

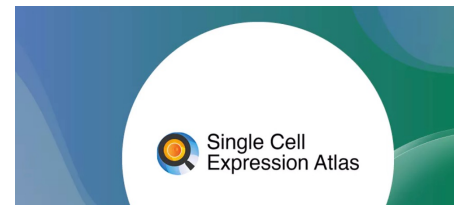


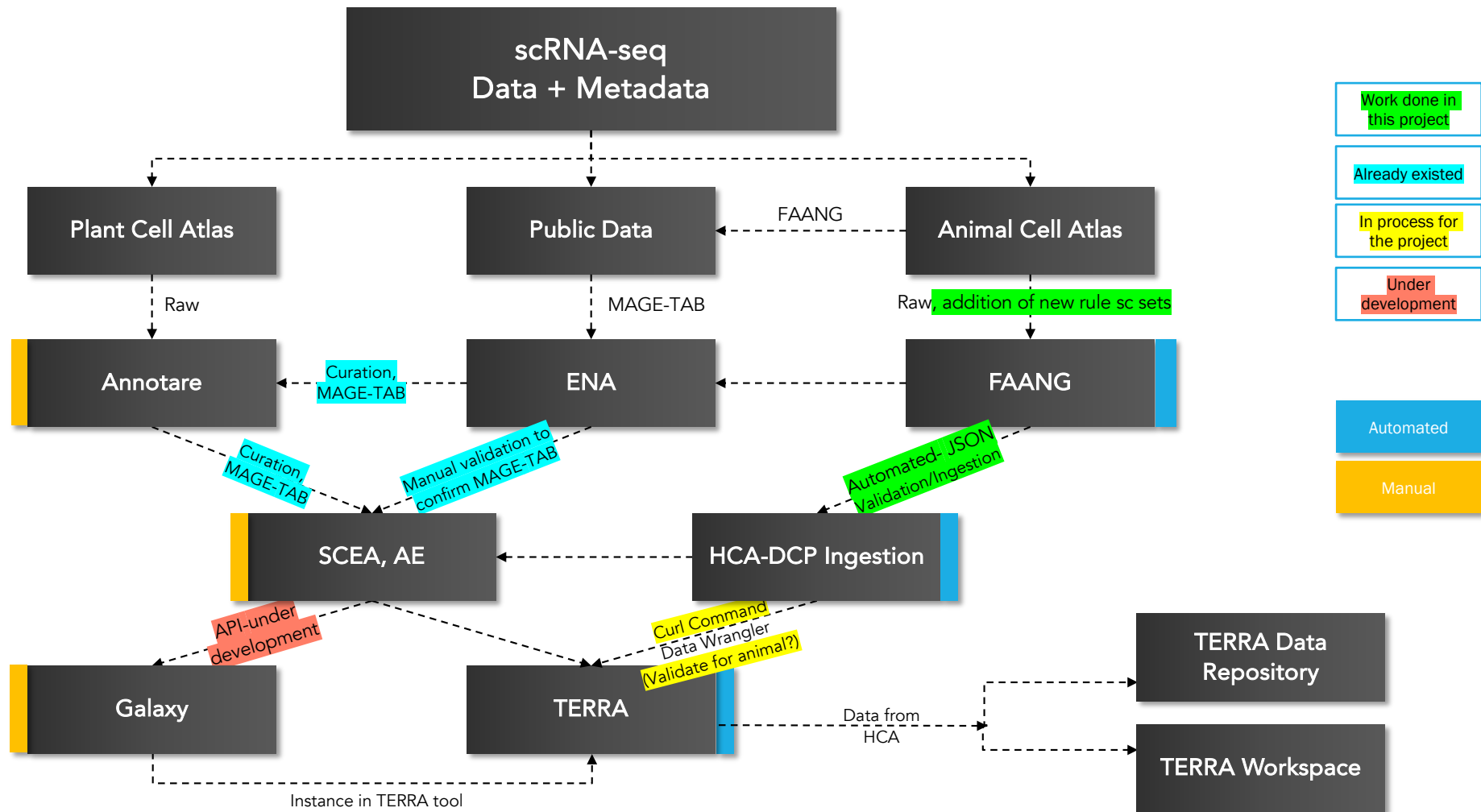
Creating a FAIR data ecosystem for
incorporating single cell genomics data into
agriculture G2P research



What is the Purpose of this study?

