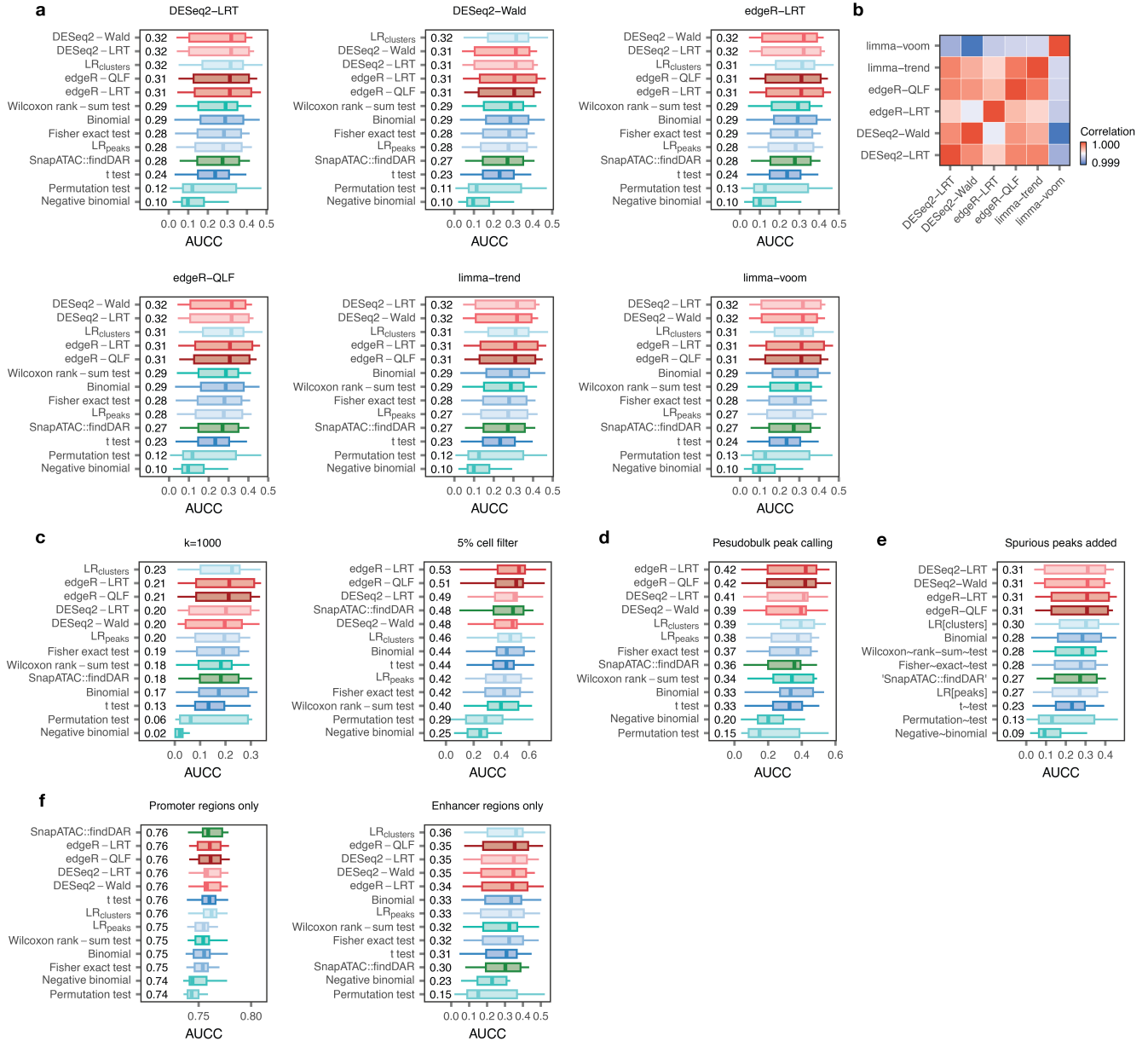


Supplementary figures



Supplementary Fig. 1 | Evaluating single-cell DA methods with matched bulk data: secondary analyses.

a, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 1, using matching bulk ATAC-seq as a reference, after omitting each of the six DA methods used to analyze the bulk ATAC-seq data in turn. Inset text shows the median AUC.

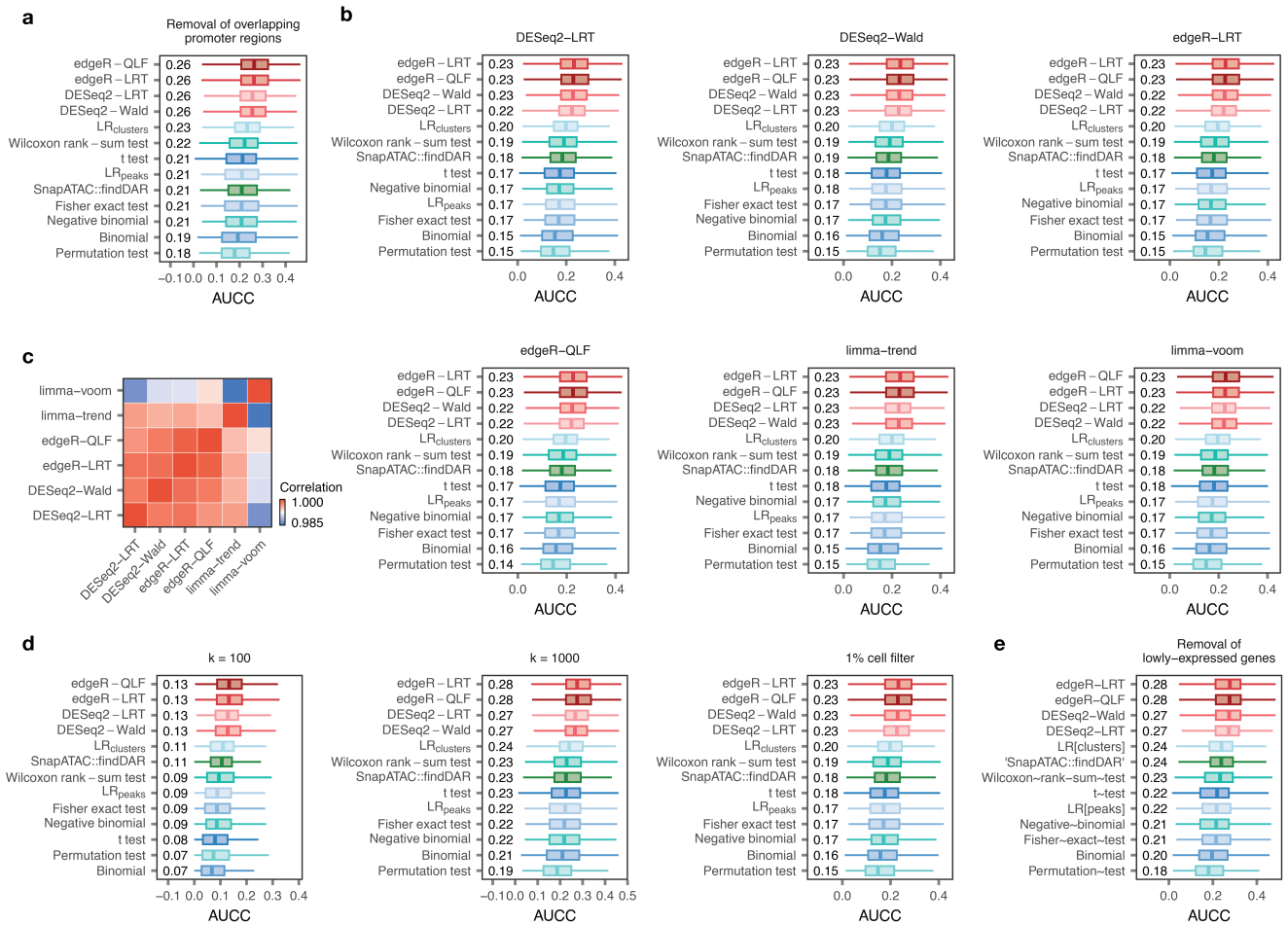
b, Correlation matrix between AUC values after omitting each of the six DA methods used to analyze the bulk ATAC-seq data in turn.

c, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 1, using matching bulk ATAC-seq as a reference, after varying the value of the parameter k used to calculate the AUC, left, or filtering peaks open in less than 5% of cells, right. Inset text shows the median AUC.

d, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 1, using matching bulk ATAC-seq as a reference, when using peaks called from pseudobulks of the single-cell data (all $p \geq 0.05$ except $p = 0.021$ for negative binomial regression, linear model with interaction term). Inset text shows the median AUC.

e, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 1, using matching bulk ATAC-seq as a reference, when injecting spurious peaks (Methods; all $p \geq 0.05$ except $p \leq 5.0 \times 10^{-5}$ for permutation test and negative binomial regression, linear model with interaction term). Inset text shows the median AUC.

f, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 1, using matching bulk ATAC-seq as a reference, shown separately for peaks in promoter vs. enhancer regions. Inset text shows the median AUC.



Supplementary Fig. 2 | Evaluating single-cell DA methods with single-cell multi-omics: secondary analyses, gene level.

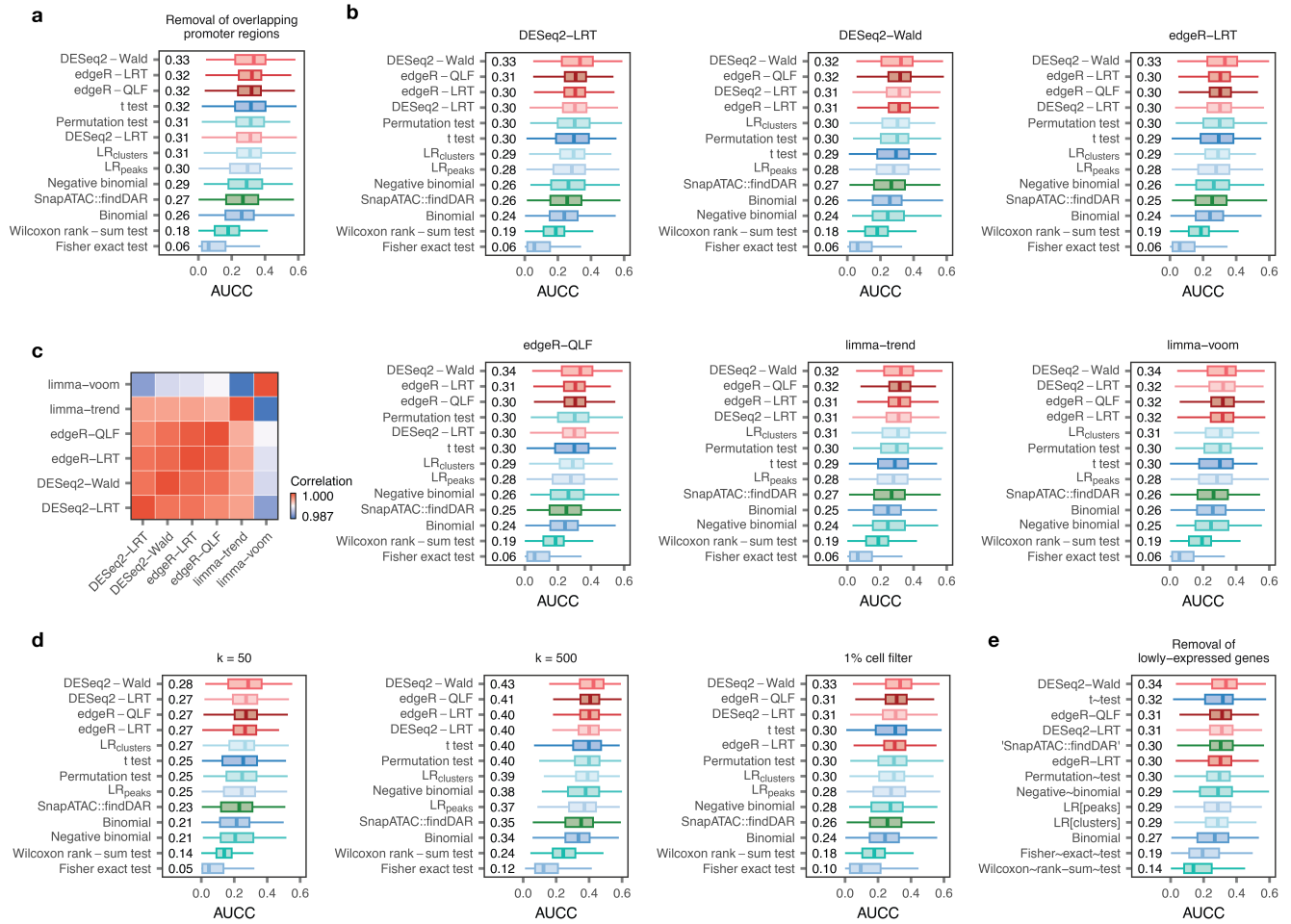
a, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 2, using matching snRNA-seq as a reference, after removing genes with overlapping promoter regions. Inset text shows the median AUC.

b, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 2, using matching snRNA-seq as a reference, after omitting each of the six DE methods used to analyze the RNA modality in turn. Inset text shows the median AUC.

c, Correlation matrix between AUC values after omitting each of the six DE methods used to analyze the RNA modality in turn.

d, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 2, using matching snRNA-seq as a reference, after varying the value of the parameter k used to calculate the AUC, left and middle, or filtering promoter regions open in less than 1% of cells, right. Inset text shows the median AUC.

e, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 2, using matching snRNA-seq as a reference, after filtering lowly-expressed genes. Inset text shows the median AUC.



Supplementary Fig. 3 | Evaluating single-cell DA methods with single-cell multi-omics: secondary analyses, GO level.

