MIDWEST SECTION MEETING
March 12-15 • Madison, Wisconsin
2023

Computational tools and resources for analysis and explorations of single-cell RNAseq data in agriculture

Author Name

Muskan Kapoor, Christopher Tuggle, Alexey Sokolov, Enrique Ventura, Galabina Yordanova, Nicholas Provart, Irene Papatheodorou, Nancy George, Doreen Ware, Sunita Kumari, Timothy Tickle, Lance Daharsh, James Koltes, Benjamin Cole, Marc Libault, Christine Elsik, Wesley Warren, Tony Burdett, Peter Harrison

- Abstract
- Introduction
- Materials & Methods
- Results
- Conclusion

Meet the Authors

TAP TO GO BACK TO KIOSK MENU

Muskan Kapoor

I am a second year PhD student in Dr Tuggle's research group. My research interests lies in Bioinformatics, Computational Biology and its applications such as tools development and data analysis.

Email: muskan@iastate.edu

Christopher Tuggle

I am a Professor of Molecular Genetics at the Department of Animal Science. I lead the USDA-funded Pig FAANG project, and my interests are in the functional analysis of the pig genome, with emphasis on immunity and disease resistance.

Email: cktuggle@iastate.edu

IOWA STATE UNIVERSITY
Bioinformatics & Computational Biology Graduate Program

FAANG Functional Annotation of Animal Genomes

AG2PI Agricultural Genome to Phenome Initiative

USDA

Computational tools and resources for analysis and explorations of single-cell RNAseq data in agriculture

MIDWEST SECTION MEETING 2023
March 12-15 • Madison, Wisconsin
Iowa State University, Department of Animal Science

TAP TO GO BACK TO KIOSK MENU

Abstract

The agriculture genomics community has numerous data submission standards available but little experience describing and storing single-cell (e.g., scRNAseq) data. Other single-cell genomics infrastructure efforts, such as the Human Cell Atlas Data Coordination Platform (HCA DCP), have resources that could benefit our community.

We will describe a pilot-scale project to determine if our current metadata standards for livestock and crops can be used to ingest scRNAseq datasets in a manner consistent with HCA DCP standards and if established resources (e.g., Terra) can be used to analyze the ingested data. For animal datasets, another EMBL-EBI portal, the FAANG portal, has been developed that provides bulk and scRNAseq data access. Currently, the most comprehensive data ingestion portal for high throughput sequencing datasets from plants, fungi, protists, and animals/humans is Annotare (located at EMBL-European Bioinformatics Institute), ensures that sufficient metadata are collected to enable re-analysis and dissemination via the Single Cell Expression Atlas (SCEA).

In an extension of these efforts, we have also created a Shiny-based web application, called Shiny-PIGGI, for the single cell-level transcriptomic study of pig immune tissues and peripheral blood mononuclear cells, which will be an important resource for improved annotation of porcine immune genes and cell types. The Shiny-PIGGI (https://shinypiggi.ansci.iastate.edu/) is implemented completely in R, runs on any modern web browser, and requires no programming skills to use. Our main goal was to develop an interactive web application that allows users such as animal scientists and immunologists to visualize and analyze biological datasets.

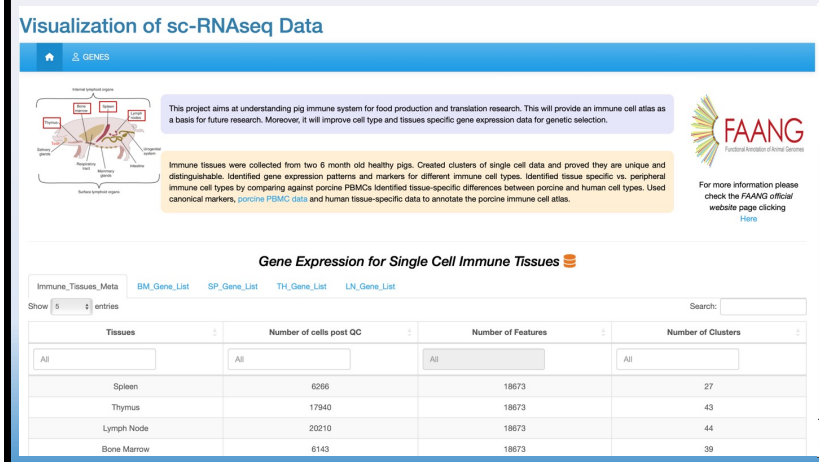
Computational tools and resources for analysis and explorations of single-cell RNAseq data in agriculture

TAP TO GO BACK TO KIOSK MENU

MIDWEST SECTION MEETING
March 12-15 • Madison, Wisconsin
2023

Introduction

- The HCA-DCP is integrated with Terra, a cloud-native workbench for computational biology developed by Broad, Verily, and Microsoft that houses tools for scGenomics analysis.
- Annotare supports user-directed annotation and processing of their data, as well as search tools via the SCEA and transferred to the Galaxy analysis space.
- Shiny tool thus increases accessibility through eliminating technical training requirements for using Seurat object and related R packages commonly used in scRNAseq analysis.



