

EpiDISH Web Toolkit Example

Shijie C. Zheng

2019-03-02

In this file, we show an example of using EpiDISH web toolkit with example data. Using the example data which we provide, you could replicate the results shown in this file.

Data preparation

Download example beta value matrix

On the **Data preparation** page, we first download the example beta value matrix. Click the text *here* as red circled in the following figure. You shall download a file named ExampleBeta.txt.

Epigenetic Dissection of Intra-Sample-Heterogeneity

EpiDISH web toolkit Home Data preparation Infer CT fraction Run CellDMC About

Data privacy statement
The data you upload will **NOT** be stored, reused, or shared by us in any form.

Beta value matrix

Here you upload your beta value matrix with rows labeling the CpGs (usually Illumina BeadArray probe IDs) and columns labeling samples. NA values are not allowed. If you only want to infer cell-type fractions (not running CellDMC later), you can upload a subset beta value matrix, which only contains cell-type specific CpGs as in the reference matrix.

You can download the example beta value file [here](#) (Tips: The first column of your data should have a name, e.g. cpG. The values of the first column will be used as feature names later.). Both **txt** and **csv** formats are acceptable. You can choose the separator (tab, comma, or semicolon).

Upload your data

Browse... No file selected

Separator

Tab
 Comma
 Semicolon

Fig 1. Download the example beta value matrix

Check beta value matrix

Next, upload the file you just downloaded. The page will automatically update, and you will see a preview of the matrix you just uploaded as shown below. The number of CpGs and samples will show up above the data table.

Beta value matrix

Here you upload your beta value matrix with rows labeling the CpGs (usually Illumina BeadArray probe IDs) and columns labeling samples. NA values are not allowed. If you only want to infer cell-type fractions (not running CellDMC later), you can upload a subset beta value matrix, which only contains cell-type specific CpGs as in the reference matrix.

You can download the example beta value file [here](#) (Tips: The first column of your data should have a name, e.g. cpg. The values of the first column will be used as feature names later.). Both **txt** and **csv** formats are acceptable. You can choose the separator (tab, comma, or semicolon).

Upload your data

Browse... ExampleBeta.txt
Upload complete

Separator

Tab
 Comma
 Semicolon

Quote

None
 Single quote
 Double quote

You have successfully uploaded the data. **1000** CpGs and **40** samples found. Here is the preview:

Show 5 entries Search:

| | cpg | Sample_1 | Sample_2 | Sample_3 | |
|---|------------|---------------------|---------------------|---------------------|-----|
| 1 | cg14882966 | 0.326242563956046 | 0.547716260779816 | 0.452204454796483 | C |
| 2 | cg19890469 | 0.00403381428195862 | 0.00544929114795679 | 0.00323926553767248 | 0.0 |
| 3 | cg19322065 | 0.0427346265291461 | 0.0318428868338324 | 0.0826528209202898 | C |
| 4 | cg10066189 | 0.020746571268345 | 0.0108183217494757 | 0.0139050528091086 | 0.0 |
| 5 | cg00442869 | 0.0130025359152673 | 0.00241733881114161 | 0.0116880670854752 | 0.0 |

Showing 1 to 5 of 10 entries Previous 1 2 Next

Fig 2. Beta value matrix uploaded

Upload and check POI

Similarly, you can download and upload example phenotype of interest (POI) file. The number of samples in POI file should match the number of samples in beta value matrix (an error message will show up if not matched).

POI (Phenotype of interest)

Here you upload your POI vector file. It will be used in CT fractions boxplot and CellDMC. This is not required for CT fraction inference.

You can download the example POI vector file [here](#) (Tips: The first column of your data should have a name, e.g. SamleName. The values of the first column will be used as sample names later.). Both **txt** and **csv** formats are acceptable. You can choose the separator (tab, comma, or semicolon).

Upload your POI

Browse... ExamplePOI.txt
Upload complete

You have successfully uploaded POI info. **40** samples found. Here is the preview:

Show **5** entries Search:

| | SampleName | POI |
|---|------------|-----|
| 1 | Sample_1 | 0 |
| 2 | Sample_2 | 0 |
| 3 | Sample_3 | 0 |
| 4 | Sample_4 | 0 |
| 5 | Sample_5 | 0 |

Showing 1 to 5 of 40 entries Previous **1** 2 3 4 5 ... 8 Next

Separator

- Tab
- Comma
- Semicolon

Quote

- None
- Single quote
- Double quote

Fig 3. POI preview

Upload and check covariates matrix

The downloading and uploading of covariate matrix are pretty much identical to the previous steps. The covariates matrix will only be used in CellDMC to adjust other known confounding factors. In this example setting, we do not use the covariate matrix. But similarly, you could use the example covariates matrix file if they have problem with file formatting.