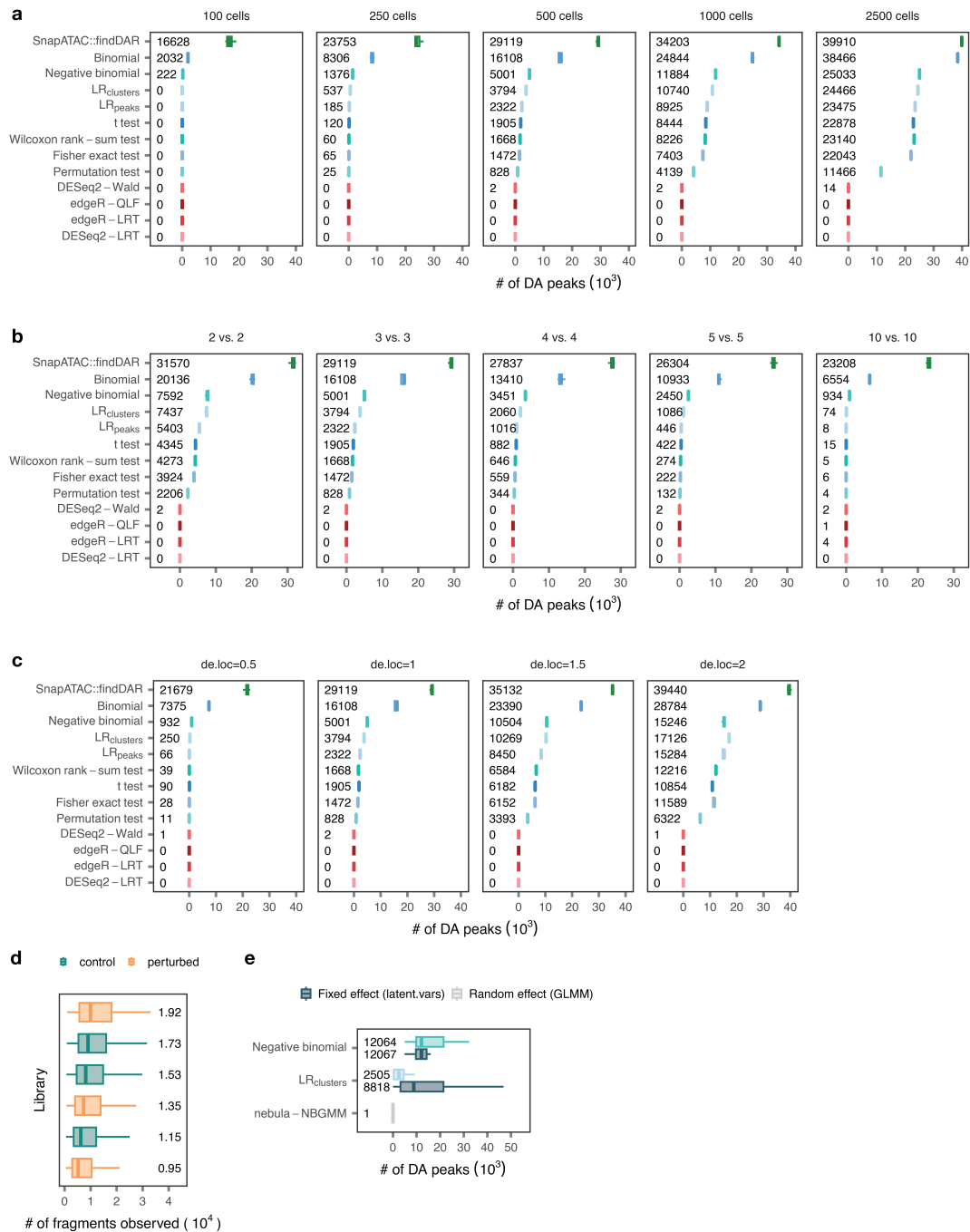
a, Area under the concordance curve (AUC) at the level of GO terms for single-cell DA methods in Experiment 2, using matching snRNA-seq as a reference, after removing genes with overlapping promoter regions. Inset text shows the median AUC.

b, Area under the concordance curve (AUC) at the level of GO terms for single-cell DA methods in Experiment 2, using matching snRNA-seq as a reference, after omitting each of the six DE methods used to analyze the RNA modality in turn. Inset text shows the median AUC.

c, Correlation matrix between AUC values at the level of GO terms after omitting each of the six DE methods used to analyze the RNA modality in turn.

d, Area under the concordance curve (AUC) at the level of GO terms for single-cell DA methods in Experiment 2, using matching snRNA-seq as a reference, after varying the value of the parameter k used to calculate the AUC, left and middle, or filtering promoter regions open in less than 1% of cells, right. Inset text shows the median AUC.

e, Area under the concordance curve (AUC) at the level of GO terms for single-cell DA methods in Experiment 2, using matching snRNA-seq as a reference, after filtering lowly-expressed genes. Inset text shows the median AUC.



Supplementary Fig. 4 | False discoveries in single-cell DA analysis: secondary analyses.

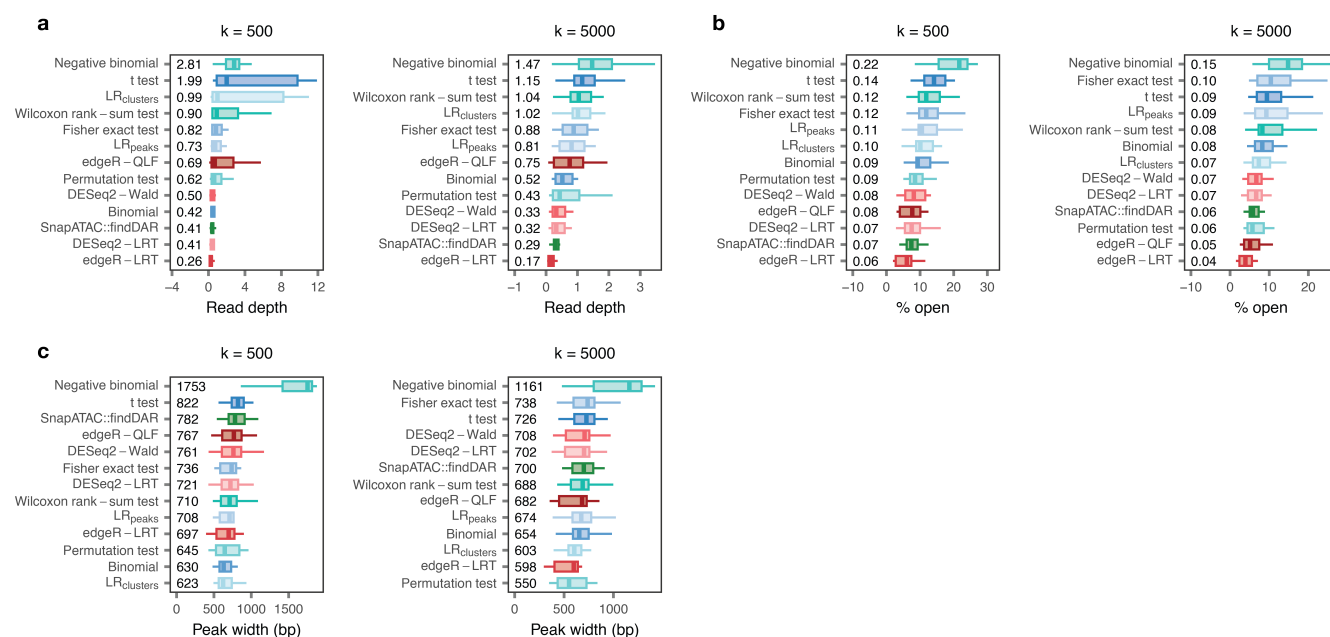
a, Number of DA peaks detected at 5% FDR in model-based simulations of scATAC-seq data without biological differences between experimental conditions, when varying the total number of cells in the dataset. Inset text shows the median number of DA peaks.

b, As in **a**, but when varying the number of replicates in the dataset.

c, As in **a**, but when varying the magnitude of the batch effects between individual replicates.

d, Number of fragment counts per cell in simulated replicates with variable sequencing depths. Inset text shows median counts per cell.

e, Number of DA peaks detected between randomly assigned replicates at 5% FDR within each cell type in random comparisons of published scATAC-seq data, when including replicate as a fixed effect in generalized linear models or a random effect in negative binomial generalized linear mixed models (nebula-NBGM). Inset text shows the median number of DA peaks.

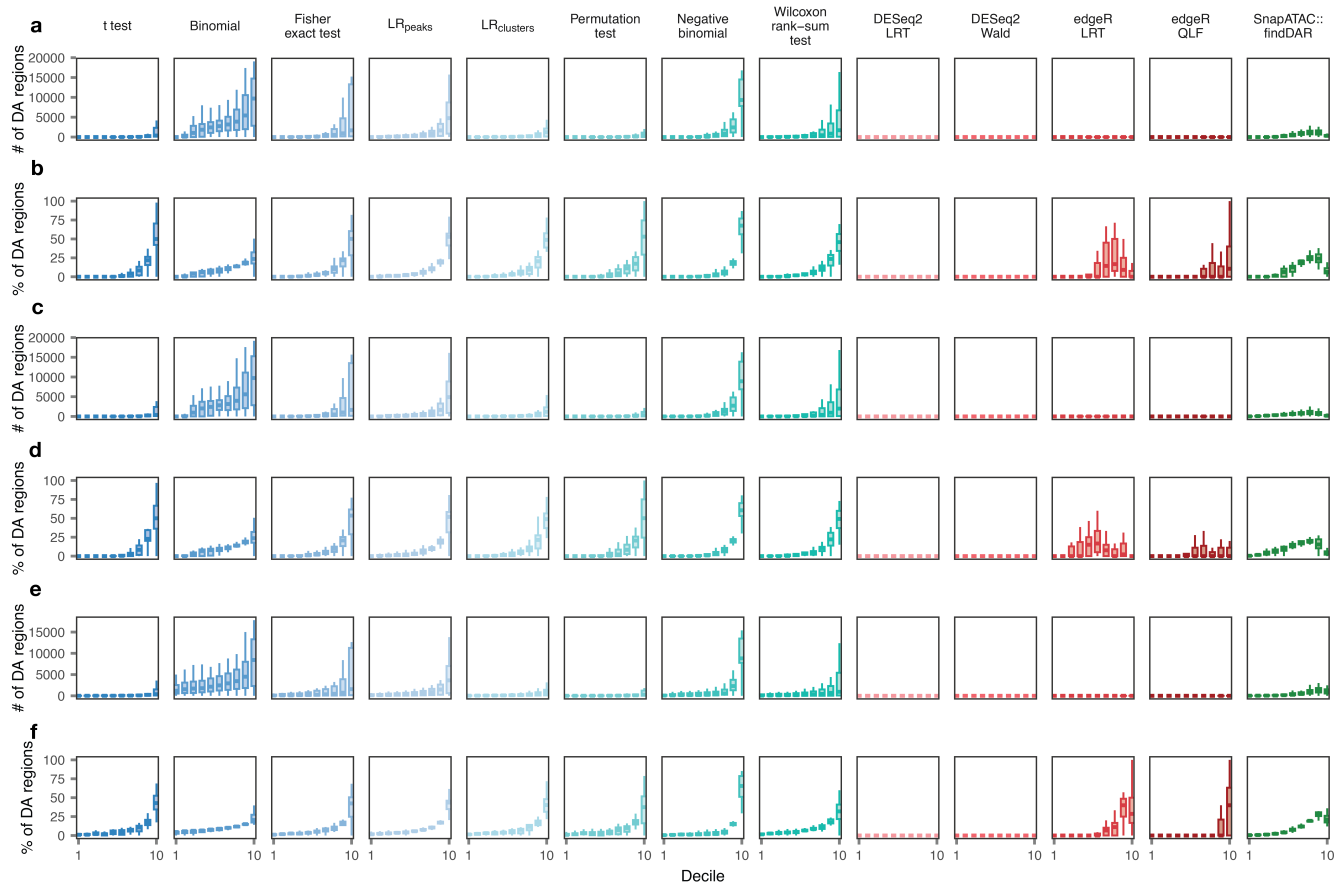


Supplementary Fig. 5 | Biases in single-cell DA analysis: secondary analyses, properties of DA peaks.

a, Mean read depth of the top-500 and top-5,000 DA peaks identified by each single-cell DA method in published scATAC-seq datasets. Inset text shows the median across comparisons.

b, As in **a**, but showing the proportion of cells in which these peaks are open.

c, As in **a**, but showing the width of each peak.



Supplementary Fig. 6 | Biases in single-cell DA analysis: secondary analyses, false discoveries in random comparisons of published scATAC-seq data.

a, Number of DA peaks detected between randomly assigned replicates at 5% FDR within each cell type in random comparisons of published scATAC-seq data, for peaks divided into deciles by mean read depth.

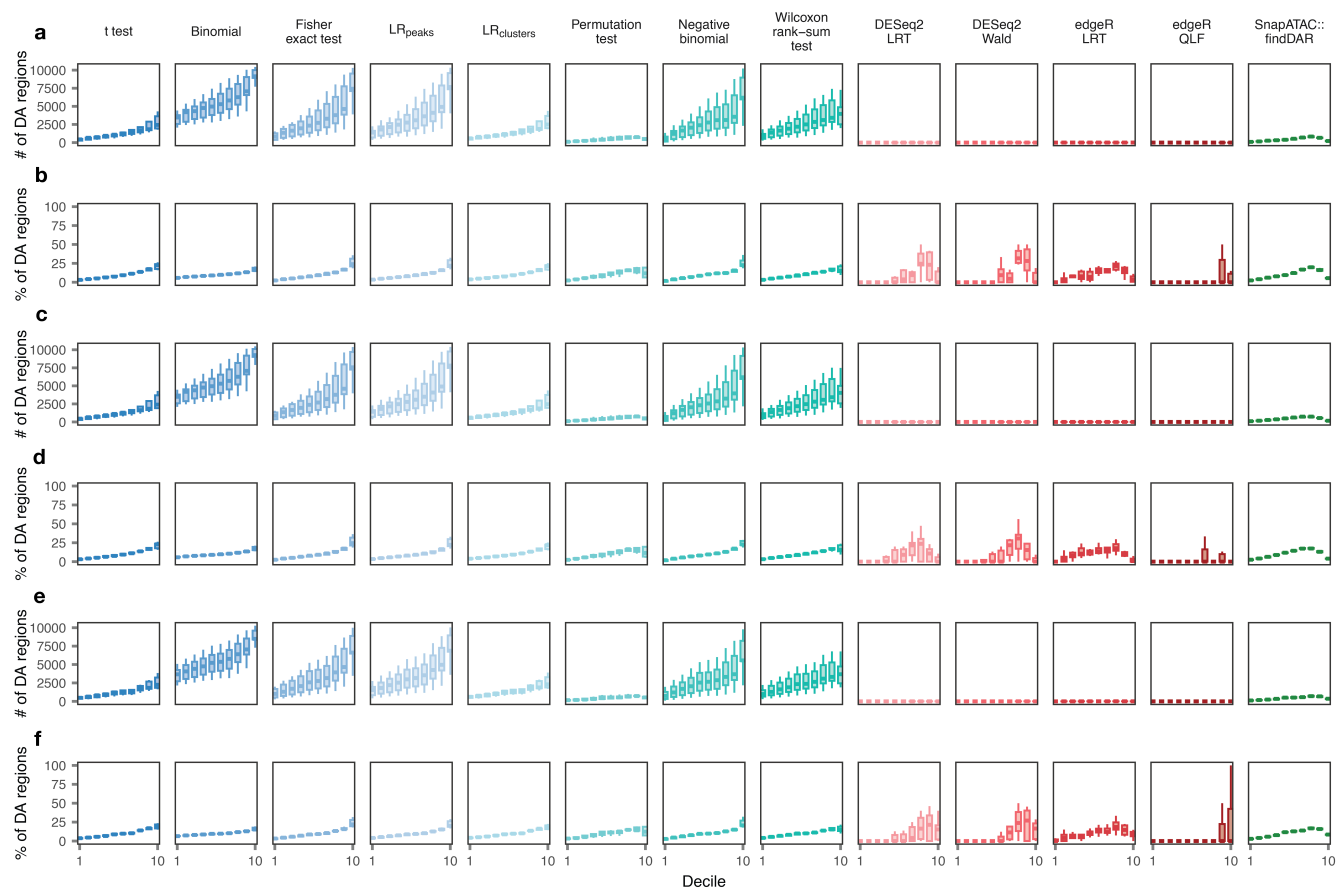
b, As in **a**, but showing the proportion of all DA peaks in each decile.

c, As in **a**, but for peaks divided into deciles by the proportion of cells in which peaks are open.

d, As in **c**, but showing the proportion of all DA peaks in each decile.

e, As in **a**, but for peaks divided into deciles by the width of each peak.

f, As in **e**, but showing the proportion of all DA peaks in each decile.