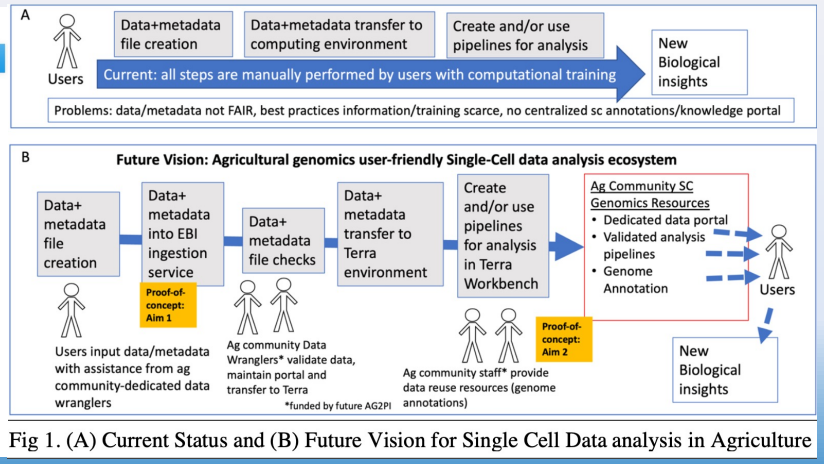


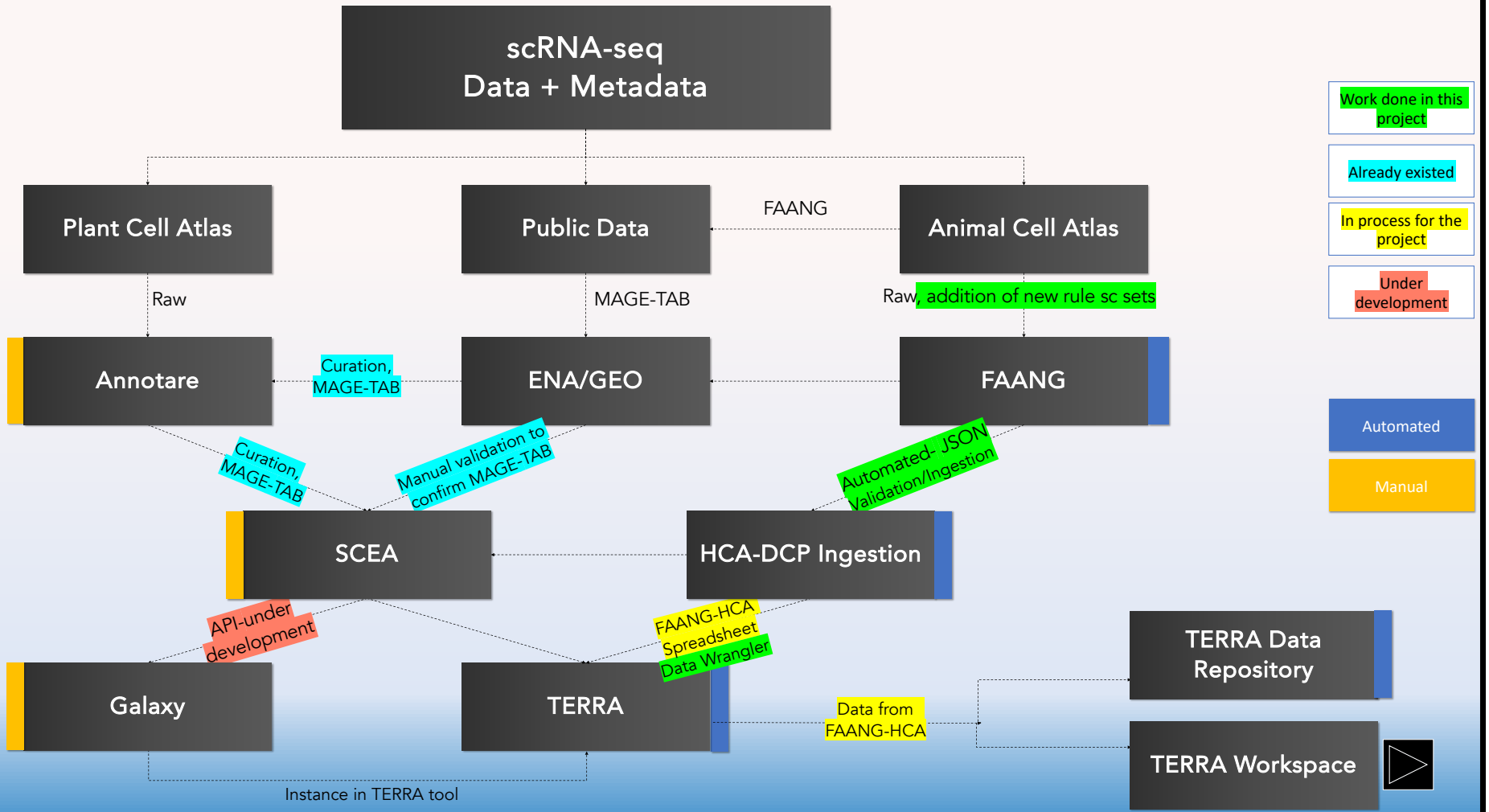
Fig 1. (A) Current Status and (B) Future Vision for Single Cell Data analysis in Agriculture

