

Classification

- [Manual classification in Partek Flow](#)
- [Automatic classification in Partek Flow](#)
- [Garnett classifiers in Partek Flow](#)
- [Classify cell type task in Partek Flow](#)
- [Train classifier in Partek Flow](#)
- [Train classifier task report in Partek Flow](#)
- [References](#)

Classification in Partek Flow can be performed manually or with automatic cell classification.

Manual classification in Partek Flow

For more help related to manual classification please see [Classifying cells](#) or watch this related help video: [Manual classification](#). Partek flow provides hosted lists from the literature for many cell types which can be added using [List management](#).

Automatic classification in Partek Flow

The Garnett¹ automated cell type classification algorithm has been wrapped into Partek Flow as the **Classification** task. As with the original Garnett tool, **Classification** in Partek Flow works on single-cell data, along with a cell type definition (marker) file, and trains a regression-based classifier. Once a classifier is obtained and published, it can be applied to classify future datasets from similar tissues. To improve the user experience, both the marker file (.txt) and the classifier file (.rds) have been implemented as library files in Partek Flow.

Garnett classifiers in Partek Flow

Partek Flow hosts a selection of pre-trained classifiers as **Managed classifiers**. The list of available classifiers can be found [here](#)². If a managed classifier exists for your data type, we recommend that you try it. In addition to the **Managed classifiers**, the **Project classifiers** trained on the same dataset prior to your classification, may also be used to classify cell type. **Project classifiers** can be promoted to **Managed classifiers** if users publish them.

Classify cell type task in Partek Flow

To classify cell type using a pre-trained classifier:

Select any non-normalized single cell data node. *Filtered counts*, which contains the raw count, is used here. Next, choose the **Classify cell type** task in the **Classification** section (Figure 1)

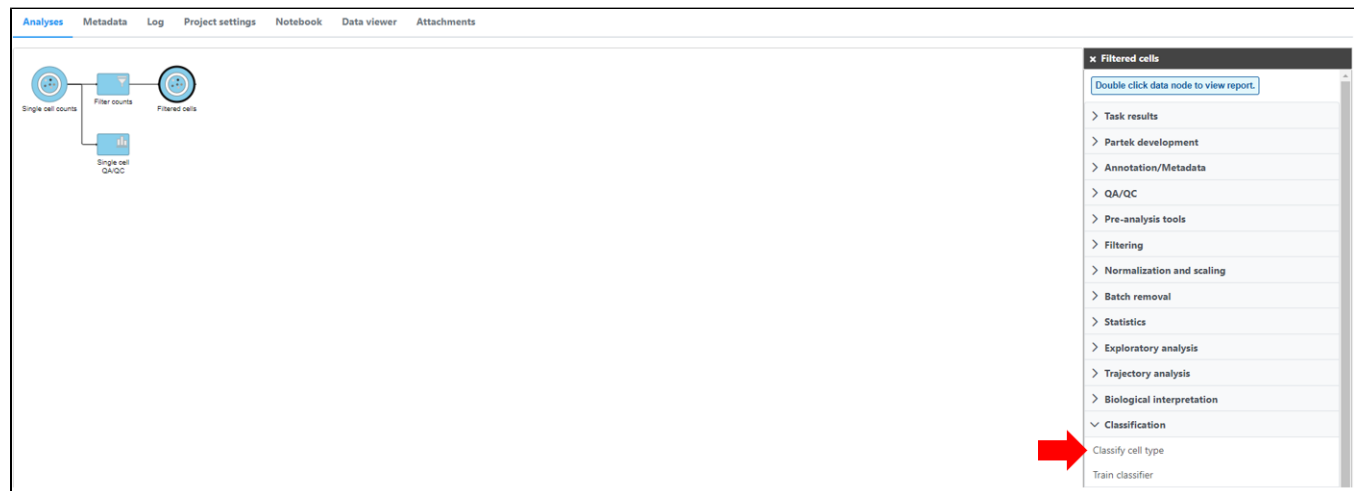


Figure 11. Selecting the Classify cell type task.