Supplementary Fig. 7 | Biases in single-cell DA analysis: secondary analyses, false discoveries in downsampled bulk ATAC-seq data.

a, Number of DA peaks detected at 5% FDR in DA analysis of downsampled bulk ATAC-seq libraries without biological differences between experimental conditions, for peaks divided into deciles by mean read depth.

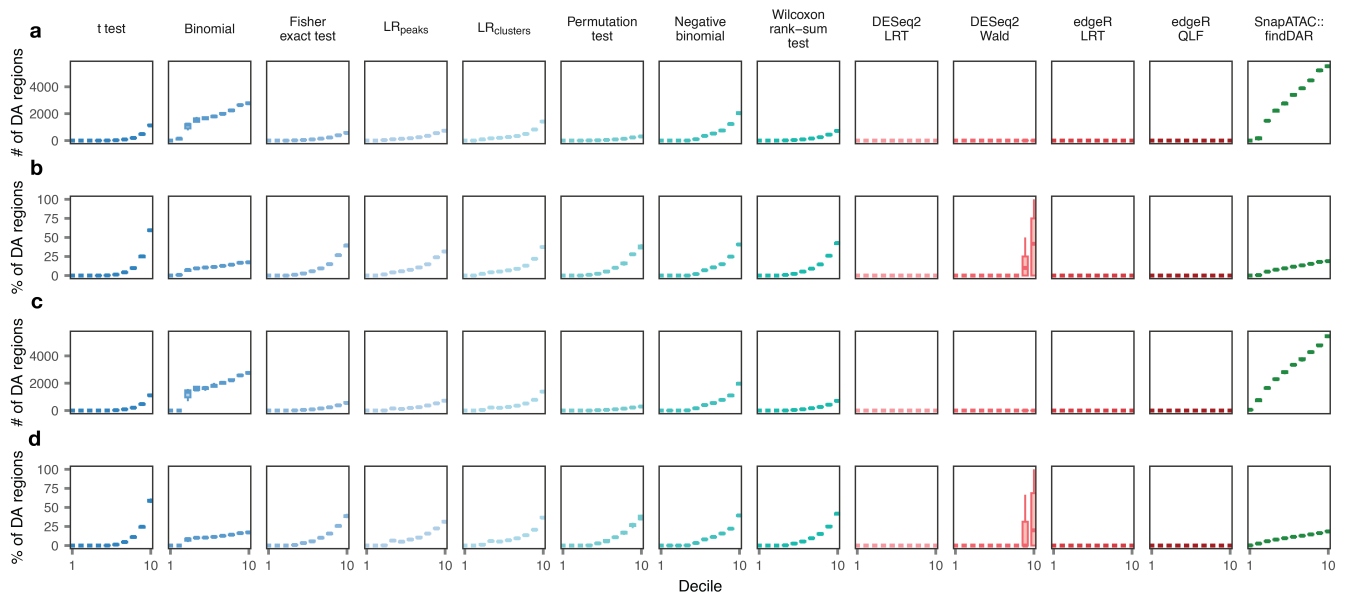
b, As in **a**, but showing the proportion of all DA peaks in each decile.

c, As in **a**, but for peaks divided into deciles by the proportion of cells in which peaks are open.

d, As in **c**, but showing the proportion of all DA peaks in each decile.

e, As in **a**, but for peaks divided into deciles by the width of each peak.

f, As in **e**, but showing the proportion of all DA peaks in each decile.



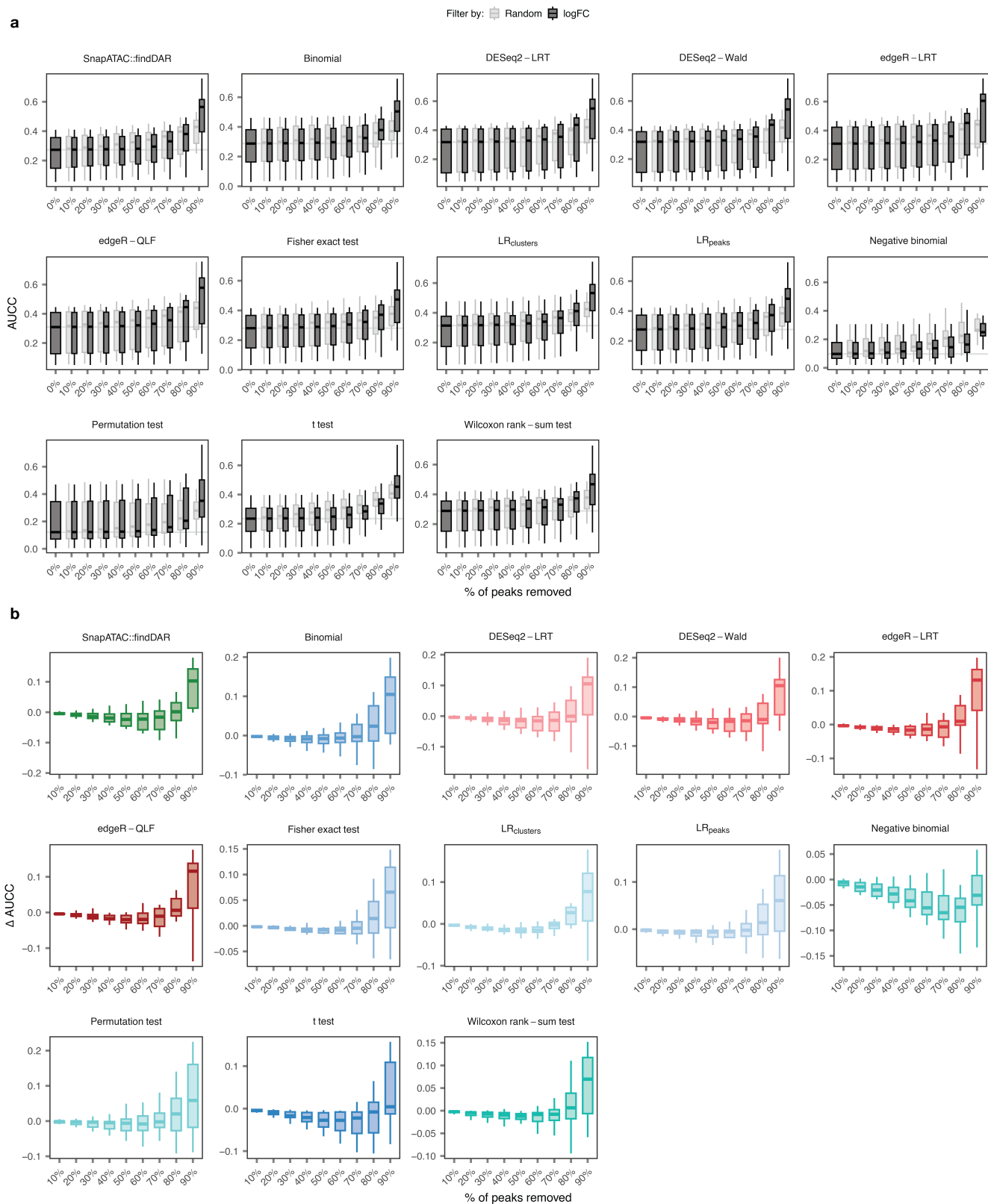
Supplementary Fig. 8 | Biases in single-cell DA analysis: secondary analyses, false discoveries in model-based simulations.

a, Number of DA peaks detected at 5% FDR in model-based simulations of scATAC-seq data without biological differences between experimental conditions, for peaks divided into deciles by mean read depth.

b, As in **a**, but showing the proportion of all DA peaks in each decile.

c, As in **a**, but for peaks divided into deciles by the proportion of cells in which peaks are open.

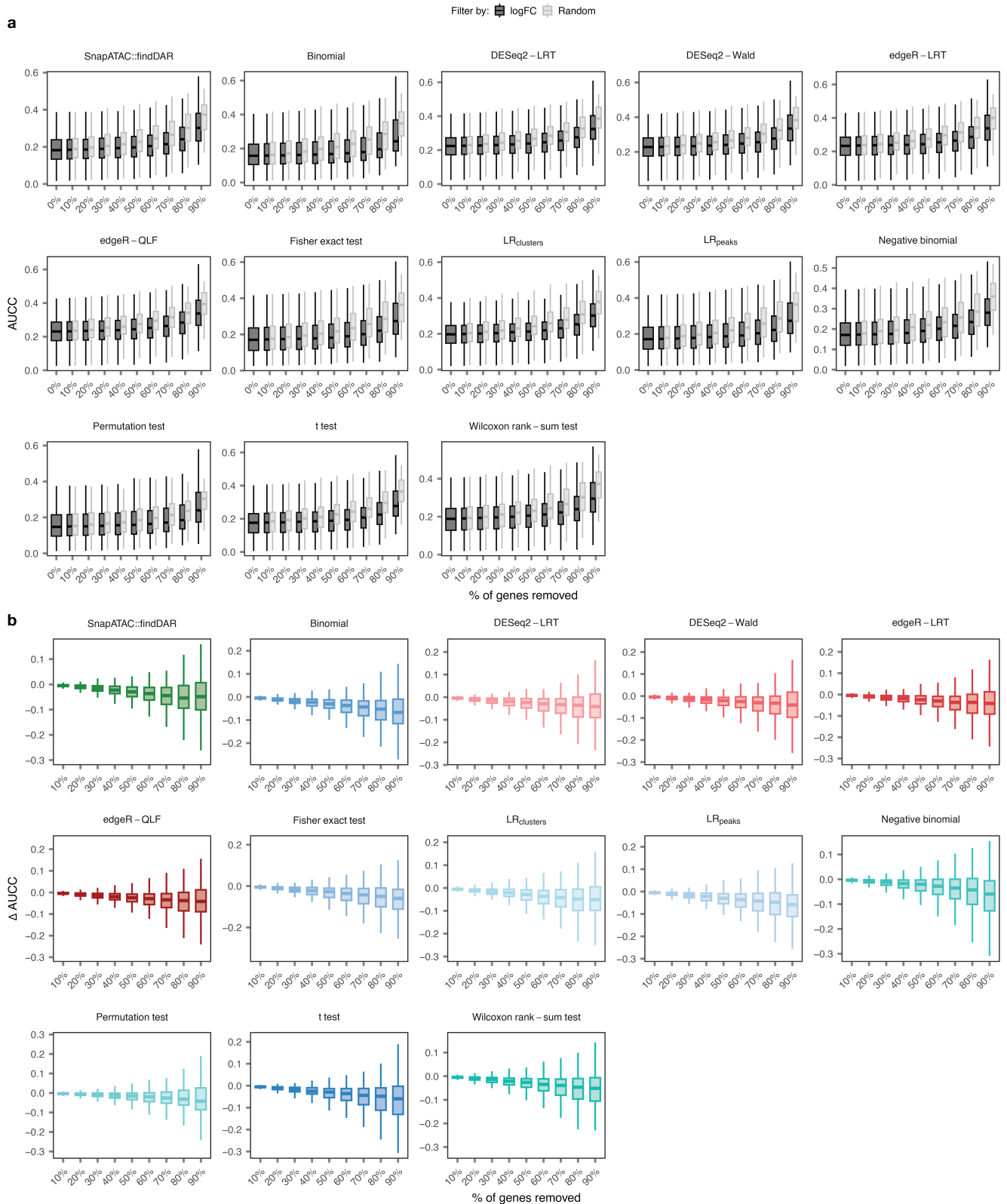
d, As in **c**, but showing the proportion of all DA peaks in each decile.

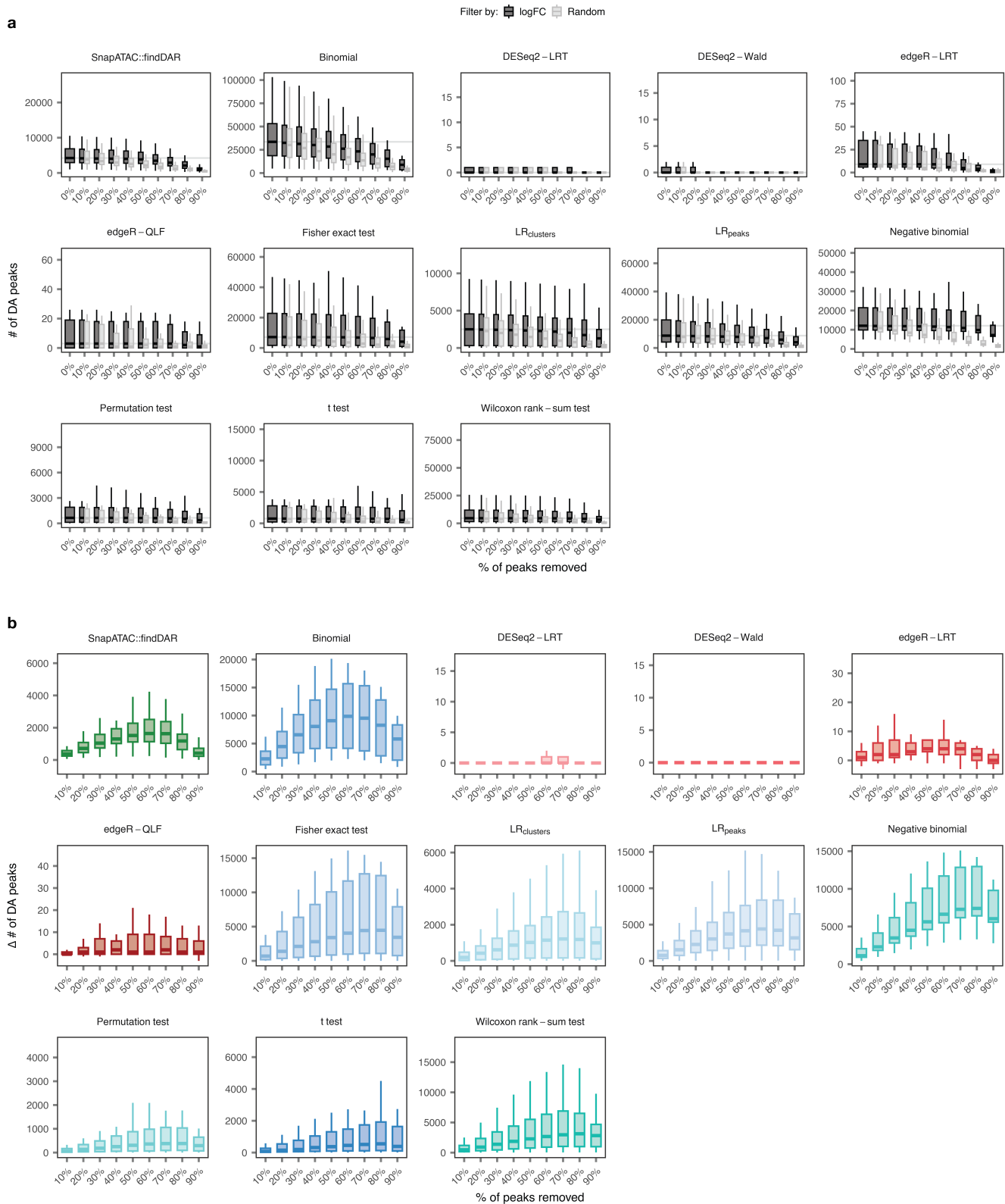


Supplementary Fig. 9 | Impact of log-fold change filtering on single-cell DA analysis: secondary analyses, matched bulk ATAC-seq.

a, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 1, using matching bulk ATAC-seq as a reference, after filtering peaks below a given quantile of log-fold change or an equivalent number of randomly selected peaks. Horizontal line shows the median AUC without any filtering.

b, As in **a**, but showing the difference in the AUC (Δ AUC) between filtering by log-fold change and filtering randomly selected peaks.

