

Supplementary Figures

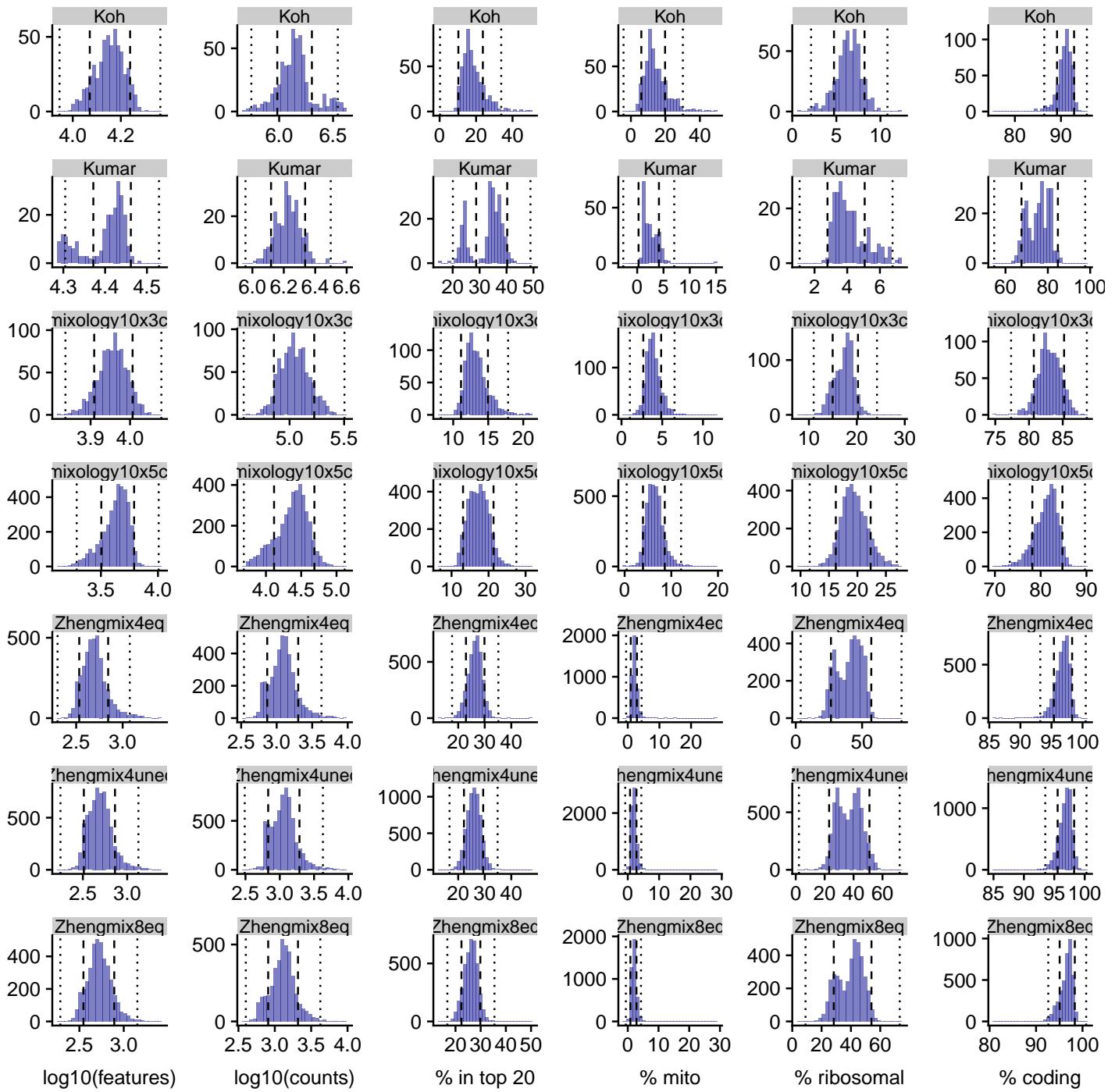
Pierre-Luc Germain

12 Mai, 2019

Contents

Supplementary Figure 1	2
Supplementary Figure 2	3
Supplementary Figure 3	4
Supplementary Figure 4	5
Supplementary Figure 5	6
Supplementary Figure 6	7

