Additional file 1: Supplementary figures

Correcting batch effects in large-scale multiomics studies using a reference-material-based ratio method
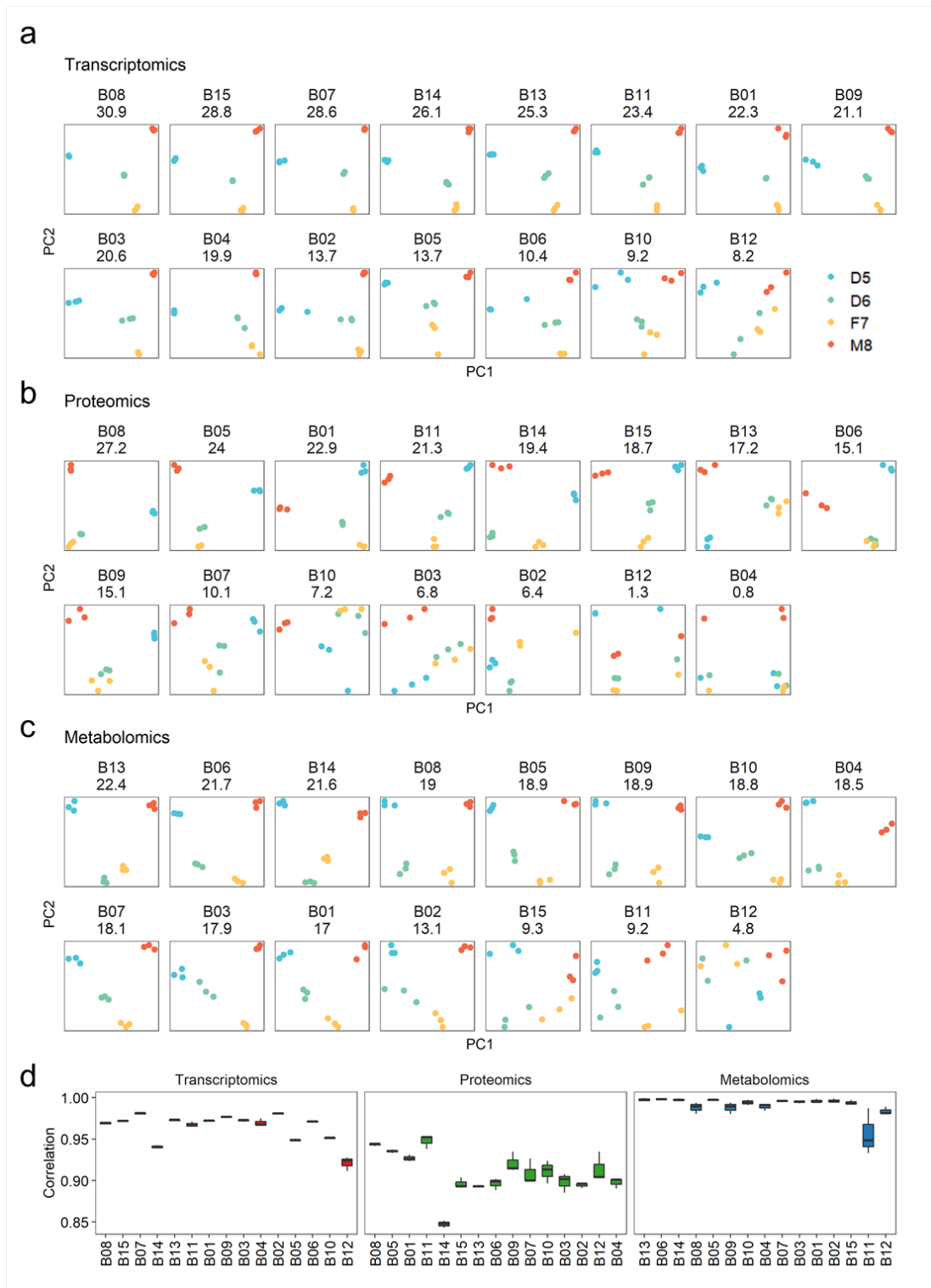
Yu Y et al.

