Cell Reports

Volume 29, Issue 6, 5 November 2019, Pages 1718-1727.e8

November 2019

Nathan Salomonis Group
Cincinnati Children's hospital, Ohio

DoubletDecon

Cell Systems

April 2019

Zev Gartner Group
University of California, San Francisco

DoubletFinder

OXFORD ACADEMIC

Bioinformatics

September 2019

Dennis Kostka Group
University of Pittsburgh

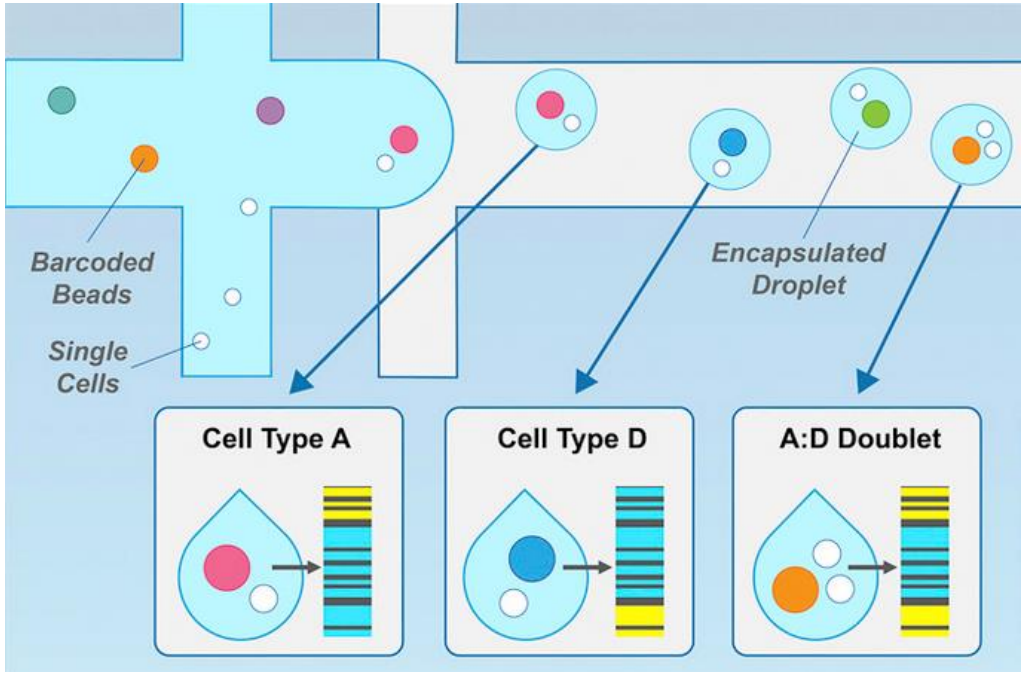
SCDS - **s**ingle **c**ell **d**oublet **s**coring

Detection of doublets in scRNAseq data

Veronika Niederlová

LAB OF ADAPTIVE IMMUNITY

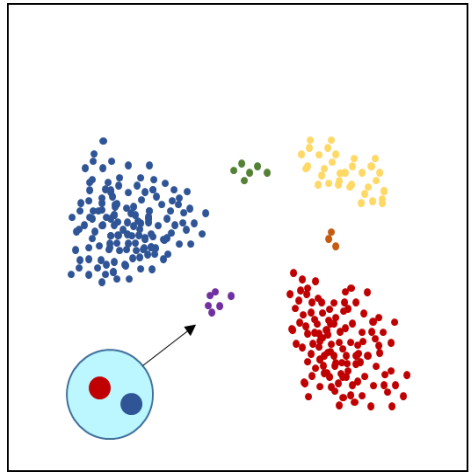
What is a doublet in scRNAseq data?



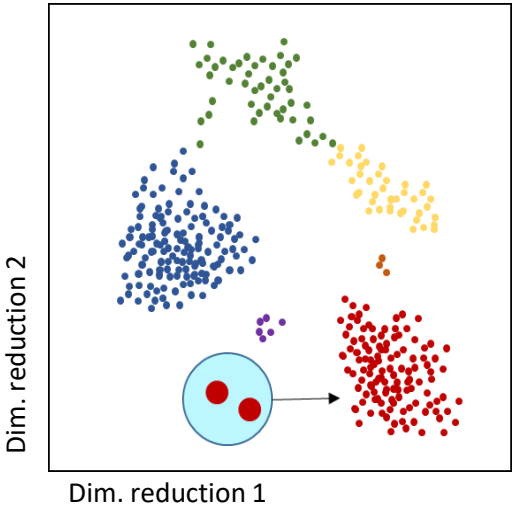
DePasquale et al. *Cell reports* (2019)

