

scverse

Foundational tools for single-cell omics data analysis

[GitHub](#)

[Discourse](#)

[Zulip](#)

[X](#)

[Bluesky](#)

[YouTube](#)

<https://scverse.org>

単一細胞オミクス解析に関連するPythonツールの開発・維持を目的に2022年に組織されたコンソーシアム。

AnnDataとScanpyをコア技術とする。

scVelo (細胞の進行方向を推定), CellRank (細胞の行き先と確率を推定)もコア技術を土台に開発されたEcosystem packageのうちのひとつ。

マルチモーダルデータ（scRNA-seq + scATAC-seq）の解析に対する拡張として MuData, Muon の開発、空間トランスクリプトーム解析のためのSquidpyの開発など。それぞれの相互運用性の改善やファイルフォーマットの統一など、一体として扱いやすいツール群の開発を目指していくコミュニティ。

CORE PACKAGES



anndata
Standard for annotated matrices



mudata
Multimodal data format



spatialdata
Spatial data format



scanpy
Single-cell analysis framework



muon
Multi-omics analysis framework



squidpy
Spatial single-cell analysis



scvi-tools
Single-cell machine learning framework



scirpy
Single-cell immune sequencing analysis framework

SnapATAC2
Single-cell ATAC analysis framework



rapids-singlecell
GPU-accelerated framework for scRNA analysis



pertpy
Perturbation data analysis framework



decoupler
Enrichment analysis framework

[View all scverse packages](#)

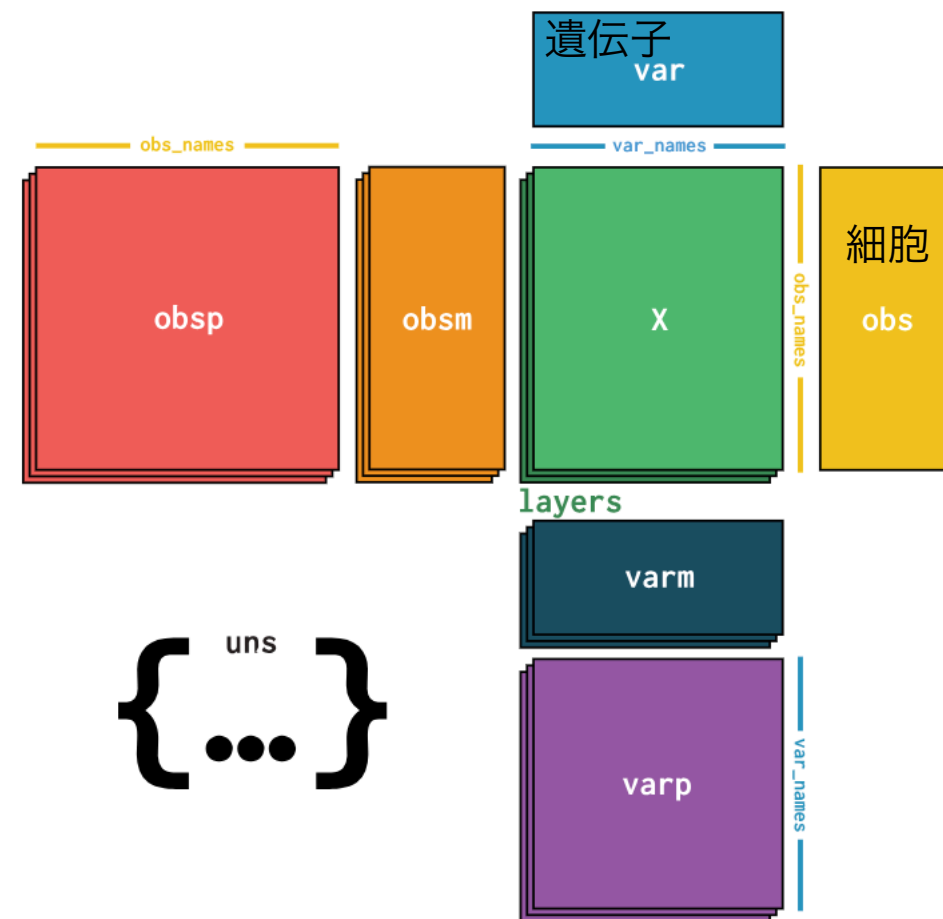
AnnData “Annotated Data”（アノテーションされたデータ）の略。