Genome Biology

Fig. S1 | Diversity of quality of original datasets.

(a-c) PCA plots per batch ordered by the SNR value for datasets in transcriptomics (a), proteomics (b) and metabolomics (c). (d). Pearson correlation coefficient of replicates of reference sample (D6) in each batch.
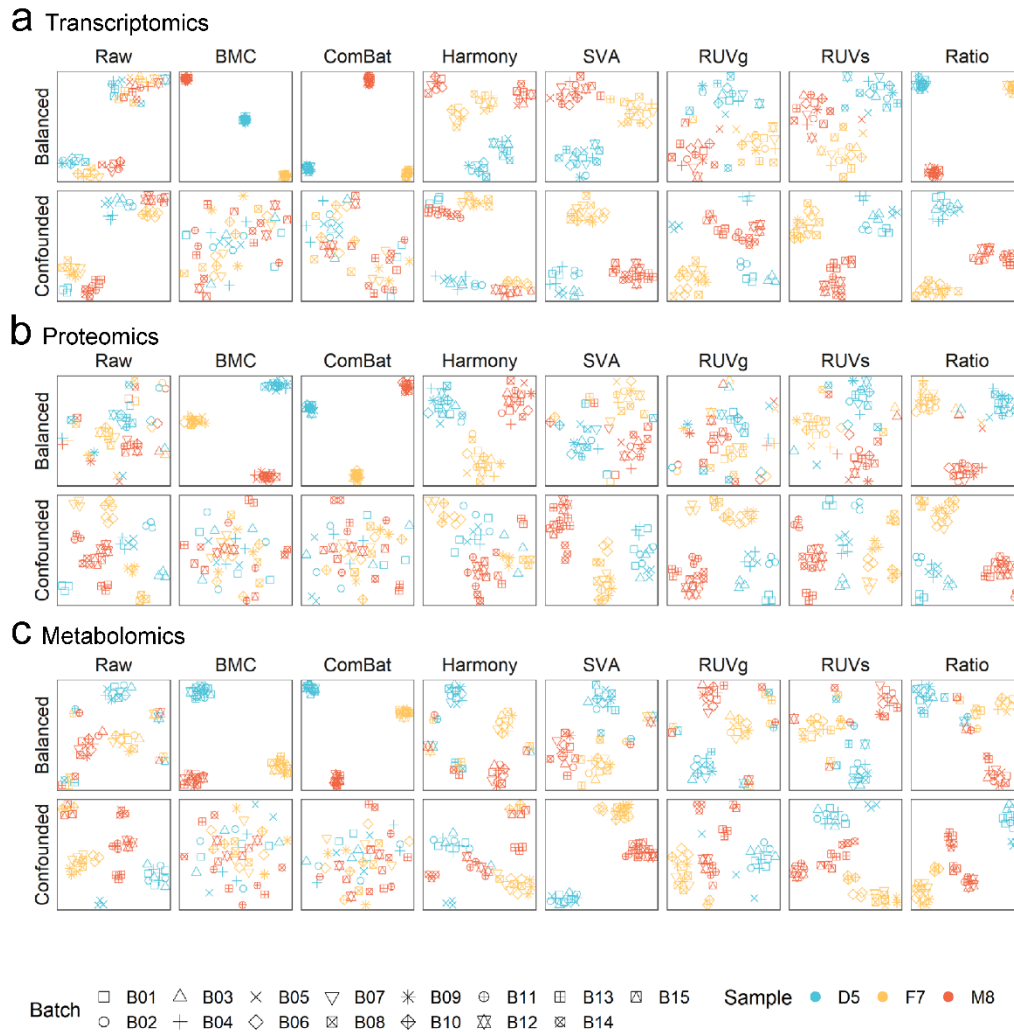
a Transcriptomics

b Proteomics

c Metabolomics

Batch □ B01 △ B03 ✕ B05 ▽ B07 ✳ B09 ⊕ B11 ⊞ B13 ▨ B15   Sample ● D5 ● F7 ● M8
○ B02 ＋ B04 ◇ B06 ⊠ B08 ⬦ B10 ▩ B12 ⊠ B14

Fig. S2 | tSNE plots based on different batch-effect correction methods.

tSNE plots based on different batch-effect correction algorithms (BECAs) in balanced and confounded scenarios, using (**a**) transcriptomics, (**b**) proteomics, and (**c**) metabolomics data. Plots were color-coded by donor (D5, F7, and M8) and shaped by batch.