If using the **Managed classifiers** tool for the first time in Partek Flow, you will be asked to create a new classifier file (Figure 2a). Users may select either the **Download Garnett classifier** that matches the species and tissue type with their dataset from a Partek Flow maintained list or the **Import Garnett classifier** that is trained out of Partek Flow (Figure 2b). Next, push the **Create** button to create the classifier file. Once the correct classifier file has been created, select **Finish** to start running the task (Figure 2d).

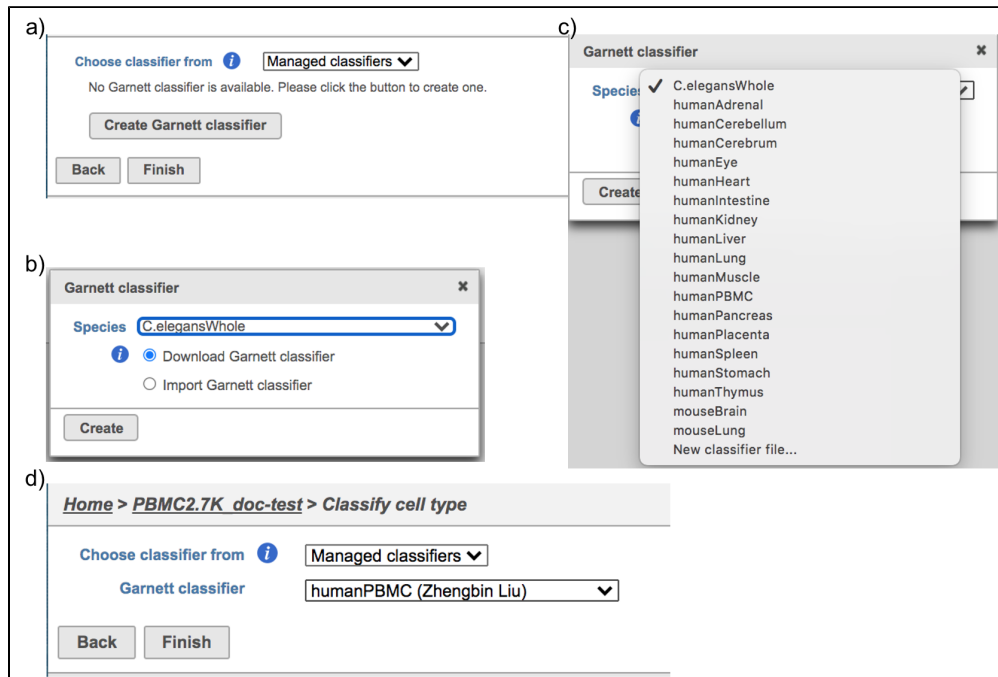


Figure 12. Create Garnett classifier from Managed classifiers and run Classify cell type task with the selected classifier.

If users would like to classify cell type with the classifier obtained from the same project, **'Project classifiers'** needs to be selected from the dropdown list. Next, use **'Select data node'** to choose the classifier before clicking the **Finish** button (Figure 3).

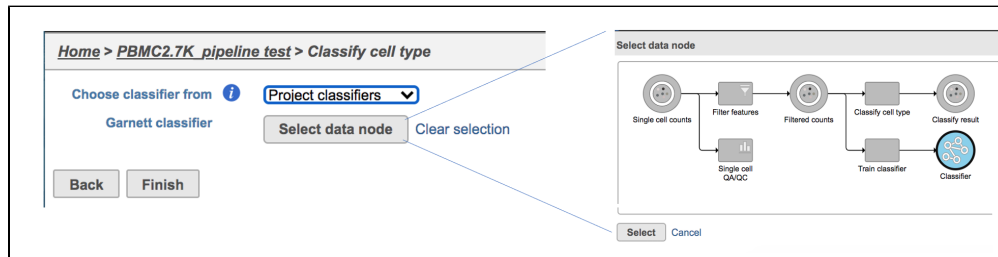


Figure 13. Run Classify cell type task with Project classifier.

Regardless of which type of classifier was applied, Partek Flow will output a new data node named **Classify result** (Figure 4). These outputs of cell type annotation are exactly the same as Garnett¹. "cell_type" is the cell type assignments directly from Garnett model. While "cluster_ext_type" is the cell type that's determined by expanding cell type assignments to nearby cells using Louvain clustering. Downstream analysis tasks such as normalization and PCA can be performed on the **Classify result** data node.

