



MIDWEST SECTION MEETING
March 12-15 • Madison, Wisconsin

2023

Abstract

Introduction

Materials &
Methods

Results

Conclusion

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 Provar, 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

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





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



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.

Visualization of sc-RNAseq Data

This project aims at understanding pig immune system for food production and translation research. This will provide an immune cell atlas as a basis for future research. Moreover, it will improve cell type and tissues specific gene expression data for genetic selection.

Immune tissues were collected from two 6 month old healthy pigs. Created clusters of single cell data and proved they are unique and distinguishable. Identified gene expression patterns and markers for different immune cell types. Identified tissue specific vs. peripheral immune cell types by comparing against porcine PBMCs. Identified tissue-specific differences between porcine and human cell types. Used canonical markers, porcine PBMC data and human tissue-specific data to annotate the porcine immune cell atlas.

Gene Expression for Single Cell Immune Tissues

Tissues	Number of cells post QC	Number of Features	Number of Clusters
Spleen	6266	18673	27
Thymus	17940	18673	43
Lymph Node	20210	18673	44
Bone Marrow	6143	18673	39

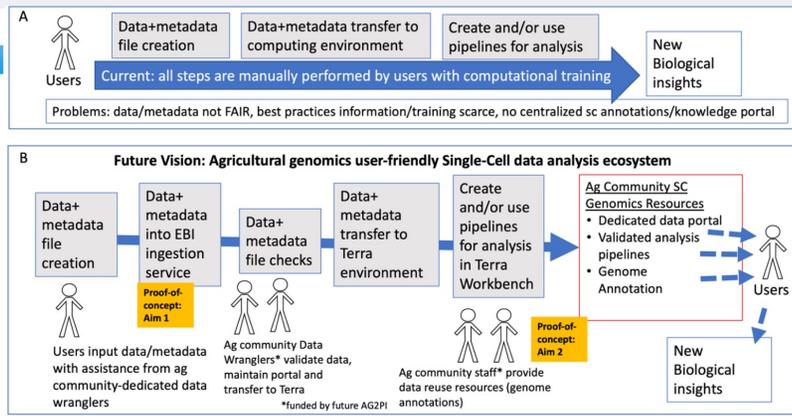
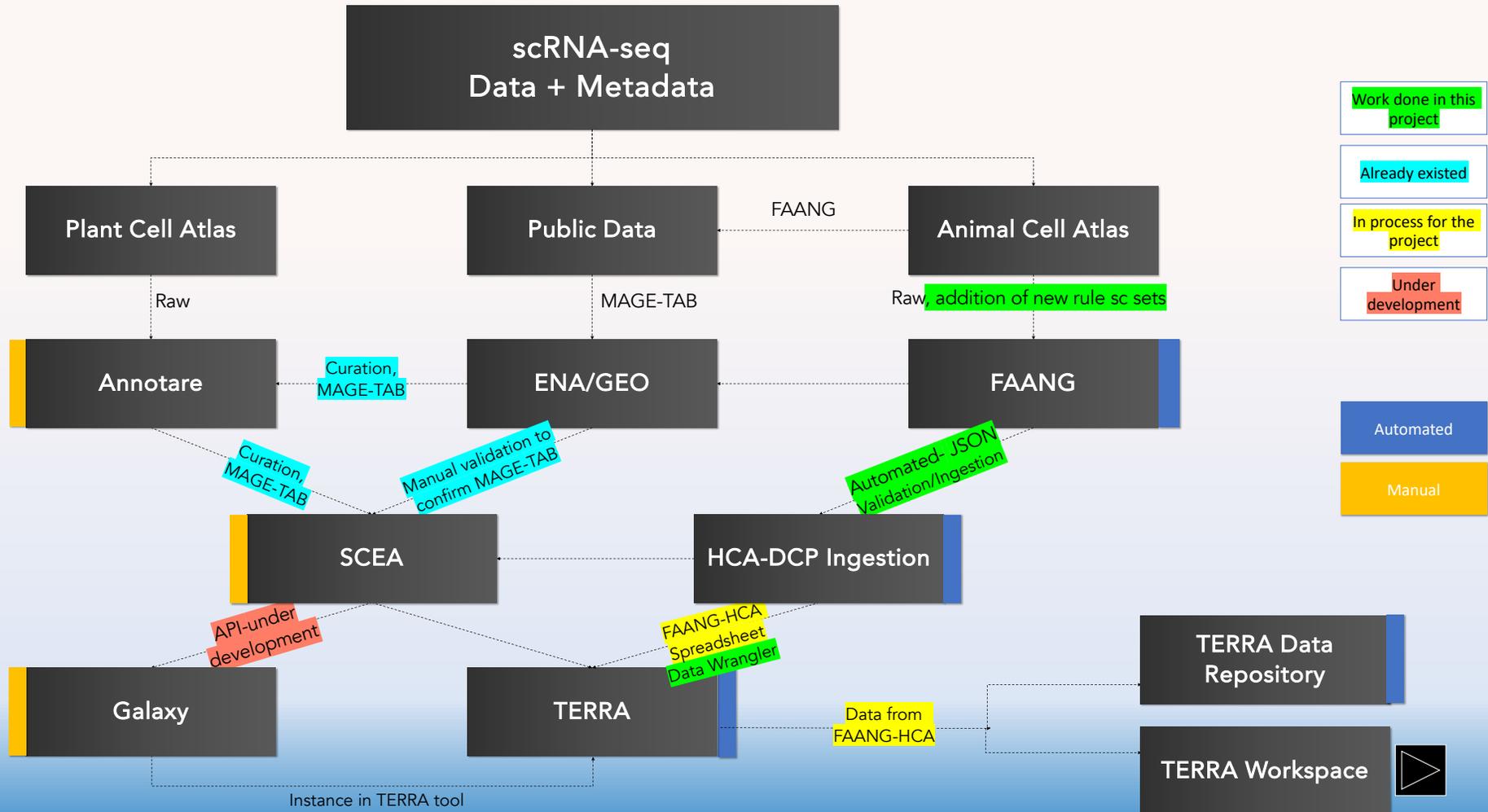