	Transcriptomics	Proteomics	Metabolomics

Detected Features

- 15 batches of full datasets in each omics type

| 58,395 | 8,150 | 984 |

Detected features by consensus
- Criteria: FPKM ≥ 0.1 for transcriptomics, FOT ≥ 0.1 for proteomics, Intensity ≥ 1 for metabolomics
- Detected in both balanced and confounded scenarios
- ≥ 5 batches for transcriptomics and proteomics; ≥ 11 batches for metabolomics

| 26,907 | 3,489 | 71 |

▽

Reference FC

Detected in both groups in each sample pair

Keep features
- $p < 0.05$ in ≥ 3/15 batches

Reference FC
- In each batch, first average three replicates and then compute FC
- Average of FC across 15 batches
- Take geometrical mean as average

	Transcriptomics	Proteomics	Metabolomics
F7/D5	15,118	971	62
M8/D5	15,922	1,136	58
M8/F7	13,678	928	52

▽

Reference DEFs & non-DEFs

- DEFs: in each batch, $p < 0.05$ & FC criteria;
- Non-DEFs: the remaining detected features

FC criteria of identifying DEF

Type	Transcript-omics	Prote-omics	Metabol-omics
Up-regulated DEFs	> 2	> 2	> 1.5
Down-regulated DEFs	< 0.5	< 0.5	< 0.667
Non-DEFs	0.5~2	0.5~2	0.667~1.5

- Generate reference up-regulated DEFs, down-regulated DEFs and non-DEFs by voting rule*

*The choice with the most first-preference votes was the final classification of the feature. Features receiving more than one first-preference vote were not included in reference datasets.

	Transcriptomics	Proteomics	Metabolomics
F7/D5	1,610	39	13
	571	20	3
	24,709	3,427	54
M8/D5	1,735	30	13
	562	25	5
	24,589	3,433	52
M8/F7	1,136	28	4
	984	35	5
	24,772	3,421	61

