CIBERSORTx Tutorials

To learn how to use CIBERSORTx, you can go through the tutorials below. You may use your own data as you go through the steps, or you can use the **Example Datasets** provided in the download links for each step, or available at the **download page** (download.php). If you use the **Example Datasets**, you should be able to reproduce the same results as if you run the Example Analysis Mode of each CIBERSORTx module. To run the Example Analysis Mode, you select "Example" under "Select Analysis Mode", as shown in the screenshot below, where the «Group Level GEPs - Follicular Lymphoma (Fig 3b-f)» example is selected for the Cell Expression Analysis Module.

More Information...

For more detailed information about how to apply CIBERSORTx, please refer to the book chapter by Steen et al. (Methods in Molecular Biology, 2020).

Tutorial 1 - Build a Signature Matrix File from Single-Cell RNA Sequencing Data

Follow this tutorial to learn how to **build and apply a custom signature matrix file from single-cell RNA sequencing data** for a given tissue of interest The resulting signature matrix file consists of barcode genes that can discriminate each cell subset of interest in that tissue type, and can subsequently be used to impute cell fractions and cell expression profiles from bulk tissue transcriptomes. Here, we demonstrate the former using the CIBERSORTx Cell Fractions Analysis Module. You can follow this tutorial using your own dataset, or the Example Dataset. This tutorial is based on the Example Dataset **«Single Cell RNA-Seq Melanoma (Supp. Fig. 2b-d)»**.

More...

Tutorial 2 - Impute Cell Fractions

Follow this tutorial to learn how to run the CIBERSORTx, Cell Fraction analysis module. Here you will learn how to **enumerate the proportions of different cell population in bulk tissue samples profiled by RNA-Seq**. This tutorial illustrates the use of the signature matrix "LM22" to deconvolve immune cell subsets from bulk RNA-Seq data of formalin-fixed paraffin embedded (FFPE) melanoma tumor biopsies. As LM22 was derived from microarray data, and the mixture file was derived from RNA-Seq profiles of FFPE specimens, this tutorial shows how one can **apply batch-correction to minimize the impact of cross-platform variation** on the results (related to Supplementary Figure 1l in Newman et al., submitted). You can follow this tutorial using your own dataset, or the Example Dataset. This tutorial is based on the Example dataset **«Melanoma (Van Allen et al., Supp. Fig. 1l)»**.

More...

1. **Prepare your signature matrix file.** Prepare your signature file with the Create Signature Matrix module, **or download the example signature matrix file used in this tutorial**.

   This tutorial uses the LM22 signature matrix. LM22 is a validated leukocyte gene signature matrix that contains 547 genes distinguishing 22 human hematopoietic cell phenotypes, including seven T-cell types, naïve and memory B cells, plasma cells, natural killer (NK) cells and myeloid subsets. For further details, please refer to Newman et al., Nature Methods, 2015 (https://www.nature.com/articles/nmeth.3337).

   If you haven't yet created a signature matrix file, please refer to the following tutorials for instructions:
   - Tutorial 1 - "Build a Signature Matrix File from Single-Cell RNA Sequencing Data"
   - Tutorial 6 - "Build a Signature Matrix File from Sorted Cell Populations (RNA-Seq data)"
   - Tutorial 7 - "Build a Signature Matrix File from Sorted Cell Populations (microarray)"

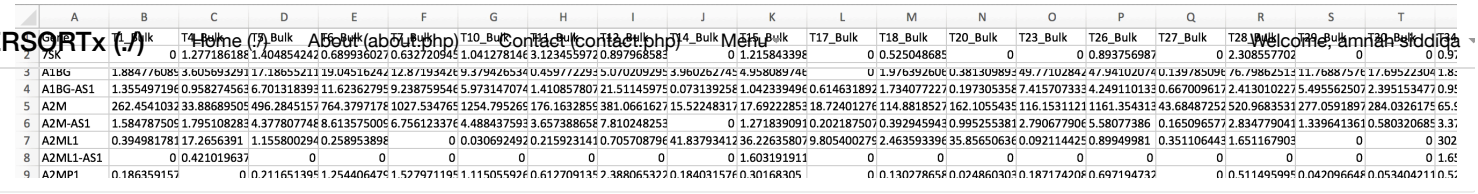
2. **Prepare your mixture file.** Prepare your mixture file according to the formatting requirements listed below under "Details". If you wish to use the example mixture file for this tutorial, it will be available in the dropdown menu when you configure your CIBERSORTx job in step 4.