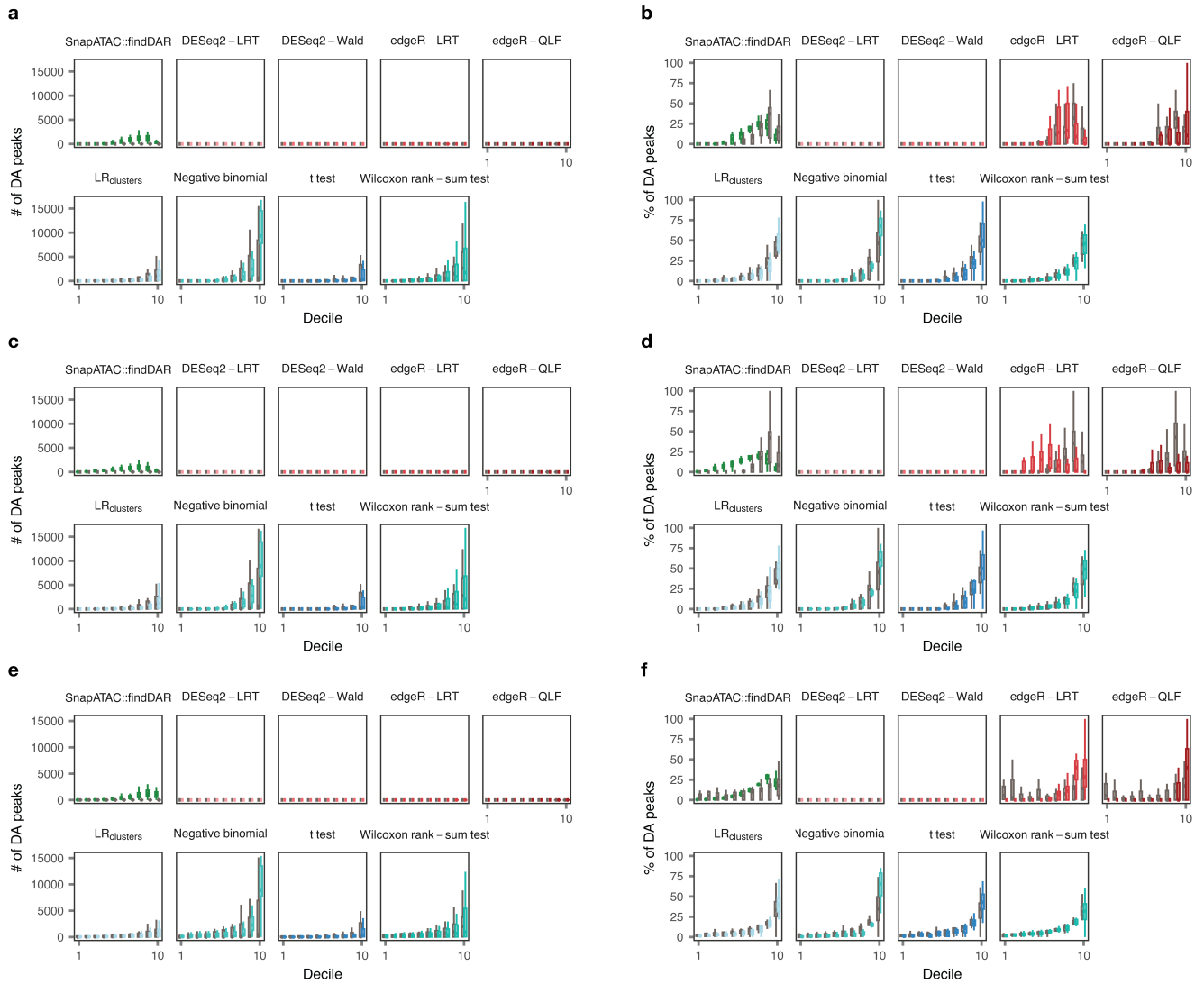




Supplementary Fig. 11 | Impact of log-fold change filtering on single-cell DA analysis: secondary analyses, false discoveries.

a, Number of DA peaks detected between randomly assigned replicates at 5% FDR within each cell type in random comparisons of published scATAC-seq data, after filtering peaks below a given quantile of log-fold change or an equivalent number of randomly selected peaks. Horizontal line shows the median number of DA peaks without any filtering.

b, As in **a**, but showing the difference in the number of DA peaks between filtering by log-fold change and filtering randomly selected peaks.



Supplementary Fig. 14 | Impact of binarization on single-cell DA analysis: false discoveries in random comparisons of published scATAC-seq data.

a, Number of DA peaks detected between randomly assigned replicates at 5% FDR within each cell type in random comparisons of published scATAC-seq data, for peaks divided into deciles by mean read depth, before and after binarization.

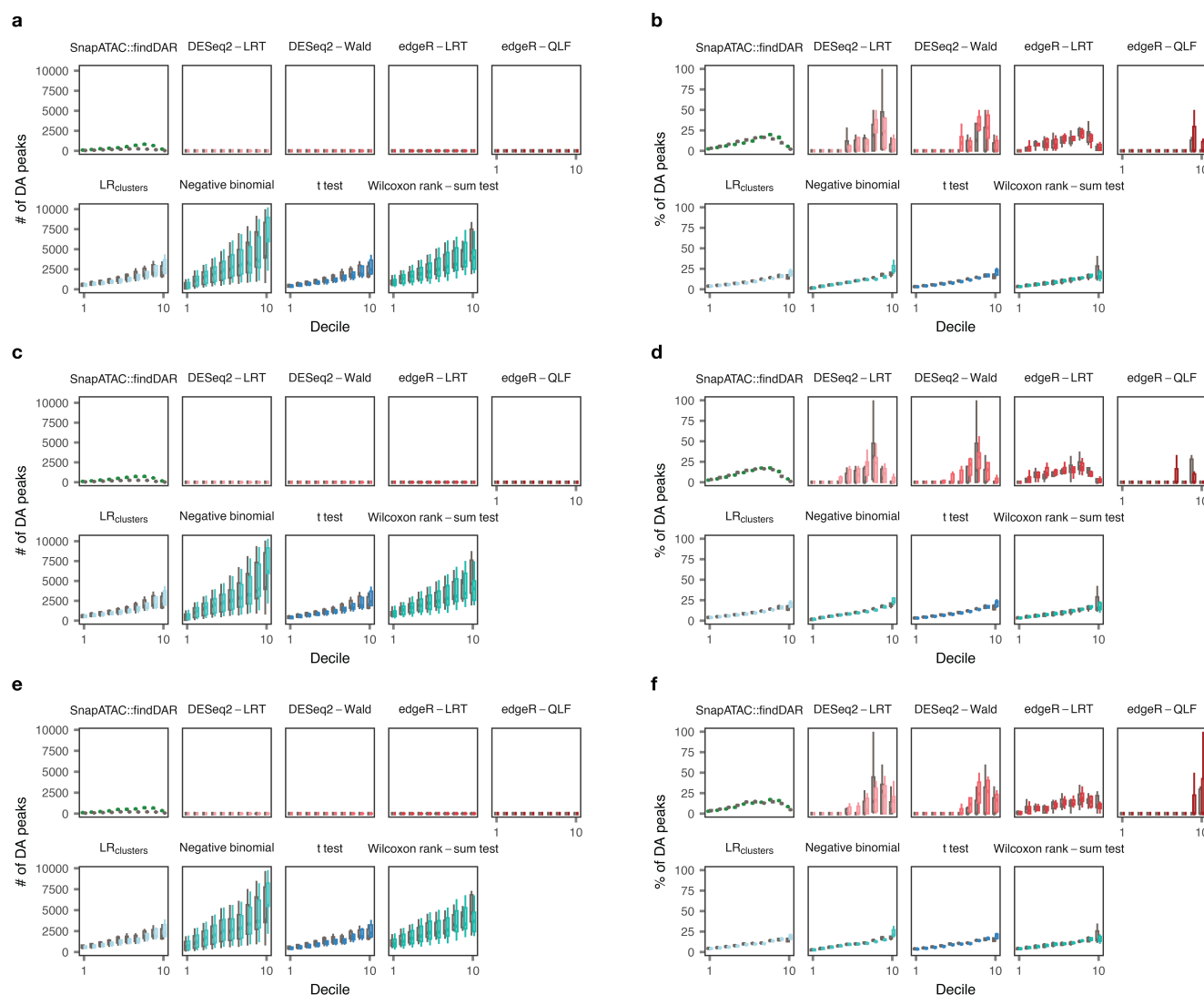
b, As in **a**, but showing the proportion of all DA peaks in each decile.

c, As in **a**, but for peaks divided into deciles by the proportion of cells in which peaks are open.

d, As in **c**, but showing the proportion of all DA peaks in each decile.

e, As in **a**, but for peaks divided into deciles by the width of each peak.

f, As in **e**, but showing the proportion of all DA peaks in each decile.



Supplementary Fig. 15 | Impact of binarization on single-cell DA analysis: false discoveries in downsampled bulk ATAC-seq data.

a, Number of DA peaks detected at 5% FDR in DA analysis of downsampled bulk ATAC-seq libraries without biological differences between experimental conditions, for peaks divided into deciles by mean read depth, before and after binarization.

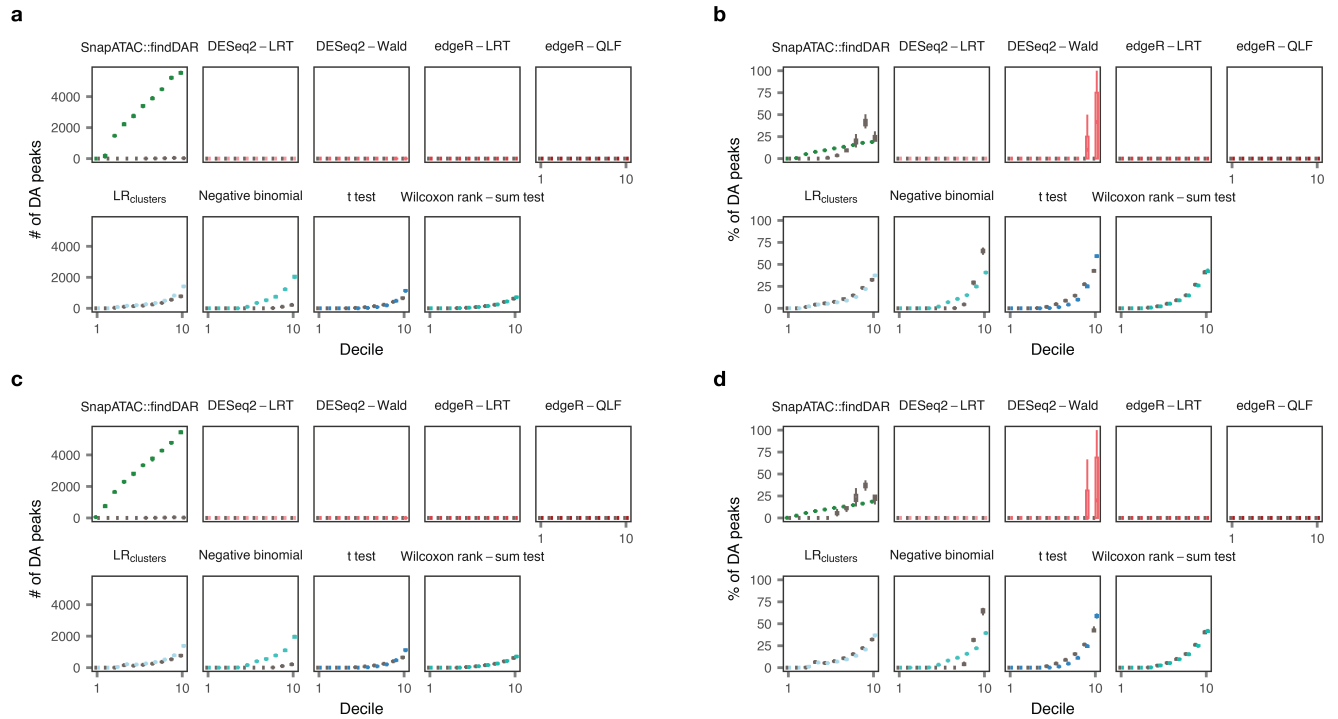
b, As in **a**, but showing the proportion of all DA peaks in each decile.

c, As in **a**, but for peaks divided into deciles by the proportion of cells in which peaks are open.

d, As in **c**, but showing the proportion of all DA peaks in each decile.

e, As in **a**, but for peaks divided into deciles by the width of each peak.

f, As in **e**, but showing the proportion of all DA peaks in each decile.