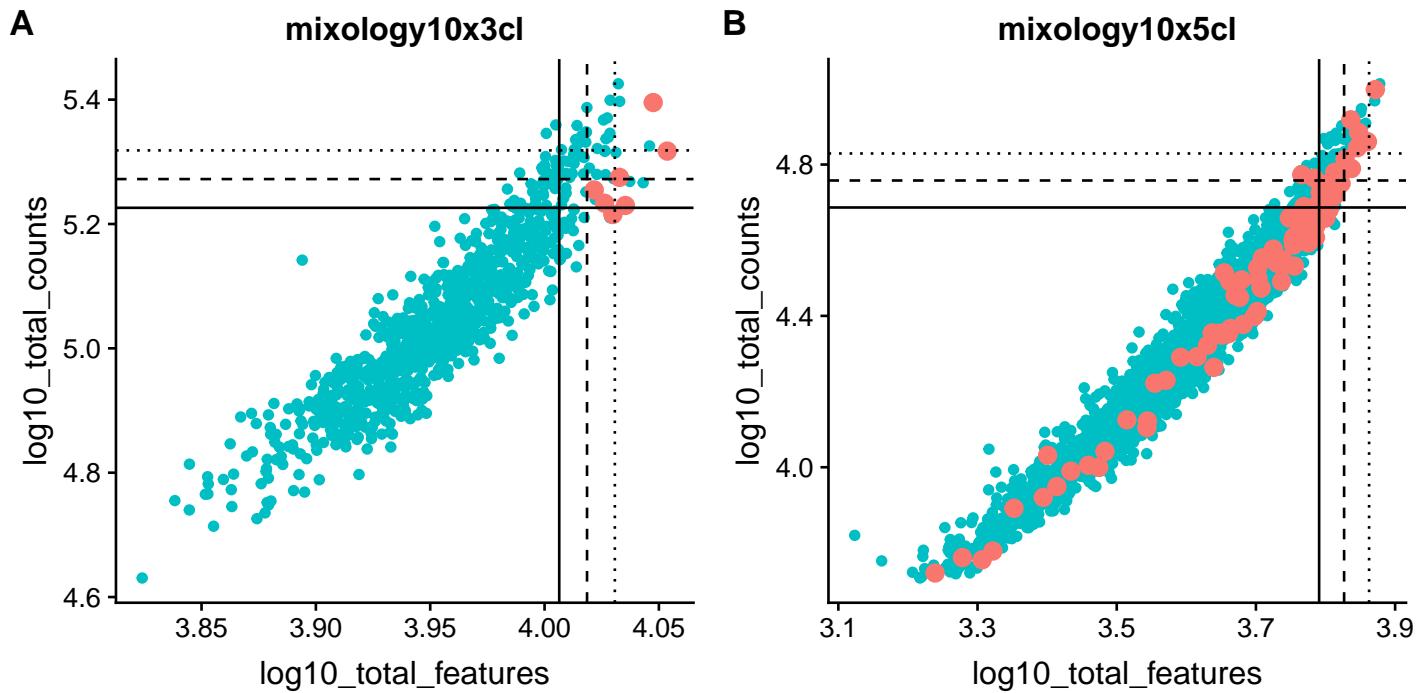
Supplementary Figure 1



Supplementary Figure 1

Distribution across cells of various control properties in the different datasets. The lines indicate respectively 2 and 5 median absolute deviations (MADs).

