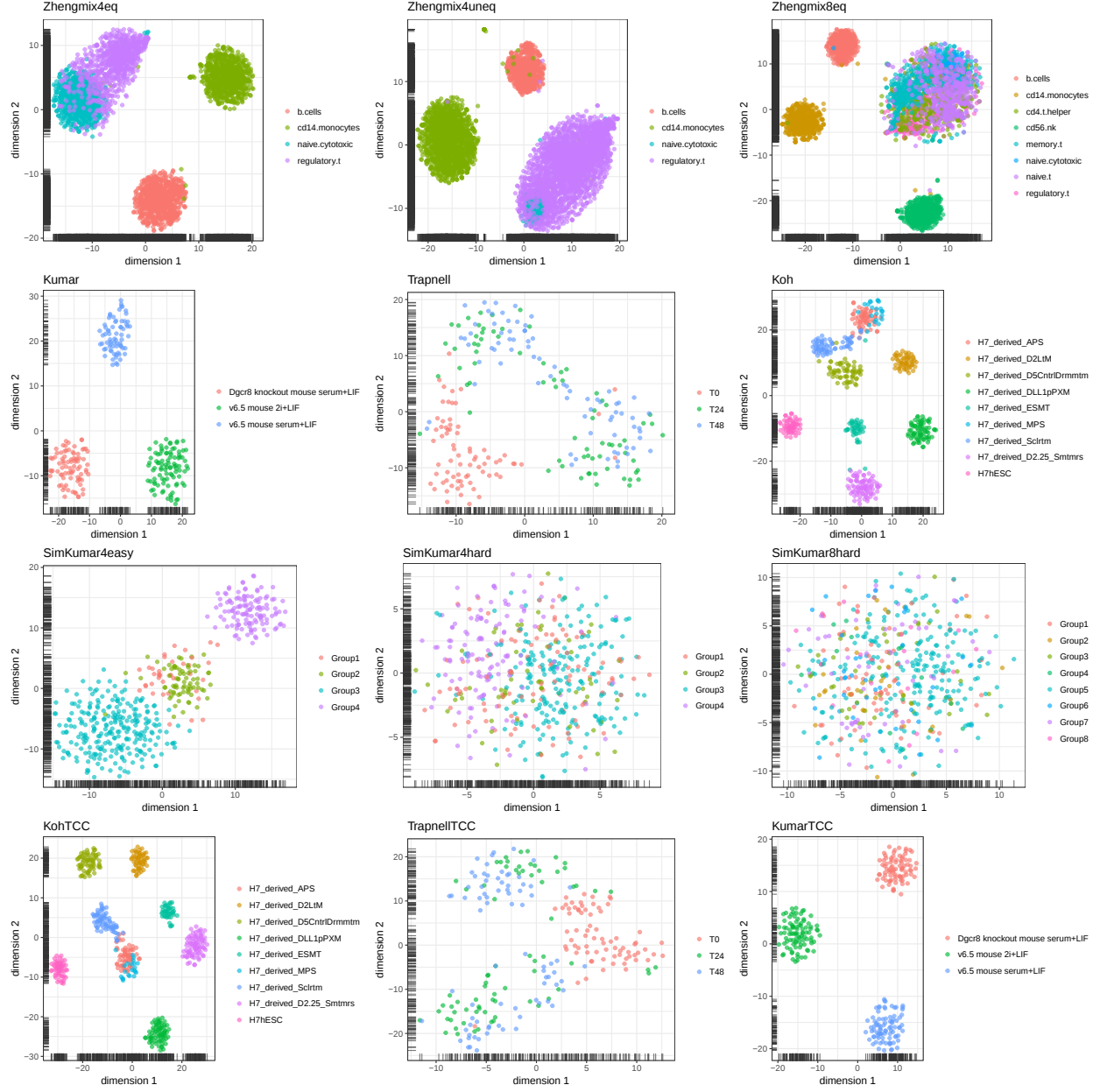


A systematic performance evaluation of clustering methods for
single-cell RNA-seq data
Supplementary Figures and Tables

Angelo Duò, Mark D Robinson, Charlotte Soneson

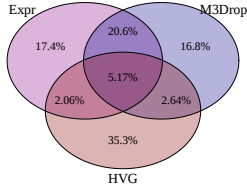
Supplementary Table 1: Overview of the datasets used in the study.

Data set	Total # cells	Total # features	Filtering	Retained # cells	Retained # features
Koh	531	48,981	Expr10	531	4,898
			HVG10	531	4,898
			M3Drop10	531	4,898
KohTCC	531	811,938	Expr10	531	81,193
			HVG10	531	81,194
			M3Drop10	531	81,194
Kumar	246	45,159	Expr10	246	4,515
			HVG10	246	4,516
			M3Drop10	246	4,516
KumarTCC	263	803,405	Expr10	263	80,340
			HVG10	263	80,340
			M3Drop10	263	80,340
SimKumar4easy	500	43,606	Expr10	500	4,360
			HVG10	500	4,361
			M3Drop10	500	4,361
SimKumar4hard	499	43,638	Expr10	499	4,363
			HVG10	499	4,364
			M3Drop10	499	4,364
SimKumar8hard	499	43,601	Expr10	499	4,360
			HVG10	499	4,360
			M3Drop10	499	4,360
Trapnell	222	41,111	Expr10	222	4,111
			HVG10	222	4,111
			M3Drop10	222	4,111
TrapnellTCC	227	684,953	Expr10	227	68,495
			HVG10	227	68,495
			M3Drop10	227	68,495
Zhengmix4eq	3,994	15,568	Expr10	3,555	1,556
			HVG10	3,300	1,557
			M3Drop10	2,430	1,557
Zhengmix4uneq	6,498	16,443	Expr10	6,414	1,644
			HVG10	5,079	1,644
			M3Drop10	3,830	1,644
Zhengmix8eq	3,994	15,716	Expr10	3,971	1,571
			HVG10	3,798	1,572
			M3Drop10	2,662	1,572

