User manual for QC & analaysis parts of the app

Louisa Hadj Abed

4/06/2022

Here, we show CellDestiny functionalities using lentiviral barcoding data studied in HadjAbed et al. 2022.

Go on the web app : https://perie-team.shinyapps.io/CellDestiny/.

In it, you see on the left a menu with the two parts: QC and Analysis. Both of them have their respective "Load data" sub-parts where you can load your matrix and metadata files.

To play with the test dataset, do as shown in the README.

QC

Load data and select dupliacte variable name

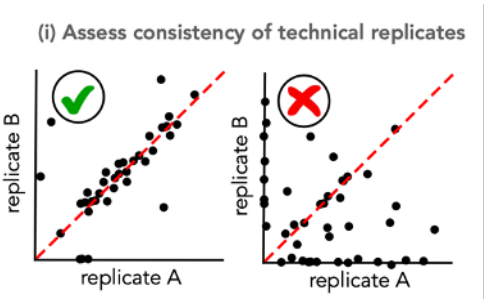
The first step of the QC part of the app is to load count and metadata matrcies and select the variable describing your *duplicates*. It corresponds to one of your metadata column name.

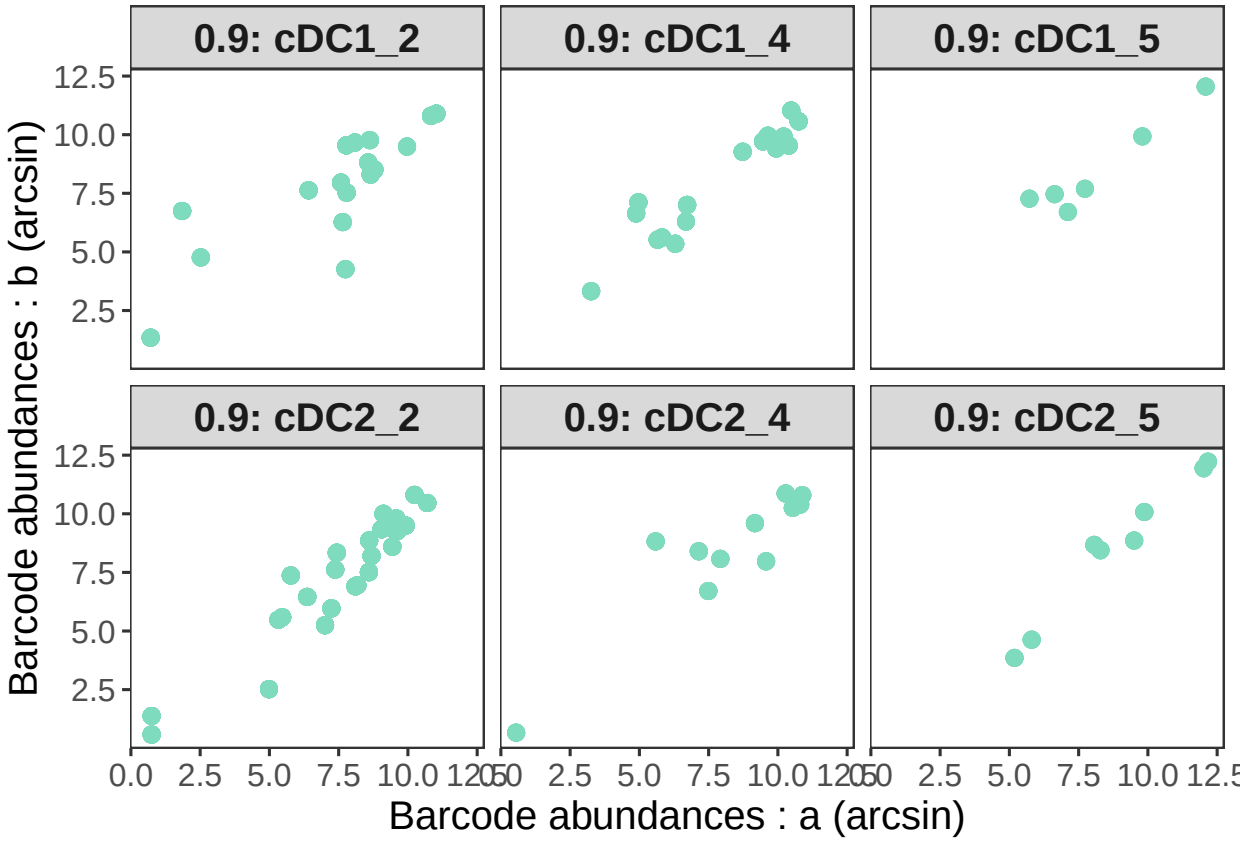
For the test dataset, it is "duplicates".

```
##   type mouse duplicates
## 1 cDC1     2          a
## 2 cDC2     4          b
## 3          5
```

Duplicates checking

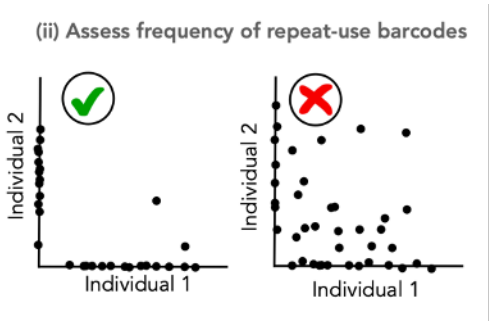
Here we look at the consistency of technical replicates.