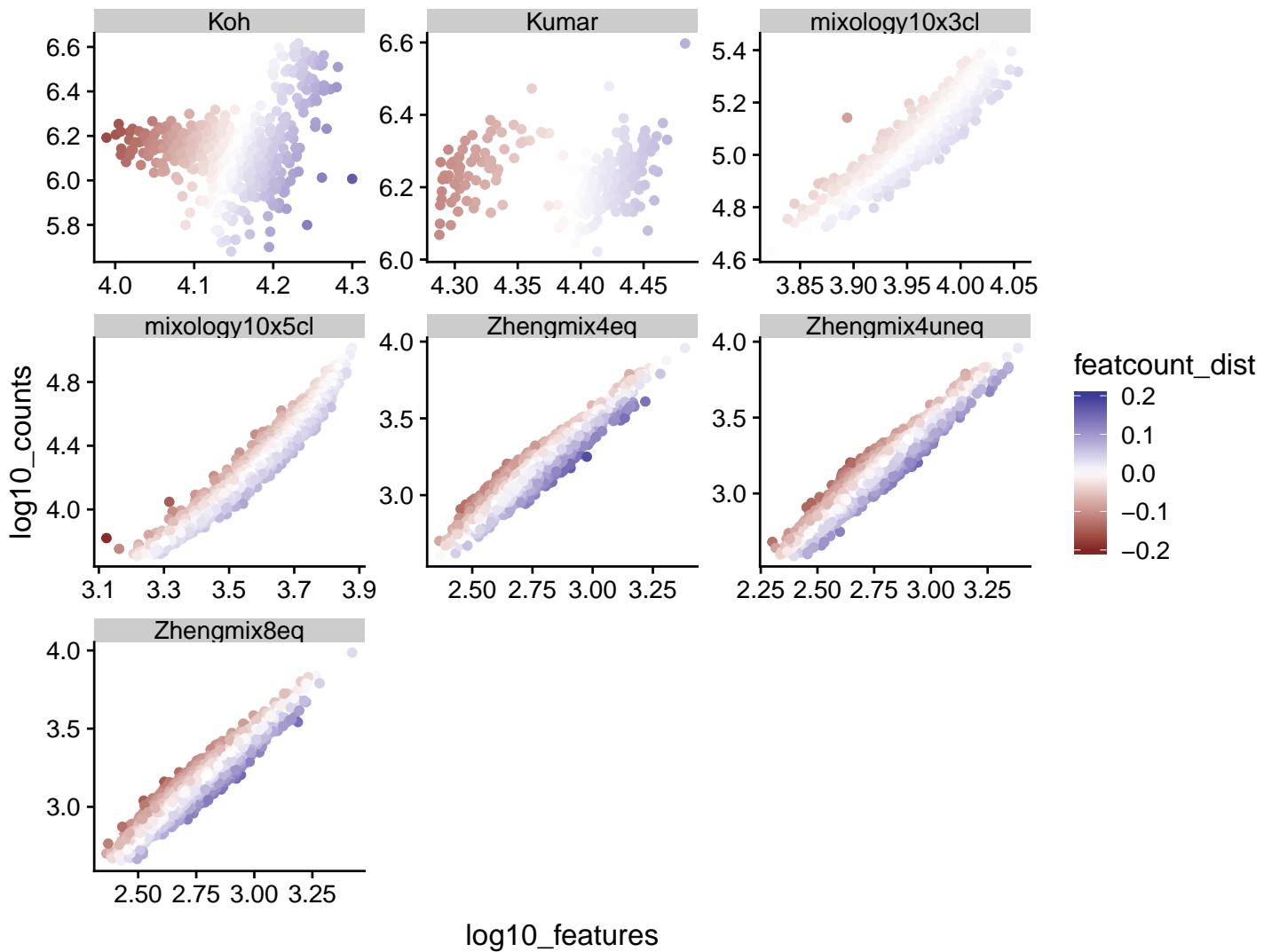
Supplementary Figure 2



Supplementary Figure 2

The total counts and total features per cell of doublets (red) versus other cells. We used the demuxlet annotation of doublets (based on SNPs) made available through CellBench. The lines indicate, respectively, 2, 2.5, and 3 median absolute deviations. While doublets tend to have a higher total count and especially number of detected features, these features alone are not always sufficient for their identification.