   Details....

   Mixture file format:

   1. Tab-delimited tabular input format (.txt) with no double quotations and no missing entries.
   2. Genes in column 1; Mixture labels (sample names) in row 1

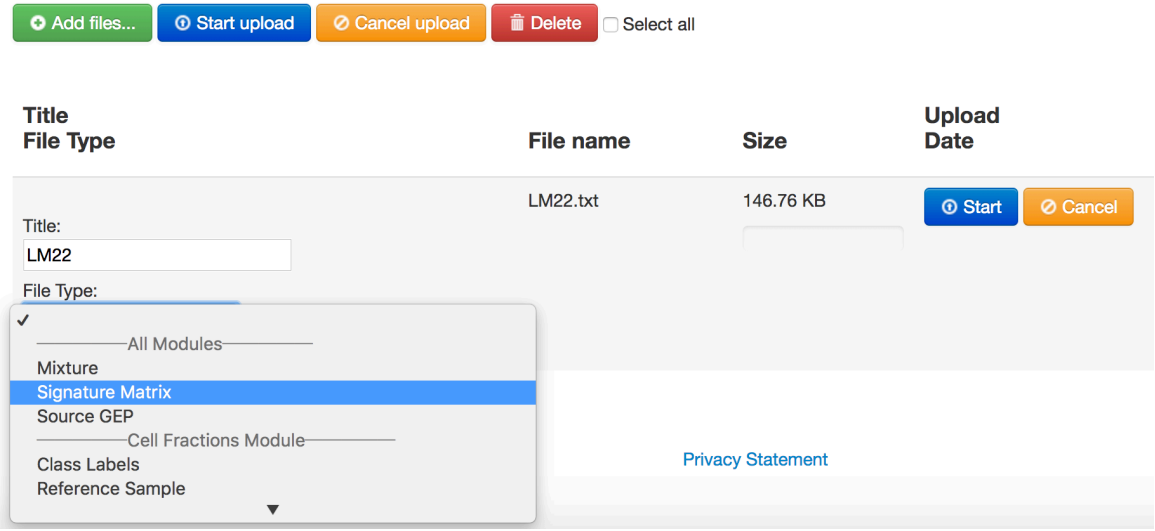
3. Given the significant difference between counts (e.g., CPM) and gene length-normalized expression data (e.g., TPM) we recommend that the signature matrix and mixture files be represented in the same normalization space whenever possible.
4. Data should be in non-log space. Note: if maximum expression value is less than 50; CIBERSORTx will assume that data are in log space, and will anti-log all expression values by 2^x.
5. CIBERSORTx will add an unique identifier to each redundant gene symbol, however we recommend that users remove redundancy prior to file upload.
6. CIBERSORTx performs a feature selection and therefore typically does not use all genes in the signature matrix. It is generally ok if some genes are missing from the user's mixture file. If less than 50% of signature matrix genes overlap, CIBERSORTx will issue a warning.



3. **Upload your file.** Go to Upload Files (upload.php).

   At the Upload Files (upload.php) page, upload your signature matrix file, by clicking on "Add files..." and selecting your file. If you are using the Example Dataset, and LM22 as signature matrix, you do not need to upload LM22 at this step. For illustration purposes, the upload form should appear like shown below if you were to upload LM22.

   Details and screenshots....



   Although optional, specify a title for your file. This name will appear in the CIBERSORTx configuration form, but if not specified, the actual file name will appear instead.

