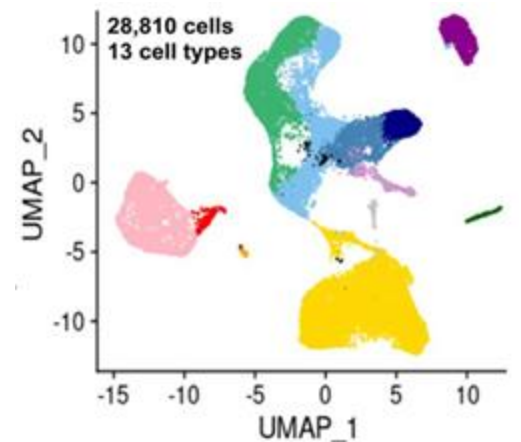
Supported by a grant from USDA
NIFA AG2PI #2024-70412-43649

First Announcement: Ag Single Cell Workshop!

“Empowering G2P in Agriculture:
building a collaborative environment
for single-cell applications”



March 29-30, 2025
Orlando, FL



To be held immediately prior to the AGBT-Ag meeting (same venue)
(<https://www.agbt.org/>)

Questions? Email the Organizing Committee: ag2pi.sc.comm.2025@iastate.edu

Aims to accomplish before SC Workshop

- **Aim: determine community needs for applying SC genomics techniques to agricultural systems.**
- We put out a survey to the agricultural research community to identify impediments to utilizing SC techniques to agricultural systems.
- Started mid-December (to > 4,600 email addresses (Angenmap, Plant phenotyping) as well as to the AgBioData mailing list.

Current survey results (preliminary- from 2 days ago...)

207 respondents, 85 from US, 36 Europe, 23 Asia, 21 Africa

108 no experience with SC sequencing, 99 with some level of experience

**AG2PI SC Survey
QR Code at end!**

Type of SC work planned: 115 “scRNAseq”, 54 “spatial transcriptomics”

Top interests in future Workshops:

“Connecting Biologists with data scientists interested in SC analysis” and

“Application of AI/machine learning in SC data analysis”

Planning to attend SC Workshop: 47 “yes”, 55 “maybe”

Aims to accomplish before SC Workshop

- **Aim: Create a community of researchers interested in applying SC genomics to elucidate genotype to phenotype in agricultural systems.**
- We will host **several** online information sessions prior to the conference.
- **Current schedule (Wednesdays 10:30-12):**
 1. Feb 5: Generating SC data (QC of sequencing data, comparison of modalities)
 2. Feb 12: Creating, submitting, sharing and accessing SC and meta-data
 3. Feb 26/Mar 5: Annotating cell types through community-based methods
 4. Mar 12/ Mar 19: Use of SC data for Genome 2 Phenome prediction

Single Cell Workshop Program March 29-30 Orlando

- **Organize and convene a conference on discussion and development of SC resources required for agriculture.**
- Purpose: develop working relationships, hear from experts working with medical and agricultural SC data and to discuss how to document and address the community's needs.
- **Saturday March 29 Session 1: Building a Transdisciplinary Community in Agricultural SC Genomics**
- **Jennifer Clarke**, University of Nebraska, NE
“Lessons learned from interdisciplinary projects: core values, orientation, and organization”
- **Sunday March 30 Session 2: Where we are in plant and animal SC genomics, as well as where we need to be?**
- **Sharon Greenblum**, Joint Genome Institute, Berkeley, CA
“Data management and infrastructure within the Plant Cell Atlas community”
- **Irene Papatheodorou**, Earlham Institute, Norwich, U.K.
“Data and meta-data management for re-using and integrating cell atlases”
- **Wes Warren**, University of Missouri, Columbia, MO
“A single nuclei and cell perspective on the chicken immune system”

Small Group Breakout Discussions

Session 3: Identifying best practices for:

1. Describing, storing and sharing FAIR data
2. Cell type annotations
3. Identifying and benchmarking relevant pipelines

Session 4: Achieving identified best practices for ag community:

1. Creating FAIR data storage/sharing
2. Creating common environments for data analysis
3. 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.

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

\$600 funding available for first 25 Early Career Researchers- will pay for all SC Workshop expenses! Email me!

We need your help!

AG2PI SC Survey



AG2PI SC Workshop info



- Take the **two-minute survey** -> -> -> ->
(will close Monday)
- ← ← **Attend the SC Workshop!**
- **Register for Workshop and main meeting.** -> -> -> ->
Deadlines for Abstract submission (Feb 5)
Deadlines for Early bird Registration (Feb 18)
- **Attend the on-line AG2PI SC Field Days (info coming)!**
- **Send us your ideas!**

AGBT-Ag meeting Reg



ag2pi.sc.comm.2025@iastate.edu

Thanks:

- **Workshop Organizing Committee**
- **AGBT for support for Workshop at the AGBT-Ag conference**
- **USDA-NIFA-AG2PI for funding # 2024-70412-43649**

AG2PI SC conference organizing Committee:

Chris Tuggle (Chair)	Iowa State University
Jennifer Clarke	University of Nebraska
Ben Cole	Lawrence Berkeley Lab
Marc Libault	University of Missouri
Fiona McCarthy	University of Arizona
Irene Papatheodorou	Earlham Institute
Sunita Kumari	USDA-ARS/Cold Spring Harbor Lab
Doreen Ware	USDA-ARS/Cold Spring Harbor Lab
Wes Warren	University of Missouri