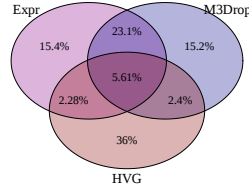


Supplementary Figure 1: t-SNE representations of all data sets used in the study (before gene filtering). The cells are colored by the annotation used to define the ground truth, to which all clusterings are compared.

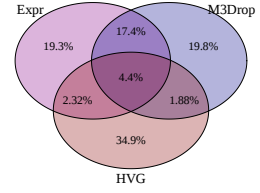
Zhengmix4eq



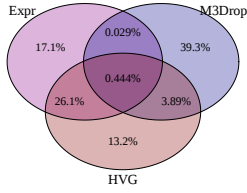
Zhengmix4uneq



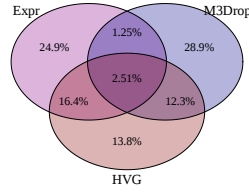
Zhengmix8eq



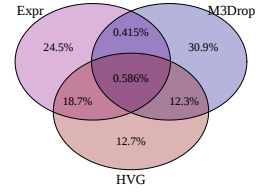
Kumar



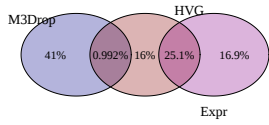
Trapnell



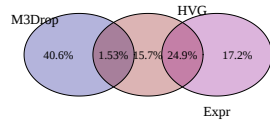
Koh



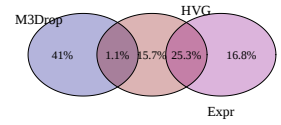
SimKumar4easy



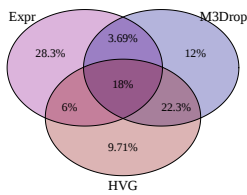
SimKumar4hard



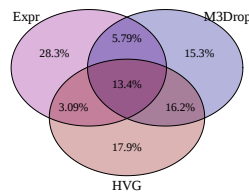
SimKumar8hard



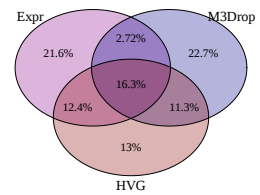
KohTCC



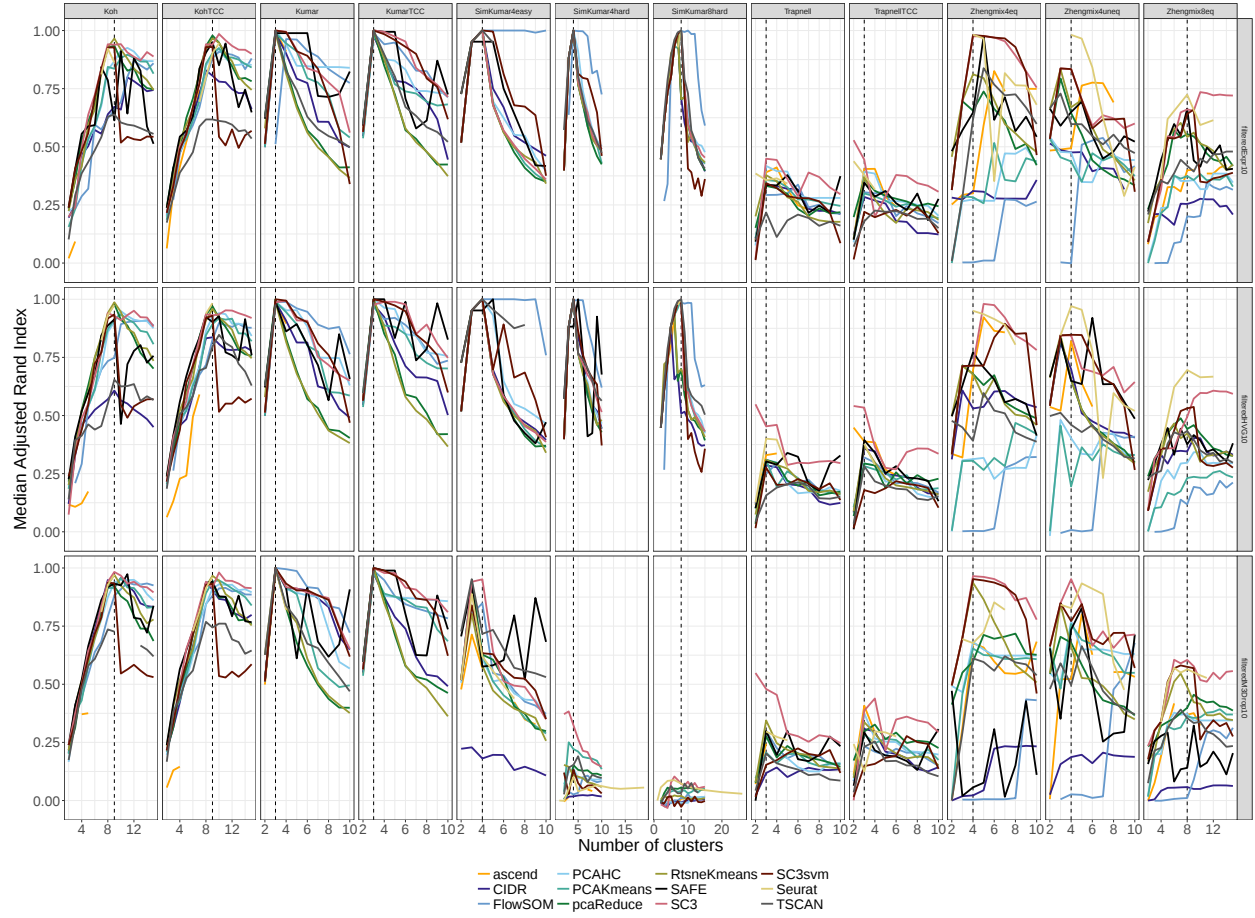
TrapnellTCC



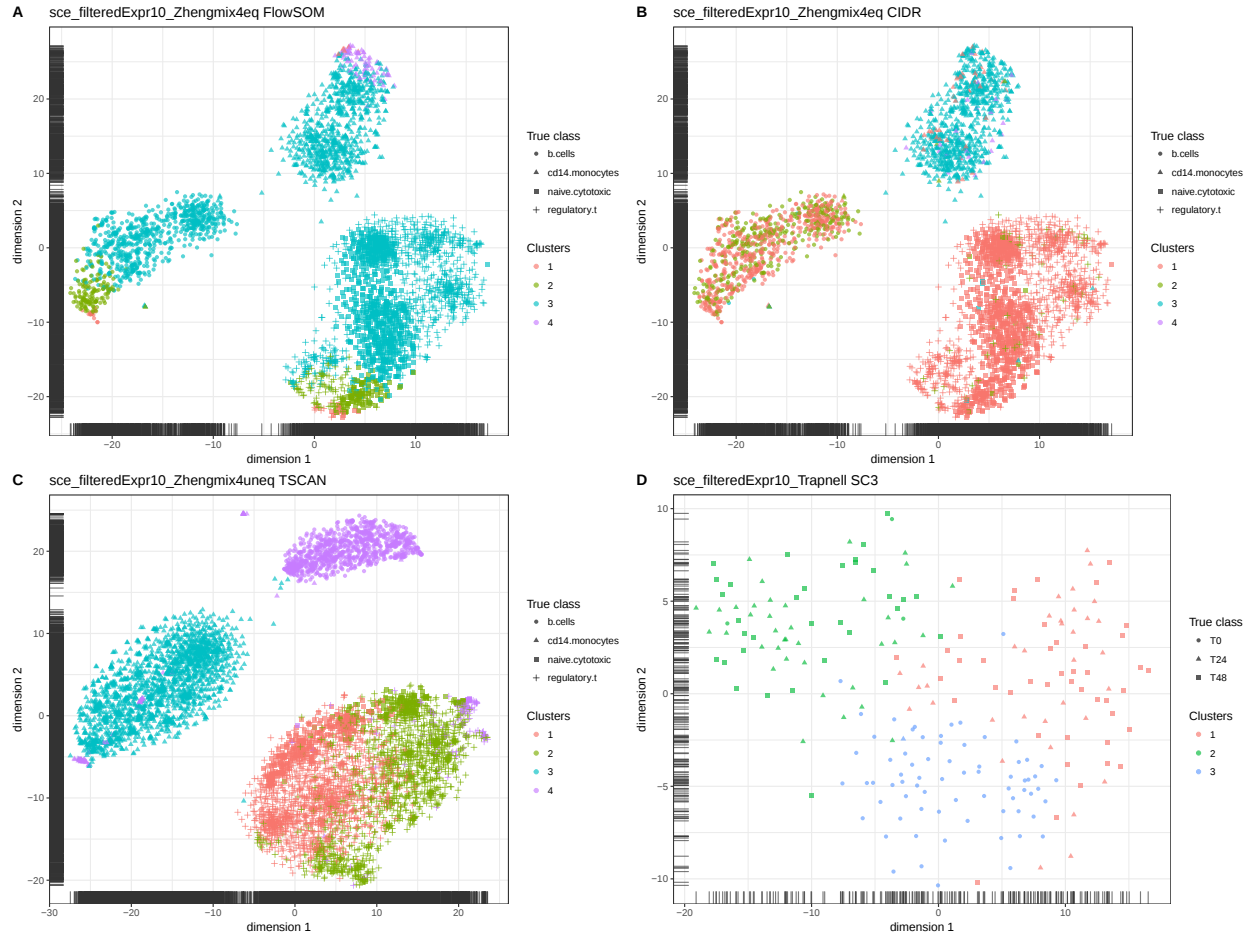
KumarTCC



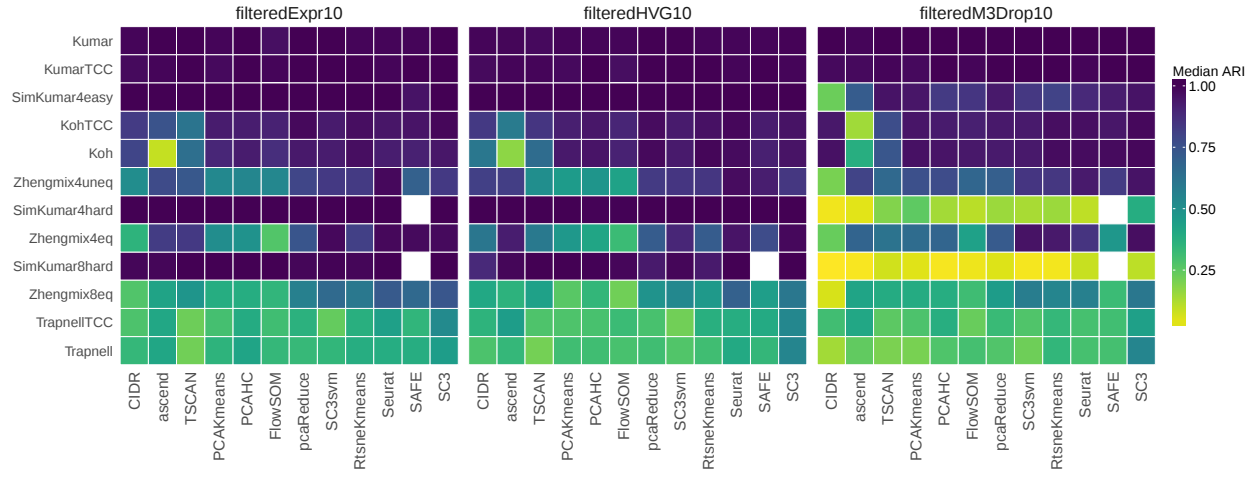