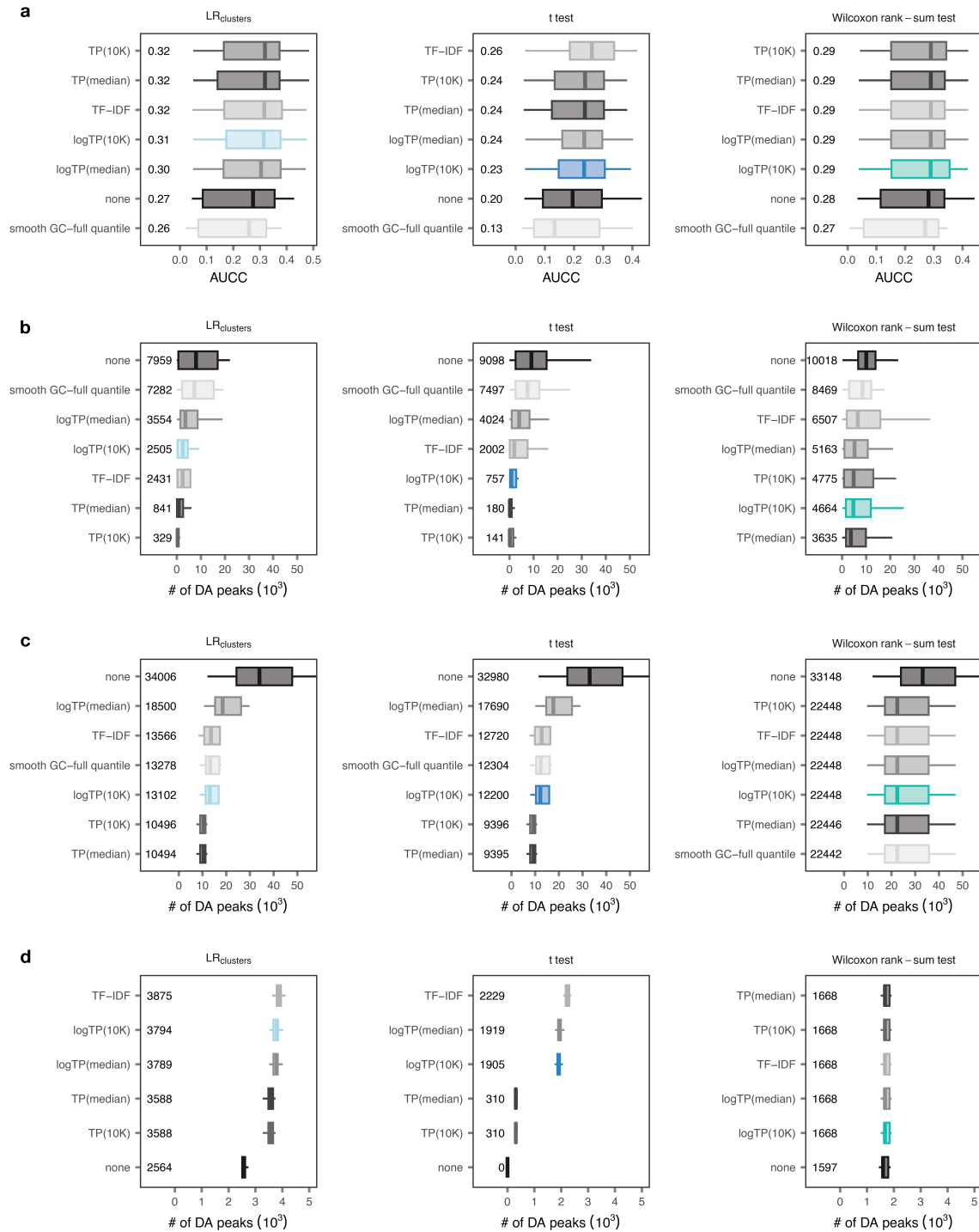
Supplementary Fig. 16 | Impact of binarization on single-cell DA analysis: false discoveries in model-based simulations.

a, Number of DA peaks detected at 5% FDR in model-based simulations of scATAC-seq data without biological differences between experimental conditions, for peaks divided into deciles by mean read depth, before and after binarization.

b, As in **a**, but showing the proportion of all DA peaks in each decile.

c, As in **a**, but for peaks divided into deciles by the proportion of cells in which peaks are open.

d, As in **c**, but showing the proportion of all DA peaks in each decile.



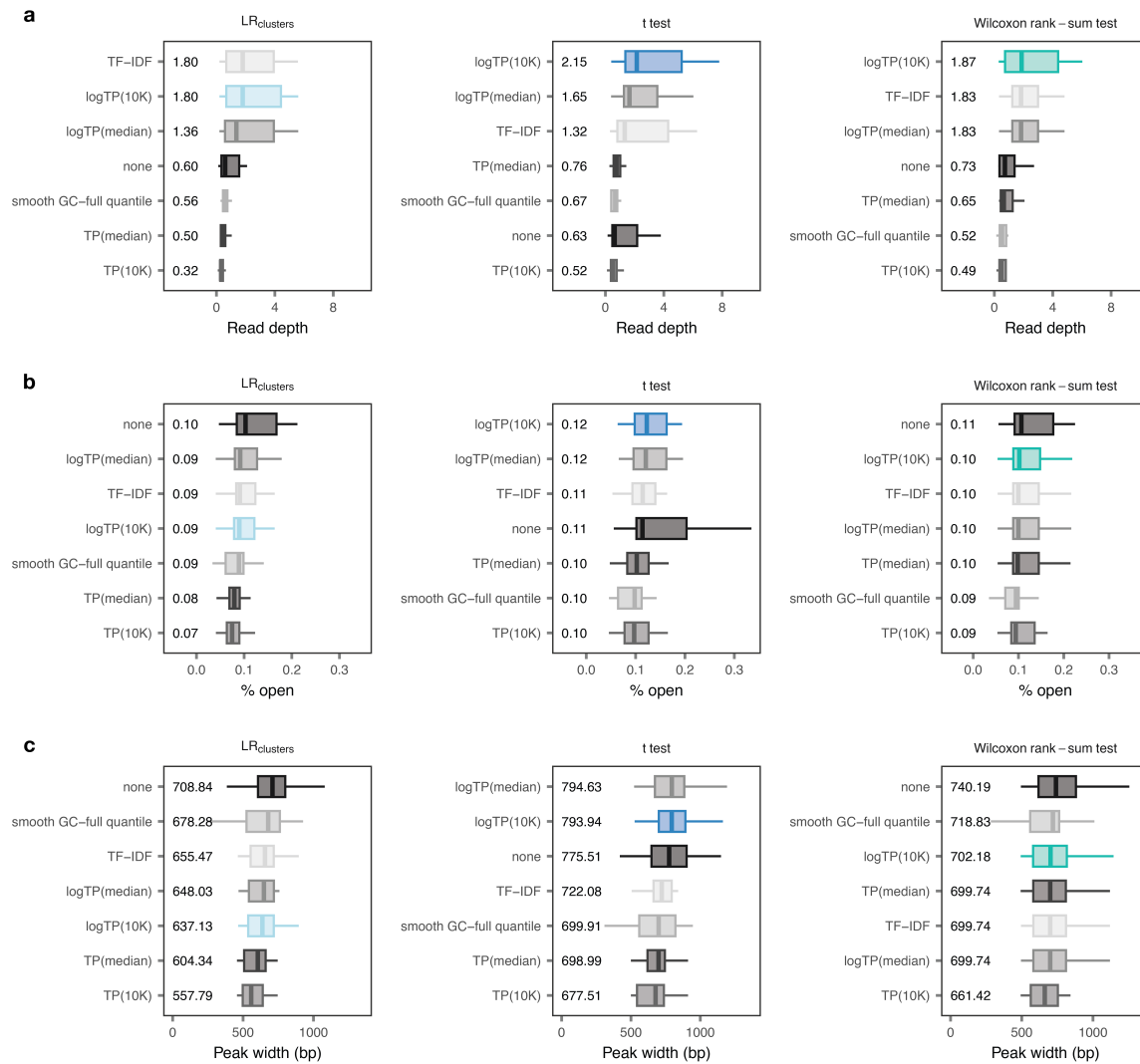
Supplementary Fig. 17 | Normalization approaches to scATAC-seq data: concordance with matched bulk ATAC-seq and false discoveries.

a, Area under the concordance curve (AUCC) for single-cell DA methods in Experiment 1, using matching bulk ATAC-seq as a reference, after applying different approaches to normalize the scATAC-seq data. Inset text shows the median AUCC.

b, Number of DA peaks detected between randomly assigned replicates at 5% FDR within each cell type in random comparisons of published scATAC-seq data, after applying different approaches to normalize the scATAC-seq data. Inset text shows the median number of DA peaks.

c, As in **b**, but for downsampled bulk ATAC-seq libraries without biological differences between experimental conditions.

d, As in **b**, but for model-based simulations of scATAC-seq data without biological differences between experimental conditions.

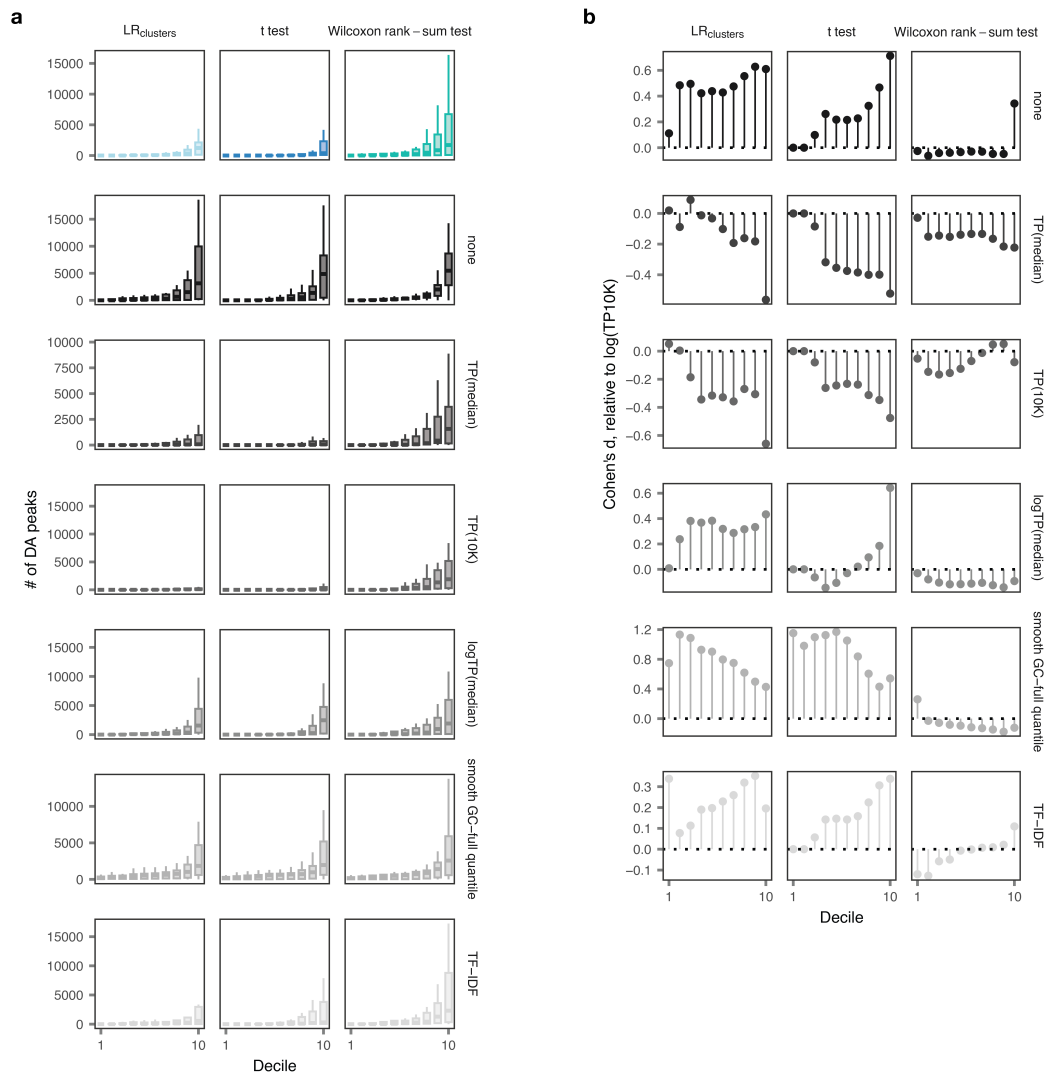


Supplementary Fig. 18 | Normalization approaches to scATAC-seq data: properties of DA peaks.

a, Mean read depth of the top-1,000 DA peaks identified by each single-cell DA method in published scATAC-seq datasets, after applying different approaches to normalize the scATAC-seq data. Inset text shows the median across comparisons.

b, As in **a**, but showing the proportion of cells in which the top-1,000 DA peaks are open.

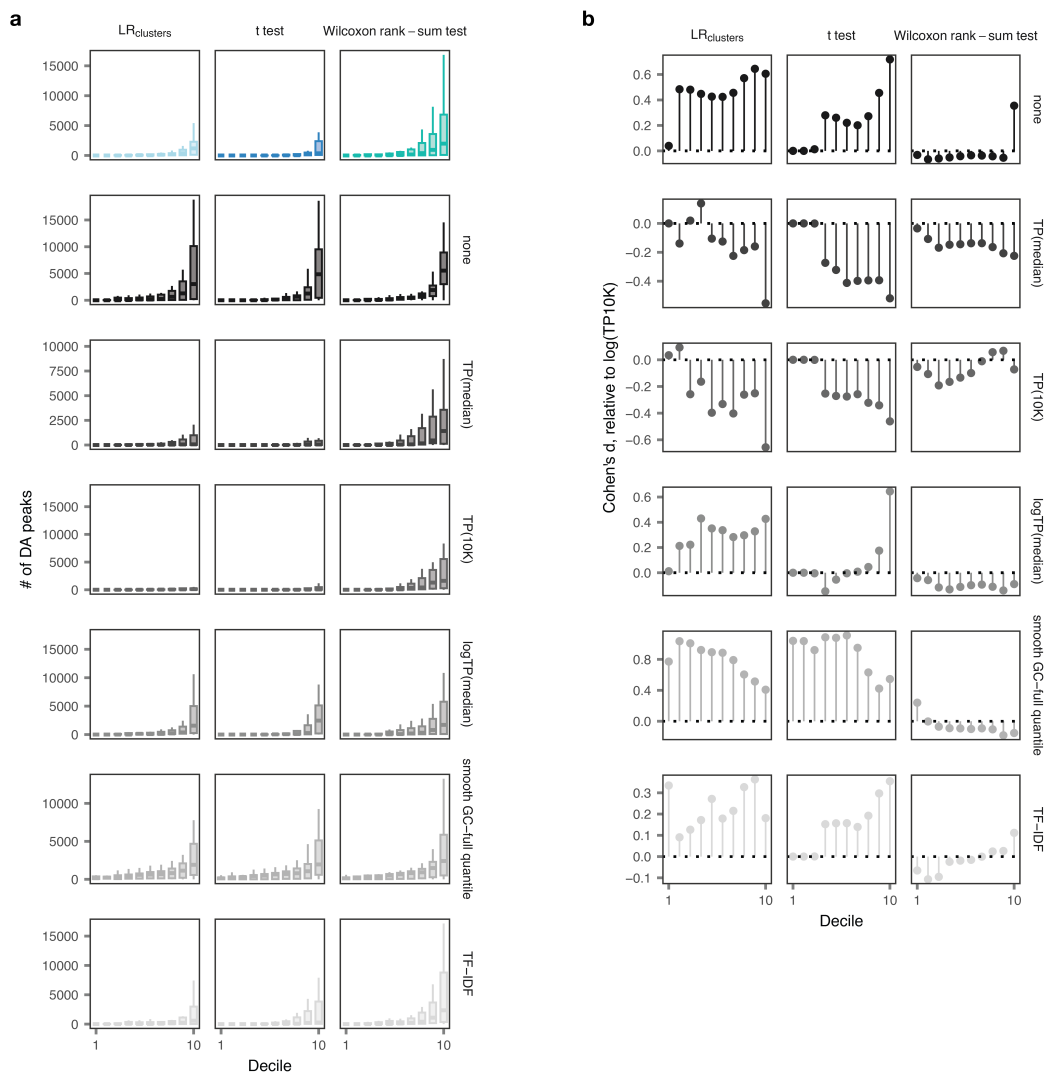
c, As in **a**, but for the width of the top-1,000 DA peaks.



Supplementary Fig. 19 | Normalization approaches to scATAC-seq data: false discoveries in random comparisons of published scATAC-seq data, by read depth.

a, Number of DA peaks detected between randomly assigned replicates at 5% FDR within each cell type in random comparisons of published scATAC-seq data, for peaks divided into deciles by mean read depth, after applying different approaches to normalize the scATAC-seq data.