Real doublets

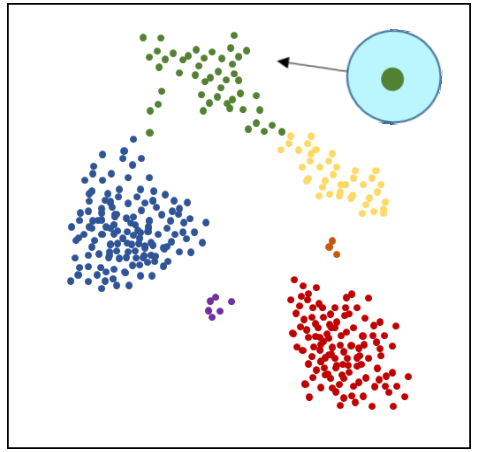
Heterotypic doublets

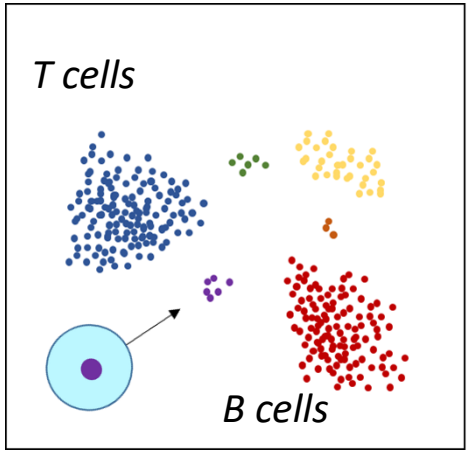


Homotypic doublets



Doublet-like cells



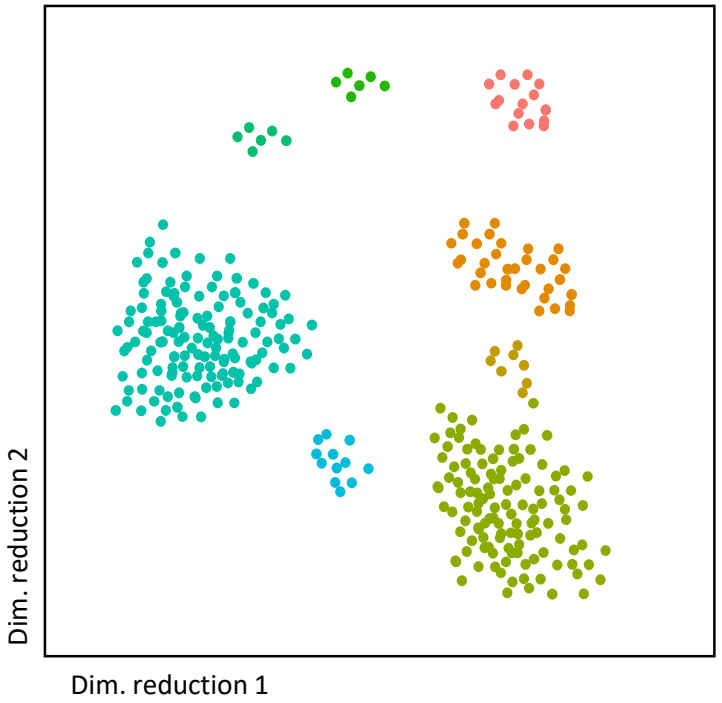


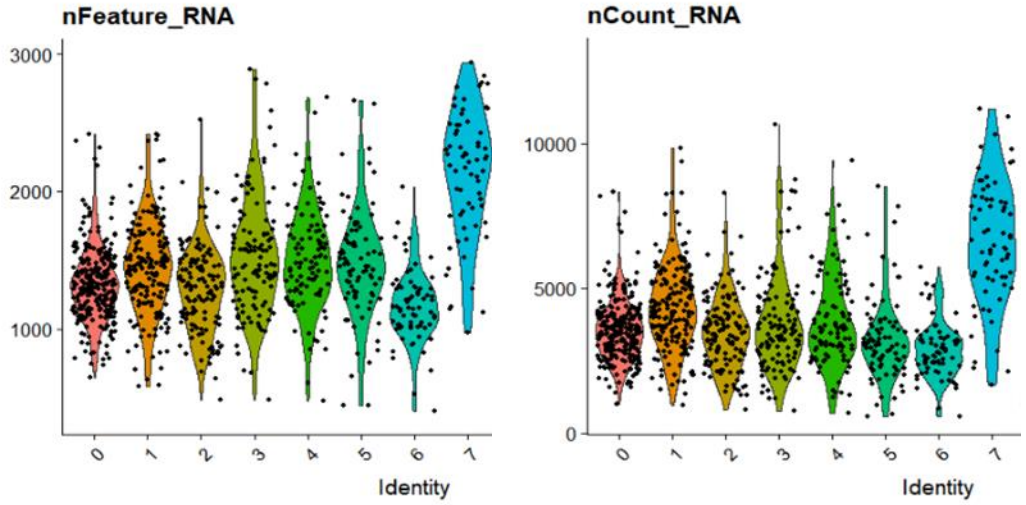
Dual Experessor cells

Revealing doublets…

Multiplet Rate (%)	# of Cells Loaded	# of Cells Recovered
~0.4%	~800	~500
~0.8%	~1,600	~1,000
~1.6%	~3,200	~2,000
~2.3%	~4,800	~3,000
~3.1%	~6,400	~4,000
~3.9%	~8,000	~5,000
~4.6%	~9,600	~6,000
~5.4%	~11,200	~7,000
~6.1%	~12,800	~8,000
~6.9%	~14,400	~9,000
~7.6%	~16,000	~10,000

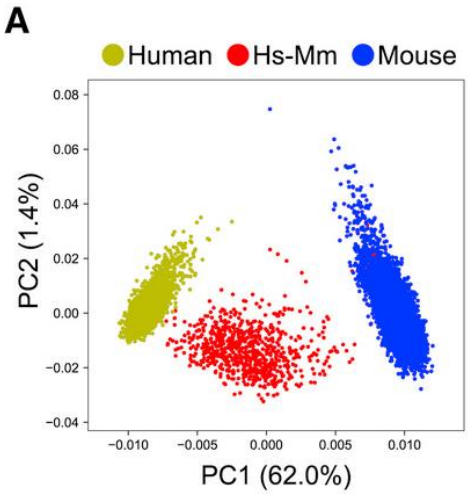
10x Genomics



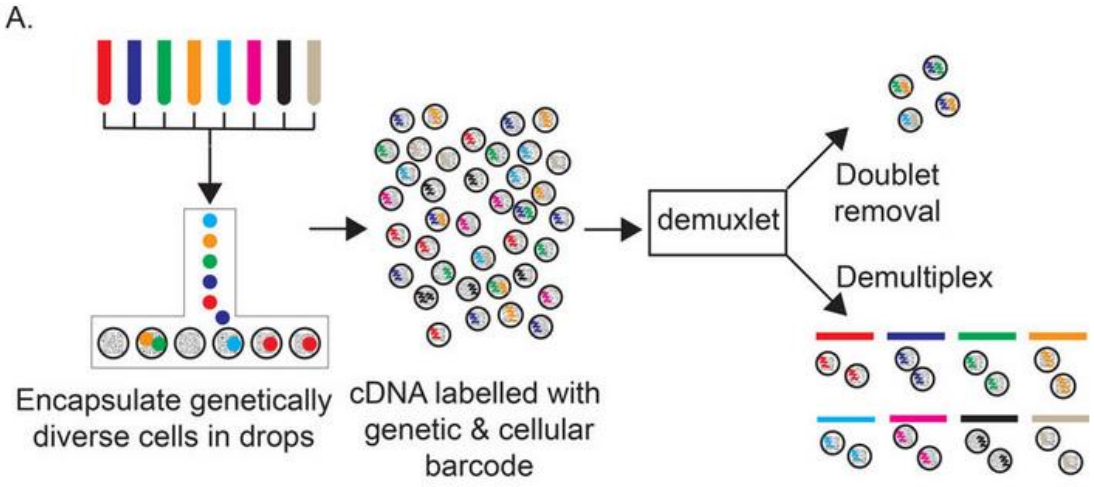


Demuxlet and cell-hashing

Mixing genetically different cells into one sample

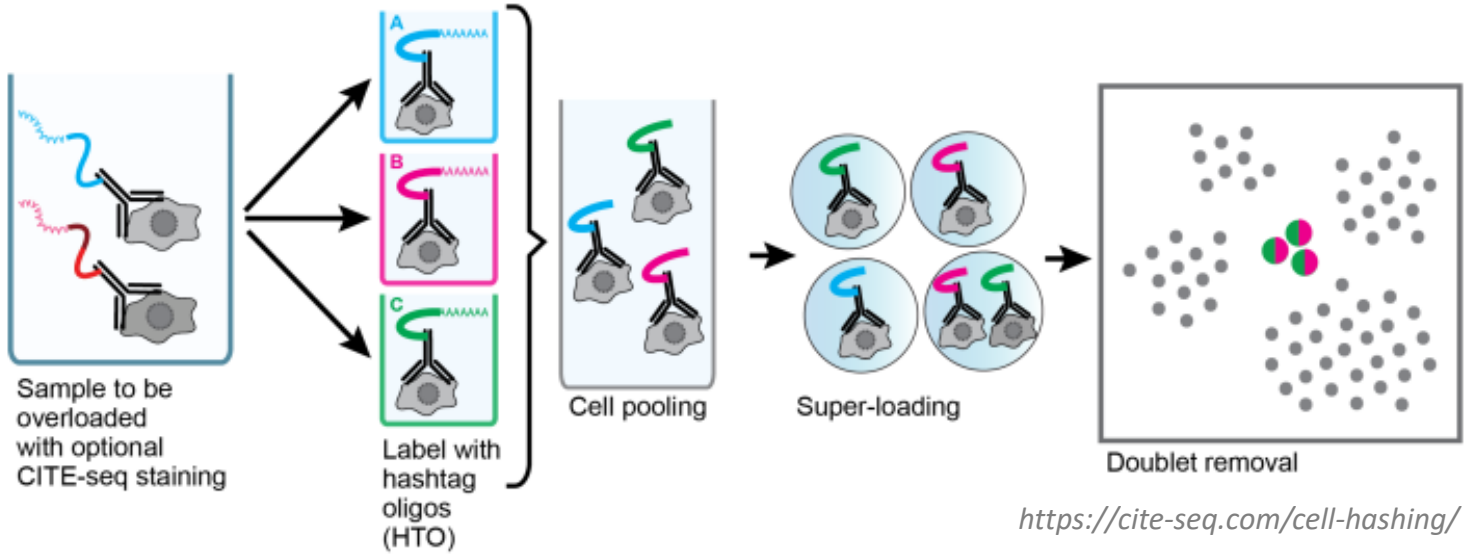


DePasquale et al. *Cell reports* (2019)



Kang et al. *Nature Biotechnology* (2018)