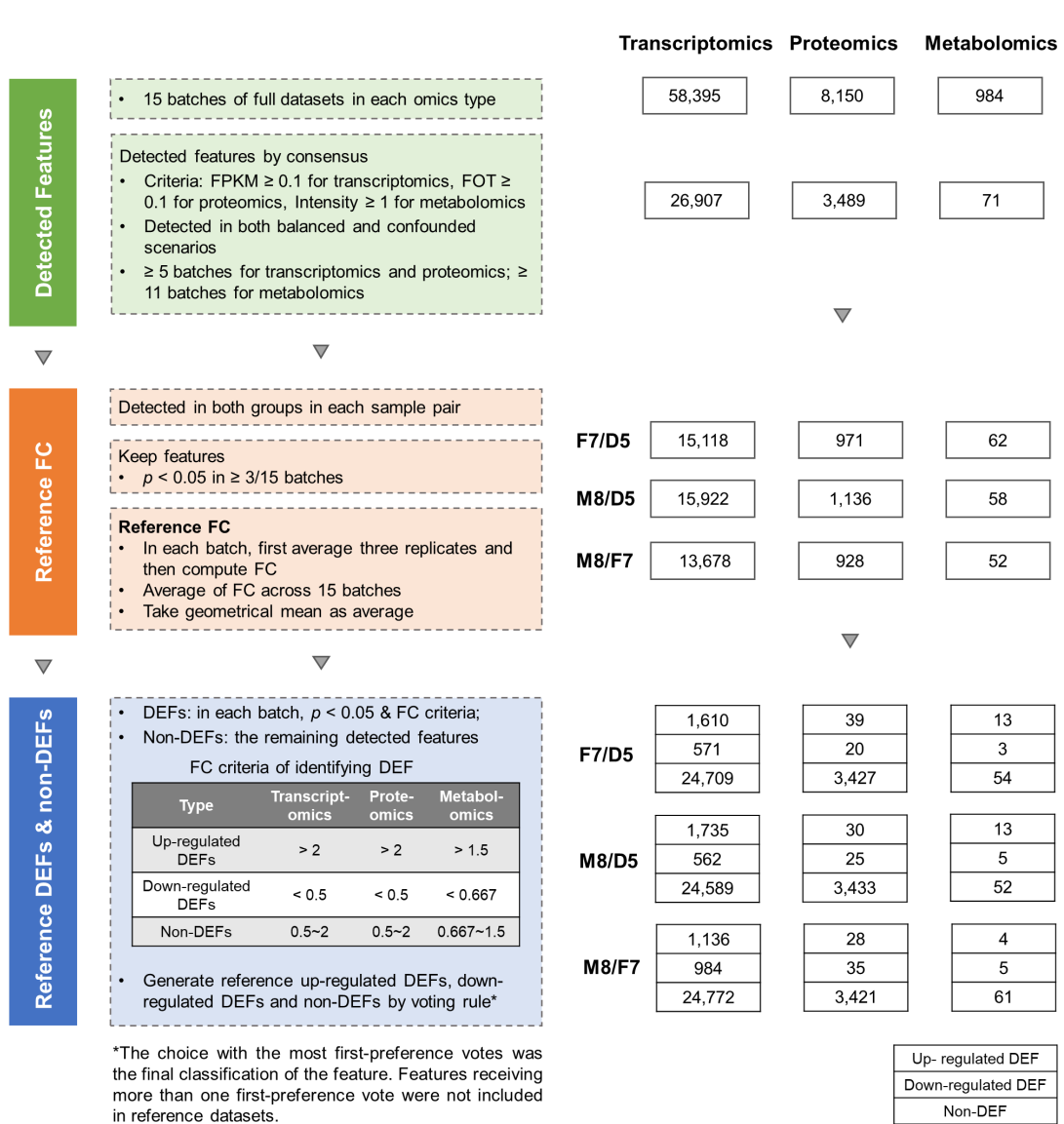| Up-regulated DEF |
| Down-regulated DEF |
| Non-DEF |

