1 Key applications of SEVtras
SEVtras stands for sEV-containing droplet identification in scRNA-seq data.

You can freely use SEVtras to explore sEV heterogeneity at single droplet, characterize cell type dynamics in light of sEV activity and unlock diagnostic potential of sEVs in concert with cells.

![Diagram illustrating SEVtras workflow and analysis](image-url)
KEY APPLICATIONS OF SEVTRAS

• recognize sEVs at single droplet resolution.
• calculate sEV secretion activity index (ESAI) for each sample at tissue.
• deconvolve sEV secretion activity into cell type level (ESAI_c).
• unlock sEV clinical potential.

1.1 Prerequisites & Installation

1.1.1 Prerequisites

SEVtras requires python $\geq 3.7$.

"numpy", "pandas", "scipy", "umap",
"statsmodels", "gseapy", "scanpy"

1.1.2 Installation

```
pip install SEVtras
```

We also suggest to use a separate conda environment for installing SEVtras.

```
conda create -y -n SEVtras_env python=3.7
source activate SEVtras_env
pip install SEVtras
```

1.2 Functions

The main functions in SEVtras are listed below:

```
SEVtras.sEV_recognizer(sample_file, out_path, input_path=None, species='Homo', 
predefine_threads=-2, get_only=False, score_t=None, search_UMI=500, alpha=0.1, dir_origin=True)
```

This function used for sEV recognizing.

• `sample_file`: the path of each sample row by row,
• **out_path**: the path for output files,

• **input_path**: if all input files in the same directory, we can use this to represent the path, default is `None`.

• **species**: the species from which the scRNA-seq sample was sequenced, default is `Homo`. For mouse samples, you can use `Mus`.

• **predefine_threads**: SEVtras uses parallel processing for acceleration, we can define how many cpu cores to use, default is all cpu cores minus two `-2`.

• **get_only**: whether to read protein information in the `adata`, default is `False`.

• **score_t**: the threshold for SEVtras score to recognize sEVs, default is `None`. If no sEVs found in the `sEVs_SEVtras.h5ad`, we can change `score_t` to a smaller threshold (str), e.g. `'10'`.

• **search_UMI**: the UMI range to search for sEVs, default is `500`, you can use `200` for stricter recognition.

• **alpha**: the parameter for identifying sEV representative genes for each sample, default is `0.01`. If you cannot detect sEVs in all samples, the parameter can be loosened to a smaller value, e.g. `0.09`.

• **dir_origin**: the path of `matrix.mtx.gz`, default is `True` assuming that the file of `matrix.mtx.gz` locates at `sample/outs/raw_feature_bc_matrix/`. If you set it as `False`, SEVtras search for `matrix.mtx.gz` in the path of each sample in the parameter `sample_file`.

```
SEVtras.ESAI_calculator(adata_ev, adata_cell, out_path, OBSsample='batch', OBScelltype='celltype', OBSev='sEV', OBSM pca='X_pca', cellN=10, Xraw=True, normalW=True, plot_cmp='SEV_builtin', save_plot_prefix='', OBSMumap='X_umap', size=10)
```

This function used for ESAI calculating.

• **adata_ev**: the path to sEV-anndata objects,

• **adata_cell**: the path to cell-anndata objects,

• **out_path**: the path for output files,

• **OBSsample**: the index represents the sample information in the `obs` of `adata`, default is `batch`,

• **OBScelltype**: the index represents the cell type information in the `obs` of `adata`, default is `celltype`,

• **OBSev**: the index represents the sEV information in the `obs` of `adata`, default is `sEV`,

• **OBSM pca**: the index represents the PCA information in the `obsm` of `adata`, default is `X_pca`,

• **cellN**: the number of neighors used for ESAI deconvolution, default is `10`,

• **Xraw**: whether or not to use the raw object in the `adata_cell`. If `adata_cell` has been filtered or normalized, please set `Xraw=True`, and `adata_cell.raw` will be used. Note: save raw `adata_cell` as `adata_cell.raw` before filtering. Default is `True`,

• **normalW**: whether or not to scale `adata_cell` in ESAI deconvolution, default is `True`,

• **plot_cmp**: the pallete used for plot different cell types in umap, default is `SEV_builtin`, you can use other pallete in matplotlib e.g. `Set2`,

• **save_plot_prefix**: the prefix name for saved files, default is `' '`,

• **OBSMumap**: the index represents the umap information in the `obsm` of `adata`, default is `X_umap`,