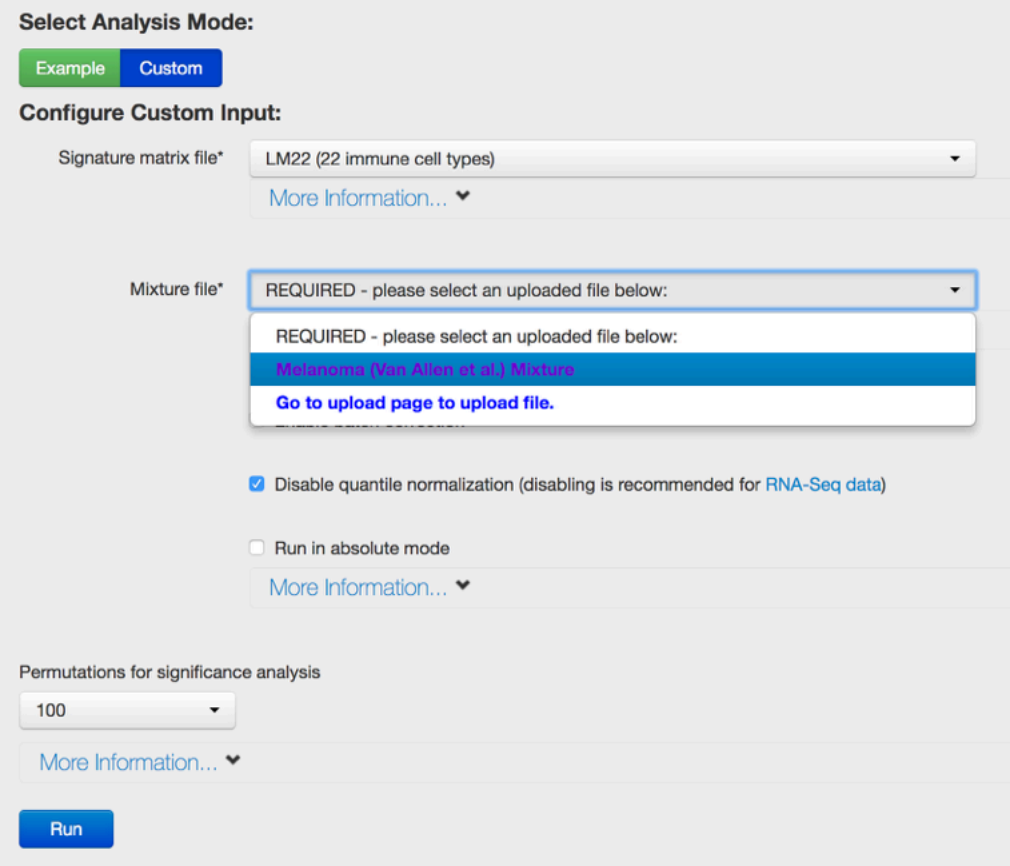
   Be sure to specify the appropriate file type in the dropdown menu, which in this case is "Signature Matrix". If no file type is specified, the file will not appear in the form.

   Click "Start" and your file should upload to the site.

4. **Configure your job.** Go to Run CIBERSORTx (runcibersortx.php). Click on "Impute Cell Fractions", and then "Custom".

   Select your Signature Matrix and Mixture files using the respective dropdown menus, as shown below under "Details". If you are following this tutorial using the Example Dataset, select "LM22" as your signature matrix, and "Melanoma (Van Allen et al.) Mixture" as your mixture file from the dropdown menu.

   Details and screenshots....



If a file is missing, it is likely due to an unsuccessful upload or omission of specifying a file type - to fix this, go back to the previous step and re-upload your files, taking care to specify the appropriate file type (i.e. "Signature Matrix") as the file type.

You can set the following parameters for your job:

- **Batch correction:** Check this box to perform batch correction, which is highly recommended when the signature matrix and mixture samples are profiled on different platforms.

  The signature matrix in this tutorial was generated using micorarray data, and the mixture file is from bulk RNA-Seq FFPE data. If your input datasets are from different platform, or if you are using the example dataset, you want to correct for the cross-platform variation. You should therefore check the box.

  When you check the box, a menu for choosing batch correction mode appears. You have the option to choose between two modes for batch correction: B-mode and S-mode. You can use the following **decision tree** to decide which batch correction mode to use.