Cell hashing



https://cite-seq.com/cell-hashing/

Three different algorithms

Cell Reports
Volume 29, Issue 6, 5 November 2019, Pages 1718-1727.e8

November 2019

Nathan Salomonis Group
Cincinnati Children's hospital, Ohio

DoubletDecon

Initial user-defined clustering

Deconvolution

Artificial Doublets

Similarity (Pearson) of cells with synthetic doublets

Rescue step

Cell Systems

April 2019

Zev Gartner Group
University of California, San Francisco

DoubletFinder

Artificial Doublets

pANN – neighborhood of synthetic doublets

Thresholding pANN based on expected doublet rate

OXFORD ACADEMIC
Bioinformatics

September 2019

Dennis Kostka Group
University of Pittsburgh

SCDS - **s**ingle **c**ell **d**oublet **s**coring

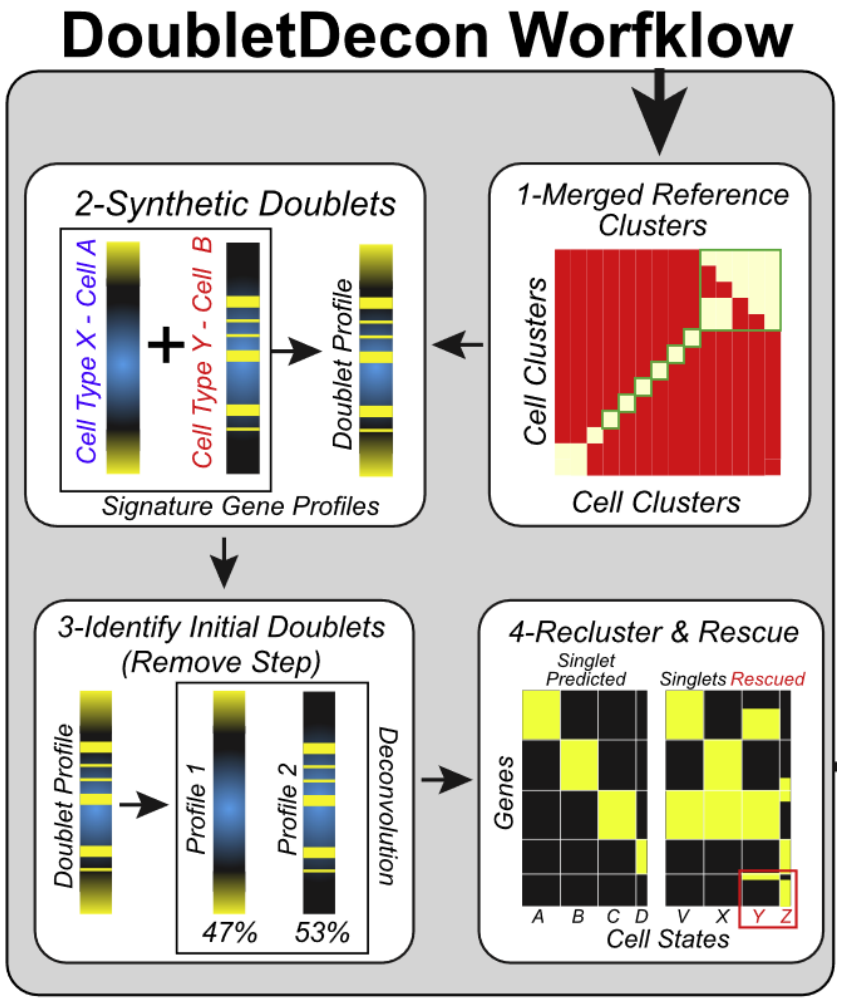
Co-expression based doublet scoring (no artificial doublets)

Binary classification based doublet scoring
Artificial doublets - adding random pairs of columns to the GEX matrix

