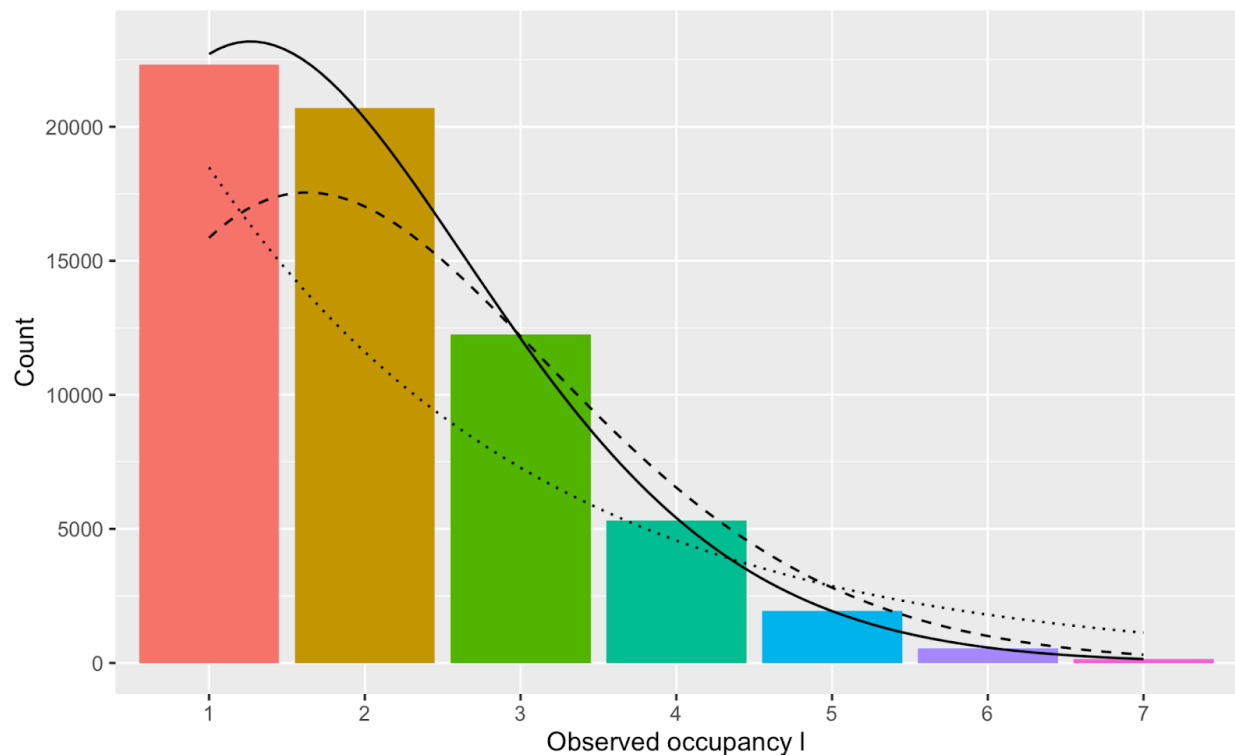


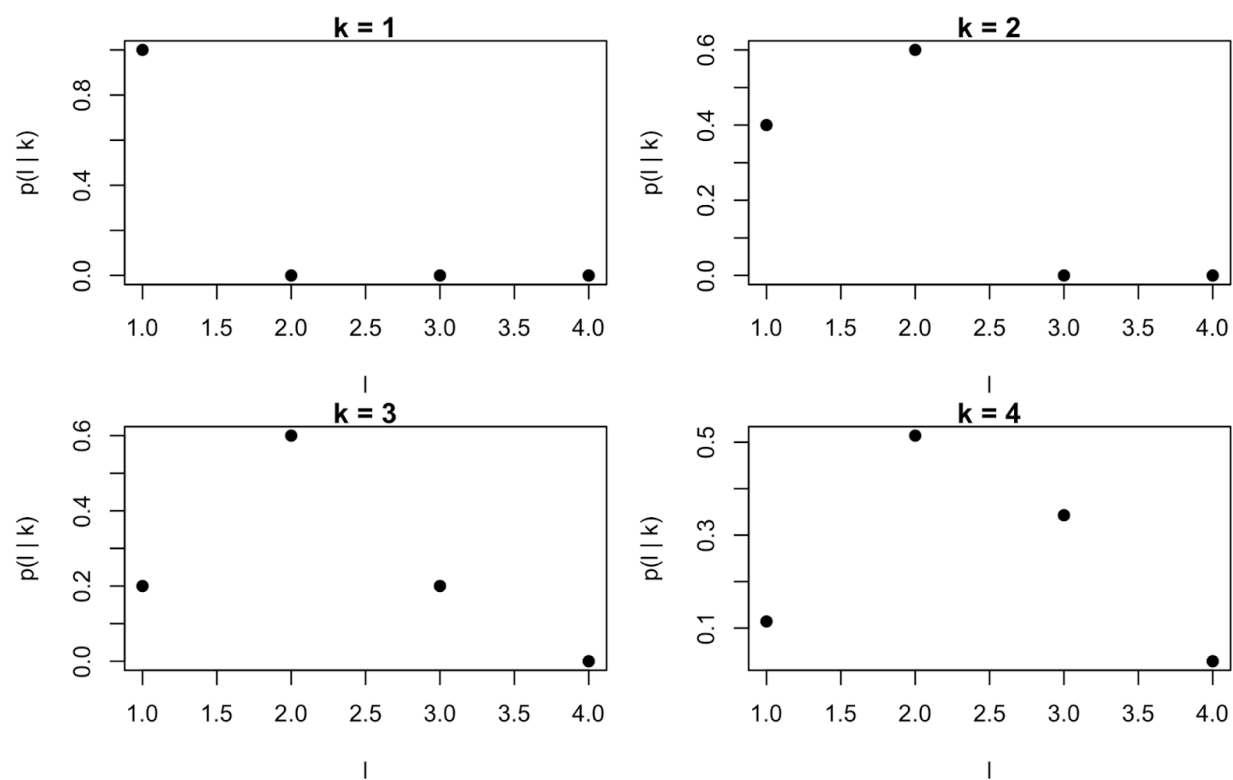
## Supplementary materials

### More cells, more doublets in highly multiplexed single-cell data

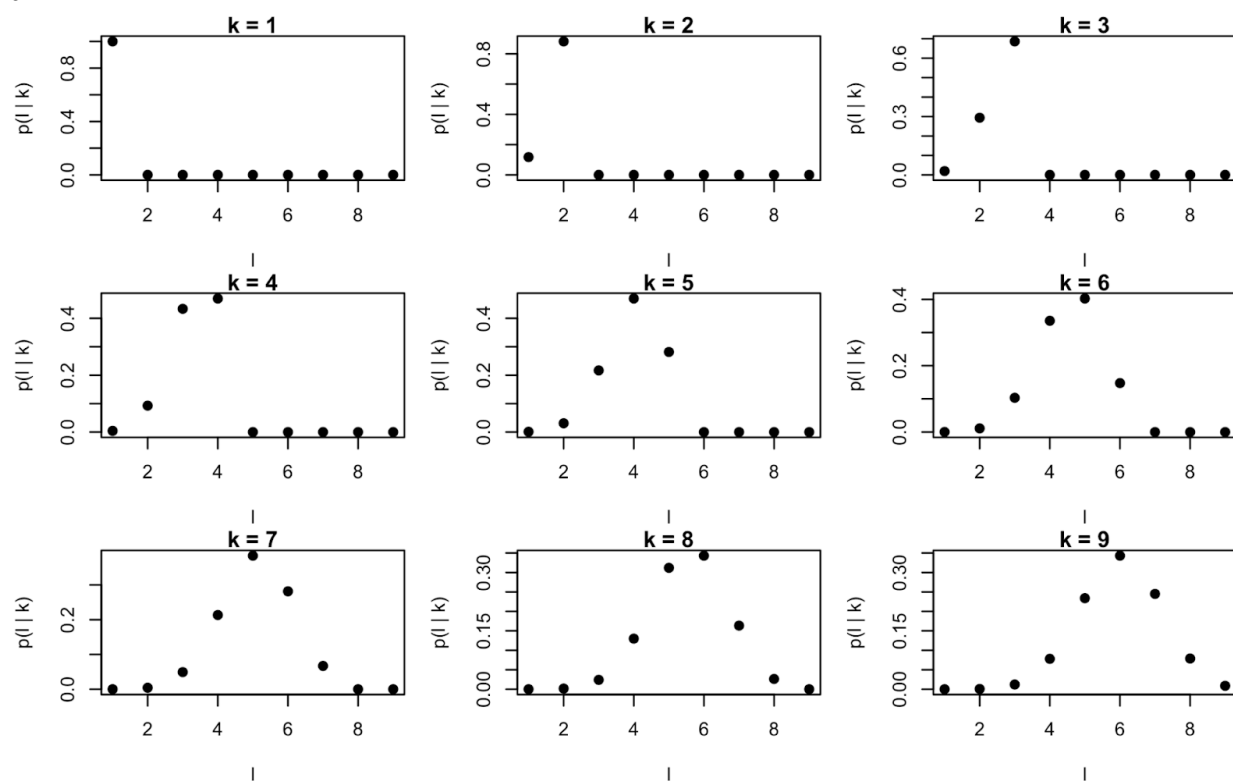


Supplementary Figure 1: Histogram showing the observed occupancy of droplets in the PBMC data set. Curves indicate fits to the data, with a Poisson fit (dashed line), zero-truncated Poisson fit (solid line) and exponential fit (dotted line).

