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. 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.

Information and things to consider when completing YOUR Data Management and Sharing Plan:
The document below includes pre-filled fields for those that will share data on the SPARC portal. These answers are provided by the SPARC Data Resource Center (DRC), and for most cases, should provide enough detail for those sections. The following checklist should be completed before submission.

- ☐ Reference NIH FAQ and NIH Writing a DMS plan if you would like more information
- ☐ Review those sections in black and modify them if they are not specific for your project.
- ☐ Add information specific to your project where indicated
- ☐ Delete ALL red text before submission
- ☐ **DO NOT include hypertext (e.g., hyperlinks and URLs)**
- ☐ NIH recommends a length of up to two pages; however, some research projects may generate a larger number of data types, which may necessitate more detail and exceeding two pages.

Element 1: Data Type

A. Types and amount of scientific data expected to be generated in the project:
Summarize the types and estimated amount of scientific data expected to be generated in the project, some examples (provided by NIH) are below:

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:
Data described in section A will be shared in a meaningful, impactful, and FAIR way on the SPARC Portal (sparc.science); including a landing page, DOI and associated information that allows researchers to quickly understand the experiment, access, download, and cite our data for reuse or reproduction.

C. Metadata, other relevant data, and associated documentation:
Data shared on the SPARC Portal will adhere to the SPARC Minimal Information Standard (MIS) DOI: 10.5281/zenodo.5348189, include additional detailed dataset metadata and associated published

protocols on protocols.io (receives DOI).

Element 2: Related Tools, Software and/or Code:

Summarize the types and estimated amount of related tools, software and/or code expected to be generated in the project. If these are created as a part of your project, the following statement can be included. Optional: add additional information as you see fit

Any specialized code or software required to access or manipulate data shared on sparc.science will be listed and will also be shared when possible. It will be accessible either directly on the dataset landing page, or via a link to the appropriate repository; o²S²PARC for simulations/models/data analysis, GitHub for code, or the software landing page.

Element 3: Standards:

Contributions to the SPARC Portal are required to adhere to the SPARC Minimal Information Standard (MIS) and the SPARC Data Structure (SDS; DOI:10.1101/2021.02.10.430563), SPARC's required method to organize and name dataset files along with mandatory dataset and file descriptions.

Element 4: Data Preservation, Access, and Associated Timelines

A. Repository where scientific data and metadata will be archived:

Data submitted to the SPARC repository is stored on NIH Managed AWS accounts leveraging the Pennsieve Data Management Platform, which is developed and supported by the University of Pennsylvania. All submitted metadata is stored in the SPARC Knowledge-base and managed at the University of California San Diego (UCSD).

Data will be available on the SPARC Portal (sparc.science).

B. How scientific data will be findable and identifiable:

Data and computational models are findable and identifiable on SPARC, which includes a DOI, dataset landing page, and key information about the experiment. SPARC requires that datasets are accompanied with rich metadata, and maps this metadata to community ontologies to enhance findability. The SPARC Portal includes multiple search options that surface datasets of interest to users.

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

The data will be available to the public on the SPARC Portal before the time of publication of the associated manuscript or the end of performance (whichever comes first). Data will be retained on SPARC for a minimum period of 10 years, or for the duration of available funding if this exceeds 10 years. In case of closure of the repository, all content including the enriched metadata will be integrated into suitable alternative repositories and DOIs will be updated to redirect users to the correct location of the data.

Element 5: Access, Distribution, or Reuse Considerations

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

Data published on the SPARC Portal are available under a CC-BY-4.0 license, allowing users to copy, modify and distribute data in any format for any purpose, including commercial use. Users are only obligated to give appropriate credit (attribution) and indicate if they have made any changes, including translations.

To ensure publication of data meets manuscript publication or grant reporting deadlines; we will consult with SPARC well in advance of these expected dates, as the time required to complete this process can vary based on multiple factors.

B. Whether access to scientific data will be controlled:

Timing of data publication on the SPARC Portal is controlled by the data contributors. However, the SPARC Data and Resource Center (DRC) requires that datasets meet a high level of quality before they are published on the SPARC Portal. To ensure that dataset contributors meet these requirements, the SPARC Portal provides documentation, tools and templates to aid dataset contribution. In addition, the

SPARC Curation team offers guidance and recommends dataset contributors touch base early in the data collection process.

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

If you do not have human research participants, please provide the following statement and delete the sentence below.

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

If you have human research participants, please delete the statement above and include the statement below. Please refer to U.S Department of Health & Human Services guidance on de-identification of protected health information (PHI) in accordance with the HIPAA Privacy Rule. Optional: Add additional information as you see fit.

SPARC requires that prior to dataset submission, all data are anonymized, therefore no longer considered to be human subjects data. SPARC dataset publication workflow includes a quality assurance process to minimize the risk of sharing protected health information (PHI). If potential PHI is detected in the repository, the affected information will be quarantined and the primary investigator will be notified to rectify the issue.

Element 6: Oversight of Data Management and Sharing:

Please add information that describes how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles). If possible, include who in your group will be the primary contact with the SPARC DRC. An example from University of Washington below:

Lead PI _[name]_, ORCID: _[ORCID ID]_, will be responsible for the day-to-day oversight of lab/team data management activities and data sharing. Broader issues of DMS Plan compliance oversight and reporting will be handled by the PI and Co-I team as part of general [campus(es)] stewardship, reporting, and compliance processes. _[name]_, ORCID will act as the primary contact for communication with the SPARC DRC regarding data management and sharing.

Data will be shared on the SPARC Portal; the SPARC Data Resource Center does not enforce data sharing policies or track data sharing milestones for investigators. SPARC dataset features such as DOIs, landing pages with detailed information (i.e., number of files, subjects and overall size of the dataset), and project association (which gives the ability to group datasets from a project) are useful tools for monitoring and oversight. A “Metrics” section on the Portal shows citations, downloads, views and forks of associated protocols to aid reporting. The SPARC data curators manage and track where datasets are in the publication process; and record status and key email communications. A summary of this information can be provided to the investigator if requested.