FAIRifying single cell RNA-seq data from Single Cell Portal

Day 2: How FAIR are our data now?

Bio-IT FAIR Data Hackathon

Tuesday, May 15, 2018

Boston, MA, USA

Recap

- Single Cell Portal provides abstracts, visualizations, and data for studies on single cell RNA-seq

- Portal abstracts and data lack rich machine-readable FAIR metadata

Goals

- Improve alignment of single cell RNA-seq data with community metadata standards from Human Cell Atlas (HCA)
  - Study metadata
  - Analysis metadata
- Accessible usage license

Working example

https://portals.broadinstitute.org/single_cell/study/-single-nucleus-rna-seq-of-cell-diversity-in-the-adult-mouse-hippocampus-snuc-seq

Study metadata: Before

Single Cell Portal

Eric

Secure | https://portals.broadinstitute.org/single_cell/study/-single-nucleus-rna-seq-of-cell-diversity-in-the-adult-mouse-hippocampus-snuc-seq#study-summary

Single Cell Portal BETA > Study Overview

Help | eweitz

Study: Single nucleus RNA-seq of cell diversity in the adult mouse hippocampus (sNuc-Seq) 1402 cells

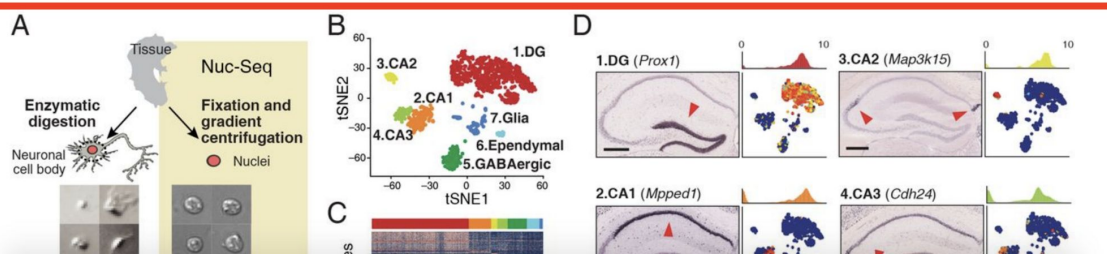
Summary | Explore | Download | Settings

Single nucleus RNA-seq of cell diversity in the adult mouse hippocampus.

Habib N, Li Y, Heidenreich M, Swiech L, Avraham-Davidi I, Trombetta J, Hession C, Zhang F, Regev A. **Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons.** *Science* 28 Jul 2016 DOI: 10.1126/science.aad7038
Contact: naomi@broadinstitute.org

Single cell RNA-Seq provides rich information about cell types and states. However, it is difficult to capture rare dynamic processes, such as adult neurogenesis, because isolation of rare neurons from adult tissue is challenging and markers for each phase are limited. Here, we develop Div-Seq, which combines scalable single-nucleus RNA-Seq (sNuc-Seq) with pulse labeling of proliferating cells by EdU to profile individual dividing cells. sNuc-Seq and Div-Seq can sensitively identify closely related hippocampal cell types and track transcriptional dynamics of newborn neurons within the adult hippocampal neurogenic niche, respectively. This study contains the sNuc-Seq analysis performed as a part of the Div-Seq method development.

Using sNuc-Seq, we analyzed 1,367 single nuclei from hippocampal anatomical sub-regions (DG, CA1, CA2, and CA3) from adult mice, including enrichment of genetically-tagged lowly abundant GABAergic neurons (*9*). sNuc-Seq robustly generated high quality data across animal age groups (including 2 years old mice), detecting 5,100 expressed genes per nucleus on average, with comparable complexity to single neuron RNA-Seq from young mice (*1*, 2, *3*). Analysis of sNuc-Seq data revealed distinct nuclei clusters (Fig. 1B-D shown below) corresponding to known cell types and anatomical distinctions in the hippocampus.



Study metadata: Use cases and tasks

User story: provide FAIR metadata using HCA community standards about studies

Use case: SCP has prose abstracts, needs to detect organism, tissue, etc. and map to ontology URL

NLP analysis: Soheil Danesh

Use case: Ontology mapping from NLP is list of key-value pairs, needs structure defined in community metadata standard

TSV-JSON transform: Tom Madden, David Managadze, Frank O

Use case: Community metadata standard uses text values, not URLs

FAIRify HCA metadata: Kenny Knecht, Adelaide Rhodes, Tom Madden

Use case: Community metadata standard uses JSON Schema, but ontologies often use OWL

JSON-OWL transform: Morgan Wahl

Use case: Tie together all these tasks in a pipeline

Pipeline: Alex Baumann, Tim Lee

Use case: Provide adapters to convert from HCA metadata standard to GEO metadata standard

Converters: Etienne Gnimpieba, Tayler Hoekstra, Isaac Hanson

Use case: NLP can't provide all the metadata we need

Augment study creation UI/UX: Anthony Dubois, Maya Bobrovitch, Kate Voss, Morgan Wahl

From NLP to metadata

Project

Project title: Atlas of human blood dendritic cells and monocytes

Project description: Dendritic cells (DCs) and monocytes consist of multiple specialized subtypes that play a central role in pathogen sensing, phagocytosis, and antigen presentation. However, their identities and interrelationships are not fully understood, as these populations have historically been defined by a combination of morphology, physical properties, localization, functions, developmental origins, and expression of a restricted set of surface markers.

