

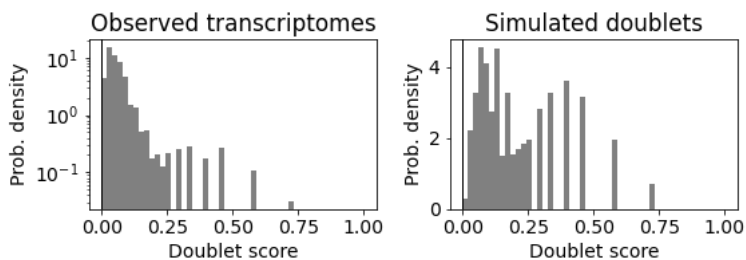
```
In [11]: %matplotlib inline
import scrublet as scr
import scipy.io
import matplotlib.pyplot as plt
import numpy as np
import os
import pandas as pd
plt.rcParams['font.family'] = 'sans-serif'
plt.rcParams['font.sans-serif'] = 'Arial'
plt.rc('font', size=14)
plt.rcParams['pdf.fonttype'] = 42
```

```
In [12]: input_dir = '/project/fsepru/PBMC5/Filtered/A/'
A_counts_matrix = scipy.io.mmread(input_dir + 'matrix.mtx').T.tocsc()
A_genes = np.array(scr.load_genes(input_dir + 'features.tsv', delimiter='\t', column=1))
print('Counts matrix shape: {} rows, {} columns'.format(A_counts_matrix.shape[0], A_counts_matrix.shape[1]))
print('Number of genes in gene list: {}'.format(len(A_genes)))
```

Counts matrix shape: 3141 rows, 16364 columns
Number of genes in gene list: 16364

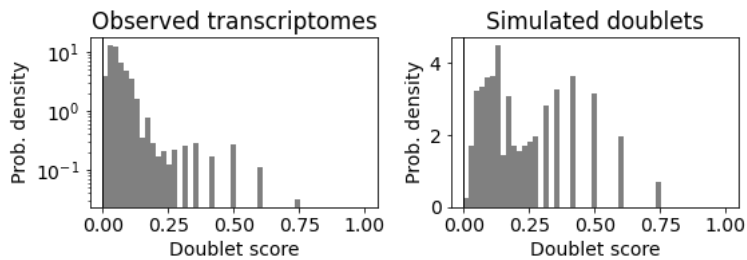
```
In [13]: A_scrub = scr.Scrublet(A_counts_matrix, expected_doublet_rate=0.06)
A_doublet_scores, A_predicted_doublets = A_scrub.scrub_doublets(min_counts=2, min_cells=3, min_gene_variability_pctl=85, n_print=10)
A_scrub.call_doublets(threshold=0.0)
A_scrub.plot_histogram();
```

Preprocessing...
Simulating doublets...
Embedding transcriptomes using PCA...
Calculating doublet scores...
Automatically set threshold at doublet score = 0.53
Detected doublet rate = 0.3%
Estimated detectable doublet fraction = 5.4%
Overall doublet rate:
Expected = 6.0%
Estimated = 5.3%
Elapsed time: 2.1 seconds
Detected doublet rate = 100.0%
Estimated detectable doublet fraction = 100.0%
Overall doublet rate:
Expected = 6.0%
Estimated = 100.0%



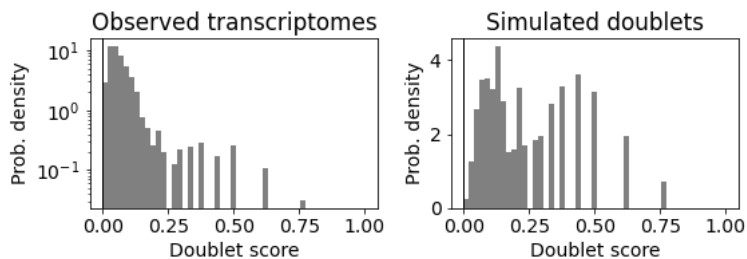
```
In [14]: A_scrub = scr.Scrublet(A_counts_matrix, expected_doublet_rate=0.065)
A_doublet_scores, A_predicted_doublets = A_scrub.scrub_doublets(min_counts=2, min_cells=3, min_gene_variability_pctl=85, n_print=10)
A_scrub.call_doublets(threshold=0.0)
A_scrub.plot_histogram();
```