MIDWEST SECTION MEETING
March 12-15 • Madison, Wisconsin
2023

Computational tools and resources for analysis and explorations of single-cell RNAseq data in agriculture

Iowa State University, Department of Animal Science

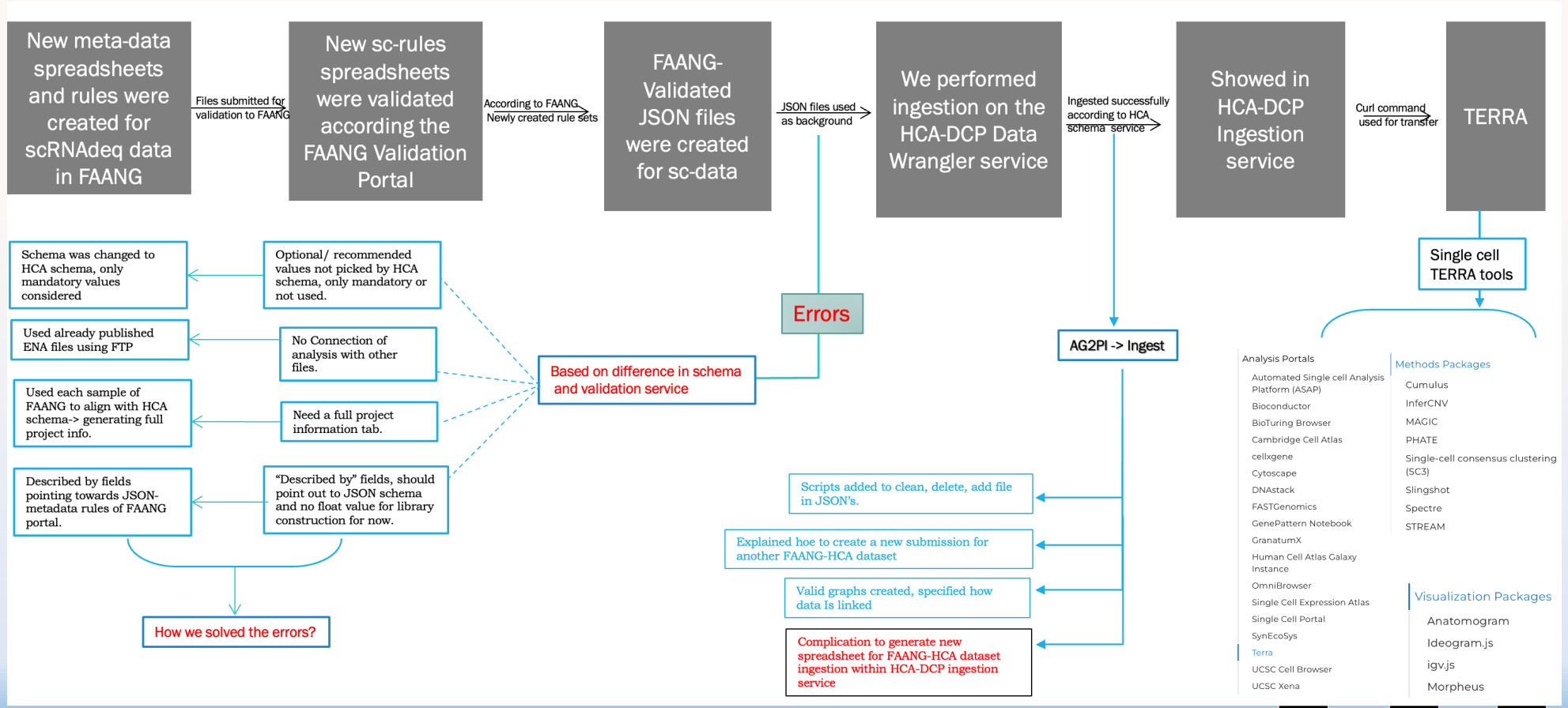
TAP TO GO BACK TO KIOSK MENU



Computational tools and resources for analysis and explorations of single-cell RNAseq data in agriculture

TAP TO GO BACK TO KIOSK MENU

Materials and Methods



Computational tools and resources for analysis and explorations of single-cell RNAseq data in agriculture

TAP TO GO BACK TO KIOSK MENU

Results-HCA-DCP Ingestion


```
"faang_hca": [
    {
        "alias": {
            "value": "Pig_Immune_Cells_RNA_FAANG_scRN
        },
        "project": {
            "value": "FAANG"
        },
        "secondary_project": {
            "value": null
        },
        "analysis_protocol_name": {
            "value": "PBMC_All Cells_10X format_scRNA
        },
        "analysis_protocol_description": {
            "value": "https://data.nal.usda.gov/datas
        },
        "analysis_protocol_type": {
            "value": "raw matrix generation"
        },
        "computational_method": {
            "value": "CellRanger v4.0.0"
        },
        "computation_method_version": {
            "value": null
        },
        "analysis_file_name": {
            "value": "matrix.mtx.gz"
        },
        "analysis_file_format": {
            "value": "mtx.gz"
        },
        "analysis_file_description": {
            "value": "Gene expression matrix"
```


Fig 1 FAANG validated JSON file.