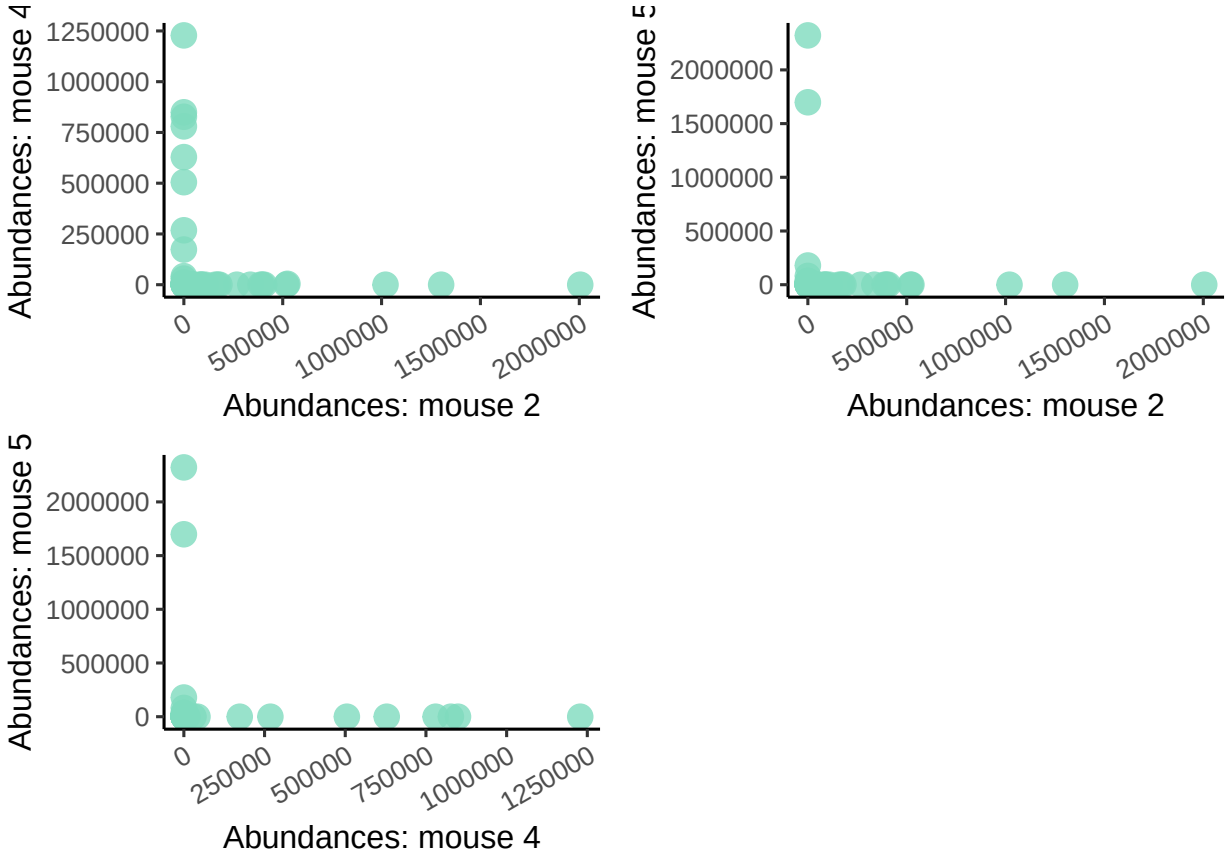




Repeat Use checking

Here we assess the frequency of repeat use barcodes in the data. Repeat used barcodes are compared between individuals.





Both duplicates and repeat use checking are ok. We can now go further in the biological analysis and use the analysis part once duplicated samples are merged.

Analsyis part

Load data and select individuals variable name

The first step of the analysis part of the app is to load count and metadata matrcies and select the variable describing your *individuals*. It corresponds to one of your metadata column name.

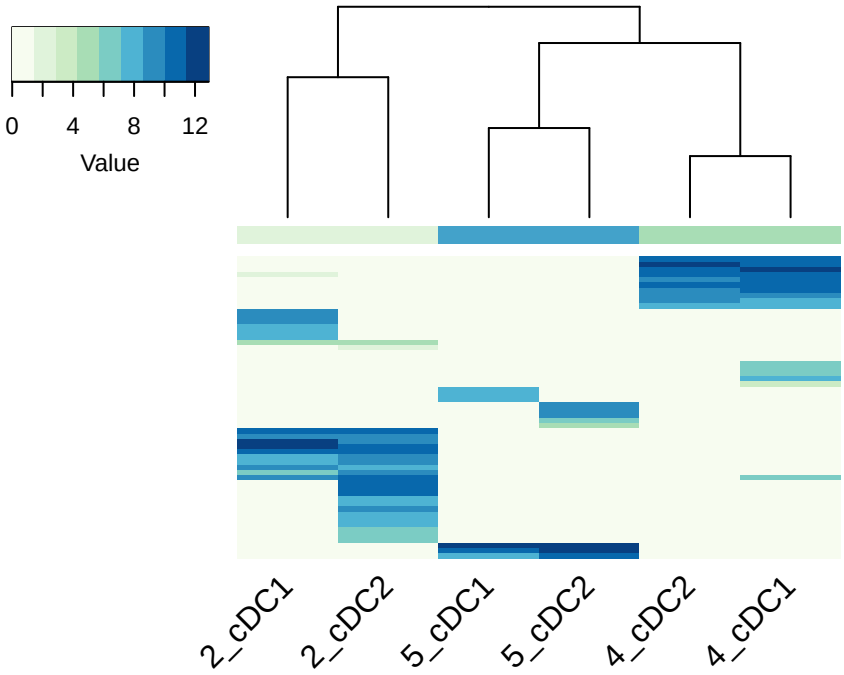
For the test dataset, it is "mouse".

```
##   type mouse
## 1 cDC1     2
## 2 cDC2     4
## 3          5
```