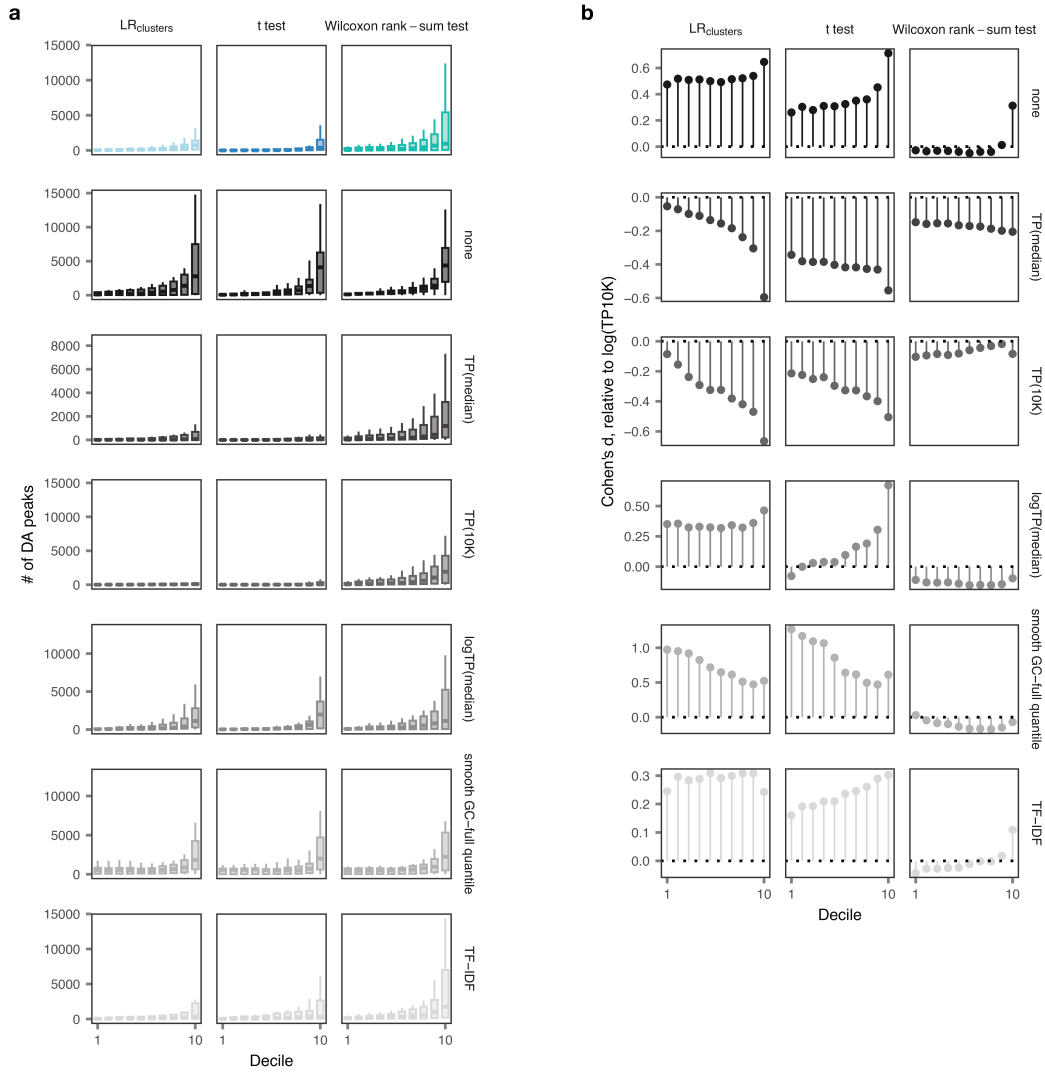
b, Effect size (Cohen's d) of different approaches to normalization of scATAC-seq data on the number of DA peaks per decile.



Supplementary Fig. 20 | Normalization approaches to scATAC-seq data: false discoveries in random comparisons of published scATAC-seq data, by accessibility across cells.

a, Number of DA peaks detected between randomly assigned replicates at 5% FDR within each cell type in random comparisons of published scATAC-seq data, for peaks divided into deciles by the proportion of cells in which each peak is accessible, after applying different approaches to normalize the scATAC-seq data.

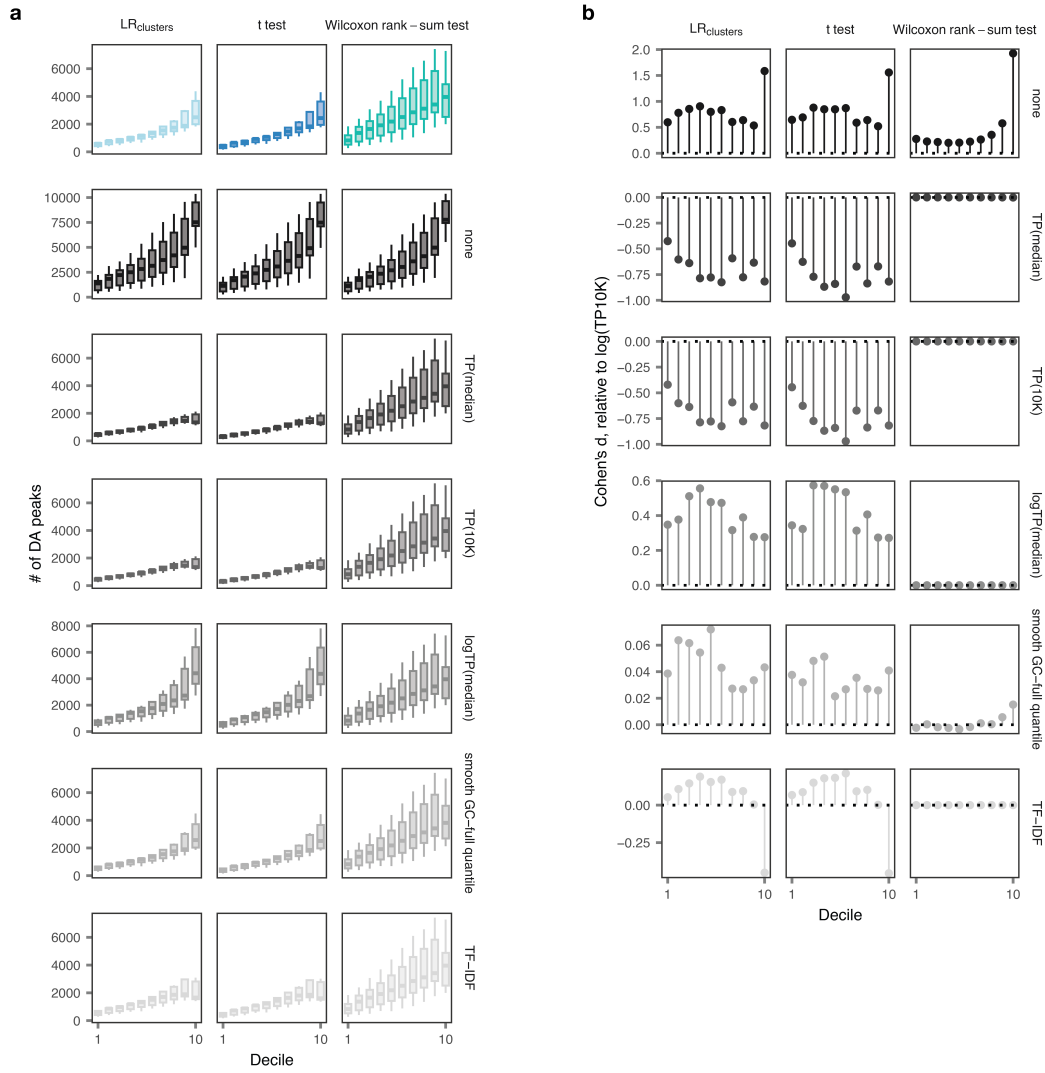
b, Effect size (Cohen's d) of different approaches to normalization of scATAC-seq data on the number of DA peaks per decile.



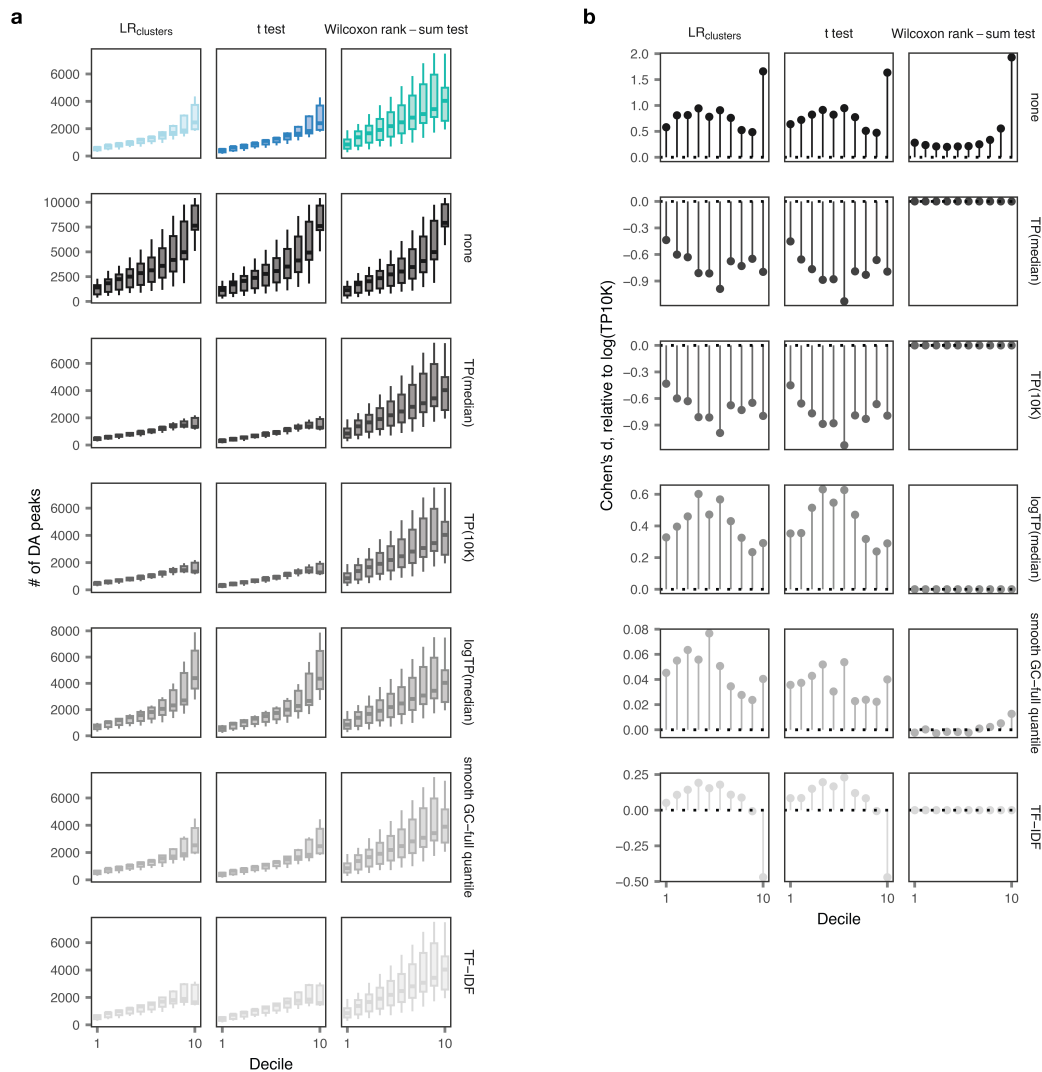
Supplementary Fig. 21 | Normalization approaches to scATAC-seq data: false discoveries in random comparisons of published scATAC-seq data, by peak width.

a, Number of DA peaks detected between randomly assigned replicates at 5% FDR within each cell type in random comparisons of published scATAC-seq data, for peaks divided into deciles by width, after applying different approaches to normalize the scATAC-seq data.

b, Effect size (Cohen's d) of different approaches to normalization of scATAC-seq data on the number of DA peaks per decile.



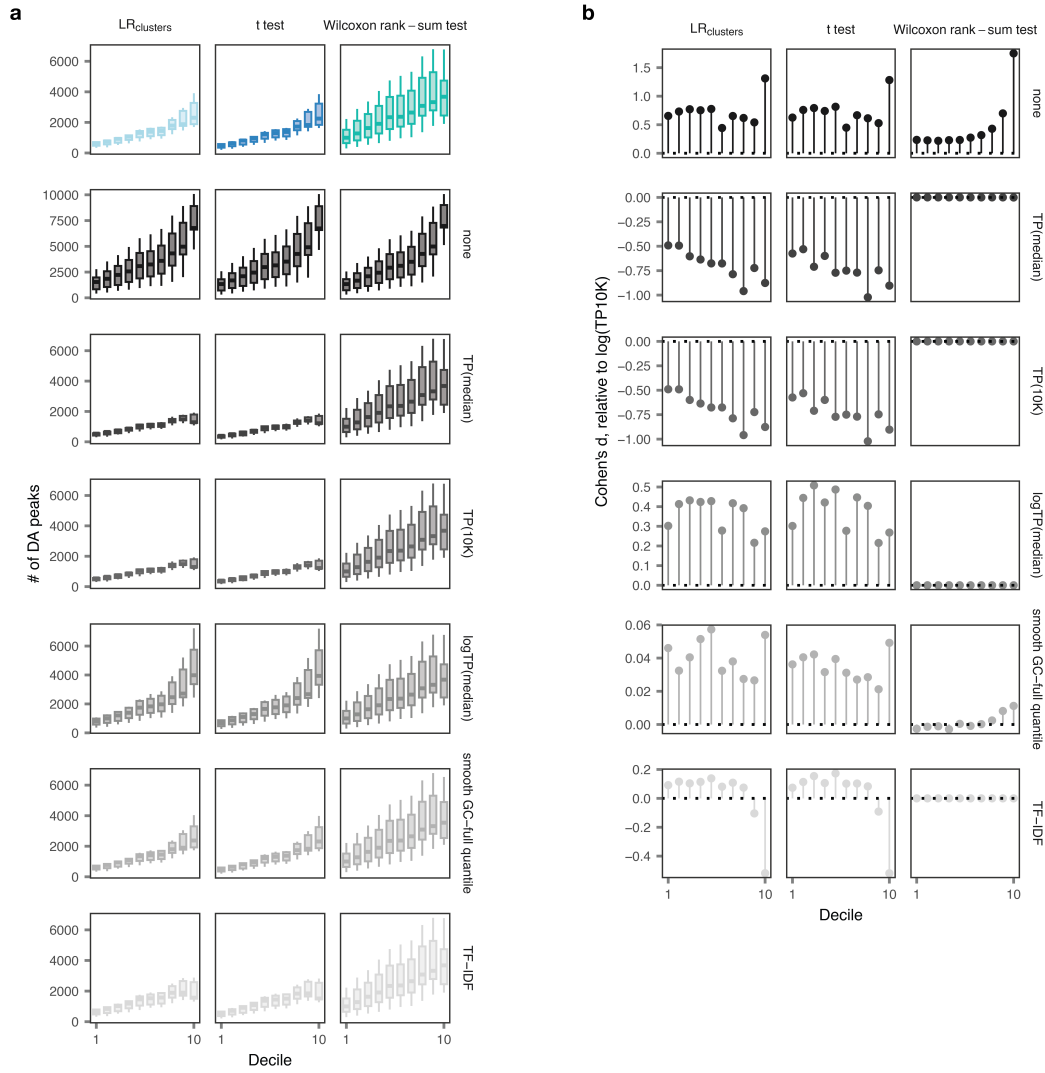
Supplementary Fig. 22 | Normalization approaches to scATAC-seq data: false discoveries in downsampled bulk ATAC-seq data, by read depth.
a, Number of DA peaks detected at 5% FDR in DA analysis of downsampled bulk ATAC-seq libraries without biological differences between experimental conditions, for peaks divided into deciles by mean read depth, after applying different approaches to normalize the scATAC-seq data.
b, Effect size (Cohen's *d*) of different approaches to normalization of scATAC-seq data on the number of DA peaks per decile.



