EXAMPLE DATA MANAGEMENT AND SHARING PLAN (in compliance with SF-424 Forms H)

EXAMPLE FOR GENE EXPRESSION ANALYSIS FROM A NON-HUMAN MODEL ORGANISM (ZEBRAFISH)

ELEMENT 1: DATA TYPE

A. Types and amount of scientific data expected to be generated in the project:

Approximately 100 wild types and 100 phox2bb:mNeonGreen reporter transgenic fish embryos will be used for experiments generating scientific data.

Whole mount in situ hybridization will be performed on phox2bb:mNeonGreen zebrafish embryos for genes of interest using digoxigenin-labeled antisense RNA probes. In situ hybridization data will be collected, managed, and analyzed as .tiff files.

Transcriptional profiling will be assessed from NeonGreen labeled cells using bulk RNA-seq analysis, single cell clustering and expression analysis, and/or single cell RNA-seq (scRNA-seq). Detailed cisregulatory and open chromatin analysis data will be generated from single cell ATAC-seq. Raw bulk and/or scRNA-seq and ATAC-seq data will be collected as fastq files. Standard workflows will be used to transform the bulk RNA-seq, scRNA-seq and ATAC-seq fastq files to matrices and bed files as is usual practice.

B. Scientific data that will be preserved and shared, and the rationale for doing so:

All scientific data including raw/measured and derived data (as described in Section 1A) will be preserved and shared, for the purposes of reproducibility and reusability.

C. Metadata, other relevant data, and associated documentation:

Project, study, sample, experimental, and file level metadata about transcriptional, single cell, and ATAC-seq profiling will be collected using templates provided by GEO and SRA data repositories. Study and file-level metadata and documentation for in situ hybridization datasets will be provided to the Zebrafish Information Network (ZFIN) as required by the repository.

ELEMENT 2: RELATED TOOLS, SOFTWARE AND/OR CODE:

All analysis routines to transform raw sequences into matrices and bed files will be repurposed from existing methods or newly created using accessible and available methods and written in standard analysis languages such as R or Python. Analyses will be captured in Jupyter lab notebooks. All methods and codes used or created will be discoverable through GitHub.

ELEMENT 3: STANDARDS:

Transcriptional profiling data and open chromatin analysis (from RNA-seq and ATAC-seq data respectively): The GEO submission procedure is aligned to the MINSEQE (Minimum Information About a Next-generation Sequencing Experiment) guidelines which outline the minimum information that should be included when describing a sequencing study. SRA will provide internationally recognized project, study, sample, and experimental accession identifiers upon data submission to the repository. In situ hybridization data: ZFIN requires that all data submitted to the repository have a minimum set of available information (metadata) that describes the data, including Genes/Markers, Figures, Expression, Phenotype, Mutation and Transgenics, Sequence Targeting Reagents, Fish, Antibodies, Orthology, Engineered Foreign Genes, Mapping, Errata and Notes. Relevant metadata will be submitted to ZFIN per the repository requirement.

Commented [JB1]: *These example DMS Plans are provided for educational purposes to assist applicants with developing Plans but are not intended to be used as templates and **their use does not guarantee approval by NIH. Do not copy/paste this Plan without modifying it to reflect the types of data that are expected to be generated through your project.***

Note that the example DMS Plans may reflect additional expectations established by NIH or specific NIH Institutes, Centers, or Offices that go beyond the DMS Policy. Applicants will need to ensure that their Plan reflects any additional, applicable expectations (including from NIH policies, ICO policies, or as stated in the FOA).

In addition, these examples may reflect resources or policies that are in place at other institutions but that are not necessarily available at the University of Iowa. If needed, investigators can contact Research Data Services (lib-data@uiowa.edu) if they have questions regarding how to best complete their DMS Plan.

Commented [BJY2]: *Example from NICHD https://www.nichd.nih.gov/sites/default/files/inline-files/ExampleDMSPlan_NonHuman_Zebrafish_NIHFormatPage_v2.pdf*

ELEMENT 4: DATA PRESERVATION, ACCESS, AND ASSOCIATED TIMELINES

A. Repository where scientific data and metadata will be archived:
The expression data will be shared through the GEO and the bulk RNA-seq, scRNA-seq and ATAC-seq fastq files will be shared in the SRA data repository. GEO will manage deposition of the raw fastq files to SRA.

The in-situ hybridization data will be shared through the ZFIN data repository and knowledgebase.

B. How scientific data will be findable and identifiable:
The expression matrices, cluster and bed files derived from processing bulk RNA-seq, scRNA-seq and ATAC-seq fastq data respectively will be findable and identifiable through GEO accession numbers. The raw sequence data will be findable and identifiable through accession numbers assigned by SRA for project, study, and experiment.

Data released through ZFIN will be assigned a ZFIN ID by the repository, which when appended to the ZFIN url can be used to find and identify the dataset.

C. When and how long the scientific data will be made available:

Data will be submitted by the project team to the identified repositories at least 3 times: 1 year after award, 4 months prior to publication dates, and 4 months prior to end of award.

Bulk and single cell RNA-seq and ATAC-seq: Data underlying publications will be shared at the time of publication by release through GEO and SRA. Remaining data will be released in GEO and SRA by the end of the funded project period.

In situ hybridization data: Once submitted to ZFIN, data will be released as part of the normal software release cycle. The PI and members of the project team will work with data curators at ZFIN to ensure all data have been released by the time of any publication using the data or the end of the funded project period, whichever is sooner.

Study data deposited in GEO and ZFIN will be available to the research community in perpetuity.

ELEMENT 5: ACCESS, DISTRIBUTION, OR REUSE CONSIDERATIONS

A. Factors affecting subsequent access, distribution, or reuse of scientific data:

There are no restrictions on subsequent access, distribution, or reuse of the scientific data from this project.

B. Whether access to scientific data will be controlled:

Data will be open-access and available publicly.

C. Protections for privacy, rights, and confidentiality of human research participants:

Not applicable as this project does not have human research participants.

ELEMENT 6: OVERSIGHT OF DATA MANAGEMENT AND SHARING:

The contact PI for the project is Dr. Investigator at the University of Somewhere. Dr. Investigator will meet monthly with members of the research project team, Dr. Collaborator (University of Somewhere, imaging data collection), Ms. Manager (University of Somewhere, data system manager), and Dr. Colleague (University of Elsewhere, collection of tissue sample) to ensure that data collection, management, and submission to the repositories occur in a manner compliant with this Data Management and Sharing Plan.