

User Manual for QC & Analysis Modules of Cell Destiny

Louisa Hadj Abed & Jason Cosgrove

4/06/2022

Overview In this vignette, we provide a walkthrough analysis of a lentiviral barcoding study of dendritic cells as described in: HadjAbed et al. 2022. Specifically, we wish to assess if a single MPP4 can produce both the cDC1 and the cDC2 subsets in the lung or whether a single MPP4 is fate-restricted to only produce cDC1 or cDCs.

To address this research question, we used a lentiviral barcoding approach focusing on the differentiation of (Lin-, Sca-1+, cKit+, Flt3+) MPP4s towards lung-resident cDCs. Murine MPP4s were purified from the bone marrow of donor mice by fluorescence activated cell sorting and infected with the LG2.2 lentiviral barcoding library. Labelled cells were then injected I.V into 3 irradiated recipient mice. 14 days later, lungs were isolated from the mice, and barcoded cDC1s, and cDC2s were purified by FACS. Samples were then processed for barcode detection in genomic DNA by deep sequencing.

Step 1: Open CellDestiny

Open the CellDestiny app using the following link : <https://perie-team.shinyapps.io/CellDestiny/>.

Once opened, you see on the left a menu with the QC and Analysis modules. Both have their respective “Load data” prompts which can be used to load countmatrix and metadata input files. Select the QC tab, and click the ‘load test data’ option.

The first step of the QC part of the app is to load count and metadata matrices and select the variable describing your *technical replicates*. It corresponds to one of your metadata column names. In this analysis our technical replicates are listed as “duplicates”.

Step 2: Using the Data QC Visualisation Module

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

Assessing the consistency of technical replicates PCR technical replicates allow to measure the degree of confidence in barcode detection to establish lineage relationship between cell types. If the replicates do not show high correlation, there may be an issue with barcode contamination or it may suggest poor recovery of daughter cells or poor expansion of progenitors. In CellDestiny, technical replicates can be uploaded and visualized for each sample.

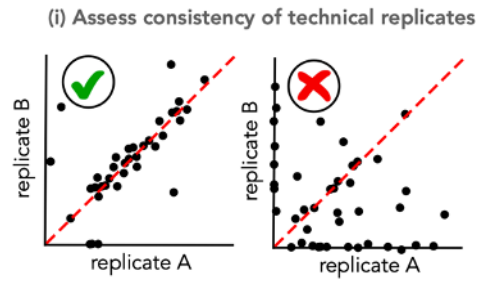
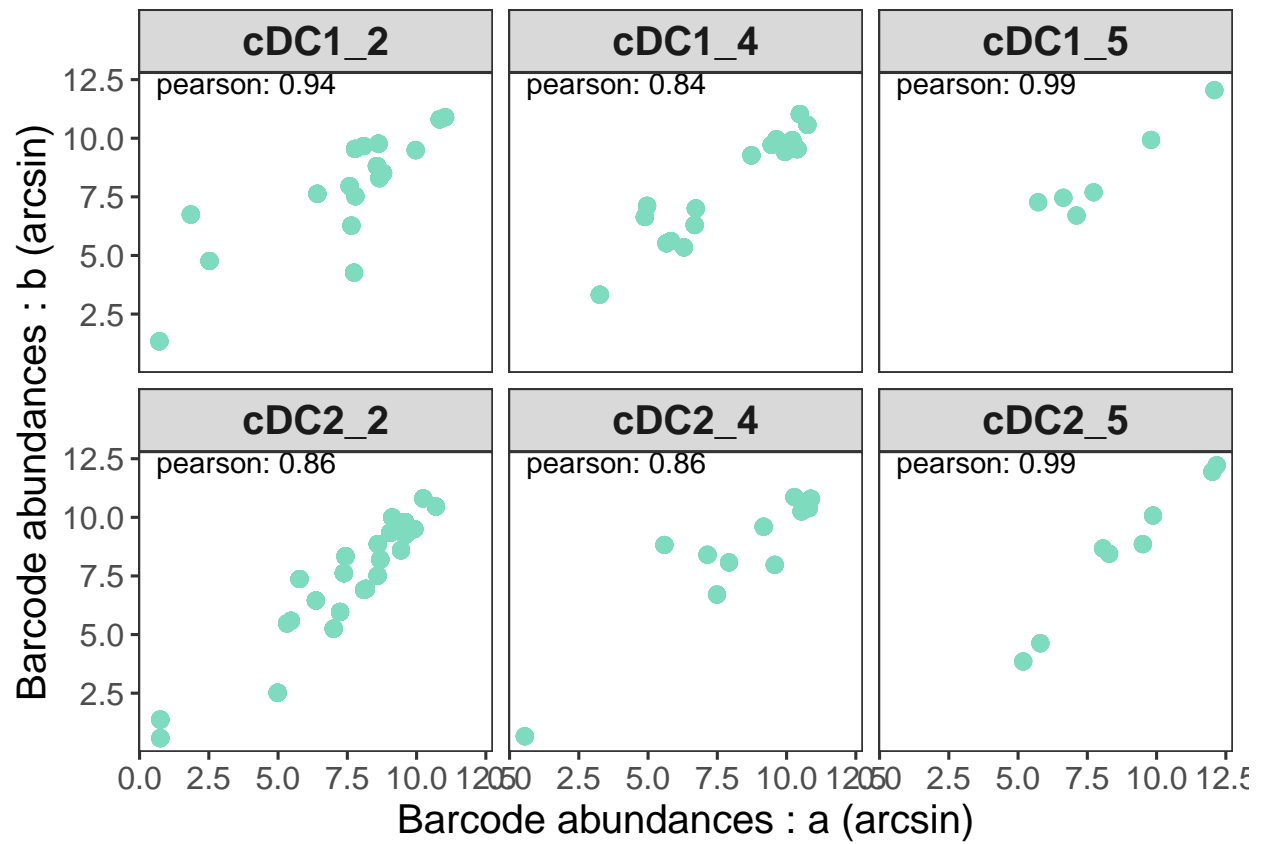