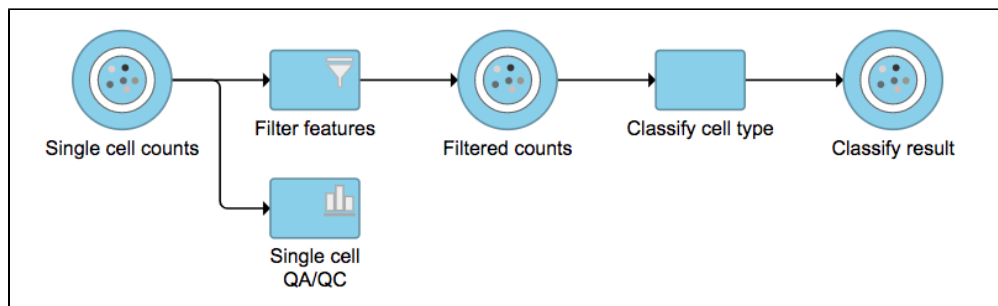


Figure 14. Classify cell type task outputs a new data node.

Train classifier in Partek Flow

To train a classifier with a list of biomarkers with your own dataset:

Select any non-normalized single cell data node, *Filtered counts* is used here. Next, choose the **Train classifier** task in the **Classification** section (Figure 1)

Similar to previously described steps, first time users will be asked to create the **Marker file**. Partek Flow does not currently host any marker files, however, users may add them as library files. Marker files should be a .txt file with the marker information in correct format. The same example in the Garnett tutorial of a simple valid **Marker file** is provided [here](#) (Figure 5).

```
>B cells
expressed: CD19, MS4A1

>T cells
expressed: CD3D
```

Figure 15. Marker file format example

Note:

- 1) Cell type name line starts with ">", followed by the name of the cell type, cell type name can NOT include hyphen (-).
- 2) Definition line starts with a keyword e.g. "expressed", "not expressed" eg. it is case sensitive. It followed by a ":" and space. Each gene name is followed by a comma and space, gene name is case sensitive. The line cannot be ended with comma or space.

For more details on how to construct a **Marker file**, please refer to Garnett tutorial³. Next, click the **Create** button and Partek Flow will then save the file with the name that users provide for future use (Figure 6a).

After the proper **Marker file** has been chosen, click the **Finish** button to start running the task as default (Figure 6b).

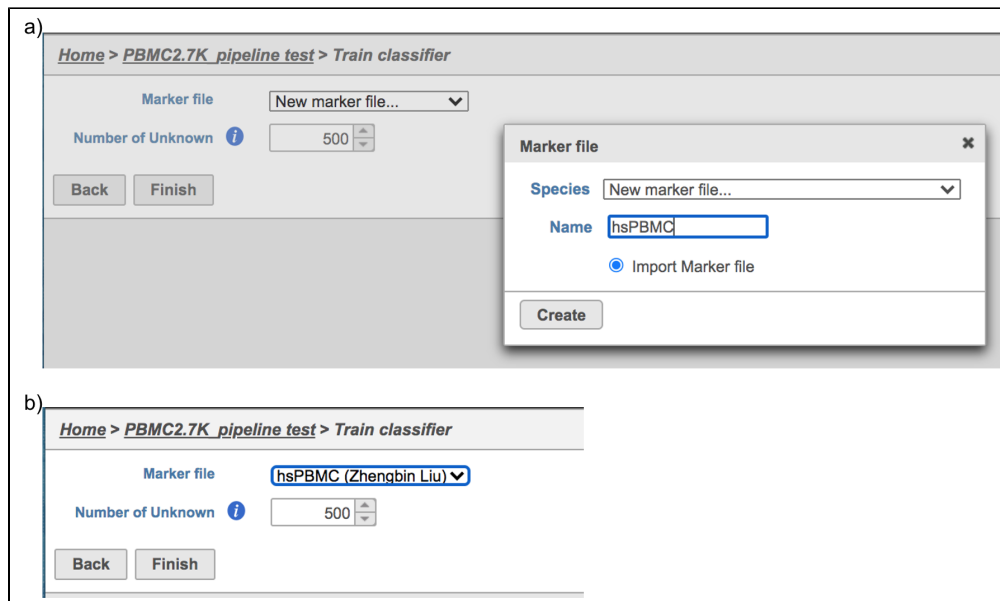


Figure 17. Train Classifier task in Flow.