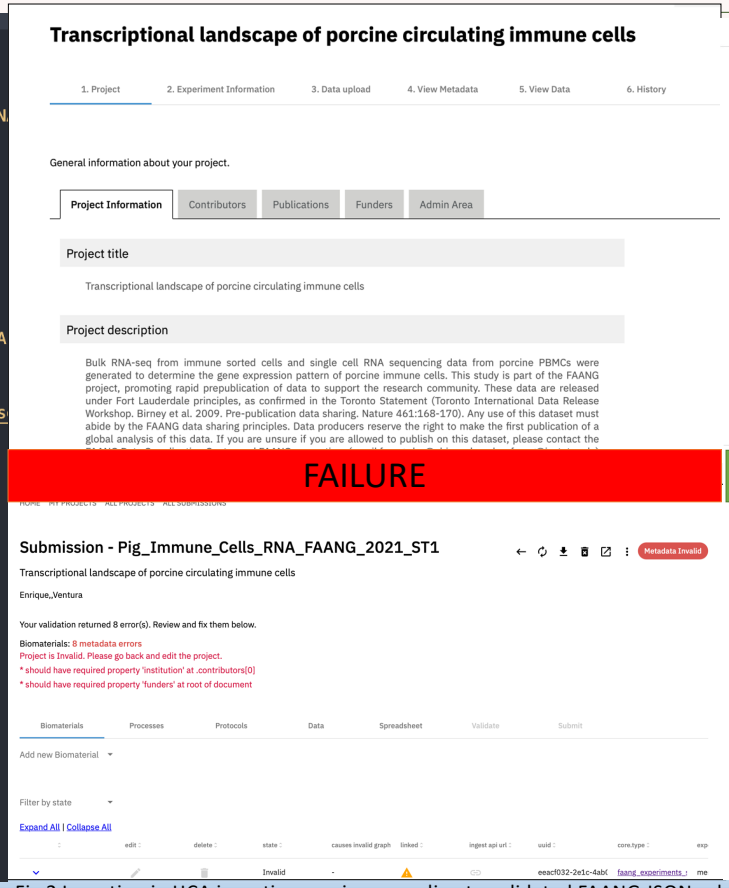


Fig 2 Ingestion in HCA ingestion service according to validated