Sample similarities

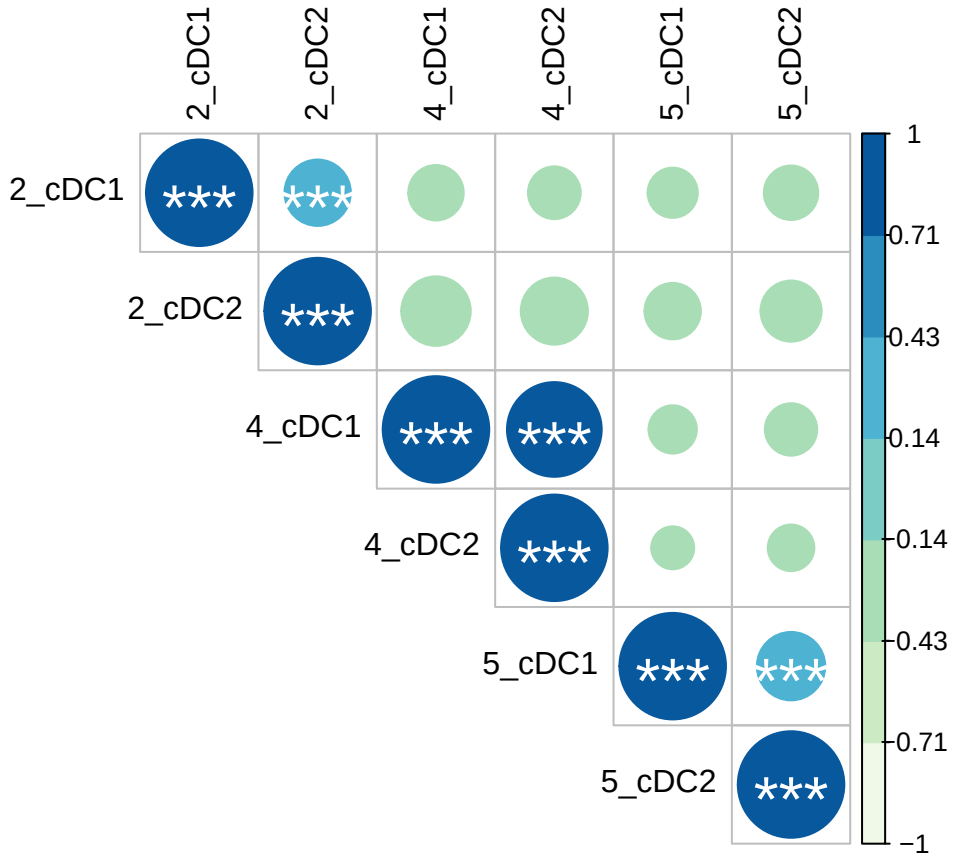
Similarities between samples can be visualized using a heatmap together with hierarchical clustering and using a correlogram.

Heatmap



Correlogram

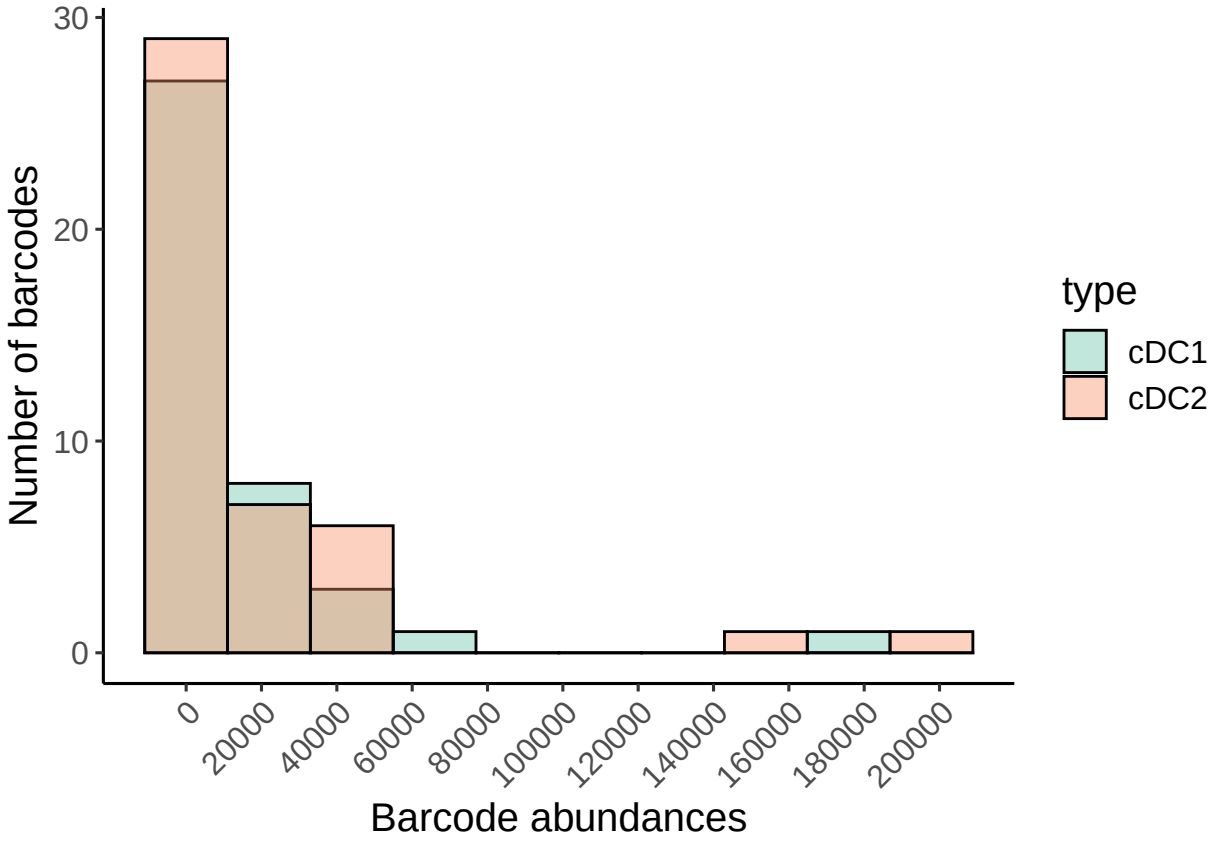
Spearman correleations are outputed.

