Bioinformatics, 2026, **42**(6), btag360
https://doi.org/10.1093/bioinformatics/btag360
Published: 4 June 2026
Applications Note | Bioimage informatics

OXFORD

Batch correction for large-scale mass spectrometry imaging experiments

Andreas A. Sparre and Ole N. Jensen*

Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, DK-5230, Denmark
*Corresponding author. Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, DK-5230, Denmark.
E-mail: jenseno@bmb.sdu.dk
Associate Editor: Virginie Uhlmann

Abstract

Summary: **We assess batch correction methods for MALDI mass spectrometry imaging experiments. ComBAT reduced batch-related technical variance, maintained biological variation, and improved the overall score by 19.4%.**

Availability and implementation: **Methods are available in R. comBAT is used through the "sva" package while Harmony, CCA, FastMNN are available in the "Seurat" package https://github.com/satijalab/seurat. scVI, scANVI are Scanorama are available in the Python programming language and through their github https://github.com/scverse/scvi-tools. Associated R code is found at DOI: https://doi.org/10.5281/zenodo.19730022.**

1 Introduction

Batch effects refer to systematic, non-biological and undesirable variations in data due to differences in experimental conditions across batches (Goh *et al*. 2022). Batch effects occur in all biological analysis regimen and are difficult to avoid, particularly in the context of large-scale omics data, e.g. single cell RNA sequencing (scRNA seq) data (Sprang *et al*. 2022). Mass spectrometry imaging (MSI) is an analytical technique that acquires spatially resolved mass spectra of biomolecules (e.g., peptides, metabolites) in a pixel-by-pixel manner by rastering across a specimen, for example, a thin tissue section. A pixel at position (*x, y*) in the specimen generates a mass spectrum where molecule ions are detected by their mass-to-charge ratio and intensity (m/z, I). The mass spectrum represents the molecular composition at position (*x, y*). Thus, the data structure in MSI is multidimensional with two spatial dimensions (*x, y*) and one spectral dimension (m/z, I) per pixel. MSI data can be considered large-scale data in a data frame that typically consists of >100.000 data points (pixels). Spatial autocorrelation is an inherent property of MSI data that, while not explicitly modeled by the batch correction algorithms, may influence the independence assumptions underlying some metrics. An experiment may cover 10's to 100's of samples and concomitant data acquisition cycles. Thus, batch effects and variations are inherent to MSI (Balluff *et al*. 2021) due to preanalytical factors including sample preparation protocols, laboratory conditions, data acquisition methods and operators. For example, accuracy of cryotome sectioning of tissue for MSI is sensitive to the tissue structure and texture, leading to variations of tissue thickness. Cumulative batch effects can mask the underlying biological information and may lead to irreproducible or inconclusive results.

We hypothesized that batch correction methods that are commonly used in the omics fields would be applicable to MSI experiments to reduce batch effects while retaining biological variation. We tested the batch correction methods comBAT, Harmony, Seurat canonical correction analysis (CCA), and fast mutual nearest neighbor (fastMNN) on a large-scale MSI dataset obtained from 128 renal cell carcinoma patients across 26 defined batches in the tissue microarray (TMA) format (Thomsen *et al*. 2026). We demonstrate successful batch correction using uniform manifold approximation and projection (UMAP) and metrics that scored successful batch integration and conservation of biological variation.

2 Methods

The clinical samples and MSI experimental setup were described previously (Thomsen *et al*. 2026; DOI: 10.64898/2026.01.13.26343935). MSI experiments were designed to profile trypsin-generated peptides from proteins in the tumor tissue sections. Briefly, the dataset consists of MSI data obtained from 541 needle cores (2- or 3-mm diameter) from 128 renal cell carcinoma (RCC) patients and controls. A total of 26 individual 3 µm tissue microarrays (TMAs, FFPE sections) were deposited onto IntelliSlides (Bruker Daltonics) to enable high-throughput MSI data acquisition from all 541 cores within 48 hours. Thus, the MSI data across 541 cores was acquired in 26 experiments, each