Classification

Genus species: Homo sapiens

Organ: Blood

Organ part: Blood dendritic cells and monocytes

Disease: Disease

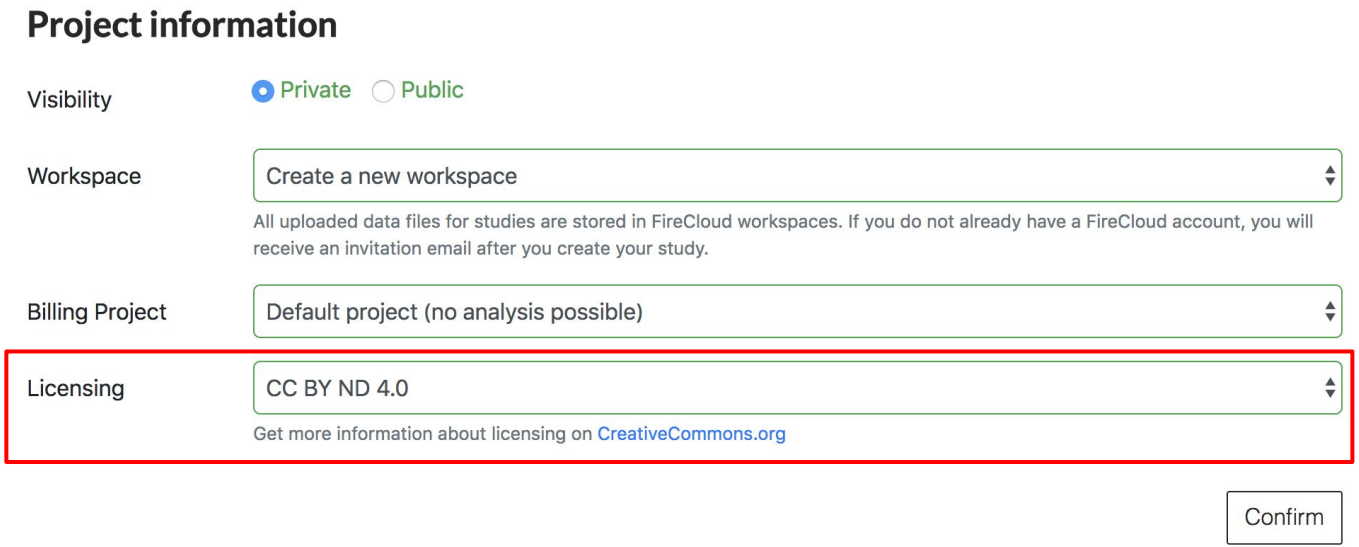
Please provide a valid disease.

1 - The user completes the project title and description.

2 - A NLP algorithm fills the form and the user can edit it using a standard dictionary (auto complete feature).

This information will be used to complete the metadata file.

Licensing and metadata generation



3 - The user fills project information and a licensing type.

4 - Create the project and generate the metadata based on the Human Cell Atlas metadata schema.

Demo

http://ec2-18-218-212-189.us-east-2.compute.amazonaws.com:5000/

Analysis metadata: Use cases and tasks

- **Use case**: Analysis metadata conforms to community standards, but audience is restricted
  - **Enable public analysis.json**: Jon Bistline

- **Use case**: Analysis metadata is not easily machine-findable
  - **Enable analysis.json search**: Jon Bistline

Relation to FAIR Metrics

- FAIR Evaluator survey is just a start
- Single Cell Portal Resources has multiple resource types
- “Study” resource type increased in Interoperability, Reusability
- “Analysis” resource type increased in Findability, Accessibility

Collaborative coding

https://github.com/BioITHackathons/single_cell_portal_core

45 commits from 12 people in < 2 days

Rapid onboarding to completely new project for 20 of 22 team members

Several people learned how to use Git and GitHub!

Thank you!

Eric Weitz
Jon Bistline
Kate Voss
Alex Baumann
Soheil Danesh
Austin Chow
Adelaide Rhodes

Maya Bobrovitch
Tim Lee
Tom Madden
Brian Kang
Brad Nissenbaum
Frank O
Morgan Wahl

Kenny Knecht
Julia Ivashina
David Managadze
Etienne Gnimpieba
Tayler Hoekstra
Isaac Hanson
Asya Lushnikova
Anthony Dubois