Fig. S3 | Workflow of construction of reference fold change and reference differentially expressed features.
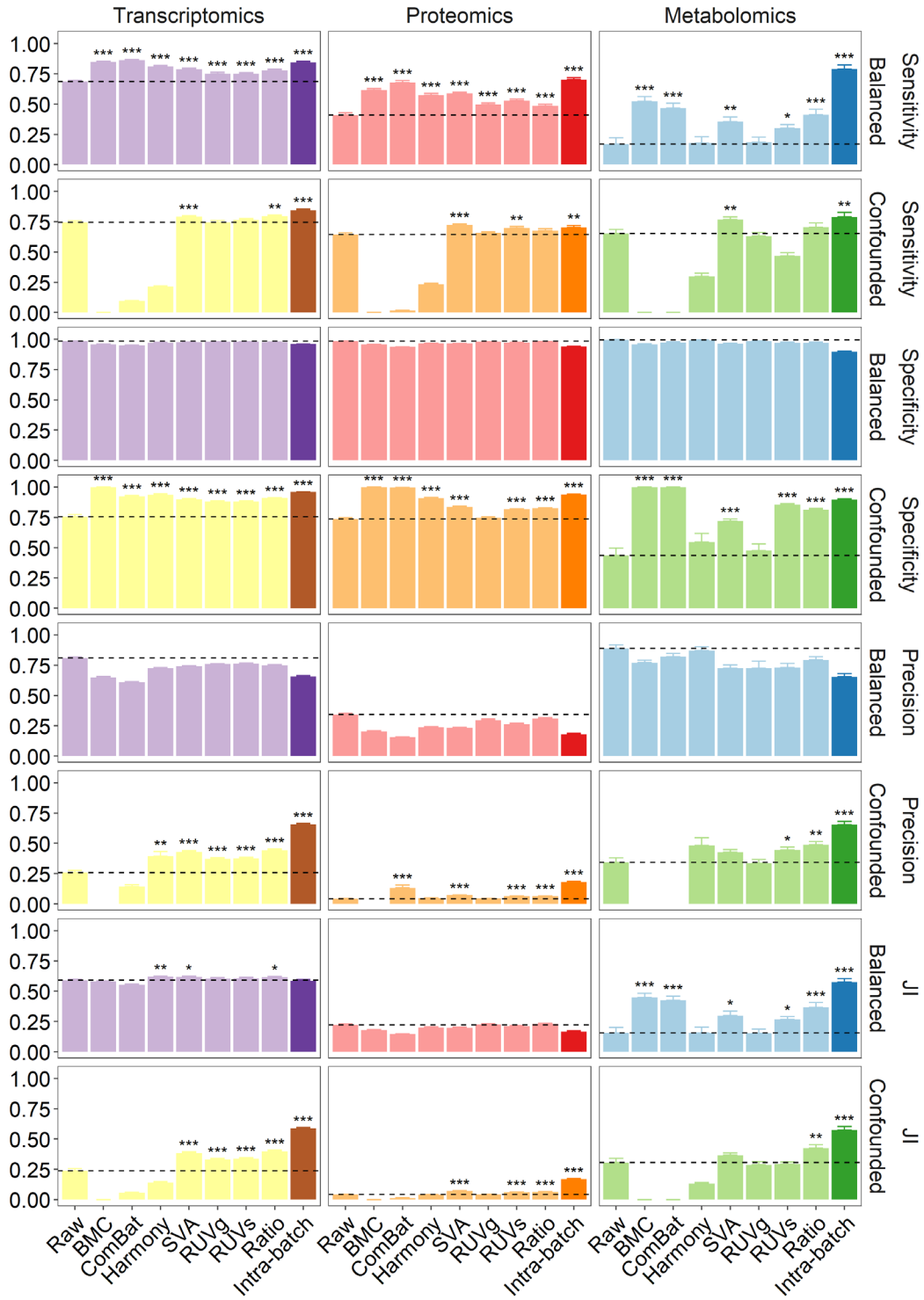
Fig. S4 | Evaluation of the performances of BECAs using sensitivity, specificity, precision, and Jaccard Index of identification of differentially expressed features.

Bar plots (mean ± s.e.) representing sensitivity, specificity, precision, and Jaccard Index (JI) of differentially expressed features with reference datasets across seven

