COSG

Release 1.0.0

Min Dai

Jun 16, 2021

CONTENTS

ONE

OVERVIEW

COSG is a cosine similarity-based method for more accurate and scalable marker gene identification.

- COSG is a general method for cell marker gene identification across different data modalities, e.g., scRNA-seq, scATAC-seq and spatially resolved transcriptome data.

- Marker genes or genomic regions identified by COSG are more indicative and with greater cell-type specificity.

- COSG is ultrafast for large-scale datasets, and is capable of identifying marker genes for one million cells in less than two minutes.

The method and benchmarking results are described in Dai et al., (2021).

TWO

DOCUMENTATION

The documentation for COSG is available here.

TUTORIAL

The COSG tutorial provides a quick-start guide for using COSG and demonstrates the superior performance of COSG as compared with other methods, and the Jupyter notebook is also available.

QUESTION

For questions about the code and tutorial, please contact Min Dai, daimin@zju.edu.cn.

CITATION

If COSG is useful for your research, please consider citing Dai et al., (2021).

5.1 Usage

Import COSG as:

```
import cosg as cosg
```

amd import Scanpy as:

```
import scanpy as sc
```

Next, load the data via:

```
adata = sc.datasets.pbmc68k_reduced()
```

then identify marker genes for each cell group by running:

```
cosg.cosg(adata, key_added='cosg', groupby='bulk_labels')
```

and the top marker genes can be visualized via:

```
sc.pl.rank_genes_groups(adata, key='cosg')
```

5.2 Installation

Runnig this package requires a Python environment (>=3.6).

5.2.1 Development Version

To use the latest version on GitHub: please clone the repository and *cd* into the root directory, and run:

```
pip install -e .
```

5.2.2 PyPI

Please run:

```
pip install cosg
```

5.3 API

Import cosg as:

```
import cosg as cosg
```

5.3.1 Marker gene identification

——

5.4 Release notes

5.4.1 Version 1.0

1.0.0

COSG is an accurate and efficient marker gene identification method for single cell sequencing data.

- COSG is applicable to single-cell RNA sequencing data, single-cell ATAC sequencing data and spatially resolved transcriptome data.
- COSG is fast and scalable for ultra-large datasets of million-scale cells (less than two minutes for one million cells with one CPU core).
- Marker genes or genomic regions identified by COSG are more indicative and with greater cell-type specificity.

Min Dai (2021-06-15)

PYTHON MODULE INDEX

C

C

cosg
    module,

M

module
    cosg,