Supplementary Fig. 23 | Normalization approaches to scATAC-seq data: false discoveries in downsampled bulk ATAC-seq data, by accessibility across cells.

a, Number of DA peaks detected at 5% FDR in DA analysis of downsampled bulk ATAC-seq libraries without biological differences between experimental conditions, for peaks divided into deciles by the proportion of cells in which each peak is accessible, after applying different approaches to normalize the scATAC-seq data.

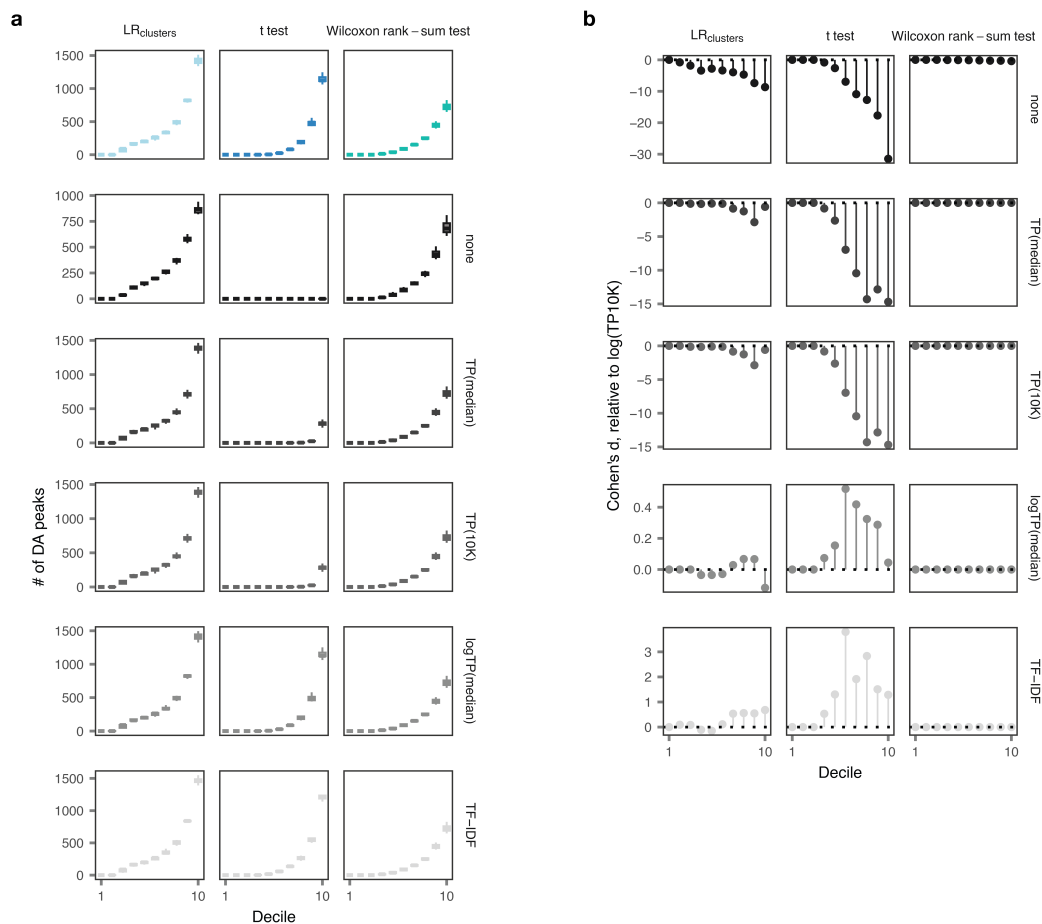
b, Effect size (Cohen's d) of different approaches to normalization of scATAC-seq data on the number of DA peaks per decile.



Supplementary Fig. 24 | Normalization approaches to scATAC-seq data: false discoveries in downsampled bulk ATAC-seq data, by peak width.

a, Number of DA peaks detected at 5% FDR in DA analysis of downsampled bulk ATAC-seq libraries without biological differences between experimental conditions, for peaks divided into deciles by width, after applying different approaches to normalize the scATAC-seq data.

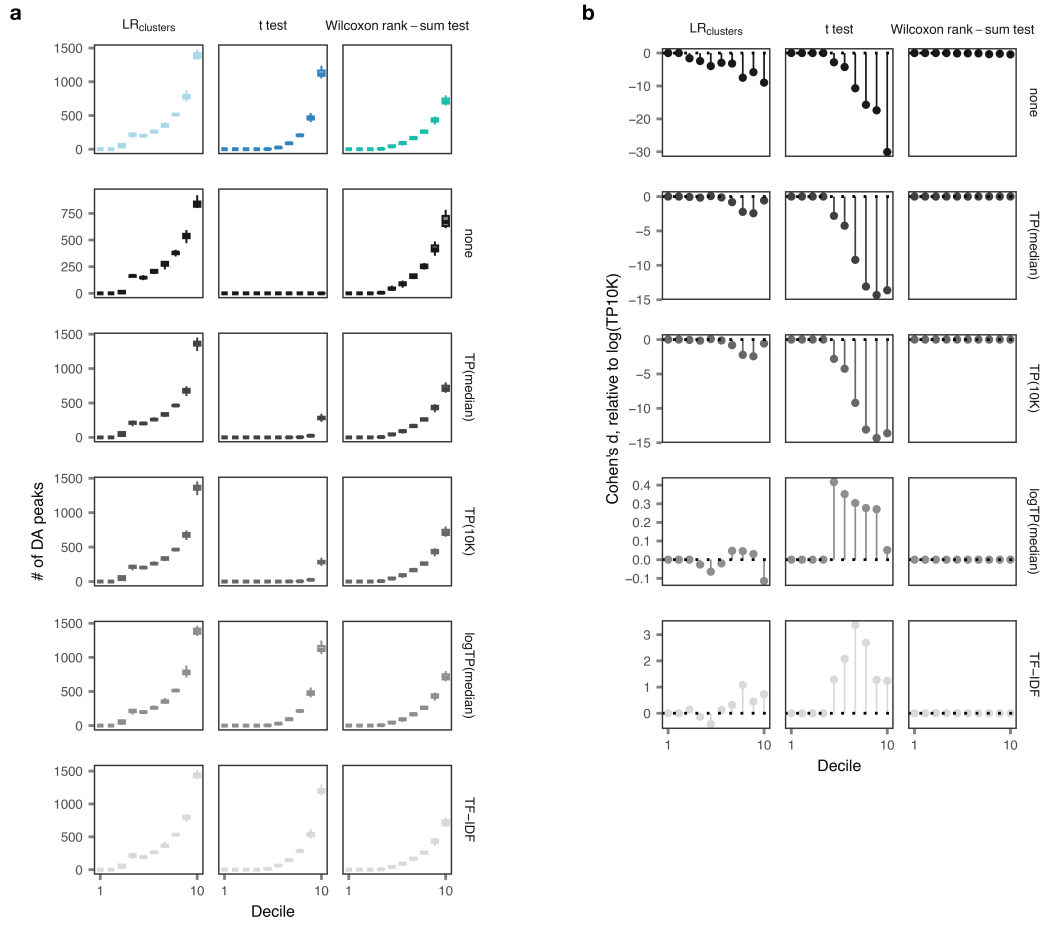
b, Effect size (Cohen's d) of different approaches to normalization of scATAC-seq data on the number of DA peaks per decile.



Supplementary Fig. 25 | Normalization approaches to scATAC-seq data: false discoveries in model-based simulations, by read depth.

a, Number of DA peaks detected at 5% FDR in model-based simulations of scATAC-seq data without biological differences between experimental conditions, for peaks divided into deciles by mean read depth, after applying different approaches to normalize the scATAC-seq data.

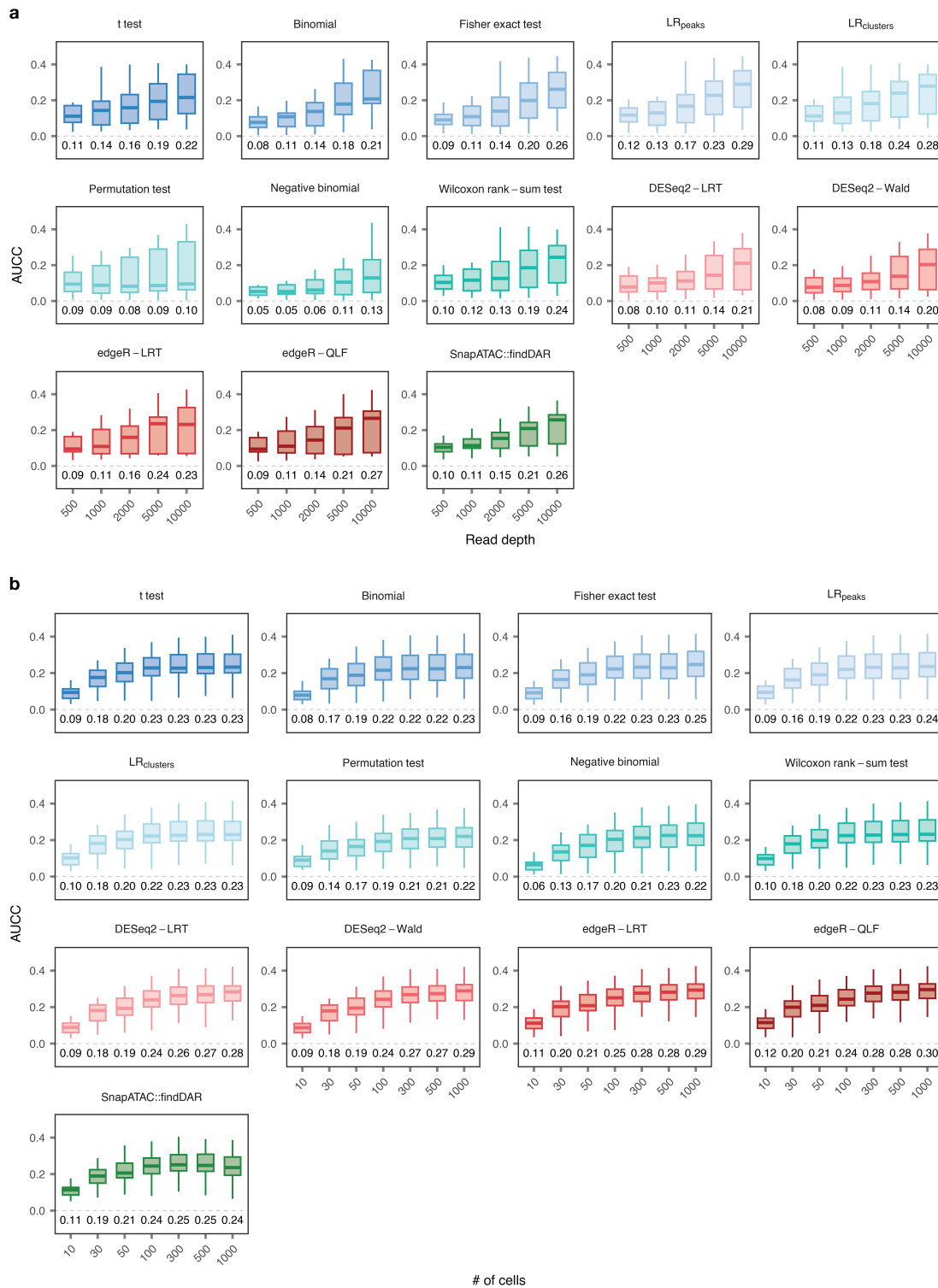
b, Effect size (Cohen's *d*) of different approaches to normalization of scATAC-seq data on the number of DA peaks per decile.



Supplementary Fig. 26 | Normalization approaches to scATAC-seq data: false discoveries in model-based simulations, by accessibility across cells.

a, Number of DA peaks detected at 5% FDR in model-based simulations of scATAC-seq data without biological differences between experimental conditions, for peaks divided into deciles by the proportion of cells in which each peak is accessible, after applying different approaches to normalize the scATAC-seq data.

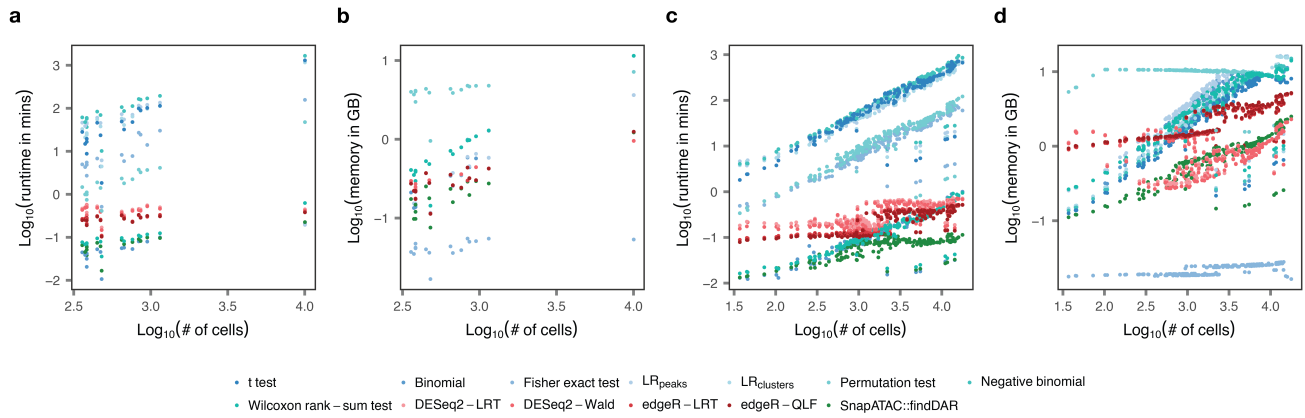
b, Effect size (Cohen's d) of different approaches to normalization of scATAC-seq data on the number of DA peaks per decile.



Supplementary Fig. 27 | Data requirements for single-cell DA analysis.

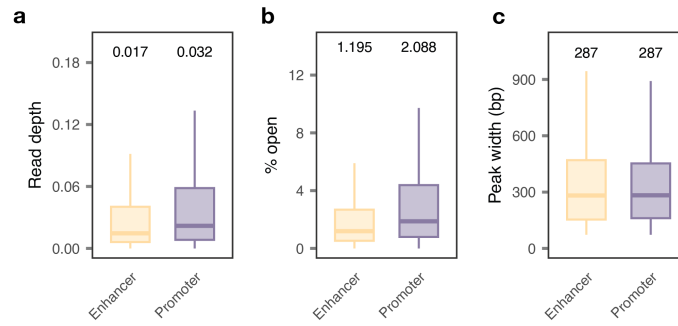
a, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 1, using matching bulk ATAC-seq as a reference, after down-sampling plate-based scATAC-seq data to a mean of 500, 1,000, 2,000, 5,000 counts per cell. Inset text shows the median AUC.

b, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 2, using matching snRNA-seq as a reference, after downsampling single-cell multi-omics data to 20, 50, 100, 200, or 500 cells per condition. Inset text shows the median AUC.