Train classifier task report in Partek Flow

Once the task has finished, click the **Classifier** data node and choose the **Task report** in the **Task results** section, or prompt it to be one of the **Managed classifiers** by clicking the **Publish classifier** task in the **Classification** section (Figure 7).

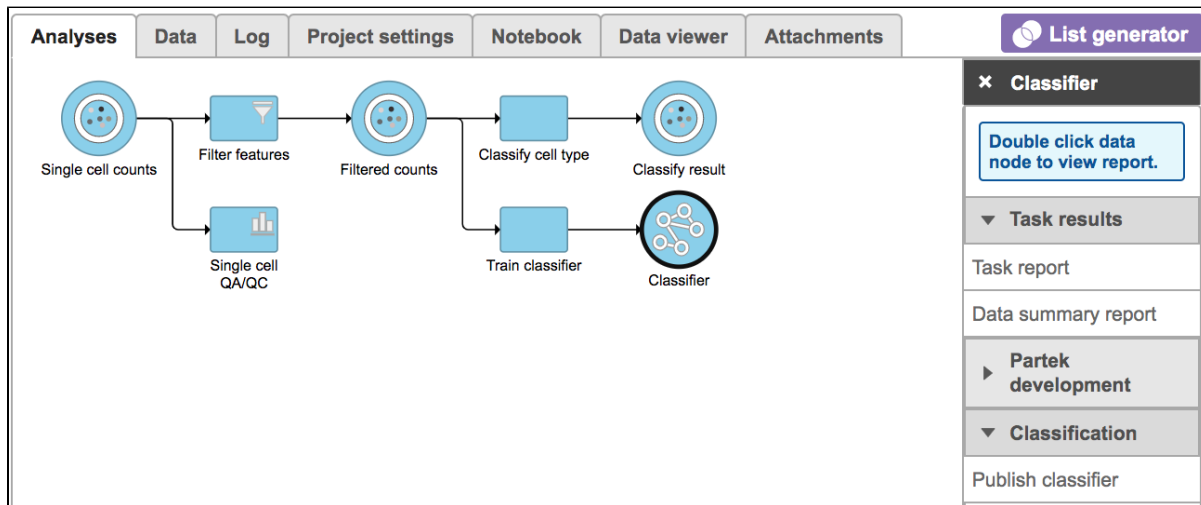


Figure 18. Train classifier task report in Task results section.

Users will find two parts in the task report: the marker evaluation plot (Figure 8) and the classification gene table (Figure 9). The marker evaluation plot provides some key information about whether the chosen markers are optimal. Ambiguity scores are calculated for each of the markers which indicates how many cells receive ambiguous labels when this marker is included. The classification gene table may give a hint to which genes are chosen as the relevant genes for distinguishing between different cell types.

