



## Fighting Childhood Cancer,

*One Cup at a Time.*

111 Presidential Blvd, Suite 203  
Bala Cynwyd, PA 19004

AlexsLemonade.org

(866) 333-1213 or (610) 649-3034

   @AlexsLemonade

## Single-cell Pediatric Cancer Atlas Request for Contributions

### About Alex's Lemonade Stand Foundation

Alex's Lemonade Stand Foundation (ALSF) emerged from the front yard lemonade stand of 4-year-old Alexandra "Alex" Scott, who was fighting cancer and wanted to raise money to find cures for all children with cancer. Her spirit and determination inspired others to support her cause, and when she passed away at the age of 8, she had raised \$1 million. Since then, the Foundation bearing her name has evolved into a national fundraising movement. Today, ALSF is one of the leading funders of pediatric cancer research in the U.S. and Canada, raising more than \$250 million so far, funding over 1,000 research projects, and providing programs to families affected by childhood cancer. For more information, visit [AlexsLemonade.org](https://AlexsLemonade.org).

### About the Single-cell Pediatric Cancer Atlas (ScPCA)

In 2019, ALSF established the Single-cell Pediatric Cancer Atlas (ScPCA) through awards for data generation and to create an atlas of single-cell gene expression profiles of pediatric cancers of different types and from different organ sites. The Childhood Cancer Data Lab, a program of ALSF, launched the ScPCA Portal (<https://scpca.alexslimonade.org/>) in 2022 to make uniformly processed, summarized single-cell and single-nuclei RNA-seq data and de-identified metadata available for download. The ScPCA Portal also supports other data modalities, such as bulk RNA-seq, CITE-seq, and spatial transcriptomics. Today, the ScPCA Portal hosts data for nearly 500 pediatric tumor and patient-derived xenograft samples from more than 30 cancer types.

### Call for Contributions to the ScPCA

ALSF seeks to expand the projects available from the ScPCA Portal by accepting submissions from researchers with existing single-cell datasets. Submitted datasets will be made publicly available on the ScPCA Portal. Researchers that submit data may be eligible to receive a one-time small grant of unrestricted funds to be used for childhood cancer research.

We will accept submissions of 10x Genomics single-cell or single-nuclei profiling of childhood and adolescent cancer (ages 0-19) data, broadly defined to include relevant animal models, patient-derived xenografts, or cell lines, as well as tumor data. Researchers must process their data using the Childhood Cancer Data Lab's production pipeline – available at <https://github.com/AlexsLemonade/scpca-nf/> – and submit the output, along with project (see [Submitting Project Metadata](#)), sample (see [Submitting Sample Metadata](#)) and cell



## Fighting Childhood Cancer,

*One Cup at a Time.*

111 Presidential Blvd, Suite 203  
Bala Cynwyd, PA 19004

AlexsLemonade.org

(866) 333-1213 or (610) 649-3034

   @AlexsLemonade

(see [Submitting Cell Metadata](#)) metadata. Submitters must also sign a Data Transfer Agreement before data transfer ([view Data Transfer Agreement template](#)).

### Award

**Up to 10 eligible researchers will receive \$5000 in unrestricted funds.** Once an eligible single-cell dataset submission has been completed and approved by the Data Lab, researchers will be invited to complete a grant application for the unrestricted funds consisting of basic contact and payment information. Eligibility will be determined at ALSF's discretion. Funds are available on a first-come-first-served basis and distributed to the researcher's institution, which must be a non-profit institution. ALSF does not allow any funds to be used for research utilizing human embryonic stem cells or nonhuman primates. Research with human induced pluripotent stem cells is permissible. **The first step in the submission process is to [fill out this form](#) to verify whether a dataset is suitable for inclusion in the ScPCA Portal.**

Researchers are encouraged to engage with the Childhood Cancer Data Lab (Data Lab) before and during the preparation of their submissions. Interested parties can join drop-in office hours sessions with Data Lab team members on Zoom or reach out to the Data Lab directly via email (See [Contact the Data Lab](#)).

### Important Dates

- **Office Hours:** June 2, 2023, from 3-4 pm Eastern and July 17, 2023, from 3-4 pm Eastern on [Zoom](#) (no registration required)
- **Dataset Submission Deadline:** September 1, 2023
- **Application Submission Deadline** (if eligible): September 15, 2023

### Contact the Data Lab

You can contact the Data Lab via email at [scpca@ccdatalab.org](mailto:scpca@ccdatalab.org). Potential submitters are also encouraged to join the #scpca-contributions channel in Cancer Data Science Slack (<https://www.ccdatalab.org/slack>) to facilitate support from the Data Lab team while preparing their submissions.