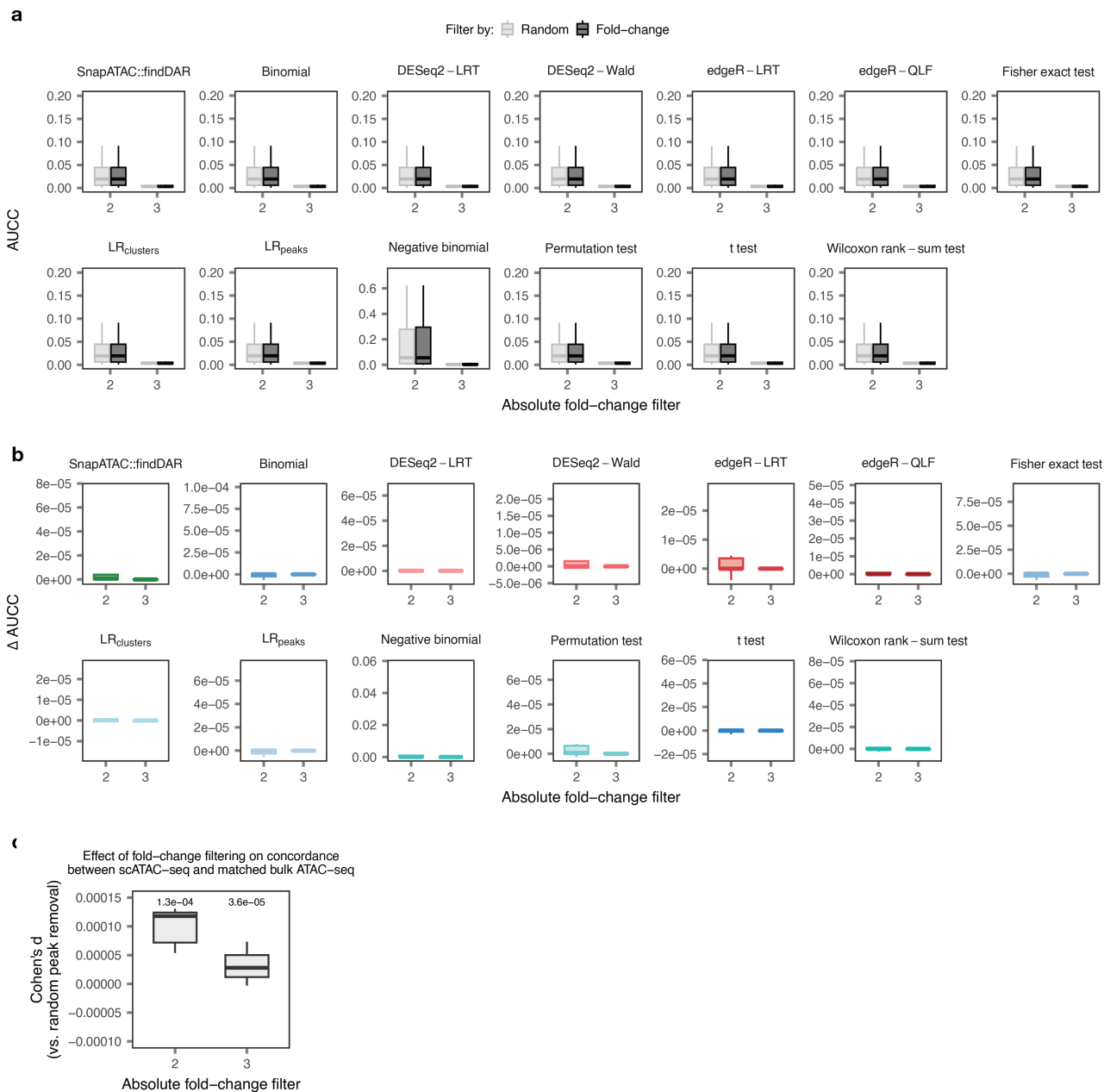


Supplementary Fig. 28 | Scalability of single-cell DA methods: secondary analyses.

- a**, Relationship between wall clock time and the number of cells per comparison in Experiment 1.
b, Relationship between peak memory usage and the number of cells per comparison in Experiment 1.
c, As in **a**, but for each comparison in Experiment 2.
d, As in **b**, but for each comparison in Experiment 2.



Supplementary Fig. 29 | Exploratory analysis: technical characteristics of promoter vs. enhancer regions.
a, Read depth; **b**, accessibility across cells; **c**, peak width in base pairs.

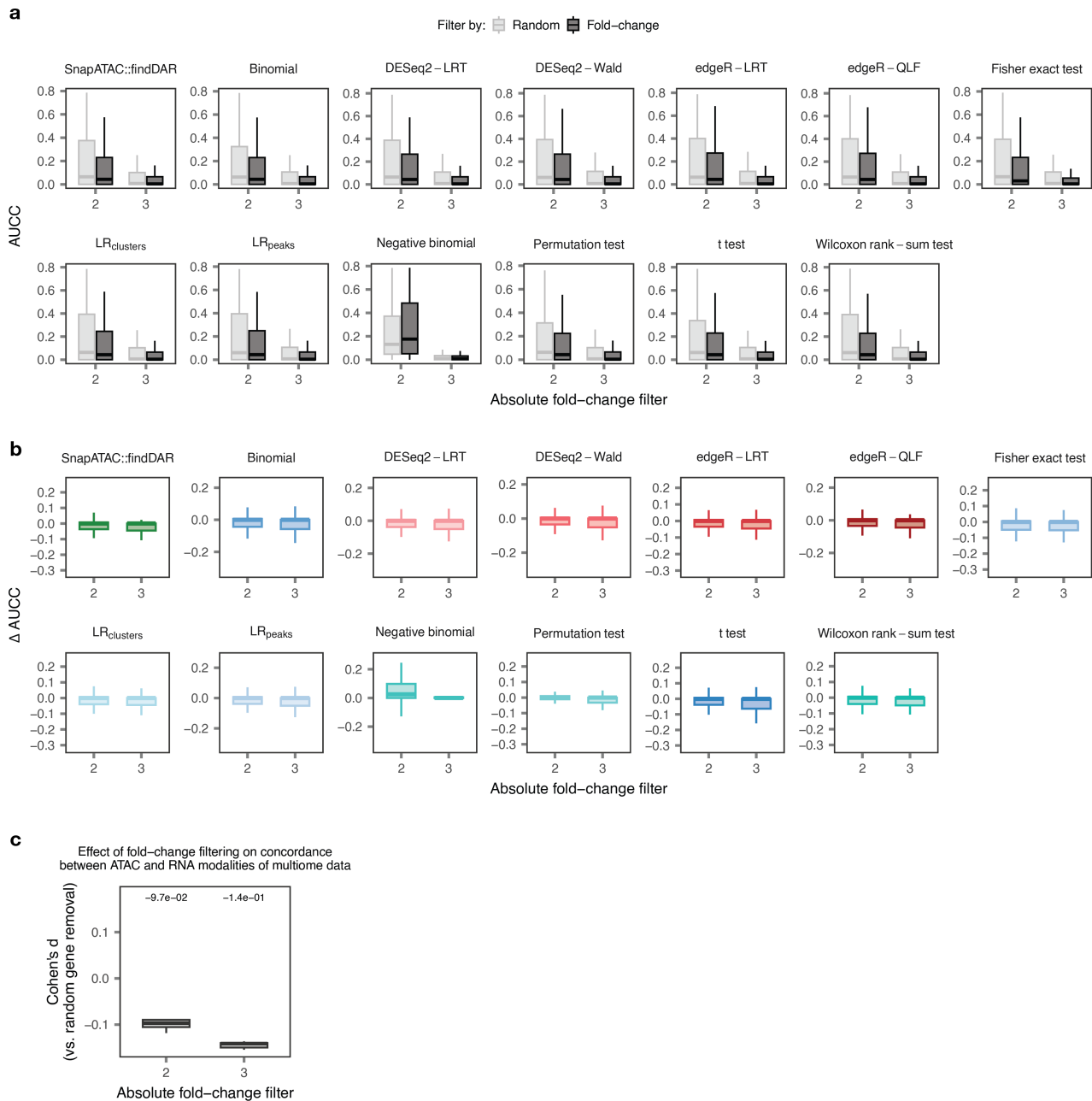


Supplementary Fig. 30 | Exploratory analysis: impact of fold-change filtering on single-cell DA analysis with fixed thresholds, matched bulk ATAC-seq.

a, Area under the concordance curve (AUC) for single-cell DA methods in Experiment 1, using matching bulk ATAC-seq as a reference, after filtering peaks with less than a two- or three-fold change between conditions or an equivalent number of randomly selected peaks.

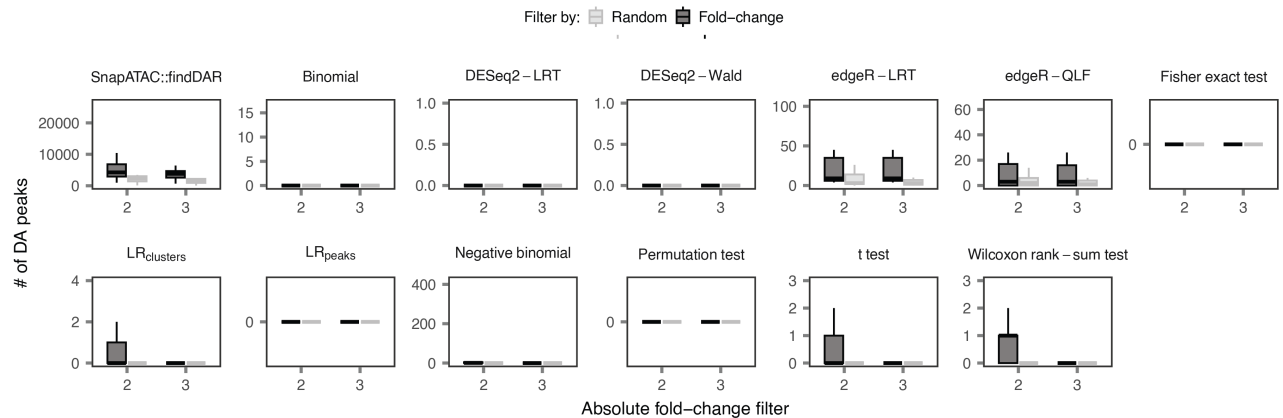
b, As in **a**, but showing the difference in the AUC (Δ AUC) between filtering by fold change and filtering randomly selected peaks.

c, Effect size (Cohen's d) of increasingly stringent fold change filtering on the AUC, relative to the removal of an equivalent number of peaks selected at random. Inset text shows the median Cohen's d .

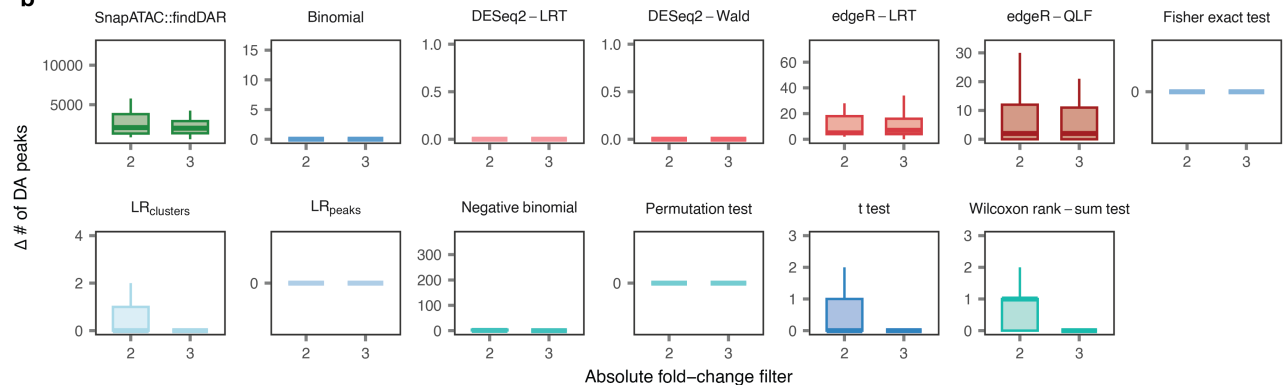


Supplementary Fig. 31 | Exploratory analysis: impact of fold-change filtering on single-cell DA analysis with fixed thresholds, multiome.
a, Area under the concordance curve (AUCC) for single-cell DA methods in Experiment 2, using matching snRNA-seq as a reference, after filtering regions with less than a two- or three-fold change between conditions or an equivalent number of randomly selected regions.
b, As in **a**, but showing the difference in the AUCC (Δ AUCC) between filtering by fold change and filtering randomly selected regions.
c, Effect size (Cohen's d) of increasingly stringent fold change filtering on the AUCC, relative to the removal of an equivalent number of regions selected at random. Inset text shows the median Cohen's d .

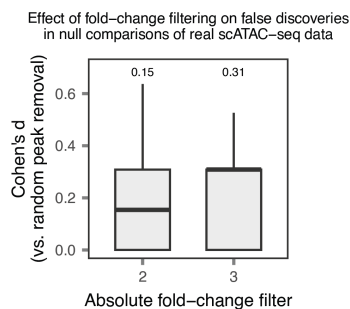
a



b



c



Supplementary Fig. 32 | Exploratory analysis: impact of fold-change filtering on single-cell DA analysis with fixed thresholds, false discoveries.

a, Number of DA peaks detected between randomly assigned replicates at 5% FDR within each cell type in random comparisons of published scATAC-seq data, after filtering peaks with less than an two- or three-fold change between conditions or an equivalent number of randomly selected peaks.

b, As in **a**, but showing the difference in the number of DA peaks between filtering by fold change and filtering randomly selected peaks.

c, Effect size (Cohen's *d*) of increasingly stringent fold change filtering on the number of false discoveries in null comparisons of published scATAC-seq data, relative to the removal of an equivalent number of regions selected at random. Inset text shows the median Cohen's *d*.



Supplementary Fig. 33 | Summary of DA method performance across Experiments 1 and 2, including primary and all sensitivity analyses. Methods were grouped into terciles of high, low, or intermediate performance on the basis of their quantitative performance on each task, as in Fig. 9, and ranked by their average performance across all primary and secondary analyses shown in Fig. 2 and Supplementary Figs. 1-3.



Supplementary Fig. 34 | Summary of DA method performance across Experiment 6, including all binarization and normalization strategies.
a, Methods were grouped into terciles of high, low, or intermediate performance on the basis of their quantitative performance on each task, as in Fig. 9, separately across each of the individual binarization and normalization scenarios evaluated in the study.
b, As in **a**, but showing only the single top-performing scenario for each DA method.