Manuscript Title

This manuscript (permalink) was automatically generated from AlexsLemonade/ScPCA-manuscript@fa36ed2 on January 25, 2024.

Authors

- **John Doe**
  XXXX-XXXX-XXXX-XXXX · johndoe · johndoe · @johndoe@mastodon.social
  Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Jane Roe** ✉
  XXXX-XXXX-XXXX-XXXX · janeroe
  Department of Something, University of Whatever; Department of Whatever, University of Something

✉ — Correspondence possible via GitHub Issues or email to Jane Roe <jane.roe@whatever.edu>.

Abstract

Introduction

- Intro to single-cell analysis and the need for a central repository with harmonized data
  - The amount of single-cell RNA-seq data has been rapidly growing
  - Unlike bulk, which averages the profiles of all cells within a sample, Single-cell RNA-seq allows for analysis and identification of individual cell populations that may play important roles in tumor growth, resistence, and metastasis
  - Single-cell RNA-seq of tumor samples also allows us to understand how tumor cells may interact with normal cells in the tumor microenvironment
  - With the growing number of single-cell RNA-seq datasets, efforts have emerged to create central, harmonized sources for datasets such as the Human Cell Atlas, which mainly contains normal tissue with a smaller proportion of samples derived from diseased tissue
  - Additionally, the Human Tumor Atlas Network hosts a collection of genomics data, including Single-cell RNA-seq, across multiple cancer types
  - By harmonizing data across multiple studies and diseases, researchers can better perform joint analysis, taking advantage of more samples to complete their analysis and illuminate previously unknown similarities
- Intro to ScPCA portal and how it fills a gap in the field
  - However, there previously was no collection of Single-cell RNA seq datasets specific to pediatric cancer
  - A sentence about why do we care about pediatric. We should mention something about the number of samples available from pediatric tumors being low compared to adult tumors and limited by institution, so it's even more important to make data available to all researchers.
  - To fill this unmet need, we developed and currently maintain the Single-cell Pediatric Cancer Atlas (ScPCA) Portal, an open-source data resource for single-cell and single-nuclei RNA sequencing data of pediatric tumors
- What is the ScPCA portal
  - The ScPCA Portal holds uniformly processed summarized gene expression for over 500 samples from a diverse set of over 50 types of cancers
  - Data comes from 10 projects funded by ALSF and additional community contributed datasets
  - In addition to gene expression data from single-cell and single-nuclei RNA sequencing, the Portal holds data obtained from bulk RNA sequencing, spatial transcriptomics, and feature barcoding methods, such as CITE-seq and cell hashing
  - Data provided on the portal is available in formats ready for downstream analysis, such as SingleCellExperiment or AnnData objects.
  - All samples contain normalized gene expression counts, dimensionality reduction results and cell type annotations (technically most will not all)
- Why is the ScPCA portal important
  - Data on the portal has been uniformly processed using scpca-nf, a Nextflow-based open-source pipeline developed by the Childhood Cancer Data Lab.
  - The scpca-nf workflow uses alevin-fry for fast and efficient processing of all data currently available on the portal, including single-cell RNA-seq data and any associated CITE-seq or cell hash data, spatial transcriptomics data, and bulk RNA sequencing.
  - This makes it easy to perform analysis across multiple samples and projects without having to do any re-processing
  - We also provide scpca-nf as a resource to the community to easily allow others to process their own samples for comparison to those on the Portal.
  - In addition to uniformly processed data across multiple cancer types, we provide comprehensive documentation about data processing and the contents of files on the portal,

including a guide to getting started working with an ScPCA dataset.
- The data included on the Portal will serve as a resource for all pediatric cancer researchers by providing uniformly processed data ready for immediate use to help researchers answer their important biological questions, without the need for time consuming data re-processing and data wrangling.

References