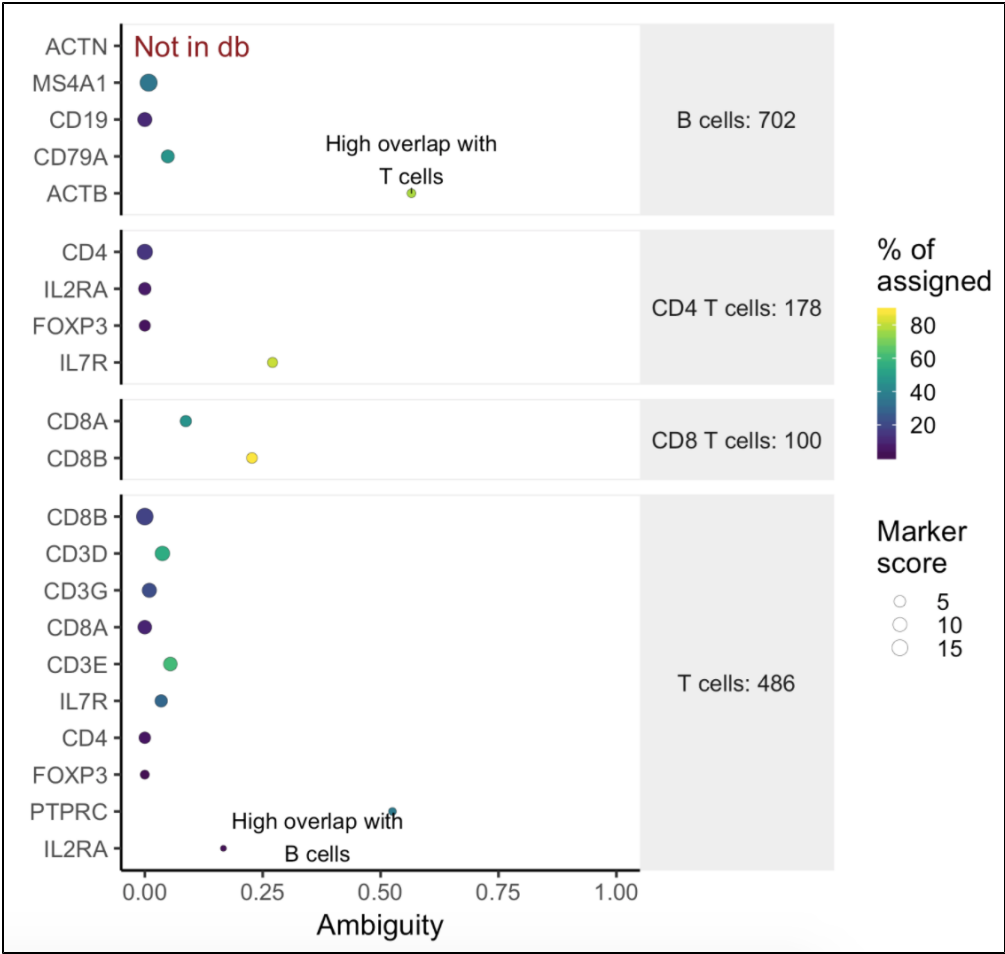


Figure 19. Example of biomarker evaluation plot in task report.

Feature ↕	B cells ↕	Dendritic cells ↕	Monocytes ↕	NK cells ↕	T cells ↕	Unknown ↕
(Intercept)	-0.40	1.06	0.67	-0.91	0.58	-0.99
CD3D	-0.11	-0.51	-0.14	-0.18	0.60	0.34
CD3E	-0.14	-0.30	0.02	-0.10	0.52	5.01E-3
CD3G	0.07	-0.39	-0.03	-0.47	0.69	0.12
CD19	2.98	-1.47	-1.73	-0.18	-2.37	2.78
CD79A	1.37	-1.96	-0.28	-0.01	-0.24	1.13
MS4A1	3.16	-0.96	-2.75	-0.46	-1.58	2.59
BATF3	-1.36	7.86	-4.64	-2.06	-2.88	3.08
CD1C	-3.03	9.41	-5.32	-1.00	-2.39	2.33
CD14	-0.62	-1.37	2.91	-0.09	-2.26	1.42
CD68	-0.76	-0.77	3.16	-1.88	-1.52	1.77
FCGR3A	-0.83	-0.33	-1.46	2.34	-1.40	1.68
FCGR1A	-0.75	-1.31	3.77	-0.44	-2.20	0.93
S100A12	-0.52	-1.16	2.50	-0.14	-1.01	0.32
NCAM1	-0.41	-0.96	-2.26	3.16	-1.14	1.61

Rows per page25(1 of 1)Download

Figure 20. Example of classification gene table in task report.

Other adjustable parameters in this task include:

Number of Unknown: this tells Garnett how many outgroup cells it should compare against; the default is 500. For a dataset with fewer cells, the number should be smaller.

References

1. <https://www.nature.com/articles/s41592-019-0535-3>
2. <https://cole-trapnell-lab.github.io/garnett/classifiers/>
3. <https://cole-trapnell-lab.github.io/garnett/docs/#constructing-a-marker-file>

Additional Assistance

If you need additional assistance, please visit [our support page](#) to submit a help ticket or find phone numbers for regional support.