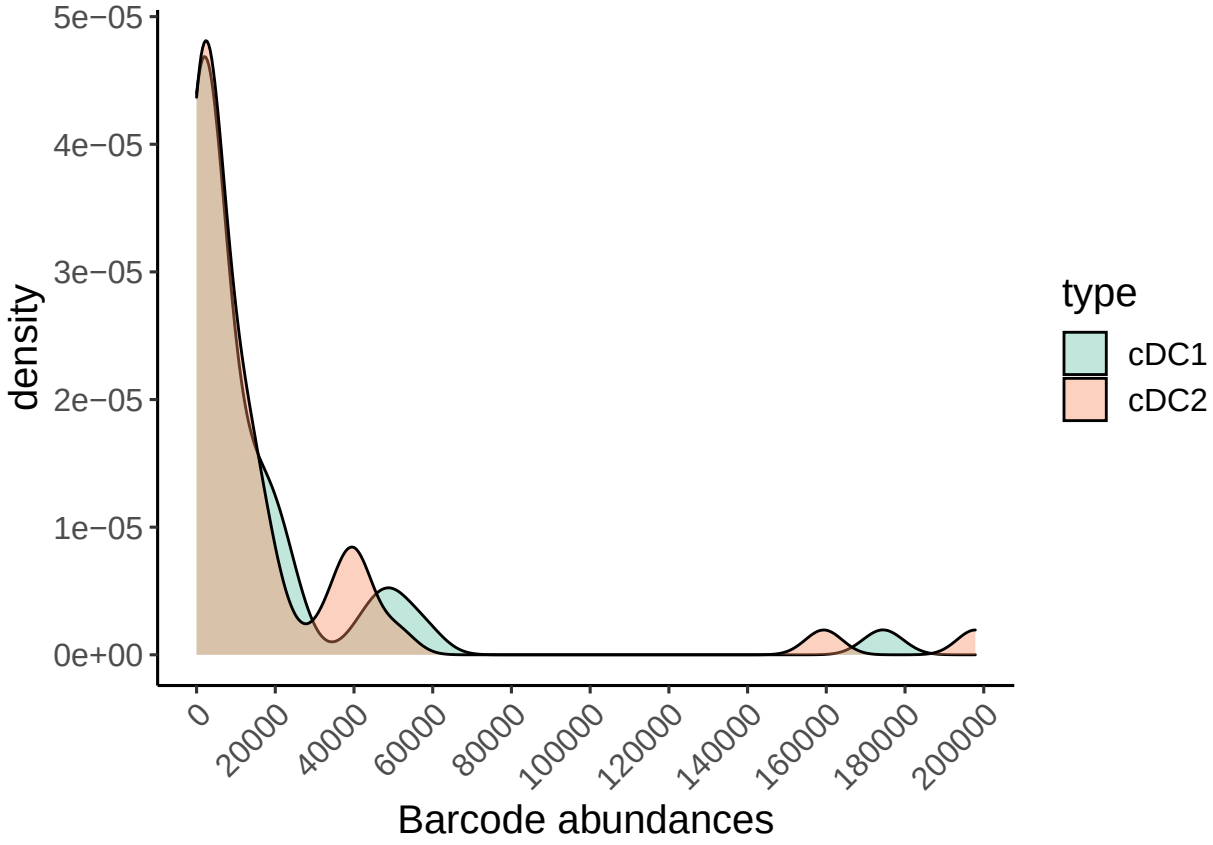


Clone sizes

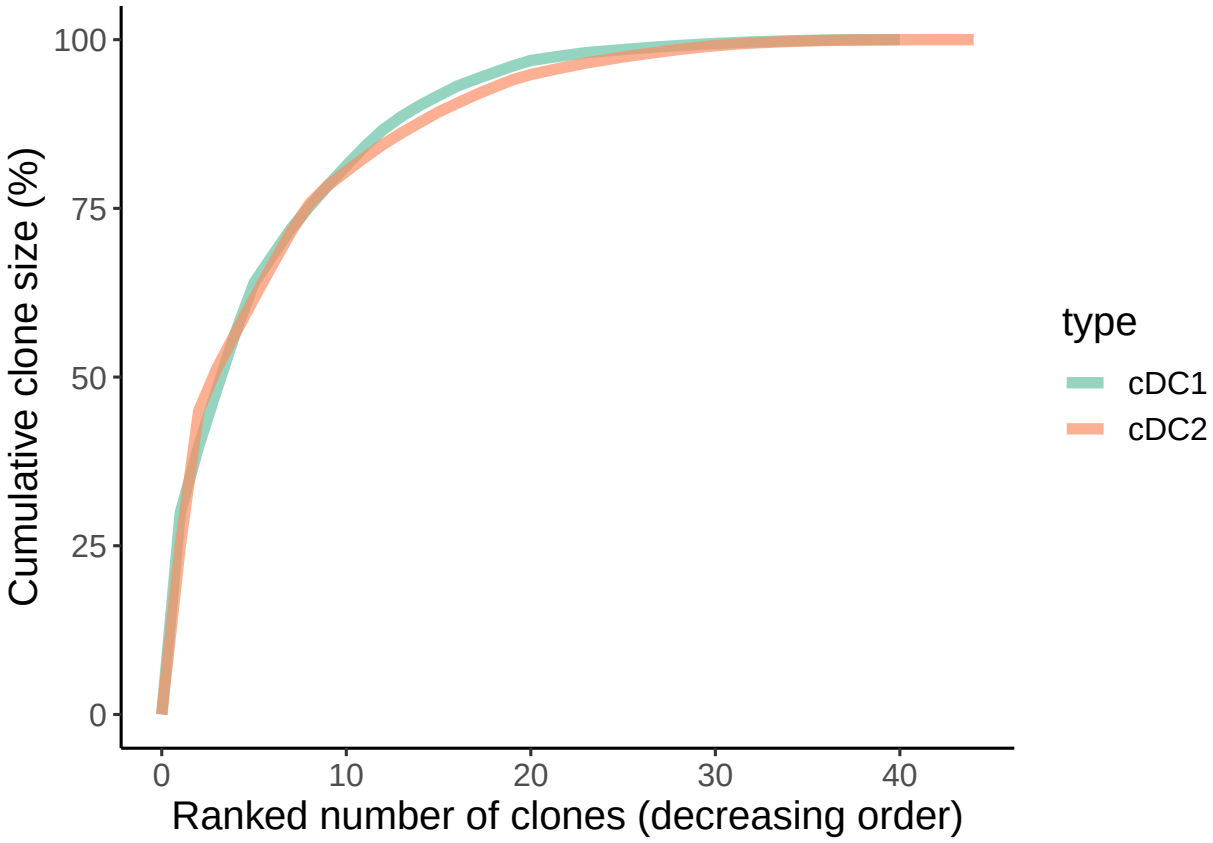
CellDestiny offers two types of clone size visualizations. The first one is a cumulative diagram. If the cumulative graph has a concave shape, it means that a cell population is dominated by a small number of large clones. On the contrary, if the shape is linear, the sample is composed of a number of clones which contribute equally to the cellularity of the population. The second type of graph is a frequency distribution plot where the user can choose between histogram or density curve-based representations of the data.

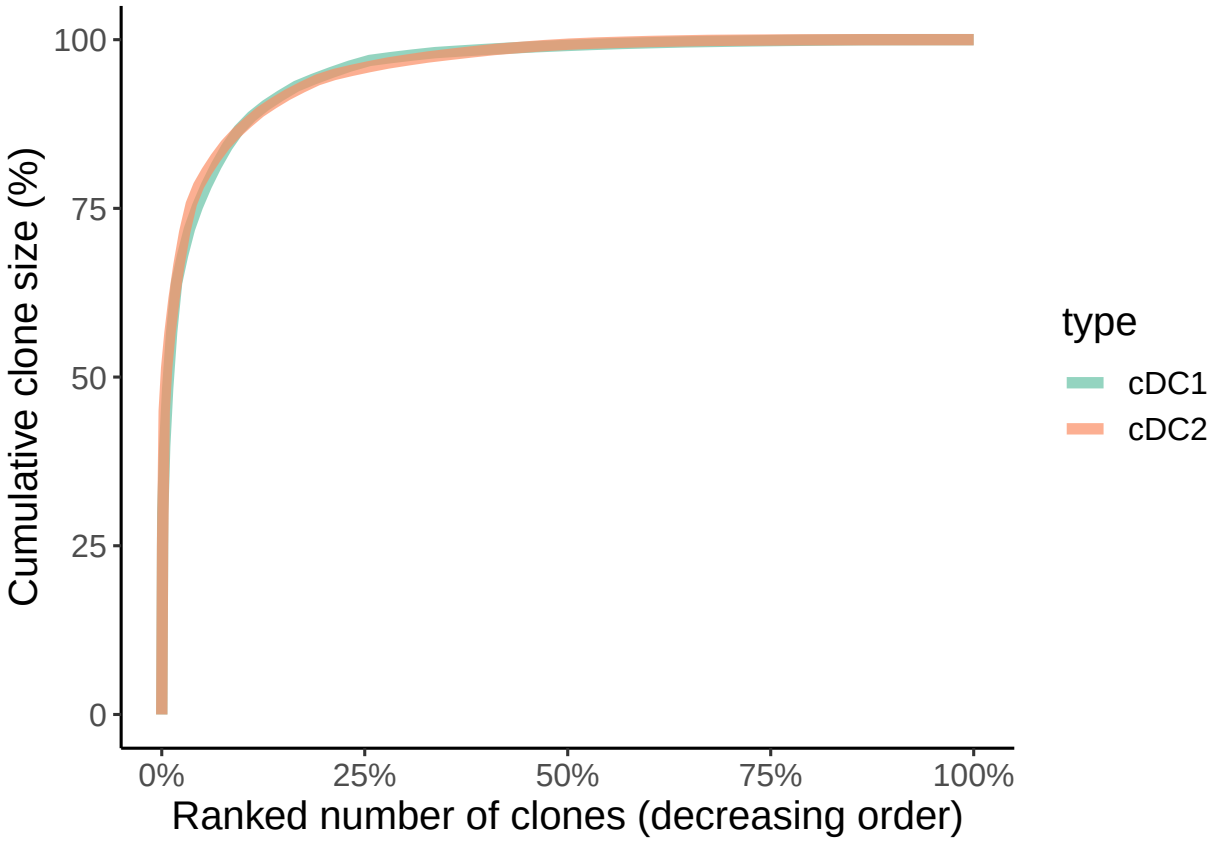
Get clone size distributions using an histogram