Figure 1: Checking the consistency of technical replicates



Assess the frequency of repeat-use barcodes The integration of the same barcode into multiple cells, called repeat usage, is also an important QC metric that should be considered in a lineage tracing analysis pipeline, as a high incidence of repeat usage may lead to false lineage relationship assignments. The transfer of progenitors from the same transduction batch into at least two separate mice, followed by subsequent comparison of the barcodes recovered from those mice, can be used to estimate the frequency of repeat barcode use within one mouse.

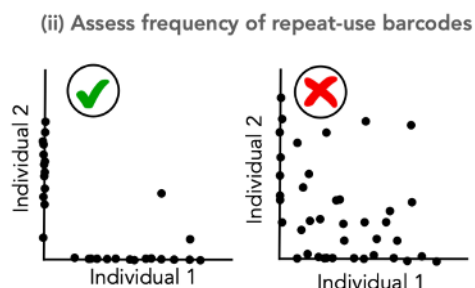


Figure 2: Assessing the frequency of repeat-use barcodes

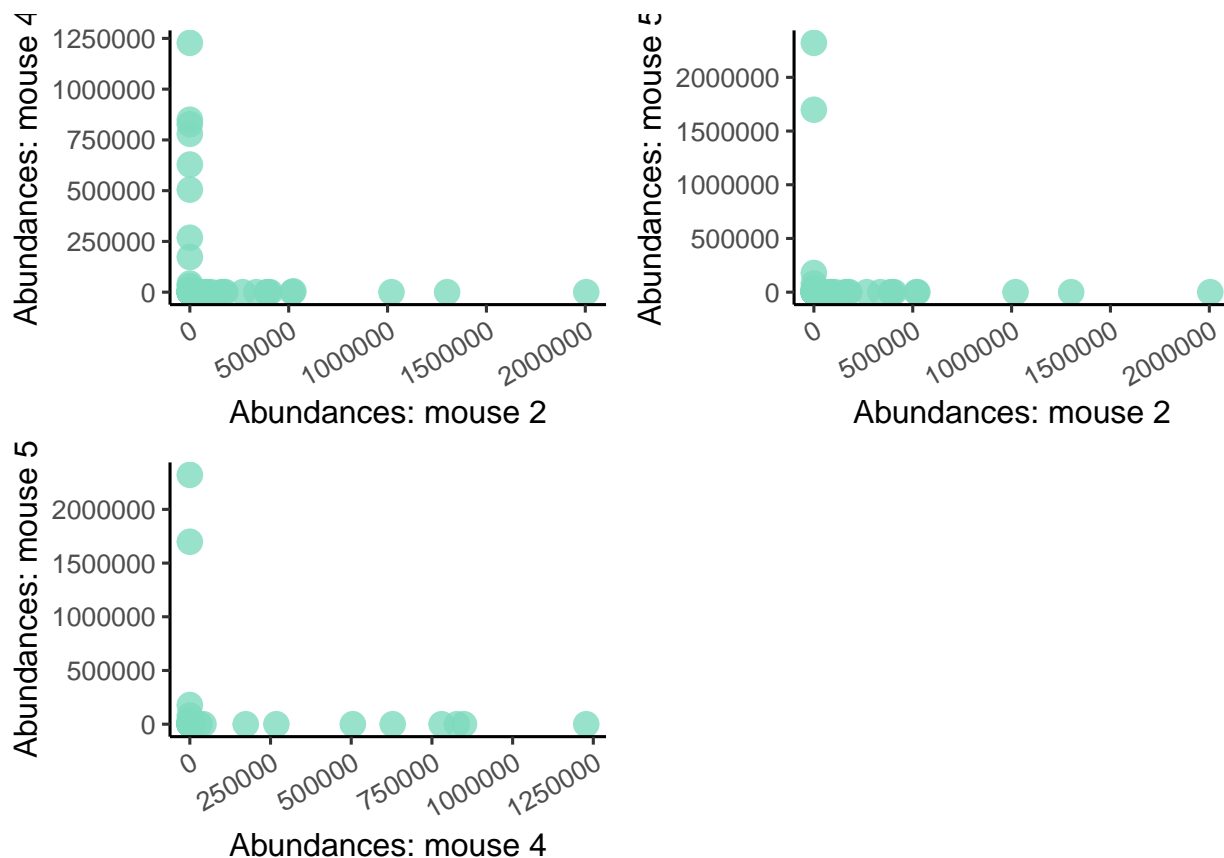


Figure 3: assessing the frequency of repeat-use barcodes

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.

Step 3: Using the Data Analysis Module

First we need to load the metadata and count matrices into the analysis module. We then 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.