オミクスデータ格納のためにPandasのDataFrameを拡張したデータ構造。シングルセル解析のためのPythonパッケージの多くが、このオブジェクトに対する計算として実装されている。

オミクスデータは実験で測定された数値テーブルのほかに、観測値（obs）、変数（var）それぞれが多様な情報を持つ。たとえばRNA-seqの場合、観測値であるサンプルは実験条件・性別・年齢など様々なメタデータを持つ。変数である遺伝子も、遺伝子IDやシンボルだけでなく、機能力カテゴリや、DEGか否かなどのメタデータを持つ。

それぞれを個別のオブジェクトとして管理するのはとても面倒。例えば、、、数値テーブルになんらかの操作を施した結果が、観測値や変数のメタデータに即座に反映されない。なので、複数のオブジェクトをいちいち行ったり来たりしなきゃならない。テーブルに対する計算の結果わかったことを観測値のメタデータに入れて、その結果に基づいて観測値をセレクションしたから今度は数値テーブルを同じようにスライスして、、、みたいな。

そういった面倒を避けるために、すべての観測と計算結果をひとつのオブジェクトに詰め込んで管理しやすくしたのが、AnnData というオブジェクトの特徴。



AnnData

- **.X**

$n_obs \times n_vars$ の数値テーブル。numpy.ndarrayやscipyのスパースマトリックス。scRNA-seqのカウントマトリックスなど、実験の根幹となるデータ。

layers に、同じshapeの複数のマトリックスを保持しておける。たとえば全体をノーマライズしたけど元々のカウントデータも残しておきたいときは別のレイヤーに入れておく。スライスの影響はすべてのlayerに作用する。

- **.obs**

observationsの略。観測値に関するメタデータ。PandasのDataFrameなのでPandasの操作は全部実行できる。

長さは必ず n_obs `adata.n_obs = 観測数（細胞数など）`

- **.var**

variablesの略。変数（遺伝子など）に関するメタデータ。PandasのDataFrame。長さは必ず n_var

- **.obsm**

multi-dimensional annotations for obs. 複数の数値のまとまりでそれぞれの観測値を表現したいときに使う。各観測値の低次元空間座標など。次元サイズは任意。 $n_obs \times$ 次元サイズの numpy.ndarray. PCA や UMAP の結果など。