Ordering

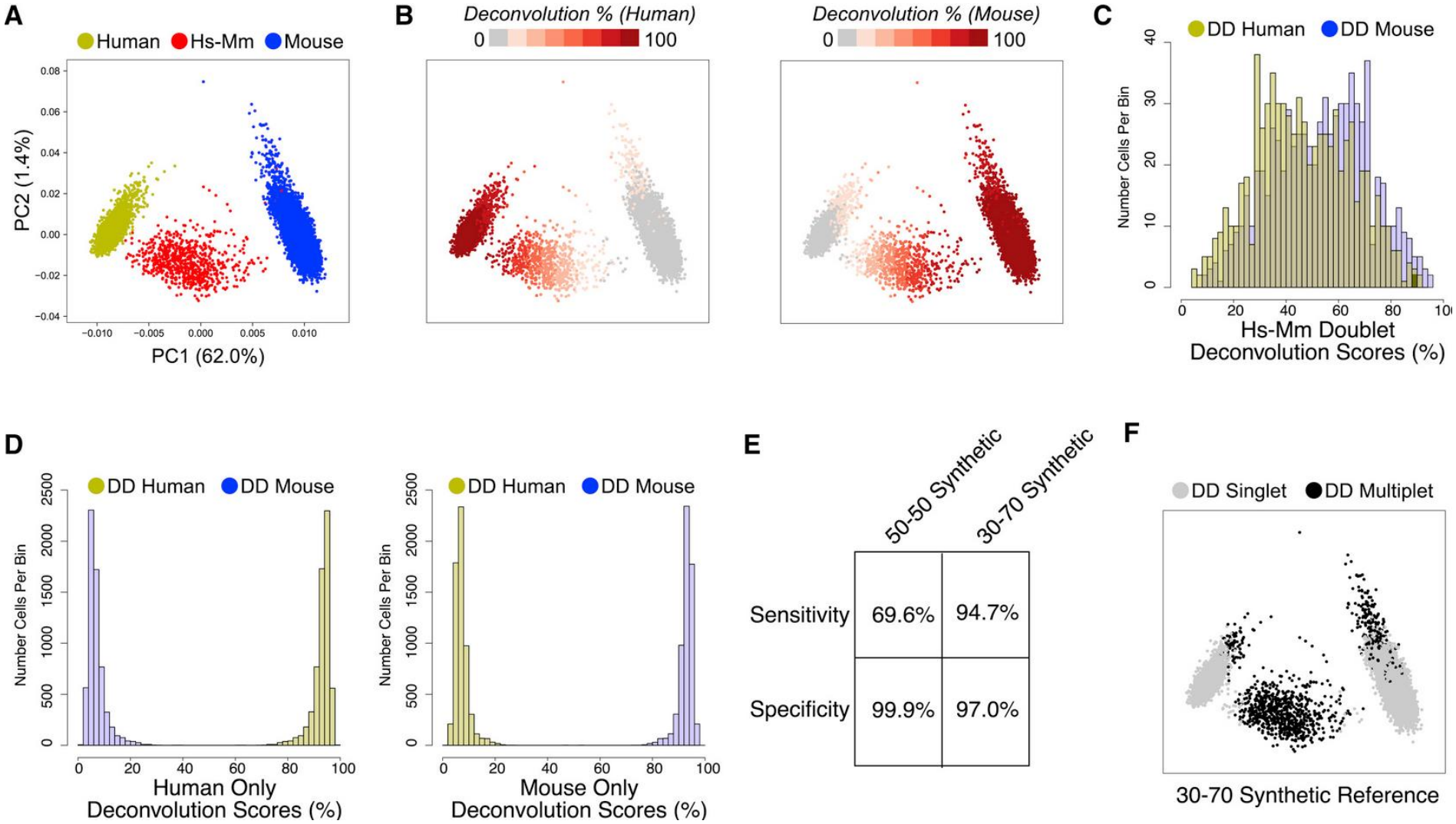
Selection of doublets

DoubletDecon overview

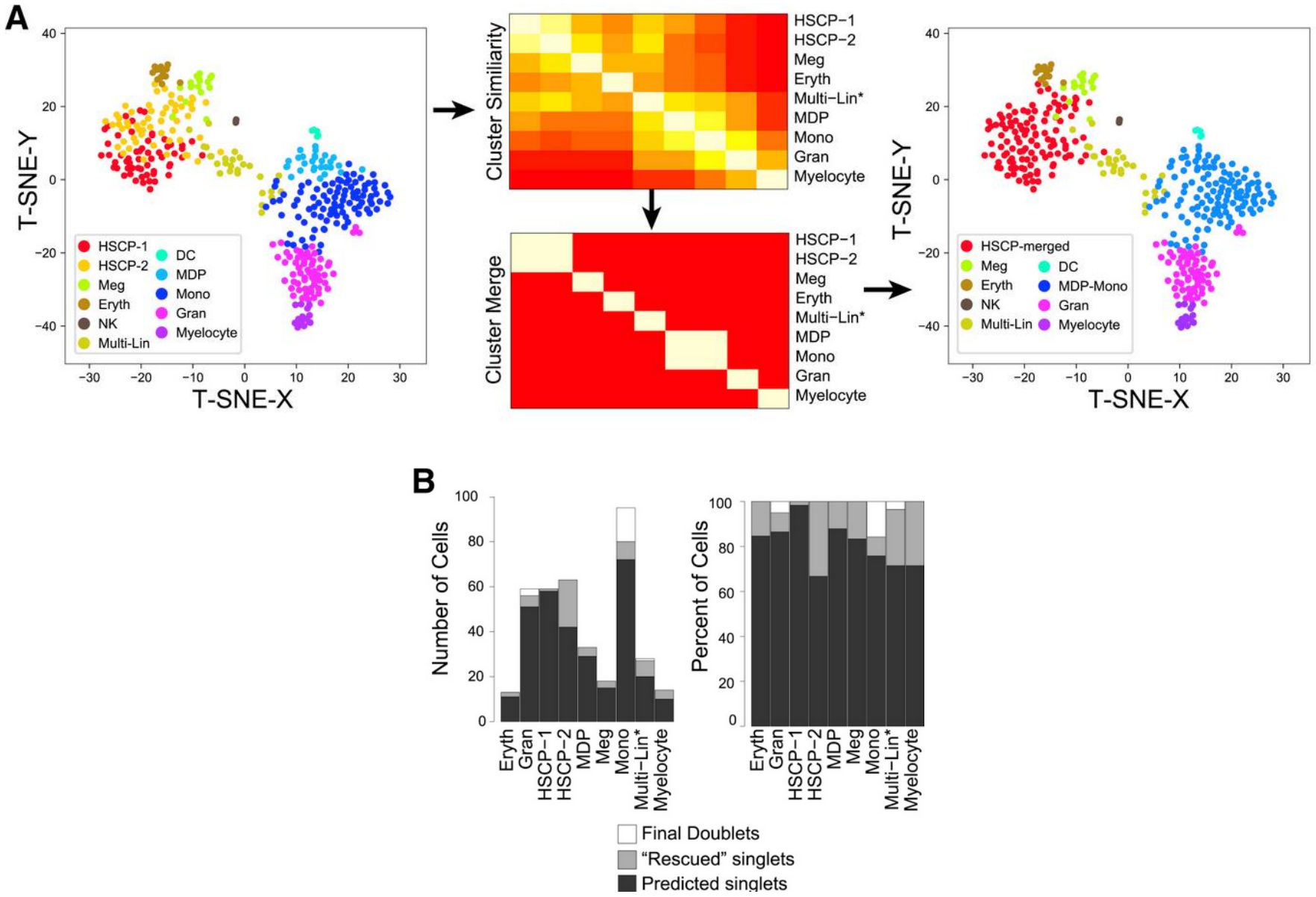


1) Seurat input: normalized GEX matrix, top marker genes, cluster identities
2) (optional cell-cycle gene removal)
3) Calculate centroids (medoids) fpor each cluster
4) Join transcriptionally similar cell clusters
5) Create synthetic doublets with either a 50%/50% contribution from two individual cells or a weighted average for 30%/70% and 70%/30% synthetic doublets
6) Deconvolution with the R package ‘DeconRNASeq’
7) Deconvolution Cell Profile (DCP)
8) Putative doublets based on DCP
9) Re-clustering, rescue

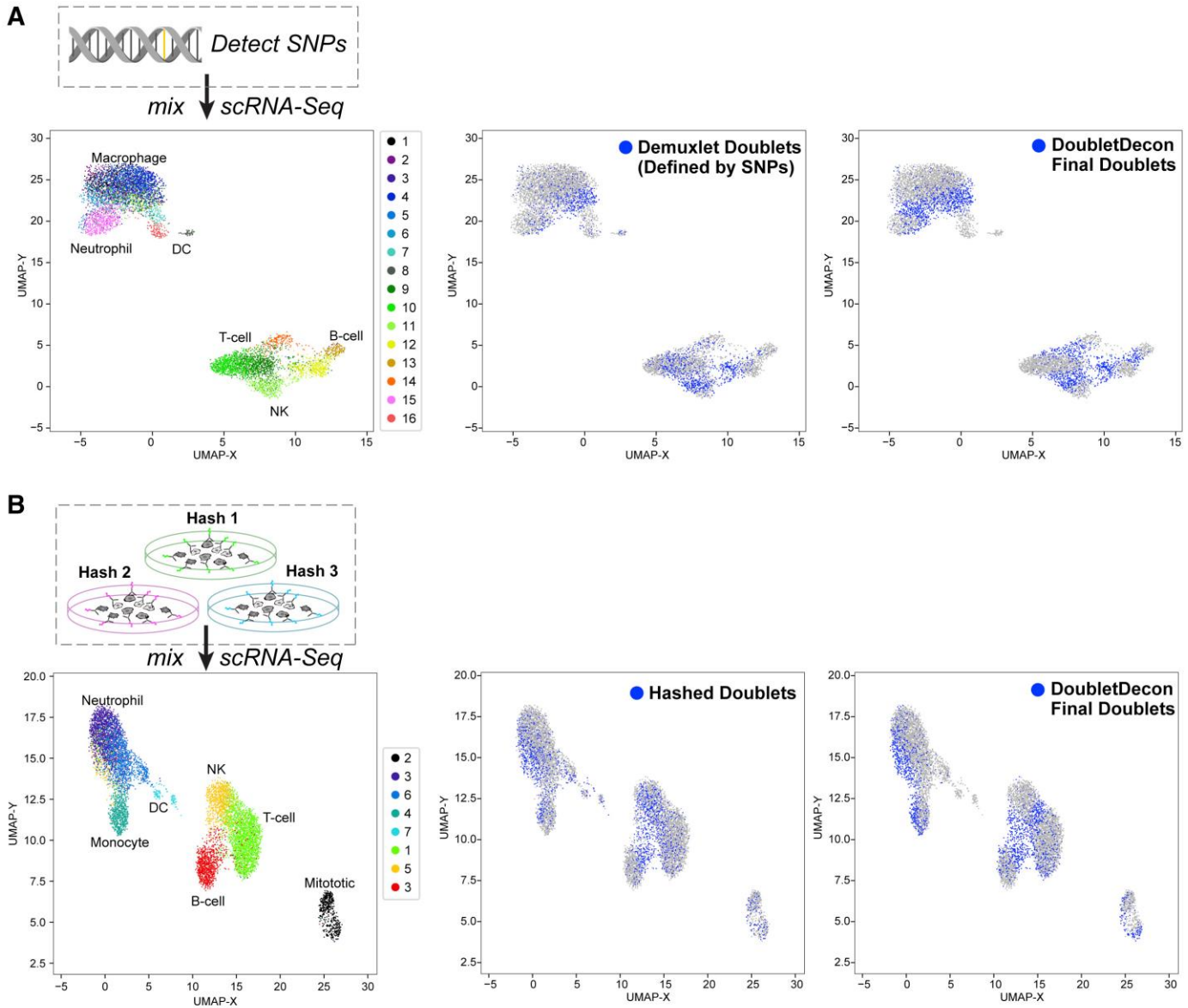
DoubletDecon performance in multiplexed dataset