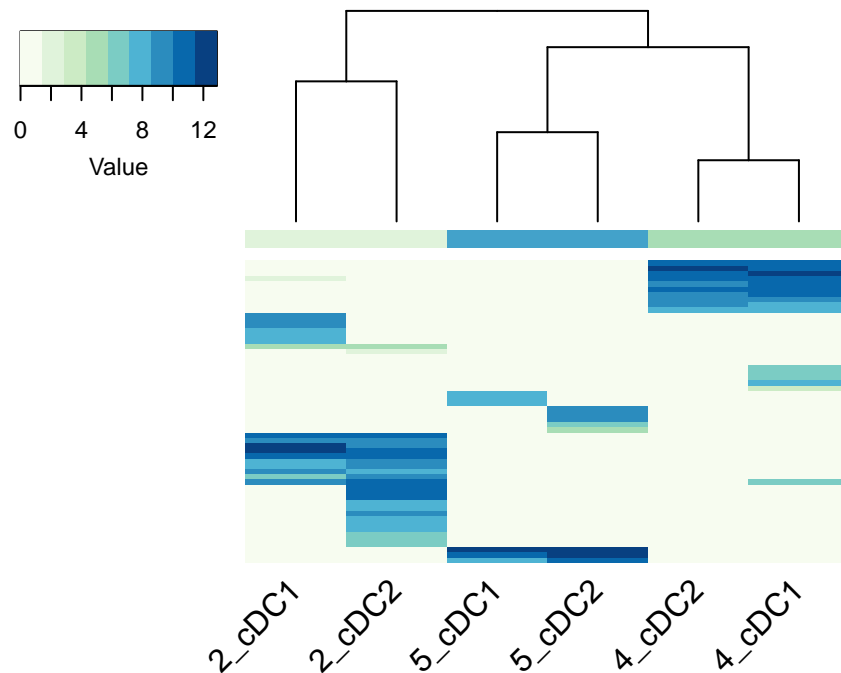


Figure 4: assessing sample similarity using a heatmap and correlogram visualisations

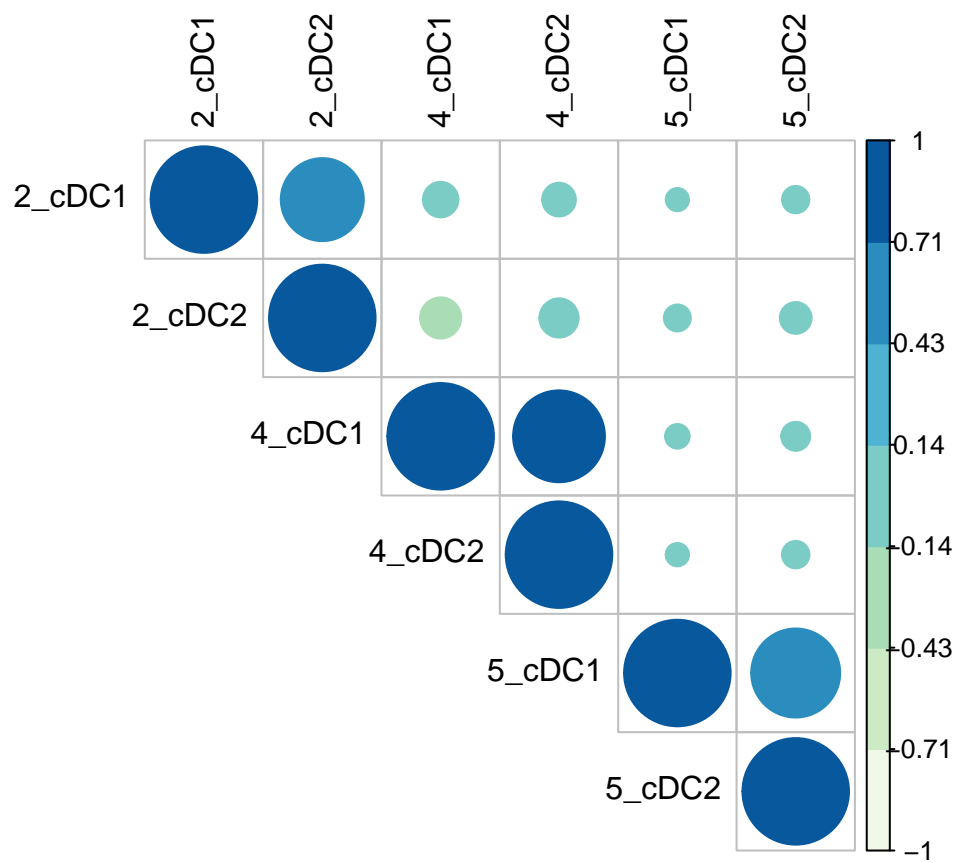


Figure 5: Spearmans correlations between technical replicates

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.