- The agriculture community has numerous data submission standards but little in describing and storing scRNAseq data
- There are existing single cell tools such as Human Cell Atlas Data Coordination Platform, TERRA and Single Cell Expression Atlas
- We want to leverage these tools to create resources for our community





Plant Data submission, visualization and analysis

Annotare

Welcome to Annotare

Annotare is a tool for submitting functional genomics data to the ArrayExpress collection in BioStudies. It generates MAGE-TAB format and supports high metadata standards in compliance with MIAME/MINSEQE guidelines.

Overview of the submission process

- Sign up & create new submission
- Fill in the metadata forms & upload/assign files
- Submit & receive accession number
(~ 24h)
- Curation review
(5-10 working days)
- Experiment loaded in BioStudies & get link for reviewers

Single Cell Expression Atlas

Single cell gene expression across species

Search across 20 species, 304 studies, 8,524,149 cells

Species: Any

Animals: Plants: Fungi: Protists:

Homo sapiens 131 experiments	Mus musculus 111 experiments	Drosophila melanogaster 18 experiments	Danio rerio 7 experiments	Gallus gallus 4 experiments	Schistosoma mansoni 2 experiments
---------------------------------	---------------------------------	---	------------------------------	--------------------------------	--------------------------------------

Galaxy / Human Cell Atlas

Tools

EBI SCXA Data Retrieval Retrieves expression matrixes and metadata from EBI Single Cell Expression Atlas (SCXA) (Galaxy Version v0.0.1+galaxy!)

SC-Atlas experiment accession
E-GEOD-100058

Choose the type of matrix to download

Raw filtered counts

Raw filtered counts or (non-filtered) TPMs

Execute

Gene expression analysis in single cells across species and biological conditions

Single Cell Expression Atlas supports research in single cell transcriptomics. The Atlas annotates publicly available single cell RNA-Seq experiments with ontology identifiers and re-analyses them using standardised pipelines available through IRAP, our RNA-Seq analysis toolkit. The browser enables visualisation of clusters of cells, their annotations and supports searches for gene expression within and across studies.

For more information check <https://www.ebi.ac.uk/gxa/sc/home>

EBI SCXA Data Retrieval

The data retrieval tool presented here allows the user to retrieve expression matrixes and metadata for any public experiment available at EBI Single Cell Expression Atlas.

To use it, simply set the accession for the desired experiment and choose the type of matrix that you want to download:

Raw filtered counts: This should be the default choice for running clustering and another analysis methods where you will introduce scaling and normalization of the data. The filtering is based on the quality control applied by IRAP prior to pseudo-alignment and quantification.

Non-programmer

Submission

- Submit meta-data forms to annotare
- An accession number is generated
- Submitted to curator or bioinformatician for review and validation

Storage & Visualization

- Atlas annotates publicly available scRNA data.
- Reanalyze them using standardized pipelines.
- Gene expression matrixes and UMAP plots.

Data Analysis

- Galaxy platform retrieves data from SCEA, HCA and other platforms.
- It mainly generates 4 output files: matrix, barcode, genes and metadata.
- Helps in analysis of sc-data using the UI driven single cell tools.

What we intend to do:

- **Plant path improvements:** reducing the manual curation and validation needed to transfer data to *Annotare* → *SCEA* and moving data from *SCEA* → *Galaxy* through an API, for analysis-- potentially taking ideas/tools from FAANG for higher throughput.
- We plan to construct a scientist friendly data resource and analytical ecosystem to facilitate single cell level genomic analyses through ingestion, storage, retrieval, reuse, visualize and compares annotation.

Useful URL's

- Annotare: <https://www.ebi.ac.uk/fg/annotare/login>
 - SCEA: <https://www.ebi.ac.uk/gxa/sc/home>
 - HCA: <https://www.humancellatlas.org>
 - TERRA: <https://terra.bio>
 - FAANG: <https://data.faang.org/home>
- 