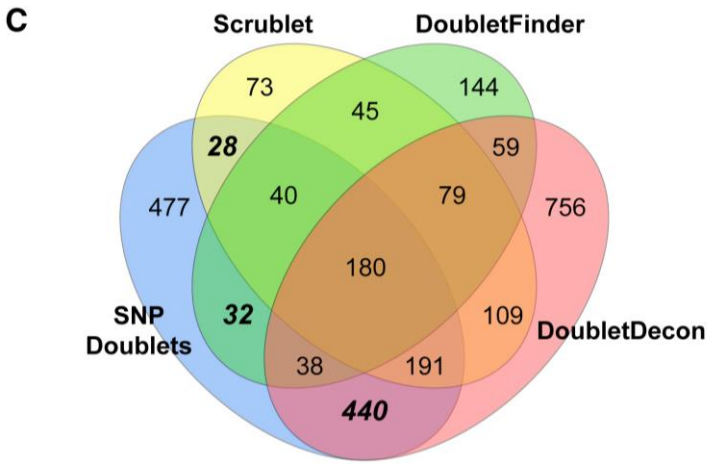


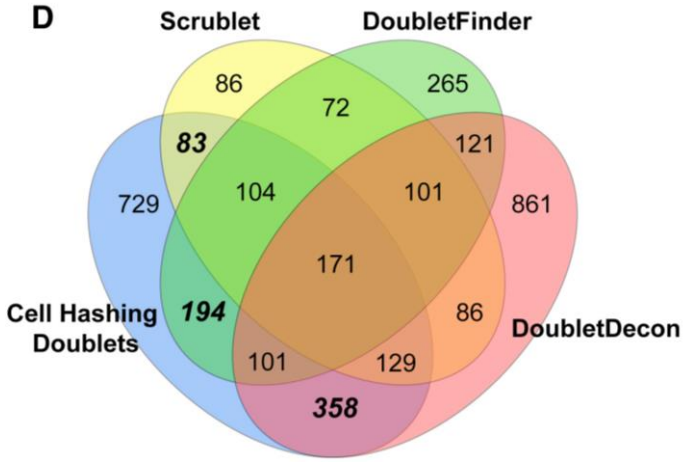
Doublet detection in a doublet-free dataset



DoubletDecon performance in Demuxlet and Cell hashing datasets







Performance of different tools

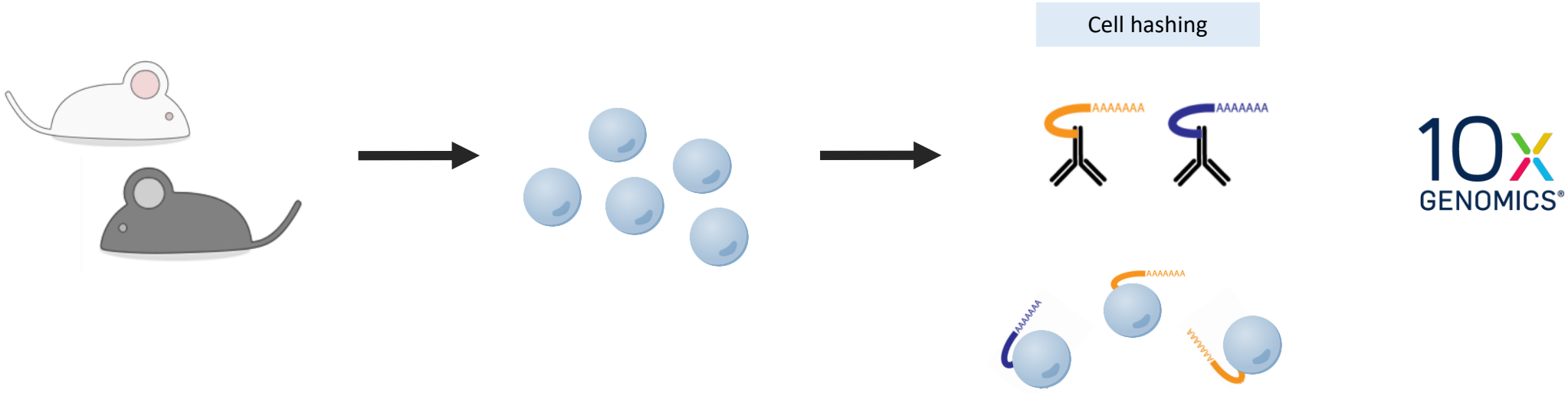
Sensitivity = True positives / True positives + False negatives
Specificity = True negatives / True negatives + False positives

Table 1. Relative Performance of Doublet Removal Tools

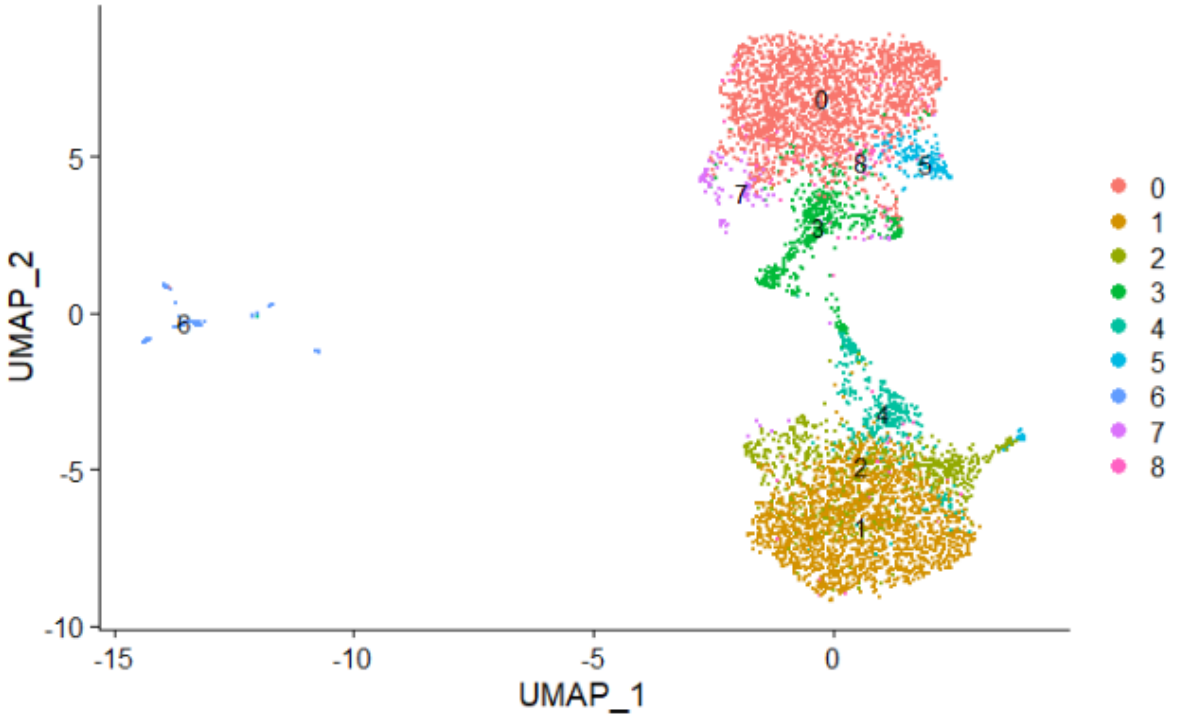
Dataset	Tool	Parameters	Sensitivity, % (SD, %)	Specificity, % (SD, %)
Demuxlet n = 6,525	DoubletDecon	rhop = 1.2, min_uniq = 4	56.90 (2.1)	81.20 (1.2)
Demuxlet n = 6,525	DoubletFinder	10 PCs, pN = 0.25, pK = auto, rate = 12.5% - homotypic adj.	21.31 (2.8)	93.87 (0.8)
Demuxlet n = 6,525	Scrublet	25 PCs, top 25% HVGs, rate = 12.5%, *Z* scores, theta = auto	31.00 (2.3)	93.94 (1.0)
Cell Hashing n = 8,402	DoubletDecon	rhop = 1.2, min_uniq = 4	38.10 (1.2)	82.40 (0.6)
Cell Hashing n = 8,402	DoubletFinder	10 PCs, pN = 0.25, pK = auto, rate = 20% - homotypic adj.	32.23 (3.8)	89.42 (1.1)
Cell Hashing n = 8,402	Scrublet	30 PCs, top 15% HVGs, rate = 20%, *Z* scores, theta = 0.37	26.16 (0.6)	94.70 (0.2)