30
20
10
0
Number of barcodes
0
20000
40000
60000
80000
100000
120000
140000
160000
180000
200000
Barcode abundances
type
cDC1
cDC2



Plot clone-size distributions using a cumulative frequency diagram.

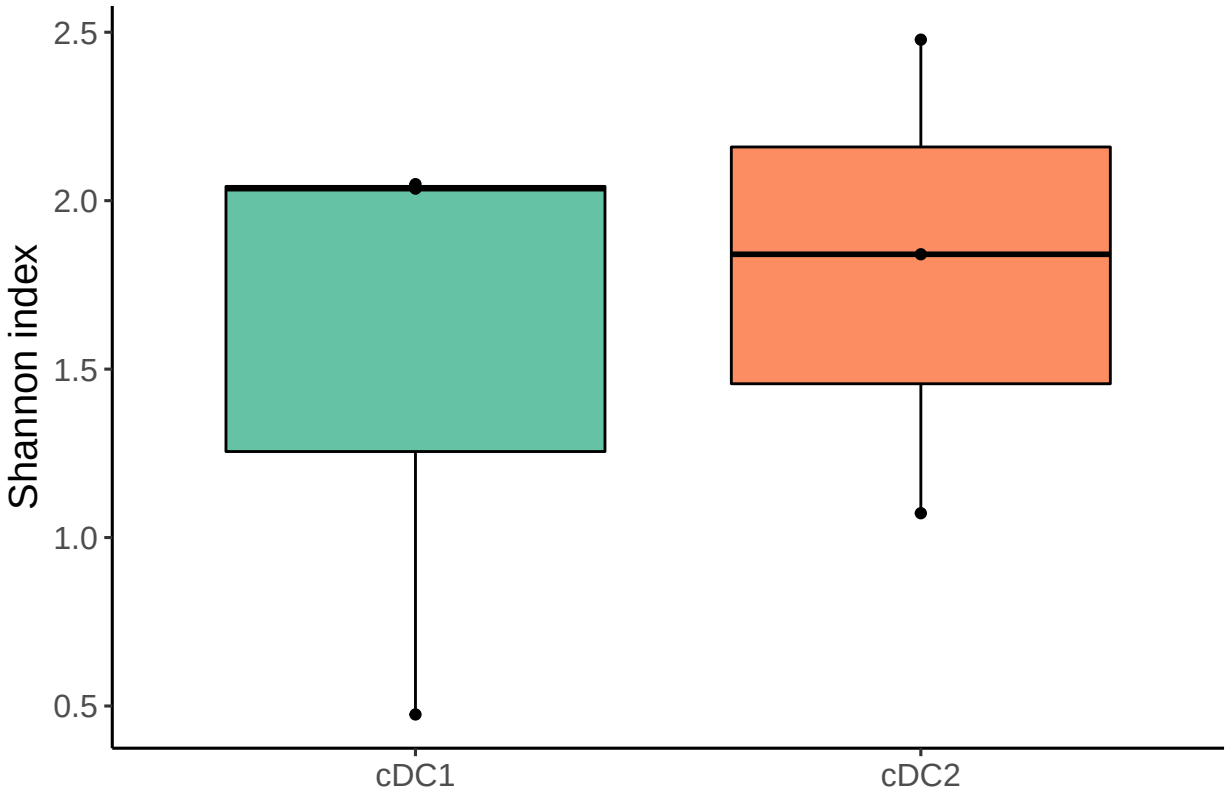

Cumulative clone size (%)
100
75
50
25
0
0
10
20
30
40
Ranked number of clones (decreasing order)
type
cDC1
cDC2