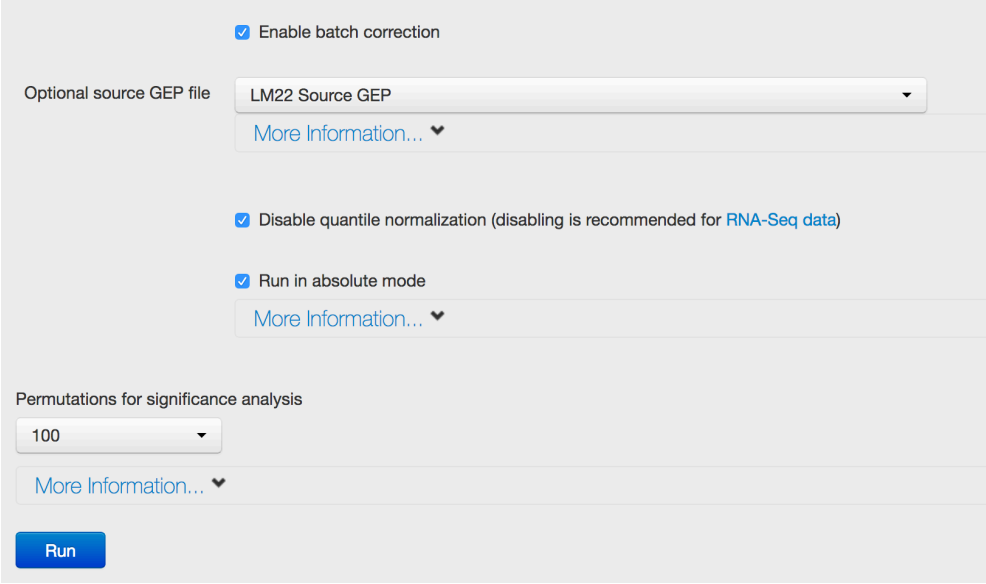
  B-mode (bulk mode) batch correction removes technical differences between a signature matrix derived from bulk sorted reference profiles (e.g., bulk RNA-Seq or microarrays) and an input set of mixture samples. The technique can also be applied to signature matrices derived from scRNA-Seq platforms, provided that transcripts are measured analogously to bulk mixture expression profiles (e.g., full-length transcripts without UMIs profiled by SMART-Seq2).

  S-mode (single cell mode) batch correction is tailored for single cell-derived signature matrices generated from droplet-based or UMI-based platforms, including 10x Chromium.

  In this tutorial, we will use B-mode batch correction, which is selected by default. With B-mode, you have the option to upload a source GEP file.

  It is optional to upload a "source GEP" file. In order to get the best results when doing batch correction, one can provide a source GEP file consisting of the transcriptome of each the cell type represented in the signature matrix. Such a file is created automatically from the reference sample file when you build the signature matrix using CIBERSORTx. By default, if no file is provided, CIBERSORTx will use the signature matrix file as the source GEP for batch correction. For this tutorial, we will use the corresponding source GEP file for LM22. The file should appear in the dropdown menu. Otherwise, you can download it here.
- **Quantile normalization:** Uncheck this box to enable quantile normalization. Disabling quantile normalization is recommended for RNA-Seq data. As we are working with RNA-Seq data in this tutorial, we leave this box checked.
- **Permutations:** Set permutations for statistical analysis, to get a deconvolution p-value (100 permutations is default). For statistically robust analyzes, we recommend setting the number of permutations to at least 100. But for the sake of efficiency, we set the number of permutations to "none" in this tutorial.

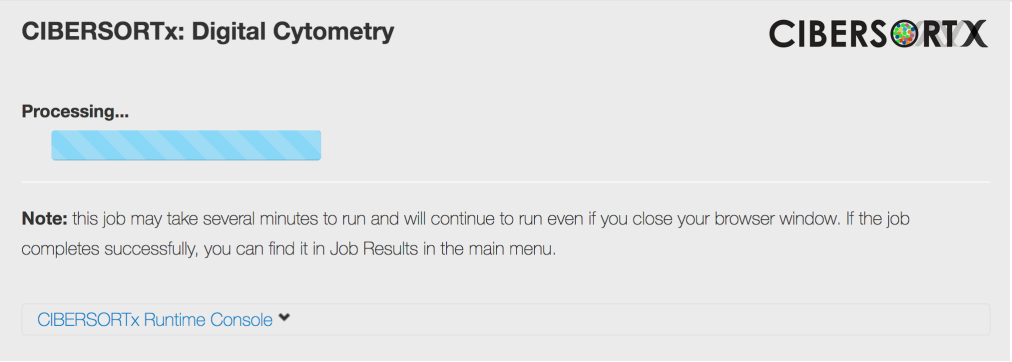
- **Absolute mode** - In contrast to the default mode (relative), which normalizes all cell fractions of the cell types in the signature matrix to 100%, absolute mode scales relative cellular fractions into a score that reflects the absolute proportion of each cell type in a mixture. If you are running the Example Dataset, check this box to run CIBERSORTx in absolute mode, and to reproduce the results presented in the Supplementary Fig. 1l of the manuscript Newman et al., submitted.



5. **Run your CIBERSORTx job.** Click the "Run" button to begin job execution.
The run will usually take a few seconds to a few minutes, based on the size of your mixture file and the number of permutations selected.

   While your job is running, the page will display a processing bar, like the one shown below.

   Details and screenshots....

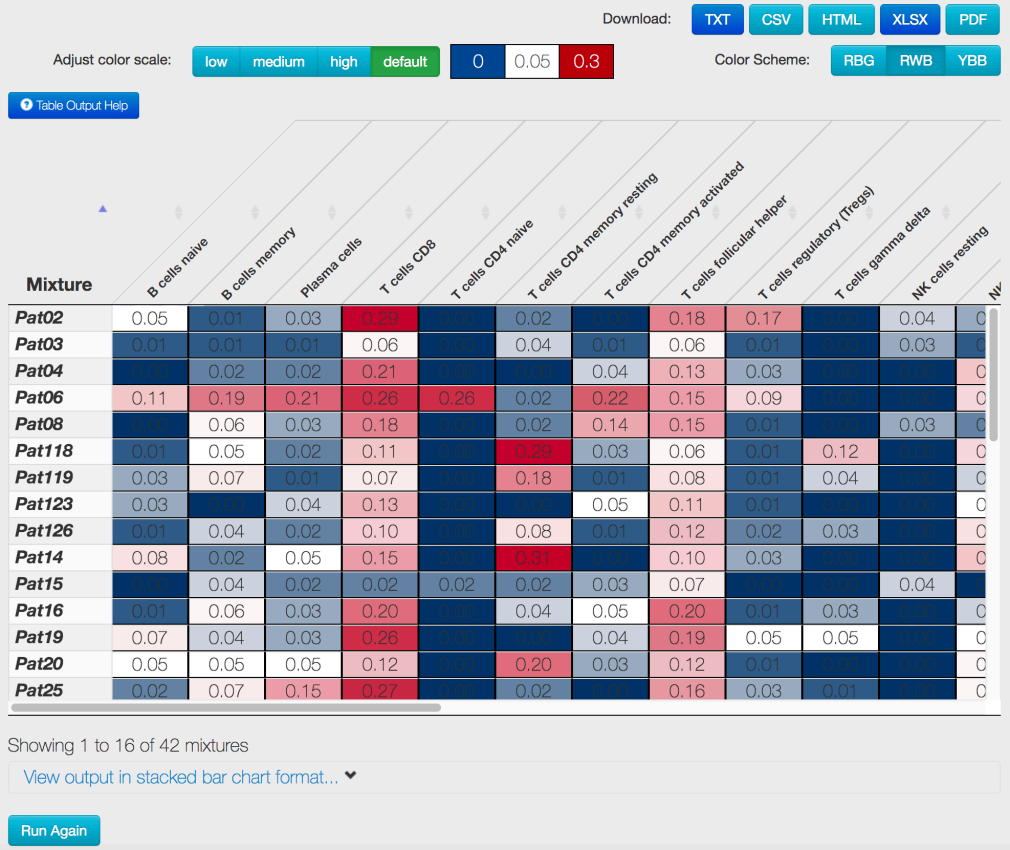


   When your job completes, the page will automatically update to display the results.

6. **Review results.** Review the results of your CIBERSORT job, which is presented in a heatmap table. The table supports dynamic search, filtering and contrast adjustment. Results can be exported in a number of file formats.