FAANG JSON rules

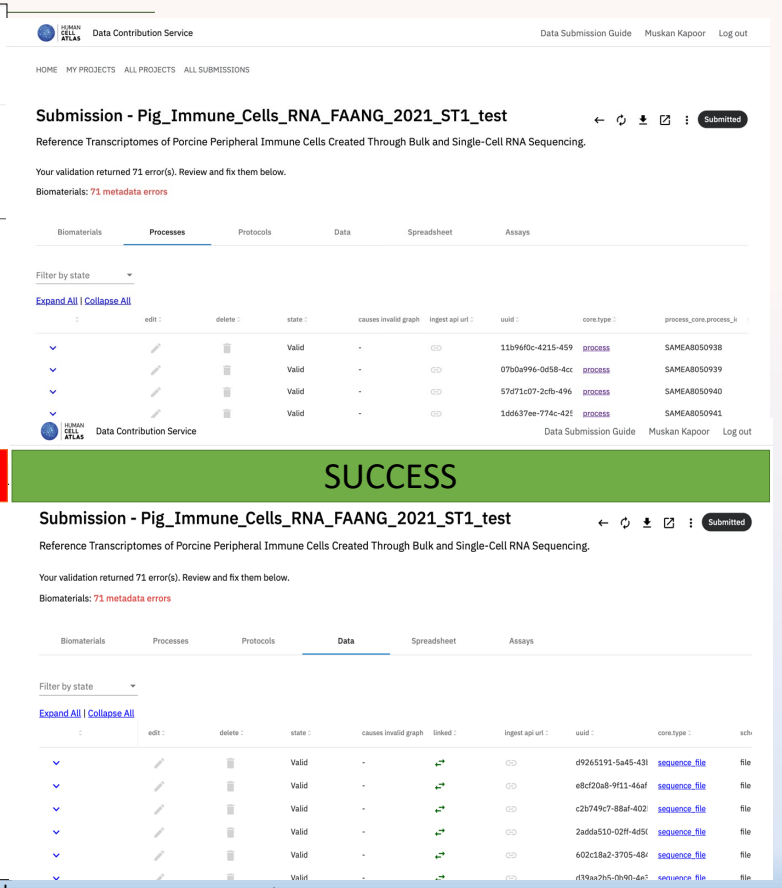
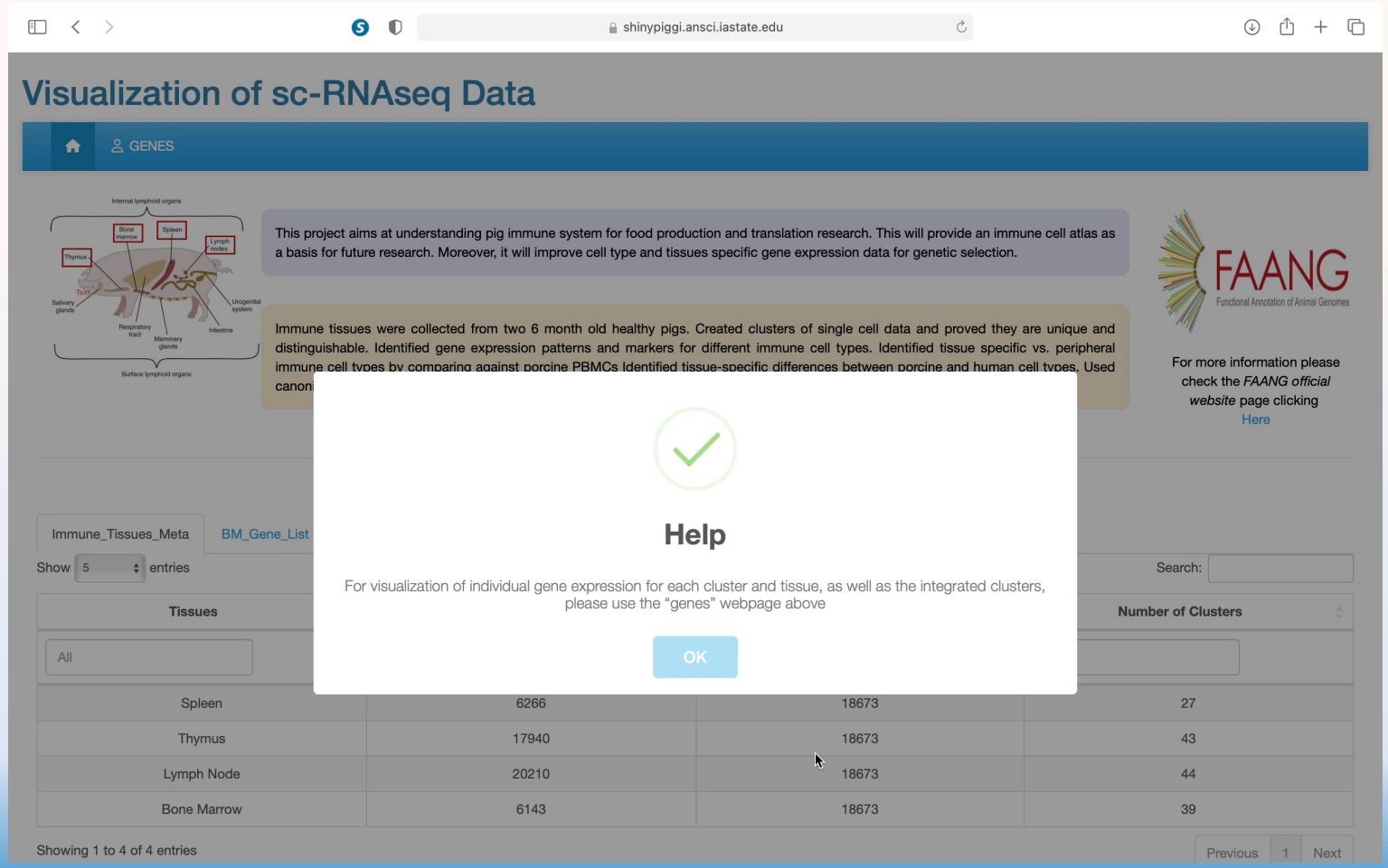