Infer CT fraction

Conduct inference

Next, we move to **Infer CT fraction** page. After selecting the reference, cell-types, deconvolution method, and corresponding parameters, you should click *Run* button as red circled in Fig 4. On the right side of the page, you will see a data table, showing inferred cell-type fractions.

Epigenetic Dissection of Intra-Sample-Heterogeneity

Select reference:
EpiFibIC

You are using general epithelial tissue reference. In this reference, Epi stands for general epithelial cells; Fib stands for general fibroblasts; IC stands for total immune cells.

Select CTs:
Epi Fib IC

Select method:
Robust partial correlation

Max iteration:
50

Hierarchical deconvolution on immune cells?

Run

Showing 1 to 5 of 40 entries

| SampleName | Epi | Fib | IC |
|------------|--------------------|--------------------|--------------------|
| Sample_1 | 0.785205601924157 | 0.214794398075843 | 0 |
| Sample_2 | 0.0837886066873242 | 0.873732512685156 | 0.0424788806275202 |
| Sample_3 | 0 | 0.878415485482975 | 0.121584514517025 |
| Sample_4 | 0.0316755620946222 | 0.170461010477185 | 0.797863427428193 |
| Sample_5 | 0.747423666792379 | 0.0869757463066017 | 0.165600586901019 |

Download estimated fractions

Fig 4. CT fraction inference

Generate CT fraction boxplot

After the last step, you will find options to generate boxplot magically appear below the *Run* button. You can brush points on the boxplot, and their sample names and fractions will be shown in the data table below.

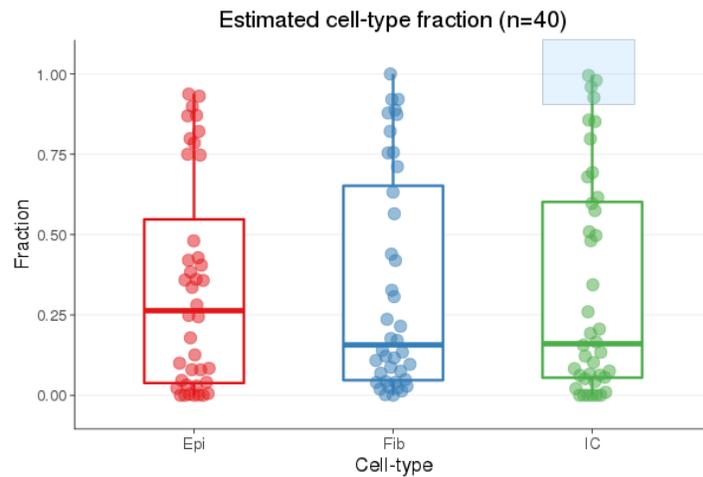
Max iteration:

Hierarchical deconvolution on immune cells?

Generate fraction boxplot?

Group by POI?

You can brush the boxplot to check data points.



[Download boxplot as pdf](#)

Show 5 entries

Search:

| leName | Epi | Fib | IC |
|--------|----------------------|---------------------|-------------------|
| _7 | 0 | 0.0404494120490641 | 0.959550587950936 |
| _20 | 0.000959271943485831 | 0.0191083428622092 | 0.979932385194305 |
| _35 | 0.0052416707124516 | 0.0685003569109161 | 0.926257972376632 |
| _40 | 0.00309936730960534 | 0.00159084509099837 | 0.995309787599396 |

Showing 1 to 4 of 4 entries

Previous Next

Fig 5. CT fraction boxplot

You can also generate the CT fraction boxplot with samples grouped by POI (not shown).

Run CellDMC

Get DMCT inference

Finally, we can identify differentially methylated cytosines in each individual cell-type. On **Run CellDMC** page, click *Run* button. The results of DMCT predictions will appear on the right side. In the table, for DMC, 1 denotes DMC, and 0 denotes non-DMC; for each cell-type, 1(-1) denotes hyper(hypo)-DMCT, and 0 denotes non-DMCT.

EpiDISH web toolkit Home Data preparation Infer CT fraction **Run CellDMC** About

Choose a method to adjust p value:
 BH FDR

Adjust p value threshold to call DMCT:
 0.05

Run

Add Illumina BeadArray annotation to the results?