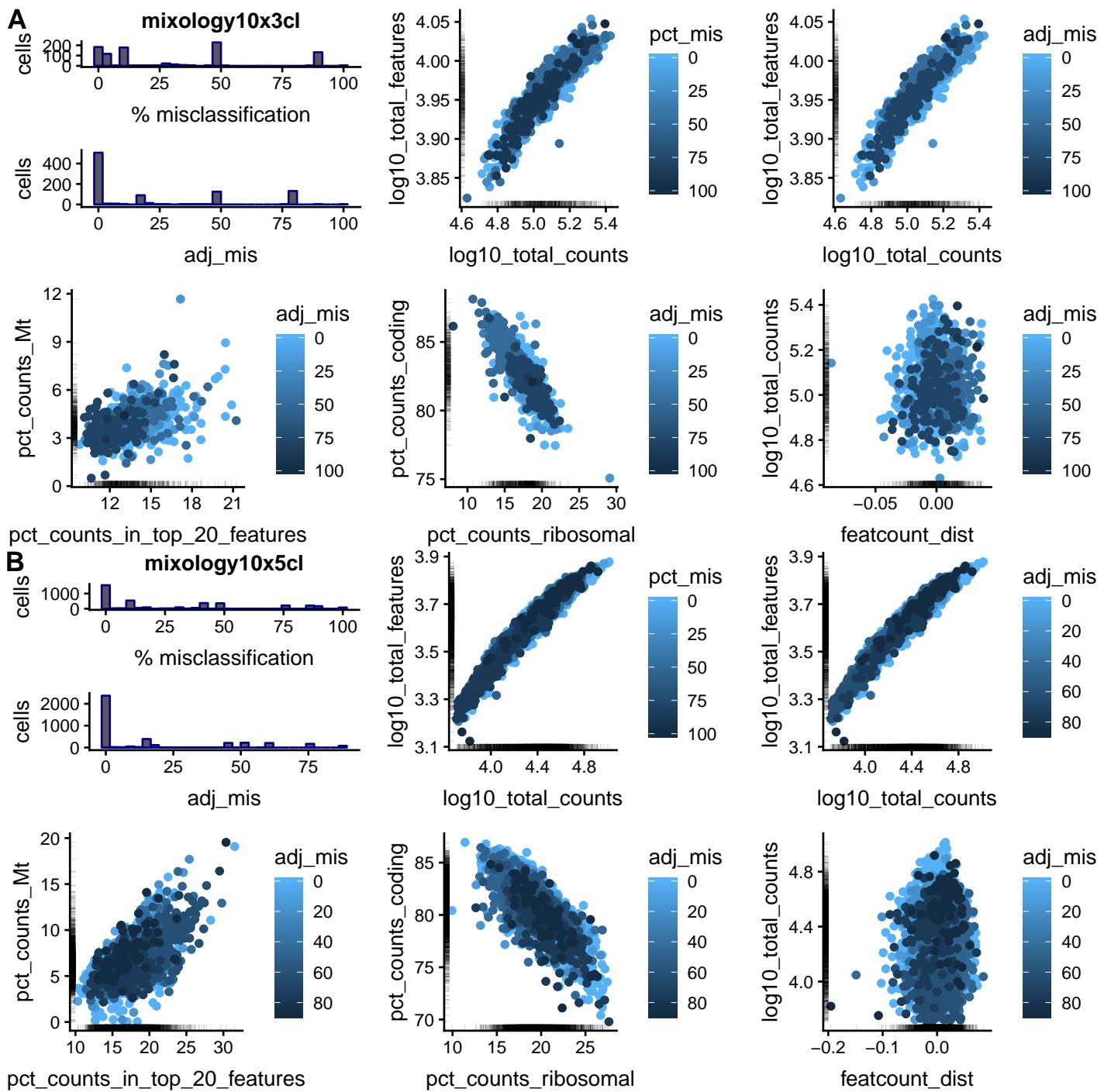
Supplementary Figure 3



Supplementary Figure 3

There is a tight relationship, in 10x datasets (i.e. not the Koh and Kumar datasets), between the total counts of a cell and its number of detected features. We therefore include, among control variables, deviation from this ratio.