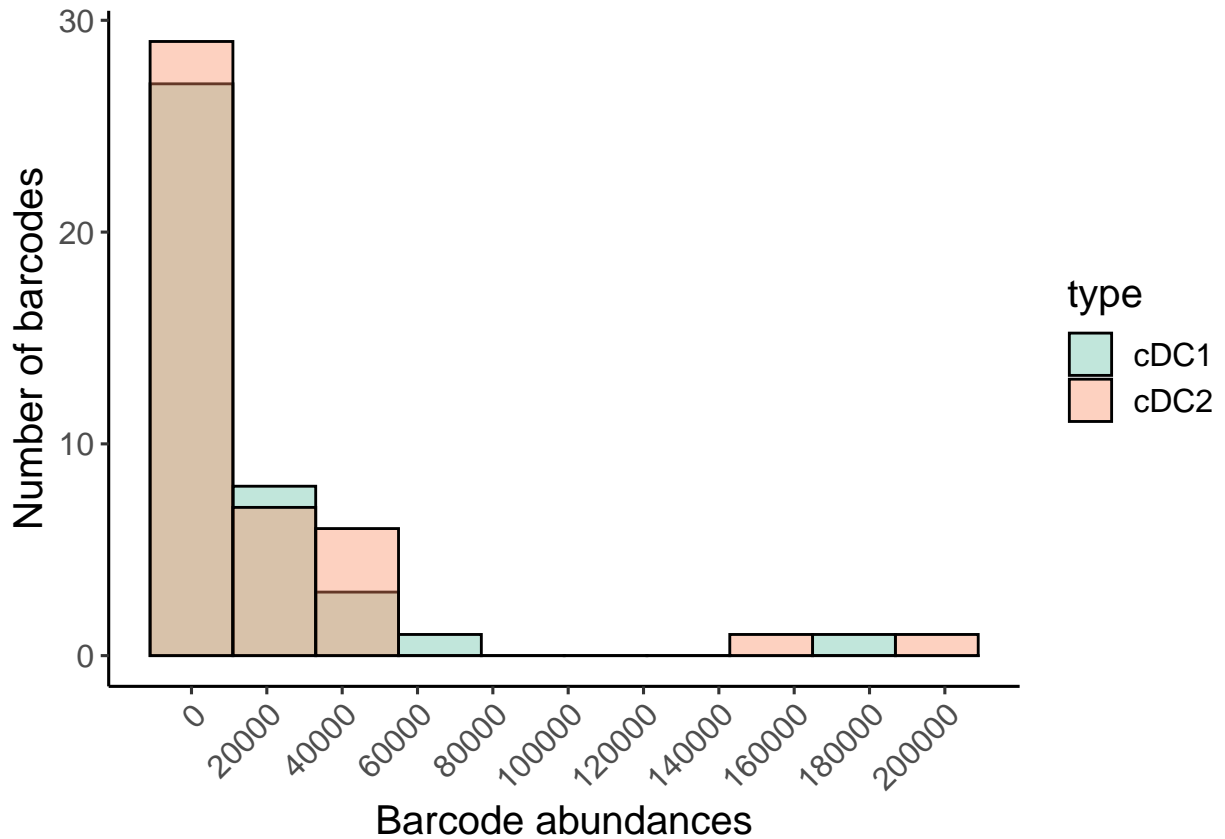


Figure 6: assessing clone size distributions

Now let's plot the clone-size distributions using a cumulative frequency diagram.