Preprocessing...
Simulating doublets...
Embedding transcriptomes using PCA...
Calculating doublet scores...
Automatically set threshold at doublet score = 0.55
Detected doublet rate = 0.3%
Estimated detectable doublet fraction = 5.4%
Overall doublet rate:
Expected = 6.5%
Estimated = 5.3%
Elapsed time: 2.3 seconds
Detected doublet rate = 100.0%
Estimated detectable doublet fraction = 100.0%
Overall doublet rate:
Expected = 6.5%
Estimated = 100.0%



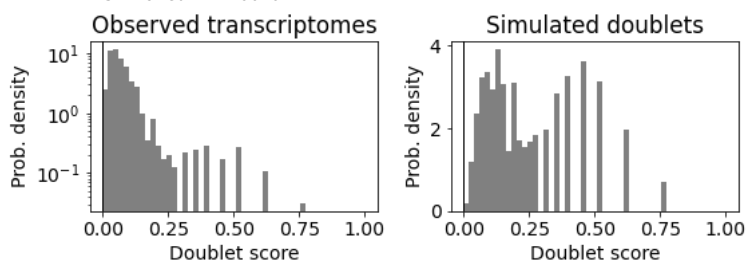
```
In [15]: A_scrub = scr.Scrublet(A_counts_matrix, expected_doublet_rate=0.07)
A_doublet_scores, A_predicted_doublets = A_scrub.scrub_doublets(min_counts=2, min_cells=3, min_gene_variability_pctl=85, n_prin_
A_scrub.call_doublets(threshold=0.0)
A_scrub.plot_histogram();
```

```
Preprocessing...
Simulating doublets...
Embedding transcriptomes using PCA...
Calculating doublet scores...
Automatically set threshold at doublet score = 0.58
Detected doublet rate = 0.3%
Estimated detectable doublet fraction = 5.4%
Overall doublet rate:
  Expected = 7.0%
  Estimated = 5.3%
Elapsed time: 2.1 seconds
Detected doublet rate = 100.0%
Estimated detectable doublet fraction = 100.0%
Overall doublet rate:
  Expected = 7.0%
  Estimated = 100.0%
```



```
In [16]: A_scrub = scr.Scrublet(A_counts_matrix, expected_doublet_rate=0.075)
A_doublet_scores, A_predicted_doublets = A_scrub.scrub_doublets(min_counts=2, min_cells=3, min_gene_variability_pctl=85, n_prin_
A_scrub.call_doublets(threshold=0.0)
A_scrub.plot_histogram();
```

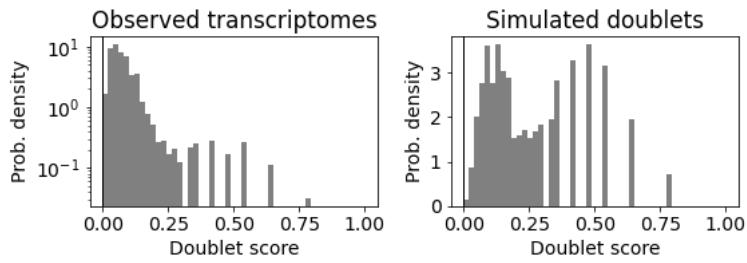
```
Preprocessing...
Simulating doublets...
Embedding transcriptomes using PCA...
Calculating doublet scores...
Automatically set threshold at doublet score = 0.33
Detected doublet rate = 2.3%
Estimated detectable doublet fraction = 31.7%
Overall doublet rate:
  Expected = 7.5%
  Estimated = 7.1%
Elapsed time: 2.0 seconds
Detected doublet rate = 100.0%
Estimated detectable doublet fraction = 100.0%
Overall doublet rate:
  Expected = 7.5%
  Estimated = 100.0%
```



```
In [17]: A_scrub = scr.Scrublet(A_counts_matrix, expected_doublet_rate=0.08)
A_doublet_scores, A_predicted_doublets = A_scrub.scrub_doublets(min_counts=2, min_cells=3, min_gene_variability_pctl=85, n_prin_
A_scrub.call_doublets(threshold=0.0)
A_scrub.plot_histogram();
```

```
Preprocessing...
Simulating doublets...
```

```
Embedding transcriptomes using PCA...
Calculating doublet scores...
Automatically set threshold at doublet score = 0.33
Detected doublet rate = 2.7%
Estimated detectable doublet fraction = 35.7%
Overall doublet rate:
  Expected  = 8.0%
  Estimated  = 7.6%
Elapsed time: 2.1 seconds
Detected doublet rate = 100.0%
Estimated detectable doublet fraction = 100.0%
Overall doublet rate:
  Expected  = 8.0%
  Estimated  = 100.0%
```



In []: