Discovery Tool (Unsupervised)

Intend

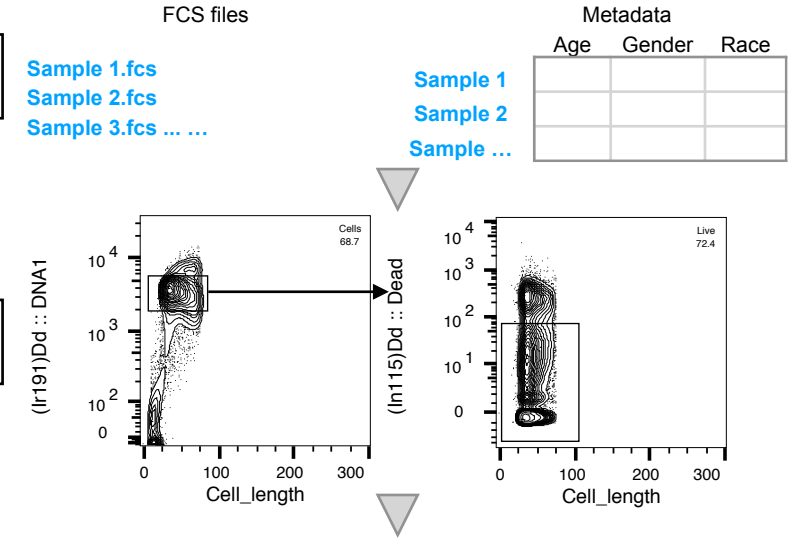
I want to compare results from open access databases (e.g. ImmPort) with dissimilar antibodies panel to the EPIC Atlas.

Pre-processing

Download fcs files and metadata file from databases

Gate out live cell events with the DNA and live-dead stain.

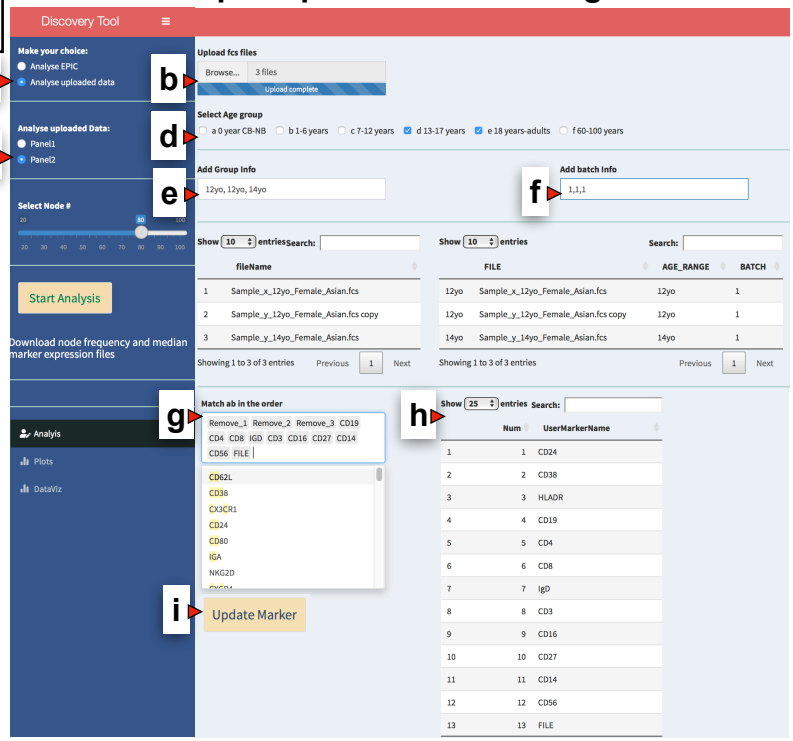
Export as fcs files (Remove markers not used in EPIC and downsample*)



Uploading

Uploading, normalisation and re-clustering with EPIC dataset

https://epicimmuneatlas.org/



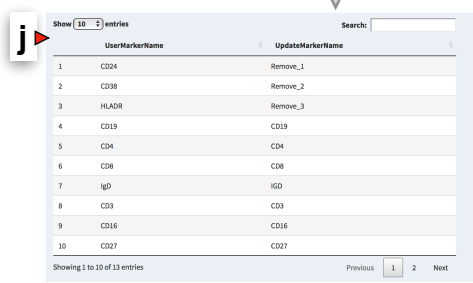
1. Select <Analyse uploaded data> **(a)**. Browse and select fcs files for upload **(b)**.

2. Select the most appropriate EPIC panel **(c)** and age range **(d)** for comparison.

3. Categorise all the uploaded fcs files sequentially using comma separated alphanumerics in the order display below “fileName” into meaningful groups **(e)**.

4. Enter the batch information using comma separated alphanumerics in the order display below “fileName” for batch correction **(f)**.

5. Markers found in the selected EPIC panel **(g)** to be ordered and matched to the markers in the uploaded data **(h)**. Once completed, clink <Update Marker> **(i)**.