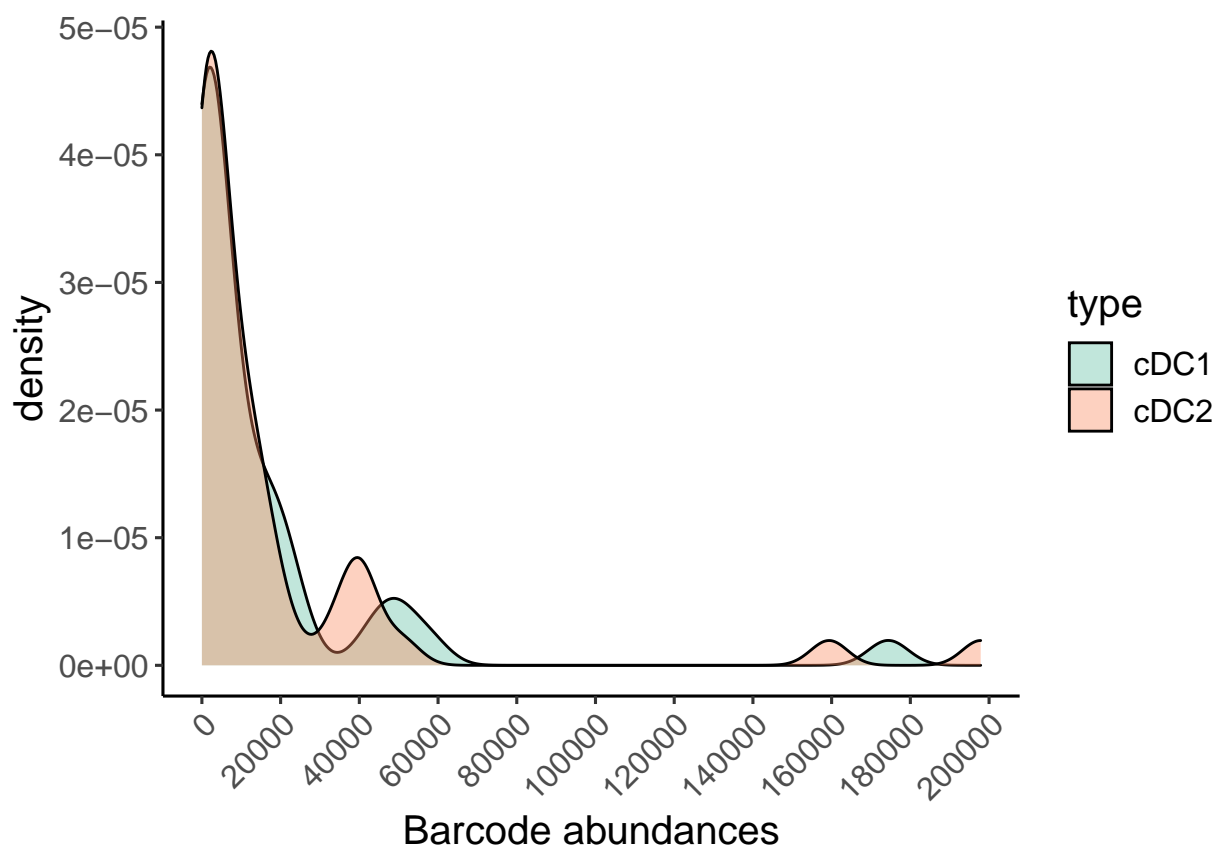


Figure 7: assessing clone-size distributions

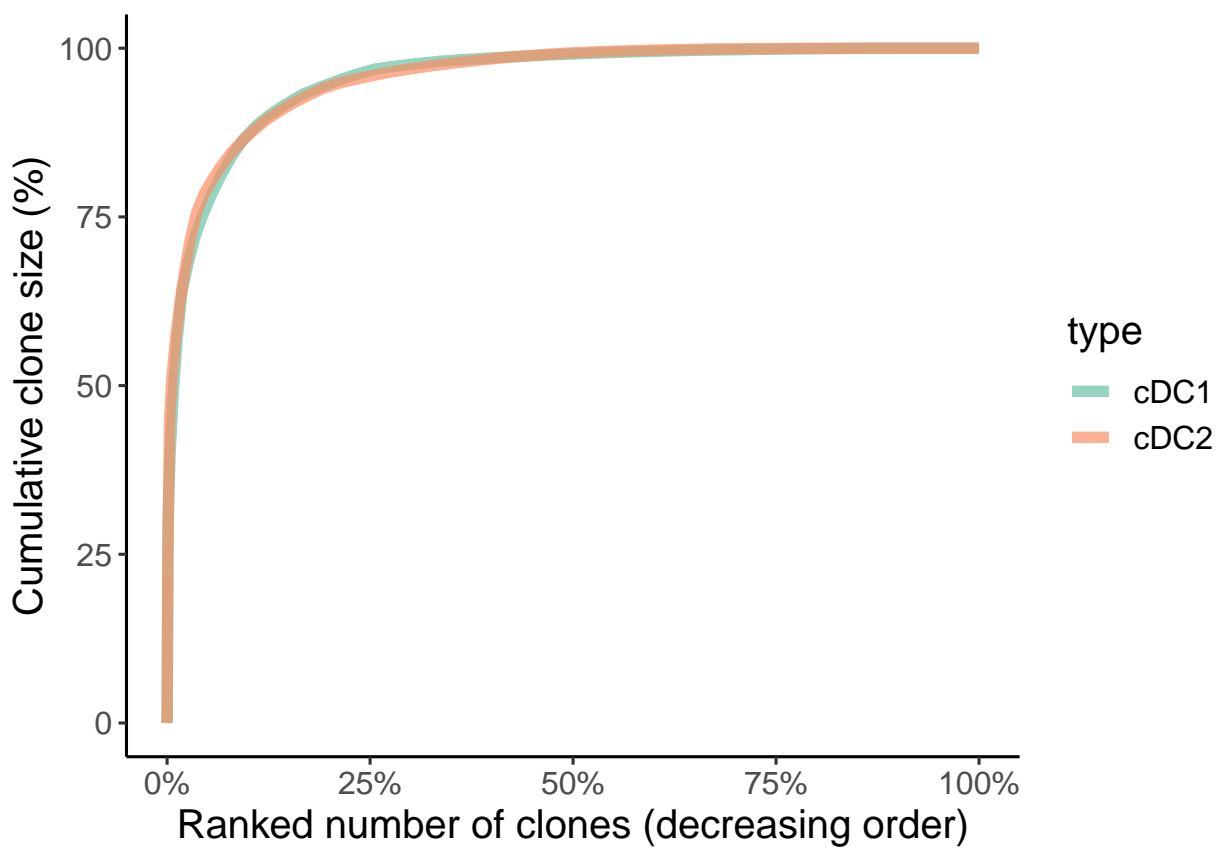


Figure 8: assessing the cumulative distribution of barcode abundances

