
DATA MANAGEMENT AND SHARING PLAN

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on [sharing.nih.gov](https://www.nih.gov/data-management/data-sharing). The Plan is recommended not to exceed two pages. Text in italics should be deleted. There is no "form page" for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format* shown below.

Public reporting burden for this collection of information is estimated to average 2 hours per response, including the time for reviewing instructions, searching existing data sources, gathering, and maintaining the data needed, and completing and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to: NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-0001 and 0925-0002). Do not return the completed form to this address.

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 .png and .jpeg 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 cis-regulatory 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.

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.