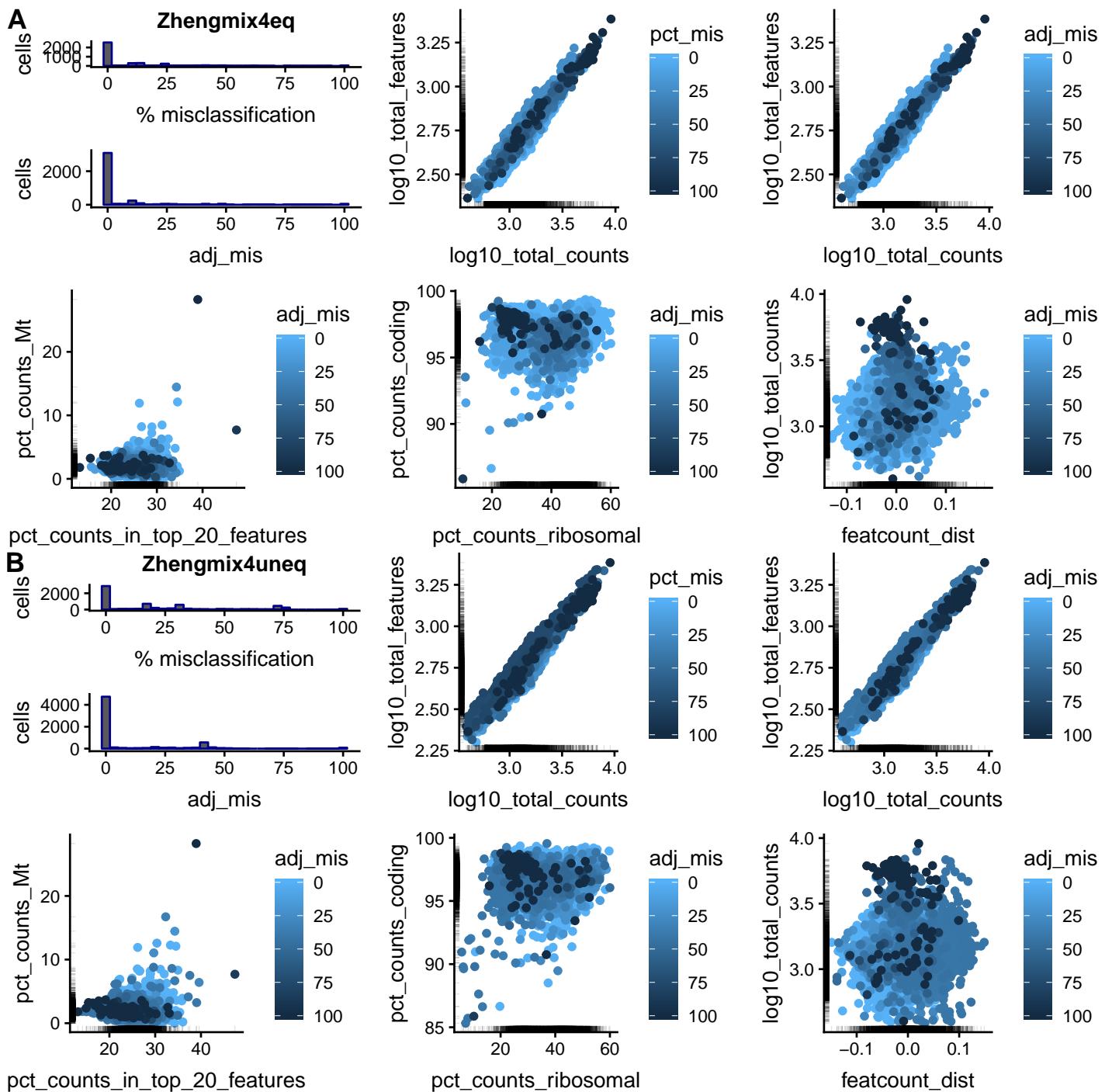
Supplementary Figure 4



Supplementary Figure 4

Relationship between various cellular properties and the frequency of cluster mis-assignment for the mixology10x3cl (A) and mixology10x5cl (B) datasets. The percentage of misclassification refers to the frequency with which a given cell is assigned the wrong cluster (using the Hungarian algorithm for cluster matching) across several hundred clustering runs with varying parameters. Since some subpopulations tend to be more misclassified than others, the adjusted rate of misclassification (adj_mis) is subtracted for the subpopulation median misclassification rate.