Received: 30 January 2026. **Revised:** 2 May 2026. **Accepted:** 18 May 2026

consisting of MSI measurements of up to 30 cores. Each RCC core consisted of at least 80% tumor. MSI data (peptide profiles, 10 ppm mass accuracy) of 541 cores (26 MSI experiments) were merged into one .imzml file in SCiLS lab v 2023a and loaded into R. The Cardinal package (Bemis *et al*. 2023) was used for data preprocessing, including total ion count (TIC) normalization, peak picking with signal-to-background ratio of 3, peak alignment with a tolerance of 0.5 Da, peak filtering, peak binning and deisotoping with a tolerance of 6 ppm, and a Pearson correlation coefficient threshold of 0.85. This resulted in a dataset consisting of 866 m/z values. Intensity data of the m/z values was then Log2-transformed and scaled (centering by mean subtraction from each observation following division by standard deviation) before batch correction. Batch correction methods included comBAT (Johnson *et al*. 2007), Harmony (Korsunsky *et al*. 2019), CCA (Butler *et al*. 2018) and fastMNN (Haghverdi *et al*. 2018). The latter three methods are integrated and available in the Seurat v5 environment(Hao *et al*. 2024), while comBAT is available through the "sva" R package.

Individual MSI pixels were treated as single observations analogous to single unit cells in transcriptomics data analysis. As a computational convention: each pixel $\{(x, y), (m/z, I)\}$ constitutes one observation without implying correspondence to a single biological cell. The underlying algorithm's assumptions concern input matrix structure rather than transcriptomic-specific features (e.g., sparsity) (Arevalo *et al*. 2024), supporting their applicability to MSI. Batches were defined at the tissue microarray section level, resulting in 26 batches, each corresponding to one MSI data acquisition cycle across one TMA section with multiple RCC cores and conditions. As a result, biological variation is partly nested within batches; while this complicates interpretation of batch mixing metrics, it reflects the practical reality of large-scale clinical MSI studies. Harmony, CCA and fastMNN work on dimensionally reduced data that required a principal component analysis (PCA) matrix as input for the correction tool, while comBAT works on the full data frame. The data was subjected to uniform manifold approximation and projection (UMAP) (McInnes and Healy 2018) with labelling of batches and conditions before and after correction. This served as an exploratory visualization rather than a quantitative comparison. Accordingly, we complement UMAP with quantitative metrics to assess the efficiency of the batch correction methods including adjusted rand index (ARI) (Steinley *et al*. 2016), normalized mutual information (NMI) (Nagel *et al*. 2024), average silhouette width (ASW) (Lengyel and Botta-Dukát 2019), PCA regression (Büttner *et al*. 2019) and Local Inverse Simpson's Index (LISI) (Korsunsky *et al*. 2019). Briefly, ARI and NMI assess biological variation by comparison of cluster assignments against ground truth. ASW considers cluster overlap and separation and depending on the use of labels (batch labels or condition labels) it assesses both data integration and biological variation. LISI scores diversity by measuring the effective number of unique labels within a local neighborhood. Depending on the label input LISI also assesses both batch mixing and biological variation. PCA regression works by linear regression of the batch variable onto each component, summarizing it for all components and assessing how well the batches are mixed. All metrics were scaled to the interval 0–1, where 1 indicates highest batch mixing or conservation of biological variation similarly to Luecken *et al*. (2022).

3 Results

We first benchmarked the algorithms by assessment metrics (Fig. 1a). Harmony scores highest in terms of conservation of biological variation with a score of 0.84, indicating superior preservation of biologically relevant structure. comBAT performs best for the batch mixing task with a score of 0.67. All batch correction methods reduced the batch effects and had higher batch mixing scores compared to uncorrected data. All tested methods retained the biological variation of the data with improvements in biological variation conservation scores. Overall, comBAT and CCA correction scored highest for performance (0.74) vs the uncorrected data (0.62) (Fig. 1a), i.e. an improvement of 19.4%.

Batch correction should result in a reduction of the variance across all batches. This is demonstrated by data shown in Fig. 1b where outlier batches 1, 17, 23 and 26 are shown before correction based on their median, density distribution and interquartile ranges (IQRs). After comBAT correction (Fig. 1c), the density distributions are similar for all batches together with the median and IQRs, which indicates that some data variation is corrected.

