Plan Overview

A Data Management Plan created using DMPonline

Title: Approximate Bayesian inference for identifying sub-populations of cells in large scale single cell RNA-Seq data

Creator: Tom Thorne

Principal Investigator: Tom Thorne

Data Manager: Tom Thorne

Affiliation: University of Surrey

Template: EPSRC Data Management Plan

ORCID iD: 0000-0002-7396-5116

Project abstract:

This project will provide solutions to identifying sub-populations of cells in large scale single-cell RNA-Seq (scRNA-Seq) data sets, by deriving variational inference schemes for single cell RNA-seq clustering models that we have developed.

ID: 69319

Last modified: 13-01-2021

Grant number / URL: EP/V032593/1

Approximate Bayesian inference for identifying sub-populations of cells in large scale single cell RNA-Seq data

Data Collection

What data will you collect or create?

Data generated will consist of open source software and notebooks documenting analyses.

How will the data be collected or created?

The data will be generated by the PI and PDRA working on the project through implementation of the methods developed in the project, and through the analysis of data. In analysing data we will produce quantitative measures of gene expression within sub-populations identified in a dataset. The process of analysis itself will also be recorded in an electronic notebook format.

Documentation and Metadata

What documentation and metadata will accompany the data?

The software implementation will include documentation explaining how data can be processed with the software. We will also provide a vignette walking through an analysis of publicly available data. Analyses will be created as Jupyter notebooks with the code to reproduce the analysis alongside text explaining the analysis being performed.

Ethics and Legal Compliance

How will you manage any ethical issues?

We do not envision any ethical issues in the generation of software and analysis of data over the course of the project.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

All software and data we propose to work with is publicly available. We do not plan to re-distribute existing software. Our software will be licensed through an appropriate open source license.

Storage and Backup

How will the data be stored and backed up during the research?

We will ensure that all data, including manuscript drafts, software, and analysis, is backed up using the University of Surrey provided OneDrive service as it is being worked on.

How will you manage access and security?

Access to data and security will be managed by the University of Surrey core IT services.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Data that should be preserved in the long term consists of the open source software we produce and the data analyses (Jupyter notebooks) we conduct, to ensure the work is reproducible.

What is the long-term preservation plan for the dataset?

To ensure that project outputs are made available for at least 10 years following UKRI policy, all outputs (Jupyter notebooks, software packages) will be archived on Zenodo, which will also provide a DOI for each output that can be cited. We will also make use of container platforms such as Singularity and nf-core to build reproducible data analysis pipelines, where appropriate.

Data Sharing

How will you share the data?

Software we produce will be shared on the open source platform Github where it is freely available, and also on the Zenodo platform which will archive software for at least 10 years.

Are any restrictions on data sharing required?

No restrictions are required.

Responsibilities and Resources

Who will be responsible for data management?

The PI will be responsible for data management during the course of the project. This includes:

- Data Acquisition and Data Capture
- Data Backup
- Data Management
- Data Retention

What resources will you require to deliver your plan?

We require services provided by the University of Surrey for backup (OneDrive), and open platforms for data sharing and long term storage (Github, Zenodo).