• **size**: the size of point in umap plot, default is `10`.

```
SEVtras.sEV_imputation(adata_sEV)
```

This function used for sEV data imputation.
SEVtras, Release 0.1

This function used for cell free droplets simulation.

```
SEVtras.cellfree_simulator(out_path, gene_exp_ev, gene_exp_cell, expect_UMI = [40, 70, 100, 130], sEV_fraction = [0.005, 0.01, 0.05, 0.10], sEV=500)
```

This function used for sEV data GO enrichment.

```
SEVtras.sEV_enrichment(adata_sEV, nBP=15)
```

1.3 Part I sEVs recognizing

Here, we used data in github tests directory as an example, and shown how SEVtras recognizing sEVs in scRNA-seq datasets. We have generated test data in h5ad format in github, and SEVtras also supports 10x_mtx and h5 data formats.

**Note1:** The input droplet-gene matrix for SEVtras should be the raw data; herein, the matrix should come from the raw_feature_bc_matrix directory in Cell Ranger outs.

**Note2:** This part requires parallel processing using the multiprocessing package. This package is currently not compatible with Jupyter on Windows 10. Running this part in Linux is preferred.

**Note3:** We don’t recommend recognizing sEVs with a single sample. Inputting more similar samples would lead to more reliable results.

We support two file input ways to run sEVs recognizing:

##1 if samples locate in one directory

```python
import SEVtras
SEVtras.sEV_recognizer(input_path='./tests', sample_file='./tests/sample_file', out_path='./outputs', species='Homo')
```

The first parameter was the path of directory that contains all samples. Because these test files exists in our tests directory, so we used ./tests. The second parameter was the name of each sample in th directory row by row. If your data format is 10x_mtx, SEVtras can automatically detect the directory of sample/outs/raw_feature_bc_matrix/ (see parameter dir_origin).

`out_path` defines the output of SEVtras that is one h5ad file, named raw_SEVtras.h5ad, with SEVtras score (‘score’) and sEV classification (‘sEV’) in the obs for all droplets, and one named sEVs_SEVtras.h5ad with only sEV-containing droplets.

And `species` represents the species from which the sample was sequenced.

If samples locate in different directories, we also supports another way to run SEVtras.

##2 if samples locate in different directories

```python
import SEVtras
SEVtras.sEV_recognizer(sample_file='./tests/sample_file', out_path='./outputs', species='Homo')
```

Here, first parameter was the absolute path of each sample row by row.

The result of sEV_recognizer can be displayed as follows:
1.4 Part II ESAI calculating

With the output of SEVtras.sEV_recognizer in Part I sEVs recognizing and cell matrix with cell type, SEVtras can track each sEV to original cell type and calculate sEV secretion activity index (ESAI).

**Note 1:** The input cell matrix should contain sample and cell type information in the obs of adata.

**Note 2:** This command can be compatible to all platform, including Jupyter on Windows.

SEVtras provides function ESAI_calculator to evaluate dynamic of cellular sEV secretion activity.

```python
import SEVtras
SEVtras.ESAI_calculator(adata_ev_path='./tests/sEV_SEVtras.h5ad', adata_cell_path='./tests/test_cell.h5ad', out_path='./outputs', Xraw=False, OBSsample='batch', OBScelltype='celltype')
```

The first two parameters represent the path to sEV- and cell- anndata objects.

The third parameter specifies the path of the ESAI_calculator outputs. The outputs include:

- an adata file combining both adata_ev and adata_cell, named SEVtras_combined.h5ad;
- two csv files calculating the sEV secretion activity index at the sample level (ESAI) and cell type level (ESAI_c), named ESAI_sample.csv and ESAI_celltype.csv;
- one pdf file embedding sEVs and cells in a umap , named SEVumap.pdf;
- and two pdf files plotting the ESAI_c in a umap, named ESAIumap.pdf and ESAIumap_sample.pdf.

The fourth parameter means whether to use the raw object in the adata_cell or not. If adata_cell has been filtered or normalized, please set Xraw=True, and adata_cell.raw will be used (Note: save raw adata_cell as adata_cell.raw before filtering).

The last two parameters define which index represents the sample and cell type information in the obs of adata. By default, SEVtras uses the index of batch and celltype in the obs of adata_cell. We can change the index with the parameters and OBSsample and OBScelltype.

The original cell type for each droplet listed in the obsm of SEVtras.sEVs.h5ad indexed as source.

The result of SEVumap.pdf and ESAIumap.pdf is similar to the following:
1.4. Part II ESAI calculating