UMAP analysis of the experimental MSI dataset of 26 batches exhibits two major pixel populations or clusters (Fig. 1d). A small cluster contains MSI pixels originating from batches 17 and 26 and a larger cluster contains MSI pixels from the remaining 24 batches. The smaller cluster formed by batches 17 and 26 group the conditions independently from the larger cluster with the remaining batches, which indicates the variation is technical and not biological (Fig. 1e). Batch correction by comBAT and harmony result in one major homogenous cluster that encompasses all batches (Fig. 1f-i), and the conditions group based on biological- rather than technical variation (Fig. 1d). Additional UMAP visualizations for CCA and fastMNN are provided in Supp. 1, available as supplementary data at *Bioinformatics* online.

We speculated that batch effects originate from size-heterogeneity of the tryptic peptides, as in situ digestion kinetics are difficult to control. Short and long tryptic peptides would be more prone to this due to lower abundance compared to peptides in the optimal range of m/z 1000–2500. Indeed, the corrected extracted mass spectra exhibited reduced relative peak intensity of long (batch 1, 15 and 17) (Fig. 1j and l) and short peptides (batch 15 and 17) (Fig. 1k and l), indicating that they vary the most across experimental batches. The same effect is also visualized by examining ion images of m/z 1198.71 matching the AVFPSIVGRPR peptide sequence originating from actin (Høiem *et al*. 2022) and m/z 2676.2669. The actin peptide is more conserved across batches and thus is less affected compared to m/z 2676.2669, which is highly affected by comBAT correction (Supp. 2, available as supplementary data at *Bioinformatics* online.).

4 Discussion

We show that batch correction methods commonly used in the omics fields are applicable to large-scale mass spectrometry imaging experiments. All the tested batch correction methods