Fig 1. (A) Current Status and (B) Future Vision for Single Cell Data analysis in Agriculture





Work done in this project

Already existed

In process for the project

Under development

Automated

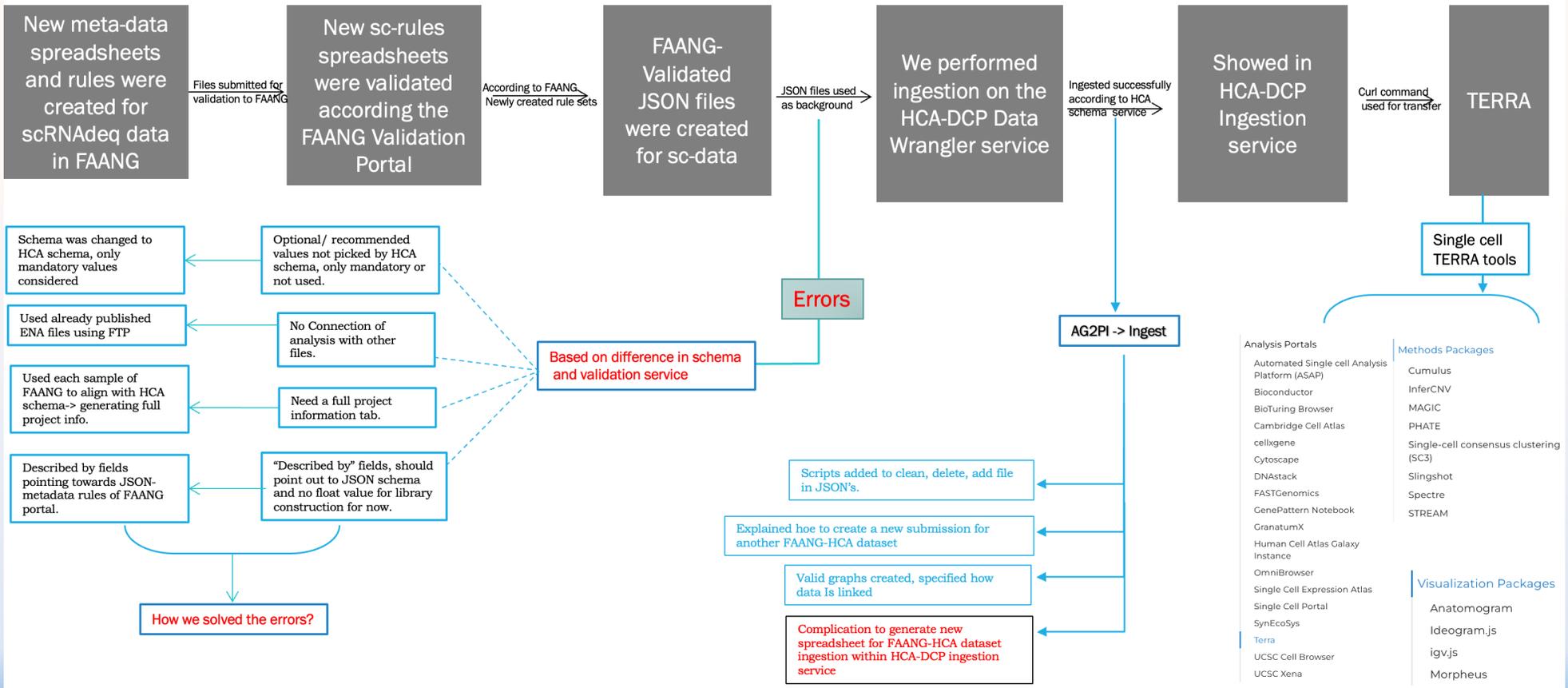
Manual



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



- | | | |
|--|--|---|
| Analysis Portals
Automated Single cell Analysis Platform (ASAP)
Bioconductor
BioTuring Browser
Cambridge Cell Atlas
cellxgene
Cytoscape
DNASTack
FASTGenomics
GenePattern Notebook
GranatumX
Human Cell Atlas Galaxy Instance
OmniBrowser
Single Cell Expression Atlas
Single Cell Portal
SynEcoSys
Terra
UCSC Cell Browser
UCSC Xena | Methods Packages
Cumulus
InferCNV
MAGIC
PHATE
Single-cell consensus clustering (SC3)
Slingshot
Spectre
STREAM | Visualization Packages
Anatomogram
Ideogram.js
igv.js
Morpheus |
|--|--|---|



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_scrRNA"
  },
  "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.

Transcriptional landscape of porcine circulating immune cells

1. Project 2. Experiment Information 3. Data upload 4. View Metadata 5. View Data 6. History

General information about your project.

Project Information Contributors Publications Funders Admin Area

Project title
Transcriptional landscape of porcine circulating immune cells

Project description
Bulk RNA-seq from immune sorted cells and single cell RNA sequencing data from porcine PBMCs were generated to determine the gene expression pattern of porcine immune cells. This study is part of the FAANG project, promoting rapid prepublication of data to support the research community. These data are released under Fort Lauderdale principles, as confirmed in the Toronto Statement (Toronto International Data Release Workshop, Birney et al. 2009. Pre-publication data sharing. Nature 461:168-170). Any use of this dataset must abide by the FAANG data sharing principles. Data producers reserve the right to make the first publication of a global analysis of this data. If you are unsure if you are allowed to publish on this dataset, please contact the