Supplementary Figure 2: Overlaps between the features retained by the different gene filtering approaches, for each of the data sets. The numbers represent percentages of the total number of features selected by at least one of the approaches.



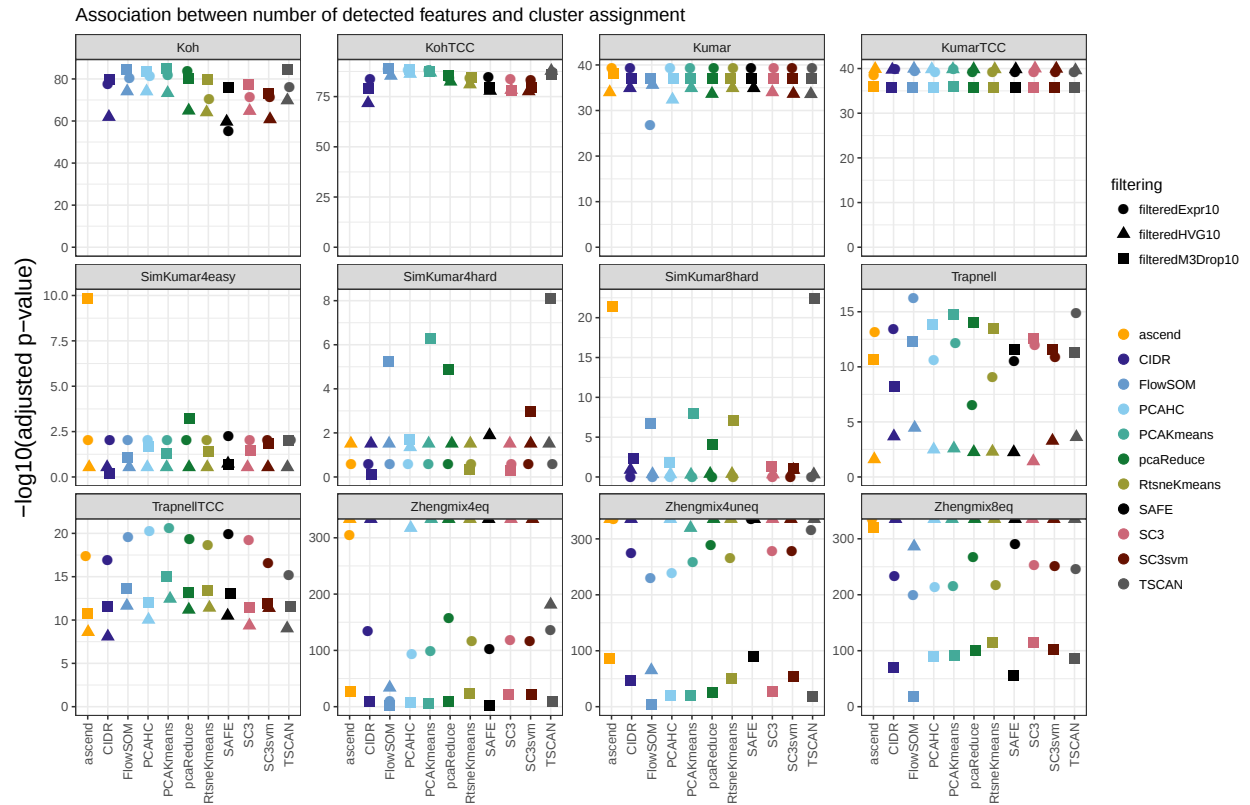
Supplementary Figure 3: Adjusted Rand Index (ARI) as a function of the number of clusters (k), for each of the data sets (columns) and gene filterings (rows). The vertical lines indicate the true number of populations in the respective data sets.