BECAs were conducted in balanced and confounded scenarios using transcriptomics, proteomics, and metabolomics data. Metrics based on intra-batch DEFs were calculated based on full datasets and used as positive controls. Mean value of the dataset without correction (raw) in each panel was plotted in dashed line. Performances between the raw group and BECA groups were compared using Student's t-test. A group with the performance significantly higher than raw group was marked with stars (*). Symbolic number coding of p-value was used as: *** ($p \leq 0.001$), ** ($0.001 < p \leq 0.01$), * ($0.01 < p \leq 0.05$).
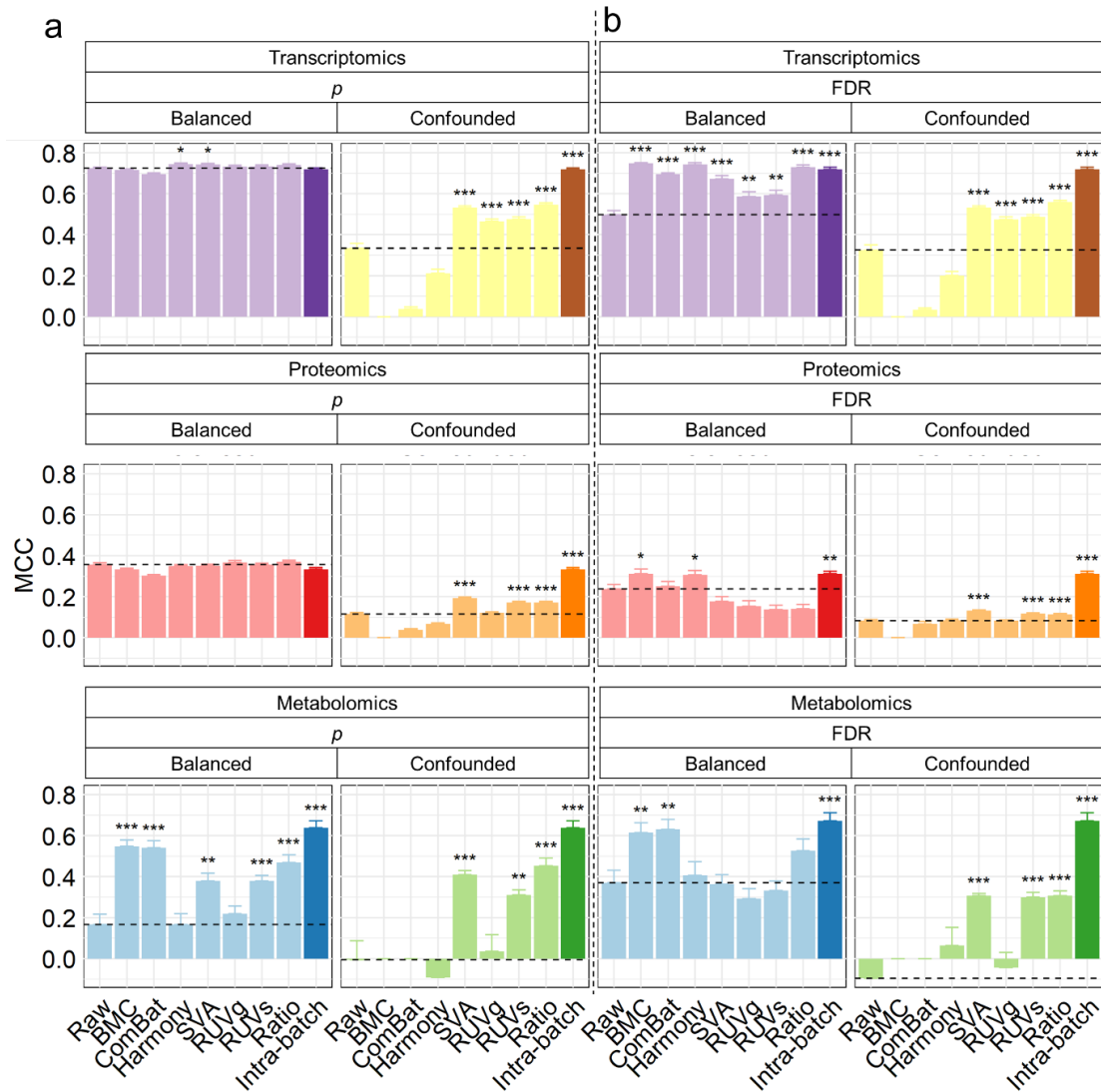
Fig. S5 | Matthews Correlation Coefficient of identification of differentially expressed features using with or without controlling multi-testing methods.