Fig 3: Ingested Data in HCA –DCP ingestion services

Computational tools and resources for analysis and explorations of single-cell RNAseq data in agriculture

TAP TO GO BACK TO KIOSK MENU

Results-Shiny App



SCAN ME

Fig 4 Created Shiny Tool for visualization of Four tissue mainly Spleen, lymph node, Thymus, Bone marrow and an integrated object Of four tissues.

The tool is hosted online at : https://shinypiggi.ansci.iastate.edu/

Computational tools and resources for analysis and explorations of single-cell RNAseq data in agriculture

TAP TO GO BACK TO KIOSK MENU

MIDWEST SECTION MEETING
March 12-15 • Madison, Wisconsin
2023

Iowa State University, Department of Animal Science

Conclusion & Future Scope

We intend to further build upon these existing tools to construct a scientist-friendly data resource and analytical ecosystem to facilitate single cell-level genomic analysis through data ingestion, storage, retrieval, re-use, visualization, and comparative annotation across agricultural species.

- Animal data: we will complete the data ingestion into Human Cell Atlas –DCP and test use of the data in the Terra environment. We will also complete and verify the ingestion of animal data into the scExpression Atlas.
- Plant path improvements: reducing the manual curation and validation needed to transfer data to $\boldsymbol{Annotare \rightarrow SCEA}$ or $\boldsymbol{ENA \rightarrow SCEA}$ and moving data from $\boldsymbol{SCEA \rightarrow Galaxy}$ through an API, for analysis-- potentially taking ideas/tools from FAANG for higher throughput.
- Shiny-PIGGI: we will add the Cluster annotation functionality to the tool and ask for user feedback.

Acknowledgement

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

[1]Bioinformatics and Computational Biology Program, Department of Animal Science, Iowa State University, Ames, IA 50011, U.S.A.
[2]EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, 12 U.K.
[3]Department of Cell and Systems Biology/Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, ON M5S 3B2, Canada.
[4]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724, USA.
[5]U.S. Department of Agriculture, Agricultural Research Service, NEA Robert W. Holley Center for Agriculture and Health, Cornell University, Ithaca, NY 14853, USA.
[6]Data Sciences Platform, The Broad Institute of MIT and Harvard, 415 Main Street, 21 Cambridge, MA 02142, U.S.A.
[7]DOE-Joint Genome Institute, Lawrence Berkeley National Laboratory, 1, Cyclotron Road, 16 Berkeley CA 94720, U.S.A.
[8]Department of Agronomy and Horticulture, Beadle Center N305, University of Nebraska-Lincoln, Lincoln NE 68588-0660, U.S.A.
[9]Division of Animal Science and Division of Plant Science and Technology, S134D Animal Science Research Center, University of Missouri-Columbia, Columbia, MO 65211
[10]Division of Animal Science, 440G/446 Life Sciences Center, University of Missouri-Columbia, Columbia, MO 65211