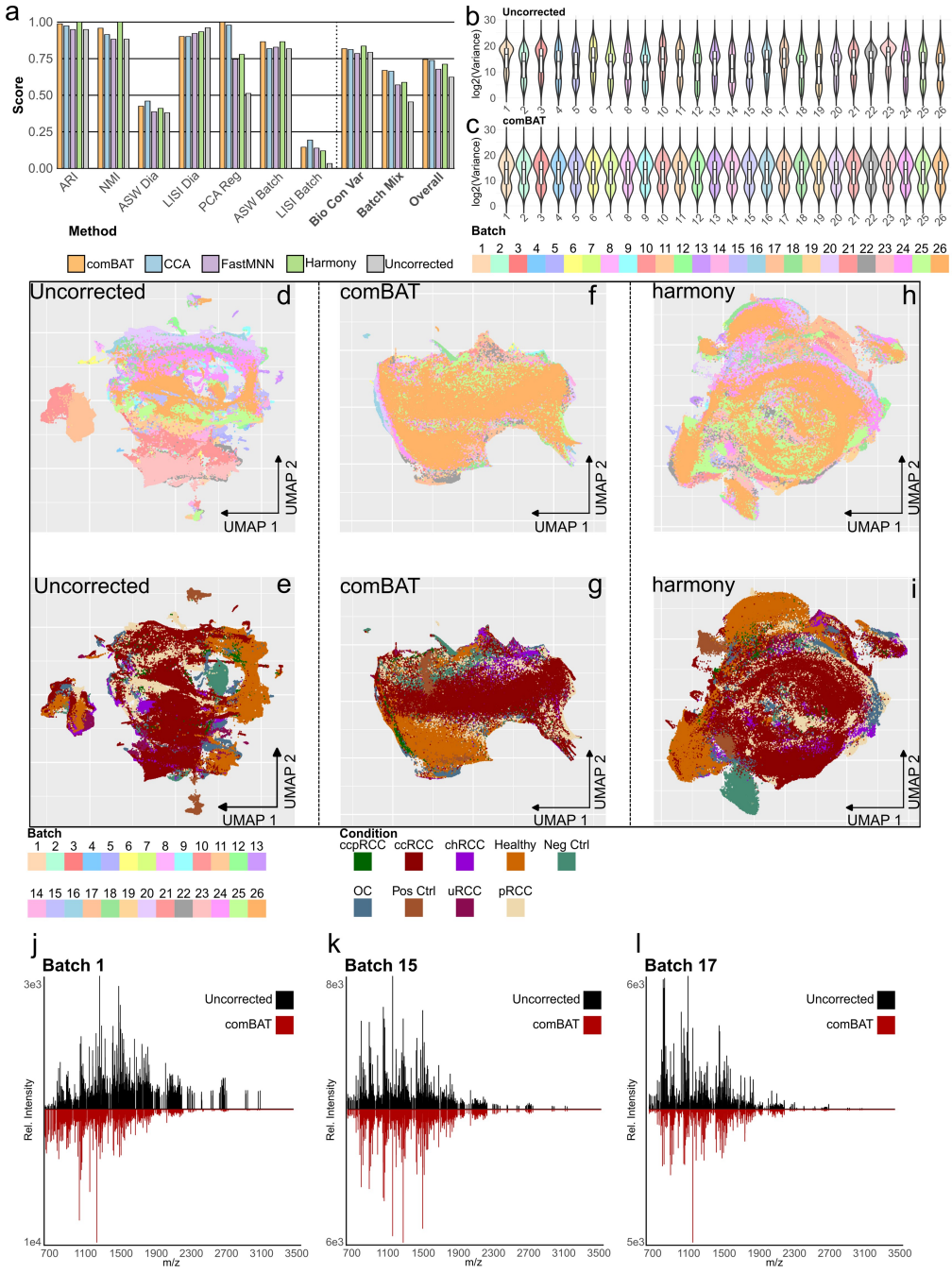

Figure 1 Assessment of batch correction methods for mass spectrometry imaging data. (a) Assessment of four batch correction methods: comBAT, CCA, FastMNN, Harmony versus uncorrected data. Barplots of metrics used to assess batch mixing and biological variation conservation. ASW Dia, LISI Dia, ARI and NMI assess conservation of biological variation using predefined sample labels. ASW Batch, LISI Batch and PCA reg assess batch mixing using predefined batch labels. Batch Mix is the mean of batch mixing metrics: ASW Batch, LISI Batch and PCA reg. Bio Con Var is the mean of biological variation conservation metrics: ASW Dia, LISI Dia, ARI and NMI. Overall is the mean of all the metrics. *x*-axis is the methods and *y*-axis is the score values are in the range of 0—1, with 1 indicating better batch mixing or bio variation conservation. (b) Violin plots of data variance (log2) of uncorrected MSI data for the 26 batches. *x*-axis shows the batch number and y-axis depicts log2-scaled variance values. (c) Violin plots of data variance (log2) of comBAT corrected MSI data for the 26 batches. *x*-axis shows the batch number and *y*-axis depicts log2-scaled variance values. (d) UMAP of uncorrected data, including all pixels in the dataset colored by batch number of origin. (e) UMAP of uncorrected data, including all pixels in the dataset colored by biological origin. (f) UMAP of comBAT corrected data, including all pixels in the dataset colored by their batch number of origin. (g) UMAP of comBAT corrected data, including all pixels in the dataset colored by the biological origin. (h) UMAP of harmony corrected data, including all pixels in the dataset colored by their batch number of origin. (i) UMAP of harmony corrected data, including all pixels in the dataset colored by the biological origin. (j) Extracted mass spectrum before (black) and after comBAT correction (red) for batch 1. *x*-axis depicts m/z and *y*-axis shows relative intensity. (k) Extracted mass spectrum before (black) and after comBAT correction (red) for batch 15. *x*-axis depicts m/z and *y*-axis shows relative intensity. (l) Extracted mass spectrum before (black) and after comBAT correction (red) for batch 17. *x*-axis depicts m/z and *y*-axis shows relative intensity.

reduced the batch effects of the MSI dataset (Fig. 1a). Harmony achieved the highest conservation of biological variation (0.84), which may be particularly advantageous in studies where preserving biological signal is the primary objective. comBAT and CCA achieved the highest overall benchmark scores (0.74) while comBAT performed best on the batch mixing task (Fig. 1a), motivating its use for detailed characterization in this study, with an overall improvement of 19.4% vs uncorrected data. Upon comBAT correction the total variance across all features is evenly distributed across the batches as compared to the uncorrected data, which contained outliers (Fig. 1b).