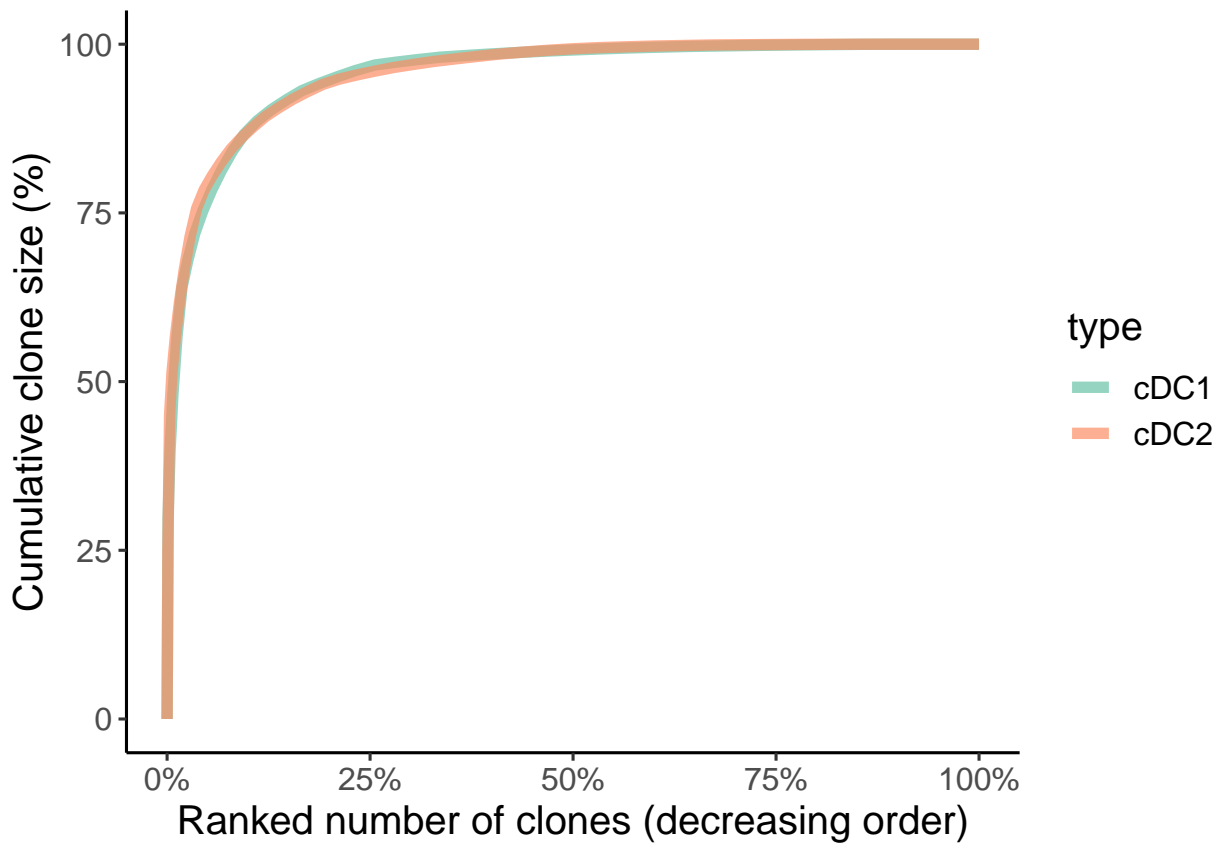


Figure 9: assessing the cumulative distribution of barcode abundances

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.