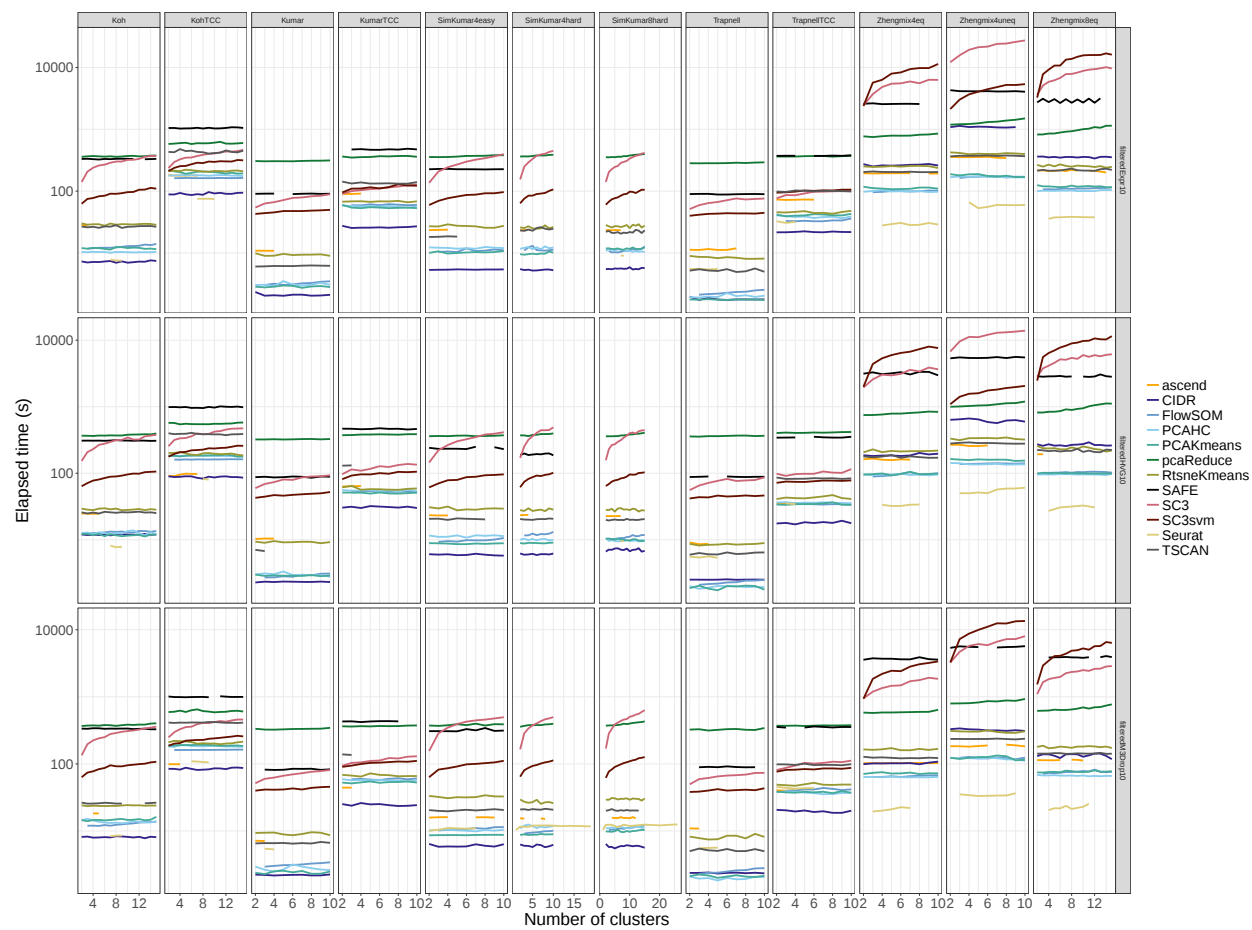
Supplementary Figure 4: t-SNE plots showing examples of data set/clustering method combinations with a large disagreement between the inferred and true clusters. The respective data sets, gene filtering methods and clustering methods are indicated above each plot. The true clusters are indicated with different shapes, and the inferred clusters in different colors.