UMAP analysis showed clear improvements of data homogeneity upon batch correction by comBAT and Harmony (Fig. 1f and i). Notably, the renal cell cancer diagnosis sub-groups clustered according to their phenotype. This becomes clear as chRCC (malign) and OC (benign) cluster together and they are often pathologically indistinguishable due to similar cell origins.

comBAT works on the intensity/counts data, which can be beneficial in terms of downstream analysis since single feature are retained after correction. The other three methods are reduced by PCA before correction and therefore makes single feature analysis challenging (e.g., identifying discriminative peaks). In the case of unsupervised data analysis, e.g. UMAP or hierarchical clustering, the tested batch correction methods are valid because the output does not rely on individual features.

We found that comBAT correction mainly affects the MSI mass range above m/z 2500 and below m/z 1000 (Fig. 1j and l, Supp. 2, available as supplementary data at *Bioinformatics* online.), indicating that ion signals in the range m/z 1000–2500 are reliable and exhibit low variation. This agrees with the fact that most tryptic peptide masses fall in the range of m/z 1000–2500.

In addition to the four batch correction methods included here, several autoencoders for single cell omics data analysis are available in the Python programming language, such as scVI (Lopez *et al*. 2018), scANVI (Xu *et al*. 2021) or Scanorama (Hie *et al*. 2024). They may find use in MSI data analysis by using the MSI data formatting and encoding presented in our study. However, MSI data has a dense data structure that may be incompatible with the fact that existing autoencoders performs best for sparse data. Autoencoder based correction may therefore require adaption for MSI data. A recent study from Farrow *et al.* showed that reComBat (Adamer *et al*. 2022) was able to remove sample-specific variation for lipid MSI data, highlighting the use of batch correction tools in the context of dense MSI data (Farrow *et al*. 2025).

All the included methods are available through R. Additional methods exist in Python code including neural networks that perform well for scRNA seq data.

5 Conclusion

We introduce an MSI data structure that allows batch correction of large-scale mass spectrometry imaging data by methods adapted from the single cell omics fields. The underlying biological variation is conserved after batch correction with a score improvement of up to 19.4%. Batch correction allows for more robust and more accurate large-scale MSI studies of sample cohorts and is an important step towards wider applications of MSI in pharmacology and clinical research, including pathology.

Acknowledgements

AAS was financed by a PhD fellowship from University of Southern Denmark (SDU), Department of Biochemistry and Molecular Biology. The authors thank Maj Rabjerg, MD and Niels Marcussen, MD, OUH Dept. Pathology for providing samples. AAS thanks DANEMO, the Danish EMBL node, for financial support of a short-term visit to Theodore Alexandrov and his research group at the European Molecular Biology Laboratory, Heidelberg, Germany, to discuss data processing in MSI.

Author contributions

Andreas A Thomsen(Conceptualization [Equal], Formal analysis [Equal], Investigation [Equal], Visualization [Lead], Writing—original draft [Equal], Writing—review & editing [Equal]), and Ole Jensen(Conceptualization [Equal], Formal analysis [Equal], Funding acquisition [Lead], Project administration [Lead], Resources [Lead], Supervision [Lead], Writing—original draft [Equal], Writing—review & editing [Equal])

Supplementary material

Supplementary material is available at *Bioinformatics* online.

Conflicts of interest

None declared.

Funding

Mass spectrometry imaging research at SDU is supported by a generous grant to the INTEGRA research infrastructure [Novo Nordisk Foundation, grant no. NNF20OC0061575 to O.N.J].

References

Adamer MF, Brüningk SC, Tejada-Arranz A *et al.* reComBat: batch-effect removal in large-scale multi-source gene-expression data integration. *Bioinform Adv* 2022;**2**:vbac071. https://doi.org/10.1093/bioadv/vbac071

Arevalo J, Su E, van Dijk R *et al.* 2024. Evaluating batch correction methods for image-based cell profiling. *bioRxiv*. https://doi.org/10.1101/2023.09.15.558001

Balluff B, Hopf C, Porta Siegel T *et al.* Batch effects in MALDI mass spectrometry imaging. *J Am Soc Mass Spectrom* 2021;**32**:628–35. https://doi.org/10.1021/jasms.0c00393

Bemis KA, Föll MC, Guo D *et al.* 2023. Cardinal v3 - a versatile open source software for mass spectrometry imaging analysis. *bioRxiv*. https://doi.org/10.1101/2023.02.20.529280.