Show coefficients for each CT?

Do you want a t-stat scatter plot?

Show **5** entries Search:

| CpG | DMC | Epi | Fib | IC |
|------------|-----|-----|-----|----|
| cg14882966 | 1 | 1 | 0 | 0 |
| cg19890469 | 1 | 1 | 0 | 0 |
| cg19322065 | 1 | 1 | 0 | 0 |
| cg10066189 | 1 | 1 | 0 | 0 |
| cg00442869 | 1 | 1 | 0 | 0 |

Showing 1 to 5 of 1,000 entries Previous **1** 2 3 4 5 ... 200 Next

[Download DMCT predictions matrix](#)

Fig 6. DMCT inference

We then check the DMCT inference in this example case. We could let the table show 10 entries each table page. In this example, the top 10 CpGs are only hypermethylated in epithelial cells as shown Fig 7.

Choose a method to adjust p value:
 BH FDR

Adjust p value threshold to call DMCT:
 0.05

Run

Add Illumina BeadArray annotation to the results?

Show coefficients for each CT?

Do you want a t-stat scatter plot?

Show **10** entries Search:

| CpG | DMC | Epi | Fib | IC |
|------------|-----|-----|-----|----|
| cg14882966 | 1 | 1 | 0 | 0 |
| cg19890469 | 1 | 1 | 0 | 0 |
| cg19322065 | 1 | 1 | 0 | 0 |
| cg10066189 | 1 | 1 | 0 | 0 |
| cg00442869 | 1 | 1 | 0 | 0 |
| cg02363317 | 1 | 1 | 0 | 0 |
| cg23084951 | 1 | 1 | 0 | 0 |
| cg10926574 | 1 | 1 | 0 | 0 |
| cg06924976 | 1 | 1 | 0 | 0 |
| cg07754492 | 1 | 1 | 0 | 0 |

Showing 1 to 10 of 1,000 entries Previous **1** 2 3 4 5 ... 100 Next

Fig 7. Epithelial hyper-DMCTs

On the second DMCT table page, we see that these 10 CpGs are only hypomethylated in epithelial compartment, as shown in Fig 8.

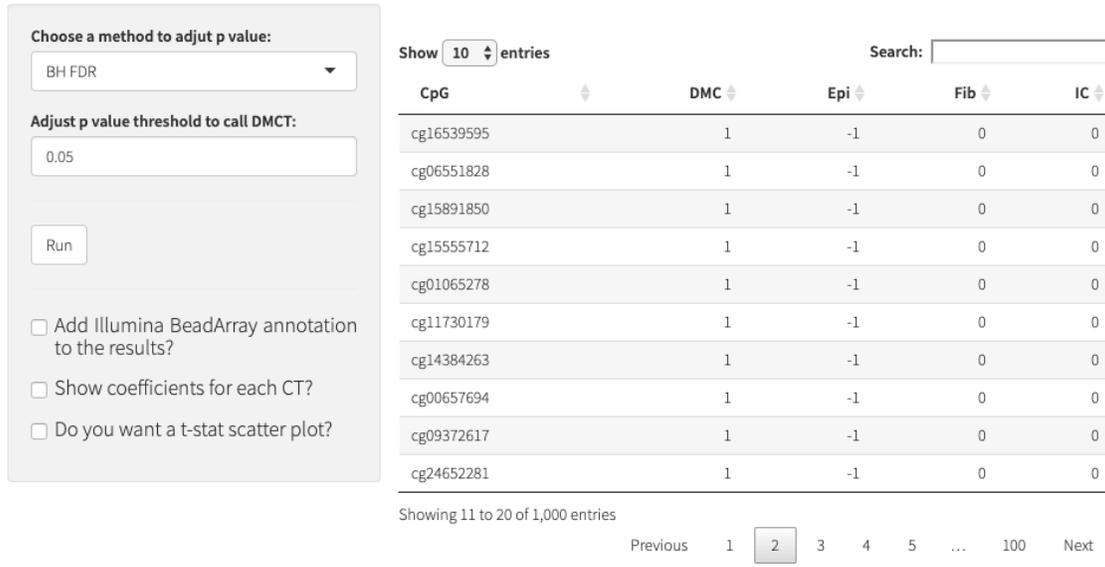


Fig 8. Epithelial hypo-DMCTs

On the third DMCT table page, these 10 CpGs are hypermethylated in both of epithelail cells and fibroblasts (Fig 9).