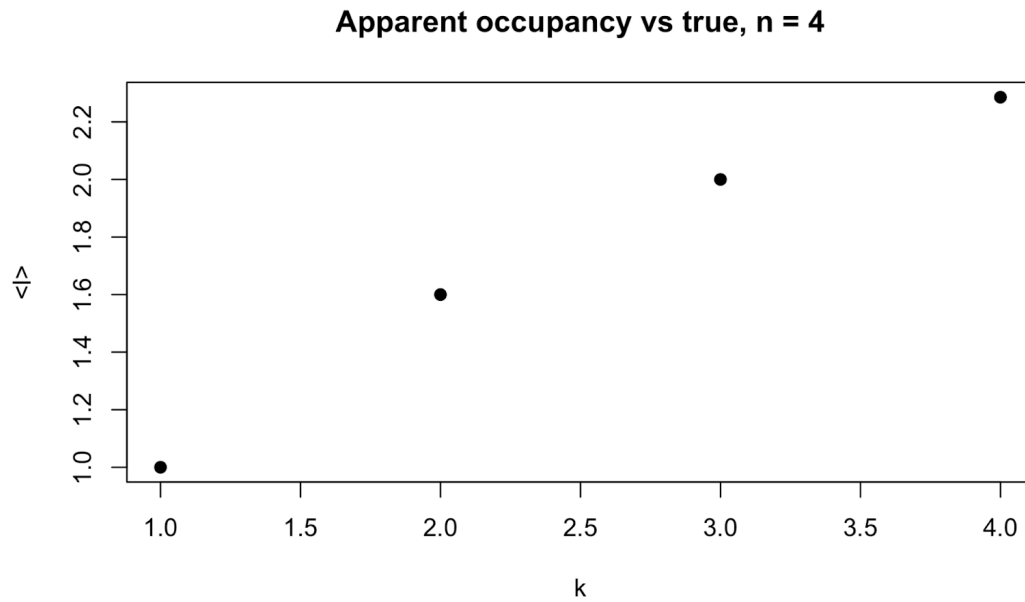
a



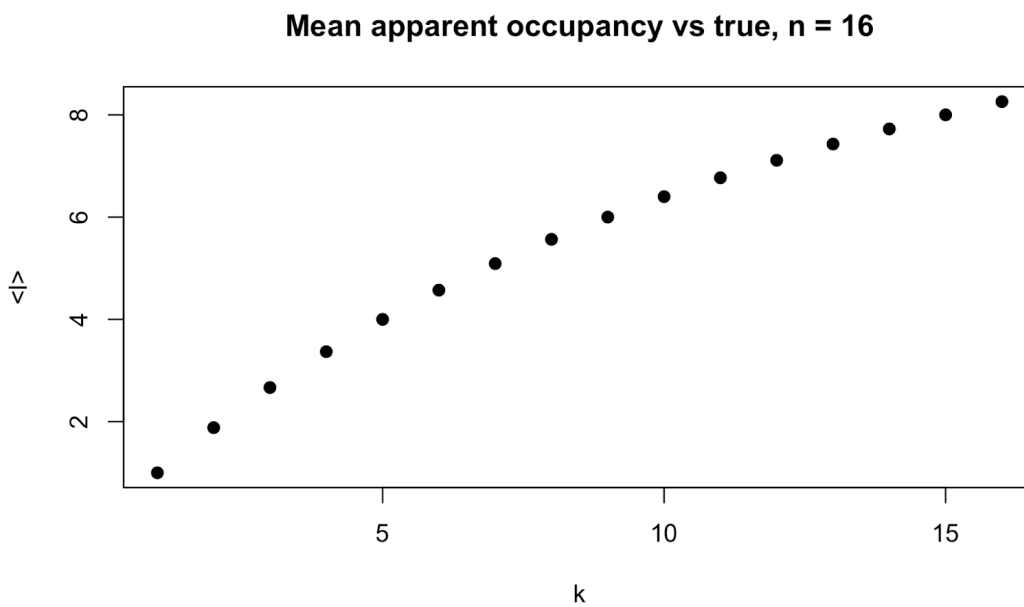
b



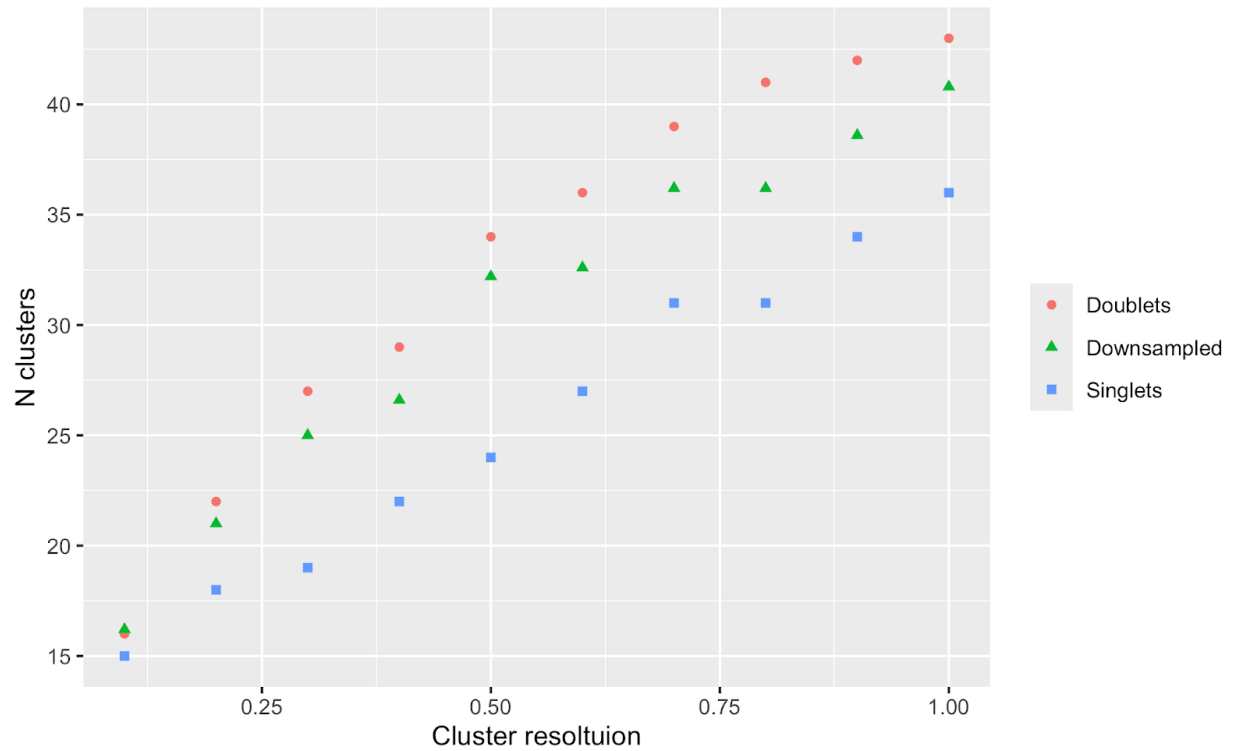
c



d



Supplementary Figure 2: Conditional probability of observing occupancy  $l$  given true occupancy  $k$  for a)  $n = 4$  and b)  $n = 16$ . Mean observed occupancy  $\langle l \rangle$  vs  $k$  for c)  $n = 4$  and d)  $n = 16$



Supplementary Figure 3: Number of clusters vs clustering resolution for: PBMC data set with doublets included (red circles), PBMC data set with doublets called by scDbIFinder removed (blue squares), Downsampled data set with same number of cells as the singlet-only dataset, but same doublet fraction as full data set (green triangles). The downsampled values shown are the mean values from 10 downsampling experiments.