Butler A, Hoffman P, Smibert P *et al.* Integrating single-cell transcriptomic data across different conditions, technologies, and

species. *Nat Biotechnol* 2018;**36**:411–20. https://doi.org/10.1038/nbt.4096

Büttner M, Miao Z, Wolf FA *et al.* A test metric for assessing single-cell RNA-seq batch correction. *Nat Methods* 2019;**16**:43–9. https://doi.org/10.1038/s41592-018-0254-1

Farrow MA, Tideman LEM, Neumann EK *et al.* A lipid atlas of the human kidney. *Sci Adv* 2025;**11**:eadu3730. https://doi.org/doi : 10.1126/sciadv.adu3730

Goh WWB, Yong CH, Wong L. Are batch effects still relevant in the age of big data? *Trends Biotechnol* 2022;**40**:1029–40. https://doi.org/10.1016/j.tibtech.2022.02.005

Haghverdi L, Lun ATL, Morgan MD *et al.* Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* 2018;**36**:421–7. https://doi.org/10.1038/nbt.4091

Hao Y, Stuart T, Kowalski MH *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* 2024;**42**:293–304. https://doi.org/10.1038/s41587-023-01767-y

Hie BL, Kim S, Rando TA *et al.* Scanorama: integrating large and diverse single-cell transcriptomic datasets. *Nat Protoc* 2024;**19**:2283–97. https://doi.org/10.1038/s41596-024-00991-3

Høiem TS, Andersen MK, Martin-Lorenzo M *et al.* An optimized MALDI MSI protocol for spatial detection of tryptic peptides in fresh frozen prostate tissue. *Proteomics* 2022;**22**:e2100223. https://doi.org/10.1002/pmic.202100223

Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;**8**:118–27. https://doi.org/10.1093/biostatistics/kxj037

Korsunsky I, Millard N, Fan J *et al.* Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods* 2019;**16**:1289–96. https://doi.org/10.1038/s41592-019-0619-0

Lengyel A, Botta-Dukát Z. Silhouette width using generalized mean-a flexible method for assessing clustering efficiency. *Ecol Evol* 2019;**9**:13231–43. https://doi.org/10.1002/ece3.5774

Lopez R, Regier J, Cole MB *et al.* Deep generative modeling for single-cell transcriptomics. *Nat Methods* 2018;**15**:1053–8. https://doi.org/10.1038/s41592-018-0229-2

Luecken MD, Büttner M, Chaichoompu K *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods* 2022;**19**:41–50. https://doi.org/10.1038/s41592-021-01336-8

McInnes L, Healy J. 2018. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. https://doi.org/10.48550/arXiv.1802.03426, preprint: not peer reviewed.

Nagel D, Diez G, Stock G. Accurate estimation of the normalized mutual information of multidimensional data. *J Chem Phys* 2024;**161**:054108. https://doi.org/10.1063/5.0217960

Sprang M, Andrade-Navarro MA, Fontaine J-F. Batch effect detection and correction in RNA-seq data using machine-learning-based automated assessment of quality. *BMC Bioinformatics* 2022;**23**:1–14. https://doi.org/10.1186/s12859-022-04775-y

Steinley D, Brusco MJ, Hubert L. The variance of the adjusted rand index. *Psychol Methods* 2016;**21**:261–72. https://doi.org/10.1037/met0000049

Thomsen AA, Rabjerg M, Marcussen N *et al.* 2026. Mass spectrometry and machine learning for classification and molecular phenotyping of renal cell carcinoma and benign tumors. medRxiv, 2026.2001.2013.26343935. https://doi.org/10.64898/2026.01.13.26343935.

Xu C, Lopez R, Mehlman E *et al.* Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol Syst Biol* 2021;**17**:e9620. https://doi.org/10.15252/msb.20209620