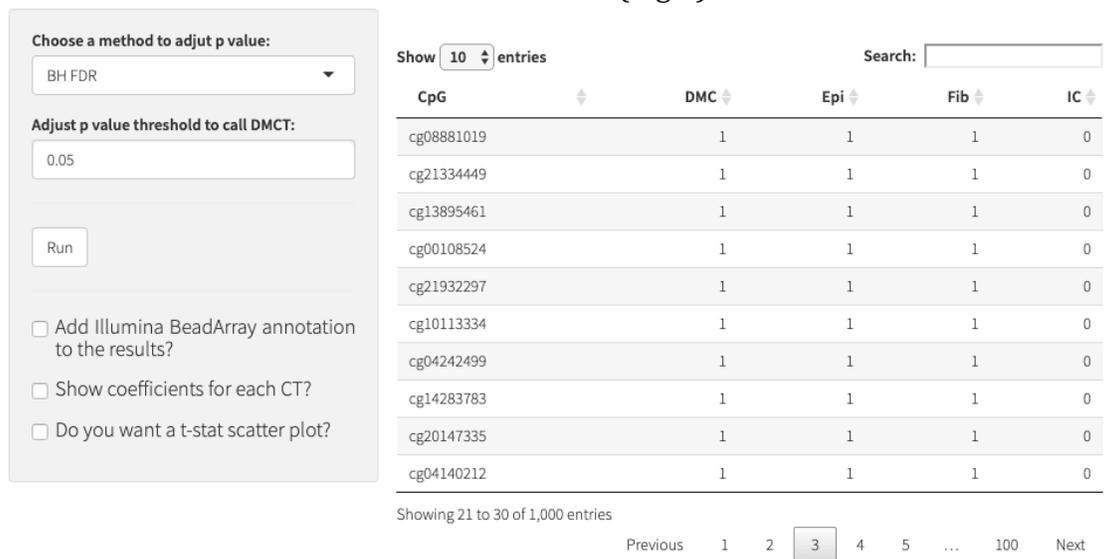


Fig 9. Hyper-DMCTs in Epi and Fib

Using Illumina BeadArray annotation

By checking the orange circled box in Fig 10, annotations will be merged into the DMCT inference table. And you can search the DMCT table (red circle in Fig 10). In this example, we input *SEPT9*, the table will automatically filter all CpGs mapped to the gene *SEPT9*.

Choose a method to adjust p value:
 BH FDR

Adjust p value threshold to call DMCT:
 0.05

Run

Add Illumina BeadArray annotation to the results?

Select BeadArray type:
 450k

Show coefficients for each CT?
 Do you want a t-stat scatter plot?

Show 5 entries

Search: SEPT9

| C | chr | pos | strand | Islands_Name | Relation_to_Island | UCSC_RefGene |
|---|-------|----------|--------|--------------|--------------------|----------------|
| 0 | chr17 | 75283800 | + | | OpenSea | SEPT9;SEPT9 |
| 0 | chr17 | 75283979 | + | | OpenSea | SEPT9;SEPT9;SE |

Showing 1 to 2 of 2 entries (filtered from 1,000 total entries) Previous 1 Next

Download DMCT predictions matrix

Fig 10. DMCT annotation search

Get coefficients for each CT

By checking the orange circled box, you will get coefficients for each CT. In the table, *Estimate* denotes the estimated DNAm changes of a CpG in a cell-type (Fig 11).

Show coefficients for each CT?

Select CT to show coefficients:
 Epi

Do you want a t-stat scatter plot?

Download DMCT predictions matrix

Here are coefficientnets for Epi:

Show 5 entries

Search:

| CpG | Estimate | SE | t | p |
|------------|-------------------|--------------------|------------------|----------------------|
| cg14882966 | 0.746706910682142 | 0.0776963336469451 | 9.61058103558921 | 3.19386618861789e-11 |
| cg19890469 | 0.762356884370576 | 0.0886084210366875 | 8.60366176771088 | 4.73221298949738e-10 |
| cg19322065 | 0.868113332609243 | 0.0615285429750569 | 14.1091157149807 | 8.97788493970543e-16 |
| cg10066189 | 0.805494247292066 | 0.0889874977988759 | 9.05176870027961 | 1.40290451680991e-10 |
| cg00442869 | 0.841375807716959 | 0.0600231950289187 | 14.0175111856607 | 1.08551350917043e-15 |

Showing 1 to 5 of 1,000 entries Previous 1 2 3 4 5 ... 200 Next

Download coefficients for Epi

Fig 11. Coefficients for each CT