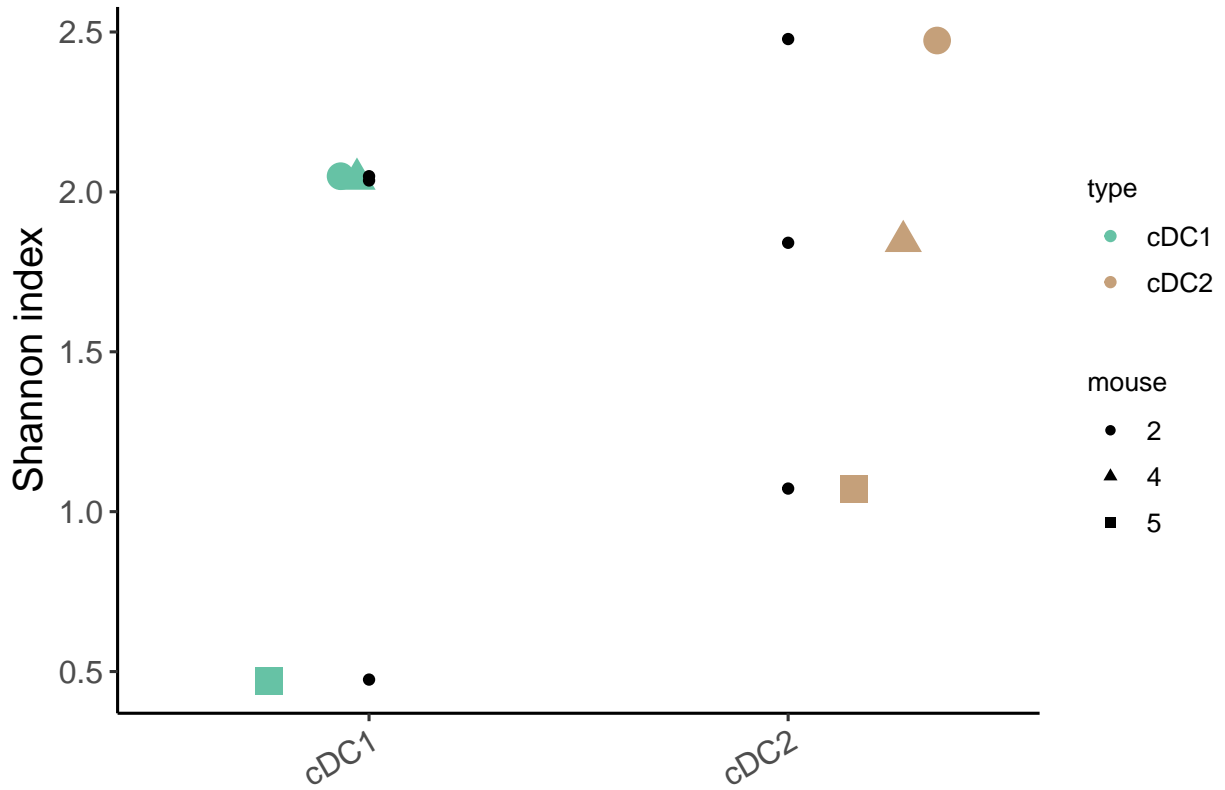


Figure 10: assessing barcode diversity

Barcode Sharing

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

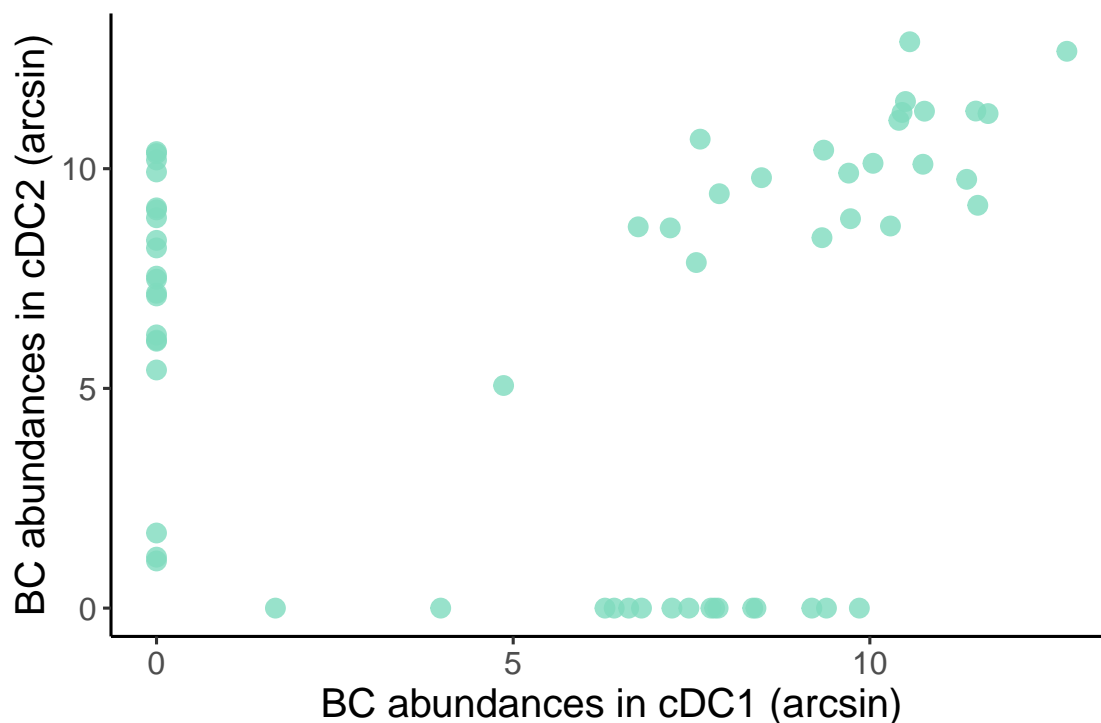


Figure 11: assessing barocde sharing

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. Here we use a fixed threshold value of 10%.

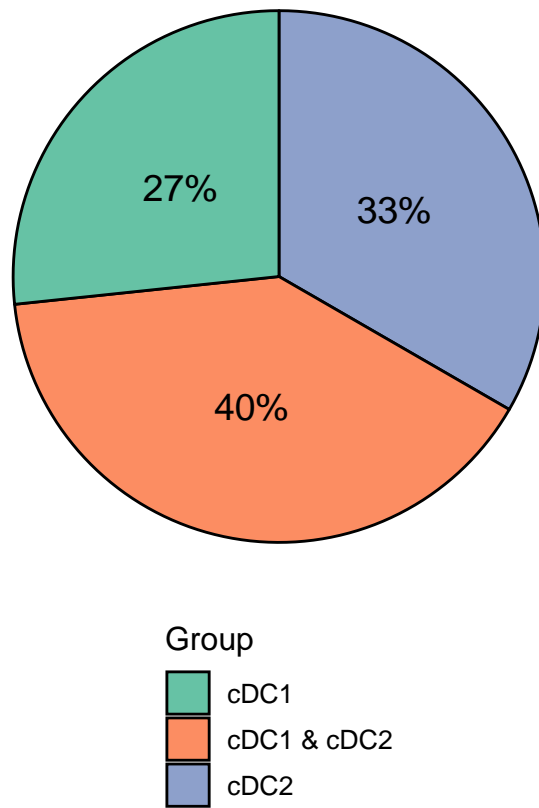


Figure 12: assessing barcode sharing using a pie chart

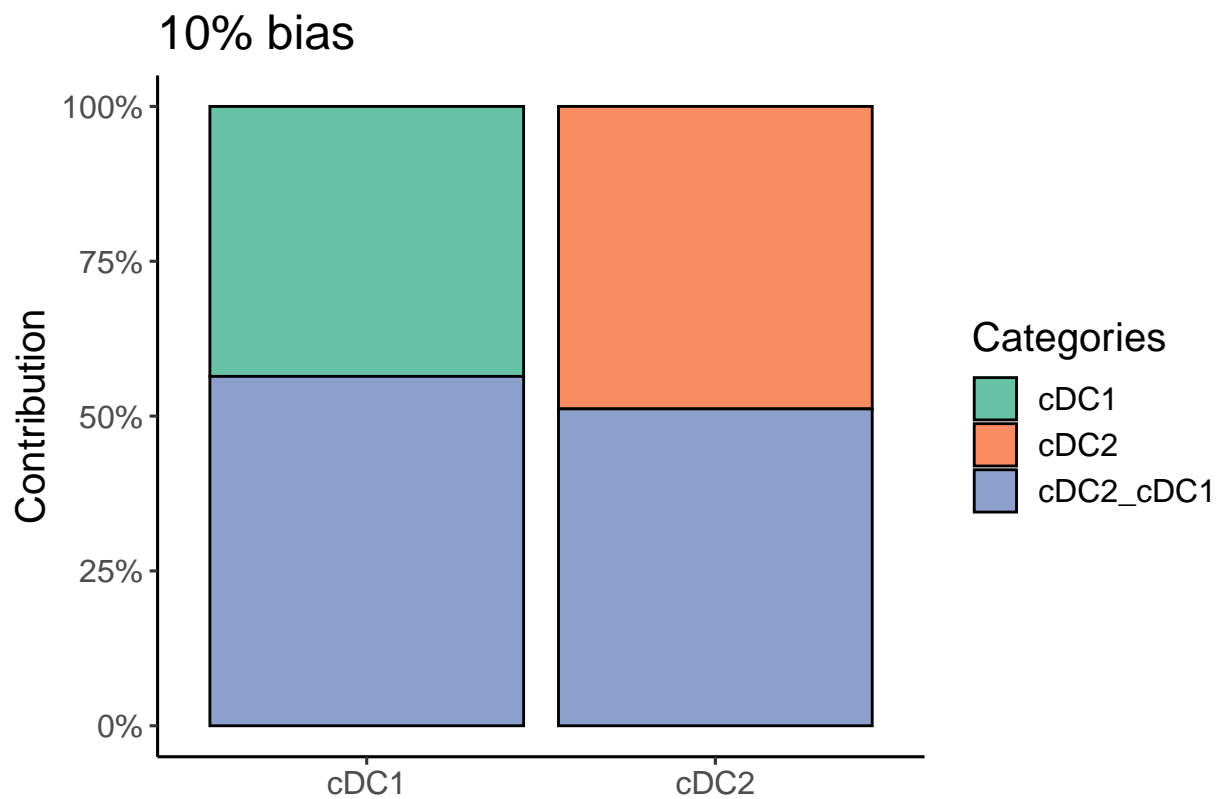


Figure 13: lineage classification analysis 10% bias

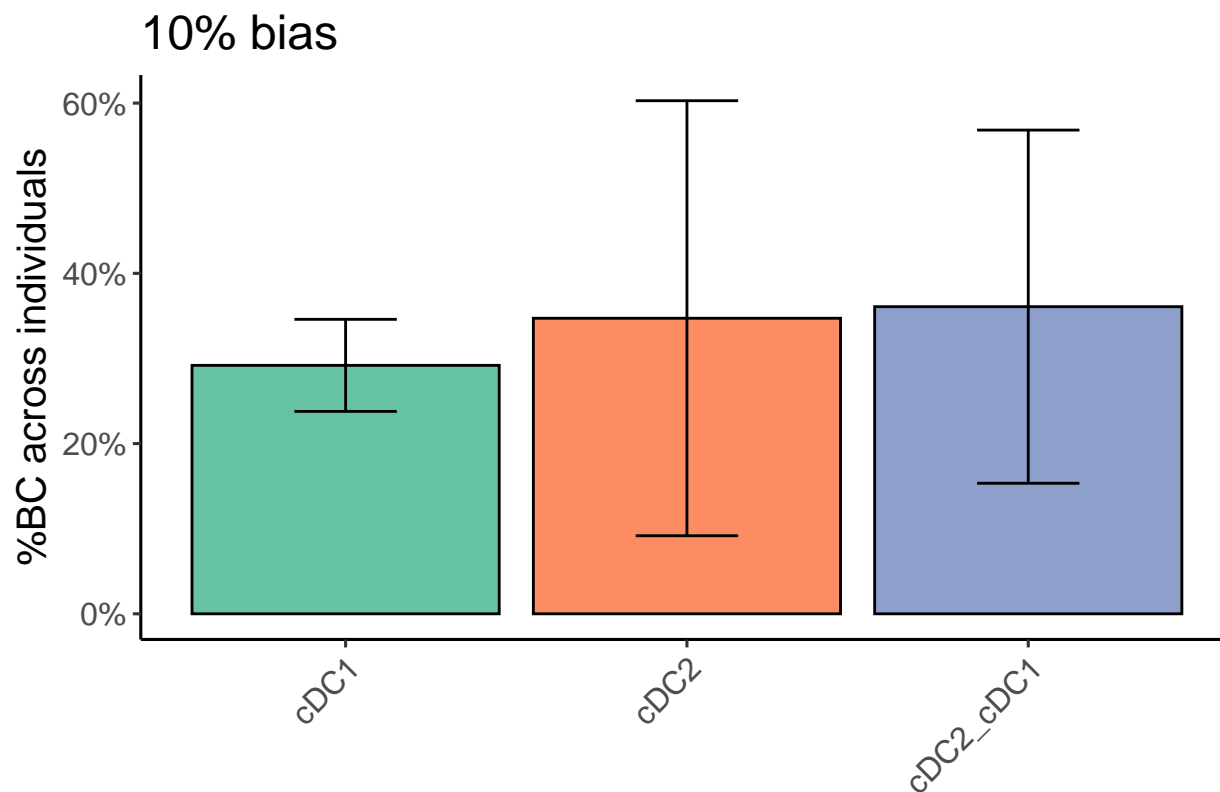


Figure 14: lineage classification analysis 10% bias