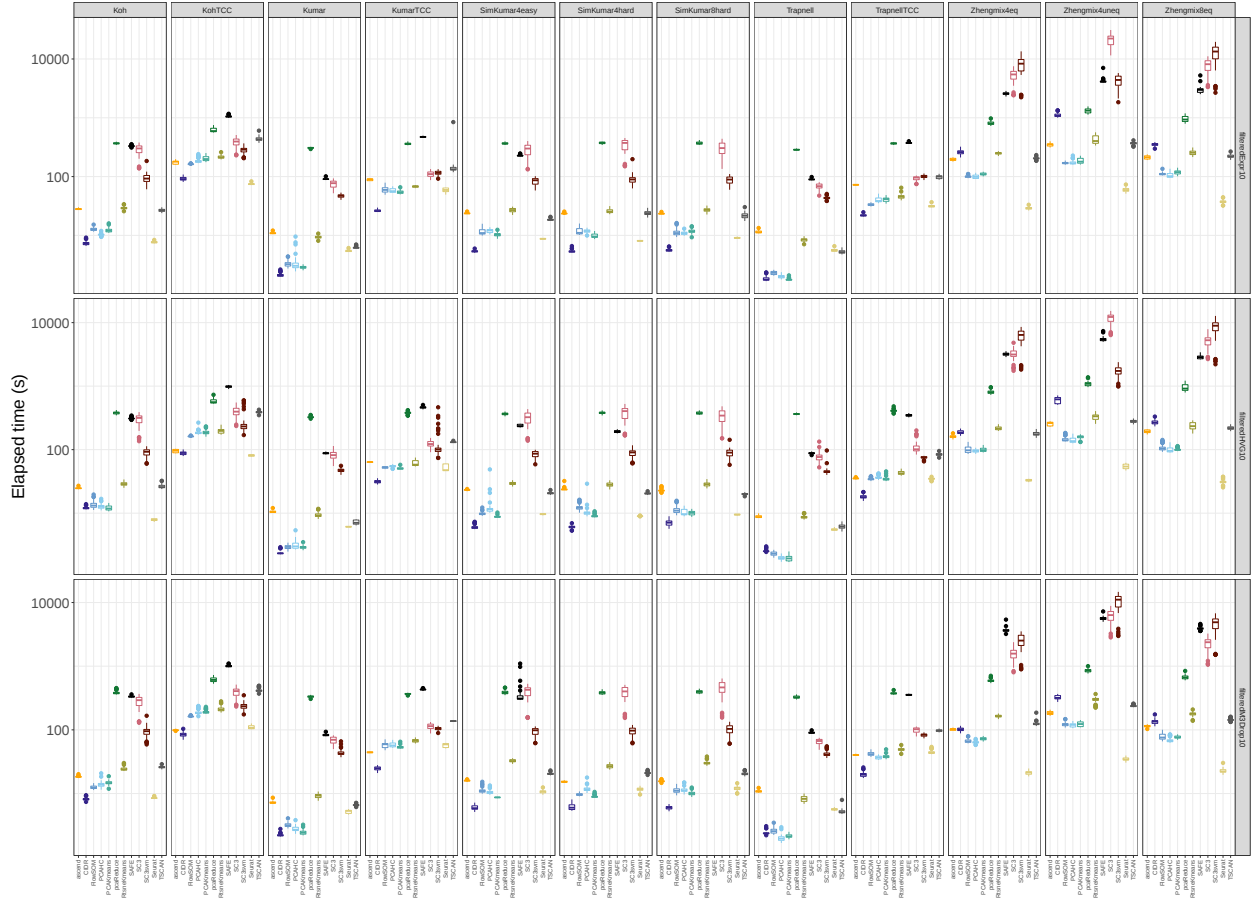
Supplementary Figure 5: Adjusted Rand Index (ARI) at the number of clusters that gives the best performance. Some methods failed to return any clustering for certain data sets (indicated by white squares).



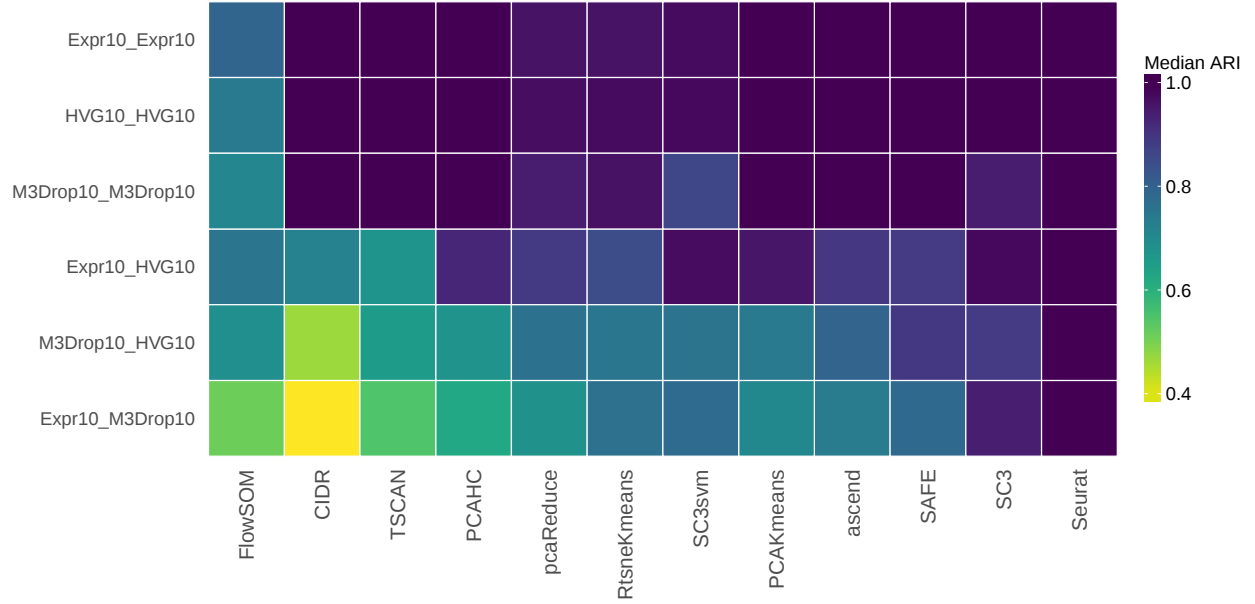
Supplementary Figure 6: Association between the number of detected features and the cluster assignment, for all data sets and filterings, with the true number of clusters. Only one of the five runs is shown. Values are adjusted p-values (using the Benjamini-Hochberg method) from a Kruskal-Wallis test.