FAILURE

Submission - Pig_Immune_Cells_RNA_FAANG_2021_ST1

Transcriptional landscape of porcine circulating immune cells

Enrique_Ventura

Your validation returned 8 error(s). Review and fix them below.

Biomaterials: 8 metadata errors

Project is Invalid. Please go back and edit the project.

- * should have required property 'institution' at .contributors[0]
- * should have required property 'funders' at root of document

Biomaterials	Processes	Protocols	Data	Spreadsheet	Validate	Submit
Add new Biomaterial						

Filter by state

Expand All | Collapse All

edit	delete	state	causes invalid graph	linked	ingest api url	uuid	core.type	exp
		Invalid				eeecf032-2e1c-4abf	faang_experiments	me

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

Submission - Pig_Immune_Cells_RNA_FAANG_2021_ST1_test

Reference Transcriptomes of Porcine Peripheral Immune Cells Created Through Bulk and Single-Cell RNA Sequencing.

Your validation returned 71 error(s). Review and fix them below.

Biomaterials: 71 metadata errors

Biomaterials	Processes	Protocols	Data	Spreadsheet	Assays

Filter by state

Expand All | Collapse All

edit	delete	state	causes invalid graph	ingest api url	uuid	core.type	process_core.process_k
		Valid	-		11b96f0c-4215-459	process	SAMEA8050938
		Valid	-		07b0a996-0d58-4cc	process	SAMEA8050939
		Valid	-		57d71c07-2c6b-496	process	SAMEA8050940
		Valid	-		1d6d37ee-774c-42f	process	SAMEA8050941

SUCCESS

Submission - Pig_Immune_Cells_RNA_FAANG_2021_ST1_test

Reference Transcriptomes of Porcine Peripheral Immune Cells Created Through Bulk and Single-Cell RNA Sequencing.