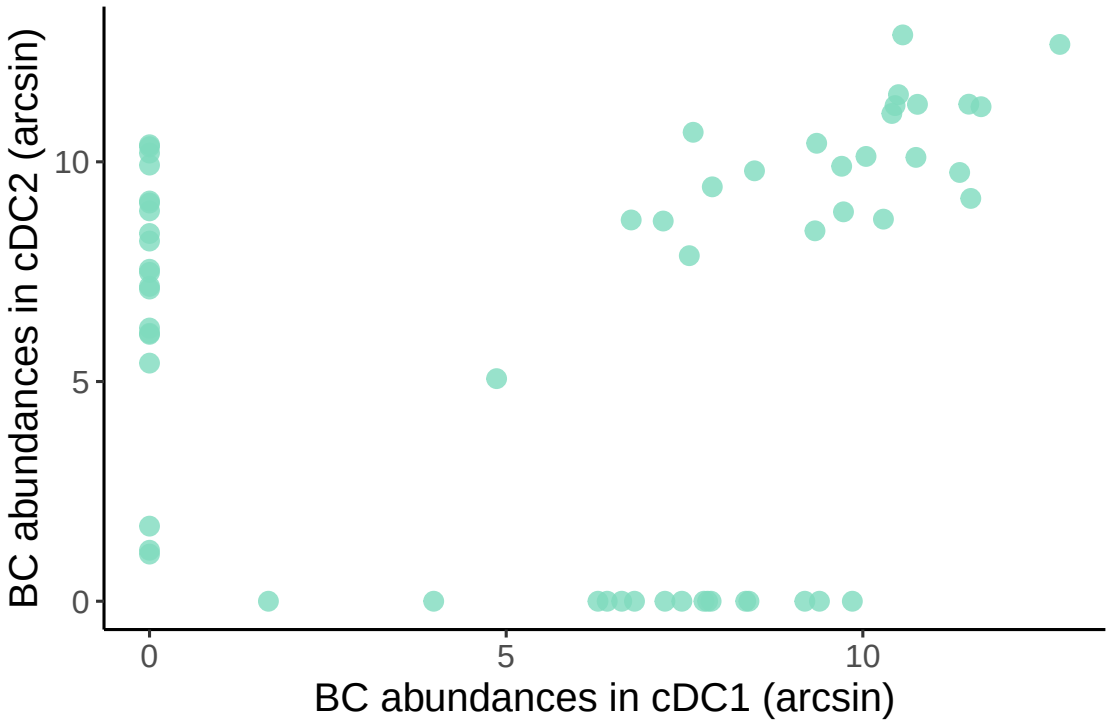
Diversity

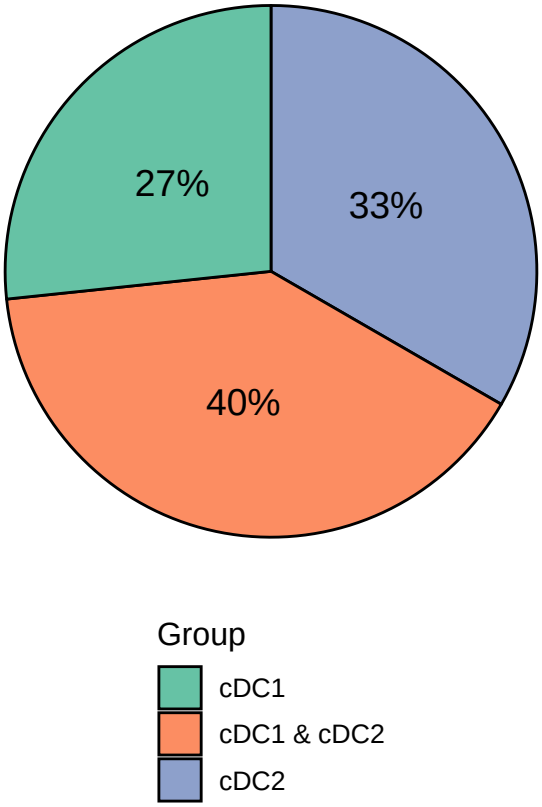
Comparing sample diversities is a common step in lineage tracing analysis. Diversity is computed using the vegan R package.

Quantify clonal diversity between cDC1 and cDC2 using the Shannon Index.



Now lets see if barcode abundances are similar between our cDC1 and cDC2 samples using scatter plot and pie chart visualisations.