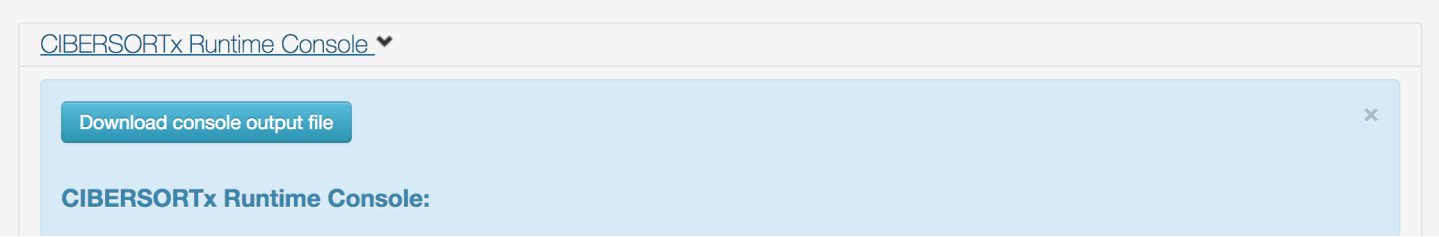
   Details and screenshots....

You should see a results page like that shown below:



Click on the "Table Output Help" button for assistance in interpreting the results and in searching, filtering and contrast adjustment of the table.

If you wish to record the parameters you used, you can click on the CIBERSORTx Runtime Console link. It will display the raw output from CIBERSORTx, as well as a Download button to save the output as a .txt file.



7. **Review past jobs.** If you run multiple jobs, you can review past job results at the Job Results (jobs.php) page.
   Details and screenshots....

Tutorial 3 - Cross-platform deconvolution

Follow this tutorial to learn how to **perform cross-platform deconvolution**. This tutorial illustrates the use of a single cell RNA-Sequencing (scRNA-Sec signature matrix to deconvolve cell subsets from bulk RNA-Seq data of whole-blood samples. As the signature matrix was derived from droplet-based scRNA-Seq data (10x Chromium), and the mixture file was derived from RNA-Seq profiles, this tutorial shows how one can **apply S-mode batch-correction to minimize the impact of cross-platform variation** on the results (related to Figure 2b,e and Supplementary Figures 1 and 2 in Newman e al., submitted). You can follow this tutorial using your own dataset, or the Example Dataset. This tutorial is based on the Example dataset **«NSCLC PBMCs Single Cell RNA-Seq (Fig. 2b) (cell fractions only)»**.

More...

Tutorial 4 - Impute Gene Expression, Group Mode

Follow this tutorial to learn how to run the CIBERSORTx, Group analysis module. Here you will learn how to **impute (i.e., "purify") cell type-specific gene expression profiles from a group of bulk tissue transcriptomes** without the need for physical cell sorting. Note that this approach will learn a single representative transcriptome profile for each cell type in the signature matrix. To infer sample-level variation, see tutorial 5. You can follow this tutorial using your own dataset, or the Example Dataset. This tutorial is based on the Example Dataset **«Group Level GEPs - Non-Small Cell Lung Cancer (Fig. 3g)»**.

More...

Tutorial 5 - Impute Gene Expression, High-Resolution Mode

Follow this tutorial to learn how to run the CIBERSORTx High-Resolution analysis module. Unlike Group Mode, High Resolution Mode is used to impute **sample-level gene expression variation of distinct cell types from a collection of bulk tissue transcriptomes**. The output is an expression matrix fo each cell type rather than a single transcriptome profile (as described in tutorial 4). You can follow this tutorial using your own dataset, or the Example Dataset. This tutorial is based on the example dataset **«High-Resolution GEPs - Block Patterns (Fig. 4a)»**.

More...

Tutorial 6 - Build a Signature Matrix File from Sorted Cell Populations (RNA-Seq)

Follow this tutorial to learn how to **build a custom signature matrix file using RNA sequencing data from sorted cell populations**. The resulting signature matrix file consists of barcode genes that can discriminate each cell subset of interest in a given tissue type, and can be subsequently used to run CIBERSORTx.

More...

Tutorial 7 - Build a Signature Matrix File from Sorted Cell Populations (microarray)

Follow this tutorial to learn how to **build a custom signature matrix file from sorted cell populations profiled by microarray**. The resulting signature matrix file consists of barcode genes that can discriminate each cell subset of interest in a given tissue type, and can be subsequently used to run CIBERSORTx.

More...