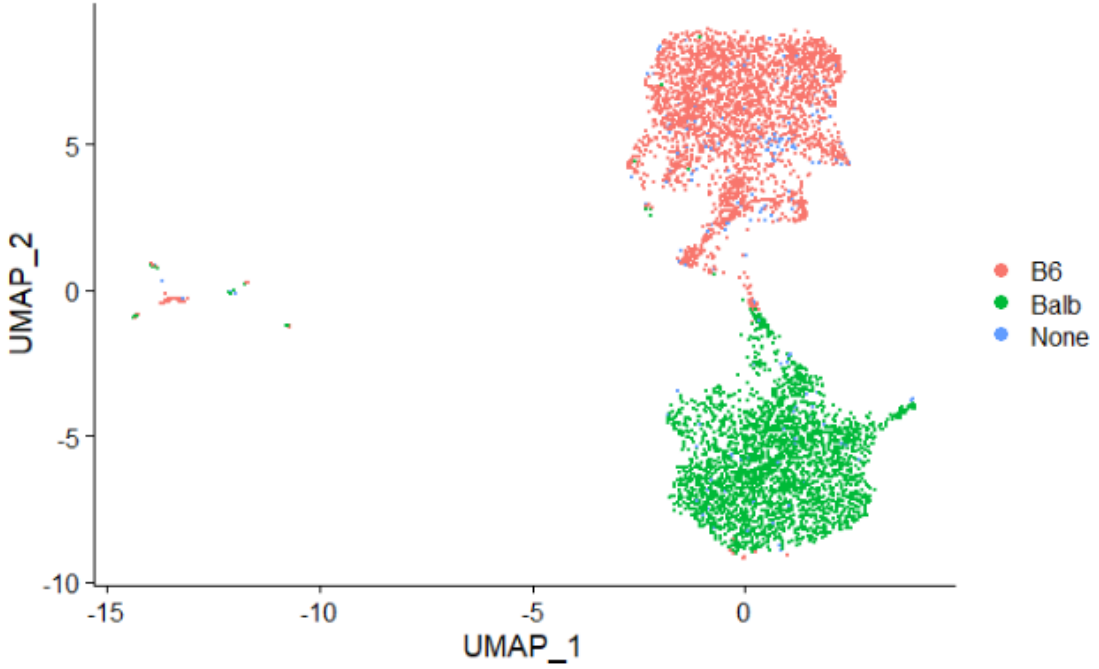
Our dataset:

- 10x genomics 5‘GEX + VDJ + Cell Hashing
- 8000 cells loaded, 5600 cells recovered
- Estimated multiplet rate 4% (320 cells)
- BALB/c and BL6 mice (Demuxlet-like setup)
- Cell hashing



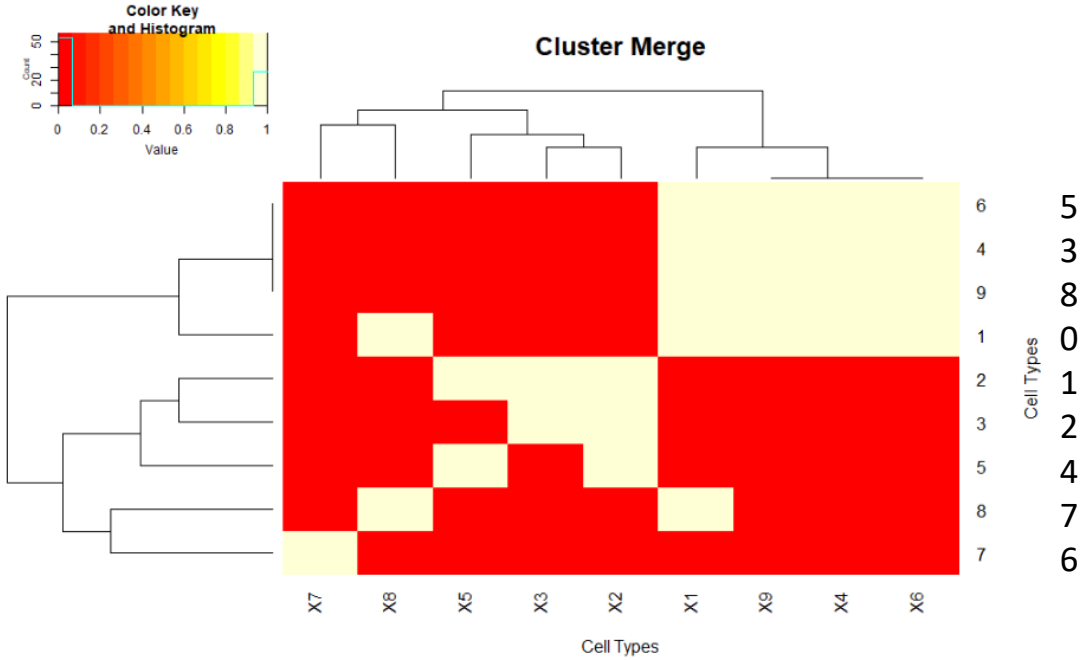
Initial clustering and doublets

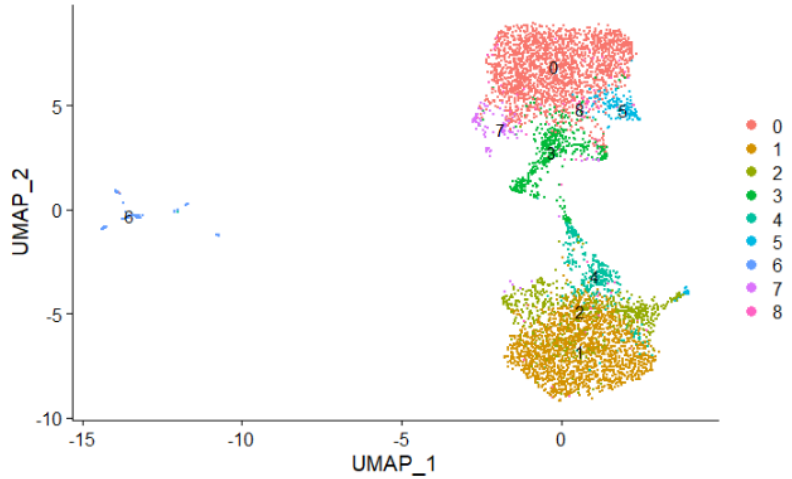


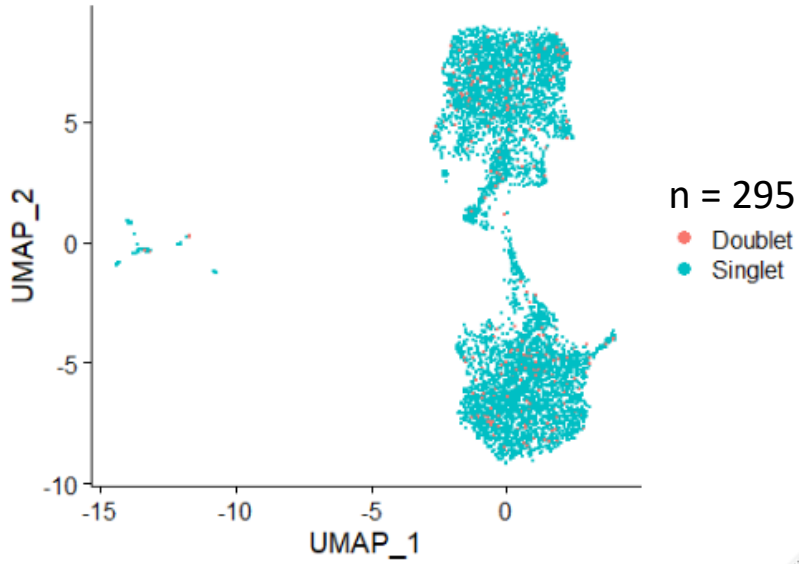



```
Orig_Doublet Orig_Singlet
         199          5497
```