### Overview of Steps for Submission

Submitters must take the following steps:

1. Complete an intake form to determine dataset eligibility:  
[https://share.hsforms.com/1V1loS-\\_hTMi3-\\_Lz107AcA336z0](https://share.hsforms.com/1V1loS-_hTMi3-_Lz107AcA336z0). A Data Lab team member



## Fighting Childhood Cancer,

*One Cup at a Time.*

111 Presidential Blvd, Suite 203  
Bala Cynwyd, PA 19004  
AlexsLemonade.org  
(866) 333-1213 or (610) 649-3034

   @AlexsLemonade

will contact you within three working days of your submission to notify you of your eligibility and provide additional information required for submission.

2. Execute a Data Transfer Agreement between your institution and ALSF. View the template agreement [here](#).
3. Process the dataset you will submit using the ScPCA processing pipeline, available at <https://github.com/AlexsLemonade/scpca-nf/> (see **About the Processing Pipeline** for more information).
4. Transfer the output of the processing pipeline and project and sample metadata to the Data Lab. The Data Lab will provide transfer details once a dataset has been determined to be eligible for an award.

### About the Processing Pipeline

The Data Lab's [Nextflow](#) workflow (scpca-nf) is used to process 10x Genomics single-cell and single-nuclei RNA-seq data for release on the [Single-cell Pediatric Cancer Atlas \(ScPCA\) project](#). Submitters are **required to use this pipeline** to ensure that all datasets available via the ScPCA Portal are uniformly processed.

The workflow processes FASTQ files with [alevin-fry](#) to create summarized gene expression matrices (gene-by-cell matrices stored as SingleCellExperiment [objects](#)). The alevin-fry pipeline requires less RAM and compute time than Cell Ranger from 10x Genomics while producing similar results in our testing. These matrices are filtered, and additional processing is performed to calculate quality control statistics, create reduced-dimension transformations, and create output reports. scpca-nf can also process CITE-seq, bulk RNA-seq, and spatial transcriptomics samples. For more information on the contents of the output files and the processing of all modalities, please see the [ScPCA Portal docs](#).

All dependencies for the workflow outside of the Nextflow workflow engine itself are handled automatically; setup generally requires only organizing the input files and configuring Nextflow for your computing environment. Nextflow will also handle parallelizing sample processing as your environment allows, minimizing total run time.

### Processing Your Data

**Please note that processing single-cell and single-nuclei RNA-seq samples with the pipeline as currently configured requires access to a high-performance computing (HPC) environment with nodes that can accommodate jobs requiring up to 24 GB of RAM and 12 CPUs.** The pipeline can be adapted to lower CPU counts if needed.



# Fighting Childhood Cancer,

*One Cup at a Time.*

111 Presidential Blvd, Suite 203  
Bala Cynwyd, PA 19004

AlexsLemonade.org  
(866) 333-1213 or (610) 649-3034

@AlexsLemonade

scpca-nf can be set up for your computing environment with a few configuration files. Please see [our instructions](#) to get started.

The Data Lab team is available to provide technical support, including video conferencing calls for troubleshooting purposes, as you prepare your submission (see [Contact the Data Lab](#) to get in touch).

## Submitting Project Metadata

Submitters will be required to submit the following to be posted on the ScPCA Portal:

- A project title
- A project abstract describing the dataset
- Contact name and email address
- *Optional:* Preprint or peer-reviewed publication portal users should cite if using the dataset for their research

Submitters will be asked to submit this information in a Word document.

## Submitting Sample Metadata

Sample-associated metadata must be included in submissions to ensure that datasets available from the ScPCA Portal are maximally useful to researchers. Submitters are encouraged to use standard terminology or, where possible, ontologies, such as the Experimental Factor Ontology, to describe samples.

Researchers are encouraged to submit as many metadata fields as they are able (please see [De-identification Standard for Human Samples](#) and the [Data Transfer Agreement template](#)) and to use their scientific expertise to determine what metadata would maximize the utility of the dataset for pediatric cancer research. For published studies, researchers should submit at least the minimal metadata required to reproduce and verify the study's reported findings.

### Minimal Metadata

The minimal metadata fields that should be included in submissions of different types are outlined below.

