

Input data requirements for SPT v0.10.6

May 8, 2022

1. Channel information file
2. HALO-exported cell table files
3. Phenotype definition table file
4. Outcomes file
5. Files manifest file

1 Channel information file

A CSV file ([example](#)) with at least the following fields:

Name

Column header fragment prefix

2 HALO-exported cell table files

A set of HALO-exported CSV files ([example](#)), each with at least the following fields:

Image Location

XMin

XMax

YMin

YMax

Cell Area

Classifier Label

In addition, for each channel, a column with the following form (`<prefix>` refers to (1)):

`<prefix> Positive`

In addition, if intensity information is provided, SPT will make use of it in some situations when requested. In this case each CSV file should contain the following fields for each channel:

`<prefix> Intensity`

3 Phenotype definition table file

A CSV file ([example](#)), with at least the following fields:

Name

Positive markers

Negative markers

The positive and negative marker values should be semi-colon delimited strings, with each entry one of the `Name` values from part (1).

4 Outcomes file

A TSV file ([example](#)), containing at least 2 columns:

Sample ID

anything

The `Sample ID` values should correspond to one of the `Sample ID` values in part (5).

5 Files manifest file

A TSV file listing all of the above files ([example](#)), containing at least the following fields:

File ID An identifier for the file, unique within this dataset. Certain values are required: `Elementary phenotypes file`, `Complex phenotypes file`, for the corresponding files (1) and (3).

Project ID A tag for this dataset or project, for downstream disambiguation.

File name The actual name to use to find this file in the dataset (without path information).

Sample ID The identifier for the sample associated with this data file, in case the data file is so associated (e.g. one of the HALO-exported cell CSVs).

Data type A tag for this file data type. The only required values are `HALO software cell manifest` and `Outcome`, for (2) and (4) respectively.