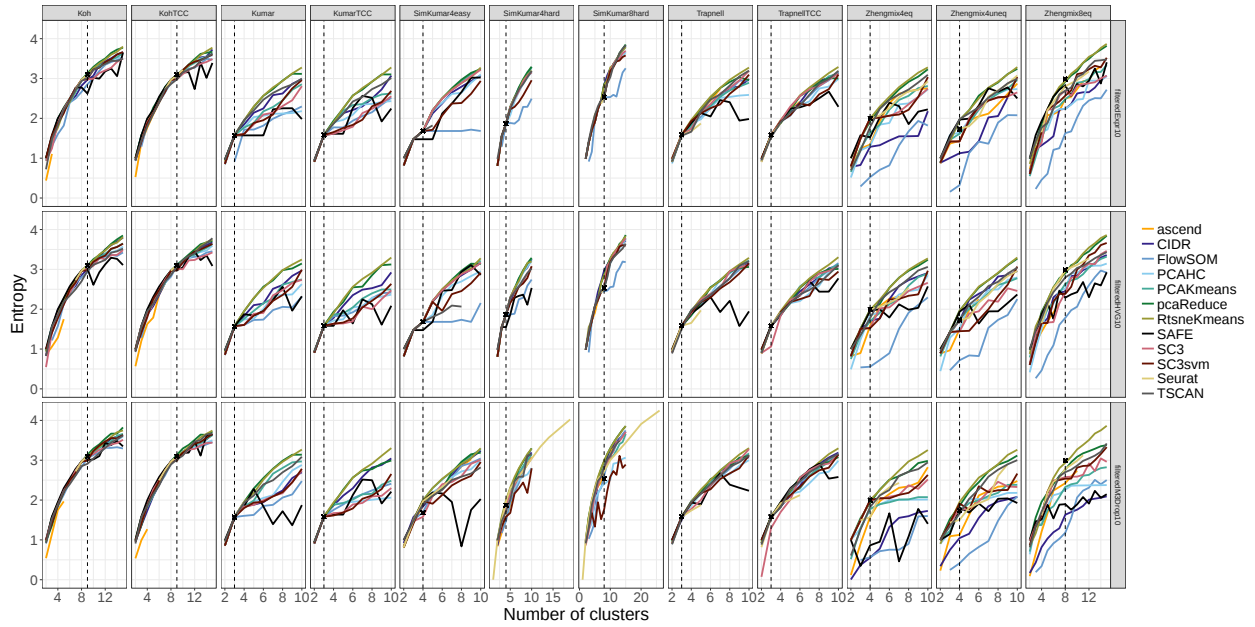
Supplementary Figure 7: Actual run times versus the imposed number of clusters, for each data set and gene filtering method.