Your validation returned 71 error(s). Review and fix them below.

Biomaterials: 71 metadata errors

Biomaterials	Processes	Protocols	Data	Spreadsheet	Assays

Filter by state

Expand All | Collapse All

edit	delete	state	causes invalid graph	linked	ingest api url	uuid	core.type	sch
		Valid	-			d9265191-5a45-431	sequence_file	file
		Valid	-			e8cf20a8-9f11-46af	sequence_file	file
		Valid	-			c2b749c7-88af-402	sequence_file	file
		Valid	-			2a6da510-02f7-4d5c	sequence_file	file
		Valid	-			602c18a2-3705-48a	sequence_file	file
		Valid	-			r19aa20f5-0b00-4e5	sequence_file	file

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

The screenshot shows a web browser displaying the 'shinypiggi.ansci.iastate.edu' website. The page title is 'Visualization of sc-RNAseq Data'. A 'GENES' tab is active. A diagram of a pig is shown with various tissues labeled: Internal lymphoid organs (Bone marrow, Spleen, Lymph nodes, Thymus) and Surface lymphoid organs (Salivary glands, Respiratory tract, Mammary glands, Intestine, Urogenital system). A text box explains the project's goal: 'This project aims at understanding pig immune system for food production and translation research. This will provide an immune cell atlas as a basis for future research. Moreover, it will improve cell type and tissues specific gene expression data for genetic selection.' Another text box states: 'Immune tissues were collected from two 6 month old healthy pigs. Created clusters of single cell data and proved they are unique and distinguishable. Identified gene expression patterns and markers for different immune cell types. Identified tissue specific vs. peripheral immune cell types by comparing against porcine PBMCs. Identified tissue-specific differences between porcine and human cell types. Used canonical gene sets for cell type identification.' A 'Help' dialog box is overlaid on the page, containing a green checkmark icon and the text: 'Help For visualization of individual gene expression for each cluster and tissue, as well as the integrated clusters, please use the "genes" webpage above'. Below the dialog box, a table displays the number of genes and clusters for four tissues: Spleen (6266 genes, 18673 clusters, 27 objects), Thymus (17940 genes, 18673 clusters, 43 objects), Lymph Node (20210 genes, 18673 clusters, 44 objects), and Bone Marrow (6143 genes, 18673 clusters, 39 objects). The page also features the FAANG logo (Functional Annotation of Animal Genomes) and a search bar.

Tissues	Number of Genes	Number of Clusters	Number of Objects
Spleen	6266	18673	27
Thymus	17940	18673	43
Lymph Node	20210	18673	44
Bone Marrow	6143	18673	39

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



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 *Annotare* → *SCEA* or *ENA* → *SCEA* and moving data from *SCEA* → *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 Kapoor¹, Alexey Sokolov², Enrique Sapena Ventura², Galabina Yordanova², Nicholas J. Provart³, Irene Papatheodorou², Nancy George², Doreen Ware^{4,5}, Sunita Kumari⁴, Timothy Tickle⁶, Lance Daharsh¹, James Koltes¹, Benjamin Cole⁷, Marc Libault⁸, Christine Elsik⁹, Wesley Warren¹⁰, Tony Burdett², Peter Harrison², and Christopher Tuggle¹

¹Bioinformatics and Computational Biology Program, Department of Animal Science, Iowa State University, Ames, IA 50011, U.S.A.

²EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, 12 U.K.

³Department of Cell and Systems Biology/Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, ON M5S 3B2, Canada.

⁴Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724, USA.

⁵U.S. Department of Agriculture, Agricultural Research Service, NEA Robert W. Holley Center for Agriculture and Health, Cornell University, Ithaca, NY 14853, USA.

⁶Data Sciences Platform, The Broad Institute of MIT and Harvard, 415 Main Street, 21 Cambridge, MA 02142, U.S.A.

⁷DOE-Joint Genome Institute, Lawrence Berkeley National Laboratory, 1, Cyclotron Road, 16 Berkeley CA 94720, U.S.A.

⁸Department of Agronomy and Horticulture, Beadle Center N305, University of Nebraska-Lincoln, Lincoln NE 68588-0660, U.S.A.

⁹Division of Animal Science and Division of Plant Science and Technology, S134D Animal Science Research Center, University of Missouri-Columbia, Columbia, MO 65211

¹⁰Division of Animal Science, 440G/446 Life Sciences Center, University of Missouri-Columbia, Columbia, MO 65211