Cluster merge, deconvolution

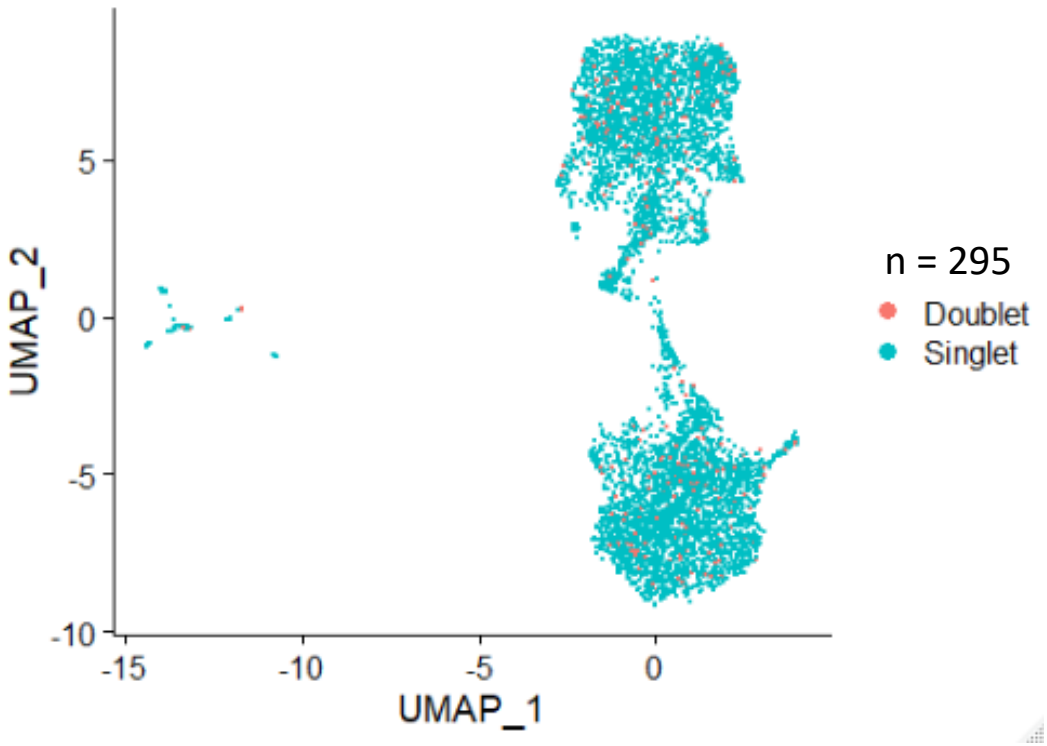


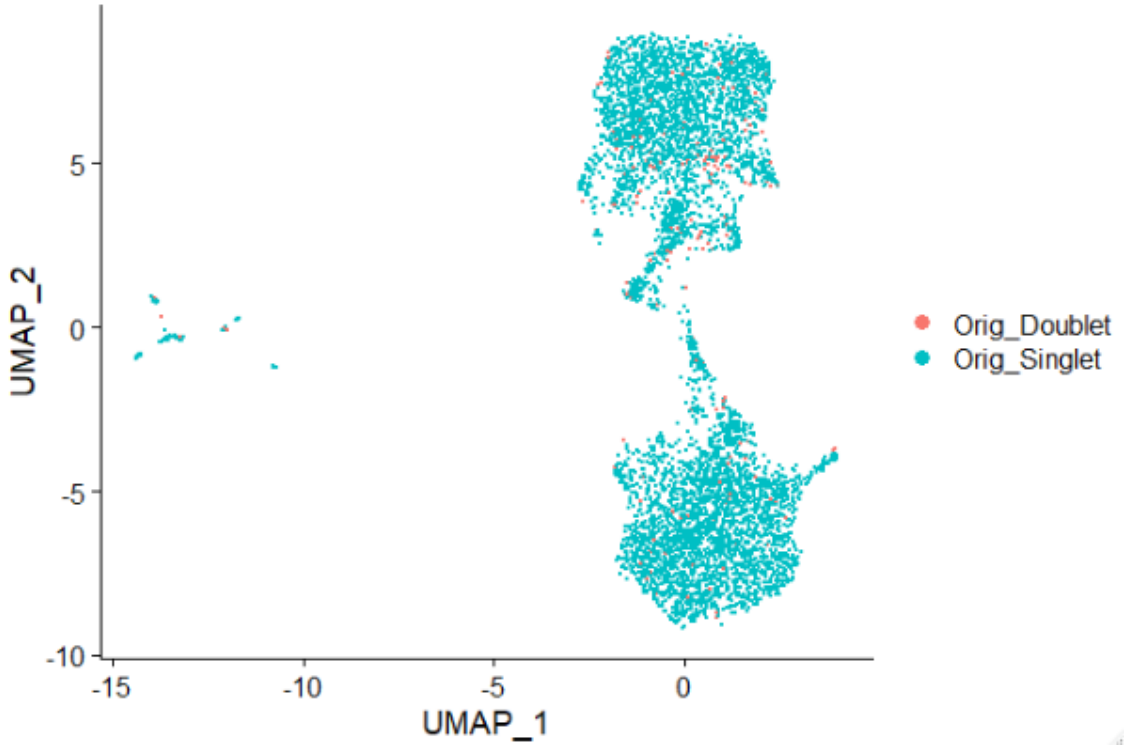




	Distance	Cell_Types	isADoublet	Group_Cluster
AAACCTGCAATGGACG	0.9990040	4	FALSE	2
AAACCTGCACGGCCAT	0.9975130	1	FALSE	2
AAACCTGGTGCGGTAA	0.9992511	4	FALSE	2
AAACCTGGTGCTTCTC	0.9997015	4	FALSE	2
AAACGGGAGCTGCAAG	0.9960734	9-6-3-4-10-8-7-2-5-two	TRUE	2
AAACGGGCAGCTGTAT	0.9704280	1	FALSE	2