- **.varm**

multi-dimensional annotations for var. $n_var \times$ 次元サイズのnumpy.ndarray

- **.obsp**

Pairwise annotation of obs. 観測値のペアに関する情報。距離行列など。

$n_obs \times n_obs$ のnumpy.ndarray

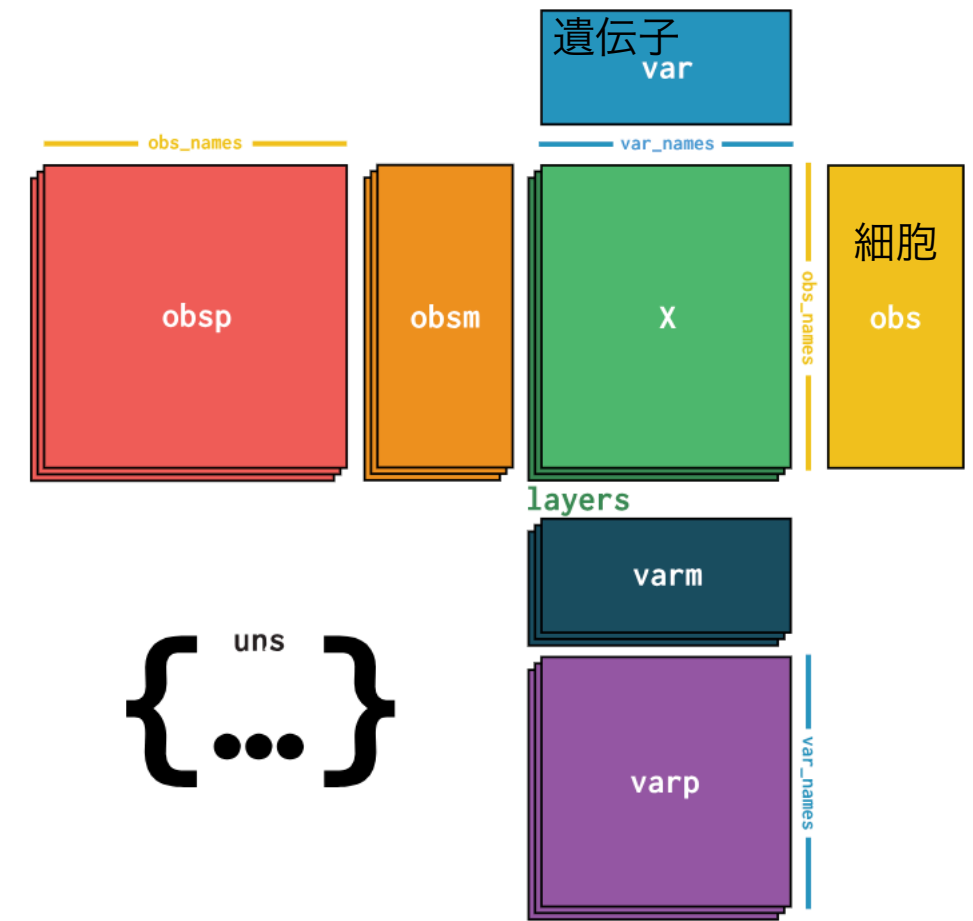
- **.varp**

Pairwise annotation of var. 変数のペアに関する情報。距離行列など。

$n_var \times n_var$ のnumpy.ndarray

- **.uns**

それ以外のデータ。とくに構造の制限はない。その他の関連データをひとまとめにしておきたいときに辞書型で放り込んでおく。クラスタの色指定とか。



Scanpy



Q Search

Installation
Tutorials
Usage Principles
How to
API
External API
Ecosystem
Release notes
Community
News
Contributing



stars 2.3k pypi v1.12 downloads 12M conda-forge v1.12 downloads 359k docs passing CI passing
discourse 6.9k posts zulip join chat powered by NumFOCUS

Scanpy – Single-Cell Analysis in Python

Scanpy is a scalable toolkit for analyzing single-cell gene expression data built jointly with [anndata](#). It includes preprocessing, visualization, clustering, trajectory inference and differential expression testing. The Python-based implementation efficiently deals with datasets of more than one million cells. For datasets too large to fit into memory, [many scanpy functions](#) are now compatible with [dask](#) (**warning: experimental**).

Discuss usage on the scverse [Discourse](#). Read the [documentation](#). If you'd like to contribute by opening an issue or creating a pull request, please take a look at our [contribution guide](#).

scanpy is part of the scverse® project ([website](#), [governance](#)) and is fiscally sponsored by [NumFOCUS](#). If you like scverse® and want to support our mission, please consider making a tax-deductible [donation](#) to help the project pay for developer time, professional services, travel, workshops, and a variety of other needs.



Pythonでシングルセル解析をする際のコアパッケージ。データの前処理や、近傍グラフ構築、t-SNEなど、標準的な解析を実行できる。

基本的に、AnnDataオブジェクトを入力して関数を実行すると、結果が同じAnnDataオブジェクトに追加されていく。新しいAnnDataを返すのではなく、**inplace**で（＝破壊的に）AnnDataが変換されていくのが特徴。

一見どこにどんな変化が生じたのかわかりにくい。観測値や変数のデータフレームにいつのまにか勝手にカラムが追加されていることがある。

Scanpyの関数

- **scanpy.pp.XXX**

前処理（**preprocessing**）に関連する関数がある。

細胞や遺伝子のフィルタリング、対数変換や、近傍グラフの構築など

- **scanpy.tl.XXX**

さまざまなツール（**tools**）のセット。

PCA, t-SNE, UMAPなどの次元削減や、Louvain/Leidenクラスタリングなど。

- **scanpy.pl.XXX**

プロット（**plotting**）用の関数。

PCA用のプロット、UMAP用のプロットなど、それぞれの可視化に適した関数が用意さ

れている。複雑な処理を書かなくても、`anndata`に含まれるメタデータから自動的に、遺伝子発現量による色のグラデーションや、クラスタごとの色分けなどをしてくれる。

注：scanpyはたいてい“sc”の短縮名で呼び出すことが多いので、以上の関数は、`sc.pp.XXX`, `sc.tl.XXX`などと呼び出す