Cutoff of identification of differentially expressed features (DEFs) was set as: (a) $p < 0.05$ and FC > 2 or < 0.5 for transcriptomics and proteomics and > 1.5 or < 0.667 for metabolomics; (b) FDR < 0.05 and FC > 2 or < 0.5 for transcriptomics and proteo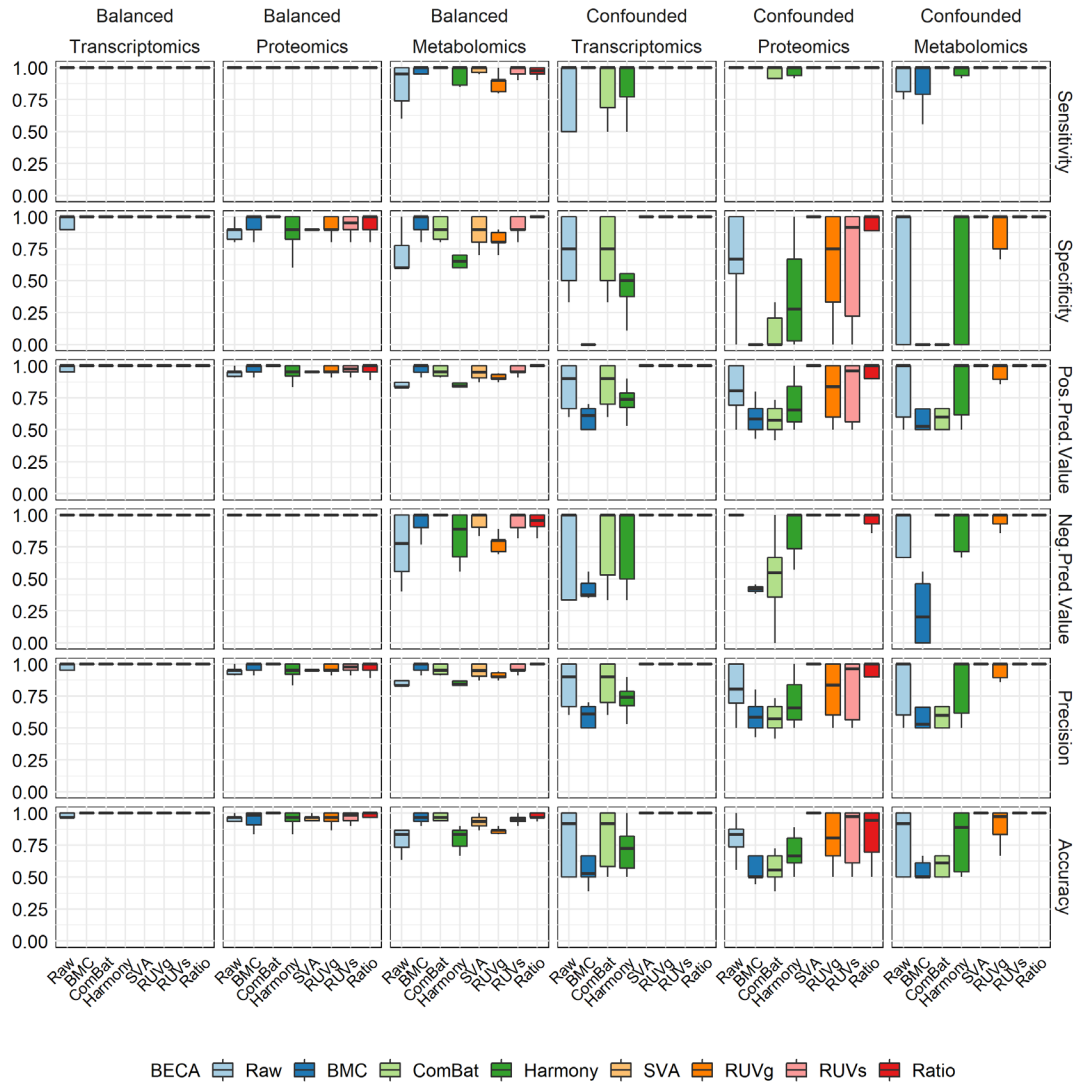mics and > 1.5 or < 0.667 for metabolomics. Matthews Correlation Coefficient (MCC) of DEFs with reference datasets were used as a metric.

Fig. S6 | Evaluation of the performances of BECAs based on model prediction results.