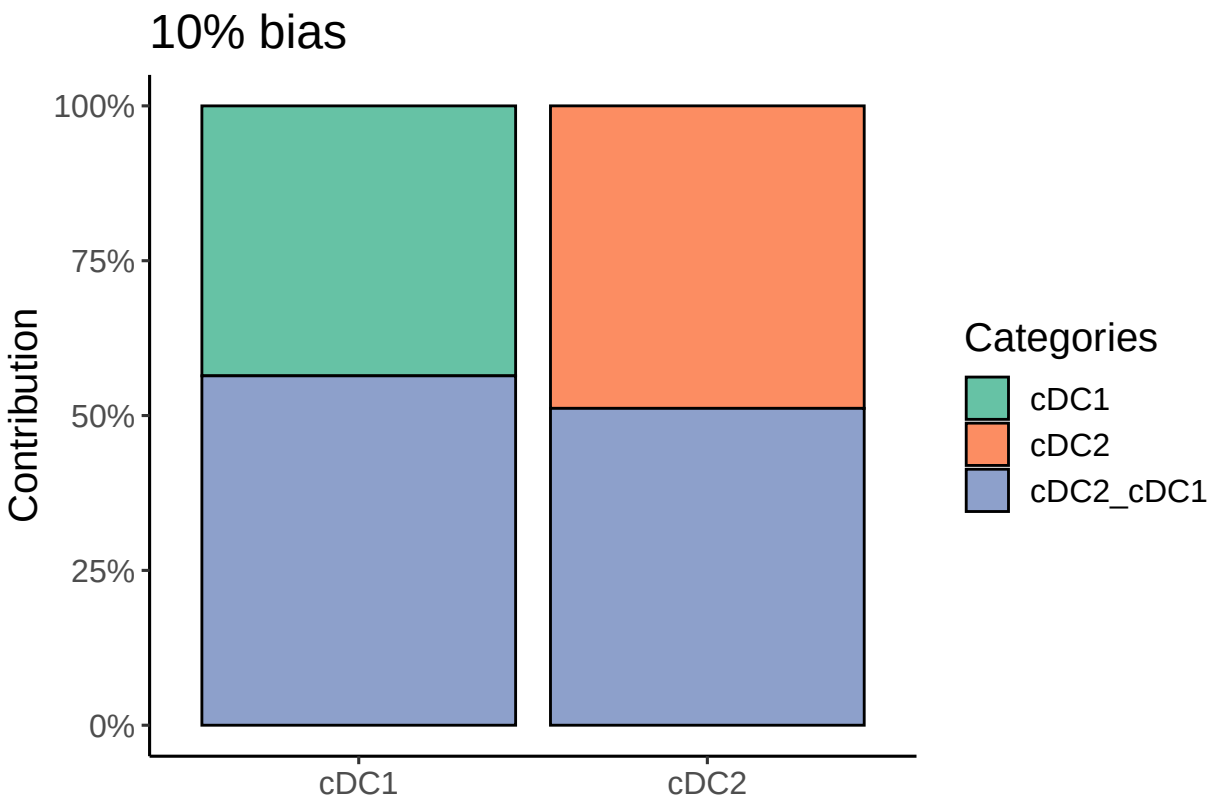

BC abundances in cDC2 (arcsin)
10
5
0
0
5
10
BC abundances in cDC1 (arcsin)

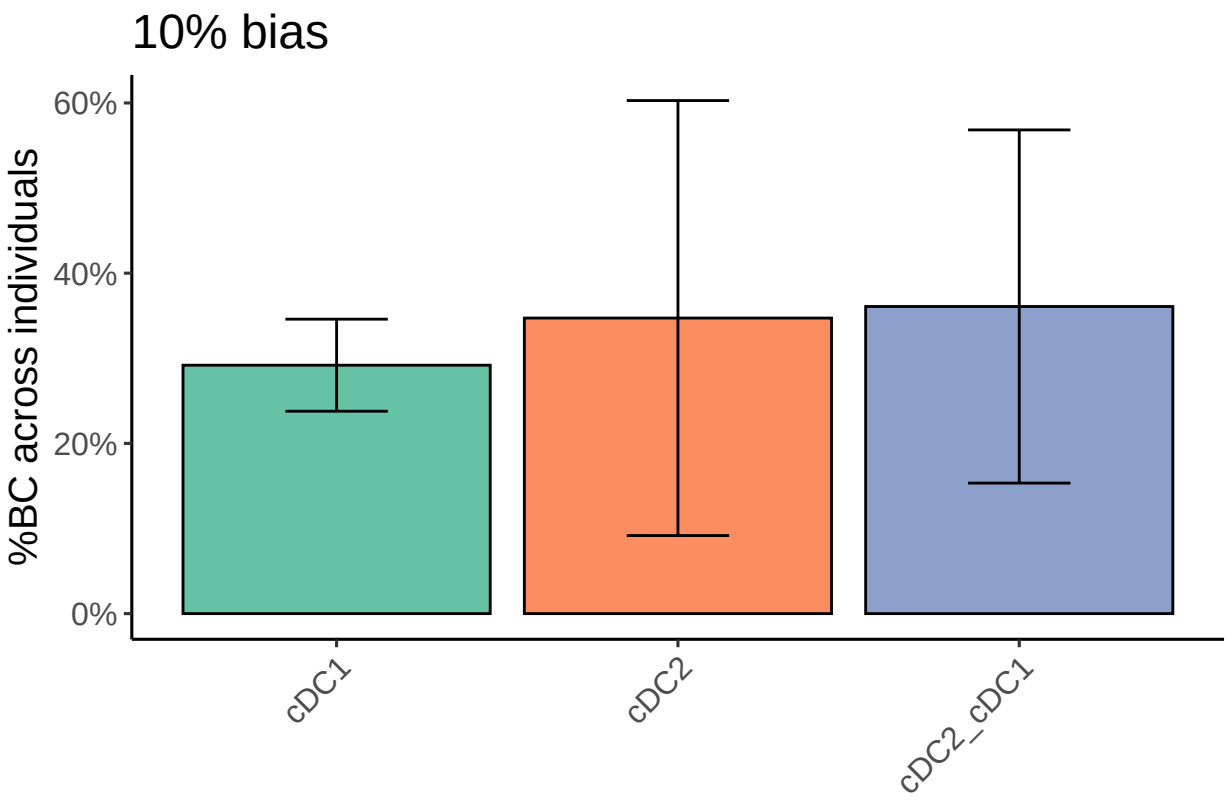


Categorisation

To classify barcodes by their lineage bias, CellDestiny uses a threshold based classifier lineage described (Naik et al. 2013b). In summary, an additional normalization step per barcode is applied in each individual, thereby enabling categorization of each barcode into classes of biased output towards the analyzed cell types. Barcodes are assigned a bias based on whether the % read abundance exceeds a threshold value. If one barcode contributes to a given lineage above the designated threshold then this barcode is assigned to be biased towards that lineage. Barcodes for which the % read abundance exceeds a threshold value across multiple lineages are classified as multi-outcome. In the CellDestiny app, the threshold used for categorization can be tuned manually.

10% bias


10% bias
Contribution
100%
75%
50%
25%
0%
cDC1
cDC2
Categories
cDC1
cDC2
cDC2_cDC1


10% bias
%BC across individuals
60%
40%
20%
0%
cDC1
cDC2
cDC2_cDC1