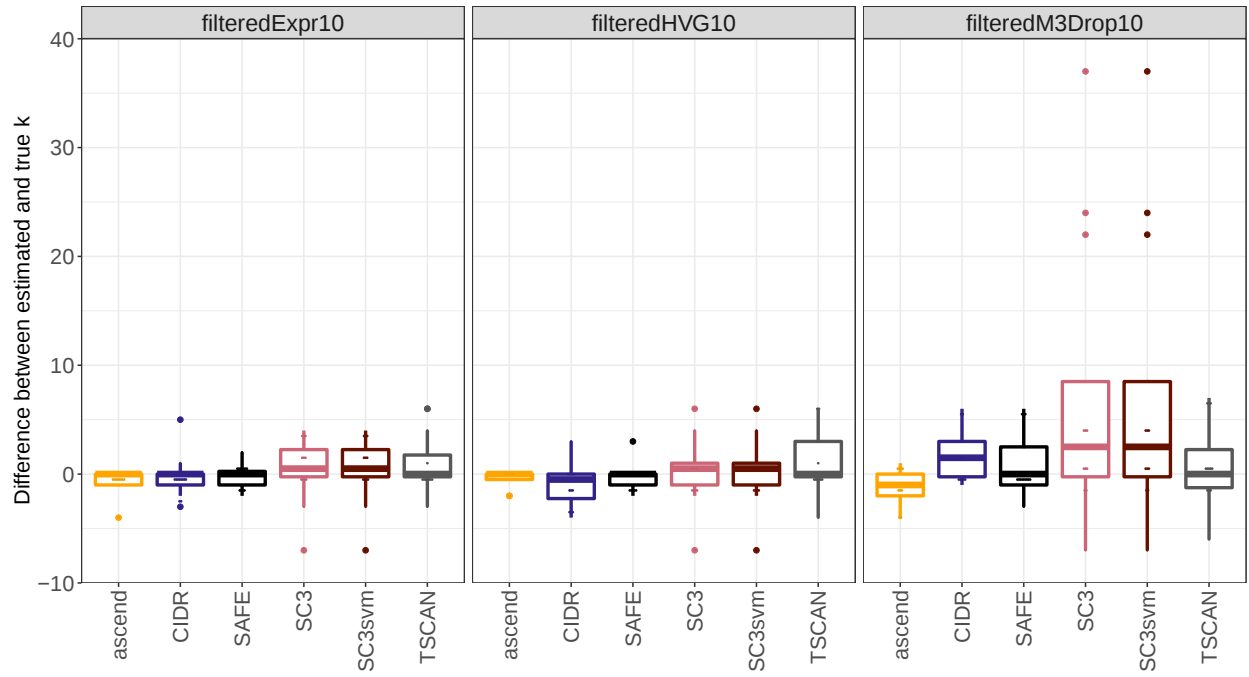
Supplementary Figure 8: Actual run times, across all numbers of clusters, for each data set and gene filtering method.



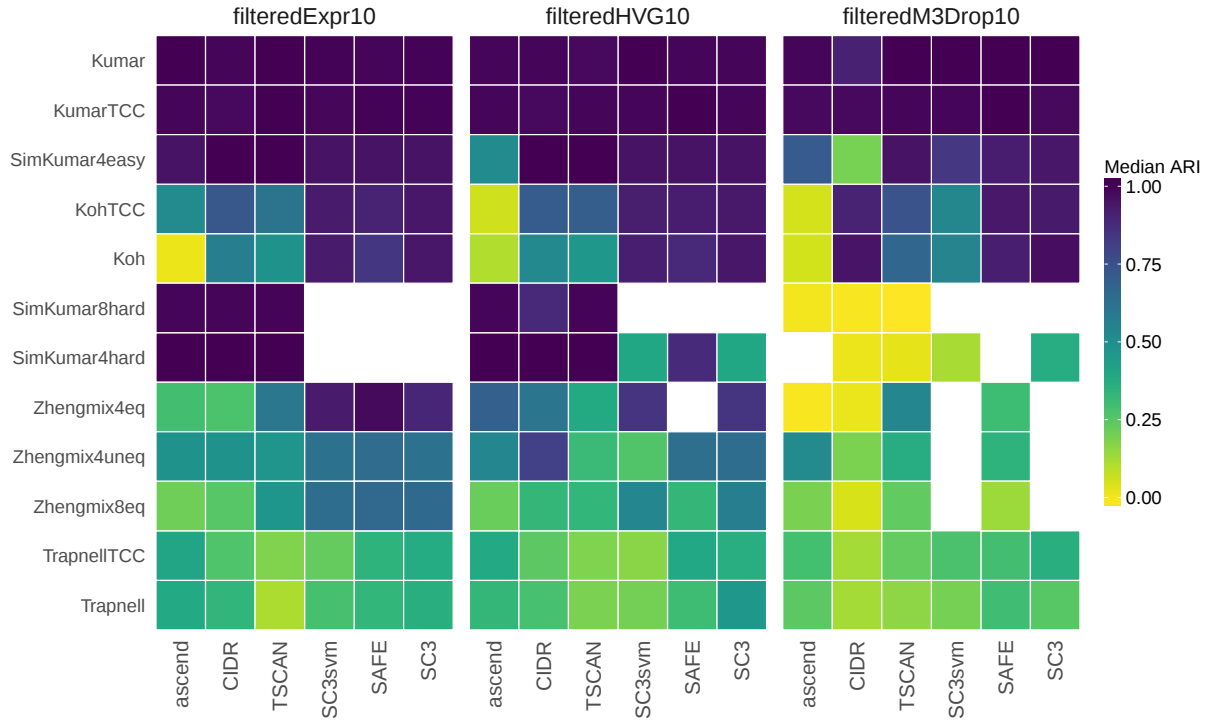
Supplementary Figure 9: Comparison of the clusterings obtained after different gene filterings. The values represent the median Adjusted Rand Index (ARI) across the data sets and filterings, for the true number of clusters, for each pair of filterings.



Supplementary Figure 10: Entropy for different imposed numbers of clusters. The true number of clusters is indicated by the vertical line, and the cross represents the entropy of the true partition.



Supplementary Figure 11: Difference between estimated and true number of clusters, for methods including a function to estimate the number of clusters.



Supplementary Figure 12: Adjusted Rand Index (ARI) at the number of clusters estimated by the indicated methods. Empty squares indicate that the estimated number of clusters was outside the evaluated range (2-10 for data sets with less than or equal to 7 true clusters, 2-15 for data sets with more than 7 true clusters).