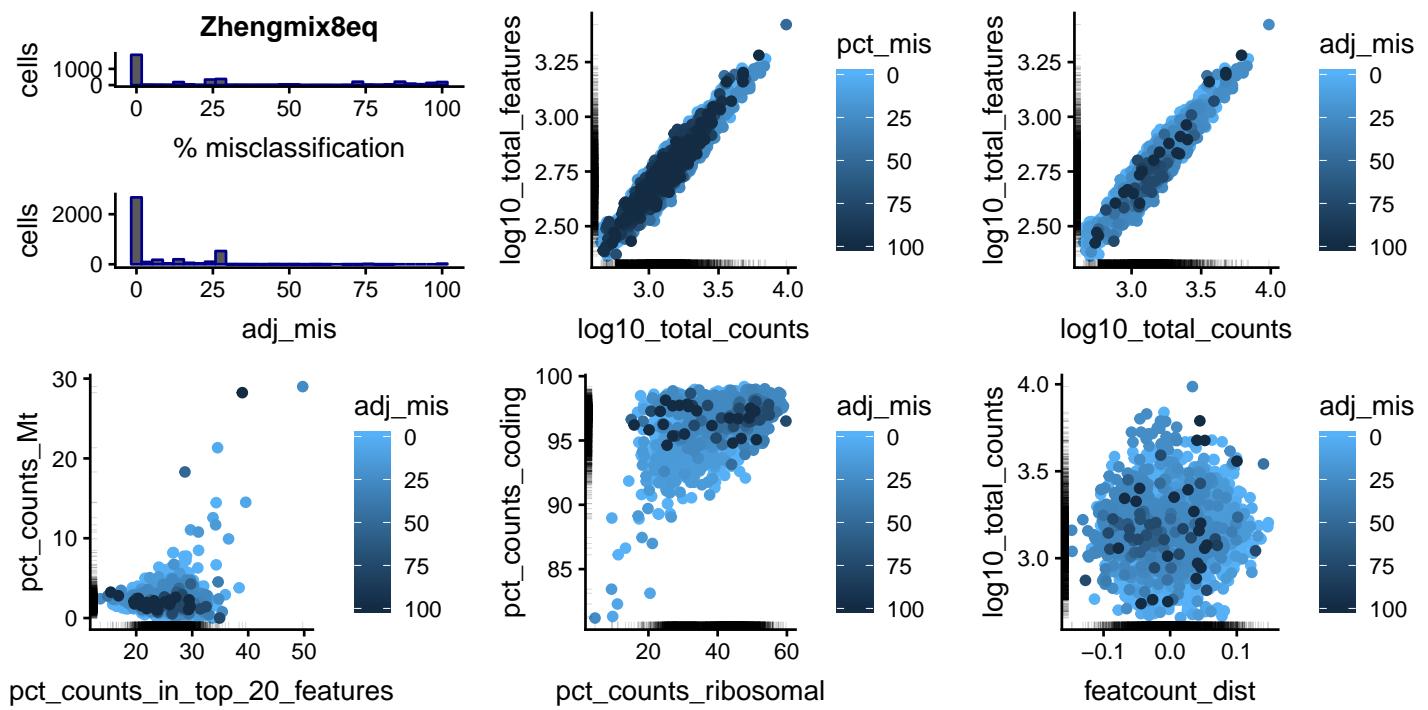
Supplementary Figure 5



Supplementary Figure 5

Relationship between various cellular properties and the frequency of cluster mis-assignment for the Zheng equal (A) or unequal (B) mixtures of four cell types. See Supplementary Figure 4 for more information. The only clear pattern is that cells with a high number of reads or features tend to have a higher misclassification rate.

Supplementary Figure 6



Supplementary Figure 6

Relationship between various cellular properties and the frequency of cluster mis-assignment for the Zheng mixture of 8 cell types. See Supplementary Figure 4 for more information.