Validation performances in predicting sex and age using five machine-learning algorithms under balanced and confounded scenarios. Model performances were measured using sensitivity, specificity, Pos.Pred.Value (positive prediction value), Neg.Pred.Value (negative prediction value), precision, and accuracy.
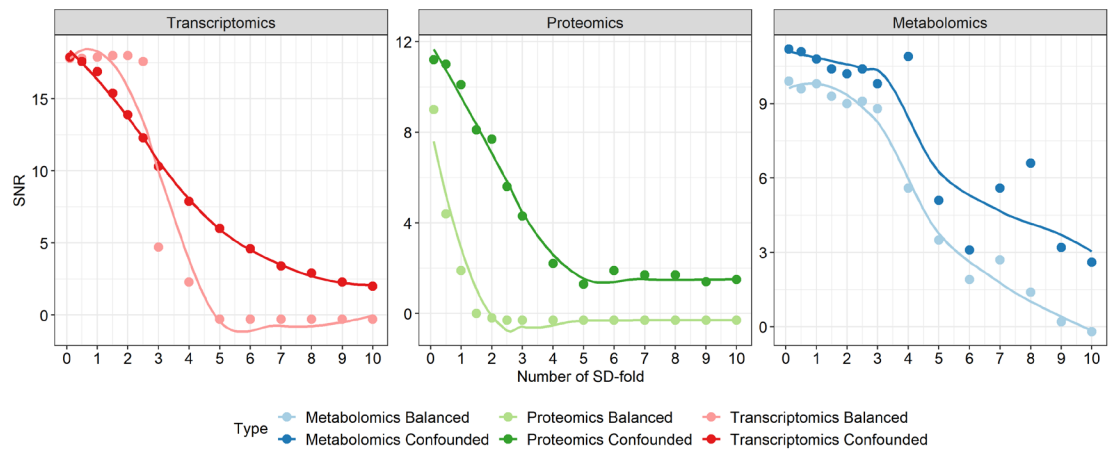
Fig. S7 | SNR of the ratio-based method under different data quality scenarios by introducing different levels of noise to the reference samples
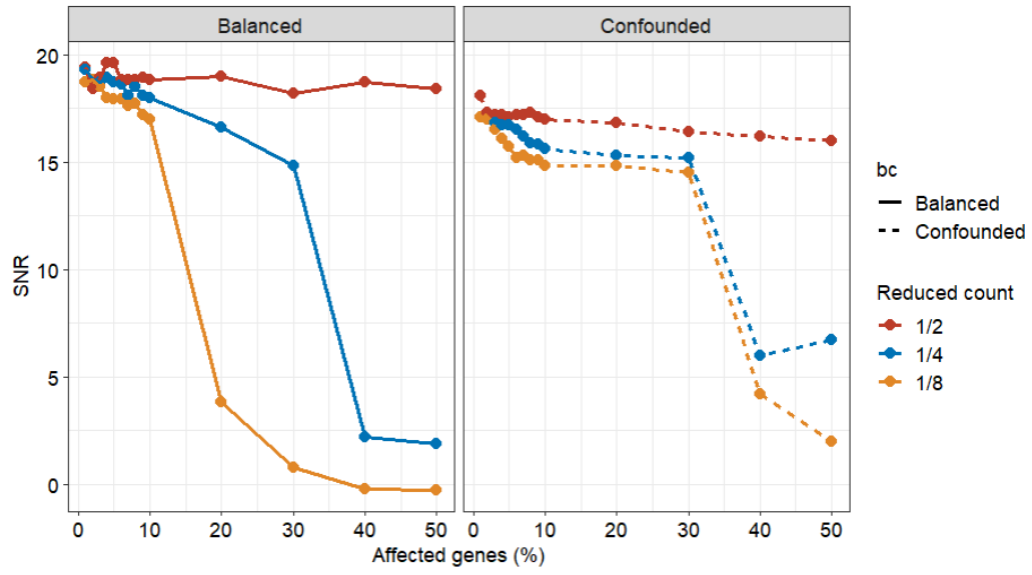
Fig. S8 | SNR of the ratio-based method under different data quality scenarios artificially reducing expression level of some genes in the reference samples