#### *All Submissions*

All submissions are required to include the following metadata fields:

Metadata	Description
----------	-------------



# Fighting Childhood Cancer,

*One Cup at a Time.*

111 Presidential Blvd, Suite 203  
Bala Cynwyd, PA 19004

AlexsLemonade.org  
(866) 333-1213 or (610) 649-3034

@AlexsLemonade

Field	
seq_unit	cell for single-cell samples or nucleus for single-nucleus samples
technology	10x Genomics kit used to process library, one of: 10Xv2, 10Xv2_5prime, 10Xv3 or 10Xv3.1
submitter_id	A research sample identifier; ideally, identifiers would allow for linkage to published results or samples in widely-used repositories for sequencing data
diagnosis	Tumor type of origin or disease being modeled
subdiagnosis	Subcategory of diagnosis or mutation status (if applicable)

## Tumor Samples

Submissions of tumor samples directly obtained from patients are required to include the following additional metadata fields:

Metadata Field	Description
age	Age at time sample was obtained
sex	Sex of patient that the sample was obtained from
tissue_location	Where in the body the tumor sample was located
disease_timing	What stage of disease was the sample obtained? At diagnosis or recurrence?

## Cell Line Samples

Submissions of human cell line samples are required to include the following additional metadata fields:

Metadata Field	Description
cell_line	Cell line name AND <a href="#">Cell Line Ontology (CLO)</a> identifier
age_at_collection	Age of the patient at the time the sample used to establish the line was obtained
sex	Sex of the patient from which the line was isolated
tissue_location	Where in the body the source of culture was derived from
disease_timing	What stage of disease was the original sample obtained? At diagnosis or recurrence?
cell_type or tissue	The type of cell (e.g., epithelial) or tissue if mixed cell types are expected



# Fighting Childhood Cancer,

*One Cup at a Time.*

111 Presidential Blvd, Suite 203  
Bala Cynwyd, PA 19004  
AlexsLemonade.org  
(866) 333-1213 or (610) 649-3034

@AlexsLemonade

passage_number	The number of times the cells were split or passaged prior to library preparation
----------------	---

Please include any perturbations (e.g., treatment with small molecule inhibitors) that were performed.

### *Mouse Model Samples*

Submissions of samples obtained from mouse models are required to include the following additional metadata fields:

Metadata Field	Description
tissue_source	The source of the tissue (e.g., tissue location or isolation methodology)
genotype	Genotype of the mouse (if applicable)
model_description	Description of the model type (e.g., syngeneic model or humanized model with specifics about engraftment)
strain	Formal name of mouse strain (i.e., “C57BL/6” not “B6”)
sex	Sex of the mouse from which the sample was isolated

Please include any perturbations (e.g., treatment with small molecule inhibitors) that were performed or any relevant phenotypic information. If a submitter has data from a non-traditional mouse model, such as cell line injections into mice, they will be required to submit additional sample data, such as injection site.

### *Patient-Derived Xenograft Samples*

Submissions of patient-derived xenograft samples are required to include the following additional metadata fields:

Metadata Field	Description
age_at_collection	Age of the patient at the time the sample used to establish the xenograft was obtained
sex	Sex of the patient from which the xenograft was isolated
tissue_location	Where in the body the source of the xenograft was derived from
disease_timing	What stage of disease was the original sample obtained? At diagnosis or recurrence?
xenograft_type	The type of xenograft (e.g., orthotopic)



## Fighting Childhood Cancer,

*One Cup at a Time.*

111 Presidential Blvd, Suite 203  
Bala Cynwyd, PA 19004

AlexsLemonade.org  
(866) 333-1213 or (610) 649-3034

   @AlexsLemonade

### De-identification Standard for Human Samples

We consider human sample-associated metadata to be de-identified when the following conditions are met:

- Elements delineated in 45 C.F.R. § 164.514(b)(2) have been fully removed (i.e., the HIPAA Safe Harbor Method for de-identification has been satisfied).
- The disclosure of any sample identifiers or accessions included in the metadata is allowable under HIPAA Privacy Rule guidelines around the disclosure of unique codes and re-identification.
- There is no reasonable basis to believe that the metadata, or the metadata in combination with the project description, is sufficient to uniquely identify individuals (e.g., a combination of values is sufficiently rare in general or at an institution documented in the project description to allow for re-identification).

### Submitting Cell Metadata

Individual cell types or labels should be submitted to ensure datasets are maximally valuable for users of the ScPCA Portal. In particular, cell types or labels assigned by submitters, e.g., as part of the underlying conclusions of a publication, should be included in submissions.

Submitters will be required to submit a table with the following information:

Field	Description
submitter_id	The research sample identifier included in your sample metadata
cell_barcode	10x Genomics cell barcode associated with each cell
cell_type_assignment	The cell type or label assigned to the cell
CLO_term	<i>(Optional)</i> The <a href="#">Cell Ontology</a> cell type term associated with the designated cell type. Although these are not required, the cell type ontology can help standardize cell type assignment across datasets available on our portal.