DoubletDecon fails in detection of doublets…





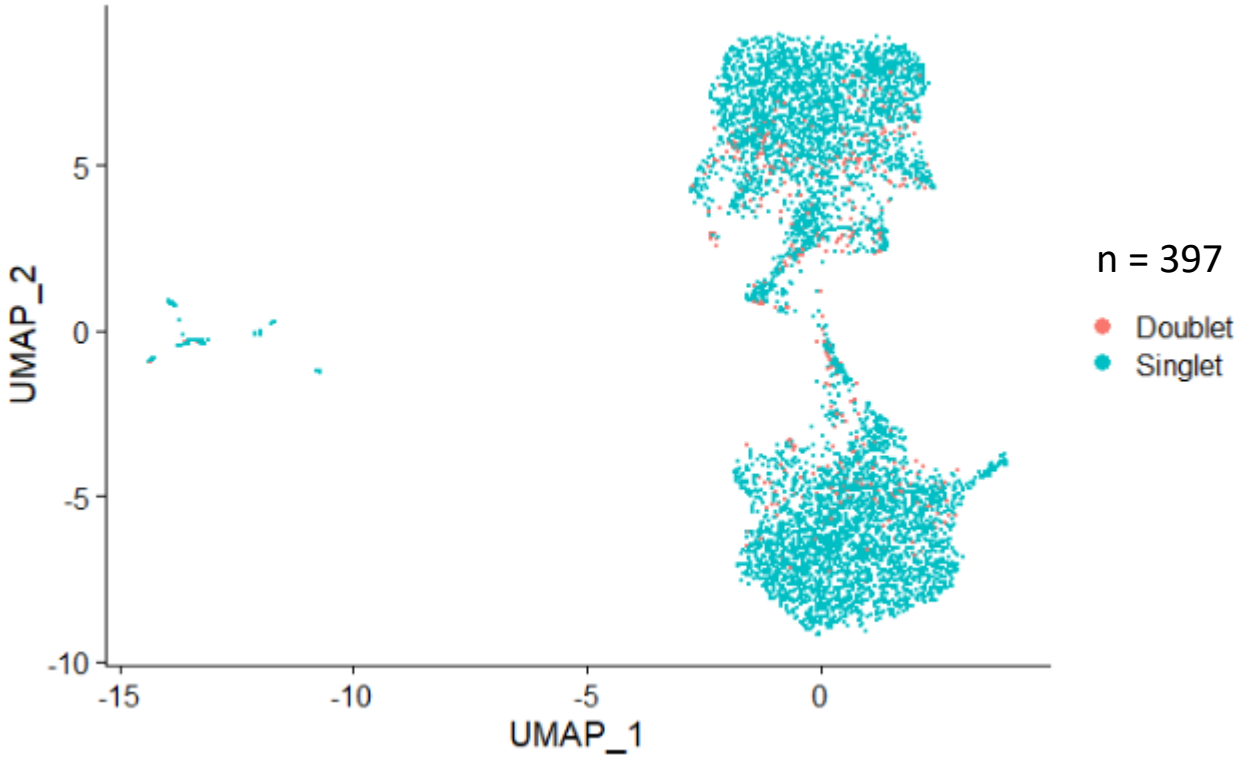
Sensitivity = 8 %

Specificity = 95 %

DoubletDecon fails in detection of doublets…

			TP	FN	FP	TN	Sensitivity	Specificity
12 clusters	100 dbl	DoubletDecon centroids 0.6	3	196	79	2649	0,015075	0,971041
12 clusters	100 dbl	DoubletDecon centroids 0.7	12	187	253	2475	0,060302	0,907258
12 clusters	100 dbl	DoubletDecon medoids 0.6	7	192	90	2638	0,035176	0,967009
12 clusters	100 dbl	DoubletDecon medoids 0.7	18	181	293	2435	0,090452	0,892595
9 clusters	100 dbl	DoubletDecon centroids 0.6	16	183	279	5218	0,080402	0,949245
9 clusters	200 dbl	DoubletDecon centroids 0.6	18	181	310	5187	0,090452	0,943606
9 clusters	500 dbl	DoubletDecon centroids 0.6	16	183	283	5214	0,080402	0,948517
9 clusters	100 dbl	DoubletDecon centroids 0.7	62	137	1314	4183	0,311558	0,760961
9 clusters	100 dbl	DoubletDecon medoids 0.6	18	181	310	5187	0,090452	0,943606
9 clusters	100 dbl	DoubletDecon medoids 0.7	26	173	687	4810	0,130653	0,875023

DoubletFinder

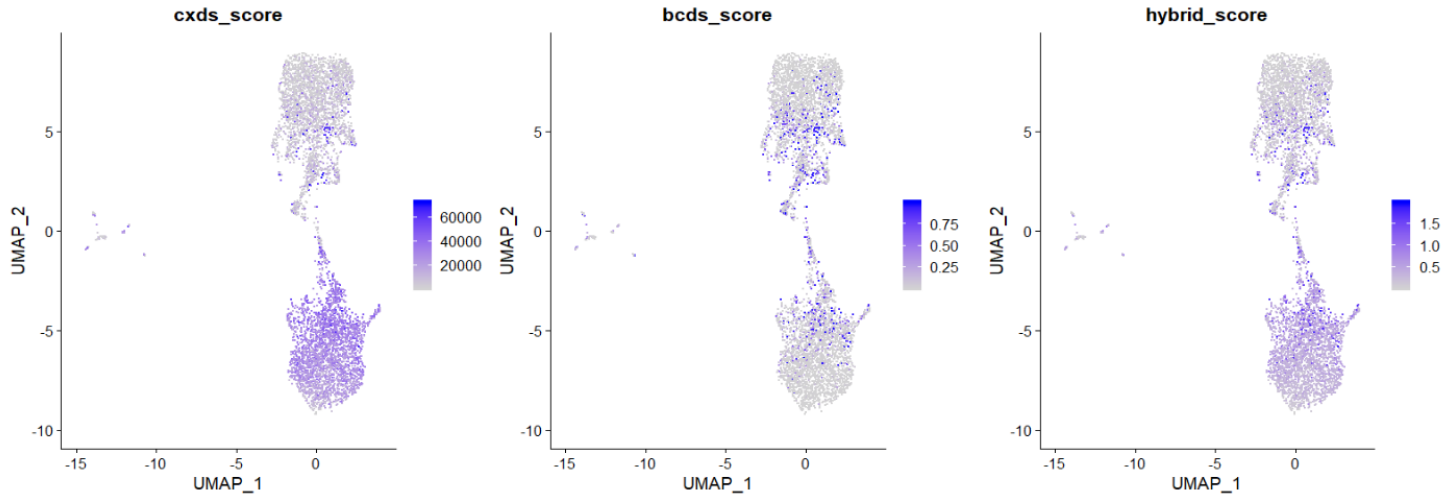


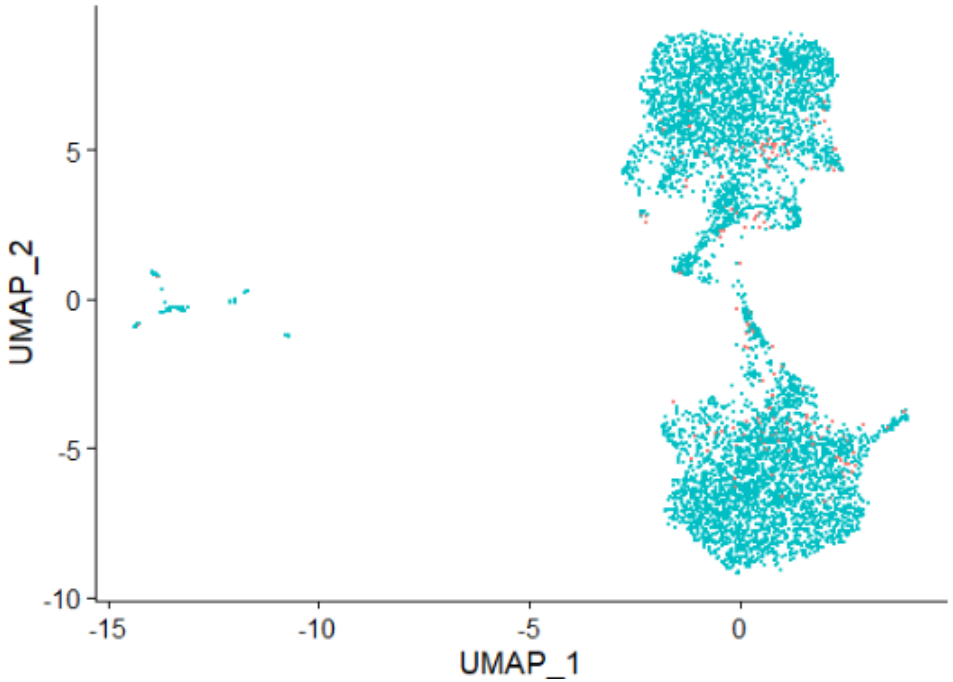
	Doublet	Singlet
Orig_Doublet	62	137
Orig_Singlet	1314	4183

Co-expression based doublet scoring (no artificial doublets)

Binary classification based doublet scoring
Artificial doublets - adding random pairs of columns to the GEX matrix

SCDS





		Doublet	Singlet
Orig_Doublet		108	91
Orig_Singlet		71	5426

Conclusion + future direction

	TP	FN	FP	TN	Sensitivity	Specificity
DoubletFinder	100	99	297	5200	0,502512563	0,945970529
SCDS	108	91	71	5426	0,542713568	0,987083864
The best result from DoubletD	62	137	1314	4183	0,311557789	0,760960524

DoubletDecon

DoubletFinder

SCDS - **s**ingle **c**ell **d**oublet **s**coring

TCR/BCR based package for detection of doublets?