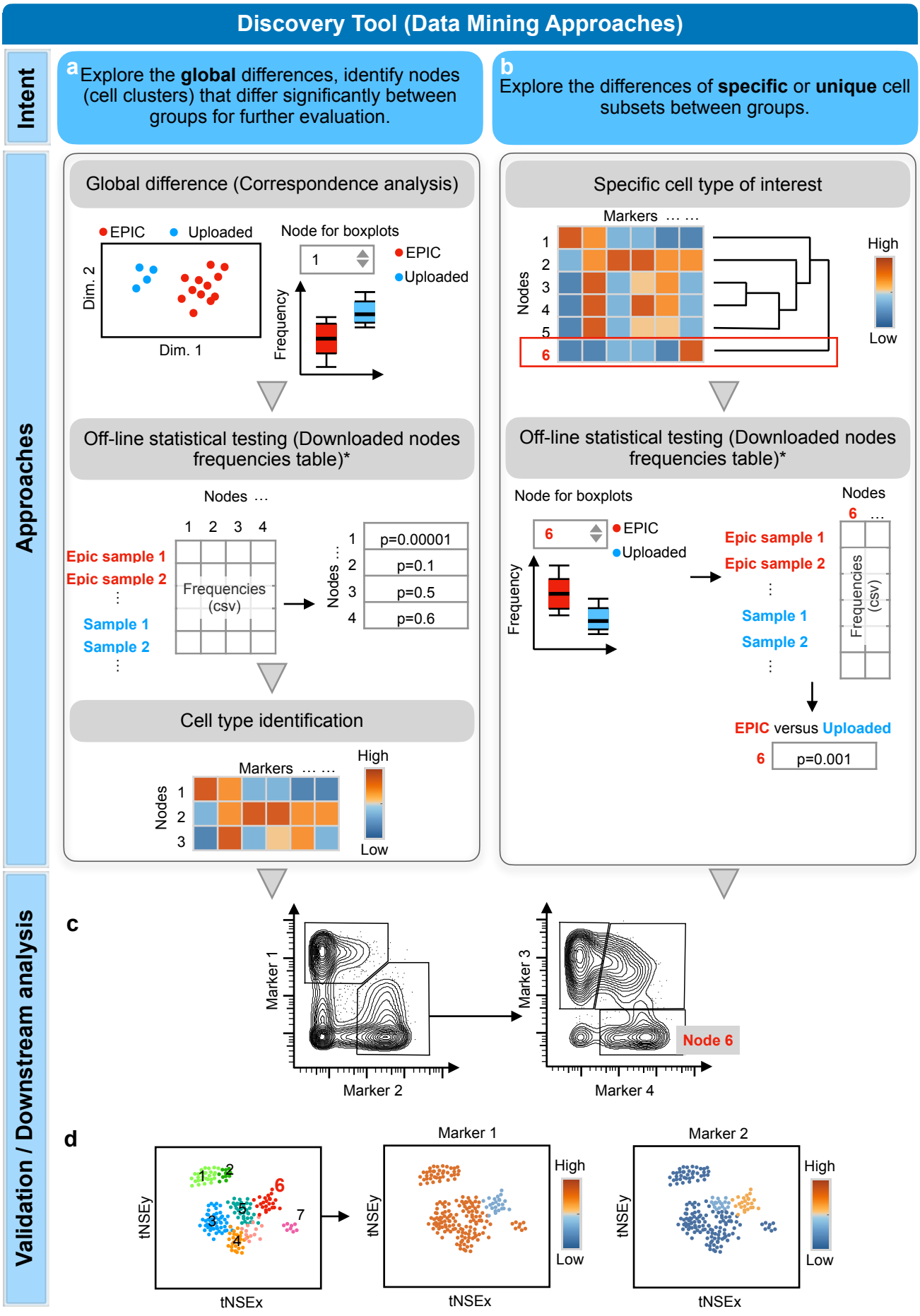
6. After the markers are updated, the matching of markers in the uploaded data set with EPIC panel can be checked **(j)**.

7. Click <Start Analysis> after selecting the desired number of nodes (clusters).

*

- Each fcs file must be <50 MB (size restriction).
- Adding important meta-data as suffixes in the filenames will enhance the user experience as these information will appear with mouse hover over the plotted graphs.

Downloading, pre-processing of flow cytometry standard (fcs) files for uploading to EPIC for comparison. Publicly available mass cytometry datasets can be obtained from data repositories such as ImmPort (URL: **https://www.immport.org/shared/home**). The meta data is required to identify useful demographics and clinical parameters (healthy or diseased) associated with each fcs file for meaningful comparison. Users can also choose to upload their own data set for comparison.



*

- Using the nodes (cell clusters) frequency table downloaded from the Discovery tool, statistical testing between two or more groups can be done to identify differences.
- Frequencies of different nodes can be merged if they are deemed to the phenotypically similar.

Data mining approaches. **(a)** The uploaded data can be compared globally with EPIC. In correspondence analysis, the cluster frequencies (Percentage of CD45+ PBMCs) of each subject is compared collectively across all samples and represented by a point plotted in two dimensions (Dim. 1 and Dim. 2) where their proximities denote greater similarity in immune cells composition. The downloadable cell frequencies file (csv, comma separate values) can be used for off-line statistical analysis and plotting. The identity of these significantly different cell subsets can then be determined based on their marker expression profiles. **(b)** A targetted analysis can be done if users are interested in specific cell subsets. The frequencies of these subsets can compared across the different groups using the downloaded cell frequencies data. **(c)** The node frequency from the unsupervised analysis can be validated with supervised bivariate gating. Schematically represented by node 6 where its unique phenotype (greatest distance in dendrogram) from other nodes single it out for further statistical evaluation and validation, from **b** to **c**. **(d)** The Discovery tool enables the downloading of single cell markers expression data embedded with the FlowSOM nodal information as a comma separate values file (csv), amenable for further off-line downstream analysis by users. Using node 6 as an example of the flow of information from **b,** the clusters obtained with FlowSOM can be visualized